PUBLIC TRANSIT SUBSIDY: FROM THE ECONOMICS OF WELFARE TO THE THEORY OF INCENTIVES

Daniele Fabbri

Department of Economics, University of Bologna,

Piazza Scaravilli 2, 40126-Bologna, Italy.

tel.: 0039-51-258126

fax: 0039-51-221968

June 1995

ABSTRACT:

Public Transit is publicly managed almost all over Europe. Public intervention in this sector is due to market failures: economies of scale and misperceptions of social and private costs may cause an insufficient supply of transit services. These arguments have been thouroughly analyzed within the *standard welfarist approach* to the theory and practice of subsidization. Ramsey rules and cost-benefit analysis emerged as useful devices for the definition of subsidy allocation. However remedies to market failures should be traded-off against government failures. Lack of incentives, X-inefficiency, regulatory capture, bureaucracy power are common facts in the internal organization of public administration. If a public sector utility, for some reasons, can not be completely privatized, it may be "almost" privatized by means of quasi-market mechanisms. Auctioning, yardistick competition, incentive schemes, auditing, regulation through competition are the keywords for the renewed public involvment in public transit. All these mechanisms can be properly studied within the theoretical format of *the incentives theory*. This approach helps us to understand past experiences of transit firms incentives scheme and to appreciate the relevance of the empirical analysis of performance indicators. An important question that emerges is the definition of operational incentive contracts. In this paper we will discuss this problem by referring to past and recent experiments with performance based subsidization programs.

J.E.L.: H71, R38.

CONTENTS

INTRODUCTION
1. THE STANDARD WELFARIST APPROACH
1.1. Cost-benefit analysis of public transit subsidy
1.2. Transit subisidization as a pricing problem
1.3. Some criticism of the standard welfarist approach
2. THE ORGANIZATIONAL CONCERN
2.1. Optimal commercial criteria for public transit firms
2.2. Positive analysis of transport firms
2.3. Why does organization matter?
3. THE INCENTIVES THEORY APPROACH
3.1. Incentives theory of regulation: the Laffont and Tirole framework
3.2. A model of transit firm regulation
4. INCENTIVE PROGRAMS IN PRACTICE
4.1. General rules for incentive contracts in the public transit sector
4.1. Some experiments with performace-based transit subsidization
4.3. Open issues
5. CONCLUSIONS
APPENDIX
REFERENCES

INTRODUCTION

A subsidy is the difference between production costs of a good and revenues from sales to final users. Its aim is to restore profitability in the production of a good, in order to make it available in quantities and qualities otherwise not provided by the normal functioning of the market. Therefore subsidization is deliberately performed by an agent in order to let some activities exist and grow.

Good reasons for subsidizing public transit have been repeatedly suggested by economists. Standard arguments are: economies of scale, second-best problems, redistribution in kind, option value, imperfect information. For all these reasons it might be socially useful to sustain the production of public transit. But to what extent should we provide this support? Standard welfarist approach gave the answers. These answers will be analyzed in section 1.

Among the several limitations of the standard welfarist approach we find that what makes it an "outdated" framework is its complete disregard of organizational concerns. In section 2 we show how these issues have been normatively treated by the former advocates of the standard welfarist approach and how relevant they are in terms of positive explanations of public firms' behaviour.

Once we explicitly introduce organizational issues we should find new theoretical paradigms allowing another answer to be answered: how to subsidize a public transit firm in order to reach some specific goals and given that it has its own rationality? The theory of incentives provides such a paradigm. In section 3 we provide a model of regulation that fits very well into the typical situations we face. Then, in the last section, it will be quite easy to analyze past and recent experiences of performance-based subsidization.

1. THE STANDARD WELFARIST APPROACH

In the welfarist approach the optimal level of subsidy is defined through a welfare calculus of costs and benefits. Its comprehensiveness is limited only by the extent that direct and indirect effects, which pass through final users of the transport system, can be measured. The literature reports two ways of performing this welfare calculus: cost-benefit analysis and approaches based on optimal pricing of public utilities.

1.1. Cost-benefit analysis of public transit subsidy

Cost-benefit analysis allows us to define subsidization rules on the basis of a welfare comparison between alternative uses of subsidies. The alternative uses are given by different combinations of adjustments in tariff and quality of transit service with respect to the initial situation. Therefore cost-benefit analysis identifies the adjustments in tariff and quality of service that maximizes social welfare given a budget constraint on the total amount of subsidy. The main difficulty in these calculations is the presence of a quality index both in the demand and cost functions[1].

Let *D(b,q)* and *C(D(b,q,),q)* be respectively the demand and the short-run cost function of transit, where *b* is the tariff and *q* is the quality index[2]. For any initial situation (b_0,q_0), define the consumer surplus *S*, and the producer surplus *P*, by:

$$S = \int_{b_0}^{\infty} D(v, q_0)\,dv \tag{1}$$

$$P = b_0 D(b_0, q_0) - C\big(D(b_0, q_0), q_0\big) \tag{2}$$

Given this initial situation (b_0,q_0), it is possible to calculate the net benefit for each amount of money spent alternatively to lower the tariff or to improve the quality of service. In the first case, the net benefit per unit of subsidy will be given by the equation:

$$\frac{dB_b}{dT_b} = \frac{\dfrac{\partial W}{\partial b}}{\dfrac{\partial P}{\partial b}} = \frac{-(b_0 - C_D)\varepsilon_b}{b_0 + (b_0 - C_D)\varepsilon_b} \tag{3}$$

where dB_b is the net benefit due to a tariff reduction, measured by the variation in total surplus $W=S+P$, dT_b is the required increase in subsidy, given by the variation of producer surplus and $_b$ represents tariff elasticity of transit demand. Equation (3) is quite easy to calculate since it is made of known ingredients: tariff, tariff elasticity of demand and marginal cost of service[3].

This is not so in the case of subsidization of an improved quality of service. In this case we have:

---

[1]These circumstances justify a "transportation" treatment of public transit subsidy.

[2]*q* might be either a scalar or a vector.

[3]If the tariff is initially set at marginal cost there is no space for increasing net social surplus. Otherwise if marginal cost is zero the increase in net social benefit will depend only on tariff elasticity of demand. Subsidizing a tariff reduction will be more welfare improving the more initial tariff exceed marginal cost and demand is elastic to the tariff.

$$\frac{dB_q}{dT_q} = \frac{\dfrac{\partial W}{\partial q}}{\dfrac{\partial P}{\partial q}} = \frac{\int_{b_0}^{\infty} D_q(v,q_0)dv + (b_0 - C_D)\varepsilon_q \dfrac{D}{q_0} - C_q}{C_q - (b_0 - C_D)\varepsilon_q \dfrac{D}{q_0}} \tag{4}$$

The numerator of equation (4) contains an integral term. It measures the variation of consumer surplus due to an increase in quality of service. No small experiment with price and quality will generate the information required to evaluate this integral term. The reason for this is that quality changes affect the welfare of the entire set of inframarginal consumers. The data generated by local changes in the parameters will not yield such estimates. Therefore in order to calculate equation (4) it is necessary to build a model that replicates the reactions of the entire set of inframarginal consumers[4]. Moreover we must rescale these effects in terms of welfare measures.

At this point the problem becomes transport-specific. In the generalized travel cost approach the problem is the building of behavioural models explaining the impact of quality of service on walking, waiting, boarding and travel time. The impact on this time requirements of the whole set of inframarginal consumers might then be translated into monetary terms by means of value of time coefficients[5].

Once made operational, equations (3) and (4) provide a sound basis for allocating public transit subsidies to different towns in the same region[6]. An improved version of this simple model has been made operational in England by Stephen Glaister. METS (Model for Evaluating Transport Subsidies)[7], was able to calculate equations (3) and (4) by solving quite complicated local mobility simulation models. It considered a multiplicity of transport modes: car, bus, rail and metro. Consumers choose on the basis of the generalized travel cost of each alternatives. Generalized travel costs are calculated through speed-flow relationships evaluated at equilibrium points. In this way METS allows to estimate modal impacts produced by pricing and quality policies by measuring their effect on congestion reduction and other second-best concerns.

[4]Spence (1975) realizes that this is the main problem faced by the regulator of a monopolist "with quality". The regulator should provide the *non-market* informations needed in order to evaluate total willingness to pay for quality improvements.

[5]This is the approach followed by Dodgson (1987). He solves equation (5) in terms of a relationship between vehicles-km and waiting-time. See also Tisato (1992) for an improved user cost model.

[6]See Evans (1985) for the definition of different equalization schemes in transit transfers. Dodgson (1987), by referring to a sample of Australian cities, suggested that an increase in vehicle-km would not be welfare improving in most of the towns.

[7]For a detailed description of METS see Glaister (1987).

Even though METS is probably the most ambitious operational model for allocating subsidies it nevertheless suffers from several internal limitations. Gwilliam (1987) provides the following ordered list of criticisms[8]:

- With respect to the objectives encompassed it has been pointed out that the use of the cost-benefit framework does not provide for any environmental or distributional objectives.
- The model is not disaggregated by time of day and therefore it does not properly assess welfare effects due to peak shifting policies.
- The high level of spatial aggregation of the model and the lack of explicit geographical content result in an uncertain estimate of the real effect of congestion.
- The short-term nature of the model does not enable the impact of subsidy over cost structure and supply capacity to be evaluated.

1.2. Transit subsidization as a pricing problem

In Welfare Economics we do not find a direct definition of subsidy. We could say that it is a derived concept. Subsidy emerges as the difference between an *opportune* unit price and the unit cost for producing the quantity of service demanded at that price. Therefore welfare economists are mainly concerned with the extension of the concept of *opportune price*. This has usually been accomplished through the definition of second and third best solution to first-best deficiencies. Production costs usually do not play any relevant role. With given technology and perfectly elastic supply of inputs it could be assumed that the service demanded is produced at minimum costs. Therefore the standard pricing problems contain a traditional cost function.

The model developed by Glaister and Lewis (1978) moves within this framework. It identifies optimal pricing schemes of urban public transit in presence of congestion and defines the second-best subsidy that allows its social costs to be internalised. The authors consider three means of transport: car, bus and train. Each mode is considered at peak (*p*) and off-peak time (*f*). Individual utility depends on personal use level and on *time losses* associated with each modes. Since these depend, through speed-flow relationships, on the levels of collective use of each one of the six means, it follows that individual utility functions could be written as $u_h(x_h,X)$, where x_h and X are respectively the vectors of individual and collective use for each modes considered. Therefore individual expenditure functions can be written as $g_h(p_h,X,u_h)$ leading to the aggregate expenditure:

---

[8]All the limits listed are escapable in theory. Dodgson and Topham (1987) introduced distributional elements in the cost-benefit calculus. Bly and Oldfield (1987) extended the model in order to consider environmental impact, capacity optimization and optimal mixing of tariff reduction and quality improvement policies. However all these extensions would imply higher costs of model building.

$$G(p, X, u_1, \ldots, u_n) = \sum_{h=1}^{n} g_h \qquad (5)$$

In order to obtain optimal tariffs Glaister and Lewis solved the following maximization problem[9]:

$$\begin{aligned} \underset{p_{Bp}, p_{Bf}}{Max} \Big\{ & G\big(\alpha_{Bp}, \alpha_{Bf}, X^{Ap}(\alpha_{Bp}, \alpha_{Bf}), X^{Bp}(\alpha_{Bp}, \alpha_{Bf}), \hat{p}, u\big) - \\ & - G\big(p_{Bp}, p_{Bf}, X^{Ap}(p_{Bp}, p_{Bf}), X^{Bp}(p_{Bp}, p_{Bf}), \hat{p}, u\big) - \\ & - \left[C^{Bp}(X^{Ap}, X^{Bp}) - p_{Bp} X^{Bp}\right] - \left[C^{Bf}(X^{Af}) - p_{Bf} X^{Bf}\right] \Big\} \end{aligned} \qquad (6)$$

where C^{Bp}, peak bus service production cost, is dependent on its own level of use, X^{Bp}, and, because of congestion, on car use during the same period, X^{Ap}; α_{Bp} and α_{Bf} are a set of base prices, considerably higher than those under consideration. Thus the difference between the expenditure function evaluated at the two points is the compensating variation, i.e. the amount of money that would be required to compensate for an increase in tariffs from p to α. The remaining two terms in square brackets are the operating subsidies required on the peak and off-peak bus services. The sum of the compensating variation and the gross revenues is the total willingness to pay from which must be subtracted the various operating costs.

Differentiating (6), equating to zero and rearranging terms we obtain the following expressions for the optimal bus tariffs:

$$p_{Bp} = S_{Bp} + \frac{S_{Ap}}{1-\rho} \frac{X^{Ap}}{X^{Bp}} \left\{ \frac{\eta^{Ap}_{Bp}}{\eta^{Bp}_{Bp}} - \frac{\eta^{Bf}_{Bp}\eta^{Ap}_{Bf}}{\eta^{Bp}_{Bp}\eta^{Bf}_{Bf}} \right\} \qquad (7)$$

$$p_{Bf} = S_{Bf} + \frac{S_{Ap}}{1-\rho} \frac{X^{Ap}}{X^{Bf}} \left\{ \frac{\eta^{Ap}_{Bf}}{\eta^{Bf}_{Bf}} - \frac{\eta^{Ap}_{Bp}\eta^{Bp}_{Bf}}{\eta^{Bp}_{Bp}\eta^{Bf}_{Bf}} \right\} \qquad (8)$$

where η_j^i represent the direct and cross elasticities of compensated demand, S_{Ap} is the marginal social cost of a peak car passenger per mile and ρ is a combination of elasticities of demand reasonably positive and smaller than 1. Assuming that the modes are substitutes, (7) and (8) indicate that both peak and off-peak prices will be below respective marginal social costs by an amount proportional to marginal social cost of car use, both because of the possibilities of attracting peak car users directly and reallocating

[9]We removed rail modes in order to simplify formal treatment.

demand between periods. Simulations performed do not bring any unexpected results. For a quite broad range of values of the parameters and marginal costs, second-best subsidy should be higher during off-peak periods.

De Borger et al. (1993) recently applied an improved version of Glaister and Lewis model of urban transport in Belgium. Social cost includes congestion costs, environmental costs and accidents. In this way the model provides an exhaustive treatment of second-best arguments for transit subsidy. Results, obtained also considering car user price as a policy variable, indicate that each tariff should be set at the corresponding marginal social cost. With a binding budget constraint optimal tariffs will be fixed on the basis of a Ramsey rule and will be higher than the respective marginal social cost, both in peak and off-peak. In the simulations performed by De Borger et al. the extent of this increase for bus is quite surprising, especially if compared to that of private cars.

1.3. Some criticisms of the standard welfarist approach

As we mentioned before the welfarist approach has been criticized for several reasons. The high level of spatial aggregation, insufficient time disaggregation, a simplified treatment of cost functions and the lack of explicit geographic content are some of the more relevant *internal limits* of models strictly based on this approach. Moreover we should not diseregard *external limits* due to the narrow definition of the mobility system. The *welfarist approach* focuses only on direct and indirect effects which pass through final users, identifying mobility needs with transport demand. In this way the empirical relevance of the approach may be quite limited in scope.

In the following we will not discuss approaches coping with these types of limitations[10]. We do believe that the standard welfarist approach has merits that largely exceed its limits. However there is one more typical external limit, concerning the organizational contents, that we consider interesting to focus on here. This is the limit that makes the approach "out-of-date".

The standard welfarist approach deliberatly disregards any references to the organizations governing supply of public transit services. It assumes that organization is simplified to a single decision maker that is a social welfare maximizer. It could be a regulator, as implicitly assumed in the cost-benefit analysis approach, or the transit firm, as is done in the second-best pricing problem. This drastic reduction changes the standard welfarist models to simple tools for evaluating transport policy options, but leaves important questions concerning the real implementation of those policy options completely unanswered.

---

[10]Two examples of such approaches are the application of Activity Analysis to public transit subsidization provided by Goodwin et al. (1983) and that of the so-called Lewes Approach, see Searle (1987).

In this last respect, as we will show in the next section, organization counts. First of all, by referring to some early papers, we will treat the problem of defining commercial criteria for public transit firms as good proxies to cost-benefit rules. Therefore we will change perspective by moving to positive analysis of public transport firms. We will show that these firms do not follow prescribed objectives.

2. THE ORGANIZATIONAL CONCERN

2.1. Optimal commercial criteria for public transit firms

The organizational concern originally emerged among the advocates of the standard welfarist approach as a problem of bounded rationality. Since the public transit firms are separate bodies from the regulator subsidizing them, it might be that they are not rational enough to implement, by themself, a welfare maximization program. Therefore it is necessary to provide them with simple commercial criteria which might approximate social welfare objectives.

In order to implement a social welfare maximizing program, i.e. a couple *(x,q)* of quantity, *x*, and quality, *q*, of transit service, each firm should solve the following problem:

$$\underset{x,q}{Max} \int_0^x b(v,q)dv - xb(x,q) + (1+\lambda)\big(b(x,q)x - C(x,q)\big) \tag{9}$$

where *b(x,q)* is the inverse demand function for transit.

From (9), the first order necessary conditions, defining price and quality of service, are the following:

$$\frac{(b-C_x)}{b} = -\frac{\lambda}{1+\lambda}\frac{1}{\varepsilon_b} \tag{10}$$

$$\frac{\int_0^x p_q(v,q)dv - C_q}{x} = \lambda\left(\frac{C_q}{x} - p_q\right) \tag{11}$$

Ramsey's rule (10) is easy to implement. The transit firm knows all the elements composing it, or can easily discover them through marginal price and quality adjustments. On the contrary it is not so simple to implement rule (11) that prescribes to increase quality as long as net average benefit over the set of inframarginal consumers exceeds the value of social loss on the last quality improvement. In this case it is impossible to evaluate the integral term by means of simple marginal experiments on price and quality. Therefore public transit firm will not be able to fix the optimal level of quality as long as it does not develop a demand simulation model. Since this is quite difficult, the regulator should provide the firm with the simple commercial rules we mentioned above.

Ridership maximization and vehicle-mile maximization, both subject to a budget constraint, have been the natural candidates examined in literature. Papers by Nash (1978), Glaister and Collings (1978) and Bös (1978) stressed the relative merits of the former in comparison to the latter. However the analysis does not yield any general conclusion. On the contrary the more general result in this stream of literature is the "negative" conclusion[11] reached by Frankena (1983).

Frankena demonstrated that, in quite general conditions, both mileage and ridership maximization under a budget constraint may lead a non profit monopolist to fix a tariff and to supply a quality higher than those implemented in the second-best allocation defined by (10) and (11). In particular this happens if marginal valuation of quality increase as long as consumers' total willingness to pay reduces[12], i.e. if $b_{xq}>0$, and marginal cost does not change with the quality of service, i.e. $C_{xq}=0$. In these circumstances a ridership maximizer will produce a level of quality higher than socially optimal because of the excessive weight he attaches to the high valuation of quality of marginal users. The same holds true for a mileage maximizer. By changing assumptions about demand, i.e. by assuming that $b_{xq}<0$, Frankena demonstrated that ridership maximization enabled the firm to choose lower tariffs and quality levels than *socially optimal*, while the previous conclusion still holds with mileage maximization. Finally, if $b_{xq}=0$ ridership maximization may lead to the *social best*. Frankena concluded that without knowing the demand and cost functions one cannot determine whether ridership maximization would lead to levels of fares and quality below, equal or above those which would be second-best efficient under the budget constraint. Moreover, unless one has information about transport firm's objective function, demand and cost function, one cannot determine which subsidy formula would be the most socially efficient[13].

[11]The narrowness of these theoretical results partially justify the relevance, in public transit economics, of a-theoretical approaches to performance analysis.
[12]In other words, willingness to pay for quality improvement is assumed to be higher among users with a lower total willingness to pay.
[13]In a previous paper, Frankena (1981), the author considered the effect of different subsidization schemes (lump sum, matching on cost and on passenger) on the performance of transit firms with different objective functions.

2.2. Positive analyses of public transport firms

All the normative models we have examined up to now contain a conventional treatment of cost relationships. It is usually assumed that the transit firm is a public enterprise that produces sustaining costs given by a canonical, well-behaved, cost function. Therefore managers are not only expected to determine level of service and standards of quality according to some welfare calculus but are also implicitly requested to define inputs mix that minimize total cost. Many theoretical and empirical positive analyses of transport firms behaviour are actually conflict with this simplified picture.

The more recent literature of public finance suggests that the behaviour of public enterprise is such that cost minimization at given prices is constrained by demand level, available technology and some political conditions which may introduce a preference for some specific inputs. Therefore public enterprise should not necessarily produce output at minimum cost. In the specific analytical approach suggested by Rees (1984) public enterprise decisions are assumed to result from a negotiation process between the firm's management and the internal labour union. The interaction of managerial preferences[14], union preferences and political control might easily lead far from standard cost minimization[15].

Several statistical analyses, for example those by Pucher et al. (1983), Pickrell (1985) and Button (1988), showed that a relationship exists between the amount of subsidy and the cost level of transit. Transit subsidy is positively correlated with unitary costs and level of service and negatively correlated with labour productivity[16]. Another piece of evidence comes from the comparative studies of private and public companies. These studies repeatedly concluded that, also in the public transit sector, the former are commonly more efficient than the latter[17].

---

[14]Potential objectives of public enterprise managers, suggested in the literature, are, among the other, budget maximization, output or revenue maximization and even price stabilization or energy consumption minimization. See Bös (1986).

[15]De Borger (1993) recently applied this framework to the case of Belgian national railroad company. In his model the company does not minimize cost with respect to market prices of inputs but with respect to shadow prices generated by political constraint and bargaining power of labour union. Empirical estimation of the shadow cost function suggests that the shadow wage varies from 67% to 86% of observed wage and that this implies a misallocation of labour such that 4,4% more labour has been employed than stricly required.

[16]Pucher et al. (1983) show that the impact on unitary costs is higher the further is the firm from the government agency subsidizing it. Similar results have been reached by Filippini, Maggi and Prioni (1992). Bly and Oldfield (1986) by using regression with lagged variables provide some evidence on the causality links from subsidy to unitary cost increase.

[17]Some analyses of the British bus deregulation experience confirm this prescription. Heseltine and Silkcock (1990) for instance report that the newly privatized transit firm lowered quality, increased prices and reduced unitary costs by cutting labour, i.e. by redefining productive mix.

These empirical findings confirm theoretical deductions about the public enterprise by suggesting that public transit firms generally have non conventional objectives and therefore follow non standard behavioural programs. In general it is therefore reasonable to assume that transit firm's objective function does not coincide with regulator's objective function. It is in this context that organization may matter.

2.3. Why does organization matter?

Organization matters as long as the agent in charge of producing and supplying the public transit services is different from the regulator who should subsidize or regulate him. This difference becomes relevant from the point of view of organization if both the following conditions hold:

1) The agent and the regulator should have different, conflicting objective functions. In this respect it is not relevant if the agent is a public or a private firm.
2) There should be some asymmetry of information that favours the agent in face of the regulator.

If 1) holds but 2) does not, it happens that, since the regulator knows everything about the agent, he could always put her in the conditions to deliberatly reach his own preferred outcome. In the reverse situation the agent would have the incentive to reveal all his private information to the regulator. If both 1) and 2) hold the regulator has an imperfect control of the regulated agent. Therefore organization matters as it becomes economically relevant to cope with the imperfectness of control.

In the early perfomance approach the agent could not share the regulator's objective function because of lack of rationality. However it was implicitly assumed that the agent would be willing to do. Therefore the regulatory problem was to identify a good substitute for his objective function that the agent could easily follow.

This organizational arrangement is quite unrealistic since there are good reasons to assume that the agent is not willing to follow the regulator's prescriptions. A private transit firm obviously has objectives that are conflicting with those inspiring the regulator. Otherwise free market arrangement would be preferable. Moreover we have seen that the public enterprise has its own specific objective functions. These might be significantly different from a "simple" social welfare function. As far as information is concerned, it is quite realistic to assume that it will be asymmetric, with a fully informed management of the regulated firm and with imperfectly or incompletely informed regulator. His lack of information results from lack of observation. Certain actions of the firm cannot be directly observed by the regulator. Alternatively or cumulatively, certain information about the state of the world can be observed only by the firm, not by the regulator. Therefore we are generally confident that organization should matter.

Within this context new questions arise concerning public transit subsidies. If organization matters, subsidy must be granted not only in order to let the sector produce in quantities and qualities otherwise not available but also to solve organizational problems, i.e. taking into account the lack of control on the regulator side. Therefore the question becomes: How can a public transit firm be subsidized in order to reach some specific goals and given that it has its own rationality? The incentives theory provides the right theoretical background for answering this question.

3. THE INCENTIVES THEORY APPROACH

The incentives theory provides a solution to the so called principal-agent problems. These problems emerge whenever a principal wants to induce an agent to take some action which is costly to the agent, as long as it will require some effort. The principal may be unable to directly observe the action of the agent, but instead observes some output that is determined, at least in part, by the actions of the agent. The principal's problem is to design an incentive payment for the agent that induces the latter to make the best decision from the principal's viewpoint. He should do this by taking into account two sorts of constraints.

Since the agent may have another opportunity available that gives him some reservation level of utility, the principal should design an incentive payment that ensures that the agent gets at least this reservation level. Otherwise the agent would not be willing to partecipate. This is the so-called *partecipation* constraint or individual rationality constraint.

The second constraint on the problem is that of *incentive compatibility*: given the incentive schedule defined by the principal, the agent will pick the best action for himself. As long as the agent attempts to maximize his own return he will exploit all his informational advantage in order to get the maximum payment and to exert the minimum effort. Since the principal cannot choose the agent's action directly but can only influence it by the incentive payment, he must set it by taking into account the utility maximizing behaviour of the agent.

In order to make these concepts clear we present in the next paragraph the theory of regulation recently developed by Laffont and Tirole (1993). Paragraph 2 provides an application of this theory that fits into our specific subsidization problem. In the last section we summarize the general rules for an incentive use of transit subsidy.

3.1. Incentives theory of regulation: the Laffont and Tirole framework

Laffont and Tirole develop a theory of regulation based on the following main features:

0. Regulation is defined through a contract between the regulator and the firm
1. The regulated firm has private information about its technology at the date of contracting, and its cost-reducing effort is unobserved by the regulator. Cost function will be written as $C = C(\beta, e, \ldots)q$, where β is a technological parameter and e is the cost-reducing effort[18]. Let $\psi(e)$ denote the firm's managers' disutility of effort expressed in monetary terms, such that $\psi'>0$ (effort is costly) and $\psi''>0$ (the cost of effort is convex). The model assumes that the regulator has incomplete information about the cost function but not about the function $\psi(e)$. The firm knows β, and the regulator has a distribution over β in an interval $\left[\underline{\beta}, \overline{\beta}\right]$.
2. Realized cost C^r, the outputs and the prices are verifiable. However the regulator cannot disentangle the various components of costs.
3. The firm can refuse to produce if the regulatory contract does not guarantee it a minimum level of utility. Let U denote the firm's expected utility. We normalize the individually rational level at zero. $U \geq 0$ will be called the firm's rent.
4. The regulator offers a monetary transfer t, to the firm.
5. The firm and the regulator are risk neutral with respect to income.
6. By accounting convention, the government receives the firm's revenue from charges to consumer, pays the firm's cost and a net transfer t. Transfers are of the linear type in realized cost C^r, i.e. $t=a-bC^r$, where a is a "fixed fee" and b is the fraction of costs incurred by the firm.
7. Firm's objective function is given by the total amount of transfer less monetary disutility of effort, i.e. $U = t - \psi(e)$
8. The regulator designs a regulatory contract in order to maximize total surplus in society, given that he faces a shadow cost of public funds $\lambda > 0$.

Ignoring for the moment output and quality decisions, the regulator has two conflicting goals: to promote cost reduction and to extract firm's rent[19]. In order to see why these two goals are in conflicts consider the two polar cases of cost-plus (i.e. $b=0$) and fixed-price (i.e. $b=1$) transfers. A fixed-price contract induces the highest amount of effort because it makes the firm residual claimant for its cost savings. Therefore the firm

[18]By convention $C_\beta > 0$, i.e. a high β corresponds to an inefficient technology, and $C_e<0$ and $C_{ee}>0$, i.e. effort reduces cost at a decreasing rate. Omitted variables in the cost function may be the vector of outputs $q_1,..q_n$, of goods 1,..., n or the level of service quality, s.

[19]In this case ex-post social welfare for an utilitarian regulator is given by:

$$S-(1+\lambda)(t+\beta-e)+t-\psi(e)=S-(1+\lambda)[\beta-e+\psi(e)]-U$$

where S is the consumer surplus due to the project. The crucial feature of this welfare function is that the regulator dislikes leaving a rent to the firm.

has the socially optimal incentive to reduce cost as it receives all the money it saved. However, a cost-plus transfer offers no incentive for cost reduction, as long as the firm does not appropriate of its cost savings.

In the case of rent extraction the logic is reversed. Under a fixed-price transfer any exogenous reduction in cost is received by the firm. The firm's rent therefore is very sensitive to the technological environment. On the other hand, a cost-plus contract is ideal for rent extraction because any exogenous variation in cost is received by the government and not by the firm.

When the regulator has perfect information about the techonology (there is moral hazard but no adverse selection), the optimal regulatory contract is a fixed-price transfer[20]. When the firm has private information about its technology optimal contracts are incentive contracts trading off effort inducement, which calls for fixed-fee transfers, and rent extraction, which calls for a cost-plus transfer. In general it is optimal for the regulator to offer a *menu* of incentive transfers, since the transfer should be tailored to the firm's information. An inefficient firm should not be regulated with the same contract as an efficient firm. The regulator discriminates among the different potential types of firm in the same way the monopolist price discriminates among consumers with different valuation for quantity or quality.

The regulatory contract emerges as the solution of a mechanism design game, that is a three step game of incomplete information, where the agent's type is private information. In step 1 the regulator designs a contract $[t(\mu),C(\mu)]$ that specifies for each announced value of the signal μ sent by the firm a net transfer to the firm $t(\mu)$, and a cost to realize $C(\mu)$. In step 2 the firm decides to accept or reject the contract. In step 3, the firm which accepts the contract plays the game specified in the contract. The *revelation principle*[21] shows that, to obtain the highest expected payoff, the principal can restrict attention to contracts that are accepted by the firm at step 2 and in which at step 3 the firm truthfully reveal its type. As we will show, the revelation principle allows the regulatory problem to be solved in terms of standard optimal control techniques.

3.2. A model of transit firm regulation

We saw that as long as organization matters, public transit subsidization should be treated as a regulation problem[22]. Organizational concern emerges in the public transit sector because of the following typical circumstances:

---

[20]In this case the fixed fee is optimally set at the lowest level consistent with the firm's partecipation provided that the firm chooses the effort that minimizes $C+\psi(e)$.

[21]We are referring here to the textbook treatment contained in Fudenberg and Tirole (1991), ch. 7.

[22]See the paper by Pedersen (1994) for the first attempt, to my knowledge, in this direction.

1. *The transit firm maximizes its own rent while the regulator is concerned with social welfare*. This latter circumstance is justified in that we are assuming the (normative) point of view of the regulator. That the transit firm maximizes its own monetary rent seems a good positive assumption at least for private companies. In the case of a public enterprise it could be intepreted as a typical managerial objective.

2. There is an asymmetry of information between the firm and the regulator in that:

   a) *The firm is better informed than the regulator about technology and about local demand conditions*. If we think of a regional authority in charge of subsidizing a multiplicity of local transit suppliers these circumstances seem quite plausible. There will be always some local specific features affecting technology and demand, that can not be properly detected by the regional authority.

   b) *The firm cost-reducing effort can not be directly observed and quality standard is not easily verifiable by the regulator*. While the first point seems obvious the second is worth some comments. Quality of transit is difficult to check as long as it depends on local conditions (demographic, geographic, social, ...). Therefore even if quality could be observed[23], in principle, it would be difficoult to quantify and to include in a formal contract.

Given these circumstances, the regulatory problem appears quite complicated as long as it is one of two-dimensional adverse selection and two-dimensional moral hazard. An important simplifying feature is that quality of transit service can usually be observed by users before use. Thus the regulator might recreate the incentives of an unregulated firm to provide quality by rewarding the regulated firm on the basis of sales. Laffont and Tirole (1993), ch. 4, provide a thorough analysis of this regulatory problem. In the following we give a brief overview of their model. We report some details in an appendix.

The transit firm, which is a local monopoly, produces the service in quantity q with quality, which can be observed by users but not by regulator, s. Cost function is given by:

$$C = (\beta + s - e)q \tag{12}$$

where β is the technological parameter and e is cost-reducing effort.

Firm's rent is given by:

$$U = t - \psi(e) \tag{13}$$

[23]This implies in any case high costs of minitoring and auditing.

Users derive from the consumption of the transit service a gross surplus:

$$S^g(q,s,\theta) = (A + ks - h\theta)q - \frac{B}{2}q^2 - \frac{(ks - h\theta)^2}{2} \tag{14}$$

where *A*, *B*, *h* e *k* are known positive constants and θ is a demand parameter[24].

The users'/taxpayer net surplus S^n, given by gross surplus net of individual expenditure and social cost of providing the service, is:

$$S^n = S^g - \left[pq + (1+\lambda)(C - pq + t)\right] \tag{15}$$

3.2.1. The complete information case

Under complete information a utilitarian regulator maximizes the sum of consumer and producer surpluses under the constraint that the firm be willing to partecipate:

$$\underset{(q,s,e)}{Max}\left\{W(q,s,e) = S^n + U\right\} \tag{16}$$

s.c. $U \geq 0$ (17)

Interior maximum is characterized by the first-order conditions[25]:

$$\frac{p - C_q}{p} = \frac{\lambda}{1+\lambda}\frac{1}{\eta} \tag{18}$$

$$\frac{\partial S^g}{\partial s} + \lambda\frac{\partial P}{\partial s}q = (1+\lambda)C_s \tag{19}$$

$$\psi'(e) = q \tag{20}$$

$$U=0 \tag{21}$$

where $\eta = p/Bq$. Equation (18) says that service is produced up to the point where the Lerner index is equal to a Ramsey index time the inverse of elasticity of demand. The

24 Quantity and quality are net complements if $k>1$ and net substitutes if $k<1$. Quantity and quality are net complements if an increase in quality raises the net marginal willingness to pay, that is the difference between price and marginal cost.

25 For *B* large enough the program [(16), (17)] is concave and its interior maximum is fully characterized by the first-order conditions.

optimal level of quality equates the marginal gross surplus plus the shadow cost of public funds times the increase in revenue to the social marginal cost of quality. Equation (20) equates the marginal disutility of effort to its marginal utility q, and equation (21) says that no rent is left to the firm.

3.2.2. Optimal regulation with asymmetric information

Now we take into account the asymmetry of information. The regulator knows neither β nor θ and cannot observe e and s. However he observes C, p and q. We assume that the regulator maximizes expected social welfare and has a prior cumulative distribution F_1 on $\beta \in \left[\underline{\beta}, \overline{\beta}\right]$, and F_2 on $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$. The firm knows β and θ before contracting.

The regulator knows that users equate their marginal utility of service to the price:

$$p = A + ks - h\theta - Bq \tag{22}$$

By using equation (22) it is possible to eliminate the unobservable quality level s in the users' gross surplus (14), which becomes:

$$S^g(p,q) = \frac{B}{2}q^2 + pq - \frac{1}{2}\left(p - A + Bq\right)^2 \tag{23}$$

Similarily the cost function becomes:

$$C = \left(\beta + \frac{h\theta}{k} - e\right)q + q\left(\frac{p - A + Bq}{k}\right) \tag{24}$$

The firm objective function will be:

$$U = t - \psi(e) = t - \psi\left(\beta + \frac{h\theta}{k} + \frac{p - A + Bq}{k} - \frac{C}{q}\right) \tag{25}$$

β and θ enter the cost function and the firm's objective function only through the linear combination $\gamma = \beta + h\theta/k$. This feature, which holds also for the regulator's objective function, reduces the model to a one-dimensional adverse selection model. We note at this point that high (low) values of β and θ, which are due respectively to a low (high) cost efficiency and a low (high) demand, imply high (low) values of γ.

The regulator wishes to maximize social welfare. From the revelation principle we can restrict the problem to the analysis of direct and truthful revelation contracts. The individual rationality constraint of the firm is:

U(γ)≥ *0* for any γ
(26)

while truth telling is guaranteed by the following first- and second- order conditions of incentive compatibility (see appendix A.1.):

$$\partial U / \partial \gamma = -\psi'(e) \tag{27}$$

$$\partial e / \partial \gamma - 1 \leq 0 \tag{28}$$

Thus the regulator designs the contract *[t(γ),c(γ),p(γ),q(γ)]* that specifies for each value of γ announced by the firm a net transfer to the firm *t(γ)*, an average cost to realize *c(γ)*, a price to charge *p(γ)* and a quantity to sell *q(γ)*, by solving the following optimal control problem[26]:

$$\max_{\{p(\cdot),q(\cdot),e(\cdot),U(\cdot)\}} \int_{\underline{\gamma}}^{\bar{\gamma}} \{W(p,q,e)\} dF(z) \tag{29}$$

under constraints (26), (27) and (28), where the welfare function *W* is a obtained by substituing out quality from the social welfare function (16), and the cumulative distribution *F(γ)* is the convolution of F_1 and F_2.

Maximizing the hamiltonian with respect to *q* and *p* we obtain conditions which correspond to (18) and (19), i.e., for a given effort *e*, the price, quantity and quality are the same as under complete information about technology and demand parameter. Moreover from (27) we see that *p* and *q* do not affect the rate at which the rent must be given up to the firm. Maximizing the hamiltonian with respect to *e* we get:

$$\psi'(e) = q - \frac{\lambda}{1+\lambda} \frac{F(\gamma)}{f(\gamma)} \psi''(e) \tag{30}$$

If we compare (30) with (20) we see that in order to extract part of the firm's rent the effort is distorted downward for a given output level, except for the most efficient type γ= $\underline{\gamma}$[27]. Equation (30) has a straightforward interpretation. If we raise effort of types in [γ,γ

[26]For *A* and *B* large enough the program is concave and the optimum is characterized by its first-order conditions [see Laffont and Tirole (1993) appendix A4.3.].

[27]Laffont and Tirole show that the level of quality is lower under incomplete information than under complete information if and only if quantity and quality are net complements. Incomplete information

$+d$] (in number $f(\gamma)d\gamma$) by de, productive efficiency increases by $[1-\psi'(e(\gamma))]de$ for these types, which yields social gain $(1+\lambda)[q-\psi'(e(\gamma))]def(\gamma)d\gamma$. However this also raise the rent of types in $[\underline{\gamma},\gamma]$ (in number $F(\gamma)$). From (27) the rent of type γ is increased by $\psi''(e(\gamma))ded\gamma$, and so is the rent of types $\gamma°<\gamma$. The social cost of the extra rents is $\lambda\psi''(e(\gamma))(de)(d\gamma)F(\gamma)$. At the optimum the marginal cost must equal the marginal benefit, which yields (30).

Differentiating (30) we get:

$$\partial e/\partial\gamma = -\frac{[\lambda/(1+\lambda)]\psi''(e(\gamma))(d/d\gamma)[F(\gamma)/f(\gamma)]}{\psi''(e(\gamma))+[\lambda/(1+\lambda)][F(\gamma)/f(\gamma)]\psi'''(e(\gamma))} \tag{31}$$

As long as the distribution F satisfies the monotone hazard rate property[28] $(d/d\gamma)[dF/df]>0$ and $\psi'''>0$ it follows that the condition (28) is confirmed, that is in the optimal allocation the effort is decreasing in the firm's type. The same is true for the rent function that is given in the optimum by:

$$U^*(\gamma) = \int_{\gamma}^{\bar{\gamma}} \psi'(e^*(\tilde{\beta}))d\tilde{\beta} \tag{32}$$

where e^* is the solution of (30). Therefore the less efficient types, those with an high cost inefficiency and low demand exerts a lower effort and receives a lower rent than the more efficient types, that is those with low cost inefficiency and high demand[29].

To implement the optimal regulatory contract the regulator must define an appropriate transfer $t(\gamma)$ offered to the firm in order to induce truthful behaviour. This transfer can be intepreted as a function $T(.)$ of the variable $z=C/q-(p-A-Bq)/k$. The first-order conditions for truthful revelation are (see appendix A.1):

$$\frac{\partial t}{\partial\gamma} + \psi'(\gamma - z)\frac{\partial z}{\partial\gamma} = 0 \tag{33}$$

This implies that:

---

makes rent extraction difficult, reduces power of incentives and leads to a decrease of effort. This raises marginal cost and reduces output. If quantity and quality are net complements, lower services are desirable, and conversly for net substitutes. Asymmetric information lowers quality when the unregulated monopolist oversupply quality.

[28]This property amounts to say that as long as the firm's type is getting "better" (more efficicient+higher demand) it is increasingly less probable that it could become even better.

[29]Type $\gamma = \bar{\gamma}$ will receive $U=0$ and exert low effort; type $\gamma=\underline{\gamma}$ will receive the maximum rent and will exert the effort e^* such that $\psi'(e^*)=q$.

$$\frac{\partial T}{\partial z} = -\psi'(\gamma - z) < 0 \tag{34}$$

Therefore the transfer is a decreasing function of *z*. Moreover it can be demonstrated (see Laffont and Tirole (1993), appendix A4.3) that differentiating (34) the transfer as a function of *z* is also a convex function. Since *T(z)* is convex it can be replaced by the family of its tangents. These tangents represent a menu of contracts that are linear in the observed value of *z*.

$$t(\gamma°,z)=t^*(\gamma°)+T'[z-z(\gamma°)]=t^*(\gamma°)+\psi'(e^*(\gamma°))[z(\gamma°)-z] \tag{35}$$

where γ° is a firm's announcement and γ* denotes solutions to the optimal regulatory contract. Thus the transfer function *t(γ)* can be replaced by the menu of linear contracts:

$$t(\gamma,z)=a(\gamma)+b(\gamma)[z(\gamma)-z] \tag{36}$$

where *z(γ)* is the announced value of *C/q-[p-A+Bq]/k*, *z* is the observed ex-post value. The transfer is therefore a function of a performance index that subtracts from the realized cost an approximation of the service quality inferred from market data. In other words, the firm is offered a choice in a menu of linear contracts and is rewarded or penalized according to deviations from an index aggregating cost data and service quality data inferred from observation of market price and quantity and from a priori knowledge of the demand function.

The menu of linear contracts (36) induces truth telling and the optimal level of effort. The firm is in fact solving the following problem:

$$\max_{\{e,\gamma°\}}\left\{U = t^*(\gamma°) + \psi'(e^*(\gamma°))\left[z(\gamma°) - \gamma + e\right] - \psi(e)\right\} \tag{37}$$

which implies that $\psi'(e^*(\gamma°))=\psi'(e)$ and therefore $e=e^*(\gamma°)$ and $(\partial t^*/\partial \gamma°)+\psi''(\partial e^*/\partial \gamma°)[z(\gamma°)-z]+\psi'(e^*(\gamma°)(\partial z/\partial \gamma°)=0$. From (33) it follows that $(\partial t^*/\partial \gamma°)+\psi'(e^*(\gamma°)(\partial z/\partial \gamma°)=0$ and then for any γ the firm reveals γ°=γ in order to have $z(\gamma°)=z$. Moreover from this follows that the optimal conditions for *q* and *p* hold. Therefore the direct revelation contract *[t(γ),c(γ),p(γ),q(γ)]* can be replaced by a contract such as *[t(z,γ),p(z),q(z)]*. That is, the regulator, instead of assigning the direct mechanism by asking to the firm for its type, γ, offers a contract *t(z,γ)* that induces the firm to implement the truthful value of *C/q-[p-A+Bq]/k* and the right level of effort.

It is interesting to note that the menu of linear contracts can be alternatively decomposed into a linear sharing of total cost overruns with a coefficient $b_1(\gamma)=\psi'(e^*(\gamma))/q^*(\gamma)$ and a linear sharing of overruns in the service quality index with coefficient $b_2(\gamma)=\psi'(e^*(\gamma))$, or

$$t = a(\gamma) + b_1(\gamma)\left[C(\gamma) - C\right] + b_2(\gamma)\left[\frac{p + Bq}{k} - \frac{p(\gamma) + Bq(\gamma)}{k}\right] \tag{38}$$

Moreover it can be demonstrated (see Laffont and Tirole (1993) that:

1) The fixed payment *a(z)* is a concave, increasing function of the parameters characterizing the power of the incentive schemes, b_1 and b_2.
2) The parameters characterizing the power of the incentive schemes b_1 and b_2, are positively correlated over the sample of types.

The more "efficient" type , faces a fixed price transfer, with $b_1=\psi'(e^*(\underline{\gamma}))/q=1/q$, $b_2=\psi'(e^*(\underline{\gamma}))=1$ and *a* set at the highest level, and is therefore residual claimant for its cost savings and sales increases. The other types will face incentive contracts that are intermediate between the fixed-price and the cost-plus contract.

4. INCENTIVE PROGRAMS IN PRACTICE

In this last section we wish to provide some general guidelines for the adoption of incentive schemes in public transit subsidization. We will not deal with the difficult questions regarding the design of "real" incentive contracts[30]. This is out of the scope of the paper and might be the subject of further research. We prefer here just to define some requirements from the previous treatment of the incentive approach. These criteria might be useful to understand old and current experiments with performance-based public transit subsidization.

4.1. General rules for incentive contract in the public transit sector

The publicly managed transit sector typically suffers from high deficits associated with low performance. It is thus quite obvious that incentive subsidization has been repeatedly advocated, and will be increasingly promoted, by local government in charge of financing these deficits. Therefore it is very important to understand, especially for the publicly managed transit firms, what makes an incentive subsidization scheme successful.

[30]See the paper by Reichelstein (1992) for a rare example of numerical computation of the incentive schemes for a government contract.

As long as we are explicitly referring to the theory of incentives as a normative paradigm, we identify conditions for the success of such schemes with the assumptions and normative prescriptions of that theory.

In order to be successful, the contract set by a regulator must be voluntarily accepted by the regulated firm. Moreover the regulator should not be able to withdraw his initial offer once it is accepted by the firm. Once these conditions have been met, the regulator and the regulated firm enter in a mechanism design game that can be solved in the way we showed in the previous section. The firm voluntarily accepts the contract as long as the regulator sets it in a way to leave her at least as well as she would have been if she had not accepted it. In order to exclude renegotiation the regulator should have a reputation incentive, i.e. it should be better for him not to reject the contract once accepted by the firm as long as he wishes to have the reputation for respecting agreements. Therefore, as a first set of prescriptions the theory suggests that the regulator:

1) quantifies the firm's reservation level of utility and considers it carefully while defining the contract;
2) invests in a reputation for respecting agreements. In this way the regulator will be commited to a non-renegotiation behaviour and will therefore be able to propose credible contracts, i.e. contracts that may induce truthful revelation.

A second general guideline is to give each firm its own contract, i.e. a contract properly tailored to the firm's information. Therefore there should be one contract for each firm. A properly defined contract induces truthful revelation, i.e. solves problems due to asymmetry of information. This implies that it should have a particular structure. We have seen that it might be defined as a transfer, linear in the realized, observed performance index, with a target rent, or fixed price component, plus a penalty, or bonus component, for performance overruns. Moreover we have seen that each contract must be defined as a *menu* of linear transfers, so that each firm might choose its own incentive transfer out of a menu of such transfers. The firm selects the target fixed price and the corresponding reward (penalty) rate by announcing its performance target. If the regulator properly calibrates these two parameters both conditional on the performace target announced by the firm, he might induce the firm to self-select, i.e. to announce the truthful performance target.

In relation to the specific regulation problem faced in the transit sector, we argue that a fundamental concern is quality. As long as transfers can not be made to depend on quality, since it cannot be oserved by the regulator, we have seen that the regulation problem can be represented as a two-dimensional moral hazard and two-dimensional adverse selection model. However we have shown that as long as the service is a search good, i.e. its quality is observed by users before use, this fairly complicated regulation problem might be solved quite easily. As long as the firm has the incentive to boost current sales, consumption is a signal of quality. In this situation, incentives to provide quality might be disconnected from those to reduce cost, if the firm is rewarded also on the basis of sales. We have seen therefore that an important role might be played by observed performance indicators directly linked to consumption. Moreover we have shown that the power of the incentive scheme should be determined by the regulator depending on the a-priori probabilistic knowledge of unobservable cost and demand parameters. High powered schemes should be given to firms with lower productive inefficiency and high demand parameters. This implies that high powered schemes should be dependent on low cost and high sales announced performance. An important question is to identify the more appropriate ex-post observable performance measures. An obvious choice would be to use only indicators that are easy to audit[31]. An alternative would be to introduce performance assessment on the basis of information provided by a third party (panel of users, firm's input suppliers, ..).

These general rules can be immediately applied by a regulator dealing with a private transit firm. On the contrary, some complications arise if the regulator wants to apply these rules to the design of performance-based subsidization for a system of publicly owned transit firms. We will briefly discuss these problems after having presented some of these experiences.

4.2. Some experiments with performance-based transit subsidization

Fielding (1992) surveys some of the experiments with performance-based transit subsidization carried out in the United States. All the cases he explicitly refers to were unsuccessful.

In Michigan the state transit authority calculated 47 indicators of transit performance. The mean, range and standard deviation for each indicator was calculated and each transit firm rewarded on the basis of its relative status in these rankings. The scheme failed because the use of so many indicators created confusion.

The state of Pennsylvania[32] was a very interesting example. Four indicators were calculated for each firm to cover the efficiency and effectiveness dimensions: cost per hour, revenue per hour, ridership per hour and the revenue-to-cost ratio. The state awarded 8.33% of the permissable deficit of each firm as a perfomance incentive, if the firm had maintained or improved performance over the previous year on the first three measures. To qualify for the incentive, a firm had to exceed a statewide revenue-to-cost

---

[31]See Pedersen (1994) on this regard.
[32] See Miller (1980).

ratio. This system, introduced in 1979, was later abandoned. Opposition arose because it was far easier for smaller firms to qualify for the incentive reward[33].

By referring to the theory of incentives, it is quite easy to see why these attempts failed: the regulator *imposed* the *same* contract on *each* firm. Therefore individual rationality and incentive compatibility constraints were not considered. This feature reveals that the underlying theory was still the standard welfarist approach, in that there was no concern for organizational aspects.

A slightly improved version of performance-based transit subsidization has been recently introduced in Emilia-Romagna. In Italy each regional transit authority receives a transfer from the central government that must be allocated to local transit firms[34] to cover their deficits. Usually this is done by political negotiations on historical sharing quotas. At the beginning of 1995, Emilia-Romagna regional authority accomplished this task by stipulating, separately with each local transit firm, incentive contracts (Contratti di Servizio). These contracts have the above-mentioned prescribed form, since they consist of a fixed payment plus a penalty, or bonus, component contingent on performance overruns.

Even if in theory this scheme is sound, in practice it suffers from at least three main drawbacks. 1) Performance is evaluated on the basis of 10 indicators calculated with data provided by each firm. 2) The penalty rate is set at the same value for each firm. In fact, the regional authority negotiated assuming that each firm should have been rewarded on the basis of the same penalty rates. 3) The penalty rate is very low, only 1% of the total eligible subsidy. Thus the contracts are all fixed-price transfers.

The first failure is quite evident since firms may give false report. The second is partly a consequence of the first: the penalty rate should not provide different incentives to false reporting. Finally the third drawback is a direct result of the second: if the penalty rate must be the same, then the firms collectively opt for the lowest one, that is for a fixed-price contract. This high powered contract allows the highest rent to be realized. From incentive theory we know that this contract induces the socially correct amount of effort in the most efficient type but may leave too high a rent to the low efficient ones. Therefore in the end this arrangement fails both in extracting rents and inducing effort at a socially desireable level. However it is still better than those surveyed by Fielding in that at least each firm deliberatly signs its own contract.

---

[33] More recently, Los Angeles county designed a similar subsidization scheme that had the same fate.
[34] In Italy local transit firms are publicly owned.

4.3. Open issues

According to the theory, all these experiences more or less failed to provide the proper incentives to the firms. This failure can be explained in two ways: 1) regulators do not know the incentives theory, 2) this theoretical test is much too demanding in that incentives theory does not account for all the complexities. Assuming that the answer lies somewhere in between, transport economists have two main tasks to accomplish: 1) to teach incentives theory to regulators and planners; 2) to improve it in order to correspond to reality. In this last case there are two main aspects of real transit subsidization schemes that must be carefully considered: dynamic contracting and endogeneity of reservation utility levels.

The first aspect, which is common to the subsidization of both private and public enterprises emerges if we consider that subsidization is usually part of a multiperiod relationship. In this context it might be quite difficult to let the agent reveal its own private information as long as the regulatory contract could be renegotiated or stipulated year by year. In such a case truthful revelation today implies to loose informational rents tomorrow.

The second issue may arise whenever subsidization involves a multiplicity of agents. The consistency of each menu of transfers might be undermined by an externality effect. The theory says that incentive contracts depend on the reservation level of utility, that is determined, for each agent, by his best alternative to signing the contract. It may be that this alternative depends on the number of agents that refuse to sign: the more agents refuse to sign the more easily they can cooperate in enforcing the regulator to implement a new more favourable subsidization scheme. Therefore reservation levels of utility may be influenced also by the extent of cooperation in refusing the contract and by the degree of rent extraction in itself.

5. CONCLUSIONS

The aim of this paper was twofold. First of all we aimed to evaluate the theories of public transit subsidization. After an early, abundance of contributions to the literature of Public Transit Economics mainly inspired by Welfare Theory and Public Economics, in the last decade the issue of transit privatisation determined a shift towards different theoretical paradigms. Transport economists debated about the opportunity of privatising public transit firms by referring very frequently to the Theory of Industrial Organization. This change has been implicitly determined by a shift from second-best to first-best arguments. Privatisation might be an answer to the enormous costs of organizations in charge of achieving the second-best.

Actually the experiences of deregulation showed that regulation through competition does not alone imply socially desireable consequences in the transit market. Moreover deregulation does not necessarily mean that transit subsidies are completely eliminated. Even in a privatized transit market, a local government may decide to give public subsidies in return for a socially superior transit service. From this point of view, the second aim of this paper was to suggest a new framework to define regulatory contracts that assign transfers to the transit firms according to the achievement of announced performance. The same reasoning is valid, a-fortiori, in the case of a publicly owned transit firm. The Theory of Incentives might provide the appropriate framework for analyzing such arrangements, i.e. to give a third-best answer to public transit problems that also take into consideration organizational constraints.

APPENDIX: FIRST AND SECOND ORDER CONDITIONS FOR TRUTHFUL REVELATION

The firm is faced with the contract *[t(γ),c(γ),p(γ),q(γ)]*. It chooses the announcement γ° that maximizes its objective function, that is, solves:

$$\max_{\gamma^\circ}\left\{t(\gamma^\circ)-\psi\left(\gamma+\frac{p(\gamma^\circ)-A+Bq(\gamma^\circ)}{k}-\frac{C}{q}(\gamma^\circ)\right)\right\}$$

$$\Leftrightarrow \max_{\gamma^\circ}\left\{t(\gamma^\circ)-\psi\left(\gamma-z(\gamma^\circ)\right)\right\} \qquad \text{(A.1)}$$

where

$$z(\gamma^\circ)\equiv\frac{C}{q}(\gamma^\circ)-\frac{p(\gamma^\circ)-A+Bq(\gamma^\circ)}{k} \qquad \text{(A.2)}$$

Thus, first order conditions of incentive compatibility is:

$$\frac{\partial t(\gamma)}{\partial\gamma}+\psi'(\gamma-z(\gamma))\frac{\partial z(\gamma)}{\partial\gamma}=0 \qquad \text{(A.3)}$$

and the second order condition is:

$$\frac{\partial z(\gamma)}{\partial \gamma} \geq 0 \qquad \text{(A.4)}$$

Equations (A.3) and (A.4) represent necessary and sufficient conditions for truthful revelation. They may be rewritten as equations (27) and (28) respectively. Let $U(\gamma)=t(\gamma)-\psi\ (\gamma-z(\gamma))$ denote type γ's rent. The envelop theorem applied to the maximization of (A.1) with respect to γ^* yields expression (27). Expression (28) derives from (A.4) as long as we note that $e(\gamma)=-z(\gamma)$.

REFERENCES

Baumol, W.J., 1975, "Payment by performance in rail passenger transportation: an innovation in Amtrak operations", *The Bell Journal of Economics*, 1, pp. 281-298.

Berechman, J., 1993, *Public Transit Economics and Deregulation Policy*, North-Holland, Amsterdam.

Bly, P.H. and Oldfield, R.H., 1987, "An analytic assessment of subsidies to bus services", in Glaister, S., edited by, *Transport Subsidy*, Policy Journals, Arrowsmith, Bristol.

Bly, P.H. e Oldfield, R.H., 1986, "The effects of public transport subsidies on demand supply", *Transportation Research A*, 20, pp. 415-427.

Bly, P.H., 1987, "Managing public transit transport: commerical profitability and social service", *Transportation Research A*, 21, pp. 109-125.

Bös, D., 1978, "Distributional effects of maximizing passengers miles", *Journal of Transport Economics and Policy*, 12, pp. 322-329.

Bös, D., 1986, *Public Enterprise Economics*, North-Holland, Amsterdam.

Button, K., 1988, "Subsidies, political control and costs of U.K. urban bus provision", *Transportation Research Record*, n. 1012, pp. 8-13.

Caillaud, B., Guesnerie, R., Rey, P. e Tirole, J., 1988, "Government intervention in production and incentives theory: a review of recent contributions", *Rand Journal of Economics*, 19, pp. 1-26.

De Borger, B., 1993, "The behaviour of public enterprise offering scheduled services", paper presented at the 38th International Conference of The Applied Econometrics Association, Athens, April 13-14, 1993.

De Borger, B., Meyers, I., Proost, S. and Wouters, S., 1993, *Social Cost Pricing of Urban Passenger Transport*, Public Economic Research Paper n. 34, CES, Katholieke Universiteit Leuven, Belgium.

Dodgson, J.S. and Topham, N., 1987, "Benefit-cost rules for urban transit subsidies", *Journal of Transport Economics and Policy*, 21, pp. 57-71.

Dodgson, J.S., 1987, "Benefits of changes in urban public trasnport subsidies in the major Australian cities", in Glaister, S., edited by, *Transport Subsidy*, Policy Journals, Arrowsmith, Bristol.

Evans, A., 1985, "Equalising grants for public transport subsidy", *Journal of Transport Economics and Policy*, 19, pp. 105-138.

Fielding, G.J., 1987, *Managing Public Transit Strategically*, Jossey-Bass, San Francisco.

Fielding, G.J., 1992, "Transit performance evaluation in the U.S.A.", *Transportation Research A*, 26, pp. 483-491.

Fielding, G.J., Babitsky, T.T. and Brenner, M., 1985, "Performance evaluation for bus transit", *Transportation Research A*, 19, pp. 73-82.

Filippini, M., Maggi, R. and Prioni, P., 1992, "Inefficiency in a regulated industry: the case of the Swiss regional bus companies", *Annals of Public and Cooperative Economics*, 63, pp. 437-455.

Frankena, M., 1981, "The effects of alternative urban transit subsidy formulas", *Journal of Public Economics*, 15, pp. 337-348.

Frankena, M., 1983, "The efficiency of public transport objectives and subsidy formulas", *Journal of Transport Economics and Policy*, 17, pp. 67-76.

Fudenberg, D. and Tirole, J., 1991, *Game Theory*, MIT Press, Cambridge, Mass.

Glaister, S. and Collings, J.J., 1978, "Theory and practice of maximizing passenger miles", *Journal of Transport Economics and Policy*, 12, pp. 304-321.

Glaister, S. and Lewis, D.L., 1978, "An integrated fares policy for transport in Greater London", *Journal of Public Economics*, 9, pp. 341-355.

Glaister, S., 1987, "The allocation of urban public transport subsidy", in Glaister, S., edito da, *Transport Subsidy*, Policy Journals, Arrowsmith, Bristol.

Glaister, S., edited by, 1987, *Transport Subsidy*, Policy Journals, Arrowsmith, Bristol.

Goodwin, P.B., Bailey, J.M., Brisbourne, R.H., Clarke, M.I., Donnison, J.R., Render, T.E. and Whiteley, G.K., 1983, *Subsidised Public Transport and the Demand for Travel*, Gower, Aldershot.

Gwilliam, K.M., 1987, "Market failures, subsidy and welfare maximization", in Glaister, S., edited by, *Transport Subsidy*, Policy Journals, Arrowsmith, Bristol.

Heseltine, P.M. and Silcock, D.T., 1990, "The effects of bus deregulation on costs", *Journal of Transport Economics and Policy*, 24, pp. 239-254.

Laffont, J.J. and Tirole, J., 1993, *A Theory of Incentives in Procurements and Regulation*, MIT Press, Cambridge.

Miller, J.H., 1980, "The use of performance-based methodologies for the allocation of transit operating funds", *Traffic Quarterly*, 34, pp. 555-585.

Nash, C.A., 1978, "Management objectives, fares and service levels in bus transportation", *Journal of Transport Economics and Policy*, 12, pp. 70-85.

Pedersen, P.A., 1994, "Regulating public transport company with private information about costs", *Journal of Transport Economics and Policy*, 28, pp.307-318.

Pickrell, D.H., 1985, "Rising deficits and the uses of transit subsidies in the United States", *Journal of Transport Economics and Policy*, 19, pp. 281-298.

Pucher, J., Markstedt, R. and Hirschman, I., 1983, "Impacts of subsidies on the costs of public transport", *Journal of Transport Economics and Policy*, 17, pp. 155-176.

Rees, R., 1984, "A positive theory of the public enterprise", in Marchand, Pestieau and Tulkens, eds., *The Performance of Public Enterpise: Concepts and Measurement*, North-Holland, Amsterdam.

Reichelstein, S., 1992, "Constructing incentives schemes for government contracts: an application of agency theory", *The Accounting Review*, 67, pp. 712-731.

Searle, G., 1987, "Value for money from rural public transport subsidies: a summary of the Lewes approach" in Glaister, S., ed., *Transport Subsidy*, Policy Journals, Arrowsmith, Bristol.

Spence, A.M., 1975, "Monopoly, quality and regulation", *Bell Journal of Economics*, 6, pp. 417-429.

Tirole, J, 1994, "The internal organization of government", *Oxford Economic Papers*, 46, pp. 1-29.

Tisato, P., 1992, "User cost minimization and transport subsidy", *Economic Letters*, 39, pp. 241-247.

Tulkens, H., 1986, "The performance approach in public enterprise economics", *Annals of Public and Cooperative Economics*, 57, pp. 429-443.

Turvey, R. and Mohring, H., 1975, "Optimal bus fares", *Journal of Transport Economics and Policy*, 9, pp. 280-286.