



Demand–supply equilibrium of taxi services in a network under competition and regulation

Hai Yang ^a, S.C. Wong ^{b,*}, K.I. Wong ^b

^a *Department of Civil Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, People's Republic of China*

^b *Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, People's Republic of China*

Received 5 September 2000; received in revised form 14 May 2001; accepted 11 July 2001

Abstract

This paper investigates the nature of demand–supply equilibrium in a regulated market for taxi service. Distinguished from conventional economic analysis, a network model is used to describe the demand and supply equilibrium of taxi services under fare structure and fleet size regulation in an either competitive or monopoly market. The spatial structure of the market such as the form of road network and the customer origin–destination demand pattern are explicitly considered. The model can determine a number of system performance measures at equilibrium such as utilization rate for taxi and level of service quality, and predict the effects of alternative regulations on system performance. The model can thus be used as a policy tool by the regulator to ascertain appropriate taxi regulations such as the selection of taxi fleet size and fare structure. A case study in Hong Kong was conducted to illustrate some interesting findings. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Taxi; Network model; Demand–supply equilibrium; Competition; Regulation

1. Introduction

In most large cities the taxi industry is subject to various types of regulation such as entry restriction and price control, and many economists have examined the economic consequences of regulatory restraints in different ways (Douglas, 1972; De vany, 1975; Shrieber, 1975; Manski and Wright, 1976; Foerster and Gorman, 1979; Beesley and Glaister, 1983; Schroeter, 1983; Frankena

* Corresponding author. Tel.: +852-2859-1964; fax: +852-2559-5337.
E-mail address: hhecwsc@hkucc.hku.hk (S.C. Wong).

and Pautler, 1986; Cairns and Liston-Heyes, 1996; Arnott, 1996). The general objective of those studies is to understand the manner in which the demand and supply are equilibrated in the presence of such regulations, thereby providing information for government decision-making in terms of such regulations (Beesley and Glaister, 1983). The general analytical framework on the economics of taxi service can be described below:

$$D = D(F, W), \quad \frac{\partial D}{\partial F} < 0, \quad \frac{\partial D}{\partial W} < 0, \quad (1)$$

$$W = W(U), \quad \frac{\partial W}{\partial U} < 0, \quad (2)$$

$$TC = c(Q + U), \quad (3)$$

where D is the demand for taxi ride (number of customers per hour), Q occupied taxi-hours (equal to the product of average ride time and demand D), U vacant taxi-hours, F the expected fare or money price of one-hour (occupied) taxi service, W expected waiting time, TC total costs, and c is the cost per taxi hour of service time. Eq. (1) states that the demand for taxi rides is a decreasing function of the expected fare and the expected waiting time (a proxy of service quality); Eq. (2) states that expected waiting time is inversely proportional to the total vacant taxi-hours; and Eq. (3) states that the cost of operating a taxi is a constant per hour.¹

This highly aggregate model was originally proposed by Douglas (1972) and has been adopted by subsequent studies on the economics of taxis, without consideration of the spatial structure of the market. It is commonly realized that there are two principal characteristics that distinguish the taxi market from the idealized market of conventional economic analyses: the role of customer waiting time and the complex intervening relationship between users (customers) and suppliers (firms) of the taxi service.

In the taxi market, the equilibrium quantity of service supplied (total taxi-hours) will be greater than the equilibrium quantity demanded (occupied taxi-hours) by a certain amount of slack (vacant taxi-hours). It is this amount of slack that governs the average customer waiting time. The expected customer 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 (De vany, 1975; Abe and Brush, 1976; Manski and Wright, 1976; Foerster and Gorman, 1979). A reduction in expected waiting time increases the demand for taxi service. However, from the point of view of each individual taxi firm, expected customer waiting time is different from the quality of the typical product. In most markets where quality is a variable, each firm can decide what quality to produce. In the taxi market, expected waiting time is usually not amenable to differentiation, but depends on the total number of vacant taxi-hours. An individual firm cannot offer customers an expected waiting time different from that offered by other firms, although a large firm may be able to affect expected waiting time (Frankena and Pautler, 1986).

¹ The additive form $(Q + U)$ in the cost function (3) is rather restrictive. One can differentiate operating costs of occupied and vacant taxi-hours, and even introduce a km-dependent operating cost.

Furthermore, the demand for and supply of taxi services are interrelated through two intervening variables: taxi availability (as measured by expected customer waiting time) and taxi utilization (as measured by expected fraction of time a taxi is occupied). On the demand side, potential customers will consider taxi availability as well as fare in making their mode-choice decisions. From the supply perspective, taxi firms will operate in response to taxi utilization rate as well as trip revenues and costs. Moreover, taxi availability, through its influence on the level of taxi use, indirectly affects the taxi utilization rate; the utilization rate, through its influence on the level of supply, in turn affects taxi availability. This demand–availability–utilization–supply relation is shown in Fig. 1 (Manski and Wright, 1976). A non-linear simultaneous equation system and a neural network model have been developed recently by Xu et al. (1999) and Yang et al. (1999) to model these complex relationships at an aggregate macroscopic level based on Hong Kong taxi service situation.

The aforementioned analytical approaches are based on abstract, simplified demand–supply models and might be useful in understanding the way taxi market operates. However, demand for and supply of taxi services take place over space in reality, and thus equilibrium modeling of the problem should be conducted in connection with a detailed road network structure and a customer origin–destination (O–D) demand pattern as done in a conventional network traffic assignment. Consideration of the spatial structure of the taxi market is particularly meaningful and essential for precise understanding and planning of urban taxi services.

This study is a continuous development in the line of the earlier work by Yang and Wong (1998) and Wong and Yang (1998) and Wong et al. (2001) that models explicitly the spatial structure of the market in terms of a road network and a customer O–D demand pattern. Yang and Wong (1998) and Wong and Yang (1998) proposed a network equilibrium model and solution algorithm of taxi movements on a network with given customer O–D demand. Wong et al. (2001) further incorporated demand elasticity and congestion effect (due to both taxi and normal

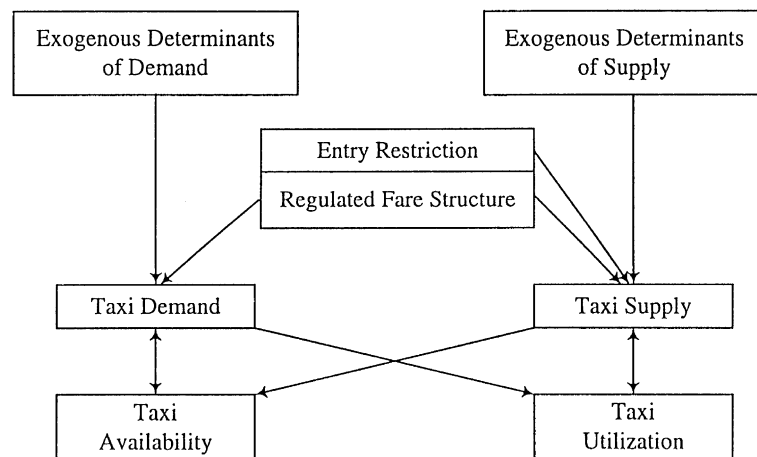


Fig. 1. The demand–availability–utilization–supply relation in a taxi market. Reproduced by permission, Manski and Wright (1976). C.F. Manski and J.D. Wright, Nature of equilibrium in the market for taxi services, *Transportation Research Record* 619, 11–15. © 1976, Transportation Research Board, National Academy of Sciences, 2101 Constitution Avenue, N.W., Washington, DC 20418, USA.

traffic) in the network model of taxi services. The demand for taxi rides is assumed to be a function of a set of exogenous variables, including fares, and of the endogenous variable, taxi availability. The profitability of taxi operations is determined by exogenous factors and by an endogenous variable, taxi utilization. The number of taxis supplied is determined by the profit function and by the industrial structure. Given the network, number of taxis operating and customer O–D demand, the movements of both vacant and occupied taxis together with the taxi availability and utilization are determined by a network equilibrium model (Wong and Yang, 1998; Yang and Wong, 1998). The demand–supply equilibrium of the market is then established until both demand and supply equations are satisfied simultaneously. An efficient algorithm for the solution of demand–supply equilibrium was developed in Wong et al. (2001).

This paper concentrates on the impacts of alternative regulatory restraints on the market equilibrium by investigating the social surplus, firm profit, customer demand at various levels of taxi fare and fleet in the regulated, competitive and monopoly markets. An intuitive graphical characterization of the various market equilibria is developed, based on a real case study for Hong Kong. In Section 2, a network equilibrium model of taxi movements and customer demand is presented. In Section 3, the demand and supply equilibrium of taxi services in either competitive or monopoly markets are modeled. In Section 4, a case study for the city of Hong Kong is presented and the model is applied to examine the impacts of alternative taxi regulations such as price and entry control on system performances. General conclusions are given in Section 5.

2. Taxi movement model

Consider a road network $G(V, A)$, where V is the set of nodes and A is the set of links in the network. In any given hour, the number of customers demanding taxi ride from origin zone i to destination zone j is D_{ij} . Let I and J be, respectively, the sets of customer origin and destination zones, and $O_i = \sum_{j \in J} D_{ij}$ and $D_j = \sum_{i \in I} D_{ij}$ be, respectively, the customer trip ends at $i \in I \subset V$, $j \in J \subset V$. Let h_a be the travel time on link $a \in A$, and h_{ij} be the travel time via the shortest path from origin $i \in I$ to destination $j \in J$, both times are measured as a fraction of an hour. For simplicity, h_a and h_{ij} are assumed to be constant (traffic congestion is not considered here).²

Suppose that there are N cruising taxis operating in the network. Given a reasonably large number of taxis, there will be no differentiation of individual taxis. It is assumed that all occupied taxis will follow an “all-or-nothing” routing behavior, which means that a taxi, once occupied by a customer at origin $i \in I$, will move to the customer’s destination $j \in J$ via the shortest path. Once a ride is completed, the taxi becomes vacant and will either cruise at the same zone or move to other zones to seek for next customer. In doing this, each taxi driver is assumed to attempt to minimize his/her expected search time required to meet next customer.

² As to be mentioned later, for the case study of the Hong Kong taxi service situation, taxi travel times between traffic zones were determined by means of another network model using EMM/2 package (INRO, 1997) and treated as constants. This assumption will greatly facilitate our analysis of the monopoly and competitive equilibria under entry and price regulations. For modeling the interaction and congestion effect between normal traffic and taxi movements in a road network, we refer the reader to Wong et al. (2001).

2.1. Taxi service time constraint

Consider a unit period (1 h) taxi operations in the network with given customer demand. The total occupied time (Q) of all taxis is the taxi-hours required to complete all $T_{ij}^o = D_{ij}$, $i \in I$, $j \in J$ trips, and is thus given by

$$Q = \sum_{i \in I} \sum_{j \in J} T_{ij}^o h_{ij}, \quad (4)$$

where T_{ij}^o is the occupied taxi movements (veh/h) from zone $i \in I$ to $j \in J$. On the other hand, the total vacant time (U) consists of moving times of vacant taxis from zone to zone and waiting (search) times at each origin zone. This time is given by

$$U = \sum_{j \in J} \sum_{i \in I} T_{ji}^v \{h_{ji} + w_i\}, \quad (5)$$

where T_{ji}^v are the vacant taxi movements (veh/h) from zone $j \in J$ to $i \in I$, w_i , $i \in I$ is the expected average taxi waiting time at zone $i \in I$.

For a given unit operating period, the sum of occupied taxi-hours and vacant taxi-hours must equal the total number of taxi-hours of service. Suppose the taxi operating period spans the demand period, then the quantity of taxi-hours of service is strictly equivalent to the number of taxis operating since an unit operation period of one hour is modeled. Thus we have the following taxi service time constraint:

$$\sum_{i \in I} \sum_{j \in J} T_{ij}^o h_{ij} + \sum_{j \in J} \sum_{i \in I} T_{ji}^v \{h_{ji} + w_i\} = N. \quad (6)$$

2.2. Behavior of taxi drivers

Once a customer service is completed at a destination zone $j \in J$, taxi driver could either stay at the same zone or move to other zones to find the next customer. It is assumed that each driver tries to minimize individual expected search time to meet a customer. The probability that a vacant taxi originating at zone $j \in J$ meets a customer eventually at zone $i \in I$ is determined by the expected search time in each zone, and specified by the following logit model:

$$P_{i/j} = \frac{\exp \{ -\theta(h_{ji} + w_i) \}}{\sum_{m \in I} \exp \{ -\theta(h_{jm} + w_m) \}} \quad \text{for any } i \in I, j \in J, \quad (7)$$

where $P_{j/j}$ when $i = j$ represents the probability of a taxi that takes a customer to zone j and then searches and meets next customer in the same zone, θ is a non-negative parameter that can be calibrated from observational data (Wong et al., 1999). The value of θ reflects the degree of uncertainty about customer demand and taxi services in the whole market from the perspective of individual taxi drivers.

In a stationary equilibrium state, the movements of vacant taxis over the network should meet the customer demands at all origin zones. Since there are D_j taxis to complete services at destination zone $j \in J$ per hour, we have

$$\sum_{j \in J} D_j \cdot P_{i/j} = O_i, \quad i \in I. \quad (8)$$

2.3. Demand assumptions

The information structure assumed is that each potential customer knows that there is some expected interval between wanting a taxi and getting one and uses this knowledge to form expected price. We consider separate demand function for each O–D pair (i, j) and assume that customer demand depends on customer waiting time and money price for taxi ride as given below:

$$D_{ij} = D_{ij}(W_i, F_{ij}), \quad i \in I, \quad j \in J, \quad (9)$$

where W_i is the expected customer waiting time at zone $i \in I$, F_{ij} , $i \in I$, $j \in J$, is the monetary cost that customer pays for a taxi ride from zone i to zone j , and D_{ij} is assumed to be monotonically decreasing with respect to both W_i and F_{ij} . To simplify our analysis, the fare of a taxi ride F_{ij} depends on the trip length (say, τ dollars per unit ride time):

$$F_{ij} = \tau h_{ij}, \quad i \in I, \quad j \in J. \quad (10)$$

Customer waiting time, an endogenous variable of the model, varies across zones. Within each zone the expected waiting time of customer can be described as a function of the density of vacant taxis in the subarea

$$W_i = W_i(n_i, w_i), \quad i \in I, \quad (11)$$

where n_i is the number of vacant taxis per hour that meet customers at zone $i \in I$ (note that $n_i = O_i$ at equilibrium), and w_i is the expected average (cruising and/or waiting) time that a vacant taxi spends in zone $i \in I$ in finding a customer. Note that specification of the customer waiting time function (11) depends on the distribution of taxi stands over individual zones. In the case of m_i identical taxi stands (in terms of arrival rates of both vacant taxis and customers) in zone $i \in I$, Eq. (11) can be specified as (see Appendix A):

$$W_i = \left\{ \frac{1}{w_i} - 1 \right\} \frac{m_i}{O_i} = \left\{ \frac{1}{w_i} - 1 \right\} \frac{m_i}{n_i}, \quad i \in I. \quad (12)$$

In the case of a continuous taxi stand distribution (taxi can pick up customer anywhere on the streets), we can assume that the vacant taxis move randomly through the street, and the expected average customer waiting time is proportional to the area of the zone and inversely proportional to the (cruising) vacant taxi hours:

$$W_i = \beta \frac{A_i}{n_i w_i}, \quad i \in I, \quad (13)$$

where A_i is the area of zone $i \in I$ and β is a common model parameter to all zones. This approximate distribution of waiting times can be derived theoretically (see Appendix B).

The above-mentioned demand–supply equilibrium model for taxi services can be formulated as a mathematical programming problem and solved by an efficient algorithm (Wong et al., 2001). In this paper, we focus on the impact of regulation on the market equilibrium, to be discussed in the following section.

3. Market equilibrium

Using the aforementioned model and assumptions, we now consider how to find the equilibrium output, capacity, and utilization of capacity in the taxi market under the following scenarios: free entry in a monopoly market with fixed price; free entry in a competitive market with fixed price; first-best social optimum and second-best social optimum.

3.1. Monopoly solution

We suppose that the cost of taxi operation is independent of the division of service time between “occupied” and “vacant” and let c be the operating and prorated capital cost per taxi per hour of service time. So the total taxi operating cost is given by $TC = cN$. Note that this assumption is not restrictive since our model can determine separately the magnitudes of occupied, vacant and waiting taxi-hours at equilibrium. It is straightforward to differentiate unit costs for different types of service time.

Assume the city government grants a single firm monopoly rights through a charter to pick up customers within a market area, and the government also sets the fare. Under this monopoly system, the single expected profit-maximizing firm would select a fleet size N^m to maximize:

$$\text{maximize } R(N^m) = \sum_{i \in I} \sum_{j \in J} F_{ij} D_{ij} - cN^m \quad (14)$$

subject to

$$F_{ij} = \tau^* h_{ij},$$

where τ^* is the regulated charge rate per unit service time, and N^m represents the monopoly solution of taxi fleet. The monopolist will operate at a solution where marginal revenue of taxis at the regulated price equals marginal cost of taxis (Douglas, 1972; De vany, 1975). If fare is unconstrained the monopolist will choose a fare–fleet combination that maximizes profits, the resultant solution depends on the price and level of service elasticities of customer demand.

3.2. Competitive solution

Assume that the market is comprised of owner-operated taxis, one taxi per owner or firm. The fare to be charged is set by the regulator, but entry is unrestricted. 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 (profits are nil). It is at this point that the individual incentive to join the taxi industry disappears. Hence equilibrium occurs at

$$\frac{\sum_{i \in I} \sum_{j \in J} F_{ij} D_{ij}}{N^c} = c, \quad (15)$$

where $F_{ij} = \tau^* h_{ij}$ and N^c represents the competitive solution of taxi fleet. Evidently, the number of taxis at the competitive free-entry market is a function of the price set by the regulator.

3.3. First-best social optimum

In the taxi market social surplus per hour is defined to be the sum of consumers' and producers' surplus. Assuming equal value of time for all customers, the first-best social optimum is derived by maximizing social surplus (SS) with respect to N^f :

$$\text{maximize } SS(N^f) = \sum_{i \in I} \sum_{j \in J} \int_0^{D_{ij}} F_{ij}(\omega, W_i(N^f)) d\omega - cN^f \quad (16)$$

subject to (4)–(11), where $F_{ij}(x, W_i(N^f))$ is the inverse of the demand function (9), and N^f represents the first-best solution of taxi fleet. Note that the consumers' surplus is obtained by integrating under a (hypothetical) demand curve in which the service level (waiting time) is held fixed while fare F varies, rather than under the market demand curve (Cairns and Liston-Heyes, 1996).

Note that at the social optimum, price or marginal willingness-to-pay for one-hour taxi service is set equal to the marginal cost per taxi-hour, and fleet size should be such that the marginal benefit from adding one additional taxi (which stems from reduced customer waiting time and thus induced customer demand) equals marginal cost (cost per taxi-hour). Therefore, the social optimum would generally generate an efficient but unfeasible (deficit) equilibrium in the sense that taxi revenues may just cover the cost of occupied taxi-hours so that, in the aggregate, taxi operation may makes a loss equal to the cost of vacant taxi-hours. This conclusion strictly holds for the abstract aggregate demand–supply model (1)–(3) (Douglas, 1972; Arnott, 1996), but may not be true for the network demand–supply equilibrium model considered here due to the complex spatial effects and specific model structure such as restricted fare function given in Eq. (10).

3.4. Second-best social optimum

Running a private industry requires restricting price and entry so that profits are non-negative. The subsequent second-best social optimum problem is that given the constraint that revenues cover costs (profits are nil), what would determine an optimal price and number of taxis supplied in this market? This second-best social optimum can be obtained by the following constrained welfare maximization problem:

$$\text{maximize } SS(\tau^s, N^s) = \sum_{i \in I} \sum_{j \in J} \int_0^{D_{ij}} F_{ij}(\omega, W_i(N^s)) d\omega - cN^s \quad (17)$$

subject to

$$\sum_{i \in I} \sum_{j \in J} F_{ij}^s D_{ij} = cN^s, \quad (18)$$

and (4)–(11), where $F_{ij}^s = \tau^s h_{ij}$, the index 's' means second-best solution. The problem becomes that social surplus (in this case social surplus equals consumer surplus since total taxi costs are held equal to total taxi revenues) is maximized subject to a zero-profit constraint and the network equilibrium model. Note that if all customers are assumed to have homogenous values of time and

a single demand function is adopted for a whole homogenous market, the problem is equivalent to the problem that maximizes total demand subject to a zero-profit constraint, and so the efficient price maximizes customer demand (Douglas, 1972; De vany, 1975). However, this may not be true in our network equilibrium model where origin and/or O–D pair-specific variables and constants appear in the separate demand functions between individual O–D pairs even though the equal customer time value is assumed.

4. A case study

4.1. Study assumptions

Hong Kong, with a land area of only 1091 km² of which approximately 15% are built-up areas, has a population of more than 6 millions people. Currently, a taxi fleet of less than 20,000 carries approximately 1.3 million customers per day. In the urban area of Hong Kong, taxis make considerable demands on limited road space even when empty and cruising for customers. Taxis currently form approximately 25% of the traffic stream overall in Hong Kong (Transport Department, 1989, 1993b). The operations of taxis are subject to various regulations on the ownership of taxis, taxi licensing system, and taxi fare policy (Hong Kong Transport Advisory Committee, 1992). In this paper, the urban taxi service in Hong Kong in 1992 is used as a case study to demonstrate the likely impacts of regulations on the market equilibrium by investigating the social surplus, firm profit, customer demand at various levels of taxi fare and fleet in an either competitive or monopoly market, when spatial effects are taken into considerations. The year 1992 was chosen for the case study because a comprehensive Traffic Characteristics Survey (TCS) (Transport Department, 1993a), covering a sample of 2% of all households in Hong Kong, was conducted in that year, which provides more complete information to support the study as compared with other later years.

Fig. 2 shows the urban area of the city of Hong Kong for the case study. The study area was divided into 15 districts, consisting of 274 zones. The taxi customer demand pattern was obtained from the TCS, whereas the taxi travel times between traffic zones were determined by means of a network model using EMME/2 package (INRO, 1997), which were calibrated also based on the TCS. For simplicity, the analysis was carried out at a district level. The demand function for all O–D pairs is assumed to take the form

$$D_{ij} = \tilde{D}_{ij} \exp \{ -\alpha(\tau h_{ij} + v_1 h_{ij} + v_2 W_i) \}, \quad i \in I, j \in J, \quad (19)$$

where \tilde{D}_{ij} and h_{ij} are, respectively, the potential customer demand and constant travel time from district $i \in I$ to district $j \in J$, v_1 and v_2 are the monetary values to the customer of unit in-vehicle and waiting time, α is a scaling parameter which indicates the sensitivity of demand to full trip price. From TCS, the monetary values v_1 and v_2 were, respectively, found to be HK\$35 and HK\$50 per hour (with US\$1 \approx HK\$7.8) and the sensitivity factor α was derived as 0.03 per HK\$. Based on the sensitivity factor and the actual taxi demand observed from TCS, together with a comprehensive survey for taxi services conducted in 1992 in which the customer waiting times were directly measured (Transport Department, 1992), the potential demands \tilde{D}_{ij} were estimated, which was found to be closely matched with the observed overall trip demand in Hong Kong.

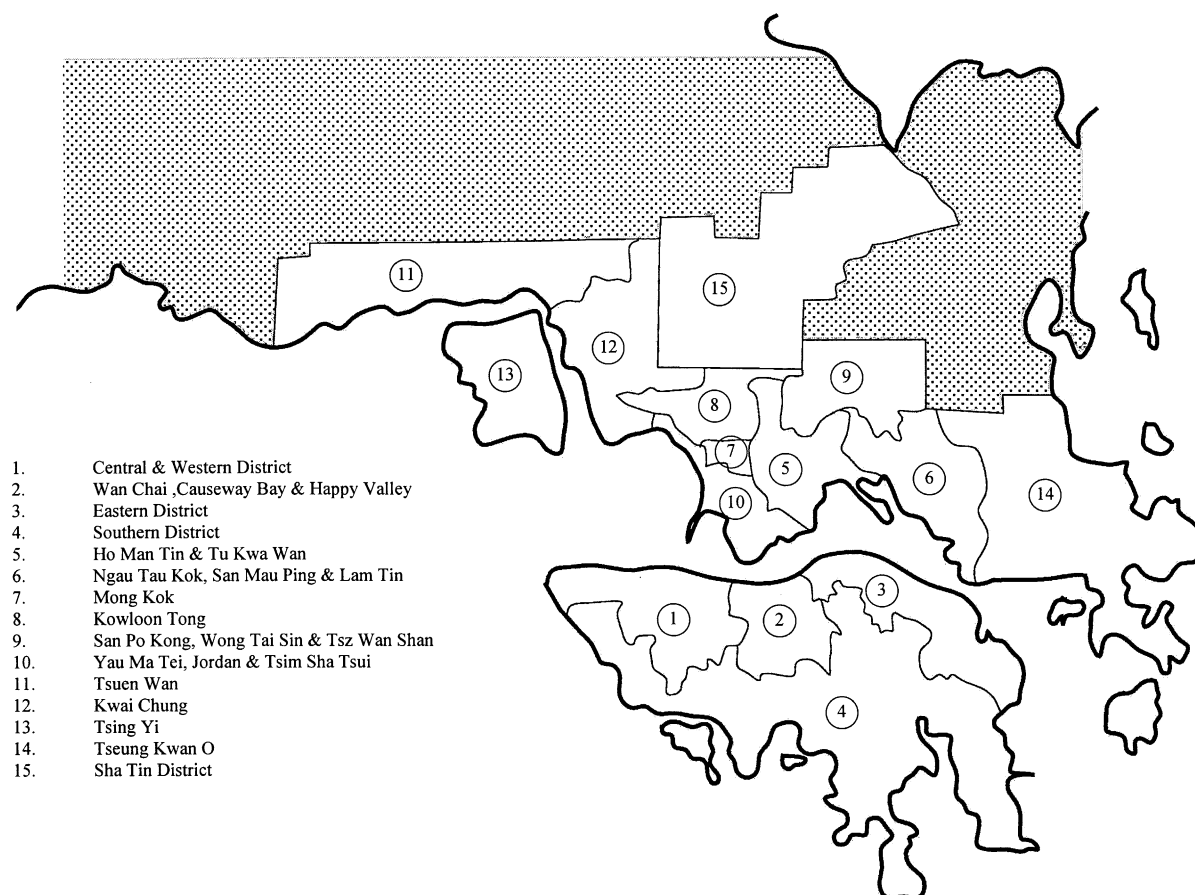


Fig. 2. The urban area of the city of Hong Kong.

One of the key model parameters, θ , was calibrated in a separate study as 11.0 per hour, in which the discrepancy between the modeled and observed taxi flows across the major screenlines and cordons in Hong Kong was minimized and the resulting taxi movement pattern was found to closely match the observed one (Wong et al., 1999). From the geometric and network configurations in Hong Kong, together with the operational characteristics of taxis, the case of continuous taxi stand distribution (taxis can pick up customer anywhere on the streets) was assumed and the parameter β was estimated as 0.4 veh h/km^2 . From the comprehensive transport study (CTS) (Transport Department, 1993b) supplemented with the statistics from the Hong Kong Annual Digest of Statistics in 1992 (Hong Kong Government, 1992), the operating and prorated capital cost of a taxi was found to be approximately HK\$70 per hour of service time.

4.2. Results and discussions

The principal operational characteristic of the taxi market is portrayed graphically in a fleet (N)–fare (τ) space shown in Figs. 3 and 4. The various combinations of fleet and fare would result

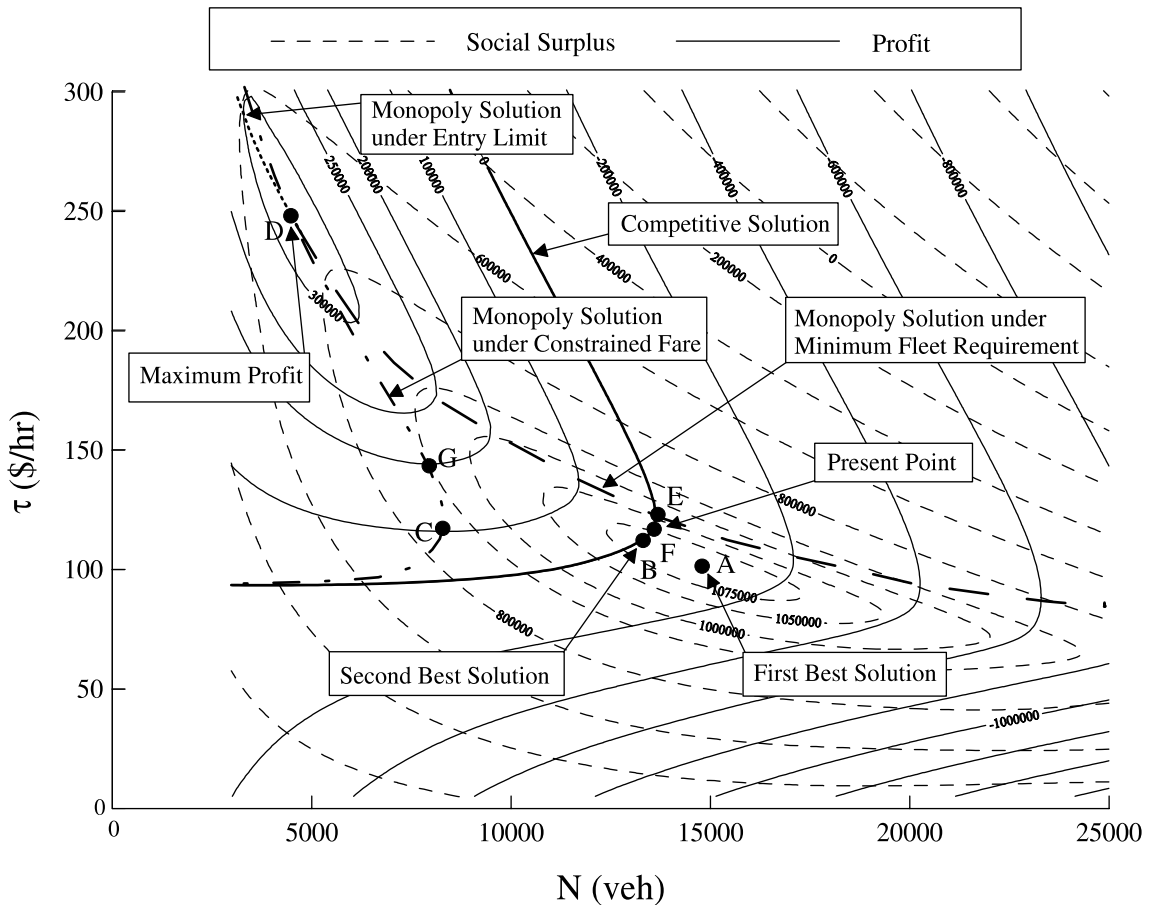


Fig. 3. Social surplus and profit of the taxi market in a fleet-fare system.

in various social surplus, taxi profits and customer demand, which are represented by iso-social surplus, iso-profit and iso-demand contours, respectively. A number of representative solutions with and without fare and/or entry controls in monopoly and competitive contexts can be identified and discussed in subsequent subsections.

4.2.1. Case 1: free entry, unconstrained fare

In this case, the *competitive* equilibrium solution is determined by the non-linear break-even equation (15). There is a feasible interval of fare within which the equation has solution. This feasible interval is given by the lowest and highest fare along the zero-profit contours. Each feasible level of fare admits two equilibrium solutions in terms of fleet size, and thus the zero-profit contour would be roughly elliptical. However, the equilibrium at smaller fleet size is unstable and inferior from a welfare perspective. Thus the lower equilibria can be excluded and are not shown in the figures. In fact, unstable solutions generally cannot be obtained from our solution algorithm, which is based on a plausible adjustment mechanism.

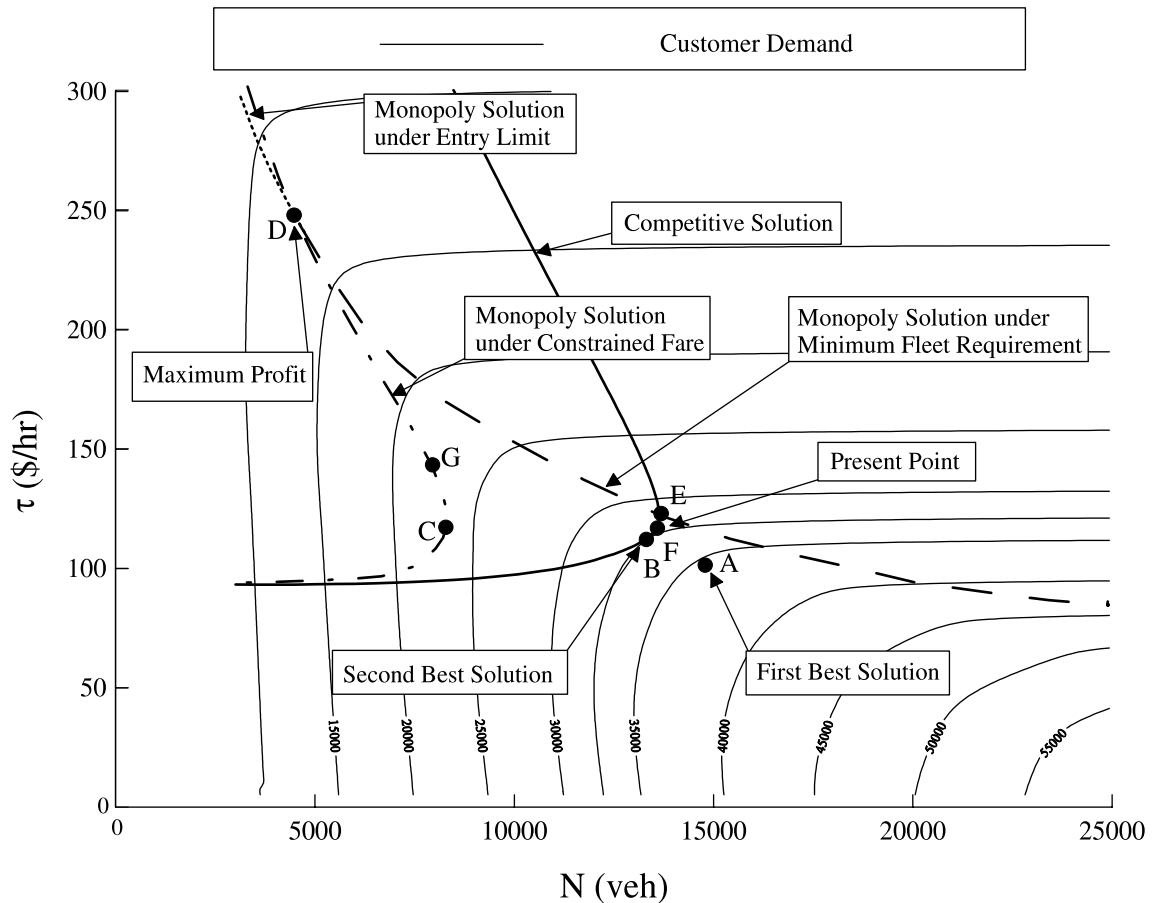


Fig. 4. Customer demand of the taxi market in a fleet-fare system.

In the fully unregulated competitive market, if the initial supply of taxis is in the left-hand side of the zero-profit curve, there will be positive profits in the industry. Supply will be expanded toward zero-profit contours until the attractiveness of this action disappears – that is until profits become zero. No service will be offered on the right-hand side of the zero-profit contours. However, there is no compelling evidence that any combination of fare and fleet solution values in the zero-profit curve will occur. Nevertheless it is conceivable that the most probable stable equilibrium occurs at point *E* where the competitive fleet size is maximized. This can be explained below. If an equilibrium occurs somewhere below point *E* at a fare lower than that at *E* (Fig. 3), a single taxi driver would be able to increase his own profit by slightly raising his fare, thus entering the positive profit region. This can be represented by a short vertical vector pointing upwards from the zero-profit curve below point *E*. This profitable action will be perceived by all drivers and more taxi firms will be attracted to join the market. The influx of additional taxis can be represented by a short horizontal vector pointing right to reach the zero-profit line. Eventually, progressions of fare and fleet increases will drive the equilibrium point to move upwards.

If an equilibrium occurs somewhere above point E at a fare higher than that at E , a single taxi driver would be able to increase his own profit by slightly reducing his fare, again entering the positive profit region. This can be represented by a short vertical vector pointing downwards from the zero-profit curve above point E . The reason for making more profit by reducing fare is that the drivers could do more business (see the iso-demand contour in Fig. 4), and hence increase the profit even with a lower fare. Again, existence of such a profitable business would attract more taxi firms to enter the market until the marginal profit of last driver becomes zero. This can be represented by a short horizontal vector pointing right to reach the zero-profit curve. As a result, progression of fare decreases and fleet increases will drive the equilibrium point to move downwards.

The only point, at which the incentive for all individual firms to change fare and/or for the number of taxis in service to change disappears, is point E , where the total derivative of profit with respect to fare vanishes. Therefore, point E is considered as a stable competitive equilibrium point in a long run. This observation indicates that a competitive market will eventually operate at point E if the market is fully deregulated.

We must emphasize that the existence of aforementioned downward and upward pressures on prices is an empirical matter and probably indeterminate, effective information communication between taxi drivers and customers about price levels is necessary to make possible upward and downward price competition.

We now consider the *monopoly* solution in the absolutely unregulated market where there exists a single firm. The profit-maximizing solution is given by solving the monopoly problem (14) with respect to both selection of taxi fleet and setting of taxi fare. The solution point to the single monopoly firm occurs at $(N^m, \tau^m) = (4000 \text{ veh}, \text{HK\$}268/\text{h})$, giving rise to a maximum profit of HK\$311,000 per hour corresponding to point D in Figs. 3 and 4. Clearly the monopoly fare is artificially high and the monopoly fleet size is low. If there were more than one firm, the firms might try to get the government to impose entry and fare regulations that would move the industry away from the zero-profit contour toward this profit-maximizing point.

4.2.2. Case 2: fixed fare, free entry

We now consider a partially regulated market where the regulator sets fare. In a *competitive* market, the solution of fleet size is unique and given by the point located at the zero-profit curve corresponding to the regulated fare. Thus the point at which the industry would operate depends on the fare set by the regulator. Two representative points are worth mentioning here. The first is the social surplus-maximizing point B located at $(N^s, \tau^s) = (13,300 \text{ veh}, \text{HK\$}113/\text{h})$. This welfare-maximizing point is referred to as the second-best solution defined by (17) and (18), which gives a maximum social surplus of HK\$1,083,000 per hour, this point also leads to a maximum realized demand (32,500 customer rides per hour), which means that social surplus and customer demand are maximized simultaneously under zero-profit constraint in our case study. Efficient fare regulation would require that fare be set at point B so that a feasible second-best solution is realized. As the fare is lowered below point B , the equilibrium utilization ratio is high, implying that the number of vacant taxis is low and the waiting time long, thus taxi demand declines because the increased waiting time more than nullifies the advantage of the lower fare for every customer. Similarly, as the fare is raised above point B , customer demand also declines, because the reduction in waiting time is insufficient to offset the higher fare for every customer. The extent of

this tradeoff depends on the value of time of customers. The second point worth of mention is the fleet size-maximizing point *E*. This point is inefficient from the society point of view because the maximum offered quantity of taxi-hours of service is not utilized efficiently. Additionally, the maximization of service quality (minimum average customer waiting time) would occur at somewhat higher fare than at point *E*.

On the other hand, in a *monopoly* market the monopolist operates where marginal revenue of taxis at the regulated price equals marginal cost of taxis, and the solution depends on the regulated fare, which is displayed by the curve of “monopoly solution under constrained fare”. There would exist a regulated fare where customer demand (and thus total consumption of service) and the social surplus are maximized, respectively. These fares need not coincide with each other, but both are located in the neighborhood of point *C* where the quantity of taxi-hours of service provided by the monopolist is maximized. Within the lower fare range below point *C*, the monopolist will increase its taxi supply as fare increases. This action also leads to the concomitant increase in social surplus, profit and realized customer demand. At higher regulated fare above point *C*, the monopolist operates where its fleet size declines with further increase in fare. This would lead to concomitant decrease in customer demand, but increase in its profit as long as the fare does not exceed point *D* where maximum profit is achieved. The social surplus will increase initially with increase in fare from *C* to point *G* (where the social surplus is a maximum along the curve of “monopoly solution under constrained fare”) and then decrease with further increase in fare above point *G*. Wherever the fare is fixed, the fleet size selected by the monopolist is always less than the fleet size that maximizes social surplus at the second-best solution point *B*.

By comparing the demand and social surplus associated with competitive and monopoly industry at the same regulated fare,³ we can observe the following. The competitive fleet size is always greater than the monopoly fleet size, thus the competitive industry demand curve (demand versus regulated fare) lies above the monopoly demand curve because of the higher service quality or lower waiting time offered by provision of service of larger taxi-hours. Therefore, the competitive outcome is superior to the monopoly outcome at every proper fare up to and including the second-best efficient level. This is mainly due to the use of different decision rules by the competitive firms and monopolist under price-fixing. Namely, the competitive firms will not select taxi-hours of operation as a function of customer demand; rather, it will behave as if it can supply an unlimited number of taxi hours at a taxi utilization rate where the marginal cost of additional taxi equals price (stay in the zero-profit curve). However, the monopoly firm will internalize the interdependencies that make the utilization rate of a given taxi dependent on the number of taxis operating to maximize its total profit at a given, constrained fare. If, on the other hand, prices are set too high (well above the second-best efficient level), the reduction in taxi supply experienced under monopoly decision-making becomes relatively efficient in terms of total market conditions (social surplus). Namely, marginal cost of taxi-hour under competition will be in excess of

³ This can be checked by drawing a horizontal line at the regulated fare, and finding its two intersections with the curve of monopoly solution under constrained fare and the zero-profit curve, respectively. Then the values of iso-social surplus and iso-demand contours going through the intersections can be identified.

marginal value. This is caused by the regulator's setting price above the efficient level, which results in wasteful competition among firms.

4.2.3. Case 3: limited entry and unconstrained fare

The case of limited entry and unconstrained fare seldom exists in reality. Figs. 3 and 4 indicate that entry controls, when set at binding, will normally have an effect on both fare and demand levels in an either competitive or monopoly market. In a *monopoly* market, if the entry limit is set below the level at point *D* (it is conceivable that no regulatory authority would like to set such a lower entry limit), the monopolist will be willing to supply as many taxi-hours as possible since profit is increasing in this range. Then the fare movement will occur along the curve of "monopoly solution under entry limit". If the taxi-hour limit is greater than level at *D*, the monopolist will choose point *D* as the optimal level of supply conditional on unconstrained fare, and thus the entry limit has no impact. Nevertheless, there might be one possibility that holding a monopoly right granted by the regulatory authority would require the monopoly firm to operate a minimum number of taxis in order to sustain a minimum service standard. In this case the monopolist will select a fare to maximize monopoly profit, and the fare movement occurs along the curve of "monopoly solution under minimum fleet requirement". The monopoly solution curve could be further extended beyond point *E* in a short run where the monopoly firm operates at a loss, but in a long run, it may make profit.

On the other hand, in a *competitive* market with entry limit, the market becomes protected, and will operate in the positive profit region at given entry limit. The fare could probably take any value in between the lowest and the highest zero-profit fares at the given entry limit. But it tends to self-adjust upwards or downwards and eventually stay at average profit-maximizing point at the entry limit. If the entry limit is set below profit-maximizing fleet size at point *D*, the competitive solution coincides with the monopoly solution under entry limit. This explains that the regulation of entry limit would create a monopoly market where individual taxi drivers coordinate and cooperate to fix their fare level (say through a taxi driver union or association) because they all share the common interest: profit maximization. If the entry limit is set in between points *D* and *E*, competitive solution will stay on the curve of "monopoly solution under minimum fleet requirement". Although in this case competitive and monopoly solutions are identical, the forces that sustain their equilibrium are different: the force for the former is due to active limit on the entry of new competitive firms (one taxi per firm), while that for the latter is due to minimum fleet size that a monopoly firm has to provide in order to hold a monopoly right from the government. Furthermore, under effective fare competition, the resultant fare is much higher and the realized demand is much lower in comparison with the second-best efficient solution, it seems fair to conclude that entry controls should only be used in conjunction with fare controls.

4.2.4. Case 4: limited entry and fixed fare

This is the usual situation with most large city taxi regulations. Regulatory authorities allocate a fixed number of operating permits, and prices are fixed on a meter basis. A single firm in a *monopoly* market will operate at a level of taxi-hours as close as possible to the profit-maximizing point *D*. The *competitive* market will produce a fleet size defined by the price and/or entry limit. The entry limit, if set in the right-hand side of the zero-profit curve, will have no effect, and, if set

in the left-hand side of the zero-profit curve, will become a determining factor. In the latter case, the competitive firms will earn supernormal profits that could be reflected by a high taxi transaction price. On the other hand, the fare constraint, if set above dashed line $D-E$, will generally have no impact, and, if set below $D-E$, will determine the competitive outcome in conjunction with entry limit. Therefore, when fare and entry are regulated, either regulation may be the determining factor.

Whenever restriction in fare and entry is binding, relaxation of such a restriction could result in either an increase or a decrease of social surplus. Now suppose that the authority initially imposes fare and entry regulations so that the taxi market is operating in the left-hand side of zero-profit curve, there would be efficiency gain if the authority reduced the fare to the second-best efficient level and eliminated entry restrictions, since the competitive market would then move to the second-best point B , which is on a higher iso-social surplus contour. If the fare was reduced but the entry restriction was not changed, the market would move down to stay in a higher iso-social surplus contour (a smaller positive efficiency gain) if the entry restriction is set relatively high (say higher than or equal to the second-best efficient level), and the market would move down to stay in an either higher or lower iso-social surplus contour (either positive or negative efficiency gain) if the entry restriction is set well below point B . In a similar manner, it can be checked that elimination of the entry restriction without a reduction in fare might either reduce or increase social surplus, depending on the initial location of market operation. Therefore, regulation of a taxi market should be implemented cautiously to improve social surplus. The second-best social optimum can be realized by restriction of fare only in an either competitive or monopoly market.

4.2.5. *Leader-follower or Stackelberg game solution*

Some cases of the taxi service regulation and market equilibrium problem examined so far can be described as a leader-follower, or a Stackelberg game (Varian, 1992) where the regulator is the leader, and the taxi firm(s) is the follower. In a competitive or monopoly market, suppose the regulator has control of either fare or entry (fleet size), but not both, in light of any control decision, taxi firm(s) decides whether to enter the market (selection of fleet size in the case of monopoly) or chooses a fare to maximize profit. Thus, the Stackelberg game solution can be characterized by that the regulator chooses an optimal control decision to maximize social surplus, while taking into account the reaction of the taxi firm(s).

We first consider the case where the regulator has control of fare only. In a competitive market, individual taxi firms (drivers) decide whether or not to enter the market for a given fare, and the aggregate reaction curve of taxi firms is the curve of “competitive solution”. Point B (the second-best point) is the Stackelberg game solution that leads to a maximum social surplus along this reaction curve. In a monopoly market, the reaction curve of the monopolist is the curve of “monopoly solution under constrained fare”, which describes how the monopolist chooses a taxi fleet size to maximize profit for a given regulated fare. The Stackelberg game solution corresponds to point G at which the reaction curve is tangent to one of the social surplus contours. Clearly, point G is near point C (the maximum monopoly fleet size solution).

We next consider the case where the regulator has control of entry or fleet size only. Then the reaction curve of taxi firm(s) will be the broken line $D-E$ in both competitive and monopoly market. As mentioned before, this reaction curve is due to the regulator’s entry limit control in a

competitive market and minimum fleet requirement in a monopoly market. The Stackelberg game solution in this case depends where the reaction curve is tangent to the social surplus contours. If tangent at point *E* or any point in its left-hand side (the non-negative profit domain), the tangent point will be the Stackelberg solution. Otherwise, point *E* will be the solution because a tangent point associated with a negative profit is infeasible. In this case study, it was found that the Stackelberg game solution is located at point *E* for the Hong Kong situation.

4.2.6. Social optimum

Finally, we briefly discuss the first-best social optimum. From the point of view of the economy as a whole, the efficient allocation would be to choose a fleet–fare combination to maximize social surplus defined by (16). The first-best social optimum is obtained at point $(N^f, \tau^f) = (14,600 \text{ veh, HK\$103/h})$, which is located at the right-hand side of the zero-profit contour. The first-best taxi pricing entails taxi operation at a loss; more specifically at an aggregate loss of HK\\$81,000 per hour, but resulting in a social surplus of HK\\$1,088,000 per hour in the market. Therefore attainment of the full first-best would require introduction of a mechanism that subsidizes taxi travel to cover the deficit that is related to the cost of vacant taxi-hours at the optimum. Regrettably, the improvement in social surplus gained by driving the market from the second-best to the first-best is very marginal (a small net increase of HK\\$5,000 per hour from HK\\$1,083,000 per hour of the second-best to HK\\$1,088,000 per hour of the first-best) and may not even be able to cover the amount needed for subsidy.⁴ This is definitely not cost-effective in our case study. However, even if it does, it raises another important question: Is it appropriate to try to attain a social optimum at the expense of introducing a complex subsidy mechanism?

4.2.7. Existing situation

It is also interesting to look at the situation in the year 1992 for the taxi services in Hong Kong, where the case study was carried out. As in most large cities in the world, both fare and entry are regulated in Hong Kong. The Government does not subsidize the taxi industry and closely monitors the taxi services by means of continuous public consultation (Hong Kong Transport Advisory Committee, 1992) and, since 1986, independent surveys on taxi services (see for example, Transport Department, 1992). In 1992, the number of taxis in service and taxi fare were around 13,500 and HK\\$115 per hour, respectively. From the fleet–fare space, this state of regulation is illustrated as point *F* in Figs. 3 and 4. It is interesting to note that the regulated fare and entry are reasonably close to the second-best solution, which reflects the fact that the policy set by the Government has acted quite effectively in response to the market outcome.

Moreover, from the figures, it can also be found that the most probable stable equilibrium at point *E* assuming a competitive market (see Section 4.2.1) is in fact very close to the second-best

⁴ This should be mentioned cautiously because the potential efficiency gains from first-best pricing may depend on the O–D matrix of customer demand. If the O–D matrix is symmetric, there will be less vacant taxi movement. Otherwise, a taxi once dropped a customer in a zone, have to go to other zones to find next customers, and thus vacant taxi movement will increase. As already shown in the aggregative model (Arnott, 1996), the required subsidy level for the first-best social optimum equals the cost of vacant taxi-hours at the optimum. The vacant taxi hours and the resulting subsidy level, in turn, imply the discrepancy between the first-best and the second-best social optimum.

solution (see Section 4.2.4). This also illustrates the fact that the Government has successfully created a socially effective but also highly competitive market environment for taxi services, through a good mix of company and individual taxi owners and a competitive bidding auction for taxi licenses (Hong Kong Transport Advisory Committee, 1992).

5. Conclusions

This paper was concerned with the modeling of demand and supply equilibrium in a regulated taxi market. Our study differs from conventional economic analysis of taxi service by explicitly considering the spatial structure of the market over which price and demand interact. The equilibrium between taxi movements and customer demand is established at a level of a road network. Therefore, our model made the operation of a realistic taxi market amenable to analysis in a detailed spatial context, and thus added realism to the work to date by a large number of economists. For example, under effective fare competition, the resultant fare is much higher and the realized demand is much lower in comparison with the second-best efficient solution. Thus it is important to look at how entry restriction can be used in conjunction with price control to achieve the second-best social optimum in a specific urban area. This is especially meaningful for modeling service area regulation where consideration of the spatial structure of the taxi market is essential.

From the case study for the city of Hong Kong, it is indeed offered some interesting insights into the equilibrium properties of the taxi services. Our graphical representation (Fig. 3) embraced nearly all the findings in the existing economic literature. This unified representation is very helpful to the intuitive understanding of the way taxi market operates, especially for non-economists since comprehension of the figure does not require a strong economic background.

Acknowledgements

The authors would like to thank Richard Arnott of Boston College and two anonymous referees for their helpful comments on an earlier version of the paper. The research described here was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project no. HKU7019/99E).

Appendix A

The relationship between waiting times for the case of discrete taxi stand distribution can be derived by treating the problem as considering m_i identical M/M/1 queuing systems with identical customer arrival and taxi service rates at the same zone $i \in I$. At a given taxi stand, the average taxi waiting time at zone i is the fraction of time or the probability of finding server idle in a stationary state. This is given by $w_i = 1 - \lambda_i/\mu_i$, where λ_i is customer arrival rate: $\lambda_i = O_i/m_i$ with $O_i = n_i$, and μ_i is taxi service rate. We thus have

$$\mu_i = \frac{\lambda_i}{(1 - w_i)} = \frac{O_i}{(1 - w_i)m_i}.$$

As a result, the average customer waiting time at zone i is given by

$$\begin{aligned} W_i &= \frac{1}{\mu_i - \lambda_i} \\ &= 1 / \left(\frac{O_i}{(1 - w_i)m_i} - \frac{O_i}{m_i} \right) \\ &= \left\{ \frac{1}{w_i} - 1 \right\} \frac{m_i}{O_i} \\ &= \left\{ \frac{1}{w_i} - 1 \right\} \frac{m_i}{n_i}, \quad i \in I. \end{aligned}$$

Appendix B

The proof follows from Douglas (1972). Let \bar{A}_i be defined as the number of street kilometers in the reference zone $i \in I$. Let $\tilde{A}_i(t)$ represent the area searched since initial time $t = t_0$. The “gross search rate” is equal to $\tilde{n}_i S$, where \tilde{n}_i is the average number of vacant taxis in the zone at each instant and S is their cruising speed. The “net search rate”, adjusting for redundant coverage, then is

$$\frac{d\tilde{A}_i(t)}{dt} = \left[\frac{\bar{A}_i - \tilde{A}_i(t)}{\bar{A}} \right] \tilde{n}_i S.$$

Let the area not searched at time t be $B_i(t) = \bar{A}_i - \tilde{A}_i(t)$. The rate of change of $B_i(t)$ is equal to the negative net search rate,

$$\begin{aligned} \frac{dB_i(t)}{dt} &= \left[\frac{\bar{A}_i - \tilde{A}_i(t)}{\bar{A}} \right] \tilde{n}_i S, \\ \frac{dB_i(t)}{dt} &= -B(t) \frac{\tilde{n}_i S}{\bar{A}}. \end{aligned}$$

Therefore, $B(t)$ is of the form

$$B_i(t) = \bar{A}_i e^{-\rho t}, \quad \text{where } \rho = \frac{\tilde{n}_i S}{\bar{A}_i}.$$

The delay distribution $P(t)$ is the probability of not having found a taxi at time t ,

$$P_i(t) = \frac{B_i(t)}{\bar{A}} = e^{-\rho t}$$

with expected value

$$W_i = \int t \cdot P_i(t) dt = \frac{1}{\rho} = \frac{\bar{A}_i}{\tilde{n}_i S}.$$

Now the rate of arrival of taxis to zone i is n_i and each taxi spends an average time w_i , we thus have $\tilde{n}_i = n_i w_i$. The expected average customer waiting time at zone i is thus given by

$$\begin{aligned} W_i &= \frac{\bar{A}_i}{\tilde{n}_i S} \\ &= \frac{\bar{A}_i}{n_i w_i S} \\ &= \beta \frac{A_i}{n_i w_i}, \quad i \in I, \end{aligned}$$

where we assumed that the total number of street kilometers, \bar{A}_i is proportional to the zone area A_i , parameter β is introduced to capture the effects of road density, taxi cruising speed and any other influencing factors.

References

- Abe, M.A., Brush, B.C., 1976. On the regulation of price and service quality: The taxicab problem. *Quarterly Review of Economics and Business* 16 (Autumn), 105–111.
- Arnott, R., 1996. Taxi travel should be subsidized. *Journal of Urban Economics* 40, 316–333.
- Beesley, M.E., Glaister, S., 1983. Information for regulation: The case of taxis. *The Economic Journal* 93, 594–615.
- 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.
- Foerster, J.F., Gorman, G., 1979. Taxicab deregulation: Economic consequences and regulatory choices. *Transportation* 8, 371–387.
- Frankena, M.W., Pautler, P.A., 1986. Taxicab regulation: An economic analysis. *Research in Law and Economics* 9, 129–165.
- Hong Kong Government, 1992. *Hong Kong Annual Digest of Statistics*. Hong Kong Government Printer, Hong Kong.
- Hong Kong Transport Advisory Committee, 1992. Consultative paper on taxi policy review. Hong Kong Transport Advisory Committee, Hong Kong.
- INRO, 1997. EMME/2 user's manual. INRO Consultants Inc.
- Manski, C.F., Wright, J.D., 1976. Nature of equilibrium in the market for taxi services. *Transportation Research Record* 619, 11–15.
- Schroeter, J.R., 1983. A model of taxi service under fare structure and fleet size regulation. *Bell Journal of Economics* 14, 81–96.
- Shrieber, C., 1975. The economic reasons for price and entry regulation of taxicabs. *Journal of Transport Economics and Policy* 9, 268–293.
- Transport Department, 1989. Second comprehensive transport study (final report). Hong Kong Government, Hong Kong.
- Transport Department, 1992. The level of taxi services. Hong Kong Government, Hong Kong.
- Transport Department, 1993a. Travel characteristics survey (final report). Hong Kong Government, Hong Kong.
- Transport Department, 1993b. Updating of second comprehensive transport study (final report). Hong Kong Government, Hong Kong.
- Varian, H.R., 1992. *Microeconomic Analysis*, third ed. W.W. Norton & Company, New York (Chapter 15, Game theory).

- Wong, K.I., Wong, S.C., Yang, H., 2001. Modeling urban taxi services in congested road networks with elastic demand. *Transportation Research Part B* 35, 819–842.
- Wong, S.C., Yang, H., 1998. Network model of urban taxi services: Improved algorithm. *Transportation Research Record* 1623, 27–30.
- Wong, K.I., Wong, S.C., Yang, H., 1999. Calibration and validation of a network equilibrium taxi model for Hong Kong. In: *Proceedings of the Fourth Conference of Hong Kong Society for Transportation Studies*, Hong Kong, December 4, 1999, pp. 249–258.
- Xu, J.M., Wong, S.C., Yang, H., Tong, C.O., 1999. Modeling the level of urban taxi services using a neural network. *ASCE Journal of Transportation Engineering* 125, 213–216.
- Yang, H., Wong, S.C., 1998. A network equilibrium model of urban taxi services. *Transportation Research* 32B, 235–246.
- Yang, H., Lau, Y.W., Wong, S.C., Lo, H.K., 1999. A macroscopic taxi model for passenger demand, taxi utilization and level of services. *Transportation* 27, 317–340.