Productive Structure, Cost Efficiency and Incentives in the Local Public Transport: A Survey of Theoretical and Empirical Issues

Massimiliano PIACENZA



Working Paper 1, 2001

© Hermes Real Collegio Carlo Alberto Via Real Collegio, 30 10024 Moncalieri (To) 011 640 27 13 - 642 39 70 info@hermesricerche.it http://www.hermesricerche.it

I diritti di riproduzione, di memorizzazione e di adattamento totale o parziale con qualsiasi mezzo (compresi microfilm e copie fotostatiche) sono riservati.

PRESIDENTE

Giovanni Fraquelli

SEGRETARIO

Cristina Piai

SEGRETERIA OPERATIVA

Giovanni Biava

COMITATO DIRETTIVO

Giovanni Fraquelli (*Presidente*) Cristina Piai (*Segretario*) Guido Del Mese (ASSTRA) Carla Ferrari (Compagnia di San Paolo) Giancarlo Guiati (ATM S.p.A.) Mario Rey (Università di Torino)

COMITATO SCIENTIFICO

Tiziano Treu (*Presidente*, Università "Cattolica del Sacro Cuore" di Milano e Senato della Repubblica) Giuseppe Caia (Università di Bologna) Roberto Cavallo Perin (Università di Torino) Carlo Corona (CTM S.p.A.) Graziella Fornengo (Università di Torino) Giovanni Fraquelli (Università del Piemonte Orientale "A. Avogadro") Carlo Emanuele Gallo (Università di Torino) Giovanni Guerra (Politecnico di Torino) Marc Ivaldi (IDEI, Universitè des Sciences Sociales de Toulouse) Carla Marchese (Università del Piemonte Orientale "A. Avogadro") Luigi Prosperetti (Università di Milano "Bicocca") Alberto Romano (Università di Roma "La Sapienza") Paolo Tesauro (Università di Napoli "Federico" II)

Productive Structure, Cost Efficiency and Incentives in the Local Public Transport: A Survey of Theoretical and Empirical Issues *

MASSIMILIANO PIACENZA*

(Università di Roma "La Sapienza", Ceris-CNR, HERMES)

February 2001

Abstract

The paper reviews both the relevant economic theory and the econometric evidence analyzing the productive structure of Local Public Transport (LPT) industry. It highlights the main implications of the empirical studies concerning: i] better knowledge of the technology (cost 'function' approach); ii] evaluation of firms' x-efficiency (cost 'frontier' approach); and iii] possible determinants of the deviations from the theoretical goal of cost minimization. A particular emphasis is put on some recent contributions (Dalen and Gomez-Lobo, 1995 and 1997; Gagnepain and Ivaldi, 1998 and 1999) which, in the light of the theoretical predictions of the new regulatory economics (Laffont & Tirole, 1993), have elaborated "structural" models incorporating the role of institutional constraints and incentive regulation into the empirical analysis of cost structure. We discuss the usefulness of this new approach for studying the specific problems faced by the Italian LPT sector, especially with regards to the ongoing change in the subsidization mechanisms implemented at local level.

Key words: local public transport, technology, x-efficiency, regulation, stochastic cost frontier.

JEL: C1, D21, L51, L92, R41.

* This study has been carried out within the Ceris-CNR/HERMES research project *I trasporti locali: dal sussidio all'incentivazione e all'efficienza*. The paper comes from the dissertation prepared for the DEA «Economic Theory and Econometrics» at University "Toulouse 1" (October 2000); I would like to thank my supervisors, Farid Gasmi and Marc Ivaldi, and Giovanni Fraquelli for their helpful comments.

* Mailing address: Ceris-CNR (Institute for Economic Research on Firms and Growth), via Avogadro 8, 10121, Torino, Italy; HERMES (Higher Education and Research on Mobility Regulation and the Economics of Local Services), Real Collegio Carlo Alberto, via Real Collegio 30, 10024, Moncalieri (TO), Italy. E-mail: m.piacenza@ceris.to.cnr.it; phone: +39 (0)11 5601241; fax: +39 (0)11 5626058.

1. Basic characteristics of technology and cost functions for the transit service

For many reasons the study of the technology and factor supplies faced by transportation providers has long been of central interest in the transportation industry. Knowledge of costs functions, which usefully summarize technology and factor supplies, enables us to answer a great many questions about the relative efficiency of various types of transportation, and about the relative importance of various parts of the production process such as capital, user time, operator wages, public facilities, and even unintended spillovers to nonusers. Moreover, costing studies represent an important tool of analysis to assist regulatory agencies in decision-making, as they provide guide-insights into such public policy questions as optimal subsidization and the desirability of competition among transit operators.

Before discussing the evolution of transportation cost studies, it is useful to begin with a brief discussion of some essential concepts peculiar to the literature on transit service. The basic description of any technology is the *production function*, which represents the relation between outputs and inputs in the production process:

$$F(y, x; \beta) = 0, \tag{1}$$

where *y* and *x* are vectors of outputs and inputs, respectively, and β is a vector of parameters which may include service-quality descriptors.¹

The expression just given can be made operational only by simplifying the complex production processes encountered in real life. For transit service, it is useful to consider two classes of output. One consists of measures of trips taken. As the ultimate purpose of transportation activity is to provide trips, and it is trips, or some aggregates of them, that are the output variables in travel-demand analysis, we may term such output measures *final outputs*.² In practice, final outputs are usually aggregated into total passenger trips, revenue passengers (the number of distinct fares paid), passenger-kilometers, or total revenues (a valid output measure if the fare structure is held constant in the analysis).

From the point of view of the transport company, however, final outputs are not under its direct control; hence it may be more interested in the cost of producing the potential for trips, as measured, for example, by vehicle-kilometers, vehicle-hours, or seat-kilometers of service. One may consider such measures to be *intermediate outputs*, because they are combined with

¹ Alternatively, services of different quality may be considered as different outputs in the vector y.

user time to produce the final output.³ Intermediate output are sometimes bought and sold as intermediate goods, for example when a local authority (Region, Province or Council) contracts to pay a private firm for the provision of a particular amount and type of bus service on a particular route.

Whether one uses the final or intermediate outputs in the transportation cost studies can not be established a priori but it will depend upon the purpose of the analysis.⁴ Whatever the type of outputs considered, the *cost function* for a given producer represents the minimum cost of producing output vector y, given the production function and the supply relations for the inputs. Usually these supply relations are assumed to consist of a fixed price vector p, in which case the problem becomes minimizing p'x subject to the technological constraint [1]. The solution, if unique, determines an optimal input vector x^* that depends on y, p, and β , so the resulting minimum cost, p'x*, can be written as $C(y, p; \beta)$.⁵ If all inputs are included in x, including those that can be varied only over a long time period, C comes to describe a *longrun cost function*. If instead one or more inputs are held fixed during the minimization, the resulting cost is called a *short-run cost function*. Typically, the fixed input is a measure of capital stock, say K; its fixed value \overline{K} becomes another argument of the resulting cost function, which we may write as $\widetilde{C}(y, p; \beta, \overline{K})$. By definition,

$$C(y, p; \beta) = \underset{K}{Min} \widetilde{C}(y, p; \beta, K).$$
^[2]

Either the short- or the long-run cost function may approach a positive constant C° as output y approaches zero. If so, C° is called the *fixed cost*, and $(C - C^{\circ})$ the *variable cost*.⁶ A short-run cost function always contains a fixed cost because it includes the carrying cost of fixed capital $(p_k \overline{K} \text{ in the case of fixed factor price})$; the rest of the short-run cost is called *operating cost*, since it characterizes ongoing operations.

² Berechman (1993) call them 'demand-oriented measures'.

³ Berechman (1993) call them 'technical output measures', because they are most directly related to the firms' exploitation of transport technology.

⁴ For example, a study of the purely technical efficiency of firms' production would probably use intermediate outputs, whereas an analysis of the effectiveness of the firms' service offerings and marketing policies would use final outputs.

⁵ However, the supply relations can also be represented by factor-supply equations, in which case p must be reinterpreted in what follows as a vector of parameters fully describing those factor equations.

⁶ Fixed cost should not be confused with *sunk cost*, a dynamic concept which expresses an irreversibility to starting a business: for example, the marketing analysis and initial advertising campaign that might initiate a new entry into the market for transit service.

Letting *C* denote either a short- or long-run cost function, we may define average and marginal cost with respect to any one output, y_j , as $AC = C/y_j$ and $MC = \partial C/\partial y_j$.⁷ Interest in analyzing the production and cost structure of a particular industry often centers on the degree of *returns to scale*, which summarizes how fast costs rise with respect to output(s). If output *y* is a scalar, returns to scale is defined simply as the inverse of the output-elasticity of cost:

$$s = \frac{AC}{MC} = \frac{C}{y.(\partial C/\partial y)} .$$
[3]

If MC < AC so that s > I (equivalently, if AC is falling in y), we have *increasing returns*, also called *economies of scale*. The opposite case (s < I) is denoted *decreasing returns* or *diseconomies of scale*; and s = 0 defines *constant returns*.⁸ Because a short-run cost function has a larger fixed cost than the corresponding long-run cost function, it is more likely to show increasing returns. In the specific context of transportation industry it is especially important at this regard to retain the distinction between expanding the *density* of output, for example by adding more vehicles or attracting more passengers on a given route, and expanding the *spatial scale* of output, for example by adding new routes with similar densities. The former often allows more intense use of equipment, thereby lowering average cost; this form of increasing returns to scale is usually called *increasing returns to density* or *economies of density*, to distinguish it from the degree of returns to scale that characterizes an expansion of the entire productive dimension, denoted *increasing returns to size* or *economies of size*.⁹

1.1. The choice of functional form

In the previous discussion on the characteristics of the production and the costs in the transit service sector it has been neglected the choice of the more appropriated functional form to use during the empirical analysis of the aforesaid relations. Indeed this point is very crucial and deserves then to be treated separately. The criteria of choice are substantially two: the statistic

⁷ One can show from [2] that the long-run marginal cost is equal to the short-run marginal cost with K set to K^* ; this implies that if capital stock is optimal, the cost of producing a small increment of output is the same whether or not that capital stock is varied.

⁸ These relationships generalize easily to many outputs, by redefining *s* in equation [3] with the denominator interpreted as the inner product between vector y and the gradient of the cost function (see Bailey and Friedlaender, 1982).

⁹ In short, through the analysis of the density economies it is possible to check if cost advantages deriving from the more intensive use of a given transport network exist, while through the measurement of the economies of size one can evaluate the reduction of the unitary costs due to a greater firm dimension or, as an example, to the fusion of two companies and the respective networks.

parsimony, that is the number of the parameters to estimate, and the generality degree of the theoretical model, that is the number of restrictions imposed by the functional form to the technology. In the remainder of this section the attention will be concentrated in prevalence on the translogaritmic functional form, being this the cost structure model that has known the widest number of applications in the literature of transportation economics, as we will see more ahead.

As it is known, the *Cobb-Douglas* function, extremely thrifty from the point of view of the parameterization, is however very rigid on a theoretical plan, as it imposes a constant degree of returns to scale and elasticity of substitution between productive factors equal to 1. *CES* function, without to overcome the first limit, is more general for that regards the substitution elasticity, that can assumes values different from the unit. In both cases, however, the substitution elasticity is constant for every level of output and every combination of input. This limit and the previous one comes exceeded resorting to the so-called flexible functional forms, of which the *translogaritmic function* is surely that most famous one and used in the practice one. The translog specification for the cost function [2] is given by:

$$\ln C = \beta_{0} + \sum_{r} \beta_{r} \ln p_{r} + \sum_{j} \beta_{j} \ln y_{j} + \frac{1}{2} \sum_{j} \sum_{k} \beta_{jk} \ln y_{j} \ln y_{k}$$
$$+ \frac{1}{2} \sum_{r} \sum_{m} \beta_{rm} \ln p_{r} \ln p_{m} + \sum_{j} \sum_{r} \beta_{jr} \ln y_{j} \ln p_{r}, \qquad [4]$$

where the indices r, m = 1..., M denote the inputs employed in the production process.

Given the regularity conditions assuring duality, the estimation of a translog cost function does not impose any other a priori restriction on the characteristics of the below technology and on the features peculiar to the cost function.¹⁰ In particular, the elasticity of substitution and the returns to scale are variable regarding both the level of the output and the combination of the inputs. This fully satisfies the criterion of model generality, however at the cost of a remarkable increment in the parameterization, which can rise serious problems of statistical efficiency of the estimation interlaced with those due to multicolinearity of the regressors.

¹⁰ In order to be consistent with cost minimization, expression [4] must satisfy the following properties only:

⁻ *C* is non-negative;

⁻ *C* is homogeneous of degree 1 in vector *p*;

⁻ *C* is non-decreasing in *y*;

⁻ C is non-decreasing in p;

⁻ *C* is concave in *p*.

A typical solution to the efficiency problem consists in increasing the freedom degrees by jointly estimating with the *SUR* method (by *Zellner*¹¹) the cost function and the related *factor-share equations*, derived applying the Shephard lemma to the expression [4]:

$$\frac{\partial \ln C}{\partial \ln p_r} = \frac{\partial C}{\partial p_r} \frac{p_r}{C} = \frac{x_r^D p_r}{C} = \beta_r + \sum_m \beta_{rm} \ln p_m + \sum_j \beta_{jr} \ln y_j$$

$$\forall r = 1, \dots, M-1,$$
[5]

where x_r^D is the input demand (conditional to the output level) relative to the r^{th} productive factor. With regard to this estimation procedure, it is important to notice that while the factorshare equations are in total M, those estimated together with [4] are only M - 1. In fact, since these factor-shares add up 1 ("adding-up" condition), one would have a system with an equation linearly depending on the others, that from an econometric point of view would imply a singular covariance matrix of the of disturbance terms. To solve this singularity problem we have to drop an arbitrary equation and estimate the remaining factor-share equations by the SUR procedure.¹²

2. The econometric costing studies: 'function' versus 'frontier' approach

Traditionally, the contribution of empirical research to the analysis of the production and cost structure of the local transportation sector has focused on the primal and dual characteristics of the production set:¹³ evaluation of the degree of technical substitution between inputs, checking of the existence of size and density economies and computation of related measures, identification of the efficient dimension for firm and plant, estimation of inputs demand functions have been for a long time the typical purpose of these studies.

In relatively more recent years the topic of productive *efficiency* has been added to the agenda of research; therefore, in parallel to the previous literature rotating around the notion of production and cost *function*, another branch of productivity analysis corresponding to the study and the estimation of the so-called production and cost *frontier* has begun to develop.

¹¹ Zellner, A., "An efficient Method of Estimating Seemingly Unrelated Regressions (SUR) and Tests for Aggregation Bias", *Journal of the American Statistical Association*, 57, 1962, pages 348-368.

¹² It should be remarked that the parameter estimates are invariant to the choice of which equation is deleted as long as the *Iterated SUR* (or *Maximum Likelihood*) estimation method is employed on the M - 1 factor-share equations.

¹³ The set of all production vectors that constitute feasible plans for the analyzed industry.

The frontier concept rises in econometrics when one takes into account that the theoretical production and cost functions [1] and [2] are functions that represent the maximum and minimum values respectively of an optimization problem. More specifically, we have seen above that, in the simple case of a single-output technology, the cost function associates to the combinations of output quantity, y, and vector of input prices, p, the minimum expenditure required to produce such an output at these prices of inputs, $C = c(y, p) = min\{p'x: y = f(x)\}$, with y = f(x) representing the maximum output, y, possible for the input set, x, given the available technology. In that sense the notion of cost function may be interpreted as a frontier relationship, i.e., a benchmark behavior, because it is impossible for the firm to achieve costs lower than the minimum requirement.

The general econometric specification of a stochastic cost frontier is the following:

$$C_i = c(y_i, p_i; \beta) \exp\{v_i + u_i\},$$
[6]

where C_i denotes the observed cost incurred by firm *i*, y_i is the vector of output quantities produced by the *i*th firm, $p_i = (p_{il}, ..., p_{iM})$ is the vector of input $(x \in R^{M_+})$ prices facing firm *i*, β is the vector of parameters characterizing the common technology, and i = 1, ..., N is an index denoting the different producers. The error term v_i reflects the statistical variability of the sample (random deviations from minimal cost) and it is usually assumed to be independent and identically distributed as a normal with zero expected value and variance σ_v^2 , $\sim N(0, \sigma_v^2)$, while the component u_i , represents the deviation of firm *i* from potential minimal cost due to *inefficiencies*, that is the shifting of the *i*th observation from the efficient cost frontier after taking into account the effect of statistical variability ($c(y_i, p_i; \beta) \exp\{v_i\}$).

Two considerations should be made with regard to this point. First, while the random disturbances v_i may be positive, zero, or negative, the disturbance due to inefficiency is always non-negative ($u_i \ge 0$) since inefficiency cannot cause cost to be less than the frontier level. Second, as the cost inefficiency may be due to both *technical inefficiency*, i.e., the inability of producers to use the minimal level of inputs given output and the input mix, and *allocative inefficiency*, i.e., the failure in making the marginal rate of substitution between any two inputs equal to the corresponding input price ratio (in other words, the firm is not using its inputs in the optimal proportions), the term u_i includes two components of inefficiency. If, on one hand, the ability to estimate both technical and allocative efficiency is just one of the advantages of a cost frontier with respect to a production frontier, on the other hand, the decomposition of cost inefficiency into its two components involves many problems, whose

solution is still object of research.¹⁴ A part from this last aspect, the *overall cost efficiency* of the i^{th} firm is usually measured by indices like that given by the ratio between potentially supportable cost and actually supported cost:¹⁵

$$CE_{i} = \frac{[c(y_{i}, p_{i}; \beta) \exp\{v_{i}\}]}{C_{i}} = \frac{1}{\exp\{u_{i}\}},$$
[7]

whose values range from 0 (for $u_i \rightarrow +\infty$ or large enough) to 1 (for $u_i = 0$).

2.1. Traditional 'cost function' analysis

In almost all of the econometrics applications the functional form of the cost model to estimate is log transformed in order to obtain a linear relationship in the logarithms of output and the other independent variables. Without loss of generality, we can then rewrite the expression for the stochastic cost frontier [6] as:

$$\ln C_{i} = \ln c(y_{i}, p_{i}; \beta) + v_{i} + u_{i}.$$
[8]

The expression above shows that observed (log) cost, $\ln C_i$, is the sum of frontier (log) cost, $\ln c(y_i, p_i; \beta)$, random deviations from minimal cost, v_i , and deviations from minimal cost due to inefficiencies, u_i . Now, if one assumes $u_i = 0$, the problem of estimating expression [8] is reduced to that of the estimation of the parameters characterizing a cost function devoid of inefficiency. This is equivalent to accepting the hypothesis that, except for stochastic disturbances, all of the observations belong to the "*best-practice*" frontier. This statement, certainly recurrent in the traditional empirical literature dealing with the cost structure analysis, obviously suffers from a conceptual weakness: it is not plausible, because it ignores the problem of cost inefficiency; consequently, the obtained estimate can not be considered a frontier but only the interpolation of the observed firms performances or, in other words, an *average cost relationship*.

On the other hand, the primary aim of the studies that utilize this approach typically is to estimate the parameters β of the cost function, or, by employing the results of dual theory, the characteristics of the production process. The problem becomes, therefore, to ascertain if the estimation of [8] given the hypothesis $u_i = 0$ (from now [8a]) suffers from having neglected

¹⁴ See Bauer (1990) and Greene (1993) for a summary of the proposed solutions.

¹⁵ According to a proposal due to Farrell (1957). Farrell was in fact the first to treat empirically the frontier efficiency comparisons hence efficiency studies have become synonymous with Farrell efficiency measurement.

the role of inefficiency; one can demonstrate that, under certain assumptions pertaining to the distribution of the error term $\varepsilon_i = v_i + u_i$, the traditional least squares estimation of [8a] (OLS technique) does not suffer from problems due to a bad specification of the model and provides consistent estimates of the parameters β .

2.1.1. Review of transportation literature

Premised that in this brief review we will only discuss contributions that have used so-called flexible functional forms,¹⁶ one can easily find that the empirical studies addressed to the analysis of the average characteristics of technology in the local public transportation (from now on LPT) industry by using the cost function approach, are very numerous and homogenous in their formulation. Indeed, in the overwhelming majority of the cases it is resorted to a SUR estimation on cross-sectional data of the system consisting of the (variable or total) cost equation [4] and the M - 1 factor-share equations [5]. The global picture of the works passed in review is illustrated in Table 1; in the continuation we will not stop on everyone of them but reference to some summarizing general result will be made.

As it can be noticed, none of the listed studies concerns the analysis of the Italian LPT sector. In fact, it can be said with certainty that, at least until the more recent years, in Italy the studies on the performance of local transportation companies more directly inspired to the microeconomic analysis and theory of production have been late to emerge and to become popular. Therefore, it is not a case if the literature which try to provide a global picture of the problems that characterize the Italian local transport by using the econometric approach to the cost analysis is relatively poor. Moreover, the very few articles published at this regard in recent years seem mainly oriented to study problems of productive inefficiency, particularly relevant for the Italian LPT sector,¹⁷ rather than to simply analyze the average characteristics of costs and technology,¹⁸ for this reason, it has been thought more convenient for the expositive purpose to postpone the presentation of these works to the following section, where it will be discussed the role of efficiency in the within of the study of cost structures and the methods used by the empirical literature to analyze such aspect.

¹⁶ They remain then outside this review the studies on LPT sector that used linear and log-linear specifications and the Cobb-Douglas form, for a discussion of which it is referred to Berechman (1993).

¹⁷ In this regard see European Commission (1998) and CER (1997), where the performance of the Italian LPT companies in relation to their operating costs and exhibited levels of productivity is analyzed.

¹⁸ The only contribution in Italy exclusively addressed to the study of the technological characteristics of the local transportation sector is that one of Petretto and Viviani (1984), in which, moreover, it is used like analysis instrument the production function rather than that of cost one.

Authors	Type of model	Data	LPT Sample	Output
Berechman, 1987	total cost + factor-shares	quarterly time series 1972-1981	urban and Extra-urban transport, Israel	- vehicle-kilometers - passenger-trips
Berechman and Giuliano, 1984	total cost + factor-shares	cross-section	urban transport, USA	- vehicle-kilometers - passenger-trips
Button and O'Donnel, 1985	tot. and var. cost + factor-shares	cross-section 1979	urban and extra-urban transport in 44 counties, UK	- revenues - served users
Caves and Christensen, 1988	tot. and var. cost + factor-shares	cross-section	urban transport, USA	- passenger-kilometers - bus-kilometers - network
De Borger, 1984	variable cost + factor-shares	yearly time series 1951-1979	regional transport company (NMVB), Belgium	- seat-kilometers
Filippini, Maggi and Prioni, 1992	total cost + factor-shares	panel 1986-1989	62 extra-urban transport companies, Switzerland	 passenger-kilometers seat-kilometers network
Gagnepain, 1998	variable cost + factor-shares	panel 1985-1993	60 urban transport companies, France	- vehicle-kilometers - network
Obeng, 1984	variable cost + factor-shares	cross-section	urban transport, USA	- passenger-kilometers
Thiry and Lawarree, 1987	variable cost + factor-shares	panel 1962-1986	5 urban transport companies, Belgium	- seat-kilometers
Viton, 1981	variable cost + factor-shares	cross-section 1975	54 urban transport companies, USA	- traveled kilometers
Windle, 1988	tot. and var. cost + factor-shares	cross-section 1978	91 urban transport companies, USA	- passenger-kilometers
Williams and Dalal, 1981	total cost + factor-shares	cross-section 1976	20 urban transport companies, Illinois	- vehicle-kilometers

Table 1. Econometric studies of the LPT industry based on the cost function approach

In order to account for the results obtained in the studies cited in Table 1, it is expedient to subdivide them into two groups: the results pertinent to the analysis of the *cost elasticity* and those derived in the context of the study of the *substitution elasticity between factors*. The picture relating to the results on the scale economies can be summarized as follows:

all the considered studies confirm the presence of *short-run economies of size*. This seems to reveal the existence of unused capacity ascribable to two circumstances that are very

relevant for the economy of the transportation companies: the massive public contribution to the investments and the importance of the so-called *peak-load problem*;¹⁹

- the evidence on the presence of *long-run economies of size* is uncertain. With regard to this aspect it seems that the type of employed sample and the way of computing the capital price are crucial elements in orienting the results. In particular, in the studies on the extra-urban transport systems it emerges the presence of remarkable economies of size which decrease with the growing of firm dimension;
- the existence of *economies of network density* is confirmed by many of the analyzed contributions, that indicate a unitary production cost for the transportation companies decreasing at the growing of the output, given the dimension of the served network;
- the sector appears finally characterized by the presence of significant *economies of use intensity*²⁰ that, in this case, they reveal the existence of capacity excess regarding the intermediate output (potential for trips) actually consumed.

As regards the analysis of the results concerning the elasticity of substitution between the productive factors (usually identified with fuel, labor, capital and maintenance), it emerges that:

- considered the very small values of the substitution elasticity, the production technology can be substantially defined as a fixed-coefficients technology;
- labor and capital turn out to be complementary inputs;
- labor and fuel are instead substitutes, even if the substitutability degree appears to be very low;
- between capital and maintenance too there is substitutability, more marked than in the previous case.

Moreover, on the basis of substitution elasticity one can directly estimate the values of the own- and cross- price elasticity of the input demands. The evidence would indicate a demand for the productive factors that is substantially inelastic to own price and very low values for the cross-elasticity.

¹⁹ With such term reference is usually made to the problem for which the maintenance of the capacity necessary to satisfy the peak demand unavoidably creates unused capacity in the low demand phases.

²⁰ This is another concept of density economies which is very recurrent in the transportation literature that uses the final output (e.g. passenger-kilometers) oriented specification of the production function. With it one means the reduction of unitary per passenger cost deriving from the increase of served users on a given transit system.

2.2. The role of efficiency: 'cost frontier' analysis

In the light of our discussion in Section 2.1, the empirical literature on the cost analysis can mainly be divided in two principal branches of research: the studies of the average characteristics of costs and technology, which still constitute the prevailing stream, and the studies more directly concerned with the efficiency issue, i.e., the cost frontier approach. This latter literature, which has refined the analysis of productive and cost structures and has been experiencing a rapid growth in recent years, has been displaying its usefulness especially when estimating cost functions for regulatory purposes, e.g., for the fixation of subsidy levels for the firms supplying local transit services. In fact, regulators should employ guidelines based on the most efficient observations; therefore they should establish cost-technology estimates based on the best observed cost-output experience rather than the average cost relationship that ordinary regression techniques yield, since the use of this latter as assistance instrument in decision-making would neglect altogether the possibility to detect and control the differing productive efficiency across an industry.

The following two subsections briefly discuss the different techniques usually employed by the literature to estimate the stochastic frontiers according to the type of available sample, cross-section or panel. Finally, in the third subsection a concise overview of cost frontiers studies carried out for the Italian LPT sector will be provided; some concluding remarks on the obtained results will also be given, emphasizing the necessity to extend the analysis of inefficiency aspects so as to take into account the effects exerted by the institutional and regulatory constraints on the behavior of transportation operators.

2.2.1. Techniques used in cross-sectional setting

The earliest literature on stochastic frontier models considered the case of a single crosssection, in which the time dimension of data is altogether ignored. In this circumstance the estimation of the complete model of stochastic cost frontier [8] hinges on distributional assumptions for the two residual terms, v_i and u_i . Since in most cases the whole point of the frontier estimation exercise is to compare productive efficiencies at the firm level, such assumptions are required in order to estimate the firm-specific inefficiency component u_i with cross-sectional data, by extracting it from the composed error term $\varepsilon_i = v_i + u_i$. The Maximum Likelihood (ML) method is the usual way estimation proceeds, although a simpler corrected OLS (COLS) estimator is also available.²¹

In the ML estimation, introduced by Aigner, Lovell and Schmidt (1977) and Meeusen and Van den Broeck (1977), one proceeds by specifying the likelihood function for the model [8] according to the assumption formulated about the distribution of the one-side disturbance u_i . Indeed, while the noise component v_i is essentially always assumed to be ~ $N(0, \sigma_v^2)$, there are several possibilities as regards the specification of the inefficiency term, which is usually assumed to follow a *truncated normal*, *half-normal*, *exponential* or *gamma* distribution,²² anyhow holding the strong assumption that it is independent of the random deviation term and the other regressors. In all these cases it is possible to derive the distribution of the composed error term ε_i and go back from this up to the likelihood function of the log-linear model [8]. The score functions allow then to derive the appropriate expressions for the numerical computation of β estimator and the variance of composed error term, σ_{ε}^2 .

Whit regards to the decomposition of the global error term ε_i into noise (v_i) and inefficiency component (u_i) , this separation can be made following the procedure suggested by Jondrow, Lovell, Materov and Schmidt (1982). They derive the distribution of u_i conditional on $(u_i + v_i)$; in this way, known for each observation the estimate of global error term, $\hat{\varepsilon}_i$, one can compute for each observation the expected value of the inefficiency term conditional on ε_i , $E(u_i|\varepsilon_i)$, evaluated at $\hat{\varepsilon}_i$ and consistent estimates of the technological parameters. This clearly allows to evaluate the component of cost inefficiency characterizing the performance of each firm in the analyzed sample. However, as a concluding remark, it should be emphasized that the inefficiency terms estimates in such a way obtained, although unbiased, are not consistent,²³ and there is no alternative consistent estimator of firm-level

²¹ This latter approach has first been proposed by Greene (1980) for the estimation of a deterministic frontier (i.e., a frontier without random disturbance, that implies $v_i = 0$). In this case the method to obtain such an efficiency measure as [7] consists in adjusting the constant term of the OLS estimated cost function in a way which allows all observations to lie above or on the frontier and at least one observation lies on the frontier. This is done by adjusting the constant term using the negative OLS residual with the highest absolute value. In the stochastic frontier case, the OLS estimate of the intercept is corrected by a consistent estimate of $E(u_i)$, which is identified through the higher-order moments of the OLS residuals. ML might be preferred on the grounds that it produces more efficient estimate of the vector of technological parameters β . However, Monte-Carlo results by Olsen, Schmidt and Waldman (1982) suggest that ML and COLS perform roughly equally well in the case of an half-normal model, which is the most common assumption about the distribution of the inefficiency component.

²² It should be remarked that the choice of distribution, especially for the u_i , is not innocuous. Evidence suggests indeed that frontier estimates are not robust to changes in distributional assumptions. See, for example, Schmidt and Lin (1984).

²³ Due to the fact that the variation associated with the distribution of u_i conditional on $(v_i + u_i)$ is independent of N, the number of analyzed productive units.

inefficiency when using cross-sectional data. This aspect, obviously, does not give evidence in favor of the estimation techniques just discussed.

2.2.2. Panel data models

As an introduction to this subsection, an important remark needs to be made. When the main goal of the research is the analysis of cost efficiency rather than the study of primal (and dual) characteristics of the production process, no entirely satisfactory estimation technique seems to exist in a cross-sectional context, such as that considered so far in the formulation [8] of a stochastic cost frontier, which altogether neglects the time dimension of data.²⁴ Nevertheless, if *longitudinal* or *panel* data sets containing observations on a large group of firms, each observed at several points in time, are available, it is possible to estimate stochastic frontiers by overcoming, or at least reducing, the most pressing limits and restrictions typically involved in the estimation of the various *cross-section models*. Those include:

- the assumption that the inefficiency level of firm *i*, *u_i*, is not correlated with the other regressors (in particular the levels of output in a cost model and the levels of inputs in a production model), despite the fact that firms may know something about their level of inefficiency and this may affect their choices of output/inputs;
- the statistical restrictions associated with the unobserved variables (specific distributional assumptions for both the statistical noise and for the inefficiency term are required to estimate stochastic frontier models and to separate the firm inefficiency from the random disturbance);
- the necessity to resort to estimators of firm-specific inefficiency that are not consistent.²⁵

In principle, the traditional methods proposed for the estimation of models of several time series,²⁶ such as, for example, Feasible Generalized Least Squares (FGLS) or ML (which ones to use in a given situation depends on the setting), can also be applied to longitudinal data sets.²⁷ However, in the typical panel, there is a large number of cross-sectional units and only a few periods. Thus, the above mentioned time-series methods may be somewhat problematic

²⁴ For a more detailed treatment of the problems involved in the cross-sectional estimation of production and cost frontiers see Cornwell and Schmidt (1996).

²⁵ Schmidt and Sickles (1984) were the first to elaborate systematically on the link between the frontier and panel data literatures and on the specific advantages of panel data to estimate firm inefficiency.

²⁶ In which the number of cross-sectional units is relatively small and the number of time periods is (potentially) relatively large.

²⁷ One usually refers to these regression techniques as *pooled* estimates.

and recent work²⁸ has generally concentrated on techniques better suited to these short and wide data sets.

The proposed models have mostly focused on cross-sectional variation, or heterogeneity, of the panel observations. In the remainder of this section, we shall examine two of the most widely used techniques, the so-called "variable intercept models". The basic framework for the discussion is a stochastic cost frontier represented through the simple regression model:²⁹

$$\ln C_{it} = \alpha + \beta_1 \ln y_{it} + \beta_2 p_{it} + v_{it} + u_{it}$$

$$i = 1, ..., N; \quad t = 1, ..., T,$$
[9]

where the component u_{it} represents the cost inefficiency faced by the i^{th} firm in period t. A fairly usual specification for this term is to consider it as consisting of two elements, a timeinvariant *firm-specific* component, u_i , and a firm-invariant *period-specific* term, r_i .³⁰ Currently, two different techniques are available for the estimation of model [9], according to the assumptions on the nature of these components and their relationship with other regressors: the *fixed-effects* models and the *random-effects* models.

The *fixed-effects* approach, which is usually referred to as the *least squares dummy* variable model (LSDV), boils down to the introduction of dummy variables capturing both the effects on the costs due to omitted variables that are firm-specific but time-invariant and the effects on the cost that are period-specific but invariant across the cross-sectional units. Then, we may write [9] as:

$$\ln C_{it} = \alpha + \alpha_i + \gamma_t + \beta_1 \ln y_{it} + \beta_2 p_{it} + v_{it}, \qquad [10]$$

where α_i is the unknown coefficient related to *firm-specific* dummy variables, γ_t is the parameter associated with *period-specific* dummy variables, while the term v_{it} includes the effects of omitted variables that are peculiar to both the firm and the period. If one assumes that the v_{it} are i.i.d. $(0, \sigma_v^2)$ and strictly exogenous with respect to other regressors,³¹ without making any particular distributional assumption neither for the v_{it} nor for the u_i (this last, moreover, is allowed to be correlated (or not) with other regressors and the v_{ii}), the model [10] can be estimated through a modified ordinary least squares estimator.³² Then, a measure of

²⁸ See Hsiao (1986).

²⁹ For simplicity of notation we use a simple model of single output-single input technology.

³⁰ It should be noted that recent work has shown that the assumption of time-invariance for the firm-specific component can be relaxed, without losing the other advantages of panel data. Models with time-varying inefficiency levels are considered in Cornwell and Schmidt (1996).

³¹ The "strictly exogeneity" assumption imply that v_{it} are uncorrelated with y_{it} and p_{it} for all *i* and *t*. ³² See Hsiao (1986), cap. 3, for the definition of "*within-units*" estimator.

inefficiency for each firm can be constructed from the obtained estimates by using the following index proposed by Green:³³

$$\hat{u}_i = \hat{\alpha}_i - \hat{\alpha}_m$$
, where $\hat{\alpha}_m = \min_{i=1...N} \{ \hat{\alpha}_i \}$ [11]

By construction the levels of firm-specific inefficiency are estimated such that the frontier is normalized in terms of the best firm in the sample. Thus the most efficient firm is defined as 100 percent efficient, i.e., $\hat{u}_m = 0$, while the remaining firms' measures of cost inefficiency assume positive values. Obviously, the explanation of omitted variables' effects in terms of firm inefficiency leads to interpret the index [11] as a measure of an inefficiency concept with a wide meaning, embracing all the circumstances that prevented the firm from minimizing the production cost. Likewise, the γ_t estimates include a variety of *shift* factors, imputable to technical progress as well as, for example, to the evolution of competition and regulation context in which the firms operate.

The fixed-effects model is a reasonable approach when we can be confident that the differences between units can be viewed as parametric shifts of the regression function. In other settings, it might be more appropriate to resort to the so-called *random-effects* models that assume firm-specific and period-specific terms to be randomly distributed across cross-sectional and temporal units respectively. In particular, this would be appropriate if one believed that sampled cross-sectional units were drawn from a large population. In the case of orthogonality between specific effects and other regressors, the textbook estimator for this type of models is the generalized least squares (GLS) or the feasible analog of GLS³⁴ when the variance components are unknown.³⁵

In the ambit of this model a simple way to estimate the inefficiency component of firm i's costs consists in computing the average of estimated residuals for this firm:

$$\hat{\alpha}_{i} = \frac{1}{T} \sum_{t=1}^{T} \left(\ln C_{it} - \hat{\alpha} - \hat{\beta}_{1} \ln y_{it} - \hat{\beta}_{2} \ln p_{it} \right).$$
[12]

³³ See Green (1993).

³⁴ If the regressors are not orthogonal to the effects (the null hypothesis of strict exogeneity can be tested through the *Hausman* (1978) test applied to the difference between the "within-units" and FGLS estimates), one can resorts to an alternative Instrumental Variable (IV)-based estimation procedure, which assumes that some, but not all, of the regressors still may be uncorrelated with the effects (a number large enough for the identification of the parameters of the model) and provides estimates that are consistent and efficient relative to the within estimator.

³⁵ In practice, this is always the case.

The more important merit of such an approach to cost frontier analysis, with respect to the fixed-effects type, is the possibility to include in the model time-invariant explanatory variables too.

So far the emphasis has been on the ways in which the use of panel data-based techniques for the estimation of firm inefficiency allows to make weaker assumptions than are necessary in the cross-sectional setting. Before concluding this subsection, however, it is worthwhile mentioning a more general advantage of using panel data samples in the context of frontier analysis. Indeed, also by proceeding with essentially the same econometric assumptions and estimation procedures illustrated for the cross-section case, the availability of panel rather than cross-sectional data sets still shows an undoubted advantage, i.e., the fact that the repetition for several periods of the observation related to the same firm *i* makes it possible to estimate its level of efficiency more precisely.³⁶

2.2.3. Main results obtained for the Italian LPT sector

In this section it will be briefly reviewed the few studies carried out in relatively recent years to improve the knowledge about the economic aspects of Italian LPT industry. This contributions find their justification in the light of the situation of stagnant crisis in which the sector appeared to be immersed for several decades, and that claimed the swift start of such a radical reform as that one introduced by the legislator in the triennium 1995-1997.³⁷ In particular, many efficiency descriptive indicators have been giving evidence of an improper use of resources within the sector, by revealing a bad trend of costs and productivity with respect to that one showed by other European realities, characterized by sharp increases in the operating costs and remarkably low levels of labor productivity.³⁸ In spite of the relevance of this problem, the literature in Italy, a part from very few exceptions,³⁹ has never attempted to go deep into the economic analysis of these aspects. More specifically, it had never been carried out any rigorous study - cost frontier based - of the cost structure of Italian LPT industry before the contributions reviewed below, while these type of works represent a fundamental step in the analysis oriented to detect the actual presence of waste of resources and consequently to prepare a regulatory policy for the sector aiming at a recovery of productive efficiency. Hence the scope of the econometric studies of Italian LPT industry,

³⁶ For more details on this aspect see Cornwell and Schmidt (1996).

³⁷ A detailed discussion of the Italian LPT Reform is provided in CER (1997), Gorla (1999) and Federtrasporto (1999).

³⁸ As for these aspects, it is referred to CER (1997) and European Commission (1998).

³⁹ For a review see Cispel (1989).

whose main characteristics (in terms of type of cost model, data structure, LPT sample and output measure) are summarized in Table 2.

Authors	Type of cost model	Data	LPT sample	Output
Fabbri, 1998	variable cost function (translog form)	panel 1986-1994	9 urban and extra-urban bus companies, Region Emilia Romagna	traveled kilometers
Fazioli, Filippini and Prioni, 1993	total cost function (translog form)	panel 1986-1990	40 extra-urban bus companies, Region Emilia Romagna	seat-kilometers
Levaggi, 1994	variable cost function (translog form)	cross-section 1989	55 urban bus companies, Italy	passenger-kilometers

Table 2. Econometric studies of the Italian LPT sector based on the cost frontier approach

As it can be noticed, all listed contributions adopted the translogarithmic form in the specification of the cost model for the reasons highlighted in Section 1.1, and focused on the bus service, being this last the dominant compartment on the whole Italian LPT sector.⁴⁰ Only one of these studies (Fazioli et al., 1993) chose to analyze the productive structure in terms of total costs, while the other two (Fabbri, 1998 and Levaggi, 1994) considered a variable cost model more appropriate to study the efficiency of public transit systems in Italy, given the strict dependence of capital purchases from the government grants-in-aid program (providing the LPT companies with funds for investments), which suggested to treat the capital stock ⁴¹ as fixed in the short run.⁴² As regards the sample of LPT companies considered in the analysis, while Fazioli et al. and Levaggi focused only on the extra-urban and urban transport respectively, Fabbri analyzed both compartments; moreover, Levaggi is the only contribution to consider a large sample of companies operating all over Italy, with the analysis by Fabbri and Fazioli et al. exclusively focusing on the firms providing the local transport service in a region of North-Italy (Emilia Romagna). Finally, the three studies differ also in the index

⁴⁰ Indeed, this transit mode accounts for over the 80 per cent of LPT services in terms of supplied seatkilometers.

 ⁴¹ Defined as the number of buses in operation owned by a company in Fazioli et al. (1993) and Levaggi (1994), and the average number of buses owned by a company weighted by the average age of the buses in Fabbri (1998).
 ⁴² More precisely, Levaggi (1994) underlines (page 71) that the main consequence of the grants-in-aid program

⁴² More precisely, Levaggi (1994) underlines (page 71) that the main consequence of the grants-in-aid program providing funds to purchase capital has been that «the price of capital the firms face is much lower than its actual price. If this is the case, the firms are no longer minimizing costs with respect to all inputs in the short run, rather they minimize costs with respect to the variable inputs». She also specifies (page 86) that «since the subsidies received from Central Government need not be repaid, the price for the capital is basically nil. However, we can consider that the capital has at least a shadow cost in terms of the other productive factors that could be bought using the subsidy».

employed to measure the output produced by LPT companies: total kilometers and seatkilometers (both "supply-oriented" measures) in Fabbri and Fazioli et al. respectively, passenger-kilometers ("demand-oriented" measure) in Levaggi.

Authors	LPT sample	Output	Economies of scale (mean point values in parenthesis)
Fabbri (1998)	9 urban and extra-urban bus companies, Region Emilia Romagna, 1986-1994	traveled kms	 high economies of size in both the short (1.66) and the long run (1.71); importance of size economies decrease with increasing firm dimension.
Fazioli, Filippini and Prioni (1993)	40 extra-urban bus companies, Region Emilia Romagna, 1986-1990	seat-kms	 high economies of size (1.70) and network density (2.61); importance of size and network density economies decrease with increasing company dimension.
Levaggi (1994)	55 urban bus companies, Italy, 1989	passenger-kms	 very high economies of use intensity in both the short (8.29) and the long run (5.40); relevant size economies in the short run (1.43) but weak size diseconomies in the long run (0.92); relevant network density economies in the short run (1.38) but weak network density diseconomies in the long run (0.89).

Table 3. Results on the *scale economies* in the econometric studies of the Italian LPT sector based on the *cost frontier* approach

In all of the contributions the results derived from cost analysis allow for a discussion of the traditional *scale inefficiency*⁴³, in terms of sub-optimal size and density of transportation system, whose relevance for the study of the industry structure and behavior has long been recognized in the economic literature. More importantly for our goals, the estimation of a frontier cost function makes also possible to discuss the results in terms of *overall cost inefficiency*; this type of analysis, through the comparison of the deviations of different firms from the best-practice cost frontier, should assist in expressing evaluations about the present

⁴³ Scale efficiency indicate the degree to which a company is producing at optimal scale. Frisch (1965) defines the optimal scale as the level of operation where the scale elasticity is equal to one. In the present context of the local transportation sector (network industry), economies of size and density are distinguished.

regulation policy and proposals of reforms eventually needed, having as a guideline the performance of the most efficient observations. For completeness of analysis both aforesaid aspects will be discussed in the remainder of this section.

The results on the economies of scale are briefly summarized in Table 3. Because of the differences illustrated above, it is difficult to fully compare the scale elasticity indicators obtained in the different studies and derive general implications in terms of changes to bring in the structure and behavior of LPT industry; therefore, we will just list the main common findings, by highlighting each time some important suggestions expressed in the single contributions. The picture, very similar to that one emerged for the studies carried out in other countries,⁴⁴ can be summarized in the following points:

- both studies using a variable cost model reveal the existence of *short-run economies of size*. This indicates that the analyzed companies were operating at an inefficient productive size which under utilized the capital stock available. Levaggi argues that the cause for this inefficient use of the capital could derive from government intervention, having the negative consequences on size efficiency of capital subsidization also been confirmed by studies for U.S.A. urban transport;⁴⁵
- as regards the presence of *long-run economies of size*, the evidence is uncertain, seeming crucially to depend on the index employed to represent the output and on the type of sample. In particular, the studies using "supply-oriented" measures of output (Fabbri and Fazioli et al.) and focusing on LPT firms operating in Region Emilia Romagna reveal the existence of significant economies of size, whose importance decrease with increasing company size; this led the authors to deduce that these firms were globally sub-dimensioned with respect to the long run equilibrium;
- also the evidence about the existence of *economies of network density* is not entirely conclusive, again depending on the measure adopted for the output and on the analyzed sample. Indeed, while in Fazioli et al. remarkable increasing returns to network density are observed at all data point, even if with a relevance weakly decreasing with firm size,⁴⁶ Levaggi found evidence of the presence of this type of returns only in the short run. On the basis of their results, Fazioli et al argued that some regional bus company in Italy operated at an appropriately low density level, which claimed that output increases

⁴⁴ See the above discussion concerning the transportation literature based on the "cost function" approach, Section 2.1.

⁴⁵ See Windle (1988).

⁴⁶ The range between 2,64 for the small companies to 2,47 for the larger companies.

through higher frequency of bus services on the existing network as well as through denser seating were put forward;⁴⁷

- finally, from the study carried out by Levaggi⁴⁸ emerges a bus urban transport industry in Italy characterized by a very high degree of *economies of use intensity*, both in the short and the long run. The substantial returns to scale for load can be interpreted in terms of considerable excess capacity regarding the potential for trips actually exploited by the passengers. This would indicate that bus companies in Italy have been facing insufficient levels of demand,⁴⁹ perhaps due to the concurrence of other vectors (tramways, subways, private cars and so on), and poses the relevant issue of partially reducing the spare capacity of buses⁵⁰ and developing in an inter-modal way the whole LPT sector, in order to make complementary the different transit vectors and improve the use of the resources within the bus-lines compartment.

As regards the analysis of the degree of overall cost inefficiency showed by the various companies, i.e., the extent to which they exhibit a productive behavior not consistent with the cost minimization, the results obtained in the three studies cannot be quantitatively compared, due to the different methods used to estimate the cost frontier and consequently to evaluate the inefficiency component specific to each firm (see Table 4).

However, it is possible to catch at a qualitative level a general finding common in all these works: there is a substantial degree of cost inefficiency through the companies crosssection due to a misuse of the inputs in the production process. This econometric evidence seems to confirm the considerations put forward by many observers of the sector about the unsatisfactory dynamics of costs and productivity of pubic transit systems in Italy, and the

⁴⁷ With respect to the intensification of the operations the authors point out that in reality, the demand met by the analyzed bus companies is often very limited and hence an intensification strategy might not be the best option to exploit the existing network density economies, whereas their finding clearly indicates a potential for a merger policy. In fact, without giving a detailed description of the structure of extra-urban transport sector in Emilia Romagna, Fazioli et al. underline the presence in their sample of several constellations of small bus companies operating through a given network with limited levels of activities and without any coordination between them, in which case mergers between small and medium companies would be feasible and desirable.

⁴⁸ This is the only contribution to use a "demand-oriented" output measure, i.e., passenger-kilometers, and therefore the only study that allow to derive an index for testing the presence of economies of use intensity, being these last related to the increase of served passengers density (i.e., the passenger-kilometers over number of seats available) on a given transport system.

⁴⁹ As highlighted in CNR (1999).

⁵⁰ However, Levaggi points out that this possibility need to be carefully examined before being put in practice. Load factor is in fact a measure of congestion on the bus; the higher load factor, the lower the probability that a person can get a seat at any given time. An increase in the load factor could then imply a reduction in the quality of the service offered. For example, if the load factor was improved through a reduction of the number of buses operating on secondary routes, the probability for a passenger to get a regular service could be seriously impaired and this would be unacceptable. As underlined by the same author (page 86), «public transport is still considered a good whose production possesses the characteristic of universal service».

existence of a scope for increasing the productivity of the operations of buses by means of a new regulatory policy aiming to improve the x-efficiency.

Authors	Data	Cost frontier estimation	Inefficiency level estimation
Fabbri, 1998	panel: - years 1986-1994 - 9 companies	fixed-effects (LSDV) model	Greene (1993) efficiency-index
Fazioli, Filippini and Prioni, 1993	panel: - years 1986-1990 - 40 companies	corrected OLS (COLS) estimator	Greene (1980) efficiency-index
Levaggi, 1994	cross-section: - year 1989 - 55 companies	maximum likelihood (ML) estimator	Jondrow, Lovell, Materov and Schmidt (1982) procedure

Table 4. Methods used in the econometric studies of the Italian LPT sector to estimate cost frontiers and *inefficiency* components

The main failing attributable to all these studies concerns the way by which they try to identify the major causes for the cost inefficiency detected through the frontier estimation. In the contribution of Fabbri the inefficiency estimates are generally ascribed to all those factors, firm specific or attributable to the context in which the companies operate, that drew the firm away from the goal of cost minimization. The author also specify that the scope for an improvement of the cost performance in terms of efficiency recovery seems to be greater for the large sized companies,⁵¹ but nowhere near he tries to go into a deeper analysis of the reasons underlying the productive inefficiency shown by LPT firms.⁵²

Levaggi and Fazioli et al. must be given credit for having pushed the discussion on the inefficiency aspects a little more ahead than Fabbri. The former argues that the way the government had been subsidizing the LPT companies was not working in terms of incentives to productive efficiency and a new type of intervention had to be envisaged, in which the amount of the subsidy to be given to each firm should be set in advance and based on the standard cost of supplying a service of average quality. In the latter, by referring to Contestability theorists (Baumol, Panzar and Willig, 1982), the necessity to resort to a new

⁵¹ At this regard it should be noted the opposite direction of this result with respect to the evidence on the size economies (whose importance decrease with increasing company size) discussed above.

⁵² He only mentions the relevance within the general context of public-owned firms (included therefore transport operators) of some political conditions that, by inducing a preference for certain types of inputs, should add another constraint besides the usual one of the available technology to the behavior of cost minimization.

policy regulation aiming to redesign the aspects of accessibility to the network is emphasized; in particular, the paper concludes that, to develop transport policies inducing efficiency, more emphasis should be put on promoting competition "for-the-market" (competitive tendering for the license is proposed as a good example); in such a case, suggest the authors,⁵³ «the new role of public authority will be to coordinate the ex-ante competition, to design efficiently that single licensing network and to enlarge the number of effective potential competitors». They also compared the relative efficiency of public and private-owned companies included in the sample, so to infer the relevance of property on the cost efficiency performances,⁵⁴ in the spirit of property rights theory that pleads for an efficiency advantage of private firms over the public ones.⁵⁵ The results indicated that there existed no significant difference in terms of cost efficiency between the public and the private enterprises in the Italian regional bus industry;⁵⁶ this, argue Fazioli et al., because «the absence of effective competition and the strong regulation of firms behaviors give no raise to important differences in efficiency incentives between public and private operators».

As it can be noted, both studies derive important implications for a regulation policy of the LPT industry aiming to improve cost efficiency of the operators, whose contents have been at some extent implemented in the recent reform process of the Italian sector started with the law n. 59/95.⁵⁷ In fact, Levaggi underlines the relevance of the type of subsidization mechanism in determining the performance of companies in terms of inputs productivity, suggesting the transition from the present cost-plus to a fixed-price reimbursement scheme in order to create an higher powered incentive environment that should lead to a significant operating costs reduction. Fazioli et al., on the other hand, highlight how a rise in efficiency incentives could depend on new pressures attached to the (potential) competition in the sector, and remark that in the context of regulated industries the ownership structure in itself has no impact on cost efficiency or, in other words (Ivaldi, 1997, page 2), «performance differential between pubic and private firms, when it exists, is due to a set of institutional constraints and incentives».⁵⁸ However, in quoting such arguments the above authors remain on a purely

⁵³ See at page 86.

⁵⁴ For this purpose, they calculated the correlation coefficient (r) between the indicator of the cost efficiency and a dummy-variable for the ownership, which takes the value of 1 if the company is public and of 0 if the company is private.

⁵⁵ See, for example, Alchian (1965).

⁵⁶ The value of the correlation coefficients indicated the irrelevance of ownership respect to inefficiency (r = 0,05).

⁵⁷ See again CER (1997), Gorla (1999) and Federtrasporto (1999).

⁵⁸ See Ivaldi (1997), "Performance Differential between Public and Private Firms: A Survey of Empirical Tests", mimeo, IDEI. In this paper the author point out that the aforesaid conclusion is widely recognized by recent

descriptive plan that merely try to justify ex-post the evidence about inefficiency obtained through cost frontiers estimation. In practice, their intuitions about the importance of the institutional and regulatory framework in inducing x-efficient behaviors by firms are not transformed into quantitative economic relationships and appropriately included within the structure of the econometric models to estimate, so as to make the measures of productive efficiency more endogenous, i.e., directly dependent on the system of incentives and institutional constraints impinging on the activity of firms.

As Ivaldi (1997) underlines,⁵⁹ by the preceding discussion, we can argue that «the frontier approach for estimating measures of efficiency can be justified [...] but one must be very cautious in using econometric results on efficiency measurement, as long as regulatory constraints are not taken into account in a proper way [...], because they are not able to provide unbiased efficiency measures». This suggestion has been motivating further research in the direction of elaborating more structural econometric models, whose developments date back to relatively recent years and are therefore still scarce. The discussion concerning these contributions will be approached in Section 3 of this work, in which we will also try to bring some conclusive considerations about the usefulness of this new approach to analyze the effects of reforms in progress in the Italian LPT industry.

3. Recent developments: institutional constraints and incentives issues

At the end of previous section, it has been underlined how recent works on the cost structure in the Italian LPT sector provided useful hints about the relevance of the institutional and regulatory environment in determining the productive performance of firms and then the extent at which they are induced to achieve efficiency goals, that is, results consistent with the cost minimization behavior predicted by standard microeconomic theory. In actual fact, this issue is not new at all. Indeed, already Leibenstein (1966) in a well-known article about the xefficiency concept appeared on the *American Economic Review* more than thirty years ago, mentioned the existence of an overall cost inefficiency due to a lack of productivity effort by managers and workers in the production process. He pointed out that this scarce will of working agents is the direct consequence of a low powered incentive environment which does

theoretical approaches, between the others Laffont and Tirole (1993, chapter 17, page 637) and Vickers and Yarrow (1991, page 116).

⁵⁹ Ivaldi (1997), pages 16-22.

not induce them to seek cost improving methods, and underlined how a rise in incentives affecting the productivity of the worker can lead to significant operating cost reductions.

As suggested in the econometric studies reviewed above and strongly reaffirmed in more recent works on productivity analysis,⁶⁰ this change in incentives may depends for instance on new competitive pressures created in the industry, or, in case the producer is a local monopoly regulated by an authority (the majority of LPT companies in the continental Europe), on the introduction of regulatory regimes providing for payment by result schemes that assure the possibility to be "residual claimant" for the firm, i.e., to appropriate the profits derived from productive efficiency recoveries. The latter case, in particular, is at the core of the so-called new theory of regulation which emerged in the early 80's with the contributions of Loeb and Magat (1979) and Baron and Myerson (1982) and has been fully evolving in the more recent works of Laffont and Tirole (1986, 1990a, 1990b, 1993). This theory, by emphasizing the impact of the power of incentive schemes adopted by the regulator on the cost performance of the regulated firm, put the regulation problem in terms of a principal-agent relationship within a framework of asymmetric information. In this way it manages to provide a methodology for designing optimal regulation mechanisms and a rigorous foundation for the welfare analysis of different institutional settings subjected to informational constraints pertaining to the production technology and the cost-reducing effort released by manages of the regulated firm.⁶¹

From the point of view of the empirical research, these developments of the theory of incentives in regulation have been leading some econometricians working in the field of the applied industrial organization to stress the importance of taking into account in a proper way the system of incentives and institutional constraints impinging on the activity of regulated firms of which one intends to analyze the global efficiency or, more generally, the productive and cost structure. Doing so would allow in the first place to reduce the sources of incorrect specification of the econometric cost model, which, in turn, should avoid bias in the estimates of technological parameters; secondly, in a policy perspective, it could be helpful in practice to design more effective incentive regulation schemes (in terms of global efficiency, industrial costs savings and social welfare) based on hints derived from the new theory of regulation. As pointed out by Ivaldi (1997), the achievement of this goals advocates the resort to a structural approach, where the term *structural* is to mean that the method used to estimate the operating

⁶⁰ See Ivaldi (1997) and Gagnepain and Ivaldi (1998).

⁶¹ We will come back more in detail on these issues in Section 3.2, where a stochastic cost frontier model including asymmetric information variables will be discussed.

costs has to account in the frame of the econometric model for the effects that regulatory mechanisms exert in the real world on the productive performance of firms.

In the following section (3.1) it will be briefly discussed the first empirical works on the cost structure of LPT companies operating in France and Norway which tried to start in the direction just mentioned, that is, towards the use of structural econometric models. However, these are still exploratory studies for a twofold reason. First, the analysis is carried out in terms of the average characteristics of costs and technology ("cost function" approach) rather than in view of assessing the overall inefficiency of single productive units ("cost frontier" approach). Moreover, in their attempt to test the proposed conjecture according to which regulatory regimes endowed with distinct contractual schemes of subsidization should have different effects in terms of incentives to productivity and cost reduction, these works does not resort to asymmetric information models. In other words, they only confine themselves to observe the cost variation connected with the adoption of different reimbursement schemes in the regulation of LPT companies, by resorting to instrumental variables to distinguish among regulatory regimes, and no attempt is done to make explicit the mechanism through which the incentive power of the regulation operates. This would require to built the econometric cost model by specifying the utility function of transport operators and the variables describing the informational asymmetries that characterize the relationship between the regulatory authority (Principal) and the LPT company (Agent). Nevertheless, succeeding empirical contributions on this subject have introduced important extensions so as to refine the structure of the analysis and go as to elaborate a "more structural" econometric model, i.e., a stochastic cost frontier model including asymmetric information variables. This will be illustrated in Section 3.2.

3.1. Effects of regulation on the cost structure of LPT systems: first studies

Following the procedure adopted in Mathios and Rogers (1989) to study the effects of different regulatory schemes on long distance telephone rates in the United States,⁶² Dalen and Gomez-Lobo (1995) and Gagnepain (1998) made use of reduced form econometric

⁶² Mathios and Rogers (1989) propose an approach which compare the tariffs level of long distance calls applied in the States where "*rate of return*" regulatory schemes (RoR) are employed with that one operative in the States adopting a "*price cap*" regulation (PC). At this purpose, they include among the variables explaining the price level also a dummy variable that assumes value 1 if the regulation fixes an upper bound to price flexibility and value 0 in the case of RoR schemes. The analysis suggets that tariffs are generally lower - from 7 to 13 per cent in the States where a PC regulatory mechanism is adopted.

models⁶³ to measure the impact of the contractual structure of subsidization on the operating costs in the Norwegian and French urban transport sectors respectively. In both contexts the content of the contract ruling the relationship between the authority in charge of the service organization and the transportation company defines, through the regulatory scheme adopted, the features of the mechanism to finance operating costs, and consequently determines the industrial and commercial risk (the former relating to costs, the latter relating to revenues) suffered by the operator and therefore, according to the discussion above, its incentive to cost reducing efforts.

As the majority of regulated LPT industries in Europe, the Norwegian urban transport sector too needs outstanding transfers from the regulatory authority to meet the budget deficit due to high operating costs. The regulation of the sector has however undergone several important changes during the last ten years, the main motivation of which was the reduction of the excessive amount of public funds spent in local transport that represented a high cost for the collectivity⁶⁴. In 1983, in particular, the old system of ex-post balancing of accounts was abolished and replaced by an ex-ante bargaining between the regulator (a county-level authority) and the company to fix subsidies before the start of the production process. Before 1986, all counties practiced company-specific bargaining over production level, fares and costs. From 1986, in connection with the introduction of a block grant system between the central government and the counties⁶⁵, regulatory practice began to develop differently in each county. Some counties adopted a system of *standard-costing* in which the county and the regulated companies agree upon a set of criteria for calculating costs of operating a busnetwork. An important aspect of this scheme is that the same standard costs apply to all companies within a county and are not object of individually negotiated contracts. Once the cost criteria are fixed, realized costs that deviate from the standardized costs will not influence the level of subsidies from the regulator,⁶⁶ and this, in the light of new theory of regulation, should create incentives for efficiency.

⁶³ Whit this term one usually refers to models which do not provide a complete design of the environment characterizing the production activity, in the sense specified above through the discussion concerning the shortcomings of these first studies about the effects of regulatory schemes that do not make explicit the role of informational asymmetries in the structure of the econometric cost model.

⁶⁴ For more details see Andersen (1992).

⁶⁵ This aspect of the regulatory reform was important because implied that counties were free to move funds between different types of activities. This change, therefore, increased the opportunity cost of public funds in the transport industry. Moreover, contemporaneously the central government removed all types of constraints on the regulatory relationship between the counties and the transportation companies, as it has been happening in Italy after the reform process of LPT sector started with the Law n. 59/1995.

⁶⁶ Given fares and timetables, the standard-cost system fixes ex-ante the level of transfers which is granted by the regulator.

Indeed, the latter type of regulatory mechanism resembles a *fixed-price regulation*, which is a high powered incentive scheme due to the fact that the amount of transfers is fixed exante on the basis of expected costs (C_0) and revenues $(R_0)^{67}$ and any extra profit derived form higher cost reducing effort accrues to the firm. Therefore, the company will increase effort until the marginal cost savings equals the disutility of effort, which happens to be also the socially efficient effort level. On the contrary, the companies in the counties which did not adopted the standard-cost system and have carried on practicing company-specific bargaining about costs and transfers are assumed to confront low powered incentives, given the similarity of this second type of mechanism with a low power *cost-plus regulation*.⁶⁸ Under this regime the public authority usually receives the commercial revenues (*R*) and pays the firm's total expost cost (*C*) and a net transfer t_0 ; hence, the firm does not bear any risk and thus it has no incentives to produce efficiently, as any cost saving from increased effort is passed on to consumers.

Given this regulatory framework, Dalen and Gomez-Lobo (1995) tried to assess the effects of a change in the regulatory scheme on the operating costs level of Norwegian bus companies.⁶⁹ To that purpose they adopted a tranlogarithmic functional form⁷⁰ and included in the econometric cost model a dummy variable that assumes value 1 if the operator was subjected to a standard-cost regulation and value 0 in case of resort to specific-company bargaining. The empirical evidence suggests that the standard-cost regime lowers operating costs on average by 3,6% compared to the individually negotiated contracts, in such a way confirming the theoretical prediction of a lesser incentive for cost reducing effort under the latter regulatory scheme.

The approach followed by Gagnepain (1998) is similar to that one of Dalen and Gomez-Lobo. In the former too a translogarithmic type reduced form model is used to analyze the operating costs structure of 60 French urban transport companies observed over the period 1985-1993. Once again, a dummy variable representing the subsidization scheme adopted for the regulation of different LPT firms is included among the regressors of the cost function, in order to capture the impact of regulatory constraints on the current costs level.

⁶⁷ This symbols will be employed later for discussing the model in Section 3.2.

⁶⁸ At this regard, studies of the transportation industry carried out by the Norwegian Ministry of Transport on this group of regulated firms concluded that the regulator's screening of network-cost is low, that the company's historical cost form the basis for transfers and that incentives for efficiency are nearly absent. Clearly, the exact measure of incentive power for individually negotiated contracts depends on the specific bargaining process regarding transfers and on the regulator's ex post reaction to deviations from the bargaining outcome.

⁶⁹ The analysis is performed on a balanced panel of 88 bus companies over the year 1987-1991.

⁷⁰ See Section 1.1.

As regards the framework regulating the French LPT sector, about six different types of contracts between local authorities and transportation companies can be identified, three of which are largely prevalent and account for around 84 per cent of regulated firms: [1] *management contract*; [2] *lump-sum payment contract*; [3] *lump-sum price contracts*. It is plausible suppose that these distinct regulatory schemes give different incentives to reduce operating costs. In fact, while the first type of contract goes quite close to a cost-plus regulation,⁷¹ the other two subsidy mechanisms are rather similar to a fixed-price scheme.⁷² Given these features of the regulatory framework, the author tried to estimate the effects on the productivity effort, thus on the operating costs, induced by the contractual environment that regulates the decisions of transportation operators. As before, the "contract" dummy variable assumes value 1 for the fixed-price type regulatory schemes, i.e., the contracts with lump-sum payment contracts.

Also in this case the empirical evidence is consistent with the theory of incentives in regulation. The dummy variable appears significant and its coefficient is negative. Indeed, the econometric analysis indicate that the LPT firms regulated through fixed-price type contracts exhibit a level of operating costs which is lower on average by 2,05% compared to companies under cost-plus type regimes. These results clearly confirm the initial hypothesis of the author according to which a transportation firm has different incentives for cost reduction depending on the contractual scheme of subsidization proposed to it by the authority in charge of the regulation; more specifically, it would not appear an optimal decision for the local authorities to submit contracts establishing that LPT company will be repay in full the ex-post operating costs.

3.2. Towards a more structural approach: the role of asymmetric information

As prefaced in the introduction to the previous section, the studies discussed above represent pioneering contributions in the field of the analysis of the effects of regulatory constraints on

⁷¹ The management contract [1] defines a relationship between authority and company according to which at the end of each financial period the operating firm transfers to the authority all the commercial revenues and the latter will refund it in exchange the total ex-post costs. Therefore, the company does not bear any risk.

⁷² Under the type [2] of regulatory regime, i.e., the lump-sum payment contract, the operator bear all the risks on costs and revenues: the company obtains a transfer in order to achieve an expected budget balance and provides managers with a net monetary payment. The class of fixed-price schemes also includes the type [3] of contract, i.e., the lump-sum price contract. Under this variant, the local authority receives expected commercial revenues and pays an ex-ante expected cost to the company. In practice, it is similar to the previous variant of fixed-price contracts.

the cost structure of LPT companies. In fact, the positive results obtained in these first works advocated the research of a more structural formalization of the mechanism by which the regulatory schemes impinge on the activity of transportation firms. In particular, the cost model to estimate should be put into the analytical framework developed by the *new theory of regulation* that stresses the importance of *informational asymmetries* and incentives in the relationship between an authority (Principal) and the company (Agent) of the sector it regulates.

The mentioned literature, according to the framework elaborated by Laffont and Tirole, adopts a general structure in which the regulated firm is assumed to hold private information about the intrinsic efficiency of its own technology (presence of *adverse selection*)⁷³ and take efforts to reduce production costs that are unobserved by the regulatory authority (*moral hazard* phenomenon).⁷⁴ From the regulator's point of view, firms should produce efficiently, in the sense that outputs are to be produced at lowest possible costs (taking into account firms' disutility of effort), and, due to the cost of public funds, firms should not earn positive rents. However, when informational constraints of the regulator are considered, there appears to be a relevant trade-off between these two goals. In such a second best world of asymmetric information the regulation problem⁷⁵ becomes consequently to define an optimal contract structure that swap some of the efficiency incentives typical of fixed transfer schemes⁷⁶ for the rent extraction properties of cost-plus regulatory mechanisms.⁷⁷ The main results from Laffont and Tirole's model⁷⁸ is that this second best optimal solution can be implemented by offering firms a menu of linear contracts which set the remuneration of the regulated company according to the following transfer function:

 $T = a - b(C - C_0), [13]$

⁷³ For instance, bus companies are generally better informed about the need for drivers in a given transport network, the fuel consumption of buses or the effect of traffic congestion on costs.

⁷⁴ Indeed, managers can spend time and efforts in such actions as monitoring bus drivers' time of rest, providing drivers with training programs or solving potential conflicts between them that are normally hard to observe for the regulator.

⁷⁵ The regulator is assumed to maximize a social welfare function subject to incentive compatibility and participation constraints.

⁷⁶ Such high-powered incentive contracts, however, generally imply that firms earn excessive rents. The reason is that when the regulator does not know exactly the real cost of production and the firm must earn a fair rate of return on capital, the fixed amount of transfers must be based on a pessimistic estimate of the firms' unobserved intrinsic efficiency, and firms that happen to be more productive will consequently enjoy positive rents.

⁷⁷ We have seen above that cost-plus contracts guarantees that, no matter what is the intrinsic efficiency of the firm, it will not earn positive rents, as any eventual cost saving resultant from increased effort is passed on to consumers.

⁷⁸ For a more complete description of this problem and the mathematical derivation of the solution it is referred to the original publications of the authors. Main references are Laffont & Tirole (1986) and Laffont & Tirole (1993).

where *C* and C_0 indicate realized and expected costs respectively, and *a* and $b \in [0,1]$ represent the parameters of the contract; more specifically, *a* is a fixed transfer to the firm, whereas *b* defines the incentive power of the contract as it measures how transfers are affected if observed costs differ from the expected ones. This menu of linear contracts holds the property that contracts designed for efficient firms have a high *b*, reaching 1 (fixed-price scheme) for the most efficient company, while low powered incentive contracts are devised for inefficient firms,⁷⁹ with the power of the contract (and the associated optimal effort level) that decreases the more the company is inefficient and takes value 0 (cost-plus scheme) for the least efficient one. The optimality characteristic of this transfer function lies in that it induces each firm to pick the contract that corresponds to its own intrinsic level of efficiency.⁸⁰

Although the theory of incentives in regulation has become a well developed theoretical field, the empirical research aiming to examine the feasibility and the effects (in terms of industrial costs saving, disbursed transfers and social welfare) of different regulatory schemes proposed by this literature has not been very extensive. Indeed, only in the recent years it has been witness to the growth of a branch of empirical studies that have been dealing with the topic of asymmetric information, incentives and regulation from an econometric perspective, mainly thanks to progress in econometrics that has provided techniques allowing to overcome the difficulties linked to the estimation of the latent variables (intrinsic technical efficiency and cost reducing effort) not observable to the econometrician.

From a normative standpoint, it is worth to be mentioned the modern strand of research undertaken by Gasmi, Laffont and Sharkey (1997a and 1997b) that combine an engineering cost model⁸¹ with some calibrated functions⁸² to study the properties of different optimal regulatory schemes in local exchange telecommunications networks and compare them with a

⁷⁹ That is, those ones using a low productivity technology.

⁸⁰ The non-optimality of cost-plus regulation is due to the fact that it offers a low powered schemes to all types of firms. Likewise, a fixed-price mechanism, such as price cap, offers a maximal powered scheme to all types of firms.

⁸¹ Rather than employing traditional econometric techniques, relied on historical field data, the authors resort to an engineering simulation model that allow them to obtain a detailed specification of the cost function for a local exchange network over the entire range of feasible outputs (whereas real world data are usually available only for a relatively narrow range of output). Moreover, the use of such a procedure enables to take into account the actual technical characteristics of existing telecommunications networks and to estimate not only current costs but also forward looking costs. See Forsund (1995) for a recent survey of the literature that uses the engineering approach to the modeling of production and cost functions.

⁸² More precisely, Gasmi et al. had to calibrate market demand, social surplus, and disutily of (cost-reducing) effort functions. In addition, they employed prices of capital and labor to simulate the technological uncertainty and the effect of managerial effort on total costs respectively in local exchange telecommunications networks. Finally, since they did not hold detailed information about the regulatory environment, the authors utilized a probability distribution for the regulator's beliefs about technology characteristics that was assumed to be uniform over the range of values for which the cost function was defined.

variety of mechanisms implemented both with traditional cost-plus regulation and new fixedprice oriented regulatory reforms.⁸³ In the same direction, even if relying on more standard econometric techniques for the cost function estimation, Wunsch (1994) tried to compute numerically second best optimal menus of linear contracts as proposed by Laffont and Tirole (1986, 1993)⁸⁴ for the regulation of mass transit firms in Europe, and made comparisons between current (observed) costs level for a given firm, expected costs under the menu of contracts, and expected costs in case of complete information (first-best).⁸⁵

Paying more attention to methodological issues concerning the empirical analysis of the production and cost structure of a particular industry, we can then distinguish another group of studies that, besides discussing the policy implications of the informational asymmetries and the effects of different regulatory schemes, they especially stresses the gain in terms of better specification of the econometric cost model (consequently smaller bias in the estimates of technological parameters) that may be achieved by means of a structural approach to cost analysis; that is, by resorting to an economic-theory-based methodology which incorporates into the econometric model the distortions on the productive activity due to the presence of regulatory constraints and the existence of informational asymmetries in the regulator-firm interaction. The pioneer contribution in this perspective is attributable to Wolak (1994). He carried out a study of the production process of a California water utility, and showed that the parameters of the water delivery technology estimated adopting an asymmetric information framework were quite different from those obtained using standard cost-function estimation procedures⁸⁶ and provided a superior description of the observed level of costs and output.⁸⁷

⁸³ The analysis also included the redistributive consequences of different regulatory schemes and the sensitivity of their relative performance to the cost of public funds.

⁸⁴ In this case the author derived a measure for the regulator's beliefs about the technological efficiency across companies from the unexplained variance of a previous cost function OLS estimation, and worked with a normal distribution of the efficiency parameters.

⁸⁵ Regarding this aspect of the analysis, he points out, in line with the theoretical predictions, how the menu of contracts is more successful in extracting a remarkable share of the anticipated cost reductions under complete information when it is proposed to firms whose current costs level is high.

⁸⁶ In particular, he found far less economies to scale in the production technology. In fact, his estimates indicate that the water delivery production process exhibits very slight decreasing returns to scale, whereas conventional techniques implied substantial scale economies as well as nonsensical elasticities of total cost with respect to input prices for most of the estimated standard models (for a discussion of reasons of why scale economies may be overestimated in regulated industries see also Feinstein and Wolak, 1991). This aspect is especially relevant from a policy point of view, as this present modeling framework may provide more useful hints into deciding questions as the presence of natural monopoly with respect to a conventional minimum-cost framework, considered the need of consistent estimates of an industry's underlying technology to approach such problems and the better fitting showed by the structural approach.

⁸⁷ The better fitting of the asymmetric information model has been verify by means of a testing procedure for non-nested hypothesis.

A step forward in respect of Wolak has been taken in an ensuing paper by Dalen and Gomez-Lobo (1997). This contribution is particularly important for a twofold reason. First of all, the authors apply the new approach developed in their work to the analysis of the Norwegian local transport industry, that is the subject of our study. Secondly, realized that the stylized depiction in Wolak of the regulator-firm interaction in terms of optimal contracts unlikely described the real regulatory process when the data were generated, they moved towards a more positive approach of analysis, and proposed a structural model in which the regulatory constraints reflect the contractual schemes actually used by the Norwegian local authorities to regulate the bus transport firms (i.e., the company-specific bargaining and the cost-standard system already illustrated above).⁸⁸ Moreover, unlike Wolak, they assumed that companies not only hold a private information about their own intrinsic technical efficiency but, in line with Laffont & Tirole's model, they can also taken unobservable cost-reducing actions in order to increase the efficiency level, thereby admitting the presence of both adverse selection and moral hazard in the relationship between the regulator and the firms.

Like Wolak, Dalen and Gomez-Lobo too show that the parameters of a traditionally estimated cost function will be biased down, because they implicitly incorporate the negative impact on costs due to an increase of the managerial effort, and underline how this fact provides an interesting reason⁸⁹ of why scale economies may be overestimated in traditional studies of regulated industries.⁹⁰ The authors used then the estimates from their cost model⁹¹ to calculate the expected costs, the expected gross transfers and expected welfare associated to the effort level induced by different regulatory schemes. In particular, they matched the

⁸⁸ In their model Dalen and Gomez-Lobo also considered that most companies producing urban public transport services operated in the unregulated and competitive inter-city transportation market too that, by itself, provides maximum incentives to increase productive efficiency. Consequently, they expected aggregate efficiency of a bus company to be linked to the output mix in these two markets and attempted to exploit this information in the econometric analysis.

⁸⁹ Alternative to Feinstein and Wolak (1991).

 $^{^{90}}$ In fact, the effect on costs of a rise in input prices, or output level, will be somewhat mitigated by a rise in effort, since (given the cost functional form adopted by the authors) the potential cost savings from increasing effort are higher when costs are higher ("Arrow effect"). This implies, among other things, that bigger firms, which produce large amounts of output, will have an incentive to increase effort more than small companies. In this case, the eventual decreasing in the average cost level for large-seized companies is not directly attributable to a scale effect but to the rise in effort induced by output expansions. The authors thereby conclude that not explicitly accounting for effort in modeling cost relationships can lead to a considerable bias in the estimated elasticities, and in particular in the estimate of economies of scale (they showed that the bias in this latter case will be between 7% and 17%).

⁹¹ Combined with a normal probability distribution of technical efficiency measures estimated from the sample residuals (similarly to the approach followed by Wunsch (1994)) and different values chosen to simulate the cost of public funds.

current regimes⁹² and compared the properties of these latter against an optimal contract and different tendering mechanisms, more precisely, the auction for the optimal contract proposed in Laffont & Tirole (1987)⁹³ and the second-price Vickrey auction for a maximal power (fixed-price) incentive scheme.⁹⁴ One important finding at this regard is that an auction for a fixed-price contract provides expected savings in transfers estimated to be between 24% and 39% of operating costs (depending on the number of bidders and the regulatory schemes used before the introduction of tendering of bus routes) and allows to achieve almost the same welfare level as the optimal tendering. Therefore, taken account of the complexity of optimal contracts, and the consequent difficulty to implement optimal auction schemes in the practice, the authors concluded that the simpler fixed-price contract auction seems to be the ideal regulatory mechanism for the Norwegian bus transport industry.⁹⁵

In spite of the remarkable refinements and the greater realism brought into their model compared to the pioneer work of Wolak, the contribution of Dalen & Gomez-Lobo is affected by the basic fault, typical of all the empirical studies based on a "cost function" approach, of not controlling for the inefficiency levels of different companies (i.e., the extent to which each firm deviates from the best practice cost frontier) in the analysis of the effects of various regulatory regimes. In fact, the companies' performance is judged by computing the expected costs and social welfare under different schemes with respect to the average sample firm.

The ensuing studies of the French urban transportation sector carried out by Gagnepain and Ivaldi (1998, 1999) have the merit of being the first to have found a common ground between the emerging empirical literature concerned with the issues of regulation under informational asymmetries in the cost structure and the econometric literature on firm's efficiency reviewed in Section 2 of this work. To achieve this goal, they developed a *structural stochastic cost frontier model*, that also allows to shed light on some important econometric questions recurrent in the practice of frontier analysis. It will be discussed below

⁹² In this case they confirmed the result obtained in their previous study (see above). Indeed, standard-cost regulation seems to be more incentive powerful than the alternative company-specific scheme, and lowers costs by about 5% compared to the individually negotiated contract.

⁹³ This optimal auction could be implemented as a direct mechanism by having firms declare their efficiency type and the authorities choosing the most efficient company as the winner. This winner would then to exert the effort allocated to him by the menu of optimal contracts and receive informational rents equal to the difference between his type and the second most efficient type's rents under the optimal contract. Truth telling represents a dominant strategy in this auction.

⁹⁴ All companies have the incentive to bid the true cost that they would incur given their type and first best effort levels.

⁹⁵ In effect, the regulatory authorities in Norway have been considering during these last years the introduction of tendering for the bus transport routes, and the results obtained by Dalen and Gomez-Lobo (1997) undoubtedly support the adequacy of this regulatory reform in terms of transfer savings and improved social welfare.

(Section 3.2.1), together with the results emerging from the application of this new approach to the study of the productive efficiency in the French LPT industry (Section 3.2.1.1) and some policy evaluations concerning the comparison between actual and optimal regulatory regimes (Section 3.2.1.2).

3.2.1. Informational asymmetries and efficiency: a stochastic cost frontier model

The model we are going to present (Gagnepain and Ivaldi, 1998) reverts in large part to the analytical structure elaborated by Dalen and Gomez-Lobo (1997). Indeed, in this work too the authors resort to an economic-theory-based methodology for estimating the effects of the regulatory constraints on the industrial cost structure. They develop, therefore, a structural econometric cost model derived by accounting for the utility function of transportation companies under alternative contract schemes and the unobservable terms concerning the intrinsic technical inefficiency linked to the quality of some inputs used in the production process (adverse selection parameter) and the effort released by managers to reduce primal inefficiency of firm (moral hazard parameter). The originality of this contribution, as we will see later, is in that it has traced the analytical structure to that one typical of a stochastic cost inefficiency⁹⁶ in which, however, the error component the literature generally attributes to cost inefficiency⁹⁷ in this case is already built-in and the econometric model exactly coincides with the theoretical model, without the necessity of adding other more than a random disturbance term to account for potential measurement faults.

More precisely, the principal innovation introduced in the field of efficiency analysis through the structural approach proposed by Gagnepain and Ivaldi is due to the fact that the global inefficiency component which usually deviates firms from the best practice frontier is made endogenous and depends on the technology characteristics and the incentives generated by the regulatory environment in a specific way. This manner of dealing with the productive efficiency issue allows then the authors to underline the serious questions (of endogeneity and identification) that econometricians might face in the traditional frontier estimation and to show the way by which their approach contribute to get over this type of problems. Moreover, the approach to cost structure analysis in terms of "frontier" rather than of "function" permits them to evaluate the inefficiency level and the cost-reducing effort for each firm in the sample, by estimating the individual adverse selection and moral hazard parameters. They are

⁹⁶ See Section 2.

⁹⁷ That is, u_i in the expression [6] at page 7.

thereby able to carry out the matching between the different actual contract schemes used in the regulation of the French urban transport industry and the comparisons between these latter and the optimal regulatory regimes with reference not to the average practice behavior (as in Dalen & Gomez-Lobo, 1997) but to the performances of single companies. In the remainder of this section we expose the structural model to estimate for a generic industry regulated by an authority, while in the next two subsections the application of this model in the context of the French LPT sector will be proposed.

The basic idea of this work is to decompose the productive inefficiency into a component of exogenous technical inefficiency (θ) and another term depending on the insufficiency of the endogenous effort activity (*e*) released by managers to remedy the intrinsic productivity lack of some inputs. In line with the *new theory of regulation*'s models, the analysis assumes that both inefficiency sources are not known by observers (regulator and econometrician) who does not take part in the production process, hence the presence of informational asymmetries in the regulator-firm interaction. Moreover, parameters θ and *e* are supposed to distort the use of one input only, from now on denoted as the r^{th} productive factor in the vector *x* of *M* inputs utilized by the firm to obtain the output *y* according to the available technology β and the capital stock *K* set by the regulator.⁹⁸ One can then represent the relation between output and input trough the following production function

$$y = f(x, K, \theta, e; \beta), \qquad [14]$$

where the inclusion of the parameters θ and *e* is to indicate that the use and the management of the inputs set is affected by productive inefficiency; more specifically, exogenous technical inefficiency θ causes a fall in the productivity of a given amount of input x_r purchased by the firm, and managers are able to reduce it by exerting a significant effort level *e*. This can be summarized through the relationship

$$x_r^* = \frac{x_r}{\exp(\theta - e)},$$
[15]

where x_r is the level of input *r* observable by the regulator and relevant for the computation of the operating costs, while x_r^* are the (unobservable) efficient units of this factor amount, decreasing with θ and increasing with *e*, from which depends the actual (observed) output

⁹⁸ The analysis is limited to industries where the stock of capital is owned by the regulator and it is therefore considered as fixed by the producer. The authors however point out the irrelevance of this assumption for the general results of their model.

level. In the best case the managerial effort completely offsets the technical inefficiency, i.e., $(\theta - e)$ tends to 0, there is no residual productive inefficiency and all units of input *r* fully contribute to the production process.⁹⁹ The specification of the functional form for the inputoutput relationship enables to represent the equations [14] and [15] by a single expression [16]. For reasons of tractability and interpretability of the results, the authors adopted a simple Cobb-Douglas technology¹⁰⁰

$$y = \beta_0 \prod_{m=1}^M x_m^{\beta_m} K^{\beta_\kappa} \exp[\beta_r (e - \theta)], \qquad [16]$$

with β_0 , β_m , β_K and β_r denoting the parameters included in the technological vector β . β_r , in particular, represents the technology associated with the factor which give rise to the global inefficiency.

From duality theory, the parallel operating cost function is defined by

$$c(y, K, p, \theta, e; \beta) = \underset{x_m}{Min} \sum_{m=1}^{M} p_m x_m \text{ subject to [16]}, \qquad [17]$$

where p_m is price of input *m*. As equation [17] indicates, this is a *conditional* operating cost function, in which the amount of total costs faced by the firm is computed for a given fixed factor *K* and level of effort *e*. It is also useful for the purposes of the analysis show the expression defining the demand for the generic input *m* derived from the program [17]:

$$\ln x_m^D = \ln \gamma_m + \ln \left\{ \prod_{j=1}^M \frac{p_j^{\beta_j/s}}{p_m} \right\} + \frac{1}{s} \ln y - \frac{\beta_K}{s} \ln K + \frac{\beta_r}{s} (\theta - e)$$
[18]

where
$$s = \sum_{m=1}^{M} \beta_m$$
 [19]

and
$$\gamma_m = \beta_m \left(\beta_0 \prod_{m=1}^M \beta_m^{\beta_m}\right)^{-1/s}$$
. [20]

From [18] we can note that the distortion $[\beta_r /s](\theta - e)$ biases upward the demand of all inputs in the same proportion, although the source of primal inefficiency regards only the use of input *r*. This happens because the lack in the factor *r*'s productivity (not offset by a

⁹⁹ It should be noticed that the ratio of observed to efficient input quantities as defined by [14] is also a direct measure of informational asymmetries between the regulator and the firm.

¹⁰⁰ In particular, they underline (Gagnepain and Ivaldi, 1999) how the computations of the optimal regulatory schemes in the application to the French LPT industry would become cumbersome if one resorts to more flexible functional forms like the translog. Moreover, they judged the Cobb-Douglas function to provide a description of the technology fairly precise for the goals of their analysis.

sufficient managerial effort) alters the marginal rate of substitution¹⁰¹ of any pair of productive factors containing input r, and leads, for given observed input price ratios, to an excess of demand for all M factors and to a consequent rise in the operating costs compared to the minimum level achievable in absence of productive inefficiency. This aspect is also well discernible in the expression [21] of the stochastic cost frontier, obtained by inserting the demand equation [18] for the *M* inputs into the minimization program [17] and adding the usual random disturbance term v:

$$\ln C = \ln c(y, K, p, \theta, e; \beta) + v =$$

$$\Omega + \sum_{m=1}^{M} \frac{\beta_m}{s} \ln p_m + \frac{1}{s} \ln y - \frac{\beta_K}{s} \ln K + \frac{\beta_r}{s} (\theta - e) + v$$
[21]

where
$$\Omega = \ln\left(\sum_{m=1}^{M} \gamma_m\right)$$
 [22]

It is interesting to compare the equation [21] with the expression for the stochastic cost frontier habitually utilized by the standard literature on the efficiency analysis and illustrated before¹⁰². To that end, we may rewrite [21] as

$$\ln C = \ln c(y, K, p; \beta) + v + u$$
[23]

where
$$\ln(c, y, K, p; \beta) = \Omega + \sum_{m=1}^{M} \frac{\beta_m}{s} \ln p_m + \frac{1}{s} \ln y - \frac{\beta_K}{s} \ln K$$
 [24]

and
$$u = \frac{\beta_r}{s} (\theta - e) \varepsilon$$
 [25]

It can be easily noticed that the last two terms in equation [23] are the same we find in the usual decomposition of global error term ε into random and inefficiency components (v + u) followed by the standard cost frontier analysis. The remarkable difference with respect to that approach is in that the asymmetric information model adopted by Ivaldi & Gagnepain (1998) enable them to clarify why this decomposition is meaningful, since the inefficiency term is not "heaven-sent" but comes from a rigorous theoretical framework and may be ascribed to the bad quality of some inputs (θ) or the insufficient endeavor of the management staff (e).¹⁰³

As in the standard productivity analysis, the component $\exp\{[\beta_r/s](\theta - e)\}$ represents the percent by which observed costs exceed the frontier minimum level after accounting for the

¹⁰¹ From now on MRS.
¹⁰² See Section 2, equation [8].
¹⁰³ In regard to these questions, see Ivaldi (1997), pages 20-22.

effect of statistical variability, i.e., $(c(y, K, p; \beta)exp\{v\})$. We can also notice that a lack of productive efficiency of $exp\{\beta_r(e - \theta)\}$ percent is turned into a deviation of actual costs from the best practice frontier of $exp\{[\beta_r/s](\theta - e)\}$ percent; in this latter term reflecting overall (or global) cost inefficiency, the component θ represents exogenous *technical inefficiency*, whereas *e* could be thought as more responsible for *allocative inefficiency*, i.e. the failure (attributable to of an insufficient managerial effort) in making the actual MRS between any two inputs equal to the corresponding input price ratio. Such interpretation clearly represents a theory-based way to approach the intricate econometric problem of decomposing global cost inefficiency into its two components. Finally, note also that an improvement of returns to scale enjoyed by the industry (measured by index *s*) reduces the overall cost excess for a given lack of primal efficiency $\beta_r(e - \theta)$.

Up to now only the technological constraints has been considered in the firm decisional program. This has therefore allows to obtain a stochastic cost frontier [21] that is preliminary, in the sense that it entails an unobservable component, $u = [\beta_r /s](\theta - e)$, which is partially endogenous since the level of cost-reducing effort *e* comes from a firms' profit-maximization program and depends on the incentive power associated to the subsidization scheme adopted by the regulator for making the company able to balance its budget.¹⁰⁴ Since only fixed-price or cost-plus schemes are mainly adopted in the regulatory practice of French LPT industry,¹⁰⁵ the authors consider these two types of contracts to characterize the regulatory environment and the incentive effects of this latter on the managerial effort activity.¹⁰⁶ Given these cost reimbursement rules and the levels of prices, quality, output and capital set by the regulator, the firm manager has only to decide upon the level of its cost reducing activity *e* through the program that maximizes the following pay-off function

$$U = t_0 + \rho[R(y) - c(y, K, p, \theta, e; \beta)] - \psi(e),$$
[26]

¹⁰⁴ Keep in mind that in this model both parameters θ and *e* are at once sources of cost inefficiency within the productive process and informational asymmetries elements in the regulator-firm interaction.

¹⁰⁵ See the discussion of Gagnepain (1998)'s work at page 59.

¹⁰⁶ Since the menu of linear contracts designed according to Laffont & Tirole's model (see equation [1] above) assigns the fixed-price and the cost-plus schemes only to the most efficient (b = 1) and the lest efficient (b = 0) firm respectively, and since in the real word there cannot exist fully efficient or fully inefficient firms only, one may deduce that observed regulatory regimes in France are not optimal, probably due to the complexity of optimal contracts and the consequent difficulty to implement them in the practice. From this point of view, Gagnepain and Ivaldi, like Dalen and Gomez-Lobo (1997) earlier, follow a more positive-oriented approach of analysis compared to the pioneer study of Wolak (1994), in which the author assumed that during the period when the data were generated there was a sophisticated regulator willing and able to offer optimal contracts.

where
$$\psi(e) = \begin{cases} 0, & \text{if } e = 0\\ \exp(\tau e), & \text{if } e > 0 \end{cases}$$
 [27]

denotes the internal cost (disutility) of effort activity and $\rho = \{0,1\}$ is a variable indicating the type of regime {cost-plus, fixed-price} faced by the company.¹⁰⁷ The solution of this program is given by the optimal effort level

$$e^* = \rho \left[\frac{\Omega + \ln \frac{\beta_r}{s} + \sum_{m=1}^M \frac{\beta_m}{s} \ln p_m + \frac{1}{s} \ln y - \frac{\beta_K}{s} \ln K + \frac{\beta_r}{s} \Theta - \tau}{\tau + \beta_r / s} \right],$$
[28]

which is nil whenever a cost-plus scheme is applied by the regulator.¹⁰⁸

The expression [28] leads the authors to point out a first important issue that econometricians should appropriately take into account in the estimation of a production or cost frontier, i.e., the endogenous origin of the effort level and then of the global inefficiency component $u = [\beta_r / s](\theta - e)$. Indeed, as equation [28] shows, optimal effort e^* depends on the available technology, the output level *y*, the inputs prices p_m , the stock of capital *K* and the parameter denoting intrinsic technical inefficiency θ ; consequently, the usual assumption of no-correlation between the inefficiency terms and other regressors which is made in the standard cost frontier analysis for applying techniques as the "Aigner et al. (1977)'s ML" procedure or "random-effects" panel data models,¹⁰⁹ is no longer admissible from an economic theory perspective and could then implies severe questions of *endogeneity* in the econometric estimation. The manner through which the authors here overcome this type of problem consists of preliminarily computing the optimal effort level in the way illustrated above, and then substituting the expression [28] for e^* in the conditional stochastic frontier [21]. So, one obtains a final *structural* cost frontier to be estimate that directly depends on the exogenous technical inefficiency θ only:

¹⁰⁷ Remember that with a cost-plus (CP) contract the regulatory authority recovers total sales R(y), reimburses in full *ex-post* costs c(y) and provides a monetary transfer t_0 , while under fixed-price (FP) regimes the regulator obtains all expected revenues R_0 and reimburses *ex-ante* costs C_0 , providing the firms with a fixed transfer $t_0 = (R_0 - C_0)$, but the firm is made residual claimant, as it bears the losses and keeps the profit. Obviously, FP schemes give more incentives for cost efficiency than CP schemes.

¹⁰⁸ Optimal effort level is obtained substituting the cost frontier [21] and the expression [27] for the effort disutility into the pay-off function [26] and computing the first order condition of the maximization program. This latter $[\psi'(e) = -\rho \partial C/\partial e]$ states that the optimal effort level is such that the marginal disutility of effort $\psi'(e)$ equals marginal cost savings $-\partial C/\partial e$ under fixed-price regimes, while in presence of cost-plus mechanisms the optimal effort level is zero.

¹⁰⁹ See Sections 2.2.1 and 2.2.2, respectively.

$$\ln C = \ln c(y, K, p, \theta, e^{*}; \beta) + v = \ln c(y, K, p, \theta, \rho; \beta, \tau) + v =$$

$$\rho \left\{ H + \xi \left[\sum_{m=1}^{M} \frac{\beta_{m}}{s} \ln p_{m} + \frac{1}{s} \ln y - \frac{\beta_{K}}{s} \ln K + \frac{\beta_{r}}{s} \theta \right] \right\} +$$

$$(1 - \rho) \left\{ \Omega + \sum_{m=1}^{M} \frac{\beta_{m}}{s} \ln p_{m} + \frac{1}{s} \ln y - \frac{\beta_{K}}{s} \ln K + \frac{\beta_{r}}{s} \theta \right\} + v$$
with $H = \xi \Omega - \frac{\beta_{r}/s}{\tau + \beta_{r}/s} \ln \frac{\beta_{r}/s}{\tau}$
[30]

and
$$\xi = \frac{\tau}{\tau + \beta_r / s}$$
 [31]

A second final remark made by the authors at this point of the analysis highlights the *identification* problem into which the econometricians would run if they attempt to estimate a structural production frontier (obtainable by introducing e^* in the initial expression [16]) instead of a cost relationship. This issue arises because of the excessively large number of clusters of parameters incorporated in the final production frontier, whereas the homogeneity property of degree one in input prices enables the econometricians to identify the parameters of the cost frontier. To recover the parameters of the underlying technology, β_0 , β_m , β_K and β_r , the authors suggest to apply the ML technique originally proposed by Aigner et al. (1977) and Meeusen et al. (1977) for the estimation of a production frontier to the model [29]. This permits to obtain estimates for the cost parameters Ω , 1/s, β_m/s , β_r/s , β_K/s and τ , from which is then possible to derive consistent and asymptotically efficient estimates of the technology vector β .

3.2.1.1. APPLICATION TO URBAN TRANSPORT REGULATION IN FRANCE

The structural model just illustrated has been then utilized by Gagnepain & Ivaldi (1998) to study the technology properties characterizing the French urban transport industry and the effects of the regulatory environment on costs and efficiency of this sector. To this purpose, they used the same data set containing annual observations about 60 transportation companies over the period 1985-1993 already employed in the previous study carried out by Gagnepain (1998).¹¹⁰

¹¹⁰ As regards the adequacy of a cost frontier model for the goals of the analysis, the authors point out that since the industry is regulated by local authorities who set the quantity of transport services to be provided and the output y may be then assumed exogenous by the operators, the use of a cost frontier to analyze the technology of this industry is appropriate. They also remark that this application is illustrated mainly to underline the potential

In order to provide the required level of service (y), transportation companies are assumed to utilize four types of inputs: labor (L), materials and energy (M), soft capital (I) and hard capital (K).¹¹¹ As in the general model above, the regulator owns the stock of hard capital and related cost are then not supported by LPT firms. Moreover, the labor input was chosen as the variable whose productivity is affected by the adverse selection parameter and the effort variable¹¹². Thus, for a given network *i* at period *t*, the stochastic cost frontier to be estimated can be implicitly represented from equation [29] as

$$\ln C_{ii} = \ln c(y_{ii}, K_{ii}, p_{ii}, \theta_{ii}, \rho_{ii}; \beta, \tau) + v_{ii}$$
whit $i = 1, ..., 60; t = 1, ..., 9$
and $\beta = \{\beta_L, \beta_M, \beta_I, \beta_K\}; r = L.$
[32]

The estimation of equation [20] was performed by the ML method. According to the procedure originally proposed by Aigner, Lovell and Schmidt (1977) for estimating stochastic frontiers,¹¹³ the two unobservable terms θ and v was assumed independent. Moreover, while for the distribution of the random disturbance v the authors adopted the usual assumption of normality $\sim N(\theta, \sigma_v^2)$, the term denoting inefficiency θ was assumed to be distributed under a Beta function with scale parameters μ and ω , as this density is conveniently defined over the interval $[0,1]^{114}$ and affords also the advantage of not binding the distribution of θ to a specific shape a priori. They specified then the likelihood function of a data point conditional to θ_{it}

$$L_{it}(\boldsymbol{\theta}_{it}) = L(\ln C_{it} \mid \boldsymbol{y}_{it}, \boldsymbol{K}_{it}, \boldsymbol{p}_{it}, \boldsymbol{\theta}_{it}, \boldsymbol{\rho}_{it}; \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\sigma}_{v}, \boldsymbol{\mu}, \boldsymbol{\omega}) = \phi \left[\frac{\boldsymbol{v}_{it}}{\boldsymbol{\sigma}_{v}^{2}} | \boldsymbol{\theta}_{it} \right],$$

$$[33]$$

usefulness of this structural approach rather than to present a meticulous study of the urban transport sector in France.

¹¹¹ More precisely: output y is measured by the number of seat-kilometers, i.e., the number of seats available in all components of rolling stock times the total number of kilometers traveled on all routes; L includes all types of workers employed in the sector; M corresponds to various inputs which are regularly renewed, at least within a year; K, which plays the role of a fixed input in the short-run cost frontier adopted by the authors, includes rolling stock (number of vehicles) and infrastructure; I refers to all materials used for performing management activities, such as commercial vehicles, computer service and office supplies.

¹¹² The reason is that in bus-transit sector drivers constitute the most important input in terms of costs (in France input labor represent roughly sixty per cent of total operating costs). Firm's efficiency-discretion through the choice of effort is expected to be linked to the training and the utilization of these drivers. Moreover, the relationship between the network structure and the need for driver-hours is reported as one of the most important sources of the regulator's informational asymmetries.

¹¹³ See Section 2.2.1.

¹¹⁴ In fact, given the relationship between the efficient and observed levels of the labor input expressed in equation [15], the intrinsic inefficiency of worker agents θ is adequately defined as a percentage.

where ϕ [.] denotes the normal density function. In order to proceed with the ML estimation, however, the authors had to compute a likelihood not conditional to θ_{it} , as this latter variable is not observable by the econometricians¹¹⁵

$$L_{it} = \int_{0}^{1} L_{it}(\theta) \theta^{\omega-1} (1-\theta)^{\mu-1} \frac{\Gamma(\omega+\mu)}{\Gamma(\omega)\Gamma(\mu)} d\theta .$$
[34]

First of all, the authors underline how the results clearly indicate an improvement in the econometric specification, with respect to the estimation of a standard fixed-effects model of a Cobb-Douglas cost frontier not explicitly accounting for regulatory constraints effects and including an individual-specific term only to detect single firm's inefficiency.¹¹⁶ This gives further support to the original intuition of Wolak (1994) about the relevance of informational asymmetries for the econometric analysis of the productive and cost structure of regulated industries, already confirmed in Dalen and Gomez-Lobo (1997)'s study.

As regards the efficiency analysis, from the estimation it emerges that, on average, the transport companies are rather efficient.¹¹⁷ However, the obvious advantage of resorting to a frontier model instead of a traditional average function approach is that the former allows to recover estimates of individual inefficiency parameters too and then to assess cost distortions of single network operators above the best practice frontier. From equations [29] and [32] it can be noticed that the estimated stochastic frontier include an aggregate error term ε_{it} which is decomposable into the two usual unobservable and independent components reflecting the statistical variability of the sample (v_{it}) and the global cost inefficiency of each firm (u_{it}). The latter term, in particular, is partially endogenous and depends on the level of managerial effort induced through the contract scheme adopted by the local authority to regulate the transport network. This can be formally represented in the following way:

$$\varepsilon_{it} = v_{it} + u_{it}$$

[35]

¹¹⁵ Assuming then the observations are independent, the log-likelihood function for the analyzed sample can be obtained just by summing all individual log-likelihood functions derived from equation [33].

¹¹⁶ This second model too was estimated by the authors and then compared to that structural-approach-based one. First, they noted that the estimated value of σ_v was lower in the asymmetric information model, and this already indicates an improved specification with respect to the fixed-effect model. Secondly, given the similar values for the parameter estimates obtained in both models, they assessed the most appropriate specification through the test for non-nested models proposed by Vuong (1989); the statistic of the asymmetric information model versus the fixe-effect one was very high, and thus strongly favored the structural approach developed in the Gagnepain & Ivaldi's work.

¹¹⁷ Indeed, the evidence implies an exponential-shaped density function for θ , with most of the surface under the density function lying for values of θ lower than 0.5.

where
$$u_{it} = \frac{\beta_L}{s} (\theta_{it} - e_{it}^*) = \begin{cases} \frac{\beta_L}{s} \theta_{it}, & under \cos t \text{-} plus \ regimes \\ \xi \frac{\beta_L}{s} \theta_{it}, & under \ fixed \text{-} price \ regimes \end{cases}$$
 [36]

The authors applied then the procedure initially suggested by Jondrow, Lovell and Schmidt (1982) to decompose the aggregate disturbance term and obtain an estimate for each u_{it} (thus for each θ_{it}).¹¹⁸ After having computed the optimal individual effort levels from equation [16], an estimate of the extent to which single firms deviates from the cost minimizing behavior can be recover by inverting the index usually adopted by the stochastic frontier literature to assess global cost efficiency, that is¹¹⁹

$$CE = \frac{[c(y, K, p; \beta)] \exp\{v\}}{C} = \frac{1}{\exp\{u\}} = \exp\left\{\frac{\beta_L}{s}(e^* - \theta)\right\}.$$
[37]

The computation of the term $\exp\{[\beta_L /s](\theta - e^*)\}$ enabled the authors to measure the discrepancy between the theoretical frontier and the observed costs for all the urban transport operators included in their database. In particular, estimated values of technical inefficiency parameters, manager's effort levels and global cost distortions are presented for the year 1993.

By distinguishing companies according to the type of regulatory contract (cost-plus or fixed-price) adopted by local authorities to subsidy them, the authors highlighted how the higher productivity-improving effort levels showed by the operators subjected to fixed-price schemes was able to significantly lower the cost distortions over the frontier compared to those ones displayed by the companies regulated through cost-plus mechanisms. This evidence is clearly consistent with the new theory of regulation which defines the fixed-price schemes as the maximal powered incentive contracts. Only for a group of operators characterized by a fairly high technical inefficiency, the type of contract adopted by the regulator appeared irrelevant for the firm's productive performance. Indeed, in this cases the cost reducing activity released by managers has a little weight in determining the global cost inefficiency; consequently, the cost distortions over the frontier remain significant also in presence of high powered incentive schemes.

¹¹⁸ This procedure to estimate firm-specific inefficiency levels has been presented at page 13.

¹¹⁹ This expression has been introduced at the beginning of Section 2 (equation [7]).

3.2.1.2. COMPARISONS BETWEEN DIFFERENT REGULATORY POLICIES

To complete their study of the French urban transport industry, Gagnepain and Ivaldi (1999) evaluated the current regulatory schemes through a comparison with two benchmarks. (1) The optimal policy in a *perfect information world*; in this case, the regulator, thanks to his/her perfect knowledge of the inefficiency level of each company, is able to achieve the maximum social welfare attainable and the related *firs-best allocation* of service prices and effort levels is implemented through a fixed-price contract. (2) The optimal regulatory policy in a word where the regulator does not know the individual levels of efficiency and can only implement a *second-best allocation*¹²⁰, by offering the menu of linear contracts illustrated above through equation [13].

To make this comparison feasible, the authors had more to estimate a model of public transit service demand¹²¹ and provide an evaluation of cost of public funds for each network. They were then able to compute, for each company in the sample, the values for the operating costs, the internal costs $\psi(e)$, the welfare costs and the total welfare associated to both optimal policies, and compared them with the evidence observed under the current regulatory regimes.

The main results the authors obtained are essentially two. Firstly, they did not find a large difference between first- and second-best optimal price levels and those ones practiced under actual regulatory regimes, to suggest that the sources of possible welfare differentials between current and optimal contract schemes have to be traced on the productive structure side only. Secondly, whereas for the group of LPT companies under cost-plus regulatory regimes both optimal policies would allow to achieve an higher social welfare, in the case of networks subjected to fixed-price contracts the sign of variations in the welfare levels associated to the implementation of optimal regulatory mechanisms was not univocally defined. In fact, the replacement of a current fixed-price scheme by a second-best contract always implied a deterioration in the social welfare, while the introduction of first-best regulatory mechanisms instead of the actual fixed-prices contracts led to welfare variations whose sign appeared to depend on the circumstance that companies were making net gains or losses under the current regimes¹²².

¹²⁰ This is obtained by maximizing the social welfare subject to an incentive compatibility constraint.

¹²¹ In order to compute the consumer surplus and to assess demand elasticities that are usually incorporated in optimal pricing rules.

¹²² More precisely, the authors found that if one replaces a current fixed-price scheme by a first-best contract, the social welfare would increase in case the company is making net positive profit, whereas the welfare level would reduce if the operator is making net negative profit.

Gagnepain and Ivaldi (1999) interpreted the above evidence in terms of a moderate impact of the informational asymmetries between local authorities and transport operators. This is probably due to the fact that a remarkable proportion (more than fifty percent) of French LPT networks under fixed-price regimes (more than sixty percent of companies in the sample) was owned by local authorities, which were therefore able to get better information about the productive efficiency of the regulated firms. In conclusion, the authors underlined the usefulness of their structural approach to help explain the choice of regulation schemes in the French urban transportation sector. Indeed, in a relevant number of cases the current regimes appeared to provide an higher social welfare level with respect to that one which could be achieved by the implementation of second- and even first-best regulatory policies.

4. Conclusion: some indications for the study of public transit systems in Italy

In the light of the evidence concerning the cost structure and productive efficiency of transit companies in Italy emerged in the cost-frontier-based studies reviewed in Section 2.2.3, we are able to outline some indications about the potential usefulness of the innovative structural approach presented in the previous section for a new empirical investigation of the Italian LPT industry.

In particular, we can draw two main observations, one pertaining to the appropriateness of the econometric models used to study the cost structure and technological properties of transport networks, and the other relative to the implications for the regulatory policy of the sector that could be derived by resorting to a structural approach such as the one developed by Gagnepain and Ivaldi (1998, 1999).

Already Wolak (1994) and Dalen and Gomez-Lobo (1997) warned against potential distortions in the estimation of technology parameters which one would incur by adopting standard econometric methods to analyze the productive and cost structure of regulated firms. They highlighted, in particular, the upward bias introduced in the scale economies estimates by models that do not account for the role played by informational asymmetries and incentives in the regulator-firm interaction. As shown above, the relevance of such issues has found further support in the study carried out by Gagnepain & Ivaldi (1998, 1999), in which the authors underlined the improvement in the econometric specification of the cost frontier model that can be achieved by means of their structural approach.

This suggests that also the results of past works concerning the estimation of technology and cost properties of the Italian LPT sector are probably affected by this problem of misspecification. Therefore, new empirical analysis incorporating the impact of regulatory constraints into the econometric model to be estimated would be desirable. Especially if we consider the need of consistent estimates of an industry's underlying technology to approach such crucial problems as the presence of natural monopoly in the provision of a given network utility.

The second important advantage offered by the asymmetric information models presented above consists in allowing to evaluate the adequacy of current regulatory regimes, in terms of global efficiency of regulated companies, amount of disbursed subsidies and social welfare level, by comparing them with the results obtainable with alternative schemes of regulation.

This possibility appears particularly appealing in the context of reforms recently started in the Italian LPT industry, in order to take the poor trend of operating costs and productivity levels more in line with the one exhibited by the other countries in continental Europe. Indeed, the adoption of a structural approach would enable us to assess if the greater financial responsibility awarded to transport operators by the introduction of the "service contract" (Law 549/1995 and Decreto Legislativo 422/1997)¹²³ has actually succeeded in achieving significant recoveries of productive efficiency. That is, if the new regulatory policy for the sector has been powerfully oriented towards incentive schemes of subsidization. Moreover, following the prints of Dalen and Gomez-Lobo (1997), it would also be possible simulate the effects in terms of expected cost reductions and welfare improvements associated with the introduction of competitive tendering mechanisms for the allotment of licenses. These were already provided for by the Decreto Legislativo 422/1997, strongly recommended in some interventions of the National Antitrust Authority (AGCM) and recently reaffirmed in the Decreto Legislativo 400/1999.¹²⁴

¹²³ Refer to Federtrasporto (1998, 1999). A detailed analysis of the regulatory framework for the Italian LPT industry and its evolution during the last decade is carried out in Piacenza (2000) and Boitani and Cambini (2001). ¹²⁴ To this regard, see Federtrasporto (1997, 1999) and European Commission (1998).

REFERENCES

- Aigner, D., Lovell, K. C. A. and Schmidt, P. G., "Formulation and Estimation of Stochastic Frontier Production Function Models", *Journal of Econometrics*, 6, 1977, pages 21-37.
- Alchian, A., "The Basis of Some Recent Advances in the Theory of Management of the Firm", *Journal of Industrial Economics*, 14, December, 1965, pages 30-41.
- Andersen, S. C., "Factors Affecting European Privatization and Deregulation in Local Public Transport: The Evidence from Scandinavia", Transportation Research, 26(2), 1992, pages 179-191.
- Andrikopoulos, A. A., Loizidis, J. and Prodromidis, K. P., "Technological Change and Scale Economies in Urban Transportation", *International Journal of Transport Economics*, 19, 1992, pages 127-147.
- Bailey, E. and Friedlaender, F., "Market Structure and Multiproduct Industries", *Journal of Economic Literature*, 20(3), September, 1982, pages 1024-48.
- Baron, D., and Myerson, R., "Regulating a Monopolist with Unknown Costs", *Econometrica*, 50, 1982, pages 911-930.
- Bauer P., "Recent Developments in the Econometric Estimation of Frontiers", *Journal of Econometrics*, 46, 1990, pages 39-56.
- Berechman, J., *Public Transit Economics and Deregulation Policy*, North-Holland, Amsterdam, 1993.
- Berechman, P. and Giuliano, G., "Analysis of the Costs Structure of an Urban Transit Property", *Transportation Research*, 18, 1984, pages 277-287.
- Berechman, P., "Cost Structure and Production Technology in Transit: An Application to Israeli Bus Transit Sector", *Regional Science and Urban Economics*, 17, 1987, pages 519-534.
- Boitani, A. and Cambini C., *La riforma del trasporto pubblico in Italia: problemi e prospettive*, Working Paper HERMES, April, Italy, Turin, 2001.
- Button, K., and O'Donnel, K. H., "An Examination of the Cost Structures Associated with Providing Urban Bus Services", *Scottish Journal of Political Economy*, 32, 1985, pages 67-81.

- Caves, D. W. and Christensen, L. R., "The Importance of Economies of Scale, Capacity Utilization and Density in Explaining Inter-Industry Differences in Productivity Growth", *The Logistic and Transportation Review*, 2, 1988, pages 3-32.
- CER, *Riforme a presa rapida*, Rapporto n. 2/1997, Italy, Rome, 1997.
- Cispel, *Metodologia e parametri di efficienza per le aziende municipalizzate*, Italy, Rome, 1989.
- CNR, *Il trasporto pubblico locale in Italia: analisi, prospettive e contributi della ricerca*, Progetto Finalizzato Trasporti 2, Italy, Rome, 1999.
- Cornwell, C. and Schmidt, P., "Productions Frontiers and Efficiency Measurement", in L. Matyas and P. Sevestre eds., *The Econometrics of Panel Data*, Kluwer Academic Publishers, 1996.
- Dalen, D. M. and Gomez-Lobo, A., "Cost Functions in Regulated Industries Characterized by Asymmetric Information", *European Economic Review*, 41(3-5), April 1997, pages 935-42.
- Dalen, D. M. and Gomez-Lobo, A., Regulation and Incentive Contracts: An Empirical Investigation of the Norwegian Bus Transport Industry, Institute for Fiscal Studies Working Paper, London, 1995.
- De Borger, B., "Cost and Productivity in Regional Bus Transportation: The Case of Belgium Case Study", *Journal of Industrial Economics*, 37, 1984, pages 35-54.
- European Commission, Improved structure and organisation for urban transport operations of passengers in Europe (ISOTOPE), Transport Research, Fourth Framework Programme, Urban Transport VII-51, 1998.
- Fabbri, D., "La Stima di Frontiere di Costo nel Trasporto Pubblico Locale: una Rassegna e un'Applicazione", *Economia Pubblica*, 3, 1998, pages 55-94.
- Farrel, M. S., "The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society*, 1957, A, 120, pp. 253-281.
- Fazioli, R., Filippini, M. and Prioni, P., "Cost Structure and Efficiency of Local Public Transport: The Case of Emilia Romagna Bus Companies", *International Journal of Transport Economics*, 20(3), October 1993, pages 305-24.
- Federtrasporto (1997), *Servizi pubblici locali: un banco di prova per la concorrenza*, Bollettino n. 6/1997, Italy, Rome, 1997.
- Federtrasporto, *Efficienza e sussidiarietà: il trasporto pubblico locale dopo la riforma*, Bollettino n. 8/1999, Italy, Rome, 1999.

- Federtrasporto, *Il contratto di servizio: norme e prassi*, Bollettino n. 7/1998, Italy, Rome, 1998.
- Feinstein, J. and Wolak, F. A., "The Econometric Implications of Incentive Compatible Regulation", in G. F. Rhodes ed., *Advances in Econometrics*, Greenwich: JAI Press, vol. 9, 1991, pages 159-204.
- Filippini, M., Maggi, R. and Prioni P., "Inefficiency in a Regulated Industry: The Case of the Swiss Regional Bus Companies", *Annals of Public and Cooperative Economics*, 63, 1992, pages 437-455.
- Gagnepain, P. and Ivaldi, M., *Incentive Regulatory Policies: The Case of Public Transit Systems in France*, mimeo, GREMAQ - IDEI, Université des Sciences Sociales de Toulouse, France, March 1999.
- Gagnepain, P. and Ivaldi, M., *Stochastic Frontiers and Asymmetric Information Models*, mimeo, GREMAQ IDEI, Université des Sciences Sociales de Toulouse, France, 1998.
- Gagnepain, P., "Structures Productives de l'Industrie du Transport Urbain et Effets des Schemas Reglementaires", *Economie et Prevision*, 135, July-Sept. 1998, pages 95-107.
- Gasmi, F., Ivaldi, M. and Laffont, J.J., "Rent Extraction and Incentives for Efficiency in Recent Regulatory Proposal", *Journal of Regulatory Economics*, 6, 1994, pages 151-176.
- Gasmi, F., Laffont, J. J. and Sharkey, W. W., "Empirical Evaluation of Regulatory Regimes in Local Telecommunications Markets", *Journal of Economics & Management Strategy*, 8(1), 1997b, pages 61-93.
- Gasmi, F., Laffont, J. J. and Sharkey, W. W., "Incentive Regulation and the Cost Structure of the Local telephone Exchange Network", *Journal of Regulatory Economics*, 12, 1997a, pages 5-25.
- Gorla, G., Trasporti Pubblici Locali, Fondazione Rosselli, Rapporto 1999, Italy, Turin, 1999.
- Greene, W. H., "The Econometric Approach to Efficiency Analysis" in H. Fried, C. A. K. Lovell and P. Schmidt eds., *The Measurement of Productive Efficiency: Techniques and Applications*, New York: Oxford University Press, 1993.
- Greene, W., "Maximum Likelihood Estimation of Econometric Frontier Functions", *Journal* of Econometrics, 13, 1980, pages 27-56.
- Hausman, J., "Specification Tests in Econometrics, Econometrica, 1978, 46, pp. 1251-1271.

Hsiao, C., Analysis of Panel Data, Cambridge: Cambridge University Press, 1986.

- Ivaldi, M., Performance Differential between Public and Private Firms: A Survey of Empirical Tests, mimeo, IDEI, Université des Sciences Sociales de Toulouse, France, March 1997.
- Jondrow, J., Lovell, K. C. A., Materov, I. and Schmidt, P., "On the Estimation of Technical Efficiency in the Stochastic Production Function Model, *Journal of Econometrics*, 19, 1982, pages 233-238.
- Laffont, J. J. and Tirole, J., "Auctioning Incentive Contracts", *Journal of Political Economy*, 95, 1987, pages 921-937.
- Laffont, J. J. and Tirole, J., "The Regulation of Multiproduct Firms, Part I: Theory", *Journal* of *Public Economics*, 43, 1990a.
- Laffont, J. J. and Tirole, J., "The Regulation of Multiproduct Firms, Part II: Applications to competitive Environments and Policy Analysis", *Journal of Public Economics*, 43, 1990b.
- Laffont, J. J. and Tirole, J., "Using Cost Observation to Regulate Firms", *Journal of Political Economy*, 94, 1986, pages 614-641.
- Laffont, J. J. and Tirole, J., A Theory of Incentives in Procurement and Regulation, MIT Press, Cambridge, 1993.
- Leibenstein, H., "Allocative Efficiency versus «X-Efficiency»", *American Economic Review*, 56, 1966, pages 392-415.
- Levaggi, R., "Parametric and Non-Parametric Approach to Efficiency: The Case of Urban Transport in Italy", *Studi Economici*, 49(53), 1994, pages 67-88.
- Loeb, M., and Magat, W., "A Decentralized Method of Utility Regulation", *Journal of Law and Economics*, 22, 1979, pages 399-404.
- Mathios, A. D. and Rogers, R. P., "The Impact of Alternative Forms of State Regulation of AT&T on Direct-Dial Long-Distance Telephone Rates", *Rand Journal of Economics*, 20(3), Autumn, 1989, pages 437-53.
- Obeng, K., "The Economics of Bus Transit Operation", *The Logistic and Transportation Review*, 20, 1984, pages 45-65.
- Olson, J. A., Schmidt, P. and Waldman, D. M., "A Monte Carlo Study of Estimators of Stochastic Frontier Production Functions", *Journal of Econometrics*, 13(1), May, 1980, pages 67-82.
- Petretto, A. and Viviani, A., "An Economic Model for Cross-Section Analysis for the Production of Urban Transport Services", *Economic Notes*, 13, 1984, pages 35-64.

- Piacenza, M., The Public Transit Systems in Italy: A Critical Analysis of the Regulatory Framework, Working Paper Ceris-CNR, 16, Italy, Turin, 2000.
- Schmidt, P. and Lin, T. F., "Simple Tests of Alternative Specifications in Stochastic Frontier Models", *Journal of Econometrics*, 24(3), March, 1984, pages 349-61.
- Schmidt, P. G. and Sickless, R. C., "Production Frontiers and Panel Data, *Journal of Business and Economic Statistics*, 1984, 2, pp. 367-374.
- Thiry, B. and Lawarree, J., "Productivité, coût et caracteristiques technologiques des sociétés belges de transport urbain", *Annales de l'économie publique, sociale et coopérative*, 4, 1987, pages 368-396.
- Vickers, J. and Yarrow, G., " Economic Perspectives on Privatization", *Journal of Economic Perspectives*, 5(2), Spring, 1991, pages 111-32.
- Viton, P., "A Translog Cost Function for Urban Bus Transport", *Journal of Industrial Economics*, 29, 1981, pages 287-304.
- Vuong, Q. H., "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses", *Econometrica*; 57(2), March, 1989, pages 307-33.
- Williams, M. and Dalal, A., "Estimation of the Elasticity of Factor Substitution in Urban Bus Transportation: A Cost Function Approach", *Journal of regional Science*, 21, 1981, pages 263-275.
- Windle, R. J., "Transit Policy and the Cost Structure of Urban Bus Transportation, in J. S. Dogson and N. Topham eds., *Bus Deregulation and Privatization*, Averbury: Aldershot, 1988.
- Wolak, F. A., "An Econometric Analysis of the Asymmetric Information Regulator-Utility Interaction", *Annales d'Economie et de Statistiques*, 34, 1994, pages 13-69.
- Wunsch, P., *Estimating Menus of Linear Contracts for Mass Transit Firms*, Université Catholique de Louvain CORE Discussion Paper: 96/40, September 1996.
- Zellner, A., "An efficient Method of Estimating Seemingly Unrelated Regressions (SUR) and Tests for Aggregation Bias", *Journal of the American Statistical Association*, 57, 1962, pages 348-368.