A diagrammatic analysis of the market for cruising taxis

J. Enrique Fernández L. a,1, Joaquín De Cea Ch a, Julio Briones M. b,*

a Departamento de Ingeniería de Transporte, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile
b Secretaría Interministerial de Planificación de Transporte (SECTRÁ), Teatinos 950 Piso 16, Santiago, Chile

Abstract

A diagrammatic approach is used to study the characteristics of the cruising taxi market. For cost modeling, both taxi operator and passenger are taken as service producers. The former provides his vehicle operation and the latter, his waiting and travel time. Market demand is defined as a function of generalized price. It is shown that under short and long run conditions a unique equilibrium exists for a deregulated industry and it corresponds to a monopolistic competition. The relations among the free market equilibrium, social optimum and second best solution are analyzed. Regulations are studied in order to find their social convenience.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Cruising taxis; Monopolistic competition; Regulation policies

1. Introduction

Cruising and dispatch taxi services often coexist, with cruising taxi service predominating in most major cities and in most developing countries. Considerable attention has been given in the literature to the economics of cruising taxi service. Different authors have identified some

* Corresponding author. Fax: +56 2 696 6477.
E-mail addresses: jef@ing.puc.cl (J.E. Fernández L.), jbriones@ing.puc.cl (J. Briones M.).
1 Fax: +56 2 686 4818.
2 In Santiago, the Chilean capital, a total of 50,000 taxis were operating in 1998; 45,000 of them providing cruising service.
peculiar features of this industry. Manski and Wright (1976), Cairns and Liston-Heyes (1996) observe that demand and supply are interrelated, or according to Beesley and Glaister (1983), subject to interdependencies. According to Shreiber (1977), there is a lack of perfect synchronization between passengers and available cabs, which makes impossible the equilibrium between supply and demand. Shreiber (1975, 1981) indicates that the lack of passengers-cabs synchronization produces, in general, an upward pressure on the service price that drives the cruising taxi system to a condition with high fares, low cab occupancy, short waiting times and price competition non-existent. Yang et al. (2002), departing from the conventional economic analysis, use a network model to analyze the supply-demand equilibrium characteristics of cruising taxi services for the case of Hong-Kong. Cairns and Liston-Heyes (1996) using a model of search, where drivers and riders search for each other, conclude that equilibrium of a deregulated industry does not exist.

To deal with these market peculiarities, regulations for cruising taxi service have been recommended and adopted almost everywhere. Fare and entry regulations have been the most commonly considered interventions analyzed by different authors (Douglas, 1972; De Vany, 1975; Shreiber, 1975, 1977; Williams, 1980; Beesley and Glaister, 1983; Frankena and Pautler, 1986; Rometsch and Wolfstetter, 1993; Cairns and Liston-Heyes, 1996; Yang et al., 2002). The previous literature has been almost completely algebraic. It turns out, however, that a complete and precise diagrammatic analysis of the economics of cruising taxi service can be developed using the approach presented in Mohring (1976) for the production of passenger transport services modeling. This approach complemented by the usual algebraic analysis is specially useful to study market equilibrium. It also allows a better understanding of the market adjustment mechanisms and the analysis of the characteristics of the outputs obtained from different regulatory policies.

Some of our basic conclusions coincide with those already presented in the literature by previous authors. As others before, we conclude that the system long run average cost function is always decreasing and, therefore, the social optimum fare produces losses to taxi operators. Nevertheless, we show that the industry economies of scale are produced by two different positive externalities, one that reduces user waiting times and a second that reduces taxi average operating cost, because, given perfect adaptation of supply to demand, taxi occupancy rate increases with industry size. Interestingly, both externalities have exactly the same analytical expression.

However, other results importantly differ from those previously published. We obtain that a unique long run equilibrium exist for free market conditions. We describe how such equilibrium is obtained from the interactions between industry demand and supply conditions. Interestingly enough this equilibrium is a monopolistic competition equilibrium.

The model developed allows to conclude that, in general, the need for regulation must be carefully considered case by case, because depending on the relative magnitudes of demand and system costs, a long run free market equilibrium could be very close to a second best solution and in the limit they could coincide.

We show that entry regulations are, in general, redundant, when applied together with fare regulations, and that they produce worse industry conditions than a free market policy, when applied alone. We also show that licensing will produce in general similar results as fleet size regulation.

---

3 Santiago de Chile has been an interesting exception for a long period between 1987 and 2000.
In Section 2, the basic model is presented; the main assumptions are stated, the individual and system production costs functions are defined and graphically represented and the demand function assumed is specified. In Section 3, we use the diagrammatic representation to show the differences between social optimum and second best solution. In Section 4, the unregulated market operation is analyzed and short and long run market equilibria are studied. In Section 5, the most commonly used regulatory policies are analyzed and their outcomes are compared. Finally Section 6 summarizes the main conclusions of the paper.

Figures presented through the paper for the diagrammatic analysis are plotted evaluating the analytical functions with real data from the Santiago de Chile cruising taxi system (see Fernández et al., 2001).

2. The model

2.1. Main assumptions

Some important assumptions made are the following:

(i) Only cruising services are considered. Taxis permanently run the streets looking for passengers. When a passenger is found he is driven to his destination after which the taxi resumes the search for a new passenger. The trip with a passenger is called a run.

(ii) Taxis operate in a given geographical area, during a given time period, for which homogeneous operating conditions are assumed. We will assume that the time period is 1 h.

(iii) It is assumed that each taxi operates independently without possibility of collusion. The driver–owner and the vehicle constitute the individual taxi firm.

(iv) Runs have an average length \( l \), and their duration is equal to a constant time, \( t \). Duration time \( t \) is a consequence of operating conditions and general congestion levels experienced on the streets of the area considered. In this paper the influence of taxis on general congestion is not considered.

(v) All taxis make the same average number of runs per period, \( q \). Therefore, if \( N \) is the number of taxis and \( Q \) is total number of runs produced by all taxis during the analysis period, then \( q = Q/N \).

In Fig. 1, we present a diagrammatic example of the main assumptions. Taxis operate over a linear city, with a network longitude equal to \( L \) (see vertical axis) and during a period \( T \) (see horizontal axis). Taxis \((i, j \text{ and } k)\) and passengers are uniformly distributed over the space and time dimensions. All taxis operate at the same speed \( s \), cruising through the entire network and facing the same number of passengers. The cruising and travel time, \( t_{\text{crui}} \) and \( t \), respectively, are similar for all operators. Passenger average waiting time is \( w \).

For the example represented in the figure, we can see that, during the period \( T \), taxi \( i \) faces the demand of three passengers \((A, B \text{ and } C)\). When taxi \( i \) reaches each of them, runs will be pro-

---

\(^4\) In most developing countries and especially in Chile, taxi operators are individuals who own one vehicle and operate independently.
duced. Hence, it is important to note that the product in this market, taxi run, is closely characterized by geographical and temporal attributes given by passenger initial location and the departure time. Therefore, we can say that all runs are different and each taxi operator has a monopoly power over each passenger demand. Based on this argument we raise that this market has a monopolistic competition behavior. This will be discussed in Section 4.

2.2. Individual and industry capacities. Occupancy rate

In addition, we will use the following concepts: The individual capacity of a taxi is the maximum number of runs that it can produce per period, and is equal to the inverse of the average run duration, \( 1/t \). Therefore the industry capacity will be given by the sum of the individual capacities over the total number of taxis, \( N/t \).

The occupancy rate, is equal to the ratio between the number of runs produced by a taxi during the analysis period and the individual capacity of the taxi, \( \alpha(q) = q \cdot (1/t)^{-1} = q \cdot t \) with \( q \leq 1/t \). Given that \( q = Q/N \), we can also define \( \alpha(N, Q) = Q \cdot t/N \), with \( Q < N/t \).

2.3. Cost modeling approach

We adopt the approach proposed by Mohring (1976) for modeling the production of passenger transport services. The service offered by taxi operators is considered a quasi-service, which in order to become a real service must have the participation of the passenger, which contributes with two important production factors: the waiting time and travel time. Therefore, both taxi operators and passengers are taken as service producers. Accordingly, the number of taxi runs really produced is equal to the number consumed.
Fig. 2 shows a temporal scheme about the production process of an individual taxi. Taxis cruise the streets looking for a passenger during a time period, $t_{crui}$; when it finds one (see point $O$), it begins a run driving the passenger to his destination (point $D$), and after that, it starts to cruise again. In the other side, passenger waits $w$ for a taxi. This cycle is repeated during the whole operation period. In this market characterized by lack of perfect synchronization between passengers and available cabs (Shreiber, 1975), point $O$ is especially important, because it is the place where taxi and passenger meet to produce a run.

According to this scheme, the production frequency of an individual taxi, $f_{\text{taxi}} = 1/(t_{\text{crui}} + t)$ and its individual demand, $f_{\text{demand}} = q$ must be equal. Thus, taxi production function can be obtained from the following expression:

$$f_{\text{demand}} = f_{\text{taxi}} \Rightarrow q = \frac{1}{t_{\text{crui}} + t} \Rightarrow t_{\text{crui}}(q) = \frac{1}{q} - t$$

Expression (1) shows that $t_{\text{crui}}$ depends on $q$, and, therefore, on $Q$ and $N$. As taxi individual demand ($q$) decreases, $t_{\text{crui}}$ increases, because is more difficult to find a passenger.

2.4. Individual taxi operating cost in the short run

The operating cost of an individual taxi includes direct costs like fuel consumption and indirect costs like vehicle cost. Assuming that the effect of a passenger on the operating cost is negligible and given that cruising taxis are permanently running, we consider as previous authors, that taxi operating cost is fixed, and proportional to the length of the operating period considered (Douglas, 1972; De Vany, 1975; Cairns and Liston-Heyes, 1996). The total operating cost incurred by a taxi during a given period of an hour can be expressed by:

$$\text{TOC}_{\text{taxi}}(q) = c \cdot (t_{\text{crui}} + t) \cdot q = c \quad 0 \leq q < \frac{1}{t}$$

It is important to note that expression (2) is equivalent to formulate the total cost as the sum of total cruising and running time during the period (1 h) due to expression (1).

The average operating taxi cost, per run produced, is equal to:

$$\text{AOC}_{\text{taxi}}(q) = \frac{c}{q} \quad 0 \leq q < \frac{1}{t}$$
The average cost function is hyperbolic and achieves its minimum value when the maximum number of runs is produced, that is, \( q = 1/t \Rightarrow \text{AOC}_{\text{taxi}}(1/t) = ct \).

The marginal operating taxi cost, per run produced, is equal to:

\[
\text{MOC}_{\text{taxi}}(q) = 0 \quad 0 \leq q < \frac{1}{t} \tag{4}
\]

Expression (4) shows that cruising taxi operating cost is independent of the number of runs produced.

2.5. Industry operating cost in the short run

We define now the average operating cost for the taxi industry (\( \text{AOC}_{\text{ind}} \)). It is obtained by horizontally adding all individual average operating taxi cost functions (3). Considering a total of \( N \) identical operators and using \( q = Q/N \) we obtain:

\[
\text{AOC}_{\text{ind}}(N, Q) = \frac{N \cdot c}{Q} \quad 0 \leq Q < \frac{N}{t} \tag{5}
\]

\( \text{AOC}_{\text{ind}}(N, Q) \) gives the minimum fare needed by taxi operators to finance all inputs used in the service production. For each possible fare value, it gives the number of runs required to finance the industry operation.

2.6. Passenger costs in the short run

Because we do not consider the effect of taxis on streets flow congestion (see assumption (iv)), the average travel cost has a constant value, determined by general traffic conditions and the passengers value of time \( \phi \). Therefore, the marginal cost is also constant and has the same value:

\[
\text{ATC} = \text{MTC} = \phi \cdot t \tag{6}
\]

The average waiting time experienced by an individual looking for a taxi, depends on the number of taxis in operation and the number of runs consumed, \( w(N, Q) \) (Cairns and Liston-Heyes, 1996) or, in other words, “the delay distribution is a function of the density of vacant taxicabs in the area” (Douglas, 1972). Then, for a given value of \( N \), as the number of runs produced increases more taxis are occupied and the probability of finding an empty taxi decreases. Based on the general expression derived by Douglas (1972), we assume the following average waiting time cost functional form:

\[
\text{AWC}(N, Q) = \theta \cdot w(N, Q) = \frac{\theta \cdot k}{N - Q \cdot t} \quad 0 \leq Q < \frac{N}{t} \tag{7}
\]

where \( \theta \) is the waiting time value, and \( (N - Q \cdot t) \) is the average number of empty taxis available during the analysis period.

The general expression derived in Douglas (1972) for waiting time is \( w = A/Vs \) with \( A \) representing the number of street kilometers in the served area, \( s \) is the average operating speed and \( V \) is the number of vacant taxis \( (V = N - Qt) \). Therefore, in (5) \( k \) should be accordingly interpreted as \( k = A/v \). Otherwise, \( k \) be considered a calibration parameter for the specific area considered.
Therefore, from (7), AWC is a convex function of \( Q \). When \( Q = 0 \), the average waiting time cost for a passenger is equal to \( k/N \). On the other side, as the number of runs produced approaches the industry capacity \( (N/t) \), the average waiting time cost asymptotically rises to infinity. Then, the probability of finding an empty taxi goes to zero.

The marginal waiting cost is obtained derivateing with respect to \( Q \), the total waiting time cost \( (TWC = AWC \cdot Q) \) experienced by all consumers \( Q \):

\[
MWC(N, Q) = \frac{\partial[TWC(N, Q)]}{\partial Q} = \frac{\theta \cdot k}{N - Q \cdot t} + \frac{\theta \cdot k \cdot t \cdot Q}{(N - Q \cdot t)^2} = \frac{\theta \cdot k \cdot N}{(N - Q \cdot t)^2} \quad 0 \leq Q < \frac{N}{t}
\]

(8)

The average and marginal passenger cost functions are obtained \( (AC_{pas}(N, Q), MC_{pas}(N, Q)) \), adding the corresponding waiting and travel time cost functions defined above.

\[
AC_{pas}(N, Q) = \phi \cdot t + \frac{\theta \cdot k}{N - Q \cdot t} \quad 0 \leq Q < \frac{N}{t}
\]

(9)

\[
MC_{pas}(N, Q) = \phi \cdot t + \frac{\theta \cdot k}{N - Q \cdot t} + \frac{\theta \cdot k \cdot t \cdot Q}{(N - Q \cdot t)^2} = \phi \cdot t + \frac{\theta \cdot k \cdot N}{(N - Q \cdot t)^2} \quad 0 \leq Q < \frac{N}{t}
\]

(10)

These functions are represented in Fig. 3.

2.7. System costs in the short run

The average system cost function, \( ASC(N, Q) \), is obtained adding the average operating cost of the industry, \( AOC_{ind}(5) \), and the average passenger cost function, \( AC_{pas}(9) \):

\[
ASC(N, Q) = AOC_{ind}(N, Q) + AC_{pas}(N, Q) = \frac{N \cdot c}{Q} + \phi \cdot t + \frac{\theta \cdot k}{N - Q \cdot t} \quad 0 \leq Q < \frac{N}{t}
\]

(11)

The ASC includes the costs corresponding to all inputs used (provided by the operator and the passenger) to produce a run. As we can see in Fig. 3(b), it has a U form. For low values of \( Q \), \( ASC \) is decreasing with \( Q \) because it is dominated by the decreasing characteristic of the \( AOC_{ind} \). However, for high values of \( Q \) it is increasing, because it is dominated by the effect of the waiting time cost, as production \( Q \) approaches capacity \( (Q \to N/t) \). In the same figure the value of \( Q \), for which the minimum value of the average system cost \( (ASC(N, Q)) \) is obtained, is represented by \( Q_{\min}(N) \).

The corresponding marginal system production cost function, \( MSC(N, Q) \) (represented also in Fig. 1(b)) has the form:

\[
MSC(N, Q) = \frac{\partial[ASC(N, Q) \cdot Q]}{\partial Q} = \phi t + \frac{\theta \cdot k}{N - Qt} + \frac{\theta \cdot k \cdot t \cdot Q}{(N - Qt)^2} = MC_{pas}(N, Q) \quad 0 < Q < \frac{N}{t}
\]

(12)

Notice that average and marginal system cost functions in Fig. 3(b), present one of the forms assumed by Mohring (1976) (Type IC) for schedules arising in transportation activities.
As we can observe, from (10) and (12), marginal system cost, MSC, is equal to marginal passenger cost, MC\textsubscript{pas}, because taxi operating costs are fixed.

2.8. **Waiting time externality in the short run**

As it is well known, when a new run is produced in the short run, the average waiting time experienced by all consumers of taxi services increases, because the availability of empty taxis is reduced. Therefore, each consumer experiences the average waiting time but produces also a negative externality over the other consumers. The social waiting time cost produced by an additional passenger is explained by the marginal waiting time cost function. Therefore, the negative externality can be readily obtained by subtracting the average to the marginal waiting time cost (Eqs. (7) and (8)):

\[
E_w(N, Q) = MWC(N, Q) - AWC(N, Q) = \frac{\theta \cdot k \cdot t \cdot Q}{(N - Q \cdot t)}, \quad 0 \leq Q < \frac{N}{t}
\]  

(13)
2.9. System costs in the long run

The long run average system cost function will be equal to the envelope of the family of short run average system cost functions. In Appendix A, the following long run total, average and marginal system cost functions are obtained:

\[
TSC^{LR}(Q) = 2\sqrt{\frac{\theta \cdot k \cdot Q \cdot c}{Q}} + Q \cdot t \cdot (c + \phi)
\]  
\[
ASC^{LR}(Q) = 2\sqrt{\frac{\theta \cdot k \cdot c}{Q}} + t \cdot (c + \phi)
\]  
\[
MSC^{LR}(Q) = \sqrt{\frac{\theta \cdot k \cdot c}{Q}} + t \cdot (c + \phi)
\]

Functions (15) and (16) are depicted in Figs. 4 and 6. We can see that: (i) the ASC\(^{LR}\) is higher than the MSC\(^{LR}\) for all finite (and positive) values of \(Q\), (ii) both functions are decreasing with \(Q\), (iii) both functions tend to the asymptotic value \(t(c + \phi)\) when \(Q\) goes to \(\infty\). Therefore, the production of taxi services presents increasing returns to scale. This makes impossible obtaining a social optimum of the cruising taxi industry (fare equal to marginal operating taxi cost) without subsidies,\(^6\) as has been noticed also by previous authors like Cairns and Liston-Heyes (1996) and Arnott (1996). This conclusion is obtained for an atomized supply of services where many small operators exist, as is the case in the main cities of the developing world and without taking into account congestion externalities. The operation based on bigger taxi companies could change this conclusion if management costs increases considerably with the company size (see Fernández et al., 2003). Also the consideration of congestion externalities could produce upward long run average costs beyond some total number of taxis in the system (see Fernández et al., 2002).

2.10. Average system cost evolution in the long run

We now analyze how the entry of new taxis to the market modifies the ASC function. No matter that this is related to the long run functions derived above we want to show the changes experienced by the short run ASC functions in addition to the decreasing property of the envelope ASC\(^{LR}(Q)\). This is an important element that is used in Section 4.3 to determine the long run market equilibrium.

The following effects are produced by the increase of \(N\), from \(N_1\) to \(N_2\) (see Fig. 4): (i) the industry production capacity increases from \(N_1/t\) to \(N_2/t\), (ii) the AOC\(_{ind}\) shifts to the right, because as the number of taxis in the market increases, the production of runs necessary to finance all the operators also increases; in other words, the average cost of producing a run increases, for any given level of production \(Q\), (iii) as new operators come into the market, the availability of taxis \((N - Qt)\) increases and therefore the average waiting time cost decreases, for any given value of \(Q\). Then, the \(AC_{pas}\) function shifts down and to the right, which means that the average passenger cost decreases as \(N\) increase, for any given value of \(Q\).

\(^6\) See Section 3.1.
Therefore, the following changes are produced in the short run average system cost function (ASC) as $N$ increases (see Fig. 4): (i) the function shifts to the right, (ii) the distance between
the left (decreasing) and right (increasing) branches of the function increases (the new function is wider), (iii) the minimum point of the function decreases.

\[
\frac{d[\text{ASC}(N, Q_{\text{min}}(N))]}{dN} < 0
\] (17)

Finally, it is easy to notice from (11), that a change in the value of taxi operating cost \(c\) will produce a vertical shift of functions \(\text{ASC}(N, Q)\): up if \(c\) increases and down if \(c\) decreases.

Therefore, a different family of such functions (one is shown in Fig. 2) will be obtained for each value of \(c\).

2.11. Waiting time externality in the long run and economies of scale

\(\text{MSC}^{\text{LR}}(Q)\) is lower than \(\text{ASC}^{\text{LR}}(Q)\) because for any finite value of \(Q\), the increase of a marginal passenger produces a positive externality in the long run. According with the optimal long run relation between taxis and passengers, \(N^*(Q)\) (see Eq. (A.2), Appendix A), the increase in the number of passengers generates an increase in the number of taxis (to keep perfect adaptation or optimum capacity). This in turn reduces the waiting time for all existing passengers in the system. The long run expression for the average waiting time cost:

\[
\text{AWC}^{\text{LR}}(Q) = \sqrt{\frac{\theta \cdot k \cdot c}{Q}}
\] (18)

is obtained replacing the optimal long run relation \(N^*(Q)\) in Eq. (7). The expression of the long run waiting time externality \(E_w^{\text{LR}}\), is obtained multiplying by \(Q\) the derivative of the \(\text{AWC}^{\text{LR}}(Q)\) with respect to \(Q\):

\[
E_w^{\text{LR}}(Q) = Q \cdot \frac{\partial \text{AWC}^{\text{LR}}(Q)}{\partial Q} = -\frac{1}{2} \sqrt{\frac{\theta \cdot k \cdot c}{Q}}
\] (19)

The minus sign appears because \(E_w^{\text{LR}}\) represents a (waiting) cost reduction (equivalent to a benefit). At the same time when both \(Q\) and \(N^*(Q)\) increase, the taxi occupancy rate, \(\alpha(N^*(Q), Q)\), also increases:

\[
\alpha(N^*(Q), Q) = \frac{Q \cdot t}{N^*(Q)} = \left[1 + \frac{1}{t} \sqrt{\frac{\theta \cdot k}{c \cdot Q}}\right]^{-1}
\] (20)

From (20) we have that when \(Q\) goes to infinity the occupancy rate goes to 1. The occupancy rate increase, obtained for a perfectly adapted system when \(Q\) increases, produces a second positive externality \(E_{\text{OC}}^{\text{LR}}\), because it makes decrease the average taxi operating cost per run. The amount of this positive externality can be obtained by replacing \(N^*(Q)\) in the average operating cost for the industry (Eq. (5)).

\[7\] It is easy to demonstrate this characteristic analyzing the slope of such trajectory. This result is consistent with the decreasing long run average system cost function obtained in Section 2.9.
\[ \text{AOC}_{\text{ind}}^{\text{LR}}(Q) = \frac{N^*(Q) \cdot c}{Q} = \sqrt{\frac{\theta \cdot k \cdot c}{Q}} + t \cdot c \]

(21)

derivating the resulting expression with respect to \( Q \) and then multiplying it by \( Q \):

\[ E_{\text{OC}}^{\text{LR}} = Q \cdot \frac{\partial \text{AOC}_{\text{ind}}^{\text{LR}}(Q)}{\partial Q} = -\frac{1}{2} \sqrt{\frac{\theta \cdot k \cdot c}{Q}} \]

(22)

Again, the negative sign indicates that \( E_{\text{OC}}^{\text{LR}} \) corresponds to a cost reduction.\(^8\)

Adding both externalities \( (E_{\text{w}}^{\text{LR}} + E_{\text{OC}}^{\text{LR}}) \) we obtain the total value of the positive externality \( E_{\text{LR}}(Q) \), produced in the long run by an additional passenger when perfect adaptation is maintained.

\[ E_{\text{LR}}(Q) = E_{\text{w}}^{\text{LR}} + E_{\text{OC}}^{\text{LR}} = -\sqrt{\frac{\theta \cdot k \cdot c}{Q}} \]

(23)

This value is equal to the difference: \( \text{MSC}_{\text{LR}}(Q) - \text{ASC}_{\text{LR}}(Q) \) (see (15) and (16)), as it should be, according to the economic theory on externalities.

It is interesting that, if \( Q \) would become extremely large (infinity), the waiting time cost for a perfectly adapted system (see expression (18)) would become equal to zero. This is because, both the density of waiting passengers and taxis available, would be so high that a perfect coordination between them would be obtained.\(^9\) Passengers would obtain taxi services without waiting \( (\text{AWC}_{\text{LR}}(\infty) = 0) \) and taxis would produce runs at the maximum theoretical capacity \( q = 1/t \). The positive externalities analyzed above \( (E_{\text{w}}^{\text{LR}} \text{ and } E_{\text{OC}}^{\text{LR}}) \) would have been exhausted and, therefore: \( \text{ASC}_{\text{LR}}(\infty) = \text{MSC}_{\text{LR}}(\infty) = t \cdot (c + \phi) \). The average system cost of service production would be equal to (only) the operator production cost at capacity, plus the travel time cost of the additional passenger. Of course such theoretical limiting value is impossible to obtain in practice.

2.12. Market demand and generalized price functions

As several authors, (Douglas, 1972; De Vany, 1975; Mohring, 1976; Cairns and Liston-Heyes, 1996; Owen, 1984), we assume that the demand for taxi runs is not only a function of the fare \( f \) charged for the service (out of pocket cost). We define a market demand function, \( X(Q) \), which represents the total willingness to pay (of a regular informed user of taxi services) for a taxi run and includes the taxi fare plus the private economic value of the waiting and travel times experienced.\(^{10}\) This function is depicted in Fig. 5. We have assumed a linear formulation.

---

\(^8\) Notice that both externalities have exactly the same expression.

\(^9\) Of course in such case there would be a critical problem of congestion, produced by the infinite number of taxis using the urban road network.

\(^{10}\) Notice that the net out of pocket willingness to pay can be obtained making: \( X(Q) - \text{AC}_{\text{pas}}(N, Q) \), see Section 4.1.
Also, we define the following generalized price function which depends on \( N \), \( Q \) and \( f \).

\[
GP(N, Q, f) = f + AC_{\text{pas}}(N, Q)
\]  

(24)

\( GP(N, Q, f) \) represents the total economic value paid by a passenger for consuming a taxi service when taxi fare is \( f \), there are \( N \) taxis in the industry and the total amount of services produced and consumed is \( Q \). Then, for a level of consumption \( Q \), when waiting and/or travel time cost increases (or decreases), the net out of pocket willingness to pay decreases (or increases) in the same amount.

Two \( GP \) functions are shown in Fig. 5 for \( N_1 \) and \( N_2 \) taxis, where \( N_2 > N_1 \). When the number of taxis \( N \) increases, from \( N_1 \) to \( N_2 \), the generalized price value (or total cost experienced by a user) decreases for any given values of \( Q \) and \( f \). This is because passenger waiting times decrease, reducing the value of the average passenger cost. Therefore, given a same fare value \( f = f' \), it is possible to have many different values of \( Q \) and \( N \) that satisfy the condition \( X(Q) = GP(N, Q, f') \). Points \( e_1 \) and \( e_2 \) of Fig. 5 are two examples. May be this lead Cairns and Liston-Heyes (1996) to argue that “for a fixed fare value \( f \) there are several possible equilibrium values of \( Q \) and \( N \)”. However, as we will see in Section 4 they do not correspond in general to market equilibria, unless important additional conditions are satisfied.

In their analysis of the Hong-Kong taxi system, Yang et al. (2002) find a zero profit curve that gives all the combinations of \( N \) and \( f \) that correspond to possible competitive solutions in a deregulated market. They argue that although, there is no compelling evidence that any combination of fare and fleet solution values in the zero profit curve will occur, it is conceivable that the most probable stable equilibrium occurs at the point where the competitive fleet size is maximized.
3. Social optimum and second best solution in the long run

3.1. Social optimum in the long run

According to basic microeconomic principles, the social optimum will be obtained at the intersection of the long run marginal system cost and market demand functions. Social optimum is shown at point $O$ in Fig. 6. As we can see, it is obtained with $Q^*$ runs and a generalized price, $GP^*$, equal to the $\text{MSC}^{LR}(Q^*) = \sqrt{\theta \cdot k \cdot c / Q^* + t \cdot (c + \phi)}$ (see expression (16)). The corresponding long run average passenger cost is obtained replacing $N^*(Q)$ in Eq. (9):

$$\text{AC}^{LR}_{\text{pas}}(Q^*) = \sqrt{\theta \cdot k \cdot c / Q^*} + \phi \cdot t.$$  

Then, making: $GP^*/Q^* = \text{AC}^{LR}_{\text{pas}}(Q^*)$ we obtain the social optimum fare, $f^* = t \cdot c$. Notice that this value is equal to the minimum average taxi cost per run, obtained when taxis occupancy rate is maximum, $q = 1/t$. As we saw above, this corresponds to a theoretical limiting case obtained only for $Q = \infty$.

For all finite value of $Q^*$, $(q^* = (Q^*/N) < 1/t)$ the social optimum fare will not fully cover taxi operating costs $(t \cdot c < c/q)$. Then, if the social optimum fare $f^*$ is imposed, operators will perceive negative profits, represented by area $L$ in Fig. 6. Notice that the lose per run is $\sqrt{\theta \cdot k \cdot c / Q^*}$, because $\text{AOC}^{LR}_{\text{ind}}(Q^*) = \sqrt{\theta \cdot k \cdot c / Q^*} + t \cdot c$. The value of $L$ can be then obtained multiplying by the number of runs performed, $Q^*, L = \sqrt{k \cdot Q^* \cdot c}$. Given that the number of vacant taxi hours is equal to $N - Q \cdot t$, multiplying this by the operating cost $c$ and replacing $N^*(Q)$ from (A.2), we obtain that $L$ is equal to the cost of vacant taxi hours at the social optimum (Arnott, 1996).

3.2. Second best in the long run

Point $S$ in Fig. 6 represents an industry condition that maximizes social welfare subject to non-negative profits; $(GP_S, Q_S)$ is obtained by the intersection of the demand and long run average cost functions. At this point, fares cover exactly taxi operating costs. A social lose is generated with respect to the optimum social solution $(GP^*, Q^*)$, represented by the area of triangle $SRO$. Therefore, we call “second best solution” to $S$. Using Eq. (A.2) (Appendix A) and the intersection condition: $X(Q) = \text{ASC}^{LR}(Q)$, we obtain the following expressions for the associated optimum fleet size and fare values:

$$N_S = \sqrt{\theta \cdot k \cdot Q_S / c} + Q_S \cdot t$$  \hspace{1cm} (25) \\
$$f_S = \frac{N_S \cdot c}{Q_S} = \sqrt{\theta \cdot k \cdot c / Q_S} + t \cdot c$$  \hspace{1cm} (26)

As we can see from (27) the second best fare, $f_S$, exceeds the value of the social optimum fare, $f^*$, by the amount $\sqrt{\theta \cdot k \cdot c / Q_S}$. As we saw above, this corresponds to the absolute value of the externality $E^{LR}(Q)$ evaluated at $Q = Q_S$. In Section 4.3 (Fig. 9) we show that, in general,

---

11 See footnote number 6.
the second best $S$ is socially preferred to the unregulated free market equilibrium. However, the magnitude of the differences will depend on the characteristics of each system analyzed and there is one case in which both could coincide.

4. Unregulated or free market equilibrium

In order to analyze the possible outcomes of a free market operation, we make the following assumptions: (i) taxi operators quote the fares, (ii) fares quoted may be freely changed but must be clearly displayed in the front of the taxi, for the information of potential passengers, (iii) all taxi operators change fares in the same way.

4.1. Individual taxi and market net willingness to pay for taxi fares

In Section 2.12 a market demand for taxi services, $X(Q)$, was defined as a function of generalized price $GP(N, Q, f)$. The main problem that we face to define market equilibrium is that consumers of taxi services set their demands on the basis of the generalized price experienced, which depends on level of service variables, as waiting time whose value depends on supply–demand interactions at the whole industry level. However, taxi operators care only for fares willingness to pay to set supply conditions. Therefore, in order to define market equilibrium conditions, individual demands faced by taxi operators must be derived from the interaction between total market demands and industry operating conditions.
Fig. 7 presents a typical back to back analysis for the taxi market (industry) and for a representative (average) individual operator. In Fig. 7(b), a system average cost function, $ASC(N, Q)$, that intersects with a market demand function $X(Q)$ at points $E_1$ and $E_2^{12}$ is shown.

In the same Fig. 7(b) the corresponding $AC_{pas}(N, Q)$ and $AOC_{ind}(N, Q)$ functions are also depicted. For each value of total runs consumed $Q$, the market net willingness to pay for taxi fares, $D(N, Q)$, can be obtained by making:

$$D(N, Q) = X(Q) - AC_{pas}(N, Q)$$

$D(N, Q)$ intersects the $Q$ axis at point $G$ and the vertical ($S$) axis at point $C$. Two additional points of interest are $A$ and $B$, that correspond to points $E_1$ and $E_2$ on the market demand $X(Q)$. At those points, for total consumptions $Q_{E1}$ and $Q_{E2}$, the net willingness to pay for taxi fares coincides with the corresponding $AOC_{ind}(N, Q)$value; therefore, fares exactly cover taxi operating costs. Notice that, for all values of $Q$ between $Q_{E1}$ and $Q_{E2}$, net willingness to pay for taxi fares exceeds average operating costs, generating a positive profit. However, for positive values of $Q$ lower than $Q_{E1}$ and higher than $Q_{E2}$ net willingness to pay for taxi fares is lower than operating costs generating a negative profit.

Once the net willingness to pay for taxi fares $D(N, Q)$ has been obtained it is easy to obtain the demand faced by an individual taxi operator. Fig. 7(a) shows a typical individual taxi operator. $AOC_{taxi}(q)$ represents the average cost per run produced as defined in (3). Individual demand $D_{taxi}(N, q)$, faced by a typical taxi operator is obtained dividing by $N$ the number of runs $Q$ corresponding to each willingness to pay value given by the function $D(N, Q)$.

\footnote{The case in which the functions do not intersect corresponds to a situation where the production cost is higher than the willingness to pay for all values of $Q$, and therefore there is no feasible production of taxi services in the long run with free market conditions.}
\[ D_{\text{taxi}}(N, Q) = \frac{D(N, Q)}{N} \]  

It is interesting to note that \( D_{\text{taxi}}(N, q) \) depends on \( N \).\(^{13}\) In Fig. 7(a) we have that points \( A \) and \( B \) that identify the two intersections between demand \( D(N, Q) \) and industry average operating cost \( \text{AOC}_{\text{ind}}(N, Q) \), translate directly to points \( a \) and \( b \) over function \( \text{AOC}_{\text{taxi}} \) and they must also be over demand \( D_{\text{taxi}}(N, q) \). Notice that \( f_{E1} \) and \( f_{E2} \) are the fares related to these points. Point \( C \) corresponds to the maximum net willingness to pay for taxi fares when \( Q = 0 \), and therefore translates to \( q = 0 \) over \( D_{\text{taxi}}(N, q) \). Point \( G \) over the industry demand corresponds to the maximum number of runs demanded \( Q_m \) when \( f = 0 \) and translates to point \( g(Q_m) \) over the demand \( D_{\text{taxi}}(N, q) \).

Individual operators face downward-sloping demands \( D_{\text{taxi}}(N, q) \) typical of imperfect (oligopolistic) competition. As in the normal case of monopoly or oligopoly, it is not possible to define supply functions, because individual operators are not price takers. This situation is typical of oligopolistic markets with product differentiation. The difference is produced here by the different availabilities of the services in time, that require of different waiting times and is similar to the differentiation produced by the location of firms in space, that generate oligopolistic markets (Saltop, 1979). This is consistent with the analysis made in Fig. 1.

4.2. Short run equilibrium

We analyze the possible outcomes obtained from a free market operation in the short run (with \( N \) fixed). Market conditions are represented in Fig. 7. From individual demand \( D_{\text{taxi}}(N, q) \), defined in (28) we can obtain the corresponding marginal income function of a individual taxi operator:

\[ MI_{\text{taxi}}(N, q) = D_{\text{taxi}}(N, q) + q \cdot \frac{\partial (D_{\text{taxi}}(N, Q))}{\partial q} \]  

Then, we apply the usual profit maximizing condition:

\[ MI_{\text{taxi}}(N, q) = \text{MOC}_{\text{taxi}}(q) \]  

Given that we have assumed that taxi costs are fixed (see (2)), then \( \text{MOC}_{\text{taxi}}(q) = 0, 0 \leq q \leq 1/t \). Therefore, individual profit is maximized for \( MI_{\text{taxi}}(q^*) = 0 \), and the optimal fare is \( f = f^* \) shown in Fig. 7(a). The profit maximizing condition for a typical individual operator is represented in the same figure as point \( h = (f^*, q^*) \) over \( D_{\text{taxi}}(N, q) \). Then, \( h \) translates to point \( H = (f^*, Q^*) \) over market net willingness to pay for taxi fares \( D(N, Q) \) and \( E^*(N) \) over market demand \( X(Q) \), which corresponds to industry profit maximizing conditions and, therefore, to a short run equilibrium when the taxi fleet is \( N \).

Notice that a positive profit \( \Delta \) (shown in Fig. 7(a)) is obtained per run produced. The profit obtained by each operator is \( \pi_{\text{taxi}} = \Delta \cdot q^* \) and the industry profit is \( \pi_{\text{ind}} = \Delta \cdot Q^* \). All the profit

\(^{13}\) This is similar to an example mentioned in Nicholson (1997).
maximizing values found \((f^*, q^*, Q^*, A^*)\) depend, however, on the number of taxis considered \(N\) (short run conditions).

In practical terms, taxi operator tests different fares where \(f_{E1} < f < f_{E2}\) (see Fig. 7a) in order to find the one that maximizes his profits. All taxi operators change fares in the same way. After one taxi change its fare all others follow such change.

4.3. Long run equilibrium

The analysis of the previous section assumed that the number of taxis \(N\) is fixed. Now we consider long run conditions. Therefore, individual operators can enter or exit the market and \(N\) will be variable. The normal microeconomic argument is that long run capacity changes are motivated by short run profits or loses. According to this, new taxis will enter the system if there are profits.

In real life situations if the market is not regulated (free fares and free entry\(^{14}\)) both short and long run adjustments will simultaneously operate until long run equilibrium is obtained. Fig. 8 shows such combination of short and long run market adjustments.

In Fig. 8, let us suppose that point 1 corresponds to a short run equilibrium with \(N_1\) taxis operating in the market and obtaining positive profits, represented by the shaded area \(A_1\). \(A_1\) represents profits because \(A_1 = \{X(Q_1) - ASC(N_1, Q_1)\} \cdot Q_1 = \{f - AC_{pas}(N_1, Q_1)\} \cdot Q_1 > 0\) This will incentive new taxis to enter the market. As the number of taxis increases from \(N_1\) to \(N_2\), market conditions will be modified as follows:

(i) The average system cost function shifts to the right (this kind of shift is represented in Figs. 4(b) and 9). In Fig. 8, this is represented through the movement from \(ASC(N_1, Q)\) to \(ASC(N_2, Q)\). This shift ends when ASC intersects the demand function at point 1. As the ASC function shifts to the right, profits obtained by operators (at points 1 and 2) are reduced because the vertical distance with the average system cost decreases. When the number of taxis in the market gets to \(N_2\), profits get to zero at point 1 and negative at point 2.

(ii) Simultaneously with ASC movement, the generalized price value decreases (see Fig. 5) because passengers waiting time is reduced. In Fig. 8, segment 1–2 represents such change; assuming that the number of runs \(Q_1\) and the fare value \(f_1\) stay temporarily constant.

(iii) The generalized price reduction will incentive an increase in the number of runs consumed. In Fig. 8, segment 2–3 represents a quantity adjustment produced because when the number of taxis increases from \(N_1\) to \(N_2\) the generalized price paid by passengers is reduced from \(GP(N_1, Q_1, f)\) to \(GP(N_2, Q_1, f)\). Therefore, for the new generalized price, users of taxi services are willing to consume a higher number of runs: \(Q_3 > Q_1\).

Once point 3 is reached, taxi operators will test new fares in order to find the one that maximizes their profits in this new situation, with \(N_2\) taxis. This will lead to a new short run equilibrium with positive profits which will incentive new taxis to enter to the market, again.

In Fig. 9 we show that, while profits exist, the adjustment mechanisms described will operate systematically and make increase the number of taxis. We can see that a number of ASC and MSC functions are shown for different taxi fleet sizes, \(N_1 < N_2 < N_3\). The corresponding short

\(^{14}\) Such free market conditions existed in the city of Santiago de Chile during the period 1990–1998.
Fig. 8. Combination of short and long run market adjustments.

Fig. 9. Change of market conditions in the long run.
run equilibra are obtained from the intersections between MSC and the MI. MI(Q) is directly obtained from X(Q) (in Appendix B is demonstrated that the maximizing condition of a taxi operator, Eq. (30), is equivalent to MI = MSC). As the number of taxis increases, point $E^*(N_i)$, where $i = 1, 2$ and $3$, moves down over the demand function; that is, the number of runs $Q_{E^*(N_i)}$ increases and the generalized prize decreases. When the number of taxis is equal to $N_3$, the average system cost function, ASC($N_3$, $Q$), is tangent to the demand function $X(Q)$. The tangency point is the short run equilibrium related to $N_3$ fleet size, that is, $E^*(N_3)$. In the following, the capital letter $T$ and the corresponding number of taxis, $N_T$, and runs, $Q_T$, will identify this tangency point. For $N$ higher than $N_T$ there is no feasible long run market solution, because operators will experience loses for any value of runs performed. Then $N_T$ corresponds to the maximum number of taxis that can operate in the long run, under free market conditions. Point $T$ is the long run market equilibrium, as long as costs and demand do not change, because the market does not bear more taxis, users are satisfied with the price paid for the quantities consumed and there is no incentive to change the current fare.

Fig. 10 represents long run equilibrium conditions both for the market an for the typical taxi operator. The existence of positive short run profits will make increase the number of taxis until $N = N_T$ ($N_3$ in Fig. 9), driving the market to the condition represented by the tangency point $T$. Then, industry demand $D(N, Q)$ becomes tangent at point $T'$ to $AOC_{ind}(N_T, Q)$ and individual demand $D_{taxi}(N, q)$ is tangent at point $T_{taxi}(q = q_T)$ to the average taxi operating costs function $AOC_{taxi}(q)$. It is interesting that equilibrium conditions correspond to a monopolistic competition equilibrium: $T_{taxi} = (f_T, q_T)$ maximizes operators profit with $\pi_{taxi} = 0$: $MI_{taxi}(q_T) = MOC_{taxi}(q_T)$ and $f_T = AOC_{taxi}(q_T)$.

### 4.4. Comparing unregulated equilibrium and second best solutions

Characteristics the average system cost functions, allow us to conclude the following (see Fig. 9):
(i) The number of taxis in operation is in general higher for the long run free market equilibrium $T$, than for the second best solution $S$: $N_T \geq N_S$ (30,780 vs. 20,367 for the case of Santiago, see Fernández et al., 2001).

(ii) The number of runs produced and consumed is higher for the second best solution than for the private market equilibrium: $Q_S \geq Q_T$ (45,330 vs. 31,848 for Santiago).

(iii) The average taxi occupancy rate is higher for the second best solution than for the private equilibrium: $a(N_T, Q_T) \leq a(N_S, Q_S)$ (28% vs. 55% for Santiago). Considering that $a(N_T, Q_T) = Q_T t / N_T$ and $a(N_S, Q_S) = Q_S t / N_S$, this is directly obtained from (i) and (ii) because: $(Q_S c / N_S) > (Q_T c / N_T)$.

(iv) Operators profits are zero for both points $T$ and $S$, but the fare charged is lower for the second best solution than for the private equilibrium $f_S \leq f_T$ (US$1.7$ vs. US$3.6$ for Santiago). Considering that $f_T = N_T c / Q_T$ and $f_S = N_S c / Q_S$, this is directly obtained from (i) and (ii) because: $(N_S c / Q_S) < (N_T c / Q_T)$.

(v) At the second best solution, average system cost is lower than at the equilibrium $T$: $\text{ASC}(N_S, Q_S) \leq \text{ASC}(N_T, Q_T)$.

(vi) The generalized price is lower at $S$ than at $T$: $\text{GP}(N_S, Q_S, f_S) \leq \text{GP}(N_T, Q_T, f_T)$.

(vii) The social welfare is higher at $S$ that $T$. Considering that taxi operator surplus is zero, social welfares is equivalent to consumer surplus. In Fig. 9, area $X(0) - S - B$ represents consumer surplus related to $S$. On the other hand, area $X(0) - T - A$ represents consumer surplus related to $T$. We can see that area $X(0) - S - B$ is higher than $X(0) - T - A$.

In their case study for Hong-Kong, Yang et al. (2002) found that the long term competitive solution and the second best solution were very close to each other. It is easy to see from Fig. 6 that the difference between both solutions will depend on the relative positions of the demand $X(Q)$ and long run average system cost functions ASC$^{LR}(Q)$. If we shift function ASC$^{LR}(Q)$ upwards relative to $X(Q)$, we could obtain, in the limit, the tangency of them in which case $T$ and $S$ coincide. Therefore, the efficiency of the competitive solution will depend on the characteristics of each case studied. This important general result is not possible to obtain from the Yang et al. case study because demand functions are implicit.

5. Regulatory policies

There has always been a general agreement among analysts that free market conditions do not produce a social optimum solution in the operation of a cruising taxi system. In particular, some authors have questioned that an unregulated market equilibrium can be obtained (Shreiber, 1977; Cairns and Liston-Heyes, 1996). Therefore, different propositions have been made in the literature with respect to regulatory policies. An important conclusion from our model is that although in general the second best $S$ is socially preferred to the unregulated free market equilibrium $T$ (see Section 4.4) the difference between them and therefore the social losses generated at $T$ will depend on the specific case studied. In particular there is a limiting case in which both solutions coincide (tangency of the market demand $X(Q)$ and long run average system cost functions ASC$^{LR}(Q)$). This finding takes away generality to the previous agreement on the need for price and entry regulations. Therefore, the need for regulation should be carefully considered case by case.
In this section we analyze the regulations most commonly proposed and used in practice, in order to find their social convenience; with that purpose we make use of the model and conclusions previously obtained.\(^\text{15}\)

### 5.1. Fare and fleet size regulation

This policy has been recommended by several authors and widely used in taxi markets. Given that a social optimum requires subsidies (Section 3), we assume that the objective of regulation is obtaining a second best solution like \(S\) (see Fig. 6). According to the analysis of Section 3, \(S\) can be obtained if generalized price is set to a value given by \(GP(N_S, Q_S, f_S)\), equal to the long run average system \(ASC^{LR}(Q_S)\) (see Fig. 6). Expressions (25) and (26) give the fleet size and fare related to the second best solution.

However, from (24), \(GP(N_S, Q_S, f_S) = f_S + AC_{pas}(N_S, Q_S)\); therefore \(S\) will be obtained only if, in addition to setting fares to \(f_S\), passengers experience an average cost equal to (see (9)):

\[
AC_{pas}(N_S, Q_S) = \phi \cdot t + \frac{\theta \cdot k}{N_S - Q_S} \cdot t
\]

Therefore, to get \(S\), starting from a different market condition, it is necessary not only fixing \(f = f_S\), but also making that \(N = N_S\). However, once \(S\) is obtained, with \((f_S, N_S)\), the operating conditions will be stable, as long as the value of \(f_S\) is maintained. At \(S\) profits are zero and the entrance of new taxis will produce losses to operators. Therefore, once the number of taxis is \(N_S\), it is enough fixing fares at \(f_S\), to maintain second best conditions \(S\). If \(N\) increases beyond \(N_S\), fare \(f_S\) will not cover taxis operating costs, \(AOC_{ind}(N, Q_S)\), because \(AOC_{ind}(N, Q_S) > AOCS_{ind}(N_S, Q_S)\) if \(N > N_S\) (see Eq. (5)).

Fig. 11, shows how a market shift from \(T\) to \(S\) can be produced. Starting from \(T\), with \((N_T, Q_T, f_T)\), if both fare and fleet size are changed to \(f_S\) and \(N_S\), a reduction of the generalized prize is obtained, from \(GP(N_T, Q_T, f_T)\), corresponding to point \(T\), to \(GP(N_S, Q_T, f_S)\), corresponding to point 1 in Fig. 11. Notice that \(GP(N_S, Q_S, f_S)\) goes through \(S\), for \(Q = Q_S\) but goes through 1, for \(Q = Q_T\). But for that value of GP, taxi users are willing to consume \(Q_S\) runs instead of \(Q_T\). Therefore an adjustment in quantity consumed, represented by segment 1 – \(S\), will take place taking market conditions to \(S\), where they will stay as long as fare value \(f_S\) is enforced. Notice that the average system cost function \(ASC(N_S, Q)\) goes through point \(S\), for \(Q = Q_S\) and therefore zero profits are experienced at that point.

If starting again from \(T\), the fare value is changed from \(f_T\) to \(f_S\), but fleet size remains equal to \(N_T\) \((N_T > N_S)\), the generalized prize value will experience a reduction from \(GP(N_T, Q_T, f_T)\), to \(GP(N_T, Q_T, f_S)\), which corresponds to point 2 (below 1). Notice that \(GP(N_T, Q_T, f_S)\) is lower than \(GP(N_S, Q_T, f_S)\), because given that \(N_T > N_S\) waiting times are lower at point 2 that at point 1. For the new value of GP, taxi users are willing to consume \(Q_R\) instead of \(Q_T\); therefore an adjustment

\(^{15}\) The importance of obtaining \(S\) instead of \(T\), will depend on the magnitude of the social lose. In a study of the Santiago (Chile) taxi system (see Fernández et al., 2001), it was found that losses associated to condition \(T\), compared with \(S\), were highly significant (43%). However for the Hong-Kong case study (Yang et al., 2002) these loses should be small given that \(T\) and \(S\) are close.
of quantity consumed, represented by segment 2 – R will occur, taking market conditions to R. However, at R, the $N_T$ taxis experience losses represented by shaded area L. Therefore if fares are forced to remain equal to $f_S$, some taxis will exit the market until the increase in occupancy rate eliminates losses. As taxis exit the market, the fleet size reduction will make the average system cost function shift to the right from $ASC(N_T, Q)$ and the generalized price function will move up from $GP(N_T, Q, f_S)$. Losses will vanish (zero profits) when the average system cost function gets to $ASC(N_S, Q)$ and the generalized price gets to $GP(N_S, Q, f_S)$; both functions intersect at S. Therefore, second best conditions S can be obtained without fleet size regulation; fare regulation can make the job, given that operators losses produced (at R) should make some taxis to exit the market until the fleet size gets to $N_S$. This means that entry can be kept free, which makes the regulation policy easier to apply.

It is important noticing that taxi operating cost $c$ is directly related to the quality of service offered: type of car used, driver qualifications, maintenance standards, etc. Also notice that $ASC(N_S, Q)$ and the position of point S are directly influenced by the value of $c$ (see Eqs. (11) and (15)). Therefore, practical implementation of regulation requires setting up the service quality standards that are assumed in the value of $c$ used to calculate the regulated fare $f_S$. Otherwise, operators can reduce $c$, through quality of service reductions; then, if $N_S$ is fixed operators will obtain profits at $(N_S, f_S)$. Then if entry is free, additional taxis will enter the industry until profits disappear; however the industry will operate at a different point $S'$, with more taxis, more runs and less service quality, than those corresponding to the values used to calculate the regulated fare $f_S$. This is a point not recognized before in the literature.
5.2. Fleet size regulation

Although fleet size regulation alone is not usual in real systems, the Chilean transport authority under the pressure of taxi operators established a fleet size freezing for the period 1998–2005. This finished with a previous period in which the system operated under free market conditions (free fare and free entry). We analyze here the consequences of this policy.

It is easy to see (Fig. 11) that starting from market conditions different to $S$, the sole application of fleet size regulation will not be enough to take the market to $S$. Assume that we fix fleet size to $N_S$; then, unless fares are simultaneously fixed to $f_S$, market condition $S$ will not be obtained. Let us assume again that market conditions are initially at $T$; if fleet size is reduced to $N_S$, generalized price will increase from $GP(N_T, Q_T, f_T)$ to $GP(N_S, Q_T, f_T)$ (corresponding to point 3 in Fig. 11). This is because the fleet size reduction ($N_S < N_T$) will make passenger waiting times increase. Market conditions will then shift from $T$ to 3. However, for the generalized price value corresponding to point 3, taxi users are only willing to consume $Q_4$ runs; therefore, the corresponding adjustment of quantity consumed will take market conditions to point 4 (away from $S$). Once system has reached this condition, taxi operators will quote their fares in order to maximize their profits. Thus, market will reach an equilibrium with fleet size regulation. This is the short run equilibrium with $N_S$ fleet size and it is shown in point $E^*(N_2)$ in Fig. 9 (fleet size $N_2$ corresponds to $N_S$).

In general fleet size regulation is equivalent to maintaining short run conditions for a given fleet size value. Therefore, in such case the short run equilibrium analysis of Section 4.2 applies. It is important noticing that market conditions represented by point $E^*(N_2)$ (see Fig. 9) are socially worse than those corresponding to point $T$, because $E^*(N_2)$ is located to the left of $T$. After reaching $E^*(N_2)$ there are no incentives to move down towards $S$. This makes an important difference with the case of price regulation that as we saw in the previous section is enough to obtain and maintain second best conditions.

Given the form and evolution of functions $ASC(N, Q)$ as $N$ increases (see Fig. 9), and the incentives to rise fares in the short run, analyzed in Section 4, we postulate that the short run equilibria obtained for $N < N_T$ will be always to the left of point $T$ over $X(Q)$. Actually they will tend to $T$ as $N$ tends to $N_T$. Therefore we conclude that fleet size regulation alone will be socially worse than a free market policy (free entry and free fares) when the service is produced by many small operators.

5.3. Licensing policy

Two ways of implementing a licensing policy are the following: (i) The authority (regulator) charges a license (a given yearly amount) to those taxis willing to enter the system, or (ii) a licenses market (medallions) is implemented as a complement of fleet size regulation.

For the first case let us assume a market equilibrium $T$, with $(N_T, Q_T)$, obtained assuming taxi operating costs $c$ (Fig. 12). Adding the license $l$ increases taxi operating costs from $c$ to $c' = c + l$. This produces an upward shift of average system costs functions $ASC(N, Q)$ (see Eq. (11) and Fig. 4) and generates a new family of relevant functions $ASC^L(N, Q)$, with operating costs $c'$. Therefore, without additional regulations, a new free market equilibrium $T^L$, will emerge at the tangency point between demand $X(Q)$ and the new average system cost function $ASC^L(N_{TL}, Q)$. At $T^L$, fleet size $N_{TL}$ must be lower than $N_T$, because some taxis will exit the system induced
by the losses produced in the short run (Fig. 12), after the license introduction. The exit of taxis will take place until losses are eliminated. Given the negative slope of the demand function, the upward move of the average system cost functions from ASC to ASC$^L$, produced by the license cost, and the shift to the left (from ASC$^L(N_T, Q)$ to ASC$^L(N_{TL}, Q)$) produced by the fleet size reduction (see Fig. 4), the new number of runs $Q_{TL}$ will be also lower than the initial equilibrium value $Q_T$.

For the second case, let us assume that fleet size is regulated to $N_{TL}$, with $N_{TL} < N_T$. This can be implemented by giving special operating permits (licenses) to $N_{TL}$ taxis. Then, each license will have an economic value equal to the present value of the profits obtained from the taxi operation. In order to obtain an efficient allocation of available licenses, among potential taxi operators, a licenses market can be established. Maintaining a license will therefore have a financial cost for the operator, making the corresponding operating cost increase. This will shift upwards the average system cost function ASC($N_{TL}, Q$), until it becomes tangent to the demand function at a point $T^L$. ASC$^L(N_{TL}, Q)$ is the average system cost function that includes the license cost and is tangent to the demand at $T^L$.

In both cases analyzed the introduction of the license produces a reduction of taxis in the system and a reduction in the number of runs performed. The equilibrium fare is always equal to the

---

16 For free entry conditions licenses will have no value. Nobody will pay for something that can be obtained for free. Here is also valid the comment with respect to the service quality requirements, made above; otherwise there will be an incentive to reduce $c$ in order to increase profits.

17 Notice that such profits exist because fleet size is restricted by regulation.
difference between the total willingness to pay \( X(Q) \) and the AC_{pas} (see Eqs. (9) and (24)) experienced by taxi passengers.

The difference between the values of ASC\(^L\)(NTL, QTL) and ASC(NTL, QTL), multiplied by the number of runs QTL, corresponding to point \( T^L \), represents the cost of financing the NTL licenses for the operating period analyzed (LC in Fig. 12). The corresponding cost per license can be obtained dividing this value by NTL. Notice that the indicated total financial cost is equal to the profits that the NTL operators would obtain if they produce and sell QTL runs with an average system cost given by ASC(NTL, QTL); in other words with a zero license value \( c' = c \). Finally, it is reasonable that a license market emerges if fleet size is regulated and operators obtain profits.\(^{18}\) Therefore, the introduction of licenses will move the market in the same direction (higher price and lower runs than \( T \)) as fleet size regulation. Notice that, licenses have an economic value only when fleet size is regulated with \( N < N_T \). However the introduction of licenses has the important consequence of producing a unique equilibrium for short run conditions (fixed fleet size), similar to that obtained under long run conditions (demand tangent to average system cost function).

6. Final comments, conclusions and extensions

A diagrammatic approach has been developed to analyze the characteristics of the cruising taxi market. This allowed us to represent and analyze different operating conditions for the cruising taxi system. System conditions are described in terms of number of taxis in operation, number of runs produced, occupancy rates, fares charged, average production costs and generalized prices. We show that for a perfectly adapted system the long run average system cost function is always decreasing, because, the increase of a marginal passenger produces two positive externalities in the long run: waiting times are reduced for all passengers and the average taxi operating cost per run is also reduced.\(^{19}\) because occupancy rate increases.

The main contribution to the literature is the result obtained about the existence of unique long run equilibrium for free market conditions, which corresponds to monopolistic competition equilibrium. Also, we describe how such equilibrium is obtained from the interactions between industry demand and supply conditions.

In previous analysis, Cairns and Liston-Heyes (1996) using a model of search, where drivers and riders search for each other, conclude that equilibrium of a deregulated industry does not exist. Yang et al. (2002), from the conventional transportation network standpoint, analyze the supply–demand equilibrium characteristics of cruising taxi services for the case of Hong-Kong. They find a free market equilibrium similar to our result considering a fix fare. Our result should not be a surprise, because there is a clear analogy between this market, where alternative supplies of

\(^{18}\) This is exactly what happened in Santiago de Chile when last year the taxi fleet size was frozen, after years of free entry. A license market was established and some 1000 taxi operators left the market during the last year (2001).

\(^{19}\) This is different in the short run. The short run average system cost function has a U form with an initial decreasing part and a final increasing section (see Fig. 1). When the number of taxis is fixed, the increase in a marginal passenger produces a negative externality because waiting times increase and a positive externality because taxi occupancy increases. For low values of \( Q \) the former is less important than the latter and vice versa for high values of \( Q \).
services are separated in time (waiting time necessary to choose a different service supplier) and spatial markets with monopolistic competition\(^{20}\) (Salop, 1979).

We use the diagrammatic analysis to show the differences between social optimum and free market equilibrium. Because the social optimum fare produces losses to taxi operators we consider a second best solution \(S\) that maximizes social welfare subject to financial constraints. An important conclusion of our analysis is that although in general the second best \(S\) is socially preferred to the unregulated free market equilibrium \(T\) (see Section 4.4), the difference between both of them and therefore the social losses generated at \(T\) will depend on the specific case studied. In particular there is a limiting case in which both solutions coincide (tangency of the market demand and long run average system cost functions).

We also show that the maximum fleet size is obtained for free market conditions, and the maximum number of runs produced corresponds to the social optimum.

Finally, we show that, if regulation is necessary because of big differences between \(T\) and \(S\), second best solutions can be obtained by price regulation alone and that entry regulation is in such case redundant. On the contrary, entry regulations (fleet size freezing, without price regulation), produce in general worse system conditions than free market \(T\). We argue that a licenses market will naturally appear when fleet size is regulated and show that the introduction of licenses will in general produce a similar result as fleet size regulation.

We also conclude that when fare is regulated, it is necessary to impose a level of service requirement consistent with the taxi operating cost implicit in the fare charged.

A logical extension of the analysis presented in the paper is the introduction of congestion, which will modify the cost functions considered. It would also be interesting to apply the analysis to other taxi operations modalities (radio taxis).

**Acknowledgment**

The results presented in this paper where obtained in a research project funded by FONDECYT of the Chilean Government and the Pontificia Universidad Católica de Chile.

**Appendix A**

We derive the optimum relation between the number of runs and taxis \(N^*(Q)\), in order to have a perfectly adapted system. \(N^*(Q)\) is obtained minimizing the total system cost \(TSC(N, Q)\) with respect to \(N\):

\[
\frac{\partial (TSC(N, Q))}{\partial N} = \frac{\partial (ASC(N, Q) \cdot Q)}{\partial N} = c - \frac{\theta \cdot k \cdot Q}{(N - Q \cdot t)^2} = 0
\]  
(A.1)

From (A.1) we obtain the long run optimal relation between \(N\) and \(Q\):

\[
N^*(Q) = \sqrt{\frac{\theta \cdot k \cdot Q}{c}} + Q \cdot t
\]  
(A.2)

\(^{20}\) It is usual the analysis of monopolistic competition by means of a spatial analogy.
Now, replacing \( N^*(Q) \) in expressions (11) and (12) we obtain the long run average and marginal system cost functions (see expressions (15) and (16)).

Appendix B

In this appendix we demonstrate that profit maximizing condition (see (30)) is equivalent to:

\[
\text{MI}(Q) = \text{MSC}(N, Q)
\]  

**Demonstration:** Condition (30) can be expressed as:

\[
\frac{\partial (D_{\text{taxi}}(q))}{\partial q} \cdot q + D_{\text{taxi}}(q) = 0
\]  

Considering expression (28) we can rewrite (B.2) as follows:

\[
\frac{\partial (X(Q) - AC_{\text{pas}}(N, Q))}{\partial Q} \cdot Q + X(Q) - AC_{\text{pas}}(N, Q) = 0
\]  

Reordering this expression:

\[
\frac{\partial (X(Q))}{\partial Q} \cdot Q + X(Q) = AC_{\text{pas}}(N, Q) + \frac{\partial (AC_{\text{pas}}(N, Q))}{\partial Q} \cdot Q
\]  

Left side is equal to \( \text{MI}(Q) \) function, while right side is equal to \( MC_{\text{pas}}(N, Q) \). This function is similar to \( MSC(N, Q) \) (see expression (12)).

References


