DIW BERLIN

Discussion Papers

960

Astrid Cullmann • Mehdi Farsi • Massimo Filippini

Unobserved Heterogeneity and International Benchmarking in Public Transport

Berlin, December 2009

DIW BERLIN

Opinions expressed in this paper are those of the author and do not necessarily reflect views of the institute.

Unobserved Heterogeneity and International Benchmarking in Public Transport

by

Astrid Cullmann (DIW Berlin)[*]

Mehdi Farsi (CEPE, ETH Zürich)

Massimo Filippini (CEPE, ETH Zürich)

December 2009

ABSTRACT

We analyze the technical efficiency of German and Swiss urban public transport companies by means of SFA. In transport networks we might face different network structures or complexities, not observed, but influencing the production process. The unobserved factors are typically modeled as separable factors. However, we argue that the entire production process is organized around different network structures. Therefore, they are inevitably non-separable from the observed inputs and outputs. The adopted econometric model is a random coefficient stochastic frontier model. We estimate an input distance function for the years 1991 to 2006. The results underline the presence of unobserved non-separable factors.

---

[*] Corresponding author: Astrid Cullmann, DIW Berlin (German Institute for Economic Research), Department of Innovation, Manufacturing, Service, Mohrenstraße 58, 10117 Berlin Germany, tel.: +49-30-89789-679, fax.: +49-30-89789-103, acullmann@diw.de

1. Introduction

Following the explosive growth of subsidy requirements for public transport services observed in the 1970s and 1980s, several European governments have introduced during the last two decades, regulatory reforms in their local transport sectors. Most of these countries, in line with the EU directives, have adopted a competitive tendering procedure for the assignment of franchised monopolies in the local transport sector. These procedures are supposed to replace the previous models with relatively low incentives for cost efficiency, commonly based on annual negotiations with individual companies over costs and transfers. However, several studies have pointed out the difficulties in the implementation of tendering procedures, which have been experienced across many European countries (Toner, 2001; Boitani and Cambini, 2002; Cambini and Filippini, 2003). These difficulties are mainly related to potential collusion among the bidders and the tendency toward auctioning small networks hence, suboptimal scale and density. An alternative approach would be incentive regulation schemes, such as yardstick competition or performance based contracts.[1] The latter regulation schemes are based on benchmarking analysis of costs and/or quality to determine the transfers and prices. In particular, Hensher and Stanley (2003) and Hensher (2007) have shown that performance based contracts can reach a greater social surplus than competitive tendering procedures.

---

[1] For a general discussion on these two approaches see Demsetz (1968), Laffont and Tirole (1993), Klemperer (1999) and Hensher and Stanley (2003).

In Switzerland and Germany the competitive tendering procedures have been introduced only partially and limited to certain areas.[2] Nevertheless, the regional authorities have been discussing the possibility of adopting high-powered contracts based on a yardstick competition model proposed by Shleifer (1985). In this context the use of production, distance or cost frontier models could be useful as a complementary control instrument in the definition of the amount of subsidies granted to the regional public transit companies.[3] Of course, the reliability of efficiency estimates is crucial for an effective implementation of those incentive mechanisms. In fact, the empirical evidence suggests that the estimates are sensitive to the adopted benchmarking approach.[4] This implies that the choice of the approach can have important effects on the financial situation of the companies.

Since urban transit companies operate in different networks and environments, and provide urban passenger services using a diversity of vehicles (bus, tramway, light rail, etc.) any method based on cost comparison has been subject to criticism. A high level of output heterogeneity is a general characteristic of network industries. Companies characterized by different share of the employed vehicles and networks with different shapes have different organization and coordination problems, thus different performances in terms of production and costs. For instance, in the transit sector the production of 100 tramway-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

---

[2] These include Swiss rural area, one German state (Hesse) and only a few large German cities (Hamburg and Munich). In most other cases, particularly, in Swiss urban areas, the incumbent companies continue to receive concessions without any competitive tendering.
[3] For an application of yardstick competition in the transport sector see Dalen and Gòmez-Lobo (2003).
[4] See Jamasb and Pollit (2003) and Estache et al. (2004) for examples.

affect the costs. Furthermore, different environmental characteristics influence the production process and therefore the costs. In general, the information is not available for all output characteristics. Many of these characteristics are therefore omitted from the production, distance and cost function specifications. Of course, this efficiency measurement problem related to the unobserved firm-specific heterogeneity becomes more serious when a regulator decides to perform a comparative efficiency analysis across several countries. This type of analysis is becoming more and more popular because regulators are interested in comparing the inefficiency level of the companies operating in the own country with the performance obtained by companies operating in another regulatory environment or just simply to increase the number of companies in the sample. It is also believed that by using international benchmarking, the regulators could limit the possibility of strategic behavior of firms within the country. The increasing use of international benchmarking analysis in network industries has raised serious concerns among regulators and companies regarding the reliability of efficiency estimates, because of unobserved firm- and country-specific heterogeneity.

Thus, our main objective is to derive and apply an appropriate SFA model, which is able to capture firm-specific unobserved heterogeneity using panel data. In recent SFA panel data models unobserved firm-specific heterogeneity can be taken into account with conventional fixed or random effects. In order to distinguish heterogeneities such as external network effects from cost efficiency, Greene (2004, 2005a,b) proposed an approach that integrates an additional stochastic term representing inefficiency in both fixed and random effects models.[5] Within this framework the unobserved factors are

---

[5] Kumbhakar (1991) proposed a similar approach using a three-stage estimation procedure. See Heshmati and Kumbhakar (1994) for an applications of this model.

widely modeled as separable factors from the production process (Greene, 2005b). However, we argue that the entire production process is organized around different network structures and shapes. Therefore, the unobserved heterogeneity is inevitably non-separable from the observed inputs and outputs. Against this background we propose a model assuming that unobserved heterogeneous factors are non-separable from the production process (see e.g. Bagdadioglu and Weyman-Jones, 2008). We show that along with the variation over time, the distinction between separable and non-separable factors can be helpful in disentangling the inefficiency from the unobserved firm-specific factors: Assuming that firm-specific factors are time-invariant but non-separable, while the inefficiency components are time-variant and separable, one can achieve a more realistic separation between the two components. In fact, being an integrated part of the technology process the unobserved network characteristics are non-separable but more or less time-invariant. Whereas it is likely that the main driving factor behind technical inefficiency namely, the management's efforts and incentives are independent from the production technology thus separable but time-variant.

The adopted econometric model is a random coefficient stochastic frontier model that allows non-separability between the firm-specific unobserved heterogeneity and the production factors. In our model the unobserved heterogeneity is treated as a stochastic network characteristic that enters as a latent variable in the distance function and can be interacted with observed inputs and outputs. The resulting specification is a random-parameter stochastic frontier model in which the individual random effects are based on a single standard normal variable.[6] The input distance function is used to test the assumption of separability between unobserved network characteristics and observed

---

[6] The econometric model bears some resemblance to the specification proposed by Alvarez et al. (2004).

production factors. The model also allows an assessment of the effect of separability assumptions on the estimates of efficiency as well as technological properties such as returns to scale and cost-complementarities.

The model is applied to a panel data sample of 56 transit companies including German and Swiss operators. The estimates of efficiency are compared between the two countries and the statistical significance of the differences is tested. From a methodological point of view, the analysis contributes to the discussion of unobserved heterogeneity that is particularly relevant for international comparisons. This study also provides an insight to the potential use of benchmarking within competitive tendering procedures that are often promoted in the ongoing reforms in the public transport sector. The rest of the paper is organized as follows: Sections 2 presents the model specification. The data and the econometric models are explained in Section 3. Section 4 presents the estimation results and discusses their implications, and Section 5 provides the conclusions.

2. Model Specification

There is a great body of literature on the estimation of production and cost frontiers for public transit operators.[7] However, the majority of these studies estimate single output production or cost frontiers. There are only a few studies that estimated a multi-output cost function. The most relevant ones in this category are Viton (1992), Viton (1993) and Colburn and Talley (1992), both of which analyzed the long run cost structure of urban multi-mode transit system in the U.S. Viton (1992) studied the cost structure of a sample of 289 urban transit companies operating in the U.S. between 1984 and 1986.

---

[7] See De Borger et al. (2002) for a detailed literature overview.

5

Six modes are distinguished: motor-bus, rapid-rail, streetcar, trolley-bus, demand responsive mode and a last mode including all other modes. Viton uses a quadratic total cost function. Colburn and Talley (1992) analyze the economies of scale and scope of a single urban multi-service company using quarterly data from 1979 to 1988. Four modes are distinguished: motor-bus, dial-a-ride, elderly service, and van pool service. Colburn and Talley used a translog total cost function. The first European analysis for multi-output firms has been performed by Farsi et al. (2006b). In this study, the authors estimate a quadratic cost function considering three modes (motor-bus, streetcar, trolley-bus) and using a dataset composed of 16 Swiss multi-mode urban transport operators observed during the period 1985–2003. All these studies did not estimate a frontier function and, therefore, did not perform an efficiency analysis. The main interest of these studies was in the estimation of the economies of scale and scope.

To measure the efficiency level of the multi-outputs Swiss and German urban transit companies we apply a parametric frontier input distance function.[8] We therefore focus on the technical inefficiency as opposed to possible inefficiencies due to suboptimal allocation of input factors. Because of the lack of consistent data on costs and input prices especially on the German side, we could not use a multi-output cost function. The latter approach, while providing a benefit in estimating the resulting effect of technical and allocative inefficiency, has an important drawback in international benchmarking. Namely, international cost comparisons would involve several empirical difficulties given the different accounting rules, depreciation standards, exchange rates etc. In addition, the choice of distance functions does not require the cost minimization

---

[8] For the use of parametric distance functions in the transport sector see Coelli and Perelman (1999, 2000).

assumption.[9] Compared to production functions the distance functions are more readily adaptable to multi-output contexts. Moreover, assuming that outputs are exogenous for given companies, we favored an input distance specification as opposed to an output distance function.[10]

The input distance function is defined on the input set as the extent to which the input vector may be proportionally contracted with the output vector held fixed (see Coelli, 2002):

$$d_I(x, y) = \max\{\rho : (x/\rho) \in L(y)\} \qquad (1)$$

$d_I(x, y)$ will take a value greater than or equal to one if the input vector x is an element of the feasible input set $L(y)$. In addition, $d_I(x, y) = 1$ if x is located on the inner boundary of the input set. ρ represents the scalar distance, so the amount by which the input vector can be deflated. It is assumed that the technology satisfies the standard axioms: $d_I(x, y)$ is non-decreasing, positively linearly homogeneous and concave in x and increasing in y.[11]

As in most empirical studies in production literature, we specify a translog functional form for the input distance function in order to satisfy flexibility while a

---

[9] For a discussion on the advantages and drawbacks of the distance-functions approach see Coelli (2002) and Coelli and Perelman (2000).

[10] An input-oriented distance function is motivated by the nature of production in the public transport sector, because it implies that efficiency is improved by reducing input usage for a given exogenous output, set by regulators or the demand side factors that are beyond the provider's control.

[11] See Coelli (2002) and Färe and Primont (1995) for more details on these properties.

straightforward imposition of the linear homogeneity restriction.[12] For the case of M outputs and K inputs the input distance function for the ith firm can be written as

$$\ln d_{Ii} = \alpha_0 + \sum_{m=1}^{M} \alpha_m \ln y_{mi} + \frac{1}{2} \sum_{m=1}^{M} \sum_{n=1}^{M} \alpha_{mn} \ln y_{mi} \ln y_{ni} + \sum_{k=1}^{K} \beta_k \ln x_{ki}$$
$$+ \frac{1}{2} \sum_{k=1}^{K} \sum_{l=1}^{K} \beta_{kl} \ln x_{ki} \ln x_{li} + \sum_{k=1}^{K} \sum_{m=1}^{M} \delta_{km} \ln x_{ki} \ln y_{mi} \tag{2}$$

To obtain the frontier surface (the transformation function) one would set $d_{Ii} = 1$, so the left hand side equals zero $\ln d_{Ii} = 0$ (see Coelli and Perelman, 2000). The restrictions for linear homogeneity in inputs can be written as:

$$\sum_{k=1}^{K} \beta_k = 1, \quad \sum_{k=1}^{K} \beta_{kl} = 0 \quad \text{and} \quad \sum_{k=1}^{K} \delta_{km} = 0 \quad \text{for all values of } l \text{ and } m \tag{3}$$

A convenient approach of imposing homogeneity constraints is to follow Lovell et al. (1994) and Coelli and Perelman (2000): Considering that homogeneity implies that for any $w > 0$

$$d_I(wx, y) = wd_{Ii}(x, y) \tag{4}$$

Therefore, one of the inputs might be arbitrarily chosen, such as the Kth input and set $w = 1/X_K$. Then one obtains

---

[12] The Cobb Douglas form is too restrictive with regard to the elasticity of substitution and scale properties.

8

$$d_I(x/X_K, y) = d_I(x, y)/X_K \qquad (5)$$

For the specification of the model used in this study we considered public transit companies characterized by a production process with three inputs, labor, number of trams and number of buses, and two outputs, seat-kilometers provided by tramways and buses respectively.[13] Considering this production process we can specify the following input distance function:

$$d = f(X_L, X_{CT}, X_{CB}, Y_T, Y_B, Z, \gamma, t) \qquad (6)$$

where x_L is labor; y_B and y_T are the numbers of seat-kilometers provided by buses and tramways respectively. Following Farsi et al. (2006a, 2006b) we decided to assume two pure supply oriented measures of the output. x_{CB}, x_{CT} are respectively two indicators of the capital input, number of buses and number of tramways; t is a time variable which captures the shift in technology, Z is the total network length (trams and bus networks) introduced in the model in order to capture part of the observable heterogeneity of the operating environment of the companies. However, in transport networks we might face different network structures and various shapes or complexities. Thus, network length only captures part of the network heterogeneity. For this reason, we included in the model a variable, γ, that captures other network structural characteristics that are constant over

---

13 We concentrate our analysis only on transit companies supplying services using the same transport modes (buses and tramways). Therefore, we excluded transit companies operating with underground system as well as small companies that use only buses. Moreover, in Switzerland some of the companies supply trolley as well as autobus services. We assumed for the empirical analysis that the trolley busses feature similar characteristics as the autobuses, therefore we sum up both singles branches to have an aggregated bus stock and aggregated supplied services.

time. These characteristics include unobserved factors related to network's shape and complexity.[14] As we will discuss in the next session, it is possible to consider γ as a latent variable in the econometric specification. In our model, we therefore assume that γ captures the time-invariant unobserved heterogeneity of the production process of the transit companies, in our case mainly the structural characteristics of the network.

Using a translog functional form and assuming non-separability of the unobserved network structural variable, γ, the model in equation (6) can be expressed as follows:

$$
\begin{aligned}
-\ln\left(x_{Lit}\right) = {} & \alpha_0 + \eta_1\gamma_i + \frac{1}{2}\eta_2\gamma_i^2 + \alpha_{CT}\ln\frac{x_{CTit}}{x_{Lit}} + \eta_4\ln\frac{x_{CTit}}{x_{Lit}}\gamma_i + \alpha_{CB}\ln\frac{x_{CBit}}{x_{Lit}} \\
& + \eta_3\ln\frac{x_{CBit}}{x_{Lit}}\gamma_i + \frac{1}{2}\alpha_{CTCT}(\ln\frac{x_{CTit}}{x_{Lit}})^2 + \frac{1}{2}\alpha_{CBCB}(\ln\frac{x_{CBit}}{x_{Lit}})^2 \\
& + \alpha_{CBCT}(\ln\frac{x_{CBit}}{x_{Lit}})(\ln\frac{x_{CTit}}{x_{Lit}}) + \beta_T\ln y_{Tit} + \eta_6\ln y_{Tit}\gamma_i + \beta_B\ln y_{Bit} \\
& + \eta_5\ln y_{Bit}\gamma_i + \frac{1}{2}\beta_{TT}(\ln y_{Tit})^2 + \frac{1}{2}\beta_{BB}(\ln y_{Bit})^2 \\
& + \beta_{BT}(\ln y_{Bit})(\ln y_{Tit}) + \delta_{TCT}(\ln y_{Tit})(\ln\frac{x_{CTit}}{x_{Lit}}) + \delta_{BCT}(\ln y_{Bit})(\ln\frac{x_{CTit}}{x_{Lit}}) \\
& + \delta_{TCB}(\ln y_{Tit})(\ln\frac{x_{CBit}}{x_{Lit}}) + \delta_{BCB}(\ln y_{Bit})(\ln\frac{x_{CBit}}{x_{Lit}}) + \alpha_Z\ln Z_{it} + \eta_7\ln Z_{it}\gamma_i \\
& + \frac{1}{2}\alpha_{ZZ}(\ln Z_{it})^2 + \alpha_{ZB}(\ln y_{Bit})(\ln Z_{it}) + \alpha_{ZT}(\ln y_{Tit})(\ln Z_{it}) \\
& + \alpha_{ZCT}(\ln Z_{it})(\ln\frac{x_{CTit}}{x_{Lit}}) + \alpha_{ZCB}(\ln Z_{it})(\ln\frac{x_{CBit}}{x_{Lit}}) + \alpha_t t - \ln d_{it}
\end{aligned}
\tag{7}
$$

Of course, by imposing separability of the unobserved heterogeneity, in our case represented mainly by the unobserved network characteristics, a simplified version of

---

[14] For instance, Filippini and Maggi (1992) have shown the importance of the inclusion of an indicator of the network complexity in a cost function for transport companies. In that study a complexity indicator based on the graph theory is used to measure the network complexity. Unfortunately, data on the shape and structure of the networks are not available for the companies included in our sample.

model (7) can be estimated.[15] However, the separability assumption is relatively strong, because the network structure influences the optimal choice of the mix of inputs and outputs. For this reason, we decided to use the non-separabilty assumption. To consider, that this assumption can have important measurement effects on the estimation of the value of the return to scale.

$\ln d_I$ is a nonnegative variable which can be associated with technical inefficiency u_{it}. Given the stochastic error v_{it} this model can be formulated in the common SFA form with the combined error term $v_{it} - u_{it}$ and the common assumption of a normal distribution for v_{it} and a half-normal distribution truncated at zero for u_{it}.[16] A radial input-oriented measure of technical efficiency can be obtained by $TE = \dfrac{1}{d_{Ii}} = \exp(-u_{it})$.

The distance function provides a promising new solution to the single output restriction of the standard production functions. One concern in the econometric estimation might be the regressor endogeneity which may introduce possible simultaneous equation bias.[17] Sickles et al. (2002) and Atkinson and Primont (2002) used methods based on instrumental variables to correct for such endogeneities. However, Coelli (2002) showed that compared to production functions, the distance functions do not face a greater risk of endogeneity bias.[18]

---

[15] With the separability assumption, all the interactions of the variable γ with observed variables will be excluded.

[16] For estimation purposes, the negative sign on the dependent variable can be ignored. This results in the signs of the estimated coefficient being reversed.

[17] This results from the fact that for instance in an input distance function, the inputs appearing on the right hand side of the equation might be correlated with the residuals.

[18] A second issue is that estimated input distance functions often fail to satisfy the concavity properties implied by economic theory. Regularity conditions could also be imposed by estimating the model in a Bayesian framework (see O'Donnell and Coelli, 2005).

3. Data and econometric specification

3.1. Data

The multi-output distance input function is estimated using annual data on 49 German and 7 Swiss companies observed over a sixteen-year period from 1991 to 2006. We use an unbalanced panel data set and had 13 years for each country (from 1991-2003 for Switzerland, and from 1994-2006 for Germany). The data for Germany is provided by the VDV Statistics.[19] For 360 public transport companies data is available; among them are 60 companies which are offering bus as well as regional rail services. We created a consistent panel data set for 49 multi-output companies offering tram and bus services in medium and larger German cities.[20]

In Switzerland operate sixteen public transport companies which cover all the local public transit services within the urban centers in Switzerland. Like in Germany the companies can be defined as independent local monopolies, given the fact that there is no overlapping between the offered transport services across the companies. For our analysis seven Swiss companies out of the sixteen are relevant, as six offer all three modal transit services, and one firm offers motor-bus and trolley-bus services, therefore can be considered as multi-output transport companies. For the years between 1991 and 1997 the Swiss data has been extracted from the annual statistics on public transport reported by the Swiss Federal Statistical Office (BFS (1991-97)). The data for the following years

---

[19] VDV (Verband Deutscher Verkehrsunternehmen, Association of German Transport Companies) which is an organization for Germany's public transport companies and rail freight transport. The VDV has approximately 440 members.

[20] In order to have in the sample companies that offer more or less the same services, we excluded from the analysis four companies that offer underground railways services in addition to bus and tram services and 3 other small single-output bus companies that have only trolley bus. We think that this type of services needs a quit different technology. Moreover, just four companies operating in four large German cities (Berlin, Hamburg, Munich and Nuremberg), are characterized by underground services.

(1998-2003) have been collected from companies' annual reports. Summary statistics of

the variables used in our models are given in Table 1.

Table 1: Summary statistics for Germany and Switzerland

Variable	Obs	Mean	Min	Max	Obs	Mean	Min	Max
German = GE; Swiss=CH	GE	GE	GE	GE	CH	CH	CH	CH
Inhabitants	616	366709	40800	1642000	91	285215	76381	421802
Number of employees	616	978	30	3996	91	953	76	2798
Network length tram in km	616	49	3	155	91	32	8	110
Network length bus in km	616	465	5	2653	91	139	42	362
Number trams	616	118	2	755	91	128	12	432
Number buses	616	135	2	470	91	167	30	314
Tram-km in 1000 km	616	5664	61	34363	91	6111	398	20518
Bus-km in 1000 km	616	7211	86	28519	91	8121	1525	18438
Seat-km tram in 1000 km	616	964943	5000	6187000	91	847835	37387	2926006
Seat-km bus in 1000 km	616	584293	4000	2303000	91	974580	121443	2283553
Area in km^2	616	171	21	405	91	169	90	275

The sample used in this empirical analysis is, therefore, composed of a sample of

Swiss and German transit companies that provides bus and tram transport services

characterized by partly different technologies, different regulation methods, different

environmental variables and in particular different network complexities γ_i. This large

output heterogeneity is not completely covered and observed in the data. This is evidently

more relevant when it comes to international cross-country efficiency analysis. Therefore,

in the choice of the econometric models this heterogeneity of the sample has to be considered in detail in order to separate the unobserved factors from inefficiencies by means of panel data.

3.2 Econometric Specification using panel data

The first use of panel data models in stochastic frontier models goes back to Pitt and Lee (1981) who interpreted the panel data random effects as inefficiency rather than heterogeneity.[21] A main shortcoming of these models is that any unobserved, time-invariant, firm-specific heterogeneity is considered as inefficiency. In order to solve this problem, the SFA model in its original form (Aigner et al., 1977) can be readily extended to panel data models, by adding a fixed or random effect in the model. Although similar extensions have been proposed by several previous authors,[22] Greene (2005a,b) provides effective numerical solutions for both models with random and fixed effects, which he respectively refers to as "true" fixed and random effects models. Several recent studies such as Greene (2004), Farsi et al. (2006b), Alvarez et al. (2004) and Tsionas (2002) have followed this line. Some of these models have proved a certain success in a broad range of applications in network industries in that they give more plausible efficiency

---

[21] Pitt and Lee (1981)'s model is different from the conventional RE model in that the individual specific effects are assumed to follow a half-normal distribution. Important variations of this model were presented by Schmidt and Sickles (1984) who relaxed the distribution assumption and used the GLS estimator, and by Battese and Coelli (1988) who assumed a truncated normal distribution. In more recent papers the random effects model has been extended to include time-variant inefficiency. Cornwell et al. (1990) and Battese and Coelli (1992) are two important contributions in this regard. In particular the former paper proposes a flexible function of time with parameters varying among firms. However, in both these models the variation of efficiency with time is considered as a deterministic function that is commonly defined for all firms.

[22] In particular Kumbhakar (1991) proposed a three-stage estimation procedure to solve the model with time- and firm-specific effects.

estimates.[23] These results raise an important question as to what extent the panel-data-adapted models can be used to have a better understanding of the inefficiencies and whether they can provide a reliable basis for benchmarking and incentive regulation systems in industries characterized by strong heterogeneity. This question is especially important when companies operate in multiple networks, entailing several network-specific heterogeneity dimensions γ_i. In the recent "true" SFA models the unobserved factors are widely modeled as separable factors from the production process (Greene, 2005a,b). However, we argue that the entire production process is organized around different network structures and shapes. Therefore, the unobserved heterogeneity is inevitably non-separable from the observed inputs and outputs. We propose a model assuming that unobserved heterogeneous factors are non-separable from the production process (see for instance Bagdadioglu and Weyman-Jones, 2008).

Along with the variation over time, the distinction between separable and non-separable factors can be helpful in disentangling the inefficiency from the unobserved firm-specific factors: Assuming that firm-specific factors are time-invariant but non-separable, while the inefficiency components are time-variant and separable, one can achieve a more realistic separation between the two components. In fact, being an integrated part of the technology process the unobserved network characteristics are non-separable but more or less time-invariant. Whereas it is likely that the main driving factor behind technical inefficiency namely, the management's efforts and incentives are independent from the production technology thus separable but, as shown by Alvarez et al. (2004), time-variant.

---

[23] See Saal et al. (2007), Farsi et al. (2005, 2006a,b) for applications in water distribution, electricity networks, bus transport and railroads respectively.

Considering the technical efficiency as a time-variant stochastic term with half-normal distribution, $u_{it} \sim N^+(0,\sigma_u^2)$, and an additive idiosyncratic symmetric error with normal distribution, $v_{it} \sim N(0,\sigma_v^2)$, the distance from the stochastic frontier ($\ln d_l$) can be specified as $v_{it} - u_{it}$. By substituting for $\ln d_l$ the stochastic frontier given in equation (7) can therefore be transformed to a random parameter stochastic frontier model with all random parameters (first order input and output terms, constants and structural variable) as functions of a single stochastic term γ_i with standard normal distribution $N(0,1)$, as follows:

$$
\begin{aligned}
-\ln\left(x_{Lit}\right) = {} & \alpha_0 + \eta_1\gamma_i + \frac{1}{2}\eta_2\gamma_i^2 + (\alpha_{CT} + \eta_3\gamma_i)\ln\frac{x_{CTit}}{x_{Lit}} + (\alpha_{CB} + \eta_4\gamma_i)\ln\frac{x_{CBit}}{x_{Lit}} \\
& + \frac{1}{2}\alpha_{CTCT}(\ln\frac{x_{CTit}}{x_{Lit}})^2 + \frac{1}{2}\alpha_{CBCB}(\ln\frac{x_{CBit}}{x_{Lit}})^2 + \alpha_{CBCT}(\ln\frac{x_{CBit}}{x_{Lit}})(\ln\frac{x_{CTit}}{x_{Lit}}) \\
& + (\beta_T + \eta_5\gamma_i)\ln y_{Tit} + (\beta_B + \eta_6\gamma_i)\ln y_{Bit} + \frac{1}{2}\beta_{TT}(\ln y_{Tit})^2 \\
& + \frac{1}{2}\beta_{BB}(\ln y_{Bit})^2 + \beta_{BT}(\ln y_{Bit})(\ln y_{Tit}) + \delta_{TCT}(\ln y_{Tit})(\ln\frac{x_{CTit}}{x_{Lit}}) \\
& + \delta_{BCT}(\ln y_{Bit})(\ln\frac{x_{CTit}}{x_{Lit}}) + \delta_{TCB}(\ln y_{Tit})(\ln\frac{x_{CBit}}{x_{Lit}}) + \delta_{BCB}(\ln y_{Bit})(\ln\frac{x_{CBit}}{x_{Lit}}) \\
& + (\alpha_Z + \eta_7\gamma_i)\ln Z_{it} + \frac{1}{2}\alpha_{ZZ}(\ln Z_{it})^2 + \alpha_{ZB}(\ln y_{Bit})(\ln Z_{it}) + \alpha_{ZT}(\ln y_{Tit})(\ln Z_{it}) \\
& + \alpha_{ZCT}(\ln Z_{it})(\ln\frac{x_{CTit}}{x_{Lit}}) + \alpha_{ZCB}(\ln Z_{it})(\ln\frac{x_{CBit}}{x_{Lit}}) + \alpha_t t + v_{it} - u_{it}
\end{aligned}
\tag{8}
$$

In generic terms, this represents a random parameters stochastic frontier model with random constant term $= \alpha_0 + \eta_1\gamma_i + \frac{1}{2}\eta_2\gamma_i^2$ where the unobserved fixed output also enters in quadratic terms, and first order terms of inputs $(\alpha_{CT} + \eta_3\gamma_i), (\alpha_{CB} + \eta_4\gamma_i)$, outputs

$(\beta_T + \eta_5 \gamma_i), (\beta_B + \eta_6 \gamma_i)$, and the structural variable network length $(\alpha_Z + \eta_7 \gamma_i)$ and nonrandom second order terms in the specified translog distance function. The unobserved network characteristics are therefore aggregated into a single time-invariant stochastic output (γ_i) interacted with observed inputs and outputs, and entering as a latent variable in the distance function model. Inspired by the Alvarez et al. (2004)'s model we can derive a stochastic frontier model in which all the random parameters are based on an identical random effect γ_i.

In summary, we see that the unobserved firm-specific heterogeneity attributed to the different network structures of the transport companies applies to marginal products represented by the coefficients of the distance function (see Section 4.1). We therefore allow firms to have different underlying production technologies caused by unobserved differences in technological conditions and networtk structures. In particular network structural characteristics play an important role in the production of transport services and cannot be fully captured by a production frontier with fixed coefficients. The proposed random coefficient frontier accounts for these differences using a single stochastic variable that is interpreted as an aggregate measure of structural characteristics that are not completely observable.

4. Empirical results

Table 2 shows the regression results of the distance function, based on the stochastic frontier model given in equation 8. The table also includes the results of an alternative specification in which the unobserved network variable (γ_i) is assumed to be

separable from all production factors. Given that all the variables are in logarithmic form, these coefficients can be directly interpreted as elasticities. For instance, the derivative of a translog input distance function with respect to a particular input is equal to the input contribution share of that input. In the interpretation of the coefficients it should be noted that a positive coefficient implies a contraction of the feasible input set thus, an increase in the distance function. Conversely, the negative effects are associated with an expansion in the input set. Therefore, outputs are expected to have negative coefficients while inputs are associated with positive effects. Similarly any positive coefficient indicates an improvement in production feasibilities, while negative coefficients can be interpreted as more resources and costs. For instance, the value of the coefficient of the time trend indicates an average technological progress of about 2 percent per year over the sample period.

Table 2: Distance function estimation results

Variable	Parameter	Random parameter model with separable unobserved heterogeneity		Random parameter model with non-separable unobserved heterogeneity	
		Coefficient	Standard error	Coefficient	Standard error
Constant	α_i	-0.090*	0.008	0.031*	0.008
Ln(x$_2$/x$_1$)	α_{CT}	0.191*	0.007	0.243*	0.007
Ln(x$_3$/x$_1$)	α_{CB}	0.365*	0.012	0.357*	0.013
Ln(x$_2$/x$_1$)^2	α_{CTCT}	-0.051*	0.016	-0.060*	0.015
Ln(x$_3$/x$_1$)^2	α_{CBCB}	0.067*	0.028	0.124*	0.023
Ln(x$_2$/x$_1$)*ln(x$_3$/x$_1$)	α_{CBCT}	0.139*	0.014	0.098*	0.012
lny$_1$	β_T	-0.334*	0.006	-0.333*	0.006
lny$_2$	β_B	-0.485*	0.007	-0.456*	0.007
lny$_1$^{2}	β_{TT}	-0.113*	0.011	-0.110*	0.012
lny$_2$^{2}	β_{BB}	-0.174*	0.018	-0.179*	0.020

lny$_1$*lny$_2$	β_{BT}	0.114*	0.014	0.091*	0.015
Ln(x$_2$/x$_1$)*lny$_1$	δ_{TCT}	0.092*	0.013	0.086*	0.013
Ln(x$_2$/x$_1$)*lny$_2$	δ_{TCB}	-0.044*	0.014	-0.017	0.015
Ln(x$_3$/x$_1$)*lny$_1$	δ_{BCT}	-0.004	0.018	0.054*	0.017
Ln(x$_3$/x$_1$)*lny$_2$	δ_{BCB}	0.007	0.018	-0.084*	0.019
Trend	α_t	0.022*	0.001	0.022*	0.001
lnz$_1$	α_Z	-0.049*	0.006	-0.032*	0.006
lnz$_1$^{2}	α_{ZZ}	0.010	0.013	-0.033*	0.014
lnz$_1$*ln(x$_2$/x$_1$)	α_{ZT}	0.159*	0.010	0.138*	0.009
lnz$_1$*ln(x$_3$/x$_1$)	α_{ZB}	-0.119*	0.014	-0.109*	0.015
lnz$_1$*lny$_1$	α_{ZCT}	-0.122*	0.009	-0.131*	0.009
lnz$_1$*lny$_2$	α_{ZCB}	0.188*	0.009	0.206*	0.010
$\sigma = \sqrt{\sigma_u^2 + \sigma_u^2}$		0.123*	0.004	0.121*	0.004
$\lambda = \sigma_u / \sigma_v$		1.927*	0.225	2.322*	0.284
	Coefficients related to latent heterogeneity				
γ_i	η_1	0.136*	0.004	0.277*	0.008
γ_i *ln(x$_2$/x$_1$)	η_3			0.125*	0.010
γ_i *ln(x$_3$/x$_1$)	η_4			-0.130*	0.015
γ_i *lny$_1$	η_5			-0.021*	0.010
γ_i *lny$_2$	η_6			-0.023*	0.010
γ_i *lnz$_1$	η_7			0.024*	0.009
$\gamma_i * \gamma_i$	η_2	0.055*	0.006	0.093*	0.011
Wald Test H_0: $\eta_3 = \eta_4 = \eta_5 = \eta_6 = \eta_7 = 0$ Chi-squared = 526.95 p-value = 0.000 **H$_0$ is rejected**					

Notes: The coefficient reported for each random parameter is the mean; (a) we report estimates of SD of normal distribution of random parameters. (*) indicates significance at the 5% level.

The estimated coefficients (means for the random parameters) of the first-order terms have the expected signs and are statistically significant at the sample median. As expected, the coefficients of first-order output variables are negative and significantly different from zero implying that the estimated distance function is decreasing in outputs. The coefficients of the first-order terms of the capital and labor inputs are as expected positive and significantly different from zero. The sum of the coefficients of the two output variables is 0.79 or 0.82 (depending on the model). This result suggests the presence of economies of density at the sample median, because, *ceteris paribus*, by increasing both outputs by 10 percent, the input requirement will increase only by about 8 percent. As for the effect of network length, the results show that the first order term is, as expected negative and statistically significant. The sum of this coefficient with the two coefficients of the two output variables is 0.87 or 0.82. This result indicates the presence of economies of scale, because by increasing both outputs and network length by 10 percent, the input requirement will increase only by about 8 (9) percent.[24]

The negative coefficients of the output square terms for both bus and tram outputs, suggest that the rate of economies of scale is decreasing in each output. The positive coefficient of the interaction of the two outputs indicates cost-complementarity between tram and bus services. For instance, the results suggest that increasing one output by 10%, will result in 0.9 or 1.1 percent (depending on the model) decrease in the marginal cost of the other output. The effect of interactions with the network length suggest that providing bus services over longer networks is relatively less costly, while for trams, longer networks are associated with higher marginal costs. This result is

---

24 Note that in translog form, any statement about sample points other than the approximation point (here, sample median), should consider the second-order terms in addition to the main effects.

consistent with the fact that in tramways, the maintenance of the network infrastructure (rails and cables) in longer network might take relatively more capital and labor resources than in bus transport.

The table shows that in both models, the coefficients of the unobserved structural variable $(\eta_1 - \eta_7)$ are significantly different from zero at conventional 5% levels of significance. This provides empirical evidence for the presence of unobserved heterogeneity. Using a Wald test we tested the hypothesis of separability. The results (also listed in the table) favor the complete model, indicating that the unobserved network characteristics are not separable form observed production factors. Comparing the results across the two models indicates a close similarity in the coefficients of the first-order terms, suggesting that the estimates of returns to scale and other technological characteristics at the approximation point (here the sample median) are not sensitive to the assumption of separability. However, most second-order terms especially those related to network length (variable *Z*), vary across the two models. This suggests that quantities such as complementarity effects between different outputs as well as substitution elasticities between inputs could be sensitive to the assumptions related to separability from the unobserved network characteristics. The differences of second-order effects across the two models also suggest that the variation of the economies of scale at different levels of output and network length is sensitive to the separability assumption.

Studying the coefficients of the latent heterogeneity can be helpful in detecting the effects captured by that variable. The positive sign of the constant (η_1) indicates that higher levels of the latent variable (γ) are associated with network and environmental characteristics that are beneficial to production. Therefore the latent variable γ can be

interpreted as an aggregate indicator of network structural characteristics with an inverse correlation with network complexity. With this interpretation in mind, namely associating lower values of γ with greater network complexity, we can explore the consistency of the regression results with our underlying assumptions about network heterogeneity. The coefficients of the interactions of the unobserved heterogeneity with both outputs, tram seat-kilometers (η_5) and bus seat-kilometers (η_6), have a negative sign, implying that the network complexity has a lower effect in higher levels of output. Similarly, the positive coefficient of the interaction of the latent variable with the network length (η_7) suggests that the network complexity has a relatively greater effect in larger networks. The positive sign of the squared term of the latent variable (η_2) can also be interpreted as an increasing marginal effect of complexity. While all these interpretations appear to be consistent with the idea of linking the latent variable to network complexity, we should recognize that alternative interpretations could equally be justified. The results however point to the fact that the time-invariant heterogeneity is not separable from observed production factors.

The results listed in Table 2 also indicate considerable variation across companies with regard to time-invariant heterogeneity. The significant effect of interaction terms of the latent variable with outputs suggest that the technological characteristics such as the economies of scope or rates of returns to scale and density show a considerable variation across different companies. These variations are ignored in the model with separability assumption. In principle, such variations can be also modeled with a random coefficients model with several random effects. However, considering an identical latent variable allows a more tangible interpretation of such variations by associating them to

22

unobserved characteristics such as network complexity. For instance, considering the latent variable as an inverse measure of the network complexity, we can interpret the negative coefficients of the output interactions as an indication that more complex networks have higher rates of economies of scale.

The inefficiency scores u_i are summarized in Table 3. The estimated values of the inefficiency vary from 0.01 to about 0.62. The values of the mean and median technical inefficiency are fairly low amounting to about 8 percent of input resources.[25] A simple calculation based on the estimated effects of γ_i and γ_i^2, indicate that the effect of heterogeneity is more considerable: Considering the estimated coefficients in Table 2 (especially η_7), one standard deviation of this heterogeneity is approximately equivalent to about 0.14 or 0.28 depending on the model. These results suggest that the effect of time-invariant heterogeneity on inputs (and costs) is significantly greater than the average estimated inefficiencies. Moreover, the results suggest that these heterogeneities tend to be underestimated should they be assumed separable from observed production factors.

---

[25] For comparison purposes, we also computed the efficiency indices using the "classical" model for panel data proposed by Pitt and Lee (1981) who interpreted the random effects as inefficiency, thus considering any unobserved firm-specific heterogeneity as inefficiency. As expected, the values of technical inefficiency are higher and have more dispersion than the inefficiencies that emerge from our models.

Table 3: Descriptive statistics of inefficiency estimates

	Model 1 with separability assumption	Model 2 with non-separability assumption
Number of Observation	707	707
Mean	0.084	0.085
Std. Dev	0.053	0.057
Min	0.012	0.012
Median	0.071	0.069
Max	0.617	0.601

Table 4 provides the descriptive statistics of inefficiency estimates by country. These results indicate more or less similar efficiency scores across the two countries. The results of Kruskal-Wallis test on the differences in inefficiency scores (p-value of .58 or .91 depending on the model) suggest that there is no significant difference between the Swiss and German transit companies. This finding is valid in both models suggesting that the variation of efficiency within each country is greater than any systematic difference between the two countries. The estimates of latent heterogeneity on the other hand point to higher average values of γ in Switzerland compared to Germany. The statistical significance of these differences is confirmed by a Kruskal-Wallis test, suggesting that the German companies operate on networks that are relatively more complex than those in Switzerland.

Table 4: Descriptive statistics of country-specific efficiency estimates

	Model 1 with separability assumption	Model 2 with non-separability assumption
	Germany	
Number of Observation	616	616
Mean	0.085	0.086
Std. Dev	0.055	0.059
Min	0.012	0.012
Median	0.072	0.069
Max	0.617	0.601
	Switzerland	
Number of Observation	91	91
Mean	0.077	0.079
Std. Dev	0.035	0.039
Min	0.026	0.025
Median	0.068	0.074
Max	0.196	0.182

5. Summary and Conclusions

In this paper we examine the technical efficiency of a sample of Swiss and German urban transit companies. These companies are characterized by a high degree of heterogeneity in environmental and network characteristics. Due to lack of data and also because many of these structural factors such as network shape and complexity are not easily measured, only part of this heterogeneity is observed and can be considered in the input distance model specification. It is evident that the unobserved firm-specific heterogeneity becomes more serious in cross-country comparative efficiency analyses. Thus, it is important to use an appropriate SFA model, which is able to capture firm-specific unobserved heterogeneity using panel data.

Modeling heterogeneity in the empirical literature is often based on certain assumptions about the separability from the observed production factors. However, we argue that in our context the entire production process is organized around different network structures and shapes. Therefore, the unobserved heterogeneity is inevitably non-separable from the observed inputs and outputs. Against this background we propose a random coefficient stochastic frontier model assuming that unobserved heterogeneous factors are non-separable from the production process.

Similar to other panel data specifications such as 'true' random effects and random parameter frontier models proposed by Greene (2005a), the econometric model used in this study can be helpful to disentangle the unobserved time-invariant heterogeneity (such as network complexities) from the inefficiency estimates. The proposed model has however a distinctive feature in that such heterogeneities are represented by a single stochastic term that are not separable from the production process, while the inefficiencies are assumed to be uncorrelated thus separable from all production factors. Such distinctions between inefficiency and network heterogeneity could be used for a better identification of time-variant inefficiencies.

The results suggest that the estimated input distance function could be a reasonable fit to the observed data and that the estimated input and output elasticities have the correct sign and magnitude. The statistical tests favor the presence of considerable network heterogeneity and reject the separability assumption. Determining the scale elasticities we see that the median company operates under both economies of density and scale. Our analysis indicates that while the first-order coefficients of the distance function are not sensitive to the separability assumption, the second-order terms

26

could differ significantly across the models. This is especially important in estimating the variation of technological properties such as returns to scale and economies of scope with output and network characteristics. In these cases, the proposed model can be used to relax the separability assumption, while allowing a possible association between the variations with tangible structural characteristics such as network complexity.

The results of a Kruskal-Wallis test on the differences in inefficiency scores between the Swiss and German transit companies indicates no significant difference between these two types of companies. However, the statistical tests suggest that the German companies operate in networks and environments that are relatively more complex regarding the unobserved factors.

In general, the results indicate considerable variation across companies in the marginal impact of the observed input and outputs. This underlines that the unobserved characteristics of the network structure play a crucial role in transport services. Thus, the proposed model can improve the estimates taking into account different unobserved network complexities. However, this study along with the previous empirical literature suggests that given possible errors in the measurement of the efficiency level, the direct use of benchmarking results in regulation could have significant and possibly undesired financial consequences for the companies. Therefore, the benchmarking results should not be directly applied to define the tariffs applied to individual companies. However, the results can be used as an instrument to minimize the information asymmetry between the regulator and the regulated companies. For instance, benchmarking can be used as a guide to classify the companies into several efficiency groups.

6. References

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

Alvarez, A., Arias, C., and W. Greene (2004): 'Accounting for Unobservable in Production Models: Management and Inefficiency', Working Paper E2004/72, Fundacion Centro Estudios Andaluces.

Atkinson, S. E., and D. Primont (2002): 'Stochastic Estimation of Firm Technology, Inefficiency and Productivity Growth Using Shadow Cost and Distance Functions', *Journal of Econometrics*, 108(2), 203-225.

Bagdadioglu, N., and T. Weyman-Jones (2008): 'Panel Data Stochastic Frontier Analysis for Energy Network Regulation' Working Paper presented at Empirical Methods in Energy Economics, CEPE ETH Züruch, August 2008.

Battese, E., and T. J. Coelli (1988): 'Prediction of Firm Level Technical Inefficiencies with a Generalised Frontier Production Function and Panel Data', *Journal of Econometrics*, 38, 387-399.

Battese, G. E., and T. J. Coelli (1992): 'Frontier Production Function, Technical Efficiency and Panel Data: with Application to Paddy Farmers in India', *Journal of Productivity Analysis*, 3, 153-169.

Boitani, A., and C. Cambini (2002): 'Il Trasporto Pubblico Locale. Dopo la Riforma i Difficili Albori di un Mercato', *Mercato Concorrenza Regole*, 1, 45-72.

Cambini, C., and M. Filippini (2003): 'Competitive Tendering and Optimal Size in the Regional Bus Transportation Industry', *Annals of Public and Cooperative Economics*, 74(1), 163-182.

Coelli, T. J., and S. Perelman (1999): 'A comparison of Parametric and Non-Parametric Distance Functions: with Application to European Railways', *European Journal of Operations Research*, 117, 326-39.

Coelli, T. J., and S. Perelman (2000): 'Technical Efficiency of European Railways: A Distance Function Approach', *Applied Economics*, 32, 1967-1976.

Coelli, T. J. (2002): 'On the Econometric Estimation of the Distance Function Representation of a Production Technology', CEPA Working Paper, Centre for Efficiency and Productivity Analysis School of Economics, The University of Queensland, Queensland, Australia.

Colburn, C. B., and W. K. Talley (1992): 'A Firm Specific Analysis of Economies of Size in the U.S. Urban Multiservice Transit Industry', *Transportation Research Part B,* 3, 195-206.

Cornwell, C., Schmidt, P., and R. C. Sickles (1990): 'Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels', *Journal of Econometrics*, 46(1-2), 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.

De Borger, B., Kerstens, K., and A. Costa (2002): 'Public Transit Performance: What Does one Learn from Frontier Studies?', *Transport Reviews*, 22(1), 1-38.

Demsetz, H. (1968): 'Why Regulate Utilities?', *Journal of Law and Economics*, 11, 55–65.

Estache, A., Rossi, M. A., and C. A Ruzzier (2004): 'The Case for International Coordination of Electricity Regulation: Evidence from the Measurement of Efficiency in South America', *Journal of Regulatory Economics*, 25(3), 271-295.

Färe, R., and D. Primont (1995): *Multi-Output Production and Duality: Theory and Applications*, Kluwer Academic Publishers, Boston.

Farsi, M., Filippini, M., and W. Greene (2005): 'Efficiency Measurement in Network Industries: Application to the Swiss Railway Companies', *Journal of Regulatory Economics*, 28(1), 69–90.

Farsi, M., Fetz, A., and M. Filippini (2006a): 'Economies of Scale and Scope in Local Public Transportation', CEPE Working Paper No. 48, Centre for Energy Policy and Economics (CEPE), Zurich.

Farsi, M., Filippini, M., and M. Kuenzle (2006b): 'Cost Efficiency in Regional Bus Companies', *Journal of Transport Economics and Policy*, 40(1), 95–118.

Filippini, M., and R. Maggi (1992): The Cost Structure of the Swiss Private Railways, *International Journal of Transport Economics*, 19, 307-327.

Greene, W. (2004): 'Interpreting Estimated Parameters and Measuring Individual Heterogeneity in Random Coefficient Models', Department of Economics, Stern School of Business, Working Paper No. 04-08, New York University.

Greene, W. (2005a): 'Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model', *Journal of Econometrics*, 126(2), 269–303.

Greene, W. (2005b): 'Fixed and Random Effects in Stochastic Frontier Models', *Journal of Productivity Analysis,* 23, 7-32.

Hensher, D., and J. Stanley (2003): 'Performance-Based Quality Contracts in Bus Service Provision', *Transportation Research Part A*, 37(6), 519-538.

Hensher, D. (2007): *Bus Transport: Economics, Policy and Planning,* Research in Transportation Economics, Volume 18, Elsevier, Oxford, UK.

Heshmati, A., and S. C. Kumbhakar (1994): 'Farm Heterogeneity and Technical Efficiency: Some Results from Swedish Dairy Farms', *Journal of Productivity Analysis*, 5(1), 45-61.

Jamasb, T., and M. Pollitt (2003): 'International Benchmarking and Regulation: an Application to European Electricity Distribution Utilities', *Energy Policy*, 31, 1609-1622.

Klemperer, P. (1999): 'Auction Theory: a Guide to the Literature', *Journal of Economic Surveys*, 13(3), 227-86.

Kumbhakar, S. C. (1991): 'Estimation of Technical Inefficiency in Panel Data Models with Firm- and Time-Specific Effects', *Economics Letters*, 36, 43-48.

Laffont, J. J., and J. Tirole (1993): *A Theory of Incentives in Procurement and Regulation* The MIT Press, Cambridge, MA.

Lovell, C. A. K., Richardson, S., Travers, P., and L. L Wood (1994): 'Resources and Functioning: a New View of Inequality in Australia', in Eichhorn, W. (ed.) *Models and Measurement of Welfare and Inequality*, Springer Verlag, Berlin, 787-807.

O'Donnell C. J., and T. J. Coelli (2005): 'A Bayesian Approach to Imposing Curvature on Distance Functions', *Journal of Econometrics*, 126, 493–523.

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

Saal, D. S., Parker, D., and T. Weyman-Jones (2007): 'Determining the Contribution of Technical Change, Efficiency Change and Scale Change to Productivity Growth in the Privatized English and Welsh Water and Sewerage Industry: 1985-2000', *Journal of Productivity Analysis,* 28, 127-139.

Schmidt , P., and R. Sickles (1984): 'Production Frontiers and Panel Data', *Journal of Business and Economics Statistics*, 2, 367–374.

Shleifer, A. (1985): 'A Theory of Yardstick Competition', *Rand Journal of Economics*, 16(3), 319-327.

Sickles, R.C., Good, D. H., and L. Getachew (2002): 'Specification of Distance Functions using Semi- and Non-parametric Methods with an Application to the Dynamic Performance of Eastern and Western European Air Carriers', *Journal of Productivity Analysis,* 17(1-2), 133–155.

Toner, J. P. (2001): 'The London Bus Tendering Regime. Principles and Practice', Presented at the VII International Conference on Competition and Ownership in Land Passenger Transport, Molde, Norway.

Tsionas, E. G. (2002): 'Stochastic Frontier Models with Random Coefficients', *Journal of Applied Econometrics*, 17, 127-147.

Viton, P. A. (1992): 'Consolidations of Scale and Scope in Urban Transit', *Regional Science and Urban Economics*, 22(1), 25-49.

Viton, P. A. (1993): 'How Big Should Transit Be? Evidence on the Benefits of Reorganization from the San Francisco Area', *Transportation*, 20, 35-57.