Regulated monopolies in urban public transport – Can we design proper regulations and incentives?

Odd I. Larsen

Institute of Transport Economics, Norway
Molde College, Norway

ABSTRACT

A model of the public transport company in Oslo is used to the design a system of price regulations and subsidies. The objective is to provide incentives for optimum provision of public transport services both for peak and off-peak demand. Optimum is defined in terms of fares, level of service and average capacity per revenue kilometre. The cost of public funds and the fact that car traffic is priced below marginal cost in peak periods are taken care of in the model. The "regulator" determines the fares, the transit operator receives a subsidy per revenue kilometre, differentiated between basic services and additional services operated only in peak periods. There is also a subsidy per passenger in peak periods due to "underpriced" car traffic. The results indicate that it should be possible to have regulated monopolies in local public transport that gives a proper supply of services based only on commercial incentives.

1. Introduction

The last decade or so we have witnessed a strong trend towards privatisation and deregulation in the market for public transport in many countries. Major reasons for the change in policy have been the inefficiencies documented in the sector and attributed to extensive regulations, public ownership and subsidies (Berechman 1996).

In local public transport in Norway this has led both to contractual reforms with some of the (old) concessionary monopolies and to a gradual replacement of contracts with tendering. Fares and the total subsidies are still determined by local authorities as part of the general budgetary process.

This paper deals with the issue of subsidies which was part of a study of contractual reform for Oslo Public Transport Company (OPTC). OPTC is a limited company owned by the municipality of Oslo. The company has a concessionary monopoly on public transport services within the City of Oslo and operates buses, trams and light rail.

Up till now the subsidies to the company have been given as a lump sum from the city treasury. In the annual budget proceedings the City Counsel determines the amount of subsidies and the level of fares. OPTC has a flat fare system with discounts for prepaid tickets and seasonal tickets and rebates for children and elderly people.

One of the tasks in the study was to propose a new mechanism for allocating subsidies and this is the main theme of this paper. Norheim (1999) gives a comprehensive overview of the whole study. The idea for the scheme proposed is based on the theory outlined in the following section.

2. The implications of a first best social optimum for public transport

Let:

$Y(X,p)$ be demand in number of boarding passenger pr unit of time as a function of fare (p) and revenue kilometres per unit of time (X)

and

$C(X,S,Y)$ = the cost of the transport system as a function of revenue kilometres (X), capacity per revenue kilometre (S) and the number of passengers per unit of time.

The number of revenue kilometres per unit of time can be taken as a proxy for the level of service. Increasing the number of revenue kilometres in a public transport system must imply higher frequency and/or better geographical coverage, i.e. denser network with shorter walking distances and more direct routes. Thus, while passengers directly respond to changes in different components of travel time, changes in the number of revenue kilometres might be a good indicator of this for a given area.

Social surplus of the transport system is defined by:

$$SSP = p \cdot Y(X,p) + \int_{P}^{p^m} Y(X,p)dp - C(X,S,Y) \qquad (1)$$

p^m is the maximal fare, i.e. the fare that will choke of all demand. We must expect p^m to depend on *X*. A first best optimum is defined by:

$$\underset{X,S,P}{Max\ SSP} \qquad st\ \ (\beta XS - Y(X,p)) \geq 0 \qquad (2)$$

The constraint on capacity can either be interpreted as a technical constraint or as a constraint related to the optimum degree of crowding. More realistically we may also divide transit riders in two groups; on group that travels on the sections of the network where the demand for capacity is determined and one group that only uses the parts where there is surplus capacity.

Maximisation gives us 3 conditions:

$$p = C_Y + \lambda \qquad (3)$$

$$pY_x - Y \cdot \frac{Y_X}{Y_p} - v(x,p) - C_X - C_Y \cdot Y_X + \lambda \cdot (\beta S - Y_X) = 0 \qquad (4)$$

$$-C_S + \lambda \cdot \beta X = 0 \qquad (5)$$

v(X,p) is defined by:

$$v(X,p) = \int_{p}^{p^m} Y \cdot \frac{\partial\left(\frac{Y_X}{Y_p}\right)}{\partial p} dp$$

The value of *v(X,p)* depends on the curvature of the demand function. With a linear demand function the term will be zero, and in general we may expect this term to be small. λ is the Lagrange-multiplier on the capacity constraint.

(3), (5) and the capacity constraint give us:

$$\lambda = \frac{C_S}{\beta X} = \frac{C_S \cdot S}{Y} \quad \Rightarrow p = C_Y + \frac{C_S \cdot S}{Y} \tag{6}$$

C_Y is mainly the cost of boarding and the second term in (6) can be interpreted as the cost of capacity per passenger. By pricing according to marginal cost and offering an optimum level of service, a public transport system will fall short of full cost recovery. Some manipulations of (3) to (5) also show that marginal cost pricing and optimum level of service also implies that the fare can be written as:

$$p = \frac{C_Y + \frac{C_X \cdot X}{Y} + v(X,p) \cdot \frac{X}{Y}}{1 - \frac{e_X}{e_P}} \tag{7}$$

(e_X and e_p are demand elasticities with respect to revenue kilometres and fare)

Except for the last term in the nominator this expression is identical to the one given in Larsen (1982).

The crucial problem when it comes to public transport operation and the level of service is related to the terms:

$$-Y \cdot \frac{Y_X}{Y_p} - v(x,p) = -\frac{pY}{X}\frac{e_X}{e_p} - v(x,p) \tag{8}$$

These terms, which is the increase in consumers' surplus from a marginal improvement in the level of service, represent an externality to a private operator which attempts to maximise profit. The operator will only be interested in the additional revenue from the increase in passengers, but the terms in (8) may be of the order 2-5 times the marginal increase in fare revenue. The terms are closely related to the Mohring-effect.

Some rearranging of terms also shows that (3)-(5) and the constraint imply:

$$-\frac{pY}{X}\frac{e_X}{e_p} - v(x,p) = C_X - \frac{C_S \cdot S}{X} \tag{9}$$

The right-hand side of (9) is approximately the cost of a revenue kilometre operated with a unit of minimum capacity. For a buss route this would come close to the cost of serving the route with a car. For railways this cost will approximate the cost per kilometre of operating a small locomotive.

(9) only holds in a “first best” world when prices are equal to marginal costs. However, (8) will be present as an externality to the operator as long as we have a system where revenue only comes from fares and the level of service is determined so at to equalise marginal revenue and marginal cost. If an operator receives a subsidy equal to the right hand side of (9) and maximise profit with respect to X and S, with fares equal to marginal cost, the operators profit should in this case be:

$$\text{Pr}\mathit{ofit} = C_X \cdot X + C_Y \cdot Y - C(X,S,Y) \tag{10}$$

With non-increasing returns to scale we should according to (10) have full cost recovery.

This simple scheme points to two crucial and well-known facts:

1. Marginal cost pricing and optimum level of service will not allow for cost recovery and the deficit will have the same magnitude as the drivers wage bill.
2. Irrespective of fares, private operators will offer a substandard level of service as long as there are no revenue matching the increase in consumers’ surplus when the level of service increases.

These conclusions should be valid for any type scheduled passenger transport, but the implications are most severe for local public transport where the subsidies in an optimum may reach 40 % of total cost or more.

If we for the moment disregard issues of second best and other complications that makes the above model unrealistically simple, it is interesting in itself to look for ways to implement a “good” solution in terms of social surplus. It is possible to imagine at least 4 broad alternatives:

1. *A planning solution.* This implies that a public authority decides on all parameters (X, S and p) and presumably also owns and operates the system. The main problems with this solution are:
   a) The information on demand needed in order to have an optimum design of the system in terms of routes, frequencies and timetables.
   b) The familiar problems of efficiency related to a publicly owned transit company receiving high subsidies.

2. *A planning solutions with many private operators.* In this case there exists several possibilities for tendering, depending on what the public authority decides in advance and what is left for tendering. The solution closest to a pure planning solution is that a public authority decides on all 3 parameters and chooses the bid with the lowest cost. However, these 3 parameters should actually depend on the cost of operating the system, i.e. on $C(X, S, Y)$ and the public authority are still left with the need for extensive information on the demand for public transport. In theory we can even also imagine that a public authority can accurately estimate the right hand side of (8). The only parameter fixed by the authority can then a subsidy to be paid per revenue kilometre, and p, X and S are left to be determined by tendering. In this case the winning bid should be the bid that implies the highest social surplus, but this presupposes that the public authority is able to undertake proper evaluation of such bids. The main problem with these solutions is the information needed by the public authority.

3. *A regulated monopoly.* In this case we can imagine that the public authority sets the fares and provide a subsidy equal to the right hand side of (8). This poses a problem of asymmetric information with respect to the operators actual costs. Information on the operators costs is a crucial input in fixing both fares and the subsidy per revenue kilometre.

4. *A competitive solution.* In this case the only parameter determined by the public authority should be the subsidy per revenue kilometre. In my opinion it is still an unresolved theoretical issue whether a competitive solution approximating a social optimum will emerge in such a setting. As opposed to a competitive solution *without subsidies* we should expect lower fares and an improved level of service. There is also evidence supporting the view that a competitive solution *without subsidies* will be unstable. A proper subsidy per revenue kilometre will facilitate exit and entry in the marked and to a large extent eliminate economies of scale that would otherwise impede competition. The crucial issue is the kind of route system that will emerge and the possible need for co-ordination of timetables.

The model presented above is too simple to actually merit a thorough evaluation of the different options with respect to organisation of public transport indicated above. The issue of X-efficiency is important when it comes to the operators cost. On the other hand, the efficiency of a public authority when it comes to the "design" of public transport services is also a problem. The competence of a public authority in this respect is an important issue when tendering is used.

A subsidy per revenue kilometre will seemingly play a major role if we want private operators to provide an optimum level of service. The drivers wages will constitute the major share of the cost that is the basis for the subsidy. In a system with routes operated at different speeds and thus having varying cost per kilometre, it might actually be simpler to have a subsidy per effective hour of revenue service.

3. A model for the Oslo Public Transit Company (OPTC)

For the study reported on here, we updated and re-calibrated a model that I have previously presented in Larsen (1994). That paper also have a more detailed description of the model, which is basically an extension and numerically implemented version of the model presented above. The extensions are of two types.

On the demand side we have three categories of passengers:

- Passengers that travels on the sections of the route network that determines the capacity (per hour) for the system (Y^{H}).
- Other peak period passengers. (Y^{L})
- Off-peak passengers (Y^{OP}).

This means that there will be a "social revenue" for each type of passengers and the model allows for different fares for the three categories of passengers. An improvement would be to also have two segments of off-peak passengers, but data did not allow for this. The demand functions used implies price elasticities that increase with the level of fares and elasticities with respect to level of service that decrease with the number of revenue kilometres. Off-peak demand has slightly higher elasticities in absolute value in the benchmark situation.

On the supply and cost side we distinguish between "basic services" (B) that is all day services and additional services (A) that is operated only in peak periods. This is the same approach as Jansson (1979,1984) The cost per revenue kilometre (for a given capacity per revenue kilometre) is much higher for the additional service than for the basic service due to higher capital and labour cost and less efficient use of rolling stock.

The model allows for different capacity per revenue kilometre for A and B and has two constraints:

$$\begin{aligned} &peak: && \beta^{p} \cdot (X^{B} \cdot S^{B} + X^{A} \cdot S^{A}) - Y^{pc} \geq 0 \\ &off-peak: && \beta^{o-p} \cdot X^{B} \cdot S^{B} - Y^{o-p} \geq 0 \end{aligned} \tag{11}$$

Both demand, the level of service and costs are initially calculated on an hourly basis and multiplied by the number of hours to arrive at annual demand and cost. Capital cost are calculated directly on an annual basis and there is also a fixed cost per hour the system is operated and a fixed cost per year that is related to the rail system.

The model also allows for a cost of public funds and for second best solutions conditioned on road traffic that is priced below marginal cost in peak periods.

A major simplification made in the model is it treats services as homogeneous while OPTC operates busses, trams and light rail services. However, this is probably not big problem with the cost structure used in the model.

By solving a problem that is actually a problem of non-linear optimisation with non-linear constraints we can numerically find the social optimum for "1. best", with cost of public funds and second best when road traffic is priced below marginal cost. It is also possible to add additional constraints.

There are two features with this model that are worth emphasising:

1. Marginal cost for Y^H is determined by the cost of expanding the capacity of the additional services that are operated only in the peak, but average cost for peak passengers is determined by the mix of basic and additional services operated in the peak.
2. If we consider the cost of expanding services off-peak it means expanding the basic services. The additional services operated in the peak can then be reduced without violating the constraint on capacity and the cost savings from this reduction should be deducted from the cost of expanding the basic services. The implication is that the marginal cost of expanding services off-peak will be lower than the average cost of the basic services, and this also carries over to the marginal cost per passenger.

The model was calibrated to the benchmark situation that is represented by cost, fares, level of service and demand in 1996.

The idea behind using the model was to estimate fares and level of service for a social optimum. Subsequently to use the estimated fares and a subsidy per revenue kilometre for basic and additional services respectively and let the model calculate the solution for a profit maximum. According to the model presented in section 2 we expected that it would be quite easy to replicate approximately a 1. best social optimum in this way, which was also confirmed by model runs. However a more realistic

case would involve both a cost of public funds and a second best solution conditioned on car traffic priced below marginal cost in peak periods. This is the actual situation in the Oslo area even with present toll ring.

Table 1: Maximising social surplus

		CPF=0		**CPF=0.25**	
	"Benchmark situation"	**1. best**	**2. best**	**1. best**	**2. best**
Public transit trips-Mill/year :	A	B	C	D	E
Peak – demand on capacity	24,75	22.4	26.8	17.7	24.9

Peak – other	8,25	11.9	12.1	9.4	9.8
Off – peak	77,0	120.7	120.7	93.5	93.5
Total number of trips	110,0	155.0	159.7	120.5	128.1
Fares – NOK per trip:					
Peak – demand on capacity	8,20	18.00	12.10	25.30	14.30
Peak – other	8,20	1.00	1.00	8.30	8.30
Off – peak	8,20	7.50	7.30	11.90	11.90
LOS-1000 rev.km per hour[1]):					
Basic services	3.0/145	10.4/64	10.4/65	8.6/61	8.6/61
Additional peak services	4.3/185	4.6/87	6.5/98	4.0/83	6.9/101
Peak total	7.6	16.1	17.9	13.4	16.4
Costs – Mill NOK per year:					
Operating costs (dep. on rev. km)	1049.0	1940.3	2005.0	1552.7	1659.5
Capital costs	206.9	194.6	231.6	154.4	214.6
Fixed cost per trip	111.7	157.3	162.1	122.3	130.0
Dependant on operating hours	32,9	32.9	32.9	32.9	32.9
Fixed annual cost	7,5	7.5	7.5	7.5	7.5
Total cost	1408,0	2332.5	2439.0	1869.7	2044.5
Average cost per trip - NOK	12,80	15.00	15.30	15.50	16.00
Fare revenue–Mill NOK/y:					
Peak – demand on capacity	203,2	402.9	325.9	445.8	354.6
Peak – other	67.7	12.1	12.3	77.8	81.0
Off – peak	632.2	903.6	902.9	1107.9	1106.9
Total revenue	903.1	1318.6	1241.1	1631.5	1542.4
Economic results:					
Profit–excl.subsidies-MillNOK/y	-504.9	-1013.9	-1197.9	-238.3	-502.1
Social surplus - Mill NOK /y		2867.8	2854.3	2180.7	2138.3

1) Revenue kilometres/Capacity per revenue kilometre. Due to the fact that the basic services are slightly higher on workdays than in the week-ends, the peak revenue kilometres comes out slightly higher than the sum of basic services and additional peak services.

Table 1 shows the benchmark situation, the 1.best social optimum, the social optimum when the cost of public funds (CPF) is 0,25 and the social optimum when we add a social revenue for passengers that transfer from car to public transport in the peak. When the social surplus is calculated with the cost of public funds, consumers' surplus is divided by 1.25.

We notice that the benchmark situation have the same average fare for the three categories of demand. On board surveys by OPTC showed that this was the case with the mix of ticket types and rebated groups actually travelling. The single ticket adult fare was 18 NOK in 1996, i.e. very close to the estimated marginal cost for Y^H. From the 1.best solution we also see that the average off-peak fare is close to the estimated marginal cost which is equal to the fare in the first best solution.

By coincidence the 1. best solution with cost of public funds greatly reduce subsidies from the present level, but the 2. best solution in this case brings the subsidies back to the present level.

We also notice from the last column in Table 1 that the fare for Y^H is below marginal cost. This points to another problem if we want to design a system of subsidies that gives proper incentives to the operator. No commercial operator will be interested in transit riders that pay a fare below marginal cost. This is an argument for supplementing a subsidy per revenue kilometre with a subsidy for each passenger in this segment.

What is clearly evident from Table 1 is that the present situation implies a "wrong" trade-off between operator costs and users costs. OPTC provide too low level of service in terms of revenue kilometres and too high capacity per revenue kilometre. This situation is not a result of poor performance by the company, but purely a result of political constraints in terms of regulated fares and allocated subsidies. When these two constraints are added to the constraint on capacity, virtually no degree of freedom is left for the company. However, as for any transit company in this type of setting, we can suspect that production efficiency can be improved.

On the other hand, as also pointed out in Larsen (1994), improved X-efficiency does not necessarily imply reduced subsidies in an optimum. Marginal cost and thus optimum fares will also decrease and the level of services should improve. These adjustments will counteract the cost reduction when it comes to the level of subsidies.

The model runs in Table 1 all show that the level of service should be improved at the expense of higher average cost per passenger. Optimum level of service will actually reduce the cost per revenue kilometre and increase the cost per passenger. This impact on costs points to a general problem of assessing the efficiency of public transport services.

The last column in Table 1 was taken as the starting point in looking for a subsidy scheme that could be applied to a profit maximising operator.

4. Transit operator with subsidies.

We assumed that OPTC would remain a monopoly. This meant that fares should be regulated. Here I will only report on three of the options we tested. The parameters are shown in Table 2.

Table 2: Fares and subsidies

	Alternative A	Alternative B	Alternative C
Fares:			
Peak – demand on capacity	15.00	15.00	8.20
Peak – other	12.00	11.00	8.20
Off – peak	12.00	11.00	8.20

Subsidy per passenger:			
Peak – demand on capacity	4.50	4.50	10.00
Peak – other	0.00	0.00	10.00
Off – peak	0.00	0.00	10.00
Subsidy per revenue km			
Basic services	10.00	11.00	0.00
Additional peak	20.00	20.00	0.00

Alternative C was tested because the management of OPTC was interested in scheme that gave more direct incentives for increasing the number of transit passengers. The fares in Alternative A and B reflect the fact that the present ticketing system has limited possibilities for differentiating the fares.

If we compare the results in alternative B with the last column in Table 1 we see that this alternative comes very close to the social optimum both with respect to demand and to the level of service. On the other hand, alternative C is a long way off with respect to the level of service and has lower demand than B even though the fares are lower (present fares). An incentive system based only on a subsidy per passenger is clearly inferior to subsidies per revenue kilometre, as should be expected from the theory in section 2.

A point to notice is the high profits after subsidies. This indicates that it is possible to reduce the subsidies. However, this should not be done by reducing the rates per revenue kilometre or passenger. The subsidy rates should be set with the objective of getting the operators marginal trade-offs right. By itself this will lead to unnecessary high subsidies. The profit in alternatives A and B is mainly due to the fact mentioned above of fares equal to marginal cost for peak traffic refer to the cost of additional services operated only in peak. A major share of the capacity in the peak comes from the basic services that have substantially lower cost. The best way to deal with this issue might be to have a long term contract that specifies the rates of subsidies and also allows for a fixed deduction in the total amount of subsidies calculated from these rates.

Table 3: Maximising profit with subsidies

	Alternative A	Alternative B	Alternative C
Public transit trips-Mill/year :			
Peak – demand on capacity	24,3	25,0	24,2
Peak – other	8,8	9,3	8,1
Off – peak	85,3	94,9	87,7
Total number of trips	118,4	129,2	119,9
LOS-1000 rev.km per hour[1]):			
Basic services	6,5/73	7,9/67	4,2/116
Additional peak services	9,0/80	9,8/71	2,3/300

Peak total	16,2	18,5	7,0
Costs – Mill NOK per year:			
Operating costs (dep. on rev. km)	1455,3	1655,7	1180,8
Capital costs	209,6	217,5	203,3
Fixed cost per trip	120,2	131,1	121,7
Dependant on operating hours	32,9	32,9	32,9
Fixed annual cost	7,5	7,5	7,5
Total cost	1825,4	2044,7	1546,1
Average cost per trip - NOK	15,40	15,80	12,90
Fare revenue – Mill NOK/year:			
Peak – demand on capacity	364,6	374,4	152,9
Peak – other	105,6	102,2	66,3
Off – peak	1023,7	1044,2	719,6
Total fare revenue	1493,9	1520,8	984,7
Profit excl. subsidies , Mill NOK/y	-331,5	-523,9	-561,4
Revenue from subsidies;			
From revenue kilometres Mill NOK/y:	551,3	711,0	
From passengers Mill NOK/year	109,4	112,3	1199,5
Profit incl subsidies, Mill NOK/year	329.2	299.4	638.1

2) Revenue kilometres/Capacity per revenue kilometre. Due to the fact that the basic services are slightly higher on workdays than in the week-ends, the peak revenue kilometres comes out slightly higher than the sum of basic services and additional peak services.

An obstacle to implementation of this type of contract is that the local authorities will not be allowed to decide on the total amount of subsidy in the annual budgets. Instead they are obliged to pay a subsidy that mainly depends on the number of revenue kilometres that the operator find profitable to supply.

The results above clearly indicate that a system of subsidies can be devised that gives proper incentives for a transit operator. Such a system will probably be much better than any system where subsidies are given as a lump sum allocation.

The regulator is still left with the problems of asymmetric information and production efficiency in a monopoly. If the subsidies per revenue kilometre are set too high, the regulator runs the risk paying excessive subsidies for a level of service that exceeds the optimum. In the case of OPTC we also deals with a publicly owned company. A system of subsidies along the lines shown here, could also facilitate privatisation of the company.

A major advantage of having one operator is that different co-ordination of timetables automatically will be taken care of without further involvement by a regulator. The real challenge in the future would be to introduce some kind of competitive pressure in system while maintaining co-ordination.

References:

Berechman J. (1993): *Public Transit Economics and Deregulation Policy*. North-Holland/Elsevier.

Jansson, Jan O. (1979): "Marginal Cost Pricing of Scheduled Transport Services - A Development and Generalization of Turvey-Mohrings Theory of Optimal Bus Fares". Journal of Transport Economics and Policy, Vol XIII, No. 3, 268 -294.

Jansson Jan O (1984): *Transport system optimization and pricing.* John Wiley & Sons.

Larsen, Odd I (1983): "Marginal Cost Pricing of Scheduled Transport Srvices"**,** Notes and Comments**,** Journal of Transport Economics and Policy, Vol. XVII No. 3, 315 -317.

Larsen, Odd I (1996): "Efficient Public Transit Subsidies" in Hensher, King and Oum (eds): *Proceedings of the 7th World Conference on Transport Research*, Vol 3, pp 225-236

Norheim, Bård (1999):"Competitive pressure as an alternative to competitive tendering? The development of a performance contract in Oslo." (This conference)

Odd I Larsen, born 1945, Norwegian citizen.

Economist (cand oecon) from University of Oslo 1970, specialising in econometrics.
Since 1980 employed at the Institute of Transport Economics (TOI).

Present position at TOI: Chief Research Officer.
Main responsibility: Passenger demand models.

Associate Professor at Molde College, Norway.

Work experience includes urban and national transport models, cost benefit analysis, modelling road pricing and econometrics. Have also worked on several projects in developing countries, mainly in Sub-Saharan Africa.