THE ECONOMICS OF TAXICAB REGULATION: A WELFARE ASSESSMENT

Jeremy P TONER
Research Fellow
Institute for Transport Studies
University of Leeds
Leeds - Great Britain

Peter J MACKIE
Senior Lecturer
Institute for Transport Studies
University of Leeds
Leeds - Great Britain

This work was sponsored by the Economic and Social Research Council, Award RD00232345

1. INTRODUCTION

The taxi industry has a long history of regulation. As early as 1635, hackney carriages in London and Westminster were restricted as to what they could and could not do because they congested the streets, and the number of vehicles was limited "to restrain the excessive and superfluous use of coaches". The scope of regulation continued apace over the years, being largely concerned with public order and protection rather than the economics of the industry. In recent years however, economists have turned their attention to the taxi industry. The result has been to establish the optimal economic performance of the trade as a purpose of regulation, alongside public safety, consumer protection and congestion management. Each of these involves the regulation of different aspects of the trade as outlined below.

Public safety involves protecting the consumer and other road users in matters concerning the physical safety and suitability of the vehicle. It implies regulation of vehicle standards (such as size, age, road-worthiness checks, identification) and driver standards ("fit and proper person").

Consumer protection is concerned with preventing operators in a strong bargaining position from exploiting weak consumers. For example, a cab driver with a vacant cab and no other vehicle in sight could, for a given journey, charge more than for an identical trip where the consumer had a choice of cabs. Or, if fares are regulated to prevent this happening, the driver could refuse a particular hiring because it was not sufficiently lucrative, having a lower than average chance of a balancing return hiring. The appropriate regulatory measures in these circumstances are compulsory metering and the setting of maximum fares and/or fare registration plus the advertisement of the fare on the side of the cab; the licensing of drivers and the plating of cabs; and compellability.

Congestion. There may be wider traffic management reasons for wanting to regulate taxis. For example, in a congested town centre it may be useful to prohibit cruising and impose return to rank rules. This requires a rank designation policy which can cope with the number of cabs and possibly a restriction on the number of licences issued.

Economic performance. There are a number of feasible price/service level combinations which could be sustained in a given taxi industry, but it is often argued that an unregulated market will not attain the best, benefit maximising one. Regulators may want to control maximum fares, cab numbers and taxi ranks in order to be able to push the market towards the best state.

The taxi market in the UK is large and growing. Consumer expenditure, at £600 million per annum, is now about half the expenditure on bus travel. For many years, local authorities outside London had absolute discretion to regulate taxicab numbers in their areas, and to set the maximum tariff which in practice operated as the actual tariff. In 1985, however, the Transport Act removed the absolute discretion to determine cab numbers, requiring authorities to issue new licences to suitable applicants unless they were satisfied that there was no "significant unmet demand" for taxi services. This new law has stimulated a need for regulators to determine whether the public is best served by an open entry policy. As an indication, about 20% of districts chose not to restrict entry before 1985, and by 1989, these had been joined by a further 20% which had deregulated entry since the Transport Act. 40% of districts had issued a limited number of new licences, but the remaining 20% had issued none at all. In general, the most rural districts chose not to control numbers, while the most urban authorities, with the largest numbers of taxis, adopted the most restrictive policies. In one third of districts, a significant licence premium or "plate value" accrues to incumbent owners on the sale of their plated vehicle.

This paper focuses on the economic performance arguments for regulation. In the next section, we set out a model of the taxi trade so as to define the problem more closely. In section 3, empirical evidence on the key parameters - values of waiting time and price and service elasticities - is presented. In section 4, the model is used together with best estimates of the elasticities to predict and compare the consequences of alternative regulatory strategies. Finally, the policy implications are briefly discussed.

2. A MODEL OF THE TAXI MARKET

The nature of markets is usually such that the location of an equilibrium between goods offered and goods demanded is identified by indicators such as excess demand or surplus stocks. If the supply of a particular commodity exceeds the demand for it at current prices, then producers will reduce the price of the good in order to eliminate their surplus stocks. Alternatively, if demand for a good exceeds the supply, consumers will bid up the price in order to be sure of obtaining the good. In equilibrium, unwanted stocks and bidding up will not occur. However, the provision of taxi services is a special case; the production and consumption of taxi trips requires time as well as material resources. On the demand side, this means potential taxi users will consider both the availability of taxis and the fare they will have to pay. For suppliers, a taxi firm will consider not only the price of a trip and its cost of production, but also the rate of utilisation of the vehicles. There is also a relationship between the rate of utilisation and the availability of cabs. This interrelationship between supply and demand is not considered in conventional models of the market and necessitates the development of a model which incorporates these features.

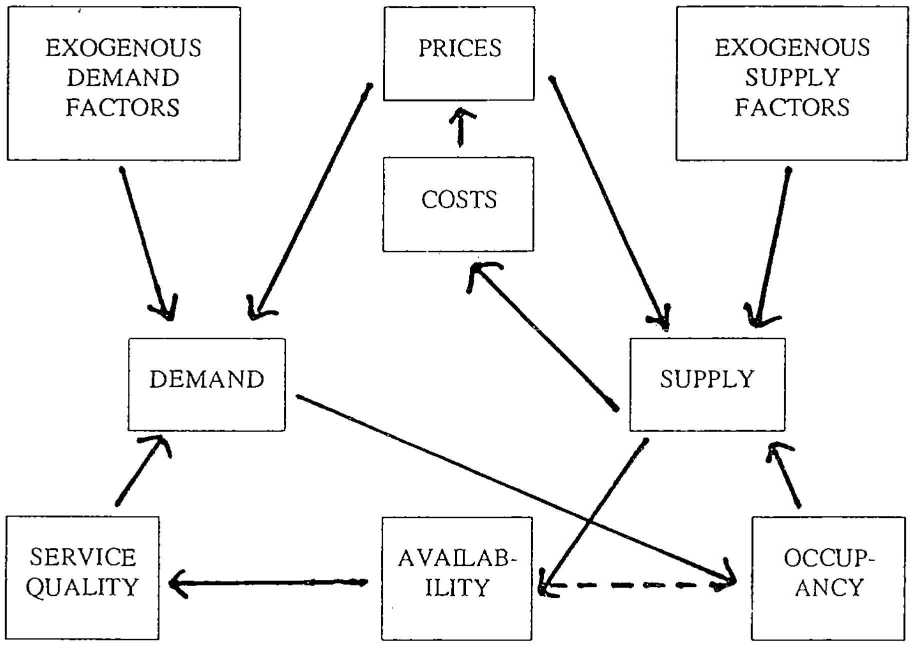

Figure 1: the taxi market. (Adapted from Manski & Wright (1976))

Another feature of the market which renders conventional supply and demand analysis inappropriate is that quantity supplied (rides offered or cab hours supplied) is not the same as quantity demanded (rides taken plus unmet demand). Generally, when a cab drops a passenger at his destination, there is unlikely to be another passenger waiting there to be carried at the time the empty taxi is generated. If there is someone waiting at the time an empty taxi is created, the chances are that the potential customer would not be in the same place as the taxi. It is therefore necessary for some "excess capacity" to be built into the system, for some taxis to run empty until supplier and consumer meet in time and space. This excess capacity needs to be paid for; it suggests that the break-even price of a trip should be higher than the cost of supplying a trip such that the price times the occupancy rate is equal to the cost.

The above features of the market have been taken on board by a number of authors who have then attempted to develop a theory of taxicabs. Much of the theoretical literature presupposes a pure cruising trade where cabs roaming the streets are hired by customers hailing them. While this is an over–simplification, given the existence of taxi ranks and radio–fitted vehicles which can be hired by telephone, it is a useful starting point to develop a theory and we present below an outline of the workings of the taxi market.

For simplicity, we assume in the first instance a taxi market with fares set by a regulatory authority, but with free entry. We shall assume further (following Douglas (1972)) that the fare is an average fare for a journey of average length and duration. We can thus use the demand for engaged cab hours, Q, as a proxy for the demand for trips. The total demand for occupied taxi hours during some reference period, say one hour, is dependent upon the trip price and the delay time encountered by a passenger wanting a taxi. Thus,

$$Q = f\,(P, T), \quad \text{with } \partial f/\partial P \text{ and } \partial f/\partial T < 0 \,,$$

with P representing the price charged for an engaged cab hour and T the mean waiting time facing a passenger.

On the supply side, cab operating costs are assumed to be related to the number of hours a cab is in service, with no distinction made between engaged and unengaged time. This is a reasonable assumption for a cruising trade, although for a rank–based trade, idle time at rank would presumably have slightly lower costs. If V is the number of cab hours offered but not taken, then the total supply of cab hours is

$$N = Q + V \,.$$

In a market with free entry and characterised by inter–firm competition, there will be entry of new firms into the market until excess profits are zero; that is, industry revenues just cover total costs. If, for the time being, we define total costs as

$$TC = h\,(Q, T) \quad \text{with } \partial h/\partial Q > 0 \,,\ \partial h/\partial T < 0$$

then,

$$PQ = h\,(Q, T)$$

The quality of service, or the time costs incurred by passengers, depend on the availability of cabs which is determined by the number of vacant cabs in the area at the time the trip is demanded. Note that it is the number of vacant cabs, not the proportion; if trips and cabs (and therefore vacant cabs) increased by the same proportion, then waiting times would fall. These user economies of scale are common in other sectors involving a service delay, and in transport are referred to as the Möhring effect. Here, we simply assume

$$T = g(V) \quad \text{with } \partial T/\partial V < 0 .$$

If we assume that costs are exogenously determined, then there exists a feasible price set P_f^* for which a solution exists. Our problem is to identify the optimal price

$$p_i^* \in P_f^*$$

to yield the solution vector (P_i, Q_i, V_i, T_i) . The system is depicted in figure 2. The contours of the function are demand curves at different expected passenger waiting times $(T_1<T_2<T_3<T_4)$. The horizontal distances between the contours represent elasticities of demand with respect to service quality. The c_i's are loci of equilibrium points at different revenues per hour of operation and, since this is an equilibrium model, different cost structures. An exogenously-determined cost structure, say c_b , determines a set of efficient prices at which solutions exist.

$$P_f^* = \{ p_i \in P \mid p_{min} \leq p_i \leq p_{max} \}$$

Under free-entry conditions, and in the absence of price regulation, the industry should expand to its greatest sustainable size. The question of interest is whether this is the optimal state of things. A suggested "pragmatic" solution to defining an optimum might be to maximise net social benefit subject to a break-even constraint, as suggested by Douglas (1972), although in a first-best world, a subsidy to taxi operators to enable consumers to avail themselves of the benefits of the quasi-Möhring effect described earlier could potentially be optimal. So, let us

$$\text{maximise} \quad W = \int P(Q,T)dQ - h(Q,T)$$

$$\text{subject to } PQ = h(Q,T)$$

where $P(Q,T)$ is the inverse function of $Q = f(P,T)$. Under certain conditions, welfare as specified is maximised at the point of maximum ridership; if, for example, demand is linear in P and T, or if a generalised cost formulation is valid. This latter case implies that every consumer has the same value of time, and the optimal solution is also that which minimises the generalised cost faced by each consumer. If we assume a

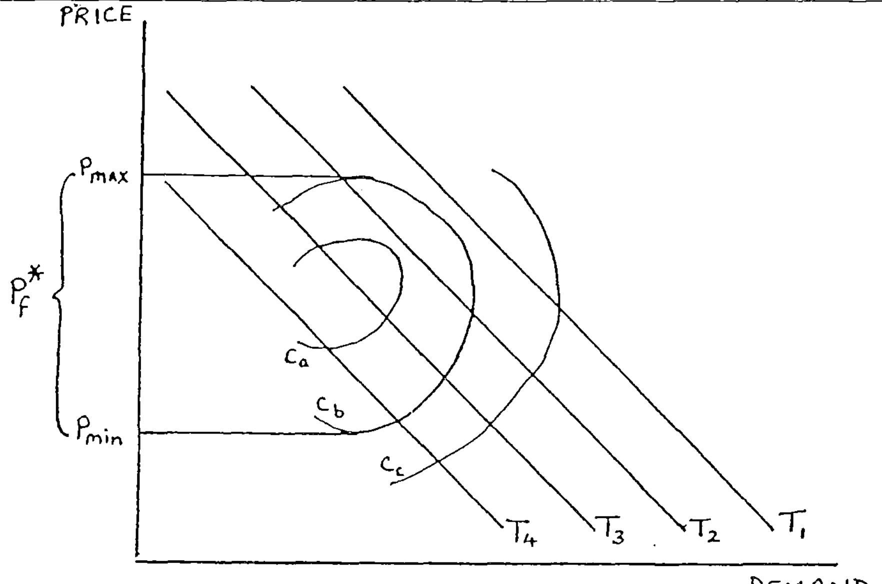

Figure 2: An equilibrium model of the taxi market.

population with heterogeneous values of time, then a single unique optimum remains undefined. We can define boundary values, being the optimal prices for the identifiable groups with the highest and lowest values of time, but within that set, assuming price differentiation is not possible, we could define a number of possible optima; for example, the price which maximises ridership overall, or the price where the sum of individual net benefits is maximised.

If we take a generalised cost formulation, and assume that everyone has the same value of time, the crucial model parameters are the initial values of price and waiting time, the demand elasticities and the value of time; these are related by the expression

$$\frac{\eta_P}{\eta_T} = \frac{P}{vT} \quad or \quad \frac{\eta_P}{P} = \frac{\eta_T}{vT} \; ;$$

and the functional form of the delay function. Previous authors (Douglas (1972), Beesley and Glaister (1983)) have used a simple inverse relationship between average waiting time and the number of vacant cabs of the form $T = g/V$. While this is most appropriate to the cruising cab and telephone markets, it is acceptable as an approximation of delays to passengers in the rank market.

The argument can be summarised as follows. If regulation of entry and/or price is to produce a better result for consumers than an open market, it is necessary to show that regulation will produce a preferred price/service level combination. This in turn requires knowledge of the relevant price and service elasticities and consumers' valuations of service (Beesley and Glaister (1983)). The principal dimension of service in this market is the time required to make the journey, including access time, waiting time and in-vehicle time. Values of time are important because they are the prime determinant of optimal regulatory behaviour. If the travellers' value of waiting time is very low, then the regulator can best serve the interests of the public by keeping fares down, allowing queues of passengers to form and waiting times to rise. If the value of time is high, then it is worth increasing service quality by ensuring a ready supply of taxis at the expense of lower occupancy rates and higher fares. There is, however, little evidence on values of time or price and service elasticities for taxi users; this is a gap which our research project has sought to fill.

3. VALUES OF SERVICE AND ELASTICITIES IN THE TAXI MARKET

There are good reasons for the lack of evidence on values of service and demand elasticities in the taxi market. Unlike, say, the bus industry, there is no data series for ridership, fares and vehicle kilometres by district which would enable elasticities to be obtained from analysis of aggregate data. Only by undertaking many case studies were we able to generate such data, and even then they suffer from serious limitations, not least because an aggregated approach fails to consider the differentiated market. The taxi caters for a wide variety of traffics. These include business travellers to/from railway stations and airports, local journeys to/from shopping centres, schools and hospitals, and the very important night time trade from pubs and clubs. Elasticities and values of time are likely to vary by journey purpose and time of day. The analysis is not made easier by the fact that in many British cities, significant waiting times are found late at night and the day time traffics are often adequately served.

For these reasons, we undertook a series of stated preference and transfer price experiments in four British cities. Stated preference methods present individuals with hypothetical scenarios and use the responses supplied to reveal information about the preferences underlying the choices made. Transfer price methods elicit from each respondent the change in an attribute level of their chosen mode (taxi) which would be just sufficient to cause a change in behaviour. We applied these techniques in four case studies. Study (iii) used SP only; the other studies used TP to obtain elasticities and SP for valuation. The different studies looked at:

(i) Waiting time, vehicle type and cost;
(ii) Waiting time and cost;
(iii) Walking time, waiting time, vehicle type and cost;
(iv) Walking time, waiting time, in-vehicle time and cost.

Studies (i) and (iii) were conducted by means of an on–street interview with people who had made a recent taxi trip. Studies (ii) and (iv) were conducted by means of a self–completion postal questionnaire handed out to passengers at taxi ranks as they boarded a taxi. Neither approach is ideal. The on–street method has the advantage of being able to include people who had made trips late at night but it is difficult for interviewers and respondents, and does not ensure a balanced sample of taxi users. The self–completion questionnaire method is better on the last two counts, but the problems involved in handing out forms at taxi ranks late at night meant it could only cover the daytime trade. The main results, and initial levels of the relevant variables, are shown in table 1.

The results are rather variable, but this is not surprising in view of the difficulties of surveying and of controlling the sample by journey type. Having said that, the overall results suggest relatively high values of walking and waiting time - well above those commonly found for bus travellers - together with low service elasticities. This seems strange, but is explained by the high fares, which are large relative to the time components of generalised cost. The very low valuation of in-vehicle time is probably due to peoples' perception of taxi in-vehicle time as being the lowest possible. Within the taxi mode, it is seen as relatively invariant, and if taxi IVT is high because of congestion, IVT for bus will be even higher. Overall, we feel it is safe to conclude that service elasticities are, on average, quite low.

In study (i), people were found to be relatively indifferent between saloon-car taxis and purpose-built London-style vehicles. However, study (iii) showed a slight preference for the London vehicle, and a fairly strong preference for large saloon cars over small saloon cars. This suggests that the minimum size/quality requirements imposed by some local authorities may be justifiable.

The results on price elasticities are the most difficult to interpret. Overall, we think that studies (i) and (ii) suffered from policy response bias and from sampling biases, and we are inclined to disregard the results as unreasonably high. The results from study (iii) were very interesting. The aggregate price elasticity was found to be –0.8, but there were clearly two distinct groups, "captives" (those travelling at night or with significant amounts of luggage), where the elasticity was –0.3, and "non–captives", where the elasticity was –1.9. Overall, it seems reasonable to conclude that the price elasticity of demand is probably at about unity for the traffic in aggregate, although this will differ according to the balance of trade in the district between day time and night time, the availability of alternative modes and the proportion of business travellers. The tariff structure (night time and luggage premia) could certainly be adapted to take account of these features.

Table 1: Service valuations and elasticities				
Study type → / Values of service↓	Study (i) On-street interview TP/SP	Study (ii) Self-completion TP/SP	Study (iii) On-street interview SP only	Study (iv) Self-completion TP/SP
Wait (p/min)	4.0	23.0	7.4	32.0
Walk (p/min)	-	-	7.8	25.0
IVT (p/min)	-	-	-	2.1
London taxi (pence)	2	-	10*	-
↓ Elasticities (initial values of variables in parentheses)				
Price	−2.2 (£1.70)	−2.5 (£2.70)	−0.8* (£3.80)	−1.0 (£3.40)
Waiting time	−0.04 (0.8)	−0.05 (0.25)	−0.1 (6.5)	−0.07 (0.75)
Walking time	-	-	−0.06 (3.7)	−0.3 (4.3)
In-vehicle time	-	-	-	−0.1 (16.5)

* See text for further commentary on these values.

4. A WELFARE ANALYSIS OF PRICE AND QUANTITY REGULATION

By applying data on the taxi trade in districts (iii) and (iv) together with our estimates of waiting time values and price and service elasticities to the theoretical model outlined in section 2, we were able to calculate the welfare effects of adopting different licensing strategies through their impact on fares, vehicle numbers and waiting times. The different strategies were evaluated against the base of the current situation, where there is a premium attached to the ownership of a plate of £30,000 in district (iii) and £20,000 in district (iv). The results are shown in table 2.

Table 2: Effects of different licensing strategies	District (iii)				District (iv)			
	% inc fares	No of cabs	% inc in CS	% inc welf	% inc fares	No of cabs	% inc in CS	%inc welf
A current case	–	262	–	–	–	120	–	–
B π=current, max rides	−5.8	251	+2.6	+1.9	−19.4	112	+26.4	+19.6
C π=0, current fares	–	365	+13.3	−18.2	–	156	+6.5	−21.0
D π=0, maintain quality	−26.2	307	+48.2	+7.0	−25.3	140	+55.3	+15.2
E π=0, max ridership	−27.0	305	+48.6	+7.3	−33.2	129	+64.3	+21.9
F π=0, fare ↓, same cabs	−33.3	262	+32.1	−4.6	−37.1	120	+59.7	+18.4
G π=0, complete dereg	+18.7	374	−15.1	−38.7	–	156	+6.5	−21.0

Row A shows the current position, with 262 cabs in district (iii) and 120 cabs in district (iv). Row B is the ridership (and welfare) maximising situation at the current values of the licence premia. The current fares level is too high, significantly so in district (iv), and the number of cabs should fall. The system should be operating at a higher occupancy ratio. To achieve this result, maintaining a licence premium, both price and entry control are required.

The remaining strategies involve eliminating this premium through reductions in fares, increases in vehicle numbers or both. There is no economic justification for licence premia, although there maybe arguments of stability and equity for maintaining a small premium, and/or allowing existing premia to erode over time, which have some validity. Strategy C is the case where entry control is abandoned, but fares regulation is maintained at current fare levels. Cab numbers increase substantially and users are better off, but insufficiently so to offset the loss of producers' surplus. Open entry at existing fares is therefore not a good strategy, and open entry with deregulation of fares (strategy G) is even worse. Entry forces the occupancy ratio even lower and fares higher. This scenario assumes that the rules of the trade - first-in, first out queuing by taxis at ranks, and no price competition between taxis - continue to hold. Further new entry at this position will result in failures of either the entrants or of some existing firms; this is the maximum sustainable industry size.

Strategy F involves regulating cab numbers at their current levels and forcing fares downwards to the break-even level. A large transfer takes place from producers to consumers, with fares falling by a third. The welfare position is mixed. This happens to be a good strategy for district (iv), which has very low waiting times in the initial situation, but clearly creates excessive waiting times in district (iii). Strategy D therefore eliminates the premium by driving down fares but issuing new plates so as to maintain the waiting times at the current level. Finally, the outright maximum ridership with zero premium is calculated in strategy E. In the two districts studied, this implies a fares cut of a quarter to a third, together with some increase in cab numbers, creating an increase in net social benefits of 7.3% and 21.9% respectively. Note, however, that these results do not take account of the potential for differential pricing by time of day, nor of the effects of the policy on road congestion or public transport. Nevertheless, the results suggest that the price, service and output levels resulting from current regulatory behaviour may be significantly sub-optimal. Nor, however, is complete deregulation advisable according to this framework, since it does not achieve the optimum price/service balance either. Figure 3 depicts the different solutions for district (iv).

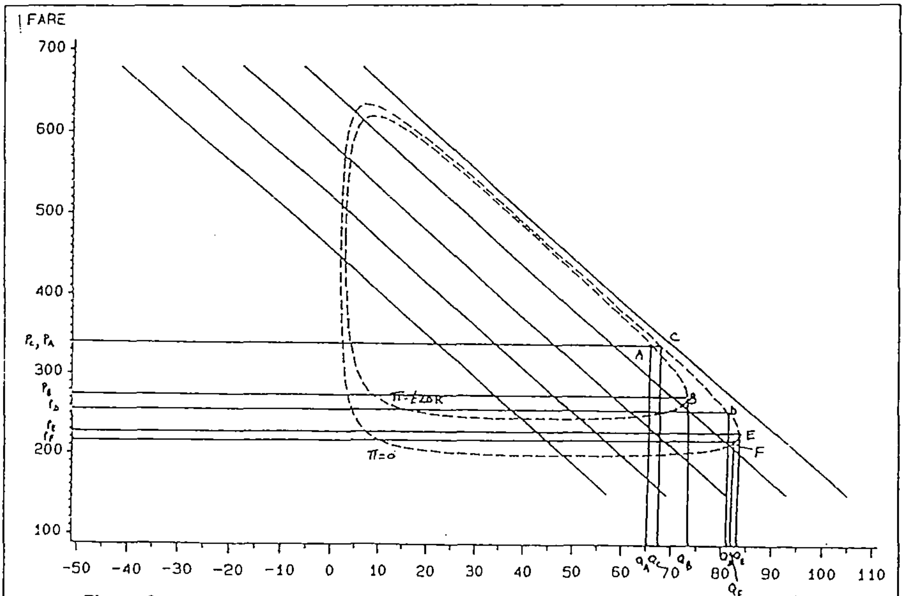

Figure 3: Model of the rank market in district (iv).

5. POLICY IMPLICATIONS AND CONCLUSIONS

The conclusion we have reached is that levels of service in the taxi trade in the UK are generally good, but waiting time elasticities are low, and price elasticities close to unity for the market as a whole. This has implications for the optimal tariff structure.

Given these overall elasticities, our conclusion is that, from the current starting point, the main issue for local authorities should be that of tariff regulation rather than entry control. Tariffs have been set on an historical basis and revised allowing for inflation. Our research suggests that fares are now significantly too high, and that the market could expand considerably in regulated districts with a combination of lower fares and some increase in cab numbers.

In contrast to this, the 1985 Transport Act has concentrated the regulators' attention on cab numbers. The evidence required to demonstrate that there is "no significant unmet demand" has been that passenger waiting times for cabs should be negligible. The result has been a higher price/ higher service/lower volume outcome than optimal, with an indefensibly high premium in the more heavily regulated districts.

One possible response to this would be to deregulate both fares and entry to the market. However, on our analysis, this could produce a poor result, with low occupancy rates, long cab waiting times and little service improvement for consumers; there would simply be more cabs chasing essentially the same business. Overall, we believe that the role of the regulators in urban areas should be focused on the balance between price and service. If price is correctly set at the net benefit maximising level, the number of cabs in service will adjust optimally through market forces.

We have argued that there are no good economic reasons for licence premia. Provided that there are no compelling social reasons either, we conclude that the regulators' role in this market should be to set the fare tariff and maintain quality standards, but to remove quantity controls; only for equity or stability reasons should entry control be used. Given the starting point, policy needs to focus much more on efficient prices and much less on efficient capacity and so-called "unmet demand".

BIBLIOGRAPHY

Beesley, ME and Glaister, S, "Information for regulating: the case of taxis", *Economic Journal*, vol 93, 1983, pp594-615

Douglas, GW, "Price regulation and optimal service standards", *Journal of Transport Economics and Policy*, vol iv, 1972, pp116–127

Manski, C and Wright, JD, "Nature of equilibrium in the market for taxi services", *Transportation Research Record 619*, Washington DC, 1976, pp11–16

Toner, JP, *The economics of regulation of the taxi trade in British towns*, (unpublished PhD thesis), University of Leeds, Leeds, UK, 1990

Toner, JP, "The demand for taxis and the value of time – a welfare analysis", *Working Paper 333*, Institute for Transport Studies, University of Leeds, Leeds, UK, 1991

Toner, JP, "The demand for taxis in Leeds and the value of time", *Working Paper 334*, Institute for Transport Studies, University of Leeds, Leeds, UK, 1991