COST EFFICIENCY IN REGIONAL BUS COMPANIES: AN APPLICATION OF ALTERNATIVE STOCHASTIC FRONTIER MODELS*

Mehdi Farsi Massimo Filippini Michael Kuenzle

Swiss Federal Institute of Technology
ETH Zentrum, WEC, 8092 Zurich, Switzerland
and
Department of Economics, University of Lugano
Via Maderno 24, 6900 Lugano, Switzerland

July 2004

* The authors are grateful to Subal Kumbhakar and Antonio Alvarez for their helpful comments. We would also like to thank participants at the North American Productivity workshop (Toronto, June 2004) for their feedback on an earlier version of the paper. Any remaining errors are solely the responsibility of the authors.

Abstract

This paper evaluates cost and scale efficiencies of Switzerland's regulated bus companies operating in regional networks. The adopted methodology can be used in benchmarking analyses applied to incentive regulation systems. Moreover, the estimations can be used to evaluate the bidding offers for the tendering processes predicted by the ongoing reform policies. Since these companies operate in different regions with various characteristics that are only partially observed, it is crucial for the regulator to distinguish between inefficiency and exogenous heterogeneity that influences the costs. A number of stochastic cost frontier models are applied to a panel of 94 companies over a 12-year period from 1986 to 1997. The main focus lies on the ability of these models to distinguish inefficiency from the unobserved firm-specific heterogeneity in a network industry. The estimation results are compared and the effect of unobserved heterogeneity on inefficiency estimates is analyzed.

1. Introduction

In many European countries the regional public bus services are being reorganized. In line with the EU policy the Swiss government has introduced important regulatory reforms in the public transport system, including regional bus companies. The new policy act predicts a tendering process for the provision of regional bus services. With the implementation of the new system, the applying companies will bid in competitive auctions and the access rights will be granted to the company with the lowest subsidies request. This system is believed to introduce greater incentives for competitive behavior. However, given the limited number of bidding companies in most regions, it is not clear to what extent the new policies lead to efficient production. Moreover, the incumbents, mostly public companies, might have an advantageous position in such auctions. Benchmarking methods can be used to evaluate the requested subsidies and proposed costs by individual companies or to adjust the minimum bidding prices.

Benchmarking analysis is based on comparing the costs of individual companies to the 'best' (most cost-efficient) observed practice. These deviations, often labeled as 'cost-inefficiency' can also be used to adjust the amount of subsidies paid to individual bus operators. Moreover, predicted costs of the benchmark practice could be used to gain information regarding the future evolution of costs incurred by the companies operating in a service area, and to re-evaluate the claimed subsidies.[1]

In order to use the efficiency estimates of individual companies in regulation, it is important to have precise measurement methods. In particular, because of considerable cost differences across various networks, it is crucial to distinguish the cost difference due to unobserved heterogeneity in external factors from the excess costs due to the company's inefficiency. Benchmarking can be conducted using econometric methods such as stochastic frontier models, which have been developed in a variety of forms during the past two decades.[2] All these models in one way or another separate the heterogeneity from cost-inefficiency. Especially, with panel data at hand, the unobserved heterogeneity can be better identified as the time-invariant elements of heterogeneity can be separately specified by firm-specific effects.

The first application of panel data models in stochastic frontier analysis was introduced by Pitt and Lee (1981). These authors formulated the firm-specific error component as a half-normal distribution, which they interpreted as inefficiency. In the following years, several models have been developed to incorporate the observed firm-specific heterogeneity. For instance, Jha and Singh (2001), Piacenza (2002) and Dalen and Gomez-Lobo (2003) use single equation models[3] proposed by Battese and Coelli (1995) to incorporate some exogenous variables to explain the determinants of the inefficiency component in the bus transportation industry. However, most of these models have a shortcoming in that they cannot disentangle firm's inefficiency from cost differences due to unobserved characteristics of the service area. Especially, transport companies operate in networks with different shapes and structures, which result in different coordination problems and thus lead to different costs. These

[1] See Farsi and Filippini (2004) for a discussion on the use of cost prediction in the regulation of public utilities.
[2] Kumbhakar and Lovell (2000) provide an extensive survey of this literature.
[3] For the advantages of single stage models, see Wang and Schmidt (2002).

characteristics are usually given and cannot be controlled by the companies. Some of these exogenous factors are either unavailable or too complex to be measured by single indicators. Unfortunately, when unobserved heterogeneity is present the inefficiency estimates can be overstated.

Greene (2003, 2004) proposes alternative panel data models, which can better distinguish between unobserved firm-specific heterogeneity and inefficiency. These models extend the previous models by adding an additional stochastic error component for the heterogeneity.[4] Such models are particularly useful in transport industries where the network and environmental characteristics are mostly unobserved or hard to measure, but play an important role on the operating costs.

The purpose of this study is to analyze the performance of different panel data frontier models with regard to estimated coefficients, inefficiency scores and estimates of economies of scale and density. Especially, we focus on the ability of different models to distinguish unobserved heterogeneity from inefficiency. Alternative models are applied to a sample of 94 Swiss rural bus companies from 1986 to 1997. It is concluded that in the studied sample, Greene's "true" random effects model has a considerable advantage over other models in separating heterogeneity from inefficiency.

The rest of the paper is organized as follows: Sections 2 and 3 present the model specification and the methodology respectively. The data are explained in section 4. Section 5 presents the estimation results and discusses their implications, and section 6 provides the conclusions.

2. Model Specification

A bus transit company can be considered as a production unit that operates in a given network and transforms labor and capital services and energy into units of transport services. Since in most cases not only the network but also the schedule of a bus operator is regulated and predetermined, it is common to estimate a cost rather than a production function.[5] Different specifications have been used in the literature.[6] Often, output is measured in terms of either passenger- or seat-kilometers. To capture some of the heterogeneity of different service areas, most specifications include additional output characteristics such as the number of stops, network length or average commercial speed. Most of these studies also include a time trend to capture the potential changes in technology.

The total cost frontier can therefore be written as the following function:

$$TC = f(Y, N, P_L, P_C, t), \tag{1}$$

[4] A similar model but with a three-stage estimation procedure has been proposed by Kumbhakar (1991) and Heshmati and Kumbhakar (1994).
[5] See Berechman (1993) for an overview of the application of cost functions in public transport.
[6] See among others Fazioli et al. (1993), Filippini and Prioni (1994, 2003), Matas and Raymond (1998), Fraquelli et al. (2001) and Fazioli et al. (2003).

where TC is the total annual cost and Y is the output represented by the total number of seat-kilometers. N represents the network length. P_C and P_L are respectively the labor and capital prices. Given that the share of the energy expenditure is relatively low (below 6%) in the majority of the cases, we consider labor and capital as the main input factors.[7] However, as we see later, the capital price is calculated from all non-labor expenses, thus includes the variations in energy prices.

It is generally assumed that the cost function given in (1) is the result of cost minimization given input prices and output and should therefore satisfy certain properties namely, linear homogeneity and concavity in input prices and monotonicity in input prices and output.[8] Input prices and output are assumed to be exogenous, thus beyond the firm's control. In the case of Swiss bus transport companies, the municipalities and the cantons specify the output by regulating the frequency of the service. The input prices can also be regarded as given, because these companies have a relatively small share in the labor and capital markets, thus cannot influence the prices through monopsony.

To estimate the cost function (1), a translog functional form is chosen. This flexible functional form is a local, second-order logarithmic approximation to any arbitrary twice-differentiable cost function. It places no *a priori* restrictions on the elasticity of substitution and allows the economies of scale to vary with the output level. The translog approximation to (1) is then written as:

$$\begin{aligned}
\ln\left(\frac{TC_{it}}{P_{L_{it}}}\right) = {} & \alpha_0 + \alpha_Y \ln Y_{it} + \alpha_N \ln N_{it} + \alpha_K \ln \frac{P_{K_{it}}}{P_{L_{it}}} \\
& + \frac{1}{2}\alpha_{YY}(\ln Y_{it})^2 + \frac{1}{2}\alpha_{NN}(\ln N_{it})^2 + \frac{1}{2}\alpha_{KK}\left(\ln \frac{P_{K_{it}}}{P_{L_{it}}}\right)^2 \\
& + \alpha_{YK} \ln Y_{it} \ln \frac{P_{K_{it}}}{P_{L_{it}}} + \alpha_{YN} \ln Y_{it} \ln N_{it} \\
& + \alpha_{KN} \ln \frac{P_{K_{it}}}{P_{L_{it}}} \ln N_{it} + \alpha_t t, \\
& \text{with } \; i = 1, 2, ..., N \;\; and \;\; t = 1,2,...,T,
\end{aligned} \qquad (2)$$

where subscripts i and t denote the company and year respectively. The technical change is specified as a linear trend and is assumed to be neutral with respect to cost minimizing input ratios.[9] The translog form requires that the underlying cost function be approximated around a specific point like the sample mean or median. Here, the sample median is chosen because it is less affected by outliers and thus the approximation will have better precision. As can be seen in equation (2), linear homogeneity in input prices is imposed by dividing total costs and input prices by labor price. The other theoretical restrictions are verified after the estimation.

[7] We considered an alternative specification including energy prices. The estimated coefficients did not change significantly and the coefficient of the energy price was generally insignificant. Moreover, because of a number of missing values for energy costs a two-input model allows a larger sample.
[8] For more details on the properties of the cost function, see Chambers (1988), p. 52.
[9] In other words the technical change does not alter the optimal input bundles.

Apart from estimating cost inefficiency, the estimation of a cost function enables us to derive important characteristics of bus supply technology such as economies of density and scale. The distinction between scale and density economies is particularly important in network industries. In such cases, a company's size is related to both its output level and its network size, which do not necessarily vary with a simple one-to-one relationship. For this reason it is important to distinguish cost changes that occur uniquely because of output changes within a fixed network and cost changes resulting from a proportional change in both network and output.

Economies of density are defined as the inverse of the elasticity of costs with respect to output that is, the relative increase in total cost resulting from an increase in output, holding all input prices and the network size fixed:[10]

$$ED := \left(\frac{\partial \ln C}{\partial \ln y}\right)^{-1} = \left(\alpha_Y + \alpha_{YY} \ln y + \alpha_{YK} \ln \frac{P_K}{P_L} + \alpha_{YN} \ln N\right)^{-1}. \qquad (3)$$

The existence of economies of density implies that the average costs of a bus operator decrease as physical output increases. Economies of density exist if the above expression (*ED*) has a value greater than one. For values of *ED* below one, we identify diseconomies of density. In the case of *ED* = 1, the company under consideration operates at the optimal output level given its network size.

Slightly different is the definition of economies of scale (*ES*). Here, the increase in total costs is brought about by an increase in company's scale that is in both output and the network size, holding the factor prices constant. However, since the changes in output and network size are inter-related, the definition of scale economies requires an assumption in this respect. The commonly used definition is the one proposed by Caves, Christensen and Tretheway (1984), which assumes that any increase in size raises the network size and the outputs with the same proportion. Based on this assumption, *ES* is defined as:

$$\begin{aligned} ES &:= \left(\frac{\partial \ln C}{\partial \ln y} + \frac{\partial \ln C}{\partial \ln N}\right)^{-1} \\ &= \left(\alpha_Y + \alpha_{YY} \ln y + \alpha_{YK} \ln \frac{P_K}{P_L} + \alpha_{YN} \ln N + \alpha_N + \alpha_{NN} \ln N + \alpha_{YN} \ln y + \alpha_{KN} \ln \frac{P_K}{P_L}\right)^{-1}. \end{aligned} \qquad (4)$$

Similarly, economies of scale exist if *ES* is higher than 1 and *ES*=1 would suggest that the company operates at an optimal scale.

It should be noted that the above definitions of scale and density economies are in terms of cost elasticity and do not necessarily correspond to the definitions derived from the production function. In fact, only in homothetic production functions, where the optimal input bundles vary proportionately, the two definitions

[10] See also Caves, Christensen and Tretheway (1984).

are equivalent. Here, we do not impose such an assumption. However, as in this paper we are interested in the cost effects of output, we define the scale and density economies as the inverse of the corresponding cost elasticities.[11]

3. Methodology

The effects of unobserved heterogeneity on inefficiency estimates are studied by a comparative analysis of four econometric models. These models are a pooled cross section model in line with Aigner, Lovell and Schmidt (1977); a random effects model as in Pitt and Lee (1981); a fixed effects model as in Schmidt and Sickles (1984); and a random intercept frontier model (also known as "true" random effects model) proposed by Greene (2003, 2004).

All models are based on the specification given in equation (2). Generally, a firm-specific effect α_i and an error term ε_{it} that can be symmetric or asymmetric depending upon the adopted econometric model are added to this core specification. The general econometric specification can therefore be written as:

$$\ln\left(\frac{TC_{it}}{P_{L_{it}}}\right) = f\left(Y_{it}, N_{it}, P_{K_{it}} \big/ P_{L_{it}}, t\right) + \alpha_i + \varepsilon_{it} \,, \qquad (5)$$

where $f(.)$ is the cost frontier as given in equation (2).

All models except the fixed effects model assume a composite normal-half-normal distribution, but they vary in the composition of the stochastic components α_i and ε_{it}. Table 1 provides a description of the stochastic terms included in the models used in this study. The first model (model *I*) is a pooled frontier model in which the firm-specific effect (α_i) is assumed to be zero. Thus the sample is considered as a series of cross sectional sub-samples pooled together. The random term is divided into two components: a normally distributed error term v_{it}, capturing general measurement errors and heterogeneity and a half-normal random term u_{it}, representing the inefficiency as a one-sided non-negative disturbance. This model is based on the original cost frontier model proposed by Aigner, Lovell and Schmidt (1977).

Model *II* is a random effects model as in Pitt and Lee (1981), which is estimated by Maximum Likelihood method. The important limitation of this model is the assumption that the firm-specific stochastic term u_i, which is assumed to include only the firm's inefficiency, is uncorrelated with the explanatory variables and is assumed to be constant over time. In fact, most frontier models assume that inefficiency is uncorrelated with explanatory variables included in the cost function. While being practical for estimation purposes, this assumption can be justified based on the fact that the apparent excess costs that are correlated with exogenous variables may be due to factors beyond the firm's control. However, the firm-specific stochastic term may contain other unobserved environmental factors, which may be correlated with explanatory variables and thus may bias the coefficients.[12] The within estimator (model *III*) can overcome this bias problem, by taking the firm-specific effects as

[11] See Chambers (1988) for more details about this issue. To avoid confusion this author refers to the inverse of cost elasticity as the "economies of size" rather than economies of scale (see page 72).
[12] In fact the results of the Hausman test performed on a GLS random effects model (not reported here) were in favor of potential correlation between regressors and firm effects.

constants, that can be correlated with explanatory variables. Thus the estimated coefficients are unbiased even in the presence of such correlations. The inefficiency estimates in this model are obtained using the procedure proposed by Schmidt and Sickles (1984).

Table 1. Econometric specifications of the stochastic cost frontier

	Model I **Pooled**	*Model II* **RE (ML)**	*ModelIII* **FE**	*Model IV* **True RE**
Firm-specific component α_i	none	none	fixed	$\alpha_i \sim \mathrm{N}(0,\sigma_\alpha^2)$
Random error ε_{it}	$\varepsilon_{it}= u_{it}+ v_{it}$ u_{it}~$\mathrm{N}^+(0, \sigma_u^2)$ v_{it}~$\mathrm{N}(0, \sigma_v^2)$	$\varepsilon_{it}= u_i+ v_{it}$ u_i~$\mathrm{N}^+(0, \sigma_u^2)$ v_{it}~$\mathrm{N}(0, \sigma_v^2)$	$\varepsilon_{it}= v_{it}$ v_{it}~iid $(0, \sigma_v^2)$ $u_i=\alpha_i\text{-}min(\alpha_i)$	$\varepsilon_{it}= u_{it}+ v_{it}$ u_{it}~$\mathrm{N}^+(0, \sigma_u^2)$ v_{it}~$\mathrm{N}(0, \sigma_v^2)$
Inefficiency	$\mathrm{E}\left[u_{it} \middle\| u_{it}+v_{it}\right]$	$\mathrm{E}\left[u_{it} \middle\| u_i+v_{it}\right]$	$\hat{\alpha}_i - \min\{\hat{\alpha}_i\}$	$\mathrm{E}\left[u_{it} \middle\| \alpha_i+\varepsilon_{it}\right]$

Aigner et al.'s model (model *I*) is formulated as a cross sectional model and thus, ignores the panel aspects of the data. This might lead to inaccurate results due to misspecification by ignoring within-firm correlations between error terms as well as firm-specific unobserved factors. Nevertheless an important advantage of this model is that it allows random variation of inefficiency over time. In both models *II* and *III*, it is assumed that the inefficiency is time-invariant and the firm specific unobserved effects are entirely due to efficiency differences. Given that in network industries, a considerable part of the firm-specific unobserved factors are related to the network complexity and are likely to be beyond the firm's control, the inefficiency estimates are likely to be overestimated in these models. In particular, the fixed effects estimator does not allow the inclusion of time-invariant regressors, which are subsumed in the fixed effects. Thus, the inefficiency estimates are likely to be overestimated.

Finally, Greene's true random effects model (model *IV*) is an extension of Aigner et al.'s frontier model that includes an additional time-invariant random term to capture the firm-specific heterogeneity. This term is assumed to be uncorrelated with the explanatory variables. Such correlations might bias the estimated parameters of the cost function. In another paper (Farsi, Filippini and Kuenzle, 2003), we proposed an adjustment based on Mundlak (1978)'s formulation to reduce the possible biases in the true random effects model.[13] However, in the present study, our analysis (not reported here) indicates that the estimation results are fairly close with or without this adjustment. Thus, we decided to focus on the model without adjustment. With two heterogeneity terms, this model is expected to provide a better distinction between inefficiency and other unexplained variations. This advantage is especially

[13] See also Farsi, Filippini and Greene (2004) for an application of this method in railway companies.

important in network industries, in which a significant part of unobserved differences is due to time-invariant factors.

Another advantage of model *IV* is that it allows a random variation of inefficiency over time. Both economic theory and empirical evidence suggest that cost-inefficiency varies with time. New technology shocks and learning are among the reasons why inefficiency varies over time and across individuals. Using a translog production function, Alvarez, Arias and Greene (2003) have shown that even in cases when the management's efficiency is constant, the technical efficiency varies with time.[14] Moreover, the assumption of time-invariant inefficiencies is not realistic in a relatively long panel such as our sample.

As shown in table 1, in models *I*, *II* and *IV*, inefficiency is estimated using the conditional mean of the inefficiency term, proposed by Jondrow et al. (1982) and for model *III*, inefficiency is calculated as the difference of the firm's effect α_i with the minimum value of these effects.

In our comparative analysis we consider two aspects of the models' performance. The first dimension is the estimation of the cost function's coefficients. In cases such as bus companies (or in general network industries), explanatory variables and thus costs can be influenced by a number of location-related factors. For instance, increasing density of stops will increase the costs due to higher infrastructure expenditures, or a widely ramified network will lead to a higher labor and capital demand than a single-line network. Additionally, longer networks are likely to be more complex. The Hausman test is used to confirm that the firm-specific effects are correlated with the explanatory variables. In this case a fixed effects estimator would be unbiased and could thus be used as a benchmark. However, this estimator does not allow for time-invariant explanatory variables such as network length and is thus used only to confirm the expected correlation.

One can argue that models with more general error structures, such as model *IV*, have lower biases because the residuals can capture a larger part of the correlations between unobserved heterogeneity and explanatory variables, thus leaving the coefficients less affected. However, the residuals are by definition uncorrelated with explanatory variables and the extent to which they may confound such correlations with errors may significantly vary from one sample to another. Especially, since the frontier estimators are non-linear, the prediction of the biases is not straightforward. This theoretical discussion is beyond the scope of this paper. Here we rather focus on the evaluation of the models with respect to our sample.

The second aspect of the models' performance concerns the estimation of inefficiency scores. Since they are based on certain interpretation of the stochastic terms included in the model, the inefficiency estimates are considered as a separate dimension of the model's performance. A high correlation between the inefficiency estimates is usually considered as an indication of the validity of individual approaches. However, in this study the correlation between the inefficiency scores estimated from different models is used to highlight the differences across models with various specifications of inefficiency term. Different models are also compared with respect to the summary statistics of the inefficiency estimates for the whole sample. As the typically overestimated inefficiency scores obtained from the fixed

[14] In translog form this time variation is due to interaction of time-invariant inefficiency with explanatory variables.

effects (model *III*) suggest, an unbiased estimation of the cost function is not a sufficient condition for consistent estimation of inefficiency.[15] In fact, there appears to be a trade-off between a reasonable estimation of inefficiencies on the one hand and the cost function parameters (slopes) on the other. That is, models that have an acceptable performance in efficiency estimates might give a biased estimation of the slopes, while models with an unbiased estimation of the slopes are likely to produce poor estimates of inefficiency. The purpose of this paper is to study whether alternative models like true random effect model can help bridge this gap. It should be noted that the validity of the results depends on the study sample and may vary from one case to another. Therefore, our purpose is not to identify a unique all-purpose model. Rather, our comparative analysis highlights in each one of the models, the relationship of inefficiency term with other sources of heterogeneity and with the explanatory variables.

4. Data

The data used in this paper is extracted from the annual reports of the Swiss Federal Office of Statistics on public transport companies. The companies operating in main urban centers are excluded from the sample. Most of these companies operate both inner-city tramways and buses, whose functioning is quite different from rural bus transport. Our data set includes information on all the 170 rural companies operating in Switzerland during the study period. However, the data is not available for all years. In several cases lack of information is due to closure or merging with other companies. We decided to exclude the companies that have fewer than four observations.[16] That is, all the companies in the final sample have at least four years of non-missing data. Therefore companies that were closed or taken over by other companies after a short period of operation are excluded. Obviously, such companies are not comparable with other companies because their closure may have been related to their excessive costs or other peculiar reasons. Moreover, since the panel models used in this study require in one way or another the estimation of firm-specific effects, four observations per firm appears to be a reasonable minimum. We also excluded Swiss Post[17] and all its sub-contractors from the sample, because a considerable part of these companies' revenues is related to package transport and other postal services. Therefore, all the companies included in the sample are mainly involved in passenger transport.

The final data set is an unbalanced panel with 985 observations including 94 operators over a 12-year period from 1986 to 1997. The number of periods per firm varies from 4 to 12 with an average of 10.5 years. The available information includes total costs, total number of employees, network length, total numbers of bus-kilometers and passenger-kilometers as well as those of buses and seats. Table 2 provides a descriptive summary of the main variables used in the analysis.

---

[15] See Farsi and Filippini (2004) for an example of overestimation of inefficiency if it is estimated from the fixed-effects.

[16] We also dropped one observation that we suspected as erroneous because of extremely low reported total costs compared to the same company's total costs reported in other years.

[17] Swiss Post, a public company funded by the federal government, mainly in charge of mail delivery and financial services, operates public transport in about 60% of Switzerland's rural bus network.

The variables for the cost function specification were calculated as follows. Total costs *TC* are calculated as the total expenditures of the bus companies in a given year. The output *Y* is measured by the number of seat-kilometers, which is calculated by multiplying the total number of bus-kilometers by the average number of seats per bus. It should be pointed out that this calculation is based on the assumption that the number of seats in a bus does not vary considerably in a given company's fleet in a given year. This is a reasonable assumption because companies usually buy their vehicles from the same supplier. The number of seats includes both sitting and standing places. In other studies such as Windle (1988), Bhattacharyya et al. (1995) and Jha and Singh (2001), the number of passenger-kilometers is used as output. However, since in Switzerland the rural buses are rarely running at full capacity and they have to run according to the frequency set by the regulators, a considerable number of seats are likely to be empty in a typical bus travel during an off-season period. Therefore, the number of passenger-kilometers is not a representative measure of output. Alternatively, several authors like Berechman (1987), Matas and Raymond (1998) and Fazioli et al. (2003), use bus-kilometers as the output measure. Given that the average vehicle size is likely to vary across different bus companies in our sample, this measure can distort the output in favor of companies with smaller thus less costly buses. The number of seat-kilometers measures the kilometers travelled by the fleet capacity, which is not sensitive to occupancy rate and at the same time account for the variation of vehicle size across companies. Therefore, we contend that in the context of Switzerland's rural bus systems, this measure is more relevant for cost estimations.[18]

Table 2: Descriptive statistics based on 985 observations

Variables	Mean	Standard Deviation	1. Quartile	Median	3. Quartile
Total annual costs (TC) Thousand CHF	3106	4802	425	1270	3410
Output (Y) Thousand seat-kilometers	47986	85556	5715	16403	53127
Network length (N) km	43	70	13	26	55
Capital price (P_C) CHF/Seat	1343	606	927	1225	1612
Labor price (P_L) CHF per employee per year	80749	27133	66417	80872	91586
Number of seats	1067	1721	184	439	1215
Number of employees	22	35	3	9	23

- All monetary values are in 1997 Swiss Francs (CHF), adjusted for inflation using Switzerland's consumer price index.

[18] In any case, the three mentioned output measures are highly correlated in our data and our preliminary estimations suggest that the results are similar regardless of the adopted measure.

Input prices are defined as factor expenditures per factor unit. Labor price (P_L) is defined as the ratio of annual labor costs to the total number of employees.[19] Following Friedlaender and Chang (1983), the capital price (P_C) is calculated as residual cost divided by the total number of seats (both standing and sitting), where residual cost is total cost minus labor cost.[20] Unfortunately, we do not have the required data to calculate the capital stock using the capital inventory method. The use of a simple indicator is justified by the fact that the bus companies do not possess a significant stock of capital apart from the rolling stock, which could be considered as a relatively uniform stock. All the costs and prices are adjusted for inflation using the Switzerland's consumer price index and are measured in 1997 Swiss Francs. The network length is also included in the explanatory variables as an output characteristic. It is expected that due to organization and coordination problems, all other factors being constant, longer networks are expected to be more costly. Other output characteristics such as the number of stops per kilometer of network were initially considered. However, given that these variables and some of their interactions proved to be highly correlated with other explanatory variables, we decided to exclude them from the equation to avoid the possibility of multicollinearity.

5. Estimation Results

The estimation results for the four models are given in table 3. These results show that the output and input price coefficients are positive and highly significant across all models. The estimated coefficient of output from the pooled model (*I*) is particularly different from those of other models. Noting that model *I* completely ignores the panel structure of the data, its estimates are likely to be biased through omitted firm-specific factors.

Since total costs and all the continuous explanatory variables are in logarithms and normalized by their medians, the estimated first order coefficients can be interpreted as cost elasticities evaluated at the sample median. For instance, the output coefficients suggest that on average a one percent increase in seat-kilometers will increase the costs by about 0.25 to 0.73 percent depending on the adopted specification. The cost elasticity of the network length is as expected positive (α_N) and significant. This implies that the increase in network length will increase total costs. This result is consistent with previous empirical studies such as Filippini and Prioni (1994, 2003) and Windle (1988).[21]

The median cost elasticities with respect to factor prices are positive and of similar magnitude in all models. The estimated coefficient for capital price (α_{PC}) represents the share of costs attributed to capital at the median production unit, which varies from 51 to 54 percent depending on the model. This result is more or less consistent with the actual data that show a capital share of about half for the sample median. Additionally, the estimated cost function is concave in input prices

[19] Given the range of variation of salaries in the data we can safely assume that a large majority of the employees in our sample are full-time.
[20] See also Filippini and Prioni (2003) for a similar approach.
[21] It should be noted that Filippini and Prioni (2003) studied the Swiss bus systems though using a different sample in a shorter time period and without cost frontier models.

suggesting that the companies have a cost-minimizing behavior in response to changes in prices.

Table 3: Regression results

Coefficient	*Model I* **Pooled**	*Model II* **RE (ML)**	*Model III* **FE**	*Model IV* **True RE**
α_Y	0.734* (0.013)	0.326* (0.040)	0.247* (0.024)	0.351* (0.009)
α_N	0.122* (0.019)	0.244* (0.023)	0.240* (0.033)	0.264* (0.018)
α_{PC}	0.512* (0.025)	0.535* (0.008)	0.540* (0.015)	0.525* (0.006)
α_{YY}	0.079* (0.018)	-0.027* (0.009)	-0.010 (0.018)	0.014* (0.004)
α_{NN}	0.083* (0.042)	0.063 (0.054)	0.027 (0.064)	0.119* (0.016)
α_{PCPC}	-0.162* (0.041)	-0.262* (0.015)	-0.264* (0.025)	-0.278* (0.013)
α_{YN}	0.003 (0.026)	-0.026 (0.023)	-0.026 (0.032)	-0.094* (0.010)
α_{YPC}	-0.067* (0.029)	-0.095* (0.009)	-0.098* (0.016)	-0.115* (0.006)
α_{NPC}	0.129* (0.042)	0.093* (0.011)	0.102* (0.024)	0.116* (0.008)
α_T	-0.005 (0.003)	0.011* (0.001)	0.015* (0.002)	0.010* (0.001)
α_0	-0.275* (0.030)	-1.085* (0.040)	-1.458* (0.108)	0.054* (0.011)
σ_α	-	-	-	0.526* (0.011)
$\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$	0.433* (0.0004)	1.306* (0.387)	-	0.163* (0.002)
$\lambda = \sigma_u / \sigma_v$	1.038* (0.090)	9.738* (3.464)	-	0.980* (0.034)

- Standard errors are given in brackets. * means significantly different from zero at least at 95%.

The coefficient of the linear time trend is significant and positive in all models except in model *I*, which shows an insignificant effect. These results suggest an annual increase of about 1% in total costs. This result can be explained by the fact that the production technology has not much changed in bus transport. The increase in costs may be related to higher quality of service and increased security requirements.

Although the Hausman specification test's results suggest that the firm specific effects in a GLS model are correlated with the regressors, the results in table

3 indicate that most of the coefficients do not vary considerably across models *II*, *III* and *IV*. In particular the estimated coefficients of model *IV* are within a reasonable range of the unbiased estimates of the fixed effects model.

Table 4 provides a descriptive summary of the inefficiency estimates from different models. These estimates represent the relative excess cost of a given firm compared to a minimum level that would have been achieved if the firm had operated as efficiently as the 'best practice' observed in the sample.[22] In comparing different models it should again be stressed that models *II* and *III* assume constant inefficiency over time. Moreover, in these models all the unobserved firm-specific differences are interpreted as inefficiency. As expected, model *II* and *III* predict rather unrealistic inefficiency scores averaging about 1.15 and 1.46 suggesting that a typical company can decrease their costs by 68 to 77 percent. As can be seen in the table, this is even accentuated in the fixed effects model. As explained before, in this model the firm specific effects are not drawn from a single distribution, thus can accommodate a wide range of variation thus a high level of inefficiency. These high values indicate that the heterogeneity across companies is an important driver of cost differences and that neglecting it may create a substantial upward bias in inefficiency scores.

In model *I* the inefficiency estimates are in a more realistic range, with an average of 0.25 and a maximum value of 0.73. These values, though still arguably quite high, are substantially lower than those predicted by models *II* and *III*. However, the inefficiency scores obtained from model *I* are likely to be overestimated, because in fact they might capture some of the network-specific unobserved heterogeneity, which is not accounted for separately. Model *IV*, which has two separate stochastic terms for inefficiency and firm-specific heterogeneity, has inefficiency estimates of about 0.09 on average, which stands for a cost saving potential of about 9 percent.[23] Although the maximum value of 0.47 appears as excessive, this model's results suggest that in 95 percent of the cases the cost saving potential is below 15 percent. The excessive estimated inefficiency in the remaining 5 percent can be explained by statistical errors. Therefore, compared to the other models the inefficiency estimates from model *IV* are plausible and remain within a reasonable range of variation.

Table 4: Inefficiency measures

	Model I Pooled	*Model II* RE (ML)	*Model III* FE	*Model IV* True RE
Mean	0.249	1.147	1.457	0.090
Median	0.228	1.070	1.408	0.082
Maximum	0.732	2.825	3.383	0.473
95th Percentile	0.472	2.316	2.854	0.153
Minimum	0.042	0.031	0	0.019

[22] Note that cost efficiency can be alternatively defined as the optimal costs divided by actual costs that is, $CE = \exp(-u)$, where u is the relative excess cost given in table 4.
[23] This result is consistent with the average inefficiency levels reported in Dalen and Gomez-Lobo (2003) for the Norwegian bus industry.

The pair-wise correlation coefficients between the inefficiency estimates from different models are listed in table 5. In order for the correlation coefficients to be comparable, they are calculated at the firm level using 94 observations (one observation for each firm). Namely, in models with time-variant efficiency, the inefficiency score is calculated as the firm's average inefficiency score over the sample period. For models with time-variant inefficiency the correlation coefficients are also given over the total of 985 observations. As expected, in most cases the correlation coefficients are rather low, suggesting substantial differences across models. Some of these differences can be explained by large sampling errors incurred for the estimation of inefficiency for individual companies, especially in cases where the inefficiency can vary with time. This problem for cross-sectional data and short panels is documented by Horrace and Schmidt (1996), Street (2003) and Jensen (2000). Obviously, to the extent that inefficiencies remain constant over time, a longer panel can help. Nevertheless, the assumption of constant inefficiency can be unrealistic in long panels.

However, the weak correlation between efficiency estimates across different models suggests that these models differ not only with respect to individual companies' inefficiency scores but also give significantly different efficiency rankings.[24] In particular, model *IV* shows a negative correlation with models *II* and *III* and a weak positive correlation with model *I*, with all three coefficients being significantly different from zero at 5% significance level. These results suggest that when unobserved network effects are not distinguished from efficiency differences, the inefficiency estimates may be completely misleading in assessing individual companies.

As table 5 shows, there is a high correlation between inefficiency estimates from models *II* and *III* (coefficient of .987). This result can be explained by the assumption of time-invariant inefficiency in these models. Additionally, although these two models impose different assumptions regarding the firm specific effects, their estimates of the cost function coefficients are fairly close (see table 3). Nevertheless, noting that the inefficiency estimates in both models *II* and *III* represent the network-specific heterogeneity, the high correlation between these two models suggests that the estimation of unobserved network effects is robust to the distribution assumption.

Table 5. Pair-wise Pearson correlation between inefficiency estimates

	Model I Pooled	*Model II* RE (ML)	*Model III* FE	*Model IV* True RE
Model l	1			
Model II	0.496	1		
Model III	0.426	0.987	1	
Model IV	0.083 (0.371)	-0.082	-0.098	1

- The correlation coefficients have been estimated over the firms (94 observations).
- Correlation coefficient based on 985 observations is given in brackets.

[24] The rank correlations show a pattern similar to table 5. These results are omitted to avoid repetition.

Table 6 shows the estimates of scale and density economies as given in equations (3) and (4), obtained from different models. For each model only the coefficients that are significantly different from zero (at p<0.20) are considered in the calculation. The results are listed for three representative companies at the first quartile, median and the third quartile outputs. We identified the median (1st/3rd quartile) company as the company that produces the sample median (1st/3rd quartile) of the number of seat-kilometers and considered that company's corresponding network length in the estimation of density and scale economies.[25] Since the factor prices are assumed to be exogenous, they are held constant at their median values for all three cases.

Table 6. Economies of scale and density estimates

	Model I Pooled	*Model II* RE	*Model III* FE	*Model IV* True RE
ED at 1st Quartile	1.537	2.819	4.052	2.224
ED at Median	1.363	3.067	4.052	2.115
ED at 3rd Quartile	1.210	3.400	4.052	3.086
ES at 1st Quartile	1.485	1.670	2.055	1.490
ES at Median	1.336	1.754	2.055	1.713
ES at 3rd Quartile	1.013	1.859	2.055	1.879

The results listed in table 6 show a considerable amount of variation between different models. As can be seen in this table, both economies of density and scale are greater than one in all three representative cases, suggesting the presence of unexploited economies in most companies in the sample. In particular, the relatively high values of density economies indicate that a more intensive use of a given network would considerably lower the average cost per seat-kilometer. However, it should be noted that the intensity of demand in a given network is beyond the company's control. An increase in output usually requires some extension in the network, which can be represented by scale economies.

The estimated scale economies from all models also suggest the existence of considerable potential for cost saving through extending the networks. As expected, the economies obtained from an increase in output density in a given network (density economies) are relatively higher than those gained by extending a company's network (scale economies). The presence of unexploited scale economies in all three representative cases suggests that most companies are smaller than the optimal size at which such economies are fully utilized. The small size of rural bus companies in Switzerland is related to the development of this industry that has been historically associated with the growth of small and fragmented user communities.

The high variation of scale and density economies across various models (see table 6) can be partially explained by the models' differences with respect to the

[25] We considered alternative definitions for representative companies, e.g. median of both output and network. However, the results are mainly the same insofar as the following discussion is concerned.

unobserved network effects. If these effects are correlated with explanatory variables (such as output and network length) the values obtained from the fixed effect model (model *III*) are unbiased and those of the other three models are biased. Particularly, the values estimated by the pooled model (model *I*) are likely to be biased downward. This model's estimates are comparable to those reported by Filippini and Prioni (2003), who applied a cross-sectional cost-frontier model to a panel data of 34 bus companies in Switzerland. In one of their specifications, the median company's size (about twice as our sample median) has been found to be very close to the optimal size. These results suggest that ignoring the unobserved firm-specific effects can bias the estimated coefficients. In fact such biases are driven by possible correlation of unobserved effects with output and network length. For instance it is plausible that larger networks are more complex in terms of unobserved factors, thus more costly. Such correlations are likely to be positive, thus lead to an underestimation of scale economies.

Theoretically, an unbiased estimation of scale economies can be obtained from the fixed effect model (*III*). However, because of the large number of parameters in this model (at least as many as the companies in the sample), the precision of the results depends on the number of periods in the sample and the companies' output variation over the sample period. For instance, as shown in table 6, model *III* predicts a constant value of scale economies for all levels of output. This unrealistic result can be explained by the relatively high standard errors in the fixed-effect model, which implies that this model cannot detect the variation of scale economies with output. In fact, as indicated in table 3, the coefficients of most of the second order terms are statistically insignificant in this model.

Interestingly, model *I* predicts a decreasing scale and density economies with output, which appears to be consistent with the common perception that sources of scale economies are exhaustible. However, models *II* and *IV* suggest that if the unobserved heterogeneity is taken into account, this result may be reversed. As table 6 shows, according to these two models, the unexploited scale and density economies are greater in relatively large companies in the sample. This result can be explained by some of the special features in relatively small network industries: Noting that the smallest companies in our sample are bus companies with a single line and a few employees, the potential gains of increasing the size are limited to savings in distributing the same fixed costs over a higher output. On the other hand, large companies have complex multiple-line networks. By increasing their size, such companies can benefit not only from savings in the fixed costs but also from a better possibility of reallocation of input factors over the network, thus reducing their variable costs.

6. Conclusions

The application of alternative cost frontier models to a panel of rural bus companies in Switzerland indicates that the inefficiency estimates are sensitive to the adopted model. From a methodological point of view, the results largely depend upon how the unobserved firm-specific heterogeneity among firms is modeled. Our comparative analysis suggests that models that do not distinguish between unobserved network effects and inefficiency can overestimate the inefficiency scores. In particular, if the inefficiency estimates are derived from the firm-specific effects (cf. Schmidt and Sickles, 1984 and Pitt and Lee, 1981), they include an important part of the unobserved exogenous factors related to the network. Our sample shows that such factors can account for a considerable part of cost differences, thus bias the inefficiency estimates to unrealistically high values.

This paper also highlights possible differences in cost function coefficients across models. A (pooled) cross-sectional model does not account for network heterogeneity. Since such heterogeneity is likely to be correlated with some of the explanatory variable, this model can give biased coefficients. A fixed effect model can solve the heterogeneity bias in the coefficients. However, because of the large number of parameters (incidental parameters problem), this model leads to relatively large estimation errors in the cost function coefficients.

This study suggests that an econometric specification that includes separate stochastic terms for firm-specific effects and inefficiency can improve the estimations regarding both inefficiencies and slopes. We considered a random-constant cost frontier model ("true" random effects model) proposed by Greene (2003, 2004). The results indicate that unlike other conventional methods, this model estimates the inefficiencies within a realistic range. Moreover, the main coefficients of the cost function are fairly close to the unbiased estimators obtained from the fixed effects model.

The results also indicate that the unexploited scale economies might be greater for relatively large companies, which can benefit from better possibilities of reallocation over larger networks. Such effects could be masked by unobserved network factors, which if neglected could lead to inaccurate results. It should be pointed out that the results of this paper are valid for the specific sample used here and cannot be directly extended to other cases. However, these results underscore the importance of modeling unobserved firm-specific heterogeneity in efficiency measurement of network industries.

From a policy point of view, this study suggests that the "true" random effects model could be a valuable alternative to the other models to set a benchmark for regulating network industries. However, it has to be emphasized that a mechanical use of any of these models in regulation could be misleading. Since each industry has its specific cost characteristics that are not equally well reproduced by these models, a careful analysis of the cost structure of the industry under consideration has to be performed before establishing a reliable benchmark. Consequently, these models should be used as one among different instruments in the assessment of subsidy requests.

References

Alvarez, A., C. Arias, and W. H. Greene (2003): "Fixed Management and Time-Invariant Technical Efficiency in a Random Coefficients Model". *Working Paper*, Department of Economics, Stern School of Business, New York University.

Aigner, D. J., C. A. K. Lovell, and P. Schmidt (1977): "Formulation and Estimation of Stochastic Frontier Production Function Models". *Journal of Econometrics*, 6, 21–37.

Battese, G. E. and T. J. Coelli (1995): "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data". *Empirical Economics*, 20, 325–332.

Berechman, J. (1987): "Cost Structure and Production Technology in Transit: An Application to the Israeli Bus Transit Sector". *Regional Science and Urban Economics*, 17, 519–534.

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

Bhattacharyya, A., S. C. Kumbhakar, and A. Bhattacharyya (1995): "Ownership Structure and Cost Efficiency: A Study of Publicly Owned Passenger-Bus Transportation Companies in India". *Journal of Productivity Analysis*, 6, 47–61.

Chambers, R. G. (1988): *Applied Production Analysis*, Cambridge University Press.

Caves, D. W., L. R. Christensen, and M. W. Tretheway (1984): "Economies of Density versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ". *Rand Journal of Economics*, 15 (4): 471–489.

Cornwell, C., P. Schmidt, and R. Sickles (1990): "Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels". *Journal of Econometrics*, 46, 185–200.

Dalen, D. M. and A. Gomez-Lobo (2003): "Yardsticks on the Road: Regulatory Contracts and Cost Efficiency in the Norwegian Bus Industry". *Transportation*, 30, 371–386.

Farsi, Mehdi and Massimo Filippini (2004): "Regulation and Measuring Cost Efficiency With Panel Data Models: Application to Electricity Distribution Utilities". *Review of Industrial Organization*, forthcoming.

Farsi, Mehdi, Massimo Filippini and William Greene (2004): "Efficiency Measurement in Network Industries: Application to the Swiss railway Companies". *Working Paper* 32, Centre for Energy Policy and Economics, Swiss Federal Institute of Technology, Zurich, Switzerland.

Farsi, Mehdi, Massimo Filippini and Michael Kuenzle (2003): "Unobserved Heterogeneity in Stochastic Cost Frontier Models: A Comparative Analysis". *Working Paper* 03-11, Department of Economics, University of Lugano, Switzerland.

Fazioli, R., M. Filippini and P. Prioni (1993): "Cost Structure and Efficiency of Local Public Transport: The Case of Emilia Romagna Bus Companies". *International Journal of Transport Economics*, 3, 305-324.

Fazioli, R., M. Filippini, and M. Kuenzle (2003): “Valutazione dell’efficienza delle compagnie di bus italiane e svizzere”, In: *L’efficienza nei servizi pubblici, BANCA D’ITALIA*, 171–210.

Filippini, M. and P. Prioni (1994): “Is Scale and Cost Inefficiency in the Swiss Bus Industry a Regulatory Problem? Evidence from a Frontier Cost Approach”. *Journal of the Economics of Business*, 2, 219–231.

Filippini, M. and P. Prioni (2003): “The Influence of Ownership on the Cost of Bus Service Provision in Switzerland - An Empirical Illustration”. Applied Economics 35 (6): 683–690.

Fraquelli, G., M. Piacenza and G. Abrate (2001): “Costs and Technology of Public Transit Systems in Italy: Some Insights to Face Inefficiency”. *Working Paper* 12/2001, Serie Speciale in Collaborazione con HERMES, Torino, Italy.

Friedlaender, A. and S. J. W. Chang (1983): “Productivity Growth in the Regulated Trucking Industry”. *Research in Transport and Economics*, 1, 149–84.

Greene, W. H. (2002a): “Alternative Panel Data Estimators for Stochastic Frontier Models”. *Working Paper*, Department of Economics, Stern School of Business, New York University.

Greene, W. H. (2002b): “Fixed and Random Effects in Stochastic Frontier Models”. *Working Paper*, Department of Economics, Stern School of Business, New York University.

Greene, W. H. (2003): “Distinguishing Between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organizations Panel Data on National Health Care Systems”. *Working Paper*, Department of Economics, Stern School of Business, New York University.

Greene, W. H. (2004): “Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model”. *Journal of Econometrics*, forthcoming.

Heshmati, Almas and Subal C. Kumbhakar (1994). “Farm heterogeneity and technical efficiency: some results from Swedish dairy farms”. *Journal of Productivity Analysis*, 5 (1): 45-61.

Horrace, W. C. and P. Schmidt (1996): “Confidence Statements for Efficiency Estimates from Stochastic Frontier Models”. *Journal of Productivity Analysis*, 7, 257–82.

Jensen, U. (2000): “Is it Efficient to Analyse Efficiency Rankings?” *Empirical Economics*, 25, 189–208.

Jha, R. and S. K. Singh (2001): “Small is Efficient: A Frontier Approach to Cost Inefficiencies in Indian State Road Transport Undertakings”. *International Journal of Transport Economics*, 18 (1): 95–114.

Jondrow, J., C. A. K. Lovell, I. S. Materov, and P. Schmidt (1982): “On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model”. *Journal of Econometrics*, 19 (2-3): 233–238.

Kumbhakar, Subal C. (1991). “Estimation of technical inefficiency in panel data models with firm- and time-specific effects”. *Economics Letters*, 36 (1991): 43-48.

Kumbhakar, S. C. and C. A. K. Lovell (2000): *Stochastic Frontier Analysis*. Cambridge University Press.

Matas, A. and J.-L. Raymond (1998): “Technical Characteristics and Efficiency of Urban Bus Companies: The Case of Spain”. *Transportation*, 25, 243–263.

Mundlak, Y. (1978): “On the Pooling of Time Series and Cross Section Data”. *Econometrica*, 46 (1): 69–85.

Piacenza, M. (2002): “Regulatory Constraints and Cost Efficiency of the Italian Public Transit Systems: An Exploratory Stochastic Frontier Model”. Working Paper N. 2/2002, Serie Speciale in Collaborazione con HERMES, Torino, Italy.

Pitt, M., and L. Lee (1981): “The Measurement and Sources of Technical Inefficiency in Indonesian Weaving Industry”. *Journal of Development Economics*, 9, 43-64.

Schmidt, P. and R. Sickles (1984): “Production Frontiers with Panel Data”. *Journal of Business and Economic Statistics*, 2 (4): 367–374.

Street, A. (2003): “How Much Confidence Should We Place in Efficiency Estimates?” *Health Economics*, 12, 895–907.

Wang, H. and P. Schmidt (2002): “One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels”. *Journal of Productivity Analysis*, 18, 129–144.

Windle, R. J. (1988): “Transit Policy and the Cost Structure of Urban Bus Transportation”. In: J. S. Dogson and N. Tophan (Eds.), *Bus Deregulation and Privatisation*, Chapter 6, Brookfield USA.