1

Optimal pricing and financing of rail passenger services 021207

Kjell Jansson

1 INTRODUCTION

Much concern has been dedicated to rail services, both from the general public and politicians in most countries. Should rail services be state-owned or private? Should they be supported financially? Are rail services financially and socially viable or are they obsolete? Paradoxically there are two simultaneous trends, rail lines are shut down and high-speed tracks and trains are introduced. Financially no entire national railway system is profitable, especially if infrastructure costs are taken into account, even though certain lines may be very successful even from financial point of view.

During the last two decades another trend has flourished: deregulation. In Western Europe this trend commenced within local and regional public transport. The privatisation of the English bus industry, even the long distance coach services, represents the "full market solution", where both supply, prices and the operation are in the hands of competing profit maximising firms. In the Nordic countries and in London, and to some extent in the US and France, the decision over local and regional public transport supply and prices has been kept in the hands of a public authority, while the actual operation is left for competition through tendering. Typically these services need local or central government grants for financing – and there are economic rationales for this.

Rail transport has in most countries so far been left in the hands of government controlled bodies, but two exceptions are Great Britain and Sweden. In Great Britain the railway has been split into Railtrack (first public sector organisation, then privatised and then public sector organisation again) and private operators. In addition two other new organisations were created: the Office of Passenger Rail Franchising and the Office of Rail Regulation. In Sweden the Swedish Rail (SJ) has been split into a social welfare oriented Railway Administration (Banverket) responsible for infrastructure investments (financed by the government) and the "commercial" new SJ, with the aim to operate the service at a minimum profit determined by the government. SJ still enjoys monopoly for the commercially viable lines for passenger transport, while the non-viable ones are put out for tender by a new government agency (Rikstrafiken). With respect to freight transport there is free competition “on the track”. The reader should keep in mind that the author of this paper basically has the Swedish organisational form in mind, where the rail service operators are commercial, but most of the analysis is supposed to be general since welfare and profit oriented regimes are contrasted.

This paper does not aim to solve or even discuss all the issues related to rail transport. First we compare optimal price and quality according to welfare maximisation optimum and profit maximisation optimum respectively. Quality is here understood as wait time and seat capacity per departure. Wait time is a function of frequency of service and seat capacity to frequency and transport unit size. The discrepancies between optima provides the ground for the basic aim: to find corrective intervening measures that could make the profit maximising operator

2

behave according to welfare optimum criteria.

The analytical platform in this paper is joint determination of price and quality of public transport from a welfare point of view, an approach that was first presented by Mohring [1972]. The approach has then been followed by e.g. Turvey and Mohring [1975] and J. O. Jansson [1979], [1984] who deal with price and service frequency, using models which are most relevant for *frequent urban services* and assuming one passenger group. Nash [1978] optimises price and output in terms of miles operated for *frequent urban bus services*, contrasting maximum profit and maximum welfare solutions and assuming demand in terms of passenger miles to be dependent on price and bus miles operated. Panzar [1979] analyses *infrequent airline services*, assuming demand to be dependent on price and service frequency and allowing for a distribution of ideal departure times. These works take into account aggregate demand from all passengers, or from one representative group travelling the average distance, with no concern for where passengers board and alight. Jansson [1991] considers and contrasts *frequent and infrequent services*, and takes into account *a variety of passenger groups*. This work follows Jansson [1991] in the basic modelling approach, for one passenger group only, but contrasts welfare and profit maximisation.

The presumably closest link to this paper in terms of the corrective intervention is found in Larsen et al. [2001]. In that paper a production-related correction of behaviour is analysed, assuming a capacity constraint, a cost function that is linear in transport unit size and pre-specification of the optimal price. Under these preconditions we find that it exists in fact no optimum since the profit function is identically zero.

Note that the reader does not necessarily have to have rail service in mind. The analysis and the results are valid for all kinds of scheduled public transport. For this reason we sometimes use number of carriages, but mostly transport unit size for the unit size variable. The variable number of carriages thus corresponds to number of seats when we think about bus, coach, ship or airline services. The policy variables that are optimised are price, service frequency and transport unit size per departure.

First the following two situations are modelled and compared:

- A welfare maximising authority determines simultaneously price, service frequency and number of carriages per train. This case corresponds to vertical integration under a government welfare maximising monopoly.
- A profit maximising operator determines simultaneously price, service frequency and number of carriages per train. This case corresponds to a free market without any government intervention.

It is found that:

Optimal price is lower and quality higher under the welfare maximisation scheme compared to the profit maximisation scheme.

Then we assume the following situation:

- A welfare maximising authority, which could be the rail track authority, determines infrastructure charges and possible consumer subsidies. This case corresponds for example to the Swedish organisation of railway services where an independent profit maximising operators on the one hand and a welfare maximising railtrack authority on the other are vertically separated.

3

The questions raised here are:

a) Is it possible to apply tendering with incentive schemes in terms of a fixed amount of money related to consumption or production such that profit maximising operators behave according to welfare maximising criteria.
b) Could such incentive schemes be applied as an alternative to procurement.

It is found that:

- A consumption-related subsidy can yield a welfare optimal solution without any constraint on the policy variables. This scheme gives the largest freedom for the operators.
- A production-related subsidy can yield a welfare optimal solution when ride time cost is assumed to be a function of capacity use, given that both optimal price and optimal transport unit size are pre-specified.
- A production-related subsidy can yield a welfare optimal solution when a capacity constraint is applied, given that optimal price is pre-specified, but *only if* the cost function with respect to transport unit size is convex.

We will first, in section 2, define the basic prerequisites and assumptions, including the definition of the passengers' price and quality attributes. Section 3 deals with optimality according to welfare and profit maximisation respectively. Section 4 analyses the possibilities for corrections of a profit maximising operator's non-optimal behaviour. Section 5 summarises the results, including comparisons of corrective subsidy policies and a discussion on possible real world applications.

2 BASIC PREREQUISITES AND ASSUMPTIONS

2.1 Notation

- p is the price for a trip,
- s is a subsidy paid to the operator per journey,
- σ is the number of seats per carriage,
- F is frequency in number of departures per hour,
- N is the number of carriages per train,
- σN is the number of seats per departure, where N=1 for other modes than train,
- I is the variable infrastructure cost per departure,
- e is the external cost per departure,
- f is the fee paid by the operator per departure, including infrastructure costs, external costs and a possible production-related subsidy,
- r is the ride time,
- c_L is a cost proportionate to number of passengers, mainly sales costs,
- φ is the vector of the travel time components,
- ϕ is the monetary time value,
- ϕ^{τ} is the monetary time value of frequency delay (wait time),
- τ is frequency delay,
- T^T is the cost of frequency delay,

4

- C_0 is the fixed cost per departure,
- $C_1[N]$ is the variable cost per departure directly related to the size of the transport unit, i.e., number of carriages, personnel costs, energy, platform size etc.,
- X is the number of passengers during a period of time, thought of as one hour.

Note that even if we primarily have railway service in mind, the analysis is valid for all kinds of public transport. N could be interpreted as number of carriages per train but also generally as vehicle size per transport unit if multiplied by the constant σ, for example size of a bus or an aircraft.

We introduce the following notation:

$$\varepsilon_p \equiv \frac{\partial X}{\partial p}\frac{p}{X}$$ for own price elasticity,

$$\varepsilon_F \equiv \frac{\partial X}{\partial F}\frac{F}{X}$$ for frequency elasticity,

$$\varepsilon_N \equiv \frac{\partial X}{\partial N}\frac{N}{X}$$ for elasticity with respect to number of carriages.

Differentials are written in two different ways, either by use of deltas or by use of sub-index, e.g.,

$$\frac{\partial X}{\partial F} \equiv X_F$$

Arguments of functions are throughout delimited by [], while polynoms are delimited by ().

2.2 Authorities

The Public Transport Authority (PTA), which could be a rail track authority, is assumed to be welfare maximising.

The PTA is assumed to be responsible for investments and maintenance of tracks, electricity distribution and allocation of slots (time spaces between departures) for rail operation. The PTA is also charging an infrastructure user charge, f.

2.3 The operating firms

The operating firms may be either welfare or profit maximising. In both cases it optimises prices, service frequency and the number of seats or carriages.

Generally we will assume efficiency in production, i.e., that any level of output is produced at minimum cost, irrespective of whether the actual producer is welfare or profit maximising. The focus is put on consumption efficiency related to determination of optimal prices, optimal frequency, optimal number of carriages and possible subsidies and infrastructure

5

charges.

Demand is assumed being specified for certain periods, such as the average weekday afternoon peak hour in wintertime, the average Saturday etc. Only one type of charge - a per-trip price - is considered.

The operating firm reaches decisions about relevant inputs and prices well ahead of implementation because of a necessary planning lag. All factors of production that are variable between decision and implementation are therefore considered relevant for the joint decision on the magnitude of policy variables. These factors include, we assume, the frequency and the size of transport unit, including the personnel required. Policy variables are thus considered to be continuous. This assumption is not very restrictive, neither with respect to frequency, nor with respect to transport unit size. A theoretical optimal number of carriages at 6.7 could in practice be 6 or 7 etc.

If C[N] denotes the cost per departure, total operating costs per hour for a line is:

$$(1) F(C_0 + C_1[N]) + c_L X$$

We interpret (1) as cost per kilometre. If one wants total cost, the distance would be multiplied by the cost per kilometre.

Infrastructure costs of rail operation is IF. Finally the operation gives rise to external costs eF. The operators are supposed to pay the infrastructure charge fF. It could be, or not, that f=I+e. There are also basic administrative and planning costs. These fixed costs are not taken into account in the model since they are not affected by the operation decisions.

6

2.4 The passengers

Aggregate consumers' surplus is expressed as a function of "generalised cost", $G = p + \phi\varphi$, where φ is the vector of the travel time components and ϕ is the vector of monetary time values, i.e., the marginal rates of substitution between price and travel time components. The vector φ is here assumed to comprise riding time and "frequency delay", which is the time interval between ideal and actual departure time.

Riding time for group i is written r. The cost of riding time is then:

$$(2) T \equiv \phi[\frac{X}{FN\sigma}]r$$

ϕ is thus the value of riding time, assumed to be dependent on the occupancy rate, $D \equiv X/FN\sigma$, which is the number of passengers per seat in a group, and where it is assumed that $\partial\phi/\partial D > 0$ and $\partial^2\phi/\partial D^2 > 0$, so that the value of riding time is progressive with the occupancy rate. The number of passengers per departure is $q \equiv X/F$. We ignore that different passenger groups travel different distances, r. We interpret r instead as the average distance travelled. Subsequently we omit the ride time r, bearing in mind that ride time costs should be multiplied by the time r. We ignore that there are several passenger groups. This simplification will substantially facilitate the expressions derived without disturbing the purpose of the analysis. The optimal welfare maximisation conditions when taking into account various distances travelled by different passenger groups are found in Jansson [1991].

The marginal ride time cost with respect to passengers is specified as:

$$(3) \frac{\partial T}{\partial X} = \frac{\partial\phi}{\partial D}(\frac{X}{FN\sigma})\frac{1}{FN\sigma} \equiv \phi_D(Z)Z \equiv Y(Z)$$

where $Z = X/FN\sigma$ and Y is a short form.

The interval between departures is 1/F hours. Ideal departure times, t, are uniformly distributed within this interval, i.e., $0 < t \leq 1/F$. Frequency delay is then $\tau \equiv 1/F - t$. The cost of frequency delay for a group is $T^\tau[\phi^\tau, F, t] = \phi^\tau[1/f-t](1/F-t)$. This is the delay multiplied by value of time, which is a function of the delay. Subsequently we ignore the distribution of t and assume that the expectation value of the frequency delay cost is a function of frequency only, i.e. $T^\tau = T^\tau[F]$. Ignoring the distribution t drastically simplifies the analysis without affecting the purpose of this paper, but the full effects of taking t into account is found in Jansson [1991]. It may sometimes be useful to think about the frequency delay cost as wait time calculated as half the headway, 1/2F, and expected wait time cost as $\phi^\tau/2F$.

If p denotes the price, the generalised cost of travel for a group (index omitted) at time t is:

$$(4) G[p, F, N, s] \equiv p + T + T^\tau[F] \equiv p + \phi\left\lfloor\frac{X}{FN\sigma}\right\rfloor + T^\tau[F]$$

so that:

$$(5) G_p = 1 + T_X X_p$$

Demand per hour, X, for each group is a function of generalised cost:

$$(6) X = X[G[p,\ F,\ N, \sigma]]$$

We know that own-price elasticities, denoted ε_p, are negative, $\varepsilon_p<0$. We assume, based on solid empirical evidence, that demand elasticities with respect to frequency and transport unit size, denoted ε_F and ε_N, are such that $0<\varepsilon_F<1$, $0<\varepsilon_N<1$, implying that $\partial q/\partial F<0$, $\partial q/\partial N<0$. This rules out the unrealistic possibility that an increase in frequency or unit size would generate so many passengers that occupancy rate is unchanged or increases[1].Consumers' surplus is:

$$(7) S[G] = \int_{G}^{G^{max}} X[\rho] d\rho$$

and

$$(8) \frac{\partial S}{\partial G} = -X$$

G^{max} is the maximal reservation price in generalised cost terms.

We already here derive a number of important relationships with respect to frequency and number of carriages.

$$(9) \frac{\partial S}{\partial G}\frac{\partial G}{\partial F} \equiv -X\frac{\partial T^{\tau}}{\partial F} - X(\frac{\partial G}{\partial T}\frac{\partial T}{\partial X}\frac{\partial X}{\partial F} + \frac{\partial G}{\partial T}\frac{\delta T}{\delta F}) \equiv -X\frac{\partial T^{\tau}}{\partial F} - X(\frac{\partial T}{\partial X}\frac{\partial X}{\partial F} + \frac{\delta T}{\delta F})$$

$$(10) \frac{\partial S}{\partial G}\frac{\partial G}{\partial N} \equiv -X(\frac{\partial G}{\partial T}\frac{\partial T}{\partial X}\frac{\partial X}{\partial N} + \frac{\partial G}{\partial T}\frac{\delta T}{\delta N}) \equiv -X(\frac{\partial T}{\partial X}\frac{\partial X}{\partial N} + \frac{\delta T}{\delta N})$$

Expression (9) shows the consumer surplus change of a marginal change of frequency. The first term on the right hand side is the effect on frequency delay. This term can also be said to represent a positive external effect of public transport due to economies of scale in consumption. The second term reflects the effect on ride time cost. This cost is composed of two terms. The first one is the effect of frequency via demand, on crowding level on board. The second effect is the direct effect on crowding of a frequency change, denoted by use of the special delta δ. By differentiating (2) with respect to F we can also write the two effects on ride time cost as follows.

$$(11) \frac{\partial \phi}{\partial D}(\frac{FX_F}{F^2 N\sigma} - \frac{X}{F^2 N\sigma}) \equiv \frac{\partial T}{\partial X}\frac{\partial X}{\partial F} - \frac{\partial T}{\partial X}\frac{X}{F}$$

---

[1] Taking elasticity with respect to frequency as an example, we have:

$$\frac{\partial q}{\partial F} \equiv \frac{F\frac{\partial X}{\partial F} - X}{F^2} \equiv \frac{X(\frac{F}{X}\frac{\partial X}{\partial F} - 1)}{F^2} \equiv \frac{X(\varepsilon_F - 1)}{F^2}$$ which is<0 only if $\varepsilon_F<1$

8

Since X/F is larger than the differential of X with respect to frequency the direct effect (second term) is larger than the indirect one (first term).

Expression (10) shows the consumer surplus change of a marginal change of unit size. This cost is composed of two terms. The first one is the effect via demand, which in turn affects the crowding in the train. The second effect is the direct effect on crowding of a change of number of carriages, denoted by use of the special delta δ. By differentiation of (2) with respect to N we can also write the two effects on ride time cost as follows.

$$(12)\frac{\partial \phi}{\partial D}\left(\frac{NX_N}{FN^2\sigma}-\frac{X}{FN^2\sigma}\right)\equiv\frac{\partial T}{\partial X}\frac{\partial X}{\partial N}-\frac{\partial T}{\partial X}\frac{X}{N}$$

Since X/N is larger than the differential of X with respect to unit size the direct effect (second term) is larger than the indirect one (first term).

The direct effects on ride time cost of a marginal change of frequency and number of carriages respectively can thus also be expressed as:

$$(13)\frac{\delta T}{\delta F}=-\frac{\partial T}{\partial X}\frac{X}{F}$$

$$(14)\frac{\delta T}{\delta N}=-\frac{\partial T}{\partial X}\frac{X}{N}$$

Below we derive the basic marginal effects on demand and generalised cost with respect to price, frequency and transport unit size and some relations between them. These expressions will be useful for the interpretations of the results that will follow.

$$(15)\frac{\partial X}{\partial p}=\frac{\partial X}{\partial G}\frac{\partial G}{\partial p}\equiv\frac{\partial X}{\partial G}\left(\frac{\delta G}{\delta p}+\frac{\partial G}{\partial T}\frac{\partial T}{\partial X}\frac{\partial X}{\partial p}\right)$$

$$(16)\frac{\partial X}{\partial p}\left(1-\frac{\partial X}{\partial G}\frac{\partial T}{\partial X}\right)=\frac{\partial X}{\partial G}$$

$$(17)\frac{\partial X}{\partial p}=\frac{\frac{\partial X}{\partial G}}{\left(1-\frac{\partial X}{\partial G}\frac{\partial T}{\partial X}\right)}$$

$$(18)\frac{\partial G}{\partial F}=\left(\frac{\delta T}{\delta F}+\frac{\partial T}{\partial X}\frac{\partial X}{\partial F}\right)+\frac{\partial T^{\tau}}{\partial F}\equiv\frac{\partial T}{\partial X}\left(-\frac{X}{F}+\frac{\partial X}{\partial F}\right)+\frac{\partial T^{\tau}}{\partial F}$$

$$(19)\frac{\partial X}{\partial F}=\frac{\partial X}{\partial G}\frac{\partial G}{\partial F}\equiv\frac{\partial X}{\partial G}\frac{\partial G}{\partial T}\left(\frac{\delta T}{\delta F}+\frac{\partial T}{\partial X}\frac{\partial X}{\partial F}\right)+\frac{\partial X}{\partial G}\frac{\partial G}{\partial T^{\tau}}\frac{\partial T^{\tau}}{\partial F}$$

$$(20)\frac{\partial X}{\partial F}\left(1-\frac{\partial X}{\partial G}\frac{\partial T}{\partial X}\right)\equiv\frac{\partial X}{\partial G}\left(\frac{\delta T}{\delta F}+\frac{\partial T^{\tau}}{\partial F}\right)\equiv\frac{\partial X}{\partial G}\left(-\frac{\partial T}{\partial X}\frac{X}{F}+\frac{\partial T^{\tau}}{\partial F}\right)$$

9

$$(21)\ \frac{\partial X}{\partial F}=\frac{\frac{\partial X}{\partial G}(\frac{\delta T}{\delta F}+\frac{\partial T^{\tau}}{\partial F})}{(1-\frac{\partial G}{\partial T}\frac{\partial T}{\partial X})}\equiv\frac{\frac{\partial X}{\partial G}(-\frac{\partial T}{\partial X}\frac{X}{F}+\frac{\partial T^{\tau}}{\partial F})}{(1-\frac{\partial G}{\partial T}\frac{\partial T}{\partial X})}$$

$$(22)\ \frac{\partial X}{\partial G}=\frac{\frac{\partial X}{\partial F}}{T_X(\frac{\partial X}{\partial F}-\frac{X}{F})+\frac{\partial T^{\tau}}{\partial F}}$$

$$(23)\ \frac{X}{X_G}=\frac{XT_X(\frac{\partial X}{\partial F}-\frac{X}{F})+X\frac{\partial T^{\tau}}{\partial F}}{X_F}\equiv XT_X-XT_X\frac{\frac{X}{F}}{X_F}+\frac{X\frac{\partial T^{\tau}}{\partial F}}{X_F}$$

$$(24)\ \frac{\partial G}{\partial N}=(\frac{\delta T}{\delta N}+\frac{\partial T}{\partial X}\frac{\partial X}{\partial N})\equiv\frac{\partial T}{\partial X}(\frac{\partial X}{\partial N}-\frac{X}{N})$$

$$(25)\ \frac{\partial X}{\partial N}=\frac{\partial X}{\partial G}\frac{\partial G}{\partial N}\equiv\frac{\partial X}{\partial G}\frac{\partial G}{\partial T}(\frac{\delta T}{\delta N}+\frac{\partial T}{\partial X}\frac{\partial X}{\partial N})$$

$$(26)\ \frac{\partial X}{\partial N}(1-\frac{\partial X}{\partial G}\frac{\partial T}{\partial X})\equiv\frac{\partial X}{\partial G}\frac{\delta T}{\delta N}\equiv-\frac{\partial X}{\partial G}\frac{\partial T}{\partial X}\frac{X}{N}$$

$$(27)\ \frac{\partial X}{\partial N}=\frac{\frac{\partial X}{\partial G}\frac{\delta T}{\delta N}}{(1-\frac{\partial G}{\partial T}\frac{\partial T}{\partial X})}\equiv\frac{-\frac{\partial X}{\partial G}\frac{\partial T}{\partial X}\frac{X}{N}}{(1-\frac{\partial G}{\partial T}\frac{\partial T}{\partial X})}$$

$$(28)\ \frac{\partial X}{\partial G}=\frac{\frac{\partial X}{\partial N}}{\frac{\partial T}{\partial X}(\frac{\partial X}{\partial N}-\frac{X}{N})}$$

$$(29)\ \frac{X}{X_G}=\frac{XT_X(\frac{\partial X}{\partial N}-\frac{X}{N})}{X_N}\equiv XT_X-XT_X\frac{\frac{X}{N}}{X_N}$$

3 WELFARE AND PROFIT OPTIMA

3.1 Objective functions

Below we present the objective functions of the welfare maximisation and the profit maximisation models. The maximisation relates to one service during a period normalised to one (hour). The analysis may then be repeated for other periods and routes. The welfare objective function is expressed as:

10

$$(30) W = S[G[p,\ F,\ N]] + pX[p,\ F,\ N] - FC[N] - c_L X - IF - eF$$

The objective function for profit maximisation includes only producer's surplus, taking into account the infrastructure paid, f:

$$(31) \pi = pX[p,\ F,\ N] - FC[N] - c_L X - fF$$

Next we derive optima for price, frequency and number of carriages (or transport unit size). Optimal values will be denoted with superindex W for welfare optimum and with superindex π for profit optimum.

3.2 Welfare optimum

3.2.1 Prices

The first-order conditions with respect to price yields:

$$(32) W_p = S_G G_p + X + pX_p - c_L X_p = 0$$

By using (5) and (8) we get:

$$(33) p^W = c_L + XT_X$$

The welfare optimal price is thus composed of the marginal production cost plus the marginal passenger cost with respect to number of passengers.

3.2.2 Frequency

The first-order condition with respect to frequency yields:

$$(34) \frac{\partial W}{\partial F} = -X\frac{\partial G}{\partial F} + p\frac{\partial X}{\partial F} - c_L\frac{\partial X}{\partial F} - C - I - e = 0$$

Use of (18) and (23) yields:

$$(35) -\frac{X}{X_G} + p - c_L - \frac{C+I+e}{X_F} = 0$$

$$(36) p^W = c_L + \frac{X}{X_G} + \frac{C+I+e}{X_F}$$

Use of (23) yields:

$$(37) (p - c_L - XT_X) X_F = C + I + e - XT_X\frac{X}{F} + XT_F^{\tau}$$

From the first-order condition with respect to price we know that the left hand side is zero. Optimal frequency can thus be expressed as:

11

$$(38) F^W = \frac{X^2 T_X}{C + I + e + X T_F^\tau}$$

Combination of the first order condition with respect to price and (38) yields:

$$(39) p^W = c_L + \frac{F(C + I + e)}{X} + F T_F^\tau$$

The welfare optimal price can thus also be expressed in three parts: a) the marginal production cost with respect to number of passengers (mainly sales cost, plus boarding time for bus mode), b) the average variable operating and external cost, c) the marginal effect of a frequency change. The last terms thus correspond to the optimal price-related subsidy, defined as the difference between optimal price and average operating cost.

3.2.3 Transport unit size

The first-order condition with respect to unit size yields:

$$(40) W_N = -X G_N + p X_N - c_L X_N - F C_N = 0$$

Use of (24) yields:

$$(41) (p - c_L - X T_X) X_N = F C_N - X \frac{X}{N} T_X = 0$$

Use of the first-order condition with respect to price yields:

$$(42) N^W = \frac{X^2 T_X}{F C_N}$$

Combination of the first order condition with respect to price (33) and (42) yields:

$$(43) p^W = c_L + \frac{N F C_N}{X}$$

The optimal price is thus the marginal production cost plus the marginal cost with respect to unit size, per passenger.

3.2.4 Profit level

The objective function (30), plus (33), (38) and (41) yields several expressions for the profit level.

12

$$(44)\pi^{W} = pX[p,\ F,\ N] - c_L X - F(C + I + e) \equiv$$
$$\equiv (c_L + XT_X)X - c_L X - F(C + I + e) \equiv$$
$$\equiv (c_L + xT_X)X - c_L X - F(XT_X \frac{X}{F} - XT_F^{\tau}) \equiv$$
$$\equiv FXT_F^{\tau} \equiv$$
$$\equiv X^2 T_X - F(C + I + e) \equiv$$
$$\equiv FNC_N - F(C + I + e) \equiv F(NC_N - C) - F(I + e)$$

The profit level is thus negative since T_F^T <0. The same result is also seen from (39) if we multiply that expression with demand X. The (negative) profit level equals the total marginal cost with respect to frequency delay. It also equals the total marginal ride time cost with respect number of passengers minus the total operating cost, infrastructure cost and environmental cost. Finally it equals the marginal cost with respect to transport unit size minus the total operating cost, infrastructure cost and environmental cost

3.3 Profit optimum

3.3.1 Prices

The first-order condition with respect to price yields:

$$(45)\frac{\partial\pi}{\partial p} = X + pX_p - c_L X_p = 0$$
$$(46)\, p^{\pi} = c_L - \frac{X}{\partial X/\partial p}$$

By use of the price elasticity concept, optimal price can also be written as:

$$(47)\, p^{\pi} = \frac{c_L}{1 + \frac{1}{\varepsilon_p}}$$

By using the development of the differential of X with respect to p in (17) we can also write (47) as:

$$(48)\, p^{\pi} = c_L - \frac{X}{X_p} \equiv c_L - \frac{X(1 - X_G T_X)}{X_G} \equiv c_L - \frac{X}{X_G} + XT_X$$

3.3.2 Frequency

The first-order condition with respect to frequency yields:

$$(49)\frac{\partial\pi}{\partial F} = p\frac{\partial X}{\partial F} - c_L\frac{\partial X}{\partial F} - C - f = 0$$

$$(50)(p - c_L)\, X_F = C + f$$

13

Use of the first-order condition with respect to price and (23) gives:

$$(51)(p - c_L - XT_X + \frac{X}{X_G})X_F = C + f - XT_X\frac{X}{F} + XT_F^{\tau}$$

Since in optimum the left-hand side is zero we achieve:

$$(52) F^{\pi} = \frac{X^2 T_X}{C + f + XT_F^{\tau}}$$

The expression for optimal frequency under profit maximisation thus looks the same as the corresponding expression under welfare maximisation, given that f=I+e.

3.3.3 Transport unit size

The first-order condition with respect to transport unit size yields:

$$(53)\frac{\partial \pi}{\partial N} = pX_N - c_L X_N - FC_N = 0$$

$$(54)(p - c_L)X_N = FC_N$$

Use of the first-order condition with respect to price yields:

$$(55)(p - c_L - XT_X + \frac{X}{X_G})X_F = FC_N - XT_X\frac{X}{N}$$

Since the left-hand side of (55) is zero according to the first-order condition with respect to price we achieve:

$$(56) N^{\pi} = \frac{X^2 T_X}{FC_N}$$

The expression for optimal number of carriages under profit maximisation thus looks the same as the corresponding expression under welfare maximisation.

3.3.4 Profit level

By use of the objective function (31), and of (23) we calculate the profit level of the profit maximiser.

$$(57)\pi^{\pi} = (C_L + XT_X - \frac{X}{X_G})X - FC - C_L X - fF =$$

$$= X^2 T_X - X^2 T_X + X^2 T_X \frac{\frac{X}{F}}{X_F} - X^2\frac{T_F^{\tau}}{X_F} - (C + f)F$$

14

By using (52) and (56) we can express the profit level in the following ways:

$$(58)\pi^{\pi} = \frac{1}{\varepsilon_F} FNC_N(1-\varepsilon_F) - \frac{1}{\varepsilon_F} FXT_F^{\tau}(1-\varepsilon_F) \equiv \frac{1}{\varepsilon_F} F(C+f)(1-\varepsilon_F)$$

3.4 Diagnostic comparison between welfare and profit optima

Below we repeat and compare the various expressions for optimal price, frequency, transport unit size and profit level according to welfare and profit maximisation. In these comparisons we assume that the infrastructure charge paid by profit maximising operators equals the marginal infrastructure and external costs, i.e., f=I+e.

3.4.1 Prices

$$(59)p^{W} = c_L + XT_X \equiv c_L + \frac{F(C+I+e)}{X} + FT_F^{\tau}$$

$$(60)\ p^{\pi} = c_L + XT_X - \frac{X}{X_G} \equiv c_L + \frac{F(C+f)}{X}\frac{1}{\varepsilon_F} + FT_F^{\tau}(\frac{1}{\varepsilon_F} - 1)$$

For the welfare maximiser the marginal passenger cost with respect to frequency is withdrawn from the average cost. The profit under welfare maximisation is thus negative. The profit under profit maximisation is positive. If we compare the second and third terms in (59) and (60) we note that in (60) the average cost in the second term is divided by $\varepsilon_f<1$ and the third term in (60) is positive while in (59) it is negative.

3.4.2 Frequency

With respect to optimal frequency we know that the expressions, repeated below, are the same for welfare and profit maximisation, given that the infrastructure charge equals the infrastructure operating cost and the external operating cost, i.e., that f=I+e.

$$(61)F^{W} = \frac{X^2T_X}{C+I+e+XT_F^{\tau}}$$

$$(62)F^{\pi} = \frac{X^2T_X}{C+f+XT_F^{\tau}}$$

3.4.3 Transport unit size or carriages

Also with respect to optimal number of carriages we know that the expressions, repeated below, are the same for welfare and profit maximisation:

$$(63)N^{W} = \frac{X^2T_X}{FC_N}$$

15

$$(64)N^{\pi} = \frac{X^2T_X}{FC_N}$$

3.4.4 Profit level

$$(65)\pi^{W} = FXT_F^{\tau} \equiv FNC_N - F(C + I + e)$$

$$(66)\pi^{\pi} = F(C + f)\frac{1}{\varepsilon_F} - F(C + f)$$

The profit level is negative in welfare optimum and positive in profit optimum.

3.4.5 Relations between welfare and profit optima

We will now examine the relations between optimal price, frequency and unit size for welfare and profit maximisation respectively. In other words we want to know whether price, frequency and transport unit size is larger or smaller for welfare or profit maximisation respectively. For this purpose we employ Topkis' theorem.[1]

We differentiate the welfare and profit functions with respect to the variables X, F and U≡FN. That is, we transform the variables F and N into one, so that σU reflects the total capacity per hour in terms of number of seats.

We differentiate the objective functions (30) and (31) with respect to demand, X, and use (48) for the profit maximisation case to get:

$$(67)W_X = -XT_X - Xp_X + p + Xp_X - c_L \equiv p - c_L - XT_X$$

$$(68)\pi_X = p - c_L + Xp_X \equiv p - c_L - XT_X + \frac{X}{X_G}$$

$$(69)W_X - \pi_X = -\frac{X}{X_G} > 0$$

We differentiate the expression for generalised cost in (4) with respect to F and U (for fixed demand X):

$$(70)G_F = p_F + T_F^{\tau}$$

$$(71)G_U = p_U + T_U$$

We will also need the following differentials of the time cost components with respect to F and U, based on expression (2).

---

[1] Harald Lang made me aware of Topkis' theorem and carried out the analysis in this section.

16

$$(72)\, T_F = 0$$

$$(73)\, T_U = \phi_D()(\frac{-X}{U^2 \sigma}) < 0$$

By definition we know that

$$(74)\, T_F^{\tau} < 0$$

We differentiate the objective functions (30) and (31) with respect to frequency, F, and use (70) to get:

$$(75)\, W_F = -XT_F^{\tau} - C[N] - I - e$$

$$(76)\, \pi_F = Xp_F - C[N] - f \equiv -XT_F^{\tau} - C[N] - f$$

As long as the infrastructure charge, f, is equal to or larger than the marginal infrastructure and external cost, we have that:

$$(77)\, W_F \text{-} \pi_F = f - I - e \geq 0$$

We differentiate the objective functions (30) and (31) with respect to U, and use (71) to get:

$$(78)\, W_U = \text{-}\, XT_U$$

$$(79)\, \pi_U = Xp_U \equiv \text{-}XT_U$$

$$(80)\, W_U \text{-} \pi_U = 0$$

We then examine the cross-effects:

$$(81)\, W_{FX} = \text{-}\, T_F^{\tau} > 0$$

$$(82)\, W_{UX} = \text{-}\, T_U > 0$$

$$(83)\, W_{UF} = 0$$

Consequently we have that:

$$(84)\, X^W \geq X^{\pi}$$

$$(85)\, F^W \geq F^{\pi}$$

$$(86)\, F^W N^W \geq F^{\pi} N^{\pi}$$

This means that demand, frequency and capacity is the same or higher for welfare optimum compared to profit optimum and at least one of these variables is higher for welfare than for profit optimum. In other words quality, in terms of frequency and seat availability, and demand, are higher for welfare optimum than for profit optimum.

17

Note that this result holds only if f≥ I+e, i.e., if there is no intervention in terms of subsidising the production cost of the profit maximising firm.

Finally we need to know how optimal prices relate for welfare and profit optimum. We assume that ride time cost is convex in number of passengers. This is reasonable since it says that congestion cost grows at an increasing rate when more passengers are boarding. Let us then differentiate (3) with respect to capacity use (number of passengers pear seat available):

$$(87)\ Y_Z(Z) = \phi_{DD}(Z)Z + \phi_D(Z) > 0$$

The total marginal ride time cost is thus growing with the crowding level. Just like the demand elasticity with respect to frequency and number of carriages respectively are below 1, the demand elasticity with respect to capacity, FN, is below 1, that is:

$$(88)\ \frac{\partial X}{\partial U}\frac{U}{X} < 1$$

Let us assume now that $p^W \geq p^\pi$, in order to see whether this assumption generates a contradiction.

We use (86) and (88) in order to achieve:

$$(89)\ Z^W = \frac{X(p^W, N^W, F^W)}{N^W F^W \sigma} \leq \frac{X(p^\pi, N^W, F^W)}{N^W F^W \sigma} \leq \frac{X(p^\pi, N^\pi, F^\pi)}{N^\pi F^\pi \sigma} = Z^\pi$$

From (59) and (60) we know that:

$$(90)\ p^\pi = c_L + Y(Z^\pi) - \frac{X}{X_G}$$

$$(91)\ p^W = c_L + Y(Z^W)$$

By use of the fact that X/X_G is negative and expressions (87) and (89) we can conclude that:

$$(92)\ p^W < p^\pi$$

This contradicts the assumption made.

To sum up so far: Welfare optimum, as compared to profit optimum, is characterised by higher quality, higher demand and lower price.

18

4 CORRECTIONS FOR NON-OPTIMAL BEHAVIOUR

In order for the welfare maximising authority to deal with the non-optimal behaviours of the profit maximising competitors we assume that two policy parameters are considered. We have already introduced an infrastructure charge or producer fee, f. The other parameter is a possible subsidy related to tickets sold, s. We ignore a possible third parameter that could be a consumption tax, VAT for example. Introduction of such a tax would, however, mean no difference to the principle outcome of the analysis since it would only serve as a fiscal parameter that would affect the magnitudes of the subsidy and the producer fee but not their principle influences.

We should keep in mind that we can in any case add a fixed subsidy or charge, which would not change the optimal solutions with respect to the policy variables p, f and N.

First we examine a subsidy on the consumption side and then two variants of a subsidy on the production side.

4.1 Subsidy on consumption side

4.1.1 Determination of subsidy

We introduce a subsidy, s, per kilometre, which means that the profit maximiser's objective function is:

$$(93)\pi = (p+s)X[p,\ F,\ N] - FC[\ N] - c_L X - fF$$

The first-order condition with respect to price:

$$(94)\frac{\partial\pi}{\partial p} = X + (p+s)X_p - c_L X_p = 0$$

$$(95)\,p^{\pi} = c_L - \frac{X}{X_p} - s$$

Use of (17) yields:

$$(96)\,p^{\pi} = c_L - \frac{X(1-X_G T_X)}{X_G} \equiv c_L - \frac{X}{X_G} - s + XT_X$$

The first-order condition with respect to frequency yields:

$$(97)\frac{\partial\pi}{\partial F} = (p+s)\frac{\partial X}{\partial F} - c_L\frac{\partial X}{\partial F} - C - f = 0$$

$$(98)(p - c_L + s)X_F = C + f$$

19

Combination of (96) and (98) yields:

$$(99)(p - c_L + s - XT_X + \frac{X}{X_G})X_F = C + f - XT_X\frac{X}{F} + XT_F^\tau$$

The first-order condition with respect to number of carriages yields:

$$(100)\frac{\partial \pi}{\partial N} = (p + s)X_N - c_L X_N - FC_N = 0$$

$$(101)(p - c_L + s)X_N = FC_N$$

Use of the first-order condition with respect to price and (29) yields:

$$(102)(p - c_L + s - XT_X - \frac{X}{X_G})X_N = FC_N - XT_X\frac{X}{N}$$

Apparently, with the assumed subsidy s, the expressions for optimal frequency and number of carriages of the profit maximiser according to (99) and (102) are the same as the corresponding expressions without a subsidy, (51) and (55).

It follows that if $s = - X/X_G$, and if $f = I+e$, then all the three first-order conditions with respect to profit maximisation coincides with the first-order conditions with respect to welfare maximisation.

4.1.2 Subsidy level

By use of (23), (52) and (56) the subsidy per passenger kilometre can be expressed as:

$$(103)s = -\frac{X}{X_G} \equiv -XT_X + XT_X\frac{\frac{X}{F}}{X_F} - \frac{XT_F^\tau}{X_F} \equiv$$

$$\equiv \frac{FNC_N}{X}(\frac{1}{\varepsilon_F} - 1) - \frac{1}{\varepsilon_F}FT_F^\tau \equiv \frac{1}{\varepsilon_F}\frac{F(C + I + e)}{X}(1 - \varepsilon_F) - FT_F^\tau$$

The total subsidy on the consumption side, s_X, is then

$$(104)s_X = -FNC_N + \frac{F(C + I + e)}{\varepsilon_F} \equiv \frac{1}{\varepsilon_F}F(C + I + e)(1 - \varepsilon_F) - FXT_F^\tau \equiv$$

$$\equiv \frac{1}{\varepsilon_F}FNC_N(1 - \varepsilon_F) - \frac{1}{\varepsilon_F}FXT_F^\tau$$

The subsidy:

- is positive
- decreases with the elasticity with respect to generalised cost
- decreases with the marginal cost with respect to transport unit size,

20

- increases with the cost per departure and the marginal infrastructure and external cost
- increases with the marginal wait time cost with respect to frequency,
- decreases with the demand elasticity with respect to frequency,

4.1.3 Profit level

The profit level including the subsidy is:

$$(105)\pi = (c_L + XT_X - \frac{FNC_N}{X} + \frac{F(C+I+e)}{X\varepsilon_F})X - c_L X - FC - F(I+e) \equiv$$

$$\equiv F(C+I+e)(\frac{1}{\varepsilon_F} - 1)$$

Apparently, with this subsidy the profit maximising operator will make the same level of profit as under free competition without subsidy intervention, given that f=I+e. Note, however, that a fixed charge may be used in order to eliminate all or most of the profit.

In the next two sections we will instead of a price related policy variable use the infrastructure charge as a policy variable.

4.2 Subsidy on production side – predetermined price and in-vehicle congestion cost function

The profit maximising firm has to set the optimal price that is predetermined by the authority. In this section we regard the case where we assume ride time cost being a function of capacity use (in-vehicle congestion), in line with the analysis made so far. If f< I+e we consider the difference to be a subsidy on the production side. In the next section, 4.3, we will assume a capacity constraint instead of an in-vehicle congestion cost.

4.2.1 Frequency and unit size are optimised

By assuming that the authority forces the profit maximising operator to charge the optimal price and by using the infrastructure charge f as a policy variable the objective function is:

$$(106)\pi = p^W X[p,\ F,\ N] - FC[N] - c_L X - fF$$

The first-order condition with respect to frequency yields:

$$(107) p^W X_F - C - c_L X_F - f = 0$$

$$(108)(p^W - c_L)X_F = C + f = 0$$

(108) can then be rewritten as:

$$(109)(p^W - c_L - XT_X)X_F = C + f - X_F XT_X = 0$$

21

Expression (37) gives the corresponding expression for welfare maximisation.

The infrastructure charge that brings equality between the right hand sides of (37) and (109) and consequently equality between welfare and profit maximisation is then:

$$(110) f = -C + X_F XT_X = I + e + X_F XT_X - XT_X \frac{X}{F} + XT_F^{\tau} = I + e + X_F \frac{X}{X_G}$$

By inserting f according to (110) we get the following profit objective function:

$$(111) \pi = p^W X[p,\ F,\ N] - FC[N] - c_L X - (I + e + X_F \frac{X}{X_G})F$$

Optimal frequency

Differentiation with respect to frequency yields:

$$(112) \pi_F = p^W X_F - C - c_L X_F - (I + e + X_F \frac{X}{X_G}) \equiv (p^W - c_L - XT_X) X_F \equiv C + I + e - XT_X \frac{X}{F} + XT_F^{\tau} = 0$$

$$(113) F^{\pi} = \frac{X^2 T_X}{C + I + e + XT_F^{\tau}}$$

If thus the profit maximiser is forced to charge the welfare optimal price and the infrastructure charge is according to (110), it seems as if welfare optimal frequency is achieved. We have to check, however, that also the transport unit size (the number of carriages) is optimal.

Optimal transport unit size (optimal number of carriages)

$$(114) \pi_N = (p^W - c_L) X_N - FC_N = 0$$

By inserting the welfare optimal price we achieve:

$$(115) (c_L + XT_X - c_L) X_N - FC_N = 0$$

$$(116)\ XT_X X_N = FC_N$$

Apparently this expression does not coincide with the first-order welfare optimum for transport unit size (number of carriages). A forced optimal price would actually be optimal only if X_N=X/N. This possibility is ruled out by our non-questionable assumption that ε_N<1.

4.2.2 Only frequency is optimised

From the section above we note, however, that if transport unit size, or number of carriages, N, were given, and only price and frequency had to be optimised, then a production-related infrastructure charge according to (110) would yield an optimal solution if price is specified at the optimal level.

22

By use of (52) we express the infrastructure charge as follows:

$$(117) f = I + e + XT_X(X_F - \frac{X}{F}) + XT_F^\tau \equiv$$
$$\equiv \varepsilon_F(C + I + e + XT_F^\tau) - C \equiv \varepsilon_F NC_N - C$$

The total production related subsidy, s_f, is then:

$$(118) s_f = F(I + e) + FC - F\varepsilon_F(C + I + e + XT_F^\tau) \equiv$$
$$\equiv F(C + I + e)(1 - \varepsilon_F) - \varepsilon_F XT_F^\tau \equiv FNC_N(1 - \varepsilon_F) - FXT_F^\tau$$

The subsidy level in this case is thus the frequency elasticity multiplied by the consumption subsidy level. The subsidy level is thus lower with the production-related subsidy compared to the consumption-related subsidy, if no fixed fees are introduced.

The profit level including the subsidy is:

$$(119) \pi = (c_L + XT_X)X - c_L X - FC - F(-C + \varepsilon_F(C + I + e + XT_F^\tau)) \equiv$$
$$\equiv F(C + I + e)(1 - \varepsilon_F) - FXT_F^\tau(1 - \varepsilon_F) \equiv FNC_N(1 - \varepsilon_F)$$

If we take the difference between profit and total subsidy we would have the following profit without subsidy:

$$(120) F(C + I + e + XT_F^\tau)(1 - \varepsilon_F) - F(C + I + e)(1 - \varepsilon_F) + \varepsilon_F XT_F^\tau \equiv$$
$$\equiv FXT_F^\tau$$

Without the subsidy this is apparently the same (negative) profit as for welfare maximisation and for the consumption-related subsidy.

4.3 Subsidy on production side – pre-specified price and capacity constraint

Here we follow Larsen et al. [2001], assuming that the authority forces the profit maximising operator to charge the optimal price but employs a constraint for maximum capacity instead of taking ride time cost to be dependent on occupancy rate. The capacity constraint is: $\sigma FN \geq X$. Since Larsen et al. specifies no infrastructure and external costs their definition of subsidy in our terms is –f, while our definition of subsidy is I+e-f.

4.3.1 Derivation of optima

The objective function for welfare maximisation is:

$$(121) L^W = S[G[p, F, N]] + pX[p, F, N] - FC[N] - c_L X - IF - eF + \lambda(\sigma FN - X)$$

The objective function for profit maximisation is:

23

$$(122) L^{\pi} = p^{W} X[p^{W}, F, N] - FC[N] - c_L X - fF + \lambda(\sigma FN - X)$$

We first derive the first order conditions with respect to welfare maximisation under this capacity constraint.

Welfare optimum – price

The first-order condition with respect to price:

$$(123) L_p^W = S_G G_p + X + pX_p - c_L X_p - \lambda X_p = 0$$

Since price now does not affect generalised cost via ride time cost, we have that G_p=1, and

$$(124) L_p^W = -X + X + (p - c_L - \lambda) X_p \equiv (p - c_L - \lambda) X_p = 0$$

So,

$$(125) p^W = c_L + \lambda$$

Compared to the unconstrained case the marginal in-vehicle congestion cost, XT_X, is thus replaced by the Lagrange multiplier λ.

Welfare optimum – frequency

The first-order condition with respect to frequency:

$$(126) L_F^W = -XG_F + pX_F - c_L X_F - C - I - e + \lambda\sigma N - \lambda X_F = 0$$

When no in-vehicle congestion cost is assumed we express XG_F by use of (18) and write the first-order condition as follows, taking into account the optimal price:

$$(127) L_F^W = (p^W - c_L - \lambda) X_F = C + I + e - \lambda\sigma N + X^W T_F^{\tau} = 0$$

Welfare optimum – number of carriages

The first-order condition with respect to number of carriages:

$$(128) L_N^W = pX_N - c_L X_N - FC_N + \lambda\sigma F - \lambda X_N = 0$$

$$(129) (p^W - c_L - \lambda) X_N = FC_N - \lambda\sigma F = 0$$

$$(130) \lambda = \frac{C_N}{\sigma}$$

We now derive the first-order conditions of the profit maximising operator with respect to frequency and number of carriages, given the assumption that the optimal price is used.

24

Profit maximisation – frequency

The first-order condition with respect to frequency:

$$(131) L_F^{\pi} = p^W X_F - c_L X_F - C - f + \lambda\sigma N - \lambda X_F = 0$$

$$(132) (p^W - c_L - \lambda) X_F = C + f - \lambda\sigma N = 0$$

Profit maximisation – number of carriages

The first-order condition with respect to number of carriages:

$$(133) L_N^{\pi} = p^W X_N - c_L X_N - FC_N + \lambda\sigma F - \lambda X_N = 0$$

$$(134) (p^W - c_L - \lambda) X_N = FC_N - \lambda\sigma F = 0$$

The first-order condition with respect to number of carriages thus looks the same for welfare and profit maximisation respectively.

In order to achieve social optimum for profit maximisation equality between (127) and (132) is required. We can also use that for both optima to be fulfilled both (130) and (134) have to be valid. The infrastructure charge in order to obtain social optimum can then be expressed in two ways:

$$(135) f = I + e + X^W T_F^{\tau} \equiv -C + N^W C_N \left[N^W\right]$$

Given that the profit maximiser is forced to charge the optimal price p^W, the capacity constraint means that $\sigma FN = X[p^W, F]$, which implies that $N = N[F]$. Implicit differentiation with respect to F yields:

$$(136) N_F = \frac{X_F - \sigma N}{\sigma F}$$

Let us now derive the profit optimum, by use of the expression for welfare optimum price according to (125) and by inserting the capacity constraint.

$$(137) \pi = p^W X - C[N[F]]F - c_L X - fF \equiv$$
$$\equiv (c_L + \lambda)\sigma FN[F] - C[N[F]]F - c_L \sigma FN[F] - fF$$

By use of (130) we get:

$$(138) \pi = C_N[N^W] FN[F] - C[N[F]]F - fF$$

4.3.2 Unit cost function is linear

25

Assume first that C[N] is linear in N, i.e.,

$$(139) C[N] = C_0 + C_1[N] = C_0 + C_1 N$$

Then, if there were an optimum for profit maximisation under these circumstances, (135) gives that the optimal infrastructure charge would be:

$$(140) f = -C + N^W C_N \left[N^W\right] \equiv C_1 N^W - C_0 - C_1 N^W = -C_0$$

Larsen et al. [2001] interpret this result so that C0 is the optimal subsidy with a linear cost function, and they say that for “a bus operation C0 will come close to the wage cost per revenue kilometre for the driver.” Note then that since we specify infrastructure and external costs as well, our corresponding expression for subsidy is I+e+ C0. This difference is, however, of no importance since one could easily add I+e.

Let us instead check whether this result of Larsen et al. [2001] really represents an optimum, by examining the first-order condition of the profit function (138) with respect to frequency. If C[N] were linear, (138) would in fact be:

$$(141) \pi = (-C_0 - f)F$$

The first-order condition with respect to frequency is then:

$$(142) \pi_F = (-C_0 - f) = 0$$

Consequently the profit function would be:

$$(143) \pi = 0 \times F \equiv 0$$

This function certainly has no optimum. Consequently there is no solution if the cost function C[N] is linear in N. Given that the cost function is linear we investigate in the appendix an alternative solution where the charge presented to the operator is not a fixed amount but a function of frequency. This solution, however, seems to be equivalent to directly specifying frequency in the first place, which is then simpler than the circumvention by use of a function.

4.3.3 Unit cost function is non-linear

Larsen et al. [2001] claims that the optimal infrastructure charge “is approximately the cost of a marginal kilometre operated with a unit of minimum capacity.” This means they claim that the charge is approximately equal to C_0 even if the cost function is not linear. In this section we examine the possibility of such an optimum and its corresponding conditions for a non-linear cost function.

Let us thus examine the possibility of finding an optimum if C[N] is not linear. We differentiate the objective function (138) with respect to F.

$$(144) \pi_F = (C_N[N^W]N[F] - C[N[F]] - f) + (C_N[N^W] - C_N[N])N_F F = 0$$

26

By assumption ε_F<1. Then, by use of (136) and the capacity constraint, we have that:

$$(145) N_F F = \frac{X_F - \sigma N}{\sigma} \equiv (\frac{X_F F}{X} - 1)N \equiv (\varepsilon_F - 1)N < 0$$

Since we are searching for a maximum of the profit function we now have to examine the second-order condition of the profit function with respect to frequency. For a maximum it is required that:

$$(146) 0 > \frac{\partial^2 \pi}{\partial F^2} = (C_N[N^W] - (C_N[N])N_F + \frac{\partial}{\partial F}(C_N[N^W] - C_N[N])N_F F$$

If we now want to obtain an optimum by use of an infrastructure charge, f, so that $N=N^W$ is optimal for the profit maximiser, then the first parenthesis in (146) is equal to zero. The second-order condition is thus that:

$$(147) Z[N] = (C_N[N^W] - C_N[N])N_F F$$

is decreasing in F in the neighbourhood of $F=F^W$.

Here $Z[N^W]$ =0, and the factor $N_F F$< 0. This means that the conditions are:

$C_N[N^W]$ - $C_N[N]$ > 0 if F > F^W and $C_N[N^W]$ - $C_N[N]$ < 0 if F < F^W

In other words, since N[F] is decreasing in F:

$C_N[N^W]$ - $C_N[N]$ > 0 if N < N^W and $C_N[N^W]$ - $C_N[N]$ < 0 if N > N^W

The conclusion is that the condition for optimum is that C[N], and consequently $C_1[N]$, is a convex function of N. This convexity criterion for optimum is not derived or mentioned in Larsen et al. [2001]. What about their claim that the subsidy is approximately equal to C_0? Well, this would hold only if the cost function is slightly convex, i.e., almost linear. As soon as the optimal size is where the convexity is more emphasised the subsidy is smaller and it might be negative. If the cost function is concave for small unit sizes and convex for larger sizes, it might occur that the optimal subsidy is higher than C_0. The possibility in practice to find an optimum with pre-determined price and capacity constraint, as well as to find the magnitude and size of the subsidy in that case, is thus strongly dependent on real-world empirical data on costs for various transport unit sizes.

Under the specific circumstances that the cost function of transport unit size is convex, (135) repeated says that the optimal infrastructure charge is:

$$(148) f = I + e + XT_F^\tau \equiv NC_N - C$$

The total subsidy is then:

$$(149) s_f = F(I + e) - F(NC_N - C) \equiv F(I + e) - (F(I + e) + FXT_F^\tau) = -FXT_F^\tau$$

The profit level is then:

27

$$(150)\ \pi = (c_L + XT_X)X - c_L X - FC - F(I + e + XT_F^{\tau}) = 0$$

The profit level is thus zero at optimum and, since optimum corresponds to maximum profit, it is below zero outside optimum.

5 SUMMARY AND INTERPRETATIONS

We will here briefly summarise the results and discuss possible practical implications of the different subsidy possibilities.

In section 5.1 we remind of the unconstrained welfare and profit maximisation optima. In section 5.2 we summarise and compare the various corrective subsidy policies. In section 5.3 we illustrate the various forms of subsidy related to costs with respect to transport unit size. In section 5.4 we summarise the results of computer simulations of reduced consumer prices. In section 5.5 finally we make some comments on possible implications for organisation of the railway sector.

5.1 Optima for welfare and profit maximisation

Price is lower and frequency and transport unit size is larger under welfare maximisation compared to profit maximisation (although theoretically it might happen that one of the variables is equal). Since frequency and transport unit size represent quality we can conclude that welfare maximisation generates both lower price and higher quality than profit maximisation.

5.2 Comparison between corrective subsidy policies

The following corrective measures that would yield an optimal solution have been analysed:

Policy a. A consumption related subsidy without any constraint.

Policy b. A production related subsidy with ride time cost being a function of capacity use, *if* optimal price *and* optimal transport unit size are stipulated.

Policy c. A production related subsidy with a capacity constraint, *if* optimal price is stipulated *and if* the cost function with respect to transport unit size is convex.

The table below summarises subsidies, infrastructure charges and profit levels for welfare maximisation, profit maximisation and three alternative corrective subsidy policies, a), b) and c), in order to achieve welfare optimum.

Note that the profit levels displayed assumes no concern for a possible lump-sum charge. For each of the policies a), b) and c) the authority can infer a positive or negative lump-sum transfer in order to alter the balance between the finances of the operator and the authority.

28

Table 5.2.1 Summary of subsidies, infrastructure charges and profit levels

Policy alternatives	**Subsidy per passenger km s**	**Infrastructure charge/departure f**	**Total subsidy**	**Profit level π**
Welfare maximisation	$-FT_F^\tau$	$I+e$	$-FXT_F^\tau$	0
Profit maximisation	0	$I+e$	0	$F(C+I+e)(\frac{1}{\varepsilon_F}-1)$
a) Correction on consumption side	$\frac{1}{\varepsilon_F}\frac{F(C+I+e)}{X}(1-\varepsilon_F)-$ $-FT_F^\tau$	$I+e$	$\frac{1}{\varepsilon_F}F(C+I+e)(1-\varepsilon_F)-$ $-FXT_F^\tau \equiv$ $\equiv \frac{1}{\varepsilon_F}FNC_N(1-\varepsilon_F)-$ $-\frac{1}{\varepsilon_F}FXT_F^\tau$	$F(C+I+e)(\frac{1}{\varepsilon_F}-1)$
b) Correction on production side Predetermined price and unit size Ride time cost is a function of capacity use	0	$\varepsilon_F(C+I+e+XT_F^\tau)-C\equiv$ $\equiv \varepsilon_F NC_N-C$	$F(C+I+e)(1-\varepsilon_F)-$ $-\varepsilon_F XT_F^\tau\equiv$ $\equiv FNC_N(1-\varepsilon_F)-FXT_F^\tau$	$F(C+I+e)(1-\varepsilon_F)-$ $-FXT_F^\tau(1-\varepsilon_F)\equiv$ $\equiv FNC_N(1-\varepsilon_F)$
c) Correction on production side Predetermined price Capacity constraint Unit cost function is convex	0	NC_N-C	$-FXT_F^\tau$	0

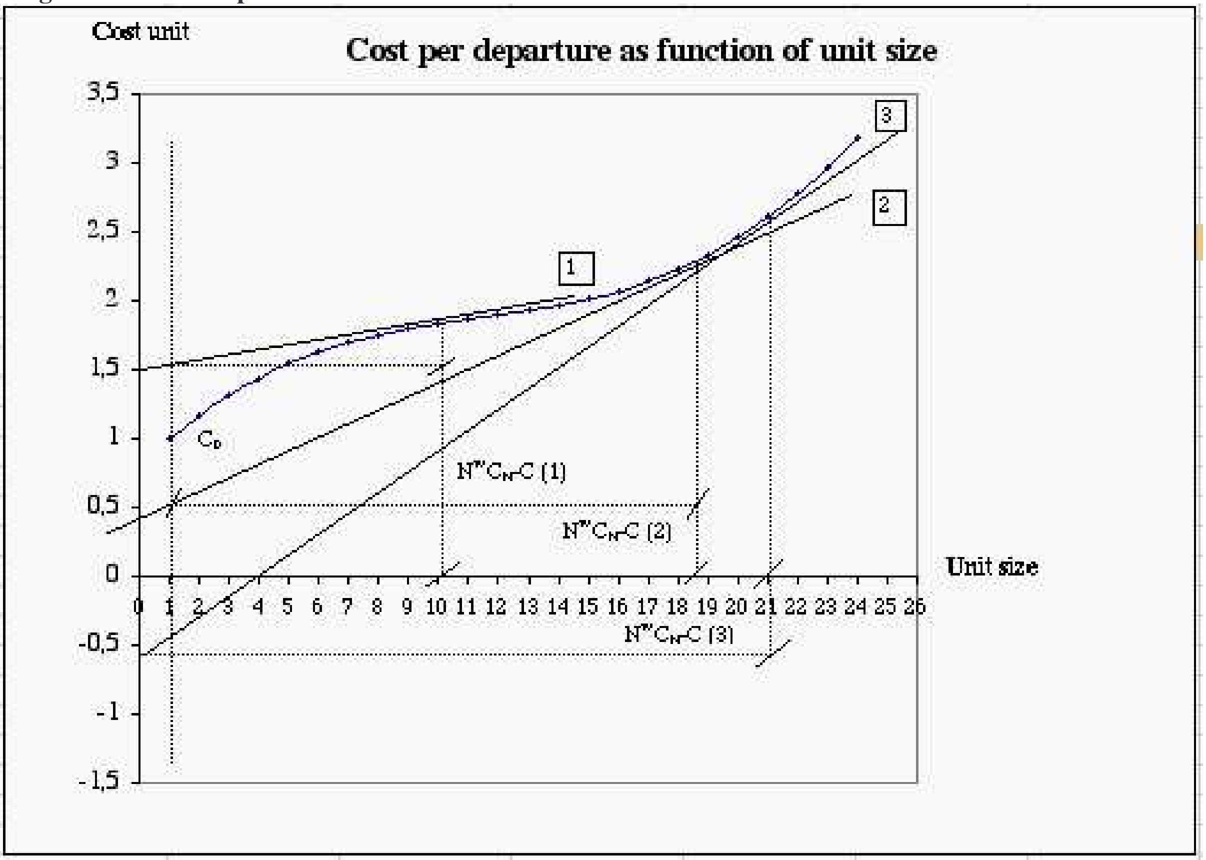
5.3 Subsidies and cost functions

We will here interpret the various forms of corrective subsidies in cost terms, in principle and by use of empirical data on cost functions.

The diagram below illustrates the principle look of this kind of cost functions per departure for bus, train or aircraft as a function of transport unit size. At a certain size the cost function shifts from concave to convex. Unit size 1 might for buses correspond to 4 seats (car) and for train to 1 carriage or some 60-90 seats dependent on carriage type.

29

Diagram 5.3.1 Principle view of unit size cost function



The three straight lines in the diagram correspond to different assumptions for the slope of the cost function, i.e., the marginal cost with respect to unit size, C_N. For these three assumptions on optimal transport unit size we illustrate $f= N^W C_N – C$. This corresponds to the infrastructure charge for the policy c) correction on production side, with predetermined price, capacity constraint and convex unit size cost. The requirement convex cost function means that this infrastructure charge is valid only for assumptions 2 and 3 concerning optimal transport unit size.

Where f is above the zero cost level the infrastructure charge is negative. However, the fact that the infrastructure charge is negative does not necessarily mean that the subsidy is negative since the subsidy per departure is $I+e-f = I+e - (N^W C_N – C)$. Where f is below the zero cost level the infrastructure charge is positive, which means that the subsidy might be negative.

If we require that all the three policy variables, price, frequency and transport unit sizes are to be optimised we notice that:

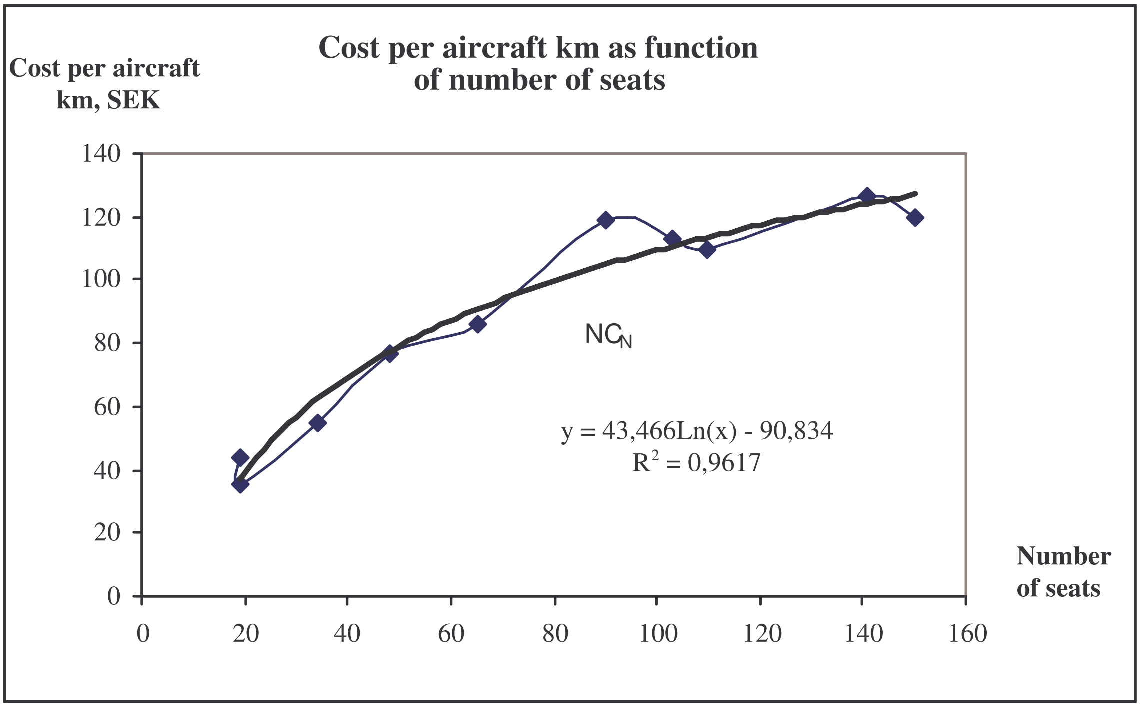
- If optimal N is where the cost grows regressively with N, only a consumption related subsidy can be optimal (see line 1).
- If optimal N is where the cost grows progressively but slowly with N, the optimal infrastructure charge could be negative (see line 2).
- If optimal N is where the cost grows progressively but steep with N, the optimal infrastructure charge could be positive (see line 3).
- Only at one point of optimal transport unit size can the optimal infrastructure charge be equal to the fixed cost C_0.

30

A crucial issue for real-world application of production-related subsidies is thus whether the cost function with respect to unit size is concave or convex. For air transport the following diagram follows from Karyd (2001), by application of logarithmic regression.

31

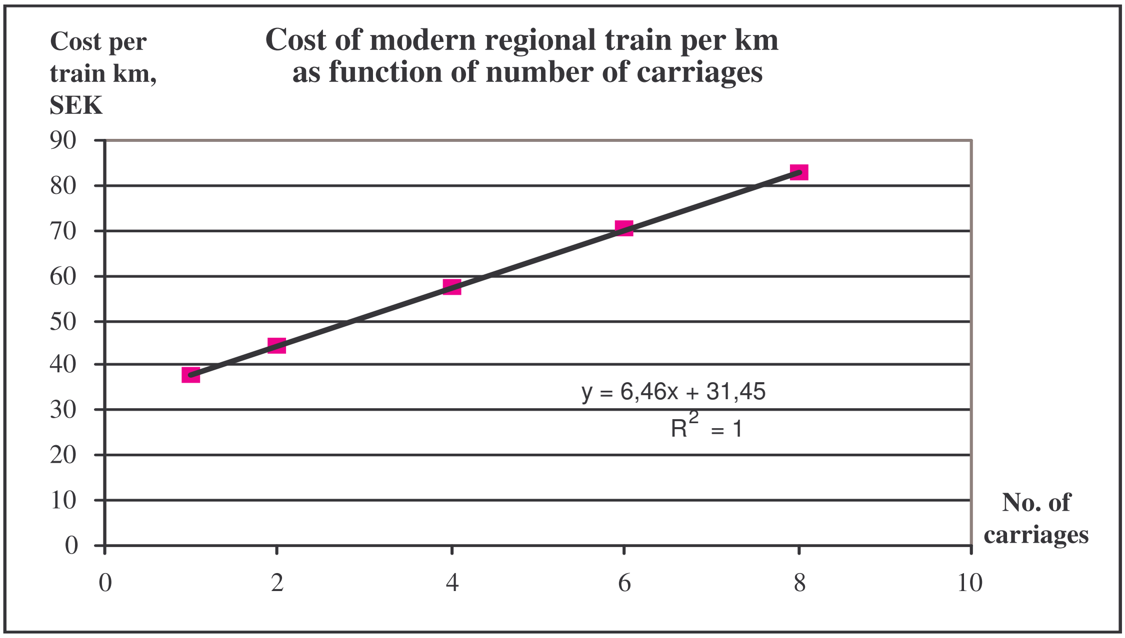
Diagram 5.3.2 Swedish Aircraft cost as function of unit size



The dots in the diagram correspond to actual observed costs according to Karyd (2001). The line is the logarithmic regression line. Apparently the cost function concerning aircraft costs with respect to unit size is concave for the size range used in Sweden.

For rail transport the empirical evidence is based on Kottenhoff (1999). See the diagram below.

Diagram 5.3.3 Example of train cost as function of unit size

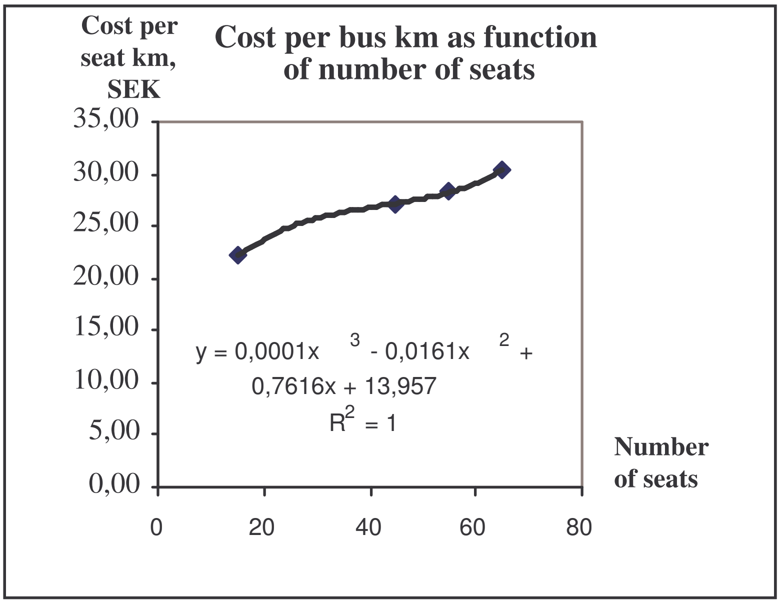


Apparently the cost function according to this source is linear.

32

For bus costs the information stems from a Swedish bus operator, Buslink. The indications given on various costs components resulted in the following third degree polynomial cost function.

Diagram 5.3.4 Indication of bus cost as function of unit size



According to this source the cost function shifts from concavity to convexity. The convex part occurs when moving from bogie bus to articulated bus.

Our conclusion from empirical data is that economies of scale in transport unit size with respect to the sum of capital and operation cost, $C_1[N]$ is concave for a wide range of demand. For the bus mode there may be a convex part when transferring to the biggest bus types. For the train mode convexity would appear when platform size is fully used or if shunting costs may increase more than proportionally with number of carriages

The other crucial issue is under what circumstances one can expect that optimum appears at the convex part of the cost function. One must then keep in mind that this optimum also depends on optimal frequency. Only if demand is high and frequency is valued low it could occur that the marginal cost with respect to transport unit size commences to increase with the size. This situation might correspond to certain peak-hour traffic only.

5.4 Simulations of consumption side subsidies

In order to test subsidies of the railway, simulations were made for the whole Swedish public transport and road network, using the network model VIPS. For large and small parts of the railway services price reduction were assumed. Small parts comprise lines that are not commercially viable and have low demand today.

In general it was found that lower prices would be socially preferable even if not commercially justified. The price cuts can thus only be introduced by use of subsidies. The simulations are thus backing up the theoretical findings in this paper. The table below summarises the welfare economic outcome of one of the simulations. This concerns price reductions of about 20 financially poor railway lines. According to the simulations the

33

railway would lose some profit (actually increase the loss) with price reductions. We assume that this loss is covered by state subsidies. Also airlines and coach operators would lose some profit. This loss is not assumed being compensated. On the other hand one could expect that these services would be modified in order to avoid losses. We have not conducted an iteration process in order to care for second- or third order response effects. On the other hand such a higher-order effect is virtually a non-issue with the low demand levels of these rail services.

Table 5.4.1 Cost-benefit result

	MSEK/year
Consumer surplus	58
whereof fare	*67*
whereof time	*-9*
Producer surplus, operators	-20
Tax correction	-1
Net State surplus	-31
Excess burden	-9
External effects	11
Net Social Benefit	**6**

The simulations indicate that price reductions financed by state grants linked to price would be socially beneficial, thus in accordance with expectations from the theoretical analysis.

5.5 Organisational aspects and possible real world applications

Let us first recapitulate the findings on the three corrective subsidy measures:

- Policy a), consumption side subsidy implies the largest freedom for the operators.

- Policy b), Production side subsidy with predetermined price and predetermined unit size implies freedom only for choice of frequency.

- Policy c), Production side subsidy with predetermined price and capacity constraint implies freedom only for choice of frequency and is valid only if the cost function is convex in transport unit size.

First of all, since a profit maximising operator would choose a price that is higher and a quality that is lower than a welfare maximising authority would do, one might question that rail services are in the hands of profit maximising bodies. The best solution might be an integrated state owned monopoly that plans, operates as well as subsidises rail services.

However, such a solution would bias the competition with air and coach services. In fact, maybe not only rail services, but also all public transport services, should be vertically integrated, in the hands of a state controlled “public transport monopoly”.

Also, a huge monopoly, state or privately controlled, might be characterised by inefficiency on the production side.

In case production inefficiency is considered a real problem or in case a monopoly is not politically accepted, the solution seems to be a supervising body and competing operators. This solution has is in fact some resemblance to the transport policy in Sweden. The national

34

rail track authority (Banverket) and the national Civil Aviation Administration (Luftfartsverket) have supervising tasks. At least the national rail track authority is also meant to make its decisions on infrastructure investments and maintenance on social economic grounds. In addition Sweden has a national procurement agency (Rikstrafiken), which arranges competitive tendering for certain rail- air- and bus services that are not commercially viable. For local and regional services in Sweden public authority planning of the network and tendering for the operation is the typical solution.

Assume now that a transport authority or a government agency considers that achievement of welfare optimum is beneficial and that the excess burden of government financial support is smaller than the welfare loss of commercial considerations only. Intervention could be through a general subsidy or through procurement.

If a general subsidy is applied we have found that a consumer related subsidy gives the largest freedom to the operators. A production related subsidy requires more parameters to be pre-specified and leaves less freedom to the operators. A production related subsidy linked with a capacity also leaves less freedom and the optimum is possible only if the transport unit size function is convex.

With intervention in terms of procurement one could in principle achieve the lowest possible subsidy level irrespective of subsidy method. For the three corrective subsidies, a), b) and c) the procurement process could be as follows.

a) Define the consumer-related subsidy, and announce the level to bidders in an auction. The bidders will deliver positive bids, i.e., pay for the commitment. With a (near) perfect competitive market the bid would (almost) equal the profit they would make under monopoly, which equals the profit they would make if the subsidy were given without any compensatory bid. The net profit would be (almost) zero.

b) Define the production-related subsidy with pre-specified price and transport unit size, and announce the level to bidders in an auction. The bidders will deliver positive bids, i.e., pay for the commitment, but the bid is lower than under a). With a (near) perfect competitive market the bid would give a net profit that is (almost) zero.

c) Define the production-related subsidy with pre-specified price, transport unit size and capacity constraint, and announce the level to bidders in an auction. With a (near) perfect competitive market the bidders will deliver (slightly) negative bids, i.e., demand a small payment for the commitment. The net profit would be (almost) zero.

We conclude that such procurement processes under “sufficient” competitive circumstances would all yield optimum at more or less the same minimum subsidy level. The decisive factor would then be what kind of intervention is most feasible in practice. All methods need information about demand and elasticities at optimum. However, the consumer-related subsidy needs the smallest number of pre-specified parameters and would at the same time provide the largest freedom to the operator.

The conclusion is that if procurement is chosen, then a consumer-related subsidy is preferable.

In any way the authority would need information from the operator about the current situation with respect to demand for each passenger segment, i.e., each origin-destination stops and

35

passenger category and to price. Such information would be a prerequisite for subsidies and procurement and should probably meet no barrier since the information is provided in exchange for possibility to be awarded a contract. In addition the authority needs to calculate elasticities and estimate how different passenger groups value ride time under different capacity use levels and how they value service frequency. However, these latter parameters are nothing new but the standard requirements if cost-benefit analyses are undertaken in the transport sector.

For the Swedish situation one might argue that since Banverket (the Rail Track Authority) is supposed to have the socio-economic perspective of the rail industry through infrastructure investments and maintenance, one should also have a word on pricing and subsidy matters. The motive is that subsidies affect price and quality and demand, which in turn affect the net benefit or loss of infrastructure measures.

The information requirements mean that one cannot immediately in practice find the correct subsidy levels. In turn this means uncertainty and risk-taking by the operators. One would therefore expect that each bid includes a risk premium. This is thus a disadvantage of procurement compared to total authority control and authority operation. Again the trade-off is between the possible production inefficiency under total authority control and the risk premium. On the other hand, the risk premium is a real cost only to the extent it affects excess burden of tax financing.

In addition one should ideally also take into account that different modes pay for external effects at various degrees. If payment for external costs according to real marginal social costs is something that for political reasons cannot be changed, one should investigate second-best pricing and then the subsidy policy would not be the same for all modes. This second-best issue has not been addressed in this paper but is worth further research that could be based on the model applied here.

Also for the Swedish situation it seems as if procurement in the way described here could be something to consider for the Swedish procurement agency Rikstrafiken, as an alternative to the current bidding for a lump-sum subsidy given a specified service level.

The incentive schemes analysed here might also be considered without a procurement procedure. Instead of an auction the authority determines a fixed charge, or a subsidy along with the incentive, and permits any operator to enter the market under these conditions. What has not been analysed here, however, is how the incentive schemes might have to be modified when several operators are competing. This is also an issue for further research.

Appendix: Pre-specified charge arrived at by use of a function

We will here make an alternative interpretation of the analysis by Larsen et al., where the unit cost function was assumed to be linear. We have proved that an optimum cannot be achieved by using a pre-specified charge equal to the fixed carriage cost c_0, at least if the carriage cost function is linear. One might, however, interpret the finding by Larsen et al. in the following manner. a) The authority has calculated the welfare optimal frequency and decided that the subsidy should be c_0. b) The charge is not presented to the operator as a specific value but instead as a function of frequency that would make the operator choose the desired frequency

36

A number of functions could be used but let us assume that this charge function has the following form:

$$(A1) a + bF + \frac{d}{F}$$

The requirement of the authority is that in optimum the charge shall be:

$$(A2) f = -c_0 = a + bF^W + \frac{d}{F^W}$$

Since the capacity constraint implies that either frequency or transport unit size is the only policy variable, we can eliminate the capacity constraint by inserting N = X/Fσ, and thus use only F as the policy variable. The profit function is then:

$$(A3) \pi = p^W X[p^W,\ F] - FC[\frac{X}{F\sigma}] - c_L X - (a + bF + \frac{d}{F})F$$

The first order condition gives:

$$(A4) \pi_F = p^W X_F - c_L X_F - C - FC_N \frac{FX_F - X}{F^2\sigma} - (a + bF + \frac{d}{F}) + F(-b + \frac{d}{F^2}) = 0$$

$$(A5) \pi_F = (p^W - c_L - \frac{C_N}{\sigma})X_F = C - \frac{XC_N}{F\sigma} + a + 2bF$$

Since if we have an optimum the left-hand side of (158) is zero we can rewrite the expression as:

$$(A6) C - NC_N + a + 2bF = 0$$

When the carriage cost function, C[N], is assumed to be linear, (159) is:

$$(A7) c_0 + a + 2bF = 0$$

If we combine (155) and (160) in optimum we achieve the following equation system:

$$(A8) \begin{cases} c_0 + a + bF^W + \dfrac{d}{F^W} = 0 \\ c_0 + a + 2bF^W = 0 \end{cases}$$

For any arbitrary chosen b>0 we get the following solution:

$$(A9) d = b(F^W)^2$$

$$(A10) a = -c_0 - 2bF^W$$

It follows that the charge function is:

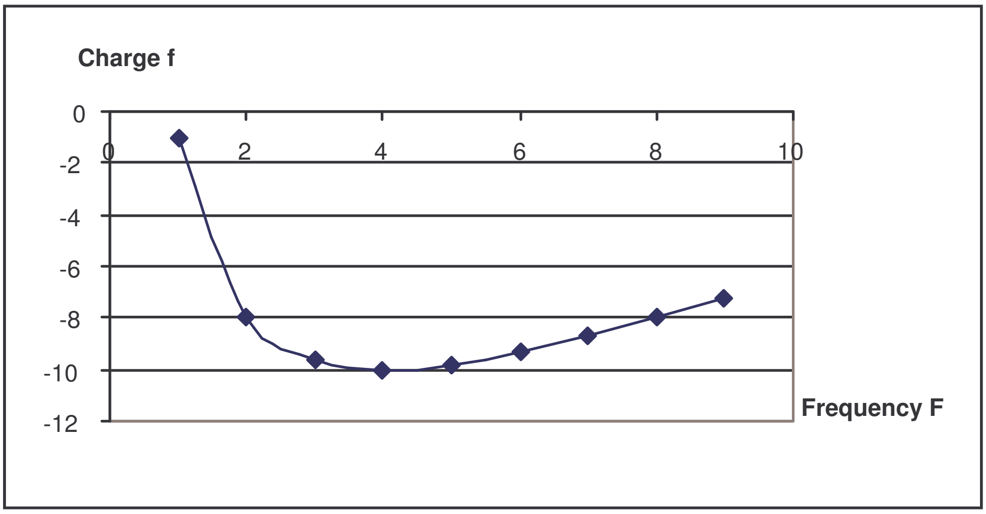
37

$$(A11) f = -c_0 - 2bF^W + bF^W + \frac{b(F^W)^2}{F}$$

It is thus possible to arrive at a charge equal to $-c_0$. Note, however that this procedure requires that the charge is chosen and pre-determined by the authority and is not a result of welfare optimisation. One could equally well have chosen $-2c_0$ or another value.

Evidently, with this kind of procedure the authority could choose among a number of charges and among a number of functions that fulfil the requirements that profit be non-negative and frequency be optimal.

In the diagram below we assume that $c_0 = 10$, $F^W = 4$ and $b = 1$. Apparently the minimum charge at optimum is -10, which means that the charge is a subsidy at level 10. At this optimum the profit level is zero.



Note finally, that use of a charge function may be difficult for an operator to comprehend correctly, and it would be much easier to pre-determine also the frequency along with the pre-determined price level.

References

Jansson, J. O. (1984): Transport System Optimization and Pricing. *John Wiley&Sons*, New York

Jansson, K. (1991): Efficient Pricing and Quality in Public Transport. Department of Economics, Stockholm University

Karyd, A. (2001): Flygplanskostnader i svenskt inrikesflyg (Air craft costs in Swedish domestic market)

Kottenhoff, K. (1999): Evaluation of Passenger Train Concepts – Methods and Results of Measuring Travellers' Preferences in Relation to Costs, Royal Institute of Technology, Department of Infrastructure and Planning

38

Larsen, O., Norheim, B, Johansen, K. W. (2001): Towards achievement of both allocative- and X-efficiency in public transport, *Journal of Transport Economics and Policy*, No.2.

Mohring, H. (1972): "Optimization and Scale Economics in Urban Bus Transportation". *The American Economic Review*, September.

Nash, C. A.: (1978): "Management Objectives, Fares and Service Levels in Bus Transport", *Journal of Transport Economics and Policy*, vol XII.

Panzar, J. C. (1979): "Equilibrium and Welfare in Unregulated Airline Markets". *The American Economic Review*, May.

Turvey, R. and H. Mohring: (1975): "Optimal Bus Fares". *Journal of Transport Economics and Policy*, vol IX.