Optimal regulation in a taxi market in a medium sized city with simultaneous radio-dispatches and taxi stand allocation

Rolf Jens Brunstad1, Kurt Jörnsten2 and Siri Pettersen Strandenes3

Preliminary and unfinished draft to be presented at the International Transportation Economics Association Annual Conference hosted by The Institute of Transport Economics – Norwegian Centre for Transport Research in collaboration with the Department of Economics at the University of Oslo June 17 - 19, 2015

Abstract
The taxi industry has historically been heavily regulated in most countries. Actual forms of regulation may be far from optimal, but the recent trend of deregulation in several countries have in many cases given increased instead of reduced fares and indicators of service quality show a mixed picture. This is probably what one should expect given the theoretical model that has been developed for the taxi market. The seminal paper in this tradition is Douglas (1972) where the demand for taxi trips depend on the fare and the value of the time the customer has to spend waiting/searching for a vacant taxi. The system allocating vacant taxis to waiting customers then becomes important in two respects. The way it operates may prevent price competition or at least make it imperfect, and the allocating system may have economies of scale indicating that a social optimum would require subsidies to the taxi industry.

Leaving the scale problem aside a second best zero profit solution, can be achieved in a highly simplified model by regulating fares leaving entry free. Either complete deregulation of fares and entry or deregulation of fares combined with entry regulation will lead to overcapacity and excessive fares. Under slightly different assumptions in modelling the waiting time De Vany (1975) reached similar conclusions. Recently these conclusions were confirmed in a more general model by Yang and Yang (2011). However, even the Yang and Yang model rely heavily on several limiting assumptions, the key one being that there is a common dispatch central or a pure cruising market. In such markets efficient price competition will not exist.

Beesley and Glaister (1983), using the Douglas model, discussed the regulator’s problem. Reaching the social optimum by regulation is only possible if the regulator is fully informed which generally is not the

1 Department of Economics, NHH Norwegian School of Economics
2 Department of Business and Management Science, NHH Norwegian School of Economics
3 Department of Economics, NHH Norwegian School of Economics
case. The paper discusses various ways of using common knowledge at least to improve on the existing situation.

In view of the regulator’s imperfect information Brunstad (1990 and 1991) raised the question whether a monopoly would perform better than unregulated free entry. With monopoly, the high fares will result in a monopoly profit, with free entry the monopoly profit will be partly wasted on excess capacity instead of lower fares and more trips.

The crucial market failure in the taxi market is the lack of price competition. Price competition requires more than one dispatch central. In this case, the conditions for Bertrand equilibrium might be met provided there are no barriers to entry, no collusion and constant returns to scale in the matching function between taxis and customers. However, with increasing return to scale the Bertrand equilibrium will no longer be optimal. Schroeter (1983) estimated the parameters of the matching function in the Minneapolis airport taxi market. He found that the scale elasticity that was not significantly higher than 1 at 10 % level of significance.

In reality, the matching system will be a mix between radio dispatches, hailing in the street and taxi stand allocation. With company specific radio dispatch and the cruising and taxi stand market open for all, this means that the matching function will be company specific. Furthermore, Brunstad et al (2012) found that in the multi central market of Bergen, the customers have a strong preference for the largest company, indicating that this company will take a disproportionate share of the radio dispatch market, leaving proportionally more of the cruising and taxi stand market to the smaller competitors. In such a setting the economies of scale question is clearly more important.

In this paper, we take as the point of departure the Yang and Yang model with the duration of a trip set to one for simplicity. We expand the model to have company specific matching functions and calibrate using available data and informed guesstimates from the Bergen taxi market. In the calibrated model, we calculate the social optimum, the second best no profit optimum, unregulated free entry and unregulated monopoly. Our preliminary results indicate that in a medium sized city like Bergen the scale problem is important. However, even assuming constant returns to scale it is likely that the current regulatory regime controlling entry and not fares is counterproductive. Our results further indicate that unregulated monopoly is better than complete deregulation and under realistic assumptions even better than the current regulatory regime.

1. Introduction.

The taxi cab industry has been heavily regulated in most countries. Either fares or entry are regulated, or both. Since this is so general, one might suspect that there is some market failure inherent in the taxi
market, so that free competition is unable to achieve an optimal result. This is the theme of this paper.\(^4\) Will the market mechanism on its own achieve an optimal solution in the taxi market, or will regulation in some form or another be necessary?

2. Characteristics of the taxi market.

The source of the market failure in the taxi market is that supply and demand are intertwined in a complicated manner. Supply will in a certain sense affect demand, whereas supply is affected by the level of demand. We have here a special case of external effects.

2.1. The demand for taxi trips.

Let us first take a look at demand. The demand for taxi trips depends of course on the fare just like the demand for any good or service will depend on the price. Thus, \(cet. \ par.,\) the demand for taxi trips will diminish when the fare is increased and \textit{vice versa}. However, as the customer does not always immediately find a vacant taxi when he wants one, the demand for taxi trips will also depend upon the expected average length of time the potential customer will have to wait before the taxi turns up. For a given fare then, the demand will depend negatively on the expected waiting time. The demand function can therefore be written as:

\[
X = D(p + qT) = A - b(p + qT) \quad D' < 0,
\]

where \(X\) is the number of trips demanded, \(p\) is the fare, \(T\) is the expected waiting time and \(q\) is the value of time for the potential passenger.\(^5\) \(p + qT\) is the full price of the taxi trip to the customer, including also his time costs. This is illustrated in figure 1, where several demand functions are drawn, each corresponding

\(^4\) The first part of this paper draws heavily on Brunstad (1991). An earlier version was published in Norwegian in 1990 (Brunstad 1990).

\(^5\) In addition, demand depends on prices and quality of alternative transport, real income and the intensity of business activity, which we do not consider here.

\(^6\) Using the linear demand function, \(X = A - b(p + qT),\) with \(A = 10; b = q = 1\)
to different waiting times.

In earlier work on the taxi market waiting time was assumed to depend negatively on the number of vacant taxis (Douglas 1972 and Beesley and Glaister 1983) or the total number of taxis in operation (de Vany 1975). Schroeter (1983) modeled waiting time as the outcome of a matching process between waiting customers and vacant taxis. Brunstad (1990 and 1991) using the same approach, demonstrated that the two earlier approaches turned out as special cases. This approach was later followed by Bergantino and Longobardi (2000) and Yang and Yang (2011).

The market demand for taxi trips per unit of time is as the number of persons who in that period decides to take a taxi. For simplicity, we then assume that there is one person per taxi and that taxi trips are all of equal length and have duration of one time unit\(^7\). Since people have to find taxis, the demand at a certain point of time equals the number of entrants to the queue of people waiting for a taxi at the same time. Exit from the queue will consist of those who find a taxi during the same period. Exit from the queue will then be equal to the number of matches between vacant taxis and waiting customers. The number of matches will be greater the more vacant taxis and the longer the queue of waiting customers. We can therefore formulate the matching function as:

\[
Y = G(Q, V), \quad G_1 > 0, G_2 > 0,
\]

where \(Y\) is the number of matches, \(Q\) is the number of persons waiting for a taxi and \(V\) is the number of vacant taxis. The actual matching process between taxis and customers will affect the form of the matching function. If the only way to obtain a taxi is to hail vacant cabs in the streets, a so-called cruising taxi market, the matching function will be different from what it would be if there is also a telephone operated radio-dispatch system, which in turn could be more or less computerized. A radio-dispatch system is probably more efficient than a cruising market, and more computerization enhances efficiency. Changes from one system to another or increasing computerization can be interpreted as shifts in the matching function.

\(^7\) These assumptions are identical to Yang and Yang (2011) except that Yang and Yang assume equal and constant duration different from 1. Our approach becomes less messy and entails no real loss of generality.
Figure 1. The demand for taxi trips.

The number of vacant taxis, $V$, is equal to the difference between the number of cabs in operation at a point of time, $N$, and the number of occupied taxis, $X$. As the duration of a taxi trip is one time unit by assumption, the number of occupied taxis will be equal to the number of matches, $Y$. If the queue shall not increase
indefinitely or disappear, exits must on the average equal entries, \textit{i.e.}: 

\begin{equation}
X = Y.
\label{eq:3a}
\end{equation}

in which case we will have a steady state. In steady state the length of the queue will be equal to the rate of entry times the average waiting time, \(T\):

\begin{equation}
Q = TX.
\label{eq:3b}
\end{equation}

Taking account of (3a) the number of vacant taxis can then be written as:

\begin{equation}
V = N - X.
\label{eq:4}
\end{equation}

\textbf{2.2. The supply of taxi trips.}

The supply of taxi trips depends of course on the total number of taxis in operation at a certain point of time. A taxi is in operation if it is driving, either with or without passenger, or if it is waiting. The number of taxis in operation will depend on revenue and cost per unit of time. Revenue per unit of time for a taxi will be the product of the fare and the taxi's occupancy ratio. The occupancy ratio is the ratio between time with passenger and total time in operation. On average for all cabs, this will be the ratio between the number of occupied taxis and the total number of cabs in operation, \textit{i.e.}:

\begin{equation}
U = X/N,
\label{eq:5}
\end{equation}

where \(U\) is the occupancy ratio. We shall further assume that cost per unit of time is not affected by whether the cab is occupied or not. This is a reasonable assumption as Norwegian data (Berthelsen 1982) show that more than 70\% of total cost do not depend on distance\(^8\), and for the remaining 30\%, taxis are in fact driving

\(^8\) Swedish data (Börjeson 1989) suggest that an even higher proportion of total cost is independent of distance.
much of the unoccupied time, in which case the cost will not be significantly reduced if the taxi is vacant.

In addition, we shall assume that these costs are constant per taxi irrespective of the number of cabs that each cab owner or taxi firm has at its disposal. This means that there are no economies or diseconomies of scale. This is a reasonable assumption as long as we are only considering the driving. Economies of scale might exist for repair and maintenance, and certainly will for the operation of a dispatch service. We will assume, however, that repair and maintenance can be bought externally at constant prices. We will also assume that the taxi firm can be connected to a dispatch service central at a constant fee per cab. For the time being, we assume that there is only one dispatch service central.

Assuming free entry, there will be no profits in the end and the long run supply function of taxi trips can be found from the no profit condition:

\[(6a) \quad pX = cN\]

or

\[(6b) \quad p = c/U\]

where \(c\) is cost per taxi per unit of time.

For a given occupancy rate the supply curve is horizontal, reflecting our assumption of constant returns to scale in the operation of a taxi fleet.

2.3. The waiting function.

In an ordinary market, demand is uniquely derived from consumer preferences while supply is uniquely derived from costs. At a given price, demand will be independent of the size of supply at that price and vice versa; not so in the taxi cab market. At a given fare, demand will be affected by the expected waiting time. But waiting time will in turn depend on the number of cabs in operation and thus on supply.
The waiting time function can be derived from the exit function in equation (2). Substituting into equation (2) from equations (3), (4) and (5), we get:

\[ (7) \quad X = G(TX, (N - X)) = G(TX, \left( \frac{1}{U} - 1 \right)X), \]

which can be solved for waiting time as a function of the number of trips and the occupancy rate. To investigate the properties of this relationship we must take a closer look at the exit function, which is in fact describing the matching process between waiting customers and vacant cabs.

Let us then define the following elasticities:

\[ \alpha_1 = G_1 \frac{TX}{X} = G_1 T \]

and

\[ \alpha_2 = G_2 \frac{(N - X)}{X} = G_2 \frac{V}{X} = G_2 \left( \frac{1}{U} - 1 \right) \]

If \( \alpha_1 + \alpha_2 > 1 \), there is increasing returns to scale in the matching function. By implicit differentiation of (7) we get:

\[ (8) \quad \frac{dT}{T} = \frac{1 - \alpha_1 - \alpha_2}{\alpha_1} \cdot \frac{dX}{X} + \frac{\alpha_2}{\alpha_1} \cdot \frac{1}{1 - U} \cdot \frac{dU}{U}, \]

Or equivalently:

\[ (9) \quad \frac{dT}{T} = \frac{1 - \alpha_1}{\alpha_1} \cdot \frac{dX}{X} + \frac{\alpha_2}{\alpha_1} \cdot \frac{dV}{V} \]

if we do not substitute for \( V \) in (2).
From (8) we see that:

\[
T = T(X,U), \quad T_x = 0 \quad \text{if} \quad 1 - \alpha_1 - \alpha_2 = 0, \quad T_u > 0
\]

Given X, the waiting time will be longer the higher the occupancy rate, whereas the sign of \( T_x \) depends on the returns to scale in the matching function. If there are increasing returns to scale, \( T_x \) will be negative; with diminishing returns to scale, \( T_x \) will be positive. If there are constant returns to scale in the matching function, then (7) will reduce to:

\[
G\left(T; \left(\frac{1}{U} - 1\right)\right) = 1,
\]

which can be solved for waiting time as a function of occupancy rate alone, \( i.e. \)

\[
T = T(U), \quad T_u > 0.
\]

We can probably rule out decreasing returns, but queue systems do in fact often have increasing returns, and earlier work in this area have implicitly assumed this by their choice of functional form for the waiting function. For example both Douglas (1972) and Beesley and Glaister (1983) assume that the waiting function has the simple form:

\[
T = kU^{-1},
\]

where \( k \) is some constant. Looking at equation (9), we see that this means that we must have \( \alpha_1 = \alpha_2 = 1 \), which gives a scale elasticity of 2 in the matching function. De Vany (1975) has the following waiting function:

\[
T = T(N) = T\left(\frac{X}{U}\right), \quad T' < 0.
\]
For this function, the partial derivative of \( T \) with respect to \( X \) is negative, which also implies \( \alpha_1 + \alpha_2 > 1 \).  

Schroeter (1983) estimated the parameters of the matching function for the Minneapolis radio dispatch taxi market, and got the following estimates (Schroeter 1983, standard deviations in parentheses):  

\[
\alpha_1 = 0.8395 \ (0.0649), \quad \alpha_2 = 0.2933 \ (0.0875)
\]

Both \( \alpha_1 \) and \( \alpha_2 \) are significantly less than 1 at the 1% level, indicating that economies of scale may be significantly less than what Douglas (1972) and Beesley and Glaister (1983) assumed. The scale elasticity, \( \varepsilon = \alpha_1 + \alpha_2 = 1.1328 \), is slightly above 1, but the hypothesis \( \alpha_1 + \alpha_2 = 1 \) against the alternative \( \alpha_1 + \alpha_2 > 1 \) cannot be rejected even at the 10% level of significance. This suggests that the economies of scale, at least for a city of the size of Minneapolis, may be insignificant or nonexistent.

In what follows we shall assume the matching function\(^{10}\)

\[(7a) \quad X = \left( \frac{T}{U} \right)^{\alpha} (N - X)^{1-\alpha} - \gamma.\]

This is a non-homogenous CES function with elasticity of substitution \( \sigma=1 \), and scale elasticity  

\[\varepsilon = \alpha_1 + \alpha_2 = 1 + \frac{\gamma}{(\frac{T}{U})^\alpha(N-X)^{1-\alpha}} = 1 + \frac{\gamma}{\gamma + X}.\]

For \( \gamma > 0 \), \( \varepsilon \) goes from 2 asymptotically to 1 as \( X \) goes from zero till infinity.

From (7a) we get the following waiting function:

\[(7b) \quad T(X, U) = B \left( \frac{X+\gamma}{X} \right)^{\frac{1}{\alpha}} \left( \frac{U}{1-U} \right)^{\frac{1-\alpha}{\alpha}}.\]

\[T_X \left( = \frac{\partial T}{\partial X} \right) = B \left( \frac{U}{1-U} \right)^{\frac{1-\alpha}{\alpha}} \left( \frac{1}{\alpha} \right) \left( \frac{X+\gamma}{X} \right)^{-\frac{1}{\alpha}} \left( -\frac{\gamma}{X^2} \right) = B \left( \frac{U}{1-U} \right)^{\frac{1-\alpha}{\alpha}} \left( \frac{1}{\alpha} \right) \left( \frac{X+\gamma}{X} \right)^{-\frac{1}{\alpha}} \left( \frac{U}{1-U} \right)^{-\frac{1}{\alpha}} < 0 \]

\(^{9}\)For this function we have that: \( \frac{\partial T}{\partial X} = -\frac{\partial U}{\partial T} \). For this to be consistent with (8) we must have: \( \alpha_1 + \alpha_2 = 1 + \alpha_2 \frac{1}{1-U} > 1 \) unless \( \alpha_2 = 0 \).

\(^{10}\)This function, which probably better reflects the scale properties of a matching process, has been suggested by Lars Mathiesen.
Alternatively, the waiting time can be written as a function of $X$ and $N$:

$$T_X = B \left( \frac{-Y}{X} \right) \left( \frac{X+Y}{X} \right)^{\frac{1-\alpha}{\alpha}} \left( \frac{U}{1-U} \right)^{\frac{1-\alpha}{\alpha}} \left( \frac{1}{X} \right)^{\frac{1-\alpha}{\alpha}}$$

$$T_U \left( = \frac{\partial T}{\partial U} \right) = B \left( \frac{X+Y}{X} \right)^{\frac{1}{\alpha}} \left( \frac{U}{1-U} \right)^{\frac{1-2\alpha}{\alpha}} \left( \frac{1}{(1-U)^2} \right) = B \left( \frac{1-\alpha}{\alpha} \right) \left( \frac{X+Y}{X} \right)^{\frac{1}{\alpha}} \left( \frac{U}{1-U} \right)^{\frac{1}{\alpha}} \left( \frac{1}{U^2} \right) > 0$$

$$T_{U} U = B \left( \frac{1-\alpha}{\alpha} \right) \left( \frac{X+Y}{X} \right)^{\frac{1}{\alpha}} \left( \frac{U}{1-U} \right)^{\frac{1}{\alpha}} \left( \frac{1}{U} \right)$$

Alternatively, the waiting time can be written as a function of $X$ and $N$:

$$(7c) \quad T(X, N) = B \left( \frac{X+Y}{X} \right)^{\frac{1}{\alpha}} \left( \frac{X}{N-X} \right)^{\frac{1-\alpha}{\alpha}}$$

$$T_X \left( = \frac{\partial T}{\partial X} \right) = B \left( \frac{X}{N-X} \right)^{\frac{1-\alpha}{\alpha}} \left( \frac{1}{\alpha} \right) \left( \frac{X+Y}{X} \right)^{\frac{1-\alpha}{\alpha}} \left( \frac{N}{(N-X)^2} \right)$$

$$T_N \left( = \frac{\partial T}{\partial N} \right) = -B \left( \frac{X+Y}{X} \right)^{\frac{1}{\alpha}} \left( \frac{1-\alpha}{\alpha} \right) \left( \frac{X}{N-X} \right)^{\frac{1-2\alpha}{\alpha}} \left( \frac{1}{N-X} \right)^{\frac{1}{\alpha}} X$$

3. The optimal solution in the taxi-market.

Optimally the number of taxi trips, $X$, and the occupancy rate, $U$, should be set to maximize social surplus, that is the sum of consumer and producer surplus. The area under the demand function is:

$$\int_0^X D^4(z)dz,$$

where $D^4$ is the inverse function of (1). To arrive at the social surplus two kinds of cost will have to be deducted; operating cost for the taxis, $cN$, and waiting cost for the customers, $qTX$. This gives the following expression for social surplus:

$$(10) \quad W(X, U) = \int_0^X D^4(z)dz - qTX - cN = \int_0^X D^4(z)dz - qTX - c\frac{X}{U}.$$
Maximizing $W$ with respect to $X$ and $U$ gives the following first order conditions$^{11}$:

(11a) \[ p = qT_X X + \frac{c}{U}, \]

and

(12) \[ \frac{c}{U} = qT_U U. \]

Equation (12) tells us that in optimum the cost of adding one more taxi, $c$, should equal the marginal social benefit in the form of reduced waiting time, $qT_U U$.

(11a) and (12) can be combined to yield

(11b) \[ p = q(T_X X + T_U U) \]

Equation (11b) tells us that in optimum the fare should equal the value of the extra waiting time that the marginal taxi trip imposes on the intra-marginal units. The right hand side of (11) contains two components; one capacity utilization effect, $qT_U U$, which is always positive$^{12}$ and one scale effect, $qT_X X$, which will be negative if there are increasing returns to scale in the matching function. From (11a), we see that in optimum the taxis will not be able to cover their costs if there are increasing returns. In the presence of increasing returns in the matching function, an optimal solution would require subsidies to the industry. If we do have access to non-distorting taxation this is unproblematic from an efficiency point of view. Otherwise, the efficiency loss from increased taxes should have been taken into account when calculating the social optimum.

\[ \frac{\partial w}{\partial x} = p + qT - qT_X X - \frac{c}{U} = p - qT_X X - \frac{c}{U} \quad \frac{\partial w}{\partial u} = -qT_U X + \frac{cX}{U^2} \]

\[ \frac{\partial w}{\partial x} = 0 \Rightarrow \quad p = qT_X X + \frac{c}{U} \]

\[ \frac{\partial w}{\partial u} = 0 \Rightarrow \quad \frac{c}{U} = qT_U U \]

$^{11}$ With our assumptions we must have $U \geq 1$. We further assume an inner solution which means that $U < 1$. 

$^{12}$
Therefore we may be interested in the second best solution given that we require that taxis should be able to cover their costs, that is: \( pX - cN = 0 \) or \( (D^{-1}(X) - qT)X - c \frac{X}{U} = 0 \).

Forming the Lagrangian:

\[
\mathcal{L} = \int_0^X D^{-1}(z) \, dz - qTX - c \frac{X}{U} - \lambda \left[ (D^{-1}(X) - qT)X - c \frac{X}{U} \right],
\]

we find the following first order conditions:

i) \( p - qTXX - c \frac{1}{U} = \lambda (\{(D^{-1})' - qT\}X + p) \)

ii) \( -qUXX - c \frac{X^2}{U^2} = \lambda \left[ -qUXX - c \frac{X^2}{U^2} \right] \)

iii) \( p = c \frac{U}{U} \)

From ii) \( \lambda = 1 \). From i) we then get: \(-c \frac{1}{U} = (D^{-1})'X = \frac{p}{D'X} \).

Considering iii) this is equivalent to:

\[
(13) \quad p \left( 1 + \frac{1}{D'X} \right) = 0 \implies -D' \frac{p}{X} = 1.
\]

The interpretation of this is as follows. The social optimum requires a number of trips beyond the point where the taxis are able to cover their costs. If we impose cost coverage, the marginal revenue from the last trip should be zero.

With constant returns to scale the waiting function will be an increasing function of \( U \) alone, \( T = T(U)^{13} \), and the first order conditions for optimum reduces to:

\[
(11^*) \quad p = qT_U U
\]

and

\[
^{13} \gamma = 0 \implies T(X, U) = B \left( \frac{a}{1 - a} \right)^{\frac{1 - a}{a}} \left( \frac{1}{U} \right)^{\frac{1}{2}} \left( \frac{1}{U} \right)^{\frac{1}{2}} \left( \frac{1}{U} \right)^{\frac{1}{2}} \left( \frac{1}{U} \right),
\]

\[
T_X = T_X X = 0, \quad T_Y = B \left( \frac{1 - a}{a} \right)^{1 - a} \left( \frac{1}{U} \right)^{\frac{1}{2}} \left( \frac{1}{U} \right)^{\frac{1}{2}} \left( \frac{1}{U} \right).
\]
\[ \text{(12*)} \quad \frac{c}{U} = q T_u U \]

In this case, as \( T \) is a function of \( U \) alone, we can solve (12*) for \( U \) as a function of \( c \) and \( q \):

\[ U_{opt} = U(c,q), \quad U_c > 0, U_q < 0. \]

The higher the cost of capacity, \( c \), the higher is the optimal occupancy ratio; the higher the cost of time, \( q \), the lower is the optimal occupancy ratio.


For constant returns to scale in the matching function, the optimal solution characterized by (11*) and (12*) does not require subsidies, but as will be shown in what follows, it will not be reached by an unregulated market.

4.1. Unregulated free entry.

With unregulated free entry, entries or quits will take place until profit is zero, equation (6a).

Combining the (1) and (6b) to eliminate \( U \), we get:

\[ \text{(14)} \quad X = D(p + q T_c(\frac{c}{p})) = E(p) \]

This function is the locus of all possible free entry equilibria of supply and demand for taxi trips and it is drawn as the curve EE in figure 1. Differentiating (14) we get:

\[ \text{Partial differentiation of (12* gives} \quad \frac{1}{qB} dc - \frac{cb}{(qB)^2} dq = \left( \frac{1-a}{a} \right) \left( \frac{1}{1-U} \right)^{1-a-1} \left( \frac{1}{1-U} \right)^2 dU \]

\[ ^{14} \text{Partial differentiation of (12*) gives} \quad \frac{1}{qB} dc - \frac{cb}{(qB)^2} dq = \left( \frac{1-a}{a} \right) \left( \frac{1}{1-U} \right)^{1-a-1} \left( \frac{1}{1-U} \right)^2 dU \]
\[
\frac{dE}{dp} = (1 - qT_u \frac{c}{p^2}) D',
\]
the sign of which is determined by:

\[p - qT_u \frac{c}{p}\]

For small enough \(p\), (15) will obviously be negative, so that the EE curve will be increasing for small \(p\). For increasing \(p\), the second term in (15) will gradually diminish in absolute value so that in the end (14) will become positive and the EE curve will bend backwards. When (15) is zero as in point SO in the figure 1, we will also have reached the social optimum as both the first order conditions, (11\(^*\)) and (12 \(^*\)), are met in this point.

More formally, we can take the second derivative of (14) to get:

\[
\frac{d^2E}{dp^2} = (1 - qT_u \frac{c}{p^2})^2 D'' + q(T_{uu} U + 2T_u) \frac{U}{p^2} D'.
\]

It follows from the discussion above that in optimum the first term on the right hand side of (15) is zero. Further, from the second order conditions for the maximum of (10), \(T_{uu} U + 2T_u\) must be positive\(^{15}\), so that (16) is negative in optimum. We may therefore conclude that if point SO in figure 1 is the social optimum, then the EE curve must bend backwards above that point.

But SO cannot be a stable equilibrium in an unregulated free entry market. To see this, consider first the zero profit curve in figure 2. This curve represents the combinations of \(p\) and \(N\) that corresponds to the points on the EE curve in figure 1. Points above this curve will give losses for the individual taxi, whereas

\(^{15}\) The second order conditions for the maximum of (10) requires: \(\frac{\partial^2 W}{\partial U^2} = -qT_{uu} X - 2 \frac{cX}{U^3} < 0\). As we have that \(\frac{c}{U} = qT_u U\) from the first order conditions, this is equivalent to \(T_{uu} U + 2T_u > 0\).
at points below the curve the taxis will run at a profit. From (6b) and (14) we have that along the no profit curve in figure 2 the number of taxis in operation can be written as: \( N = \frac{pE(p)}{c} \).

Differentiating with respect to \( p \) yields:

\[
(17) \quad \frac{dN}{dp} = \frac{X}{c} + \frac{p}{c} \frac{dE}{dP}.
\]

(17) will be positive as \( \frac{dE}{dP} \) is positive for \( p \) small enough. The no profit curve will therefore first be increasing. \( \frac{dE}{dP} \) will remain positive for fares below the optimal fare \( P_{SO} \) corresponding to the point SO in figure 1. At \( p_{SO} \), \( \frac{dE}{dP} \) will be zero, but (17) will still be positive because the first term on the right hand side is positive. The curve will therefore go on increasing for a while for fares higher than \( p \). First when \( \frac{dE}{dP} \) have become sufficiently negative, will (17) be zero and the number of taxis will have reached the maximum number that the market can bear.
Figure 2. The relationship between fare level and number of taxis in operation with constant returns to scale in the matching function\textsuperscript{16}.

\textsuperscript{16} In figure 2 we assume $\gamma=0$, $\alpha=0.5$, $A=10$; and $b=q=c=1$. 
Now suppose that we by chance have arrived at a combination of fare and number of taxis that corresponds to the social optimum, (SO), in figure 2. As we shall see this point cannot be a stable equilibrium with unregulated free entry.

Let $\pi$ be profit for each individual taxi, so that:

\[(18) \quad \pi = pU - c = \frac{pX}{N} - c = \pi(p, N).\]

Differentiating (18)$^{17}$ with respect to $p$ and $N$, we get:

\[(19) \quad \frac{\partial \pi}{\partial N} = -\frac{pU}{N - qD'T_u} < 0\]

and

\[(20) \quad \frac{\partial \pi}{\partial p} = \frac{X + (p-qT_u)D'}{N - qD'T_u}\]

At the point $(p_{SO}, N_{SO})$, (20) will be positive, implying that higher fares would increase profit, whereas

$^{17}$ In this expression $X$ is given by:

\[
(i) \quad X = D(p + qT_u)\frac{X}{N}.
\]

Differentiating (i) to get:

\[
dX = D'dp + D'qT_u \frac{1}{N}dX - D'qT_u \frac{X}{N^2}dN,
\]

and setting $dp$ and $dX$ alternatively to zero, we get the partial derivatives

\[
\frac{\partial X}{\partial p} = \frac{D'N}{N - D'qT_u} \quad \text{and} \quad \frac{\partial X}{\partial N} = -\frac{DqT_u}{N - DqT_u},
\]

These derivatives are then used to obtain (19) and (20).
(19) tells us that new entrants would reduce profit.

If this is the case, Douglas (1972) argues that fares will rise, as the individual taxi owner will always have an incentive to charge a little above the going fare. The reason for this is the following: Assume the going fare is $p_{so}$. An individual driver could then charge $p_{so} + \delta$ for some positive $\delta$ and increase his income, provided $\delta$ is less than the cost to the customer of his extra wait should he reject the taxi (Douglas 1972, p. 122). If this is profitable for one taxi owner, it is profitable for all. The taxis will run at a profit and this will in turn attract new entrants. Therefore, market forces will always push both $p$ and $N$ upwards as long as (19) is positive. (19) will become zero when the following condition is met:

$$\left(21\right) \quad p\left(1+\frac{X}{pD}\right) = qT_U U.$$ 

In figure 2, this occurs at point FE corresponding to FE on the backward bending part of the EE curve in figure 1. At this point we also have the highest number of taxis that the market can bear, as we see from (17) that $\frac{dN}{dp}$ will be zero if (21) holds. We can therefore rule out points on the zero profit curve to the left of FE.

For points on the zero profit curve to the right of FE in figure 2, (20) will be negative. The argument above is not symmetric, however. For fares higher than $p_{FE}$ the taxi owners as a whole would gain by lowering the fare. The individual taxi owner, however, would have no incentive to do so. If a single taxi reduced the fare without the others following, this would only bring a loss to this individual taxi and give its customers a pleasant surprise, but it would have no effect on the demand for taxi trips. Only if a significant proportion of the taxis did reduce their fare, the demand for taxi trips would increase enough to compensate for the reduction in income per trip. Therefore, the market fare will not go down unless there is some sort of concerted action or cooperation among the taxi firms. It would seem then that points on the solid curve to the right of $p_{FE}$ could not be ruled out. It should be noted, however, that the individual incentive to increase the fare will decrease with increasing $p$ because waiting times and consequently the cost to the customer of rejecting a taxi will diminish.

Douglas (1972), however, seems to argue that the existence of taxi associations (which presumably would
be necessary to assure the operations of an efficient dispatch service even in a competitive situation) would bring about fare reductions if (20) is negative, so that \( p = p_{FE} \) might become the eventual outcome. In what follows we shall refer to \((p_{FE}, N_{FE})\) as the unregulated free entry solution, bearing in mind, however, that the free entry solution might be indeterminate.

At point FE in figure 2,

\[
\frac{\partial W}{\partial X} = p - \frac{c}{U} = 0
\]

and

\[
\frac{\partial W}{\partial U} = X \left( \frac{c}{U} - qT_u U \right) > 0
\]

From this, we see that the unregulated free entry solution leads to inoptimally low capacity utilization. Points on the EE curve above FE have lower occupancy ratios, so they are even less efficient. We can conclude that unregulated free entry will lead to a situation where the fare is too high and there are too many taxis, each of which is producing too few taxi trips.

4.2. Unregulated monopoly.

In monopoly, the monopolist will maximize total profits:

\[
\Pi(p, N) = pX - cN
\]

or equivalently:

\[
\Pi(X, U) = \left( D^{-1}(X) - qT(X, U) \right) X - \frac{cx}{U}
\]

Maximizing (22b) with respect to \( p \) and \( N \) gives the following first order conditions\(^{18}\):

\[
\frac{\partial \Pi}{\partial X} = \frac{A-2X}{b} - qT - qT_X X - \frac{c}{U}
\]

\[
\frac{\partial \Pi}{\partial U} = -qT_u U + \frac{cx}{U^2}
\]

\(^{18}\)
(23) \[ p \left( 1 + \frac{1}{D' \frac{p}{X}} \right) = qT_X X + \frac{c}{U}, \]
where \( D' \frac{p}{X} = \frac{1}{(D^{-1})'} \frac{p}{X} \) is the fare elasticity of demand,
and
(24) \[ \frac{c}{U} = qT_U U \]

In this point
\[ \frac{\partial W}{\partial X} > 0, \]
implying that there are too few taxi trips. However,
\[ \frac{\partial W}{\partial U} = 0, \]
so that, given the number of trips, the occupancy rate is optimal.

In fact, (24) is identical to (12*), so that for constant returns to scale in the matching function the occupancy ratio in monopoly is the same as in optimum,

In figure 1 this means that the monopoly solution must lie on the same demand curve as the optimum. We can find it as the point on this demand curve corresponding to the intersection between the marginal revenue curve and the line \( p_{SO} = c/U_{SO} \) = marginal cost

The optimal solution is the intersection between the demand curve corresponding to the optimal occupancy ratio and the EE curve given by the no profit condition. The free entry solution is on the EE curve, but on a “wrong” demand curve. The monopoly solution is on the correct demand curve, but too few taxi trips are supplied. The monopoly solution then clearly gives higher fare, fewer trips and fewer taxis than the optimal solution. The free entry solution on the other hand also involves higher fare and fewer trips than the optimal

\[ \frac{\partial n}{\partial X} = 0 \Rightarrow \frac{A-2X}{b} - qT = qT_X X + \frac{c}{U} \]
\[ p + \frac{X}{b} = qT_X X + \frac{c}{U} \]
\[ \frac{\partial n}{\partial U} = 0 \Rightarrow \frac{c}{U} = qT_U U \]
solution, but there are more taxis than optimal. The free entry solution clearly means more taxis and lower occupancy rates than monopoly, but nothing certain can be said about fare and trips. In figure 2, the solid curve shows those combinations of fare and number of taxis that are compatible with the no profit condition. Both the optimal and the free entry solutions are located on this curve, at the points \((p_{SO}, N_{NO})\) and \((p_{FE}, N_{FE})\) respectively. The dotted blue curve is an iso profit curve. The curves shift downwards as profit gets higher. At the top point of each iso profit curve (13), which is identical to (24), will hold. The monopoly solution must therefore be located at the end of the solid line connecting these points.

As neither free entry nor monopoly is optimal, one might wonder which of them would better if regulation is infeasible. This will depend upon the importance of the cost of waiting. If there were no waiting cost, the optimal occupancy rate would be equal to its technical maximum (which by definition must be less than or equal to one) and the externalities would disappear. The free entry solution would therefore coincide with the optimum, and would obviously be superior to the monopoly solution. For small but nonzero values of the cost of time this may be the case, but if the cost of time becomes high enough the free entry solution might involve so much excess capacity that the monopoly situation may preferable.

In table 1 we compare the condition for social optimum, second best, free entry equilibrium and monopoly. With constant return to scale in the matching function, \(T_X = 0\). Then social optimum is compatible with non-negative profit, so no second best solution is required. Free entry fulfills the first but not the second condition for social optimum, whereas monopoly fulfills the second but not the first.

<table>
<thead>
<tr>
<th>Social optimum</th>
<th>Second best</th>
<th>Free entry</th>
<th>Monopoly</th>
</tr>
</thead>
<tbody>
<tr>
<td>(11) ( p = qT_XX + \frac{c}{U} )</td>
<td>( p = \frac{c}{U} )</td>
<td>( p = \frac{c}{U} )</td>
<td>(23) ( p \left( 1 + \frac{1}{\theta} \right) = qT_XX + \frac{c}{U} )</td>
</tr>
<tr>
<td>(12) ( \frac{c}{U} = qT_U U )</td>
<td>(13) ( p\left( 1 + \frac{1}{\theta} \right) = 0 )</td>
<td>(21) ( p \left( 1 + \frac{1}{\theta} \right) = qT_U U )</td>
<td>(24) ( \frac{c}{U} = qT_U U )</td>
</tr>
</tbody>
</table>

**Table 1.** Social optimum, second best, free entry and monopoly compared

With increasing returns to scale in the matching function waiting time will depend not only on
the occupancy rate, \( U \), but also depend negatively on the number of trips, \( X \), as we have seen above. Social optimum will therefore require negative profit. See figure 3.

![Figure 3](image)

**Figure 3.** The relationship between fare level and number of taxis in operation with increasing returns to scale in the matching function

In the second best without subsidies is where a welfare contour is tangent to the zero profit curve at SB in figure 3.

5. **Optimal regulation.**

We have seen that neither unregulated free entry nor unregulated monopoly will be optimal in the taxi market. With constant returns to scale in the matching function, unregulated free entry would fulfill the
first but not the second of the two conditions for social optimum, whereas unregulated monopoly would fulfill the second but not the first. With increasing returns to scale, unregulated monopoly would still fulfill the second of the optimality conditions, whereas unregulated free entry would fulfill neither of them.

If we instead aim at a second best solution, the best solution compatible with non-negative profit, the first of the conditions will be replaced by a zero profit condition. This condition will be met by free entry but not by monopoly.

If authorities had full information on cost and demand conditions, they could stipulate $p_{SO}$ in figure 2 as the maximum fare. In a free entry situation this restriction would be binding, as $\frac{\partial \pi}{\partial p} > 0$ for $p \leq p_{SO}$. In the same way, one could reach the second best solution under increasing returns to scale by imposing $p_{SB}$ as the maximum fare. With a maximum fare equal to $p_{SB}$ profit new entrants would come in until profit is zero. Compared to the unregulated free entry solution, the lower regulated fare would squeeze some of the taxis out of the market and thereby increase the occupancy ratio of the remaining cabs until $N_{SB}$ is reached. Consequently, if the regulator were fully informed, the optimal way to regulate the taxi market would be to regulate the fare and leave entry free. Unfortunately, however, the regulator would normally not possess this information. It is of course cold comfort to the authorities to know that all they have to do to assure an optimum in the taxi market is to stipulate the fare at the optimal level, if they do not have the necessary information to do so.

A monopoly facing a maximum fare at $p_{SB}$ will select the corresponding point on the $\frac{\partial n}{\partial N} = 0$ curve, and end up in point B in figure 3.

Restrictions on entry will not bring about an optimal or second best situation unless accompanied by a stipulated maximum fare, but then the restrictions on entry are unnecessary.

Assume that the authorities stipulate the number of licenses at $N_{SB}$, and that the fare is $p_{SB}$ at the outset.
As long as fares are not regulated, we will end up in point A on the $\frac{\partial n}{\partial p} = 0$ curve in figure 3. Such licensing, leaving fares free, will therefore give to high fares and to low occupancy ratios compared to the social optimum, even if the number of taxis is right\(^{19}\).

### 6. Effects of deregulation.

Assume now that both fares and entry are regulated. Assume further that the number of licenses is stipulated at $N_{SB}^{20}$. If the regulation of fares is binding, the maximum fare must then lie somewhere between $p_{SB}$ and the fare corresponding to point A in figure 3\(^{21}\). A total deregulation would lead to the free market solution FE. Deregulation from a situation with binding regulation of both entry and fares will therefore clearly mean an increase in the number of taxis. The effect on the fare is more uncertain depending on the actual size of the regulated fare. There exist, however, a clear possibility that fares would also increase as a result of deregulation.

Data from some American cities during the great wave of deregulation in the eighties are summed up in table 2, which is taken from Teal & Berglund (1987).

From the first panel we can see that the number of taxis have increased in all deregulated cities, and for some of the cities the increase is rather spectacular. From the second panel we can see that the level of fares has increased relative to a control group of regulated cities. If both the fare and the number of taxis increase, it follows from (1) and (5) that this must lead to a reduction in the occupancy ratio. From the third panel we can see that this seems to be the case.

In present regulatory regime in urbanized areas in Norway, fares are no longer regulated. Taxis are licenced but the taxis must be connected to a central. No central can have more than 50 % of the total licences. New licences have been issued. However, the development since this regime was introduced seems to be consistent with a movement along the $d\Pi/dp=0$ curve in figure 3 from A towards FE.

---

\(^{19}\) That is, equal to $N_{SB}$ (or $N_{SO}$ if $v=0$).

\(^{20}\) It is of no importance to the argument whether $(p_{SB}, N_{SB})$ is the social optimum or not.

\(^{21}\) Assuming that the taxi owners will reach some sort of agreement not to raise their fares above the fare corresponding to point A.
### Change in Size of Taxi Industry since Deregulation

<table>
<thead>
<tr>
<th>City</th>
<th>Increase in Number of Taxis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seattle</td>
<td>33</td>
</tr>
<tr>
<td>San Diego</td>
<td>127</td>
</tr>
<tr>
<td>Sacramento</td>
<td>56</td>
</tr>
<tr>
<td>Kansas City</td>
<td>18</td>
</tr>
<tr>
<td>Phoenix</td>
<td>83</td>
</tr>
<tr>
<td>Tucson</td>
<td>33</td>
</tr>
<tr>
<td>Oakland</td>
<td>38</td>
</tr>
</tbody>
</table>

### Average Annual Inflation-Adjusted Indexed Fares in Regulated and Deregulated Cities

<table>
<thead>
<tr>
<th>Fares Adjusted by</th>
<th>CPI</th>
<th>Cost Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
</tr>
<tr>
<td>Deregulated Cities</td>
<td>0.910</td>
<td>0.957</td>
</tr>
<tr>
<td>Regulated Cities</td>
<td>0.960</td>
<td>0.930</td>
</tr>
<tr>
<td>Regulated Cities</td>
<td>0.956</td>
<td>0.905</td>
</tr>
<tr>
<td>Regulated Cities</td>
<td>0.953</td>
<td>0.918</td>
</tr>
</tbody>
</table>

(1971 = 1.0)

\( ^a \) Uses June 1979 as after/before point.

\( ^b \) Uses June 1982 as after/before point.

\( ^c \) Average of (a) and (b)

### Trends in Taxi Productivity after Deregulation

<table>
<thead>
<tr>
<th>City</th>
<th>Trips per Shift</th>
<th>Trips per Cab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phoenix</td>
<td>-23(^a)</td>
<td>-34(^a)</td>
</tr>
<tr>
<td>San Diego</td>
<td>N/A</td>
<td>-37(^a)</td>
</tr>
<tr>
<td>Seattle</td>
<td>-35(^b)</td>
<td>-48(^b)</td>
</tr>
<tr>
<td>Tucson N/A</td>
<td>-33</td>
<td>-38(^a)</td>
</tr>
</tbody>
</table>

\( ^a \) One year after regulatory change.

\( ^b \) Two years after deregulation (based on trip sheets which may be incomplete).

*Table 2. Effects of deregulation in the USA. (Source Teal and Berglund (1987))*
7. Competition between dispatch centrals

So far we have assumed that there is only one radio dispatch service and that the taxis are buying its services at a constant cost. It is hard to imagine any efficient price competition between taxis belonging to the same dispatch central. When the customer phones the central, his choice is made. Efficient price competition probably requires competing dispatch centrals.

With more than one dispatch central, there will be one matching function for each central, matching incoming customers and taxis connected to that central. With constant returns to scale in the matching function, no economies of scale in the operation of the central itself, and if a taxi trip can be treated as a homogeneous product, the classical conditions for Bertrand competition might apply.

As there is no profit in equilibrium, however, this requires that there are no irreversible costs of entry, as no one would challenge the incumbent unless they can expect to earn enough profit to allow them to recover the entry costs.

With increasing returns to scale in the matching functions, things become a lot more complicated. When dispatch centrals operate different number of taxis, matching will be influenced. The smaller central will have to have a larger fleet relative to its share of the trips to be able to deliver the product at the same full price, \( p + qT \), as the larger one.

In reality, the matching system will be a mix of radio dispatches, hailing in the street and taxi stand allocation. With company specific radio dispatch and apps, and the cruising and taxi stand market open for all, the matching function will be company specific. See figure 4.

*Figure 4. Market shares in the different segments of the taxi market.*
Brunstad et al (2012) found that customers in the Bergen multi central market, have strong preference for the largest company, hence this company will take a disproportionate share of the radio dispatch market, leaving proportionally more of the cruising and taxi stand market to the smaller competitors. While the economies of scale in the cruising and taxi stand segments are probably small, they may be much greater in the dispatch and apps segment.

In such a setting the economies of scale question is clearly even more important.

In a multi central setting with n centrals the matching function for central i is

\[ x_i = \left( \frac{TX_i}{B} \right)^\alpha (s_i N - x_i)^{1-\alpha} - \gamma \]

where \( x_i \) is the number of trips for central i, and \( s_i \) is its share of the total number of taxis.

Assume for simplicity that the centrals are of even size. Then \( s_i = \frac{1}{s} \), \( x_i = \frac{X}{s} \), the matching function becomes

\[ X = \left( \frac{TX}{B} \right)^\alpha (N - X)^{1-\alpha} - s\gamma. \]

Equation (26) is identical to (7a) except that the parameter \( \gamma \) is multiplied by \( s \). This illustrates that with increasing return to scale, going from a one central to a multi central setting, leads to a clearly inferior matching process.

Even if we have two centrals with the same number of taxis, the competing centrals would not be able to replicate the second best solution with only one central, as the average matching process would be less efficient for \( \gamma > 0 \). The gain in competition may be outweighed, fully or partly, by reduced efficiency. The efficiency loss will depend on the degree of economies of scale.

With reference to figure 3, it is then obvious that the monopoly outcome and social optimum will not be affected as the optimal number of centrals is clearly 1. However, if price competition requires more than one central the second best optimum SB is beyond reach.
References


Brunstad, R. J.," Deregulering av drosjenæringen” («Deregulation of the taxi industry»), *Sosialøkonomen*, nr. 6 1990.

Brunstad, R. J. *The taxi market – excess capacity and insufficient supply*. Discussion paper 11/91. Norwegian school of economics and business administration December 1991,


