EFFICIENCY MEASUREMENT IN NETWORK INDUSTRIES: APPLICATION TO THE SWISS RAILWAY COMPANIES

Mehdi Farsi† Massimo Filippini† William Greene‡

† Center for Energy Policy and Economics
Federal Institute of Technology
ETH Zentrum, WEC, 8092 Zurich, Switzerland
and
Department of Economics, University of Lugano
Via Ospedale 13, 6900 Lugano, Switzerland

‡ Department of Economics, Stern School of Business
New York University
44 West 4th St., New York, NY 10012, USA

May 2004

ABSTRACT

This paper examines the performance of several panel data models to measure cost and scale efficiency in network industries. Network industries are characterized by a high degree of heterogeneity, much of which is network-specific and unobserved. The unaccounted-for heterogeneity can create bias in the inefficiency estimates. The stochastic frontier models that include additional firm-specific effects, such as the random-constant frontier model proposed by Greene (2004), can control for unobserved network effects that are random but time-invariant. In cases like railway networks the unobserved heterogeneity is potentially correlated with other exogenous, but observed, factors such as network size and density. In such cases the correlation with explanatory variables may bias the coefficients of the cost function in a random-effects specification. However, these correlations can be integrated into the model using Mundlak's (1978) formulation. The unobserved network effects and the resulting biases are studied through a comparative study of a series of stochastic frontier models. These models are applied to a panel of 50 railway companies operating over a 13-year period in Switzerland. Different specifications are compared regarding the estimation of both cost frontier coefficients and inefficiency scores.

1. INTRODUCTION

The railroad system in Switzerland consists of two sectors. The first sector includes the international and inter-regional transports. This sector is monopolized by the Swiss Federal Railways, which operates more than half of the railway networks in Switzerland. The second sector provides regional and local transport services that account for about a third of Switzerland's railway passengers. This sector consists of 57 small private and regulated companies.[1] These companies have regional monopoly in that they have an exclusive access to their assigned networks and different companies' networks do not overlap with each other. Most of these companies have long-term contracts and are strongly subsidized by the cantons and the federal government. Given that most of Swiss cantons have financial problems, there is an increasing interest in the possibility of reducing the allocated subsidies by improving productive efficiency.

The measurement of cost and scale efficiency in railway industry has been an important policy issue for the past several years in Switzerland. However, since these companies operate in different networks and environments, any method based on cost comparison has been subject to criticism.[2] A high level of output heterogeneity is a general characteristic of network industries. Networks with different shapes have different organization and coordination problems, thus different costs. For instance, in the railway sector the production of 100 train-kilometers on a simple linear network is less costly than the same output in a *Y*-shaped network. Other factors such as the density of stops can also affect the costs.

---

[1] See Filippini and Maggi (1993) for more information.

[2] For instance, Filippini and Maggi (1993) estimated the efficiency level of the Swiss railways companies using Corrected OLS method, which in presence of a high heterogeneity in the production process, can lead to inaccurate inefficiency estimates.

Furthermore, different environmental characteristics influence the production process and therefore the costs. For instance, railway operation is more costly in a mountainous region than in a flat area. In general, the information is not available for all output and environmental characteristics. Many of these characteristics are therefore omitted from the cost function specifications.

Unobserved firm-specific heterogeneity can be taken into account with conventional fixed or random effects in a panel data model. However, in these models all the unobserved time-invariant heterogeneity is considered as inefficiency. In order to distinguish heterogeneities such as external network effects from cost efficiency, Greene (2003, 2004) proposed an approach that integrates an additional stochastic term representing inefficiency in both fixed and random effects models. As shown in those papers, assuming that the inefficiency term follows a distributional form, both models can be estimated. In this paper we use a 'true random-effects' model, which is a random-constant frontier model, obtained by combining a conventional random-effects model with a skewed stochastic term representing inefficiency. The extended model includes separate stochastic terms for latent heterogeneity and inefficiency. Therefore, it should in principle, be able to provide better estimates of inefficiency. In addition, since many of the unobserved factors, especially those related to the network's shape, are likely to be correlated with the output and perhaps other explanatory variables, the random-effect estimators of the cost function coefficients could be biased. To overcome this shortcoming, the 'true random-effects' model has been adjusted for correlation between unobserved heterogeneity and explanatory variables using Mundlak's (1978) formulation.[3]

[3] The application of Mundlak's adjustment in frontier models has been proposed by Farsi et al. (2003).

The success of these recently developed panel data models[4] could lend certain support to the application of benchmarking methods in the regulation of strongly heterogeneous network industries such as Swiss railways. Provided that they can sufficiently control for the unobserved heterogeneity across firms, these methods can be used to estimate an order of magnitude for the sector or individual companies' cost-inefficiency. In addition, in the case of railway networks, such analyses can be used to evaluate the bidding offers for the future tendering processes predicted by the new public transport policy.[5]

The purpose of this paper is to study the potential advantages of these extended models in an application to Switzerland's railway companies. In particular, our eventual interest is in models that can exploit the advantage of a fixed-effects model to have an unbiased estimate of the cost function without compromising the estimates of inefficiency scores. The models are estimated for a sample of 50 railway companies operating in Switzerland from 1985 to 1997. The alternative models are compared regarding the cost function slopes and inefficiency estimates. The conventional FE estimators of the cost function coefficients are assumed to be unbiased, thus used as a benchmark to which other models are compared. For the inefficiency estimates, the correlation between different models and the effect of econometric specification have been analyzed. The results suggest that the inefficiency estimates are substantially lower when the unobserved network effects are taken into account.

---

[4] Greene (2003), Farsi et al. (2003) and Alvarez et al. (2003) are examples of application of such models in efficiency analysis.

[5] In line with the EU policy the Swiss government has introduced important regulatory reforms in the public transport system. The new policy act predicts a tendering process for the provision of regional railway services. The new 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 these measures lead to efficient production.

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 5 provides the conclusions.

2. MODEL SPECIFICATION

A railway company can be considered as an aggregate production unit that operates in a given network and transforms labor and capital services and energy into units of transport services such as passenger-kilometers of public transport and ton-kilometers of freight. Given the extremely high number and types of different transport services, the measure of output requires an aggregation of outputs in one way or another.[6] A practical way of getting around this approximation is to include output characteristics such as network length or average haul in the model. Different strategies have been used in the literature. Caves et al. (1985) used passenger-miles and freight ton-miles as output, and controlled for the average lengths of trip for freight and passengers and the number of route miles as output characteristics. Filippini and Maggi (1993) have considered a single-output production function with the number of wagon-kilometers as a measure of output and included the network length in their model specification. In their international analysis, Cantos et al. (1999) considered the aggregate number of passenger-kilometers and ton-kilometers as two outputs. Todani (2001) considered three types of wagon-miles (high-valued, bulk and others) as three main outputs and accounted for average length of haul and the number of road miles as output characteristics.

---

[6] In the case of railways each relation between any two points in the space could be defined as an output type. From a practical point of view it is not possible to estimate a multi-product cost function with so many outputs. Therefore, an aggregation process is inevitable.

In this paper a two-output production process is assumed. The outputs are transported passengers measured by the total number of passenger-kilometers in a given year, and the transported freight measured as the aggregate number of ton-kilometers. The length of network is included in the model as output characteristics. Three input factors are considered: labor, capital and energy. A total cost function has been considered.

Based on the above specification the total cost frontier can be represented by the following cost function:

$$TC=f\,(Y, Q, N, P_K, P_L, P_E, \boldsymbol{d_t}) \qquad (1)$$

where TC is the total annual costs; Y and Q are the numbers of passenger-kilometers and ton-kilometers respectively; P_K , P_L and P_E are respectively the prices of capital, labor and energy; N is the length of network and $\boldsymbol{d_t}$ is a vector including 12 year dummies from 1986 to 1997 (year 1985 is the omitted category). The year dummies capture the cost changes associated with technical progress as well as other unobserved year-specific factors.[7]

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.[8] Mainly, this function must be non-decreasing, concave, linearly homogeneous in input prices and non-decreasing in output. To estimate the cost function (1), a Cobb-Douglas (log-linear)

---

[7] In the cost function estimations it is common to use a linear trend for technical progress. However, our preliminary regressions indicated that the time-variation of costs is strongly non-linear. In fact there is a gradual increase in the beginning of the sample period followed by a decrease in costs. These variations can be explained by many unobserved factors (such as changes in collective labor contracts or seasonal composition of the demand) that change uniformly across companies.

[8] For more details on the functional form of the cost function see Cornes (1992), p.106.

functional form is employed.[9] The concavity assumption is automatically satisfied in this functional form. The linear homogeneity restriction can be imposed by normalizing the costs and prices by the price of one of the input factors. Here we considered the energy as the numeraire good. The other theoretical restrictions are verified after the estimation. The cost function can therefore be written as:

$$\ln\left(\frac{TC_{it}}{P_{E_{it}}}\right) = \alpha_0 + \alpha_Y \ln Y_{it} + \alpha_Q \ln Q_{it} + \alpha_N \ln N_{it} + \alpha_S \ln S_{it} + \alpha_K \ln \frac{P_{K_{it}}}{P_{E_{it}}} + \alpha_L \ln \frac{P_{L_{it}}}{P_{E_{it}}} + \sum_{t=1986}^{1997} \alpha_t d_t + \alpha_i + \varepsilon_{it} \quad (2)$$

with $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T_i$

Subscripts *i* and *t* denote the company and year respectively, α_i is a firm-specific effect and ε_{it} is an *iid* error term. As we will explain in the next section, in the recent models proposed by Greene (2004), the stochastic term ε_{it} is composed of two parts: a skewed component representing inefficiency and a symmetric part for the random noise.

3. ECONOMETRIC MODELS

Stochastic frontier models have been subject of a great body of literature resulting in a large number of econometric models to estimate cost functions. Kumbhakar and Lovell

[9] As an alternative form we also evaluated the possibility of applying a translog functional form that can account for variation of scale economies with output. However, we decided to exclude this model because our study focuses on the efficiency estimates rather than scale economies. Moreover, the translog model requires a relatively large number of parameters, which creates certain numerical problems in the simulated likelihood maximization for the random-constant model.

(2000) provide an extensive survey of this literature. The main models used in this paper are based on Greene's (2004) extension of the original frontier approach proposed by Aigner et al. (1977). In this framework, ε_{it} as given in specification (2), is assumed to be a composite stochastic term with a normal-half-normal distribution, including both idiosyncratic effects and inefficiencies. The additional firm-specific term α_i (see equation 2) represents the unobserved network heterogeneity and is assumed to have a normal distribution. This model is actually a stochastic frontier model in line with Aigner et al. (1977) with a random constant. This model is developed by Greene (2004) and is referred to as a "true" random-effects model.[10] The estimation method is based on simulated maximum likelihood.

The results are compared with other alternative models such as the fixed-effects model proposed by Schmidt and Sickles (1984) and the random-effects model proposed by Pitt and Lee (1981). Both these models are covered by the general form given in (2) with the difference that in the former model α_i is a fixed effect and ε_{it} is a zero-mean error term with no distribution restriction, and in the latter (Pitt and Lee) model α_i is a random effect with half-normal (or truncated normal) distribution and ε_{it} is a normal random error term.

A summary of the five models used in the paper is given in table 1. The first model is a fixed effects (FE) model. In this model the firm-specific effects are considered as constant parameters that can be correlated with the explanatory variables. The coefficients are estimated through "within-firm" variations and therefore, are not affected by heterogeneity bias.[11] In the cost frontier literature the inefficiency scores are estimated as the distance from

---

[10] The name "true" is chosen to show that the model keeps the original frontier framework and the extension is done only by including an additional heterogeneity term.

[11] The term "heterogeneity bias" was used by Chamberlain (1982) to refer to the bias induced by the correlation between individual effects and explanatory variables in a random-effects model. See also Baltagi (2001) for an extensive discussion of fixed-effects (within) estimators.

the firm with the minimum estimated fixed effect, that is $\hat{\alpha}_i - \min\{\hat{\alpha}_i\}$, as proposed by Schmidt and Sickles (1984).

Table 1. Econometric specifications of the stochastic cost frontier

	Model I **FE**	*Model II* **RE**	*Model III* **Pooled**	*Model IV* **True RE**	*Model V* **True RE with Mundlak adjustment**
Firm-specific component α_i	Constant	Half-normal $N^+(0, \sigma_\alpha^2)$	None	$N(0, \sigma_\alpha^2)$	$\alpha_i = \gamma \bar{X}_i + \delta_i$ $\bar{X}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} X_{it}$ $\delta_i \sim N(0, \sigma_\delta^2)$
Random error ε_{it}	$iid\ (0, \sigma_\varepsilon^2)$	$iid\ (0, \sigma_\varepsilon^2)$	$\varepsilon_{it} = u_{it} + v_{it}$ $u_{it} \sim N^+(0, \sigma_u^2)$ $v_{it} \sim N(0, \sigma_v^2)$	$\varepsilon_{it} = u_{it} + v_{it}$ $u_{it} \sim N^+(0, \sigma_u^2)$ $v_{it} \sim N(0, \sigma_v^2)$	$\varepsilon_{it} = u_{it} + v_{it}$ $u_{it} \sim N^+(0, \sigma_u^2)$ $v_{it} \sim N(0, \sigma_v^2)$
Inefficiency	$\hat{\alpha}_i - \min\{\hat{\alpha}_i\}$	$E\left[\alpha_i \mid \omega_{i1}, \omega_{i2}, \ldots\right]$ with $\omega_{it} = \alpha_i + \varepsilon_{it}$	$E\left[u_{it} \mid u_{it} + v_{it}\right]$	$E\left[u_{it} \mid \alpha_i + \varepsilon_{it}\right]$	$E\left[u_{it} \mid \delta_i + \varepsilon_{it}\right]$

Model *II* is a random effects (RE) model proposed by Pitt and Lee (1981), which is estimated using the maximum likelihood method. The firm's inefficiency is estimated using the conditional mean of the inefficiency term proposed by Jondrow et al. (1982),[12] that is: $E\left[\alpha_i \mid \omega_{i1}, \omega_{i2}, \ldots\right] = E\left[\alpha_i \mid \bar{\omega}_i\right]$ where $\omega_{it} = \alpha_i + \varepsilon_{it}$ and $\bar{\omega}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} \omega_{it}$. The important limitation of this model is the assumption that the firm-specific stochastic term α_i, which represents the firm's inefficiency, is uncorrelated with the explanatory variables. Although it could be reasonable to assume that the firm's cost-inefficiency is not correlated with exogenous variables,[13] the firm-specific stochastic term may contain other unobserved environmental

12 See also Greene (2002b).

13 Note that here the cost-efficiency does not include scale efficiency.

factors, which may be correlated with explanatory variables. Moreover, in both models (*I* and *II*), inefficiency indicators may include unobserved environmental factors, thus may overstate the firms' inefficiency. There are however two factors that may exacerbate this problem in the FE model. First, unlike the RE model, the firm-specific effects do not follow a single distribution, thus can have a relatively wide range of variation. Secondly, these effects can be correlated with the explanatory variables, thus can also capture the heterogeneity factors that are correlated with the regressors. Whereas in the RE model in which the firm-specific effects are by construction uncorrelated with the regressors, these factors are suppressed at least partially through the "between" variations, into the regression coefficients.

In the first two models (*I* and *II*), the firm's inefficiency is assumed to be constant over time, thus captured by the firm-specific effects, while in other models inefficiency can vary across years. Model *III* is a pooled frontier model in that the sample is considered as a cross-section and its panel aspect is neglected. The random error term is divided into two components: a normal error term v_{it} capturing the noise 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 et al. (1977). The firm's inefficiency is estimated using the conditional mean of the inefficiency term $\mathrm{E}\left[u_{it} \middle| u_{it} + v_{it}\right]$, proposed by Jondrow et al. (1982).

Models *IV* and *V* are extensions to model *III* that include an additional firm-specific random effect (α_i) to represent the unobserved heterogeneity among firms. Model *IV* is Greene's (2002a,b) true RE model.[14] In this model it is assumed that the unobserved cost differences across firms that remain constant over time, are driven by network-related unobserved characteristics rather than inefficiency. Given the relatively long period covered

[14] This model is a special case of a stochastic frontier model with random parameters (in this case random intercept).

in the data (12 years on average), this is a realistic assumption. The inefficiency term is assumed to be an *iid* random variable with half-normal distribution. This implies that the inefficiency is not persistent and each period brings about new idiosyncratic elements thus new sources of inefficiency. This is a reasonable assumption particularly in industries that are constantly facing new technologies. Therefore there are two justifications for such a specification in network industries: The first one is a practical assumption that persistent cost differences are related to unobserved heterogeneity across networks and the second one is based on the conjecture that the sources of inefficiency in network industries are dominated by new technology shocks and the incomplete adaptation of managers facing them.

Model *V* is an extension of model *IV* that uses Mundlak's (1978) specification to account for the potential correlation of unobserved network heterogeneity with the explanatory variables. Mundlak's adjustment[15] can be written as an auxiliary regression given by:

$$\alpha_i = \gamma \bar{X}_i + \delta_i \qquad \bar{X}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} X_{it} \ , \ \delta_i \sim N(0, \sigma_\delta^2) \qquad (3)$$

where X_{it} is the vector of all explanatory variables and γ is the corresponding vector of coefficients. Equation (3) actually divides the firm-specific stochastic term into two components: The first part can be explained by exogenous variables, whereas the remaining component (δ_i) is orthogonal to explanatory variables. The advantage of this model is that it allows for a time-variant inefficiency term while minimizing the heterogeneity bias. The heterogeneity bias can be avoided to the extent that the auxiliary equation can capture the correlations.[16]

---

[15] See also Hsiao (2003), pp. 44-46, for an extensive discussion of Mundlak's formulation.

[16] Notice that the heterogeneity bias would be zero if the error term (ε_{it}) were symmetric. In this case it can be shown that Mundlak's adjustment turns the estimator into a within estimator.

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 railway companies the operating costs are affected by network characteristics, which may be correlated with explanatory variables such as network's size and input factor prices. For instance, larger networks are more likely to have more complex shapes. Denser networks are usually located in areas with higher population density, where wages are relatively high. Such relationships imply a positive correlation between the output level and labor price with the network complexity, which is not fully captured by the included factors in the model. The Hausman test is used to confirm that the firm-specific effects are correlated with the explanatory variables. In this case the FE estimators (model *I*) are unbiased, thus provide a benchmark to which other models can be compared.

The second aspect of the models' performance concerns the inefficiency estimates. It is important to note that the consistency of slopes (coefficients) does not necessarily imply that inefficiency estimates are unbiased. Interestingly, the empirical results suggest that there is a trade-off in estimations. Namely, models (like the FE model) with a good performance on slopes have strongly biased inefficiency estimates.[17] Roughly speaking, the heterogeneity bias may be suppressed into the slopes as it appears in the RE model, or into the efficiency estimates as observed in the FE model. Farsi et al. (2003) provide a discussion on this issue. The results of that study on a sample of nursing homes suggest that Mundlak's formulation can be helpful to reduce the heterogeneity bias in both slopes and inefficiency estimates at the same time. In this paper we use a similar approach to study if such a conclusion can be applied to a network industry.

One should bear in mind that the inefficiency estimation requires a certain interpretation of the stochastic terms in the model. In the frontier literature, starting from Aigner et al.

---

[17] See Farsi et al. (2003) for a discussion of this point.

(1977), it is commonly accepted that the skewed stochastic term with a certain distribution represents inefficiency. For instance a half-normal distribution through its zero mode, implies that any company is most likely to be completely efficient. Moreover, implicit in this model is the assumption that inefficiency is uncorrelated with all exogenous variables and also with the idiosyncratic variations reflected in the symmetric error term.[18] This is a legitimate and helpful assumption from a practical point of view. In fact, through this assumption all the inefficiencies that are somehow related to exogenous variables such as factor prices and output are excluded from the firm's productive inefficiency. Later studies like Cornwell et al. (1990) and Battese and Coelli (1992) extended the original framework to include exogenous variables in the distribution of the inefficiency term. However, in this paper we maintain the original assumption such that the efficiency measures are restricted to the sources that are completely uncorrelated with all exogenous variables, which by definition are beyond the firm's control. The only exception is the FE model (model *I*) that allows any correlation of inefficiency scores. Furthermore, we assume that the inefficiency can vary over time, thus for the inefficiency estimates we focus on models *III*, *IV* and *V*.

4. DATA

The data set 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 inner-city tramways and buses, whose functioning is quite different from trains. We also excluded one other company whose extremely low total costs and energy expenses suggest the possibility of a

[18] Here, cost inefficiency is defined as the excess costs due to the firm's technical problems or to suboptimal allocation of resources. Thus, scale inefficiencies, which are related to suboptimal output, are excluded.

reporting error. The final sample includes 50 railway companies over a 13-year period from 1985 to 1997. The sample is an unbalanced panel with number of periods (T_i) varying from 1 to 13 and with 45 companies with 12 or 13 years, resulting in 605 observations in total.[19] The available information for any given year includes total costs, labor and energy expenses separately, total number of employees, the quantity of consumed electricity, network length, total number of seats, and total number of train-kilometers, passenger-kilometers and ton-kilometers.

Capital costs are calculated as the residual costs after deducting the labor and energy expenses from the total costs. These costs are mainly related to equipment and materials. Total number of seats is used as a proxy for capital stock.[20] Thus, the capital price is calculated as the residual expenses per seat. The passenger and freight outputs are respectively measured by the number of passenger-kilometers and ton-kilometers. In Switzerland, each railway company is required to run a certain minimum number of trips per day for any given connection, specified by the cantonal regulators. Therefore, the number of train-kilometers or wagon-kilometers could be also an appropriate measure of passenger output. However, in order to be consistent with the recent literature[21] and also given that there is a high correlation between train-kilometers and passenger-kilometers (a correlation coefficient of 0.97 in our sample) we adopted the number of passenger-kilometers and ton-kilometers. All the costs and prices are adjusted for inflation using the Switzerland's global price index and are measured in 1997 Swiss Francs.

---

[19] The average number of periods in the sample is 12 years. For 37 companies, the data are available for 13 years. Eight other companies have 12 years available. The number of years available for the remaining five companies is respectively 1, 3, 7, 7 and 10.

[20] See Filippini and Prioni (2003) for a similar approach.

[21] Some recent examples are Mancuso and Reverberi (2003), Estache et al. (2002), Cantos et al. (1999) and Banos-Pino et al. (2002).

Table 2. Descriptive statistics (605 observations)

	Mean	Standard Deviation	Median	Min.	Max.
Total annual costs (TC) CHF million	26.73	49.88	8.83	2.12	307.43
Passenger output (Y) $\times 10^6$ passenger-kms	30.80	55.10	10.00	0.41	311.00
Average cost (CHF per passenger-km)	1.20	0.76	1.09	0.33	5.98
Goods output (Q) $\times 10^6$ ton-kilometers	10.20	52.70	0.27	0.00015	477.00
Network length (N) (km)	39.43	56.64	22.82	3.90	377.00
Capital price (P_K) per seat (CHF '000)	4.53	2.13	4.03	1.04	14.47
Average labor price (P_L) per employee per year (CHF '000)	86.05	6.48	86.09	60.93	104.93
Energy (electricity) price (P_E) CHF/ kWh	0.157	0.023	0.158	0.076	0.265

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

Table 2 provides a descriptive summary of the main variables used in the analysis. As it can be seen in this table, the total costs show a high variation in the sample. The average cost of a passenger-kilometer varies from 0.3 to about 6 Swiss Francs. There is also a considerable variation in input prices and both outputs in the sample. Given the importance of within variations in most panel data models (especially the fixed-effect model), it is helpful to distinguish these variations from the variations across companies. Table 3 gives a summary of "within" and "between" variations for the main variables used in the regressions. As it can be seen in this table, the dependent variable and most explanatory variables show a fairly considerable amount of within variation, supporting the use of a fixed-effect model. As expected, the within variation of network length is relatively low (limited to 7 percent).

Table 3. Within and between variations (50 companies and 12 years on average)

	Mean	Standard Deviation			Fraction of within variation
		Overall	Between	Within	
$\ln\left(\frac{TC}{P_E}\right)$	11.31	1.10	1.12	0.15	0.14
ln (Y)	16.32	1.34	1.34	0.12	0.09
ln (Q)	12.49	2.72	2.78	0.61	0.22
ln (N)	3.20	0.91	0.93	0.06	0.07
$\ln\left(\frac{P_K}{P_E}\right)$	10.18	0.44	0.39	0.19	0.43
$\ln\left(\frac{P_L}{P_E}\right)$	13.22	0.16	0.13	0.10	0.62

- For each variable (X) the between standard deviation is based on companies' average values that is: $\bar{X}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} X_{it}$; and the within standard deviation is based on deviations from companies' averages ($X_{it} - \bar{X}_i$). The overall and within statistics are calculated over 605 company-years and the between statistics are calculated over 50 companies. The fraction of within variation is calculated as the ratio of within to overall standard deviation.

5. ESTIMATION RESULTS

The estimation results for the five models are given in table 4. These results show that the output and input price coefficients are positive and highly significant across all models. The estimated coefficients show a considerable variation across different models. The estimates from the pooled model (*III*) are particularly different from those of other models. The year dummies are mostly significant and suggest that the cost variation over time is not linear. Again, the pooled model is an exception in which none of these dummies show any statistically significant effect. Noting that model *III* completely ignores the panel structure of the data, its estimates are likely to be strongly biased by omitted firm-specific variables. On the other hand the fixed-effects estimators (model *I*) are derived from the within-firm variations and thus unbiased.

The year dummy coefficients (excluding model *III*) show that the total costs of railway companies rose almost linearly from 1985 to 1992 with an average annual growth rate of about 1.6%, but declined after 1992 with an average rate of about 1.5% per year. Since total costs and all the continuous explanatory variables are in logarithms, the estimated coefficients can be interpreted as average cost elasticities. For instance, the output coefficients suggest that on average a one percent increase in passenger-kilometers will increase the costs by 0.11 to 0.49 percent depending on the adopted specification. The marginal effect of ton-kilometers is about 10 times lower, suggesting substantially lower variable costs for goods transportation. The coefficient of network length indicates that the marginal cost of a one percent extension in the network keeping the output constant, is approximately equivalent to 0.4 percent increase in costs. These results are consistent with the previous empirical results regarding Switzerland's railroad industry (cf. Filippini and Maggi, 1993) in that they suggest increasing returns to scale. Following Caves et al. (1985), parameters α_Y, α_Q and α_N can be employed to calculate the value of the economies of scale and density.[22] All the results obtained from

[22] Economies of density (ED) are defined as the proportional increase in total costs brought about by a proportional increase in both outputs, holding all input prices and the size of the network fixed. Economies of scale (ES) are defined as the proportional increase in total costs resulting from a proportional increase in both outputs and the size of the network, holding all input prices fixed. See Jara-Díaz and Basso (2003) and Oum and Waters (1996) for a discussion on the definition and interpretation of scale and density economies.

different models suggest that the Swiss railway companies do not fully exploit the potential scale and density economies.

Table 4 also indicates that if the pooled model is set aside, the input price coefficients do not vary significantly across different models. The coefficient of labor price, varying between .55 and .57 (bar model *III*), is actually comparable to the average share of labor expenses, which is about 52% in the sample. The capital price coefficient varies between .31 and .32 (model *III* excluded), which is considerably below the average share of capital costs in the sample (44%). This result may suggest that the companies are not so responsive as a constantly cost minimizing behavior should be, to the changes in capital prices. This can be explained by the fact that in the short run railway companies cannot vary much of their capital stock such as equipment and machinery.

Comparing the results from different models in table 4 shows that excluding model *III*, all other models have reasonably comparable coefficients. In model *III* (pooled model) variations over time and within firms are treated exactly similar to those between different firms. Moreover, the unobserved firm-specific effects are completely neglected, which may bias the estimations. A Lagrange Multiplier test on an OLS model strongly rejects the hypothesis that the residuals of a given company are uncorrelated (test statistic of 2990 for a chi-square with 1 degree of freedom), suggesting that the pooled model is mis-specified. Moreover, the Hausman test rejects the hypothesis that the firm-specific effects are uncorrelated with the explanatory variables (test statistic of 61.5 for a chi-square with 17 degrees of freedom). This result suggests that models that do not account for these correlations can give biased results. Given the relatively high number of periods (on average 12 years) and the reasonable within-company variations (see table 3) in the sample, the fixed effects model's results can be considered as unbiased estimates of the cost function parameters. Therefore, the coefficients estimated from model *I* are used as a benchmark for assessing the potential heterogeneity bias in other models.

Compared to model *I*, the parameter estimates in the pooled model (*III*) have the highest differences. The estimated coefficients in the remaining models are fairly close to those of the FE model, suggesting that heterogeneity biases in the coefficients are not substantial. This statement does not apply to the inefficiency estimates, which as we will see later, show considerable biases. As seen in table 4, there is no clear distinction between models *II* and *IV* concerning the heterogeneity biases. While in certain coefficients model *IV* is closer to the unbiased estimates (model *I*), in some others model *II* shows a 'better' performance.

The random effects specification in both models *II* and *IV* has however a shortcoming in that the firm-specific heterogeneity terms (u_i in model *II* and α_i in model *IV*) are assumed to be uncorrelated with the explanatory variables. If we put any trust in the Hausman specification test, this assumption is not realistic. Moreover, as discussed earlier, it is plausible that some of the unobserved network characteristics be correlated with the network length. Such correlations are taken into account in model *V* through the auxiliary coefficients (γ_x). The results in table 4 indicate that model *V* shows the smallest differences with the unbiased estimators of model *I*. This suggests that applying Mundlak's (1978) adjustment to the TRE model (model *IV*) can decrease the heterogeneity biases. As shown in the table, the auxiliary coefficients (γ_x) are all significant. These coefficients can be interpreted as the correlation effect between the unobserved firm characteristics and the corresponding explanatory variable. For instance, the positive signs of γ_Y and γ_Q suggest that keeping all observed factors fixed, networks with higher outputs are more likely to belong to the 'high-cost' or 'difficult' networks; and the negative signs of γ_N, γ_K and γ_L suggest that larger networks and companies that have higher input prices are more likely to be in the 'low-cost' category.

Table 4. Regression results

	Model I **FE**	*Model II* **RE**	*Model III* **Pooled**	*Model IV* **True RE**	*Model V* **True RE + Mundlak**
α_Y	.114* (.032)	.200* (.030)	.492* (.015)	.133* (.023)	.106* (.034)
α_Q	.014* (.006)	.021* (.003)	.030* (.006)	.038* (.004)	.017* (.003)
α_N	.448* (.051)	.485* (.039)	.393* (.026)	.432* (.015)	.488* (.035)
α_K	.318* (.017)	.310* (.010)	.171* (.032)	.312* (.008)	.315* (.009)
α_L	.546* (.037)	.548* (.029)	.592* (.074)	.568* (.036)	.562* (.034)
γ_Y	–	–	–	–	.159* (.050)
γ_Q	–	–	–	–	.090* (.013)
γ_N	–	–	–	–	-.150* (.056)
γ_K	–	–	–	–	-.189* (.067)
γ_L	–	–	–	–	-.193 (.180)
α_{1986}	.010 (.015)	.009 (.041)	.009 (.056)	.022 (.027)	.017 (.035)
α_{1987}	.020 (.015)	.012 (.031)	.003 (.056)	.032 (.025)	.029 (.031)
α_{1988}	.039* (.015)	.028 (.044)	.010 (.057)	.051 (.037)	.049 (.050)
α_{1989}	.065* (.016)	.052 (.046)	.036 (.057)	.076* (.033)	.074 (.050)
α_{1990}	.084* (.016)	.068 (.036)	.024 (.058)	.097* (.034)	.94* (.044)
α_{1991}	.098* (.017)	.078* (.029)	.030 (.058)	.114* (.028)	.111* (.035)
α_{1992}	.111* (.017)	.094* (.034)	.046 (.058)	.130* (.026)	.122* (.034)
α_{1993}	.100* (.017)	.081* (.034)	.015 (.057)	.119* (.026)	.112* (.034)
α_{1994}	.082* (.017)	.063 (.040)	-.001 (.056)	.103* (.037)	.093* (.039)
α_{1995}	.059* (.016)	.048 (.032)	.019 (.057)	.081* (.023)	.064 (.034)
α_{1996}	.037* (.017)	.028 (.024)	.027 (.057)	.066* (.022)	.043 (.025)
α_{1997}	.038* (.018)	.030 (.032)	.019 (.060)	.063 (.039)	.042 (.032)
α_0	–	-4.90* (.57)	-8.31* (.98)	-3.89* (.51)	-1.89 (2.66)
σ_α	–	–	–	.783* (.027)	.751* (.058)
$\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$	–	.807* (.14)	.464* (.001)	.109* (.005)	.095* (.005)
$\lambda = \sigma_u / \sigma_v$	–	11.37* (3.81)	2.88* (.30)	2.58* (.56)	1.59* (.031)

- Standard errors are given in brackets. * means significant at less than 5%.
- The sample includes 605 observations (50 railway companies).

Table 5 provides a descriptive summary of the inefficiency estimates from different models (see table 1, last row). These estimates represent the relative excess cost of a given firm compare to a minimum level that would have been achieved if the firm had operated as efficiently as the 'best practice' observed in the sample. In comparing different models it should be noted that in the first two models (*I* and *II*), the inefficiency is assumed to be constant over time. Moreover, in these models all the unobserved firm-specific differences are interpreted as inefficiency. As expected, both models *I* and *II*, especially the FE model, predict rather unrealistic inefficiency scores averaging about .7 to .8 and up to a maximum of 2 to 2.5. According to these models, a typical company can save about a third of its costs by a more efficient allocation of resources. 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.

Table 5. Inefficiency measures

	Model I **FE**	*Model II* **RE**	*Model III* **Pooled**	*Model IV* **True RE**	*Model V* **True RE with Mundlak**
Mean	0.813	0.696	0.343	0.078	0.063
Median	0.676	0.662	0.289	0.061	0.053
Maximum	2.507	1.992	0.848	0.386	0.311
95 percentile	1.723	1.470	0.848	0.187	0.134
Minimum	0.000	0.160	0.060	0.011	0.012
N	605	605	605	605	605

In model *III* the inefficiency estimates are in a more realistic range, with an average of .34 and a maximum value of .85. These values though still too high to be convincing, are substantially lower than those predicted by models *I* and *II*; and this despite the fact that the pooled model (*III*) does not account for unobserved heterogeneity. This attenuation of inefficiency estimates can be explained by the structure of the inefficiency term in model *III*. Given that the inefficiency term (u_{it}) is assumed to be independently and identically distributed over time and across companies, it cannot fully capture the firm-specific differences that are time-invariant, thus such differences are partly suppressed into and bias the model's coefficients.

Both models *IV* and *V*, which have separate stochastic terms for inefficiency and firm-specific heterogeneity, have quite reasonable inefficiency estimates about 6 to 8 percent on average and 31 to 38 percent on maximum. The substantial decrease in these values compared to other models, suggests that these models can separate to a considerable extent, the heterogeneity from the inefficiency. To understand the reasons behind these results, it is helpful to note that the sole difference between models *III* and *IV* is that model *IV* includes an additional firm-specific random term (α_i). This term represents the variations across firms, which are about 7 times larger than the variation within firms (compare σ_α to σ in the lower panel of table 4).

Given that the unobserved heterogeneity is potentially correlated with the explanatory variables, and that these correlations are not taken into account in model *IV* the resulting inefficiency scores may capture some of these differences. This issue can be explored by comparing models *IV* and *V*. In model *V* the time-invariant cost differences across companies are separated from inefficiency estimates (as in model *IV*). In addition, the possible correlations with explanatory variables are mitigated through auxiliary coefficients. The results in table 5 show that when such correlations are controlled for (model *V*), the inefficiency estimates slightly decline (by about .015 on average and by .075 on maximum). According to this model the average (median) company is only 6.3 (5.3) percent inefficient, and the maximum inefficiency in 95 percent of the sample is limited to 13.4 percent. These results suggest that model *V* not only provides unbiased, or close to unbiased, estimates of the cost function's coefficients, it can also better separate the heterogeneity from inefficiency.

The pair-wise correlation coefficients between the inefficiency estimates from different models are listed in table 6. In order for the correlation coefficients to be comparable, they are calculated at the firm level using 50 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 605 observations.

As shown in table 6, models *I* and *II*, and models *IV* and *V* show a relatively high correlation.[23] However, except a few cases the correlation coefficients are quite low, suggesting substantial differences across models.[24] Especially, models *IV* and *V* show a negative correlation with all other models. Given that the correlation coefficients are calculated on company-average inefficiency scores, the weak (and negative) correlations may suggest that the inefficiency estimates vary considerably from one year to another, in which case the correlation between models with constant and time-variant inefficiency should be weak. However, this can only partly explain the observed correlations. In fact the positive and fairly strong correlation between the pooled model *III* (with time-variant efficiency) and both

---

[23] These results are consistent with Farsi et al. (2003) who used a similar methodology for a sample of nursing homes.

[24] The rank correlations show similar patterns. These results are omitted to avoid repetition.

models *I* and *II* (with time-invariant efficiency) indicates that averaging cannot explain the negative correlations.

Table 6. Pair-wise correlation between inefficiency estimates

	Model I **FE**	*Model II* **RE**	*Model III* **Pooled**	*Model IV* **True RE**	*Model V* **True RE with Mundlak**
Model I	1				
Model II	.932*	1			
Model III	.497*	.614*	1		
Model IV	-.247	-.256	-.158 [.092*]	1	
Model V	-.334*	-.320*	-.197 [.105*]	.948* [.971*]	1

- The correlation coefficients have been estimated over the firms (50 observations) that is, average values over the sample period are used in models with time-variant inefficiency (*III*, *IV* and *V*).
- Correlation coefficients based on 605 observations are given in brackets.
- * means significant at 5%.

The negative correlation coefficients (table 6) point to a striking distinction between the models *IV* and *V* and all other models, which do not distinguish unobserved heterogeneity from inefficiency. The negative correlations manifest especially in model *V* in which the correlations with observed factors are taken into account. These values suggest that some of the unobserved network characteristics may actually be negatively correlated with company's average inefficiency. One interpretation is that the relatively complex thus costly networks are more likely to be operated by an efficient management. This is a plausible explanation because the companies with complex networks are more likely to have a general awareness and perhaps the required expertise for technical problems. Such expertise can directly or indirectly contribute to the firm's efficiency. The results in table 6 highlight the importance of unobserved heterogeneity, as failure to account for such factors can result in a completely misleading and even reverse picture of inefficiencies.

6. CONCLUSION

Alternative cost frontier models applied to a panel of Swiss railway companies indicate that the estimations particularly the inefficiency estimates, are sensitive to the adopted specification. The data show a considerable unobserved firm-specific heterogeneity that is likely to be correlated with explanatory variables. In such cases unbiased coefficients can be obtained from the fixed effects model. This model's estimates of inefficiency are however unrealistic. In fact, comparing the results across different models suggest that the inefficiency estimates largely depend upon how the unobserved heterogeneity across firms is specified. Panel data models such as Pitt and Lee (1981) and Schmidt and Sickles (1984) that do not distinguish between unobserved firm-specific heterogeneity and inefficiency can overestimate the overall inefficiencies or even give misleading patterns of inefficiency. The cost frontier random effects model labeled as 'true' random-effects model (Greene, 2004) provides reasonable estimates of inefficiency confirming that the inefficiency estimates in other models are confounded with unobserved heterogeneity such as network effects. However, the problem of this model is that because of potential correlation between heterogeneity and explanatory variables, the cost function coefficients may be biased (heterogeneity bias), especially as the Hausman specification test confirms the presence of such correlations.

Using an auxiliary equation in line with Mundlak (1978) can be helpful in this regard. This adjustment has been applied to the 'true' random effects. The resulted specification not only proves a very low level of heterogeneity bias, it slightly reduces the inefficiency estimates. The high correlation between the inefficiency estimates across the two models suggests that in so far as the heterogeneity is accounted for, the correlation between heterogeneity and explanatory variables does not considerably affect the inefficiency estimates.

References

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

Alvarez, A., C. Arias, and W. Greene (2003) 'Fixed management and time-invariant technical efficiency in a random coefficients model', Working Paper.

Banos-Pino, Jose, V. Fernandez-Blanco, A. Rodriguez-Alvarez (2002) 'The allocative efficiency measure by means of a distance function: The case of Spanish public railways' *European Journal of Operational Research* 137 (2002) 191-205.

Baltagi, Badi H. (2001) *Econometric Analysis of Panel Data*, 2nd edition, John Wiley and Sons.

Battese, George E., and Tim J. Coelli (1992) 'Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India', *Journal of Productivity Analysis*, 3 (1): 153-169.

Cantos, P., J. M. Pastor, and L. Serrano (1999) 'Productivity, efficiency and technical change in the European railways: a non-parametric approach' *Transportation* 26: 337-357.

Caves, Douglas W., L. R. Christensen, M. W. Tretheway, and R. J. Windle (1985) 'Network effects and the measurement of returns to scale and density for U.S. railroads', in *Analytical Studies in Transpot Economics*, editor. A. F. Daugherty, Cambridge University Press, pp. 97-120.

Chamberlain, Gary (1982) 'Multivariate regression models for panel data' *Journal of Econometrics*, 18: 5-46.

Cornes, R. (1992) *Duality and Modern Economics,* Cambridge University Press.

Cornwell, Christopher, Peter Schmidt, and Robin C. Sickles (1990) 'Production frontiers with cross-sectional and time-series variation in efficiency levels' *Journal of Econometrics*, 46 (1990): 185-200.

Estache, Antonio, M. Gonzalez, and L. Trujillo (2002) ‘What does privatization do for efficiency? Evidence from Argentina’s and Brazil’s Railways’ *World Development*, 30 (11): 1885-1897.

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.

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): 683-90.

Filippini, Massimo, and Rico Maggi (1993) ‘Efficiency and Regulation in the case of the Swiss Private Railways’, *Journal of regulatory Economics*, 5, 199-216.

Greene, William (2004) ‘Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model’, *Journal of Econometrics*, forthcoming.

Greene, William (2003) ‘Distinguishing between heterogeneity and inefficiency: Stochastic frontier analysis of the World Health Organization’s panel data on national health care systems’ Working paper, Stern School of Business, New York University (April 2003).

Greene, William (2002a) ‘Fixed and random effects in stochastic frontier models’ Working paper, Stern School of Business, New York University (October 2002).

Greene, William (2002b) ‘Alternative panel data estimators for stochastic frontier models’ Working paper, Stern School of Business, New York University (September 2002).

Hsiao, Cheng (2003) *Analysis of Panel Data*, 2nd edition, Cambridge University Press.

Jara-Díaz, Sergio R., and Leonardo J. Basso (2003) ‘Transport cost functions, network expansion and economies of scope’, *Transportation Research Part E*, 39 (2003) 271-288.

Jondrow, J., I. Materov, K. Lovell, and P. Schmidt (1982) ‘On the estimation of technical inefficiency in the stochastic frontier production function model’ *Journal of Econometrics*, 19, 2/3 (August), 233-238.

Kumbhakar, Subal C. and C. A. Knox Lovell (2000) *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge.

Mancuso, Paolo, and P. Reverberi (2003) ‘Operating costs and market organization in railway services, The case of Italy, 1980-1995’ *Transportation Research Part B*, 37 (2003) 43-61.

Mundlak, Yair (1978) ‘On the pooling of time series and cross section data’ *Econometrica*, 64 (1): 69-85.

Oum, T. H., and W. G. Waters II (1996) ‘A survey of recent developments in transportation cost function research’, *Logistics and Transportation Review*, 32 (4): 423-463.

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.E. Sickles (1984) ‘Production Frontiers and Panel Data’, *Journal of Business and Economic Statistics*, 2, 367-374.

Todani, Khathutshelo R. (2001) *Determinants of Price-Cost Margins in the U.S. Railroad Industry: An Econometric Investigation*, PhD Dissertation, University of Minnesota.