PERFORMANCE-BASED SUBSIDIES –

AN ALTERNATIVE SUBSIDY REGIME FOR PASSENGER RAIL TRANSPORT

Nils Fearnley and Jon-Terje Bekken,
Institute of transport economics, Oslo, Norway

Summary

The Norwegian Ministry of Transport's (MoT) 2002 White Paper on Public Transport paved the way for the introduction of performance-based contracts for rail passenger service obligations (PSO). The rationale was that such contracts provide a better allocation of risk and responsibilities between authorities and the operators, and will give further incentives for cost and market efficient solutions in passenger rail operations.

We have used a simulation model for public transport in order to estimate socially optimal subsidies, which encourage Norwegian State Railways (NSB) to increase patronage and to take external costs and benefits into account. By optimal subsidy regime we mean that the arrangement combines the social surplus maximising objectives of the MoT with the NSB's commercial goals, such that net social surplus (NSS) is maximised compared to the present situation. Within the model NSS comprises: (1) change in NSB's profit (producer surplus); (2) change in passengers benefit (consumer surplus); (3) changes in environmental and congestion costs; and (4) resource cost of public funds.

Through a series of model iterations we have designed a combination of subsidies per passenger, per train kilometre and for seat capacity. With these incentives, NSB will, on a commercial basis, strive towards service levels that resemble the social surplus maximising levels of service. The subsidies internalise the benefits to existing and new passengers and the effect on road congestion of rail service improvements into NSB's decision criteria. They also reflect the fact that rail fares are regulated. The performance-based subsidies will bring about a welfare gain in the region of NOK 8 million (€1 equals about NOK 8). The subsidies received under this regime together with the passenger revenue, may exceed the costs of the operator. To cope with this, we recommend that a fee is charged for the right to operate under this contractual regime.

There is a risk that the incentives will motivate the operators to focus their efforts entirely on train kilometre, seat capacity and passenger numbers to the degree that they overlook other important aspects of service quality, like punctuality, cleanliness, information etc. As a safety net, we recommend a set of additional conditions and bonus/malus arrangements. Most important are (1) bonus/malus for train cancellations and delays, based on the principles of internalising passengers' benefits or costs into NSB's decision criteria; and (2) the threat of tendering if quality measurements and performance levels fall below certain levels.

Background and purpose

The existing contracts between the Norwegian Ministry of Transport and Communications (MoT) and the Norwegian State Railways (NSB, state-owned limited company) consist of three documents: a principal ("timeless") agreement, a 4-year framework agreement and the annual agreement on purchase of passenger service obligation (PSO) services. For the negotiations over a new framework agreement 2003-2006 the MoT wanted to introduce elements of financial incentives in the contract. The aim was to launch a performance-based contract (PBC) in which the incentives combine the social benefit maximising objectives of the MoT with NSB's commercial goals. The intention is that this brings about more efficient resource allocation.

The Norwegian Institute of Transport Economics (TOI) was commissioned to develop the main principles of such a contractual regime for the three Norwegian intercity lines (all originating in Oslo and about 180 km each). The aim was to design a contract in which the incentives combine the social surplus maximising objectives of the MoT with the Norwegian State Rail's (NSB) commercial goals. This is a novel and promising approach to contracting for public transport services. The project has been reported in Fearnley, Bekken and Norheim (2002a and 2002b)

Towards achievement of both allocative efficiency and X-efficiency in Passenger rail services

The problem of X-inefficiency has been a driving force behind the increased use of competitive tendering in public transport including passenger rail services. This problem is related to continuously making an operator produce a given output in a cost efficient manner (also referred to as technical efficiency). In Norway, the threat of tendering by itself seems to have had more or less the same impact on X-efficiency in urban public transport (Johansen et al 2001). Thus, the problem to facilitate X-efficiency seems to be well handled by competitive tendering or even by the threat of such. The main problem at present is to assure allocative efficiency without creating problems of X-inefficiency.

The problem of allocative efficiency is well known. When operators adjust service levels according to their own business considerations they will improve service levels only to the point where the cost of the improvement is exactly offset by the increased revenue (i.e., private marginal cost equals private marginal revenue). This will maximise the profits of the operator. From a social point of view this will lead to an inefficient resource allocation. The reason is that external costs and benefits are not accounted for. Most important is the fact that the operator only considers the extra revenue raised and not the fact that quality improvements benefit existing passengers as well. Further, there are external costs and benefits of private car and rail transport. The car is regarded as a substitute for train, such that changes in rail service levels or fares will have an impact on car use.

The incentive contract model which we have developed seeks to internalise the external effects into the operators commercial decision criteria, and thus to stimulate to a socially optimal level of service. In particular, the benefit to existing passengers and the cost of congestion is internalised into the operator's commercial decision criteria. Allocative efficiency should also reflect the facts that public funds have a resource cost and alternative use and hence a shadow price. This is also included in our model.

TOI's simulation model for public transport systems

TOI has previously developed a simulation model for public transport, which has been applied to local public transport in several Norwegian urban areas (see e.g. Johansen and Norheim (1999 and 2000), Carlquist et al (1999) and Larsen (1993)). Its base principles are documented in Johansen et al (2001). We have further developed this simulation model in order to make it suitable for the Norwegian intercity rail market.

The model is based on and simulates the behaviour by a profit-maximizing operator. However, within the model, we have internalised different external effects. In particular, the benefit from existing passengers is included as well as additional costs and benefits related to transfer of car traffic. Furthermore, the model allows for inclusion of additional constraints related to capacity, fares, total amount of subsidies and minimum levels of service. This is important to make the model more credible.

The model calculates changes from a reference point. Thus the model estimates the *changes* in net social surplus rather than the overall social surplus. Formally, it is a matter of non-linear programming with non-linear constraints. Within the model, changes in net social surplus (NSS) comprises of:

1. Change in NSB's profit (producer surplus);
2. Change in passengers benefit (consumer surplus);
3. Changes in environmental and congestion costs (externalities), and
4. Resource cost (shadow price) of public funds.

The model considers three "periods" of demand: (i) Demand in peak periods at sections of the route where the capacity is at it's limit (Design capacity demand); (ii) Other peak-period demand; and (iii) Off-peak demand. Further, service levels are separated into two distinct categories: a base service level which runs throughout the operating hours, and the additional peak services that add to the basic services during peak periods.

A full NSS maximisation means that the model determines social optimal levels of 7 variables. Equally, profit maximization determines the profit maximizing levels of the same variables. The variables determined in the model are:

- Fare levels for the 3 periods of demand
- Train-kilometres produced in basic services and additional peak services
- Capacity provided per train kilometre in basic services and additional peak services

The model can next be used to identify a socially optimal subsidy regime. The optimal subsidy regime is the one that makes the profit maximising behaviour by the operator resemble the situation of maximum social surplus. To make the operator behave optimally, different incentives must be applied. By identifying the socially optimal adjustments first, and then to run the model to simulate a profit maximising operator's behaviour under different incentives, we can adjust the incentives until the profit maximising operator acts in a socially optimal way.

Model specification and calibration

The model looks at changes from a reference point. This reference point refers to a certain budget year for NSB. Both a cost function and a demand function have to be calibrated as to calculate the changes from this reference situation.

The cost function consists of a few but important elements. These are:

- Mileage-related costs, which include personnel, maintenance, energy and cleaning. These costs are assumed to increase proportionally with train size. Mileage-related costs are higher for extra services (during peak) mainly because of less efficient utilisation of labour and capital.
- Passenger-related costs, which are related to ticketing costs.
- Design capacity costs. This is the capital cost of the train fleet that is needed in order to cater for the peak demand periods. This cost depends on the fleet size, its re-purchase value, and the amortisation factor. An important feature is that the marginal design capacity cost will be much higher during peak hours compared to off-peak periods.
- Costs depending on operating hours. These are fixed costs of keeping the system up and running and do not vary with service levels.

The calibration of the cost function is based on data made available by NSB. The starting point to which the cost function is calibrated, resembles the actual costs reported by NSB for a reference year. The calibrated cost function is different between the basic services offered and the additional peak services. This is primarily related to the design capacity cost, which is designed to cater for the peak demand periods and the less efficient utilisation of labour and capital during peak services.

Demand for intercity rail services is determined by the following aspects of the journey; price, service frequency, crowding, and to some extent travel time. Conditions and circumstances outside the railway (like petrol prices) are not explicitly part of the model, but are mirrored in the elasticities that form the basis for calibration of the model. Demand is a function of generalised travel costs, G. We have chosen the following functional form.

$X(G) = V * \exp[(-a/b) * G^b] = X(G) = V * e^{(-a/b)G^b}$

where X is number of passengers per route per hour
V, a and b are constants to be calibrated. a, b > 0

This function makes the elasticity of demand increase when G increases. This is a property of public transport demand which we find in many market studies, see e.g. Carlquist and Fearnley (2001).

A special feature of this function, however, is that it only implicitly defines X, as G is again a function of X itself. Increased demand increase travel time and also crowding, both leading to higher generalised costs of travel. This implies that X is in fact an explanatory variable for X(G), and occurs both on the left and the right hand side of the equation.
The following demand elasticities can be derived.

The demand elasticity w.r.t. G increases as G increases:

$\varepsilon_G = \delta X/\delta G * G/X = -a * G^b < 0$

The demand elasticity w.r.t. the fare level, P, increases proportionally with the fare level, and at the same time the indirect effect via G is determined by the term (b-1):

$$\varepsilon_P = \delta X/\delta P * P/X = -P*a*G^{b-1} < 0$$

The demand function has been calibrated to the three different periods of demand. We have used data from NSB as input to the demand function, and used demand elasticities taken from the Norwegian national transport model to calibrate it.

In addition to the cost and demand functions, the simulation model takes into account the effect of changes in the railway on car use, and the resource cost of public funds. It is assumed that rail traffic lost to the car increases road congestion during the peak period (remembering that all the three Norwegian intercity lines originate in Oslo), and hence represents a cost to society (negative externality). This cost, however, is partly internalised and corrected by the fact that most car users into Oslo pass toll roads. The resource cost (shadow price) of public funds is set to 25 percent of the subsidies paid.

Different optima

The simulation model has been run to estimate optimal adjustments in NSB's intercity market given a number of assumptions about NSB's freedom to determine fares and service levels. *In theory* there are several possible scenarios to be estimated. In practice, however, there are a limited number of options for the model runs. For example, fare levels are, and are likely to continue to be, capped at 1 percentage point above the retail price index (i.e. RPI+1). The relevant model runs should therefore use this as a constraint on the fare level. Table 1 shows selected model runs. All figures are changes relative to the current situation with respect to fares, service levels, and financial performance.

Column I in Table 1 shows the social surplus maximising solution when fare levels are set to today's average of NOK 76. In this situation the peak service frequency is increased by app. 1 departure per hour. The seating capacity is also slightly higher than today. These changes bring about some increase in demand and a benefit to passengers of NOK 29m. The value of congestion relief resulting from the modal shift to train is estimated to NOK 2.2 m. The social surplus is reduced by NOK 22 m by the cost of raising public funds. The total estimated welfare gain compared to today's situation is thus NOK 9m per year. This is obtained with only a marginal change in subsidy level from the current situation.

Column II shows the results when the simulation model is specified to run a profit maximising, rather than net benefit maximising, scenario. Fares are not restricted in this model run. The resulting adjustment by NSB is characterised by particularly high fares, service levels that are dramatically reduced, and reduced seating capacity per train; a typical monopolistic behaviour. NSB will run a substantial financial surplus, but the cost to society exceeds this profit considerably. The change in social welfare is therefore negative, NOK -321m. The unregulated profit maximising solution is not at all desirable. Incentives, which motivate the profit-oriented operator to increase frequency and capacity, are needed.

Table 1: Model runs (all figures are changes from the starting point)

	I	II	III
	W-max, P=76	Profit max	Performance based subsidies
Fare, design capacity rush, NOK	0	+150	0
Fare, other rush traffic, NOK	0	+121	0
Off peak fare, NOK	0	+195	0
Services per hour, basic services	-0,2	-1,6	-0,1
Services per hour, rush hours	+1,1	-1,5	+1,3
Seats per train (capacity), rush	+29	-3	+3
Seats per train off peak	+19	-80	+2
Million passengers per year	+0,2	-3,3	+0,2
Total costs NOK millions	+19,4	-217,3	+15,9
Total revenues NOK millions	+12,5	+152,7	+11,4
Operating surplus NOK millions	-6,9	+370	+399,5
Change in consumer surplus NOK m.	+29,2	-756,4	+25,6
Congestion relief NOK millions	+2,2	-13	+2,2
Change in cost of public funds, NOK m.	-22,4	+448,7	-19,5
Total welfare gain NOK millions	+9,0	-320,7	+8,4
Subsidy for train kilometre NOK m.			+145,4
Subsidy for rush passengers NOK m.			+60,9
Subsidy off-peak passengers NOK m.			+89,6
Subsidy for seat capacity rush NOK m			+17,4
Subsidy for seat capacity offpeak NOKm			+90,8
Sum performance based subsidies, NOK m			+404,1

(1 Euro is about 8 Norwegian Crowns (NOK))

Optimal incentives

Given the difference between the social optimal solution and the profit maximizing solution, the question is then; how can the MoT give the profit oriented operator incentives to strive towards socially optimal levels of service?

We know that fares are capped, and hence that NSB do not earn the full potential revenue from new passengers. Further, it is an explicit goal for the MoT to increase rail patronage. A subsidy related to patronage should therefore be part of the PBC scheme. Further and evidently, NSB can do little to reduce journey times, because this is largely determined by the quality and capacity of the rail track; a responsibility which lies with the Norwegian Rail Administration. Hence, passengers' travel time is not suitable for financial incentives. NSB's decisions may, however, have an impact on waiting time (changes in service frequency) and crowding (changes in seating capacity), which are important determinants of passengers' generalised costs, as described above. A subsidy arrangement which internalises these costs/benefits to existing and new passengers into NSB's decision criteria, will therefore be a step in direction of optimal incentives.

Column III, in the table above, shows how performance based subsidies can combine the social benefit maximising objectives of the MoT with NSB’s commercial goals. This is a result of trying and failing with incentives that are related to passenger numbers, train kilometre (a proxy for service frequency since the network is fixed) and seat capacity per train. The model is run in the same way as in column II, i.e. profit maximising, but with performance-based subsidies and pegged fare levels. The results in column III are obtained when NSB is offered:

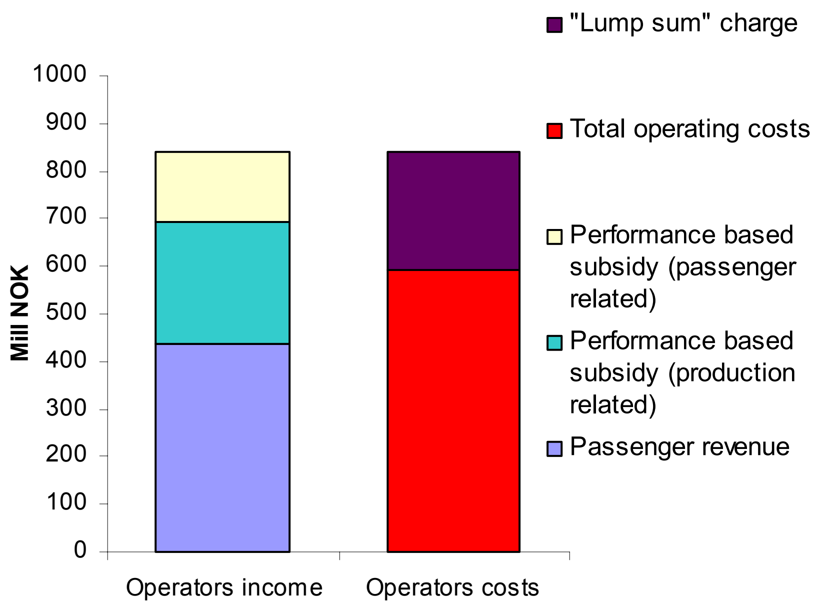
- NOK 30 (€3.75) per passenger during the rush hours,
- NOK 24 (€3) per passenger off peak,
- NOK 45 (€5.6) per train kilometre for *extra* peak services,
- NOK 29.5 (€3.7) per train kilometre for basic services,
- NOK 30 (€4) per 1000 seat kilometre during the rush hours, and
- NOK 80 (€10) per 1000 seat kilometre off peak,

With these incentives NSB, on a commercial basis, will strive towards service levels that to a large degree resemble the social benefit maximising levels in column I. The passenger compensations can alternatively be NOK 26 (€3.25) per passenger regardless of peak/off-peak. While being far easier to administrate this will only alter the result marginally.

We see that NSB will gain a large operating surplus, which mainly reflects the transfers from the MoT. Performance-based subsidies amount to NOK 404m per year. By charging NSB a lump sum fee equal to the operating surplus, for the right to operate under these performance-based subsidies, the net subsidies at the starting point will be more or less equal to the current subsidies paid to NSB

Illustration 1 gives an overview of this performance based subsidy regime at the starting point. The income of the operator, NSB, is equal to the passenger revenue plus the performance-based subsidies. The costs are equal to the ”lump sum” fee paid by NSB to the authorities for the right to operate under this regime plus the actual operating costs. The net subsidies paid by the authorities are equal to the difference between the performance-based subsidies paid and the “lump sum” payment received.

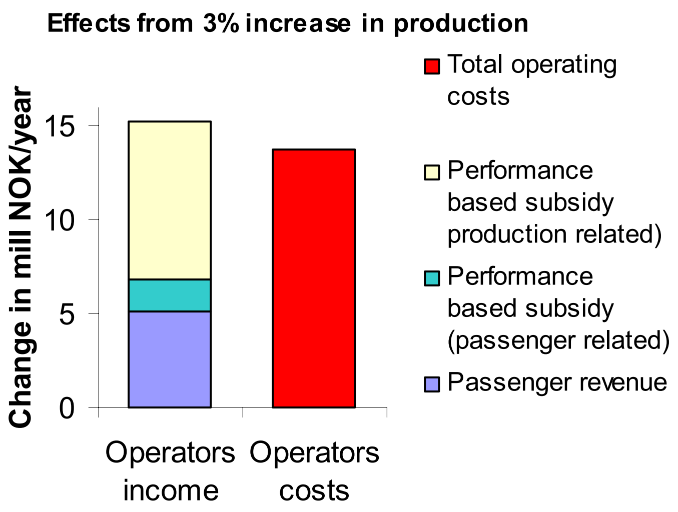
Illustration 1: The performance based subsidy regime at the starting point



By having the ”lump-sum” charge defined in the contract over a period of time, this regime gives the operator great discretion to adjust towards the social optimum. Reducing costs can increase profits; hence the regime avoids major problems of X-efficiency. Furthermore, by moving the production in the direction of the welfare maximizing solution, increased performance based subsidies can be achieved; hence allocative efficiency can be achieved.

The goal of the performance based subsidy regime is to make the operator adjust towards the social optimum. In Illustration 2, the effects from a 3% increase in production (vehicle kms) are illustrated. The increased passenger revenue does not offset the increased operating costs. Hence, without the performance based subsidies there will be no incentives to increase production. However, with the performance-based subsidies, the operator will be in a position to increase his profits by adjusting in the “right” direction.

Illustration 2: Example of changes for the operator related to a 3% increase in production



Safety net

There is a risk that the incentives will motivate the operator to focus their efforts entirely on train kilometre, seat capacity and passenger numbers to the degree that they overlook other important aspects of service quality, like punctuality, cleanliness, information, staff friendliness etc.

NSB already operate near the track capacity limit on several of the lines. If they choose to put more trains on the network at certain times of the day in order to pocket more subsidies for train-kilometres, then this may jeopardise punctuality levels. We have therefore recommended a bonus/malus arrangement for punctuality in addition to the above arrangement. This should be based on passenger delays, in order to internalise their valuation of changes in punctuality. Our recommended bonus/malus is calculated as follows: change in aggregated minutes of train delays from the base year *multiplied by* an agreed average number of passengers per train *multiplied by* passengers’ valuation of delay time. Our estimate for the latter, based on previous Norwegian studies (Norheim and Stangeby 1993 and Kjørstad 1995), is NOK 1.67 per minute.

Cancelled trains are particularly exasperating for passengers. Norwegian studies (Ruud and Frøysadal (2002), Nossum (2003), Norheim (1996)) suggest that passengers place a value of waiting time for cancelled and delayed services which is three times higher than the value of ordinary waiting time. Therefore NSB should be given strong incentives against train cancellations. We recommend, therefore, a malus per cancelled train-kilometre that is equal to 3 times the subsidy per train-kilometre.

As a "safety net" NSB and MoT should agree on a set of threshold-values, which entitle them to renegotiate the contract. For example, the contract should be renegotiated if the malus reaches a certain limit. Customer satisfaction surveys are also important in this respect. They can indicate the operator's performance on a range of quality aspects, and their overall performance. We recommend that customer satisfaction scores below a certain percentage of a base value (to be decided) should lead to renegotiation of the contract, or tendering.

Conclusions

We have developed a new and promising way to combine the social goals of the Ministry of Transport, who regulates and purchases train services, with the rail operator's profit maximising objective. This is called a performance-based contract. A main feature of the contract is the fact that incentives paid per train kilometre and per seat kilometre internalises existing passengers' benefit from increased service frequency and reduced crowding, respectively, into the operator's financial decision criteria. Further, the passenger incentives help to internalise the external costs of car use during rush hours: a transfer of traffic from the roads to rail will benefit other car users. Passenger incentives also correct for the fact that the Ministry is regulating fares.

This performance-based contract, however, will result in performance-based subsidies well above the current level of subsidies. This is necessary to make the incentives strong enough. To cope with this increased subsidy level, we recommend that a lump-sum fee is charged for the right to operate on this contract. This approach combines the necessary strength of the incentives with the budget constraint of the authorities.

Future expansions of this subsidy regime may be to tender such a performance-based contract. A promising way to do this would be to grant the right to operate under such a contract to the operator willing to pay the highest fee.

References

Carlquist, E. and N Fearnley (2001)
Welfare maximisation in public transport? An analysis of the developments in 7 Norwegian cities. TOI report 508/2001. Oslo: Institute of Transport Economics

Carlquist, E, T Hagen, A Hoelsæter, O I Larsen and B Norheim (1999)
Quality Contracts in Hordaland County - Discussion of alternative contract forms. TØI report 452/2000

Fearnley, N, J-T Bekken and B Norheim (2002a)
Developing quality contracts for the Norwegian State Railways' intercity market. Documentation report. TØI report 608a/2002

Fearnley, N, J-T Bekken and B Norheim (2002b)
Developing quality contracts for the Norwegian State Railways' intercity market. Summary report. TØI report 608/2002

Johansen, K.W., O I Larsen and B Norheim (2001)
Towards achievement of both Allocative Efficiency and X-Efficiency in Public Transport. In Journal of Transport Economics and Policy, Vol 35 Part 3, September 2001, pp 491-511

Johansen, K.W. and B Norheim (1999)
Quality Contracts - Discussion of output-based subsidy contracts. TØI report 455/2000

Johansen, K.W. and B Norheim (2000)
Output-based funding of urban public transport . Model simulation of alternative incentives and financial framework. TØI report 484/2000

Kjørstad, K. (1995)
Bus Passengers' Valuation of Service Improvements. Experiences from Moss, Grenland, Kristiansand, Tromsø and Ålesund. TØI report 312/1995

Larsen, O I (1993)
Samfunnsnytte av tilskudd til kollektivtrafikk. Oslo, Transportøkonomisk institutt. TØI report 208/1993

Norheim, B. (1996)
Better Public Transport - Documentation and Testing of the Stated Preference Method. TØI report327/1996

Norheim, B and I Stangeby (1993)
Bedre kollekivtransport. Oslo-trafikantenes verdsetting av høyere standard. TØI report 167/1993

Nossum, Å. (2003)
Public transport in the Oslo region - Travellers' vaulation of time. TØI report 633/2003.

Ruud, A and E Frøysadal (2002)
Targeted public transport: Handbook. TØI report 587/2002