Document de travail du LEM / Discussion paper LEM
2019-08

Inference in Aigner and Chu Cost Frontier Estimation: The Impact of Infrastructure on Bus Transportation in France.

Pierre OUELLETTE

Department of Economics, Université du Québec à Montréal / ouellette.pierre@uqam.ca

Stéphane Vigeant

LEM UMR 9221 / stephane.vigeant@univ-lille.fr

Linjia ZHANG

International Business School Suzhou / linjia.zhang@xjtlu.edu.cn

http://lem.cnrs.fr/

Inference in Aigner and Chu Cost Frontier Estimation: The Impact of Infrastructure on Bus Transportation in France.

Pierre Ouellette
Department of Economics,
Université du Québec à Montréal
PO Box 8888, Station Centre-Ville
Montréal, Québec
Canada H3C 3P8
E-mail: ouellette.pierre@uqam.ca

Stéphane Vigeant
Université de Lille, LEM – CNRS (UMR 9221)
Faculté des Sciences Économiques et Sociales
Université de Lille
59655 Villeneuve d'Ascq Cédex, France
and
IESEG School of Management
E-mail: Stephane.Vigeant@univ-lille.fr

and

Linjia Zhang
International Business School Suzhou,
Xi'an Jiaotong-Liverpool University,
111 Ren'ai Rd
Suzhou Dushu Lake Science and Education Innovation District
Suzhou Industrial Park
Suzhou
P.R. China
linjia.zhang@xjtlu.edu.cn

March 2019

Abstract

In this work, we estimate a cost frontier using a statistical semi-parametric approach. The development of this model is based on the frontier estimation approach proposed by Aigner and Chu (1968). The approach consists in specifying a functional form for the frontier, in this case a cost function, and then to compute the cost inefficiency based on an optimization criterion (either a linear or a quadratic program). Although it has been shown that semi-parametric models produce maximum likelihood estimators of the parameters, no standard errors for the coefficients are available. Ouellette *et al.* (2016) have shown the consistency of the estimator, however. Then, inference is conducted using a subsampling argument to derive empirical distributions for the parameters of the cost frontier and for all technology measurements of interest using simulation techniques similar to the bootstrap. An empirical application of the method to the French bus transportations industry has been made. A firm specific infrastructure index has been constructed for each firm based on an accessibility index capturing the road network and its congestion. Our results show that French bus transportation firms are relatively well behaved and operate under increasing returns to scale. Infrastructures are shown to be a substitute to labor as an increase in the infrastructure index (more roads and less congestion) reduces the time spent by buses on the roads. As a consequence, infrastructure has a negative effect on variable cost for these firms. The large firms are on average more cost efficient than the small ones. These results are statistically significant at either size 5% or 1% based on our simulation results.

Keywords: Cost efficiency, Semi-parametric methods, Subsampling, French bus transportation, Quasi-fixed inputs, Infrastructure.

JEL Codes: C51, C61, D24, L25

1. Introduction

Roughly, the measurement of cost-efficiency of firms consists in enveloping the observed cost, thereby creating a frontier to a set containing all the realized costs, and then measuring the distance between the observed performance of the firm and this estimated frontier. This estimated cost frontier is assumed to be supported by a true frontier that bears a large number of economic concepts describing the technology of the firm. For instance, the true frontier gives the lowest cost possible to produce a given output, and doing so it separates the feasible productions from those that are not. It also reveals some information on the optimal input mix given the price of the production factors and, of crucial importance for analyst of an industrial structure, it can be used to measure the returns to scale of firms. In an environment where all the inputs are under the direct control of the firms at decision time, information on input prices, output and cost is enough to assess the performance of firms. However, it is rarely the case that firms control all their inputs at decision time, some are following a complex investment process and cannot be set to their optimal value instantaneously (capital often corresponds to this characterization) and sometimes they just do not control the inputs at all (public infrastructures have this characteristic). In this paper, we are interested in measuring the cost efficiency of firms with these characteristics. Our task is to assess the cost-efficiency and technology (in particular the returns to scale) of an intercity bus transportation industry. In this industry, the firms can usually control at decision time labor, gas and the likes but plan overtime the number of buses they will use and are totally dependent on the road network. Because the infrastructures are supplied by the State, it seems natural to try to determine to what extent the quantity is optimal or not. To do so, returns to scale would turn out to be a good indicator of the performance of the industry.

The characterization of the frontier rests on an axiomatic that provides some indication on what we should expect from the frontier itself. This cost-frontier is not directly observable however and must be estimated. This paper proposes to use a semi-parametric approach to estimate the frontier based on the method proposed by Aigner and Chu (1968). This method in its original setting is fully deterministic and was mostly applied to production function. Here we propose to estimate a cost function using this method. The deterministic represents an important limitation, but it has been shown that this estimator is in fact a maximum likelihood estimator. Despite the fact that it was a maximum likelihood estimator, the method was abandoned in favor of stochastic frontier (SFA) because essential statistical properties were not proved and were shown to be difficult to implement. However, Ouellette *et al*.

(2016) have shown the estimator to be consistent and they also proposed a way to conduct inference for the relevant statistics using sub-sampling. We use here their approach to the production frontier to apply it to cost-frontier.

We also address the infrastructure measurement in detail in this paper. It is clear that the location of a firm in the network would impact its ability to deliver transportation service. It determines its ability to get to the clients to transport and then to deliver them where they have to go. Consequently, infrastructure is a compound of accessibility to the network and the extent of the network availability (including the congestion time). To capture the firm specific effect of the infrastructure network we use an accessibility index that would account for all these features. Application of our method on standard firm data with our infrastructure index shows that most of the firms operate under increasing returns to scale and that result is statistically significant. This suggests that the firms in this industry are not operating at their optimal size and mergers are probably required. It also suggests that infrastructures are the source of theses returns to scale, as they are not constraining the development of the industry. In fact, the source of the problem might be more in the regulated environment of the firms.

2. The dual semi-parametric model.

In this section we present the dual semi-parametric model. It is an adaptation of the Aigner and Chu (1968) production frontier model to the dual environment where a cost frontier is estimated instead of a production frontier. Suppose that firms produce an output, y, using P variable inputs, denoted x_p for $p = 1,\dots,P$ and Q quasi-fixed inputs, denoted k_q for $q = 1,\dots,$Q. The technology of the firm is given by a transformation function $f\left(y,\boldsymbol{x},\boldsymbol{k}\right) £ 0$. Suppose that the price of the variables factors are available, we denote them as w_p for $p = 1,\dots,P$ and that the input markets are competitive, then this technology can be represented in the dual space by a variable cost function given by the solution to the following problem:

$$C\left(\boldsymbol{w},\boldsymbol{k},y\right) = \min_{x}\left\{\boldsymbol{w}^T\boldsymbol{x} \mid f\left(y,\boldsymbol{x},\boldsymbol{k}\right) £ 0\right\}. \qquad (2.1)$$

The conditional factor demands are obtained from Shephard's Lemma:

$$\boldsymbol{x} = \boldsymbol{x}\left(\boldsymbol{w},\boldsymbol{k},y\right) = C_w\left(\boldsymbol{w},\boldsymbol{k},y\right). \qquad (2.2)$$

The cost returned by the cost function is the smallest cost reachable given the technology.

We need now to relate these theoretical quantities to their empirical or observed equivalent. It is important at this stage to single out the optimal behavior from the realized performance by a firm. Suppose that we can observe n firms, identified by the subscript i=1,…,n and define the observed or realized cost as:

$$C_i^{obs} = \sum_{p=1}^{P} w_p^{obs} x_{ip}^{obs}, \tag{2.3}$$

where C_i^{obs} and x_{ip}^{obs} are the observed cost and quantity of input p of firm i respectively and w_p^{obs} is the observed price of input p. As expected, in practice there will be a difference between the minimal theoretical cost and the observed cost. There are two sources possible to explain this difference: measurement errors and optimization errors. As both of these errors can have a purely stochastic component, it is clear that optimization errors could be associated to repeated bad management practices in an environment flawed by defective incentives. Suppose that the error structure is additive, the observed prices, inputs and the observed output are given by:[1]

$$w_p^{obs} = w_p + \nu_{w_p} \tag{2.4}$$

$$x_{ip}^{obs} = x_{ip} + \varepsilon_{x_p} + \nu_{x_p} \tag{2.5}$$

$$y_i^{obs} = y_i + \nu_y \tag{2.6}$$

$$k_{iq}^{obs} = k_{iq} + \nu_{k_q}, \tag{2.7}$$

where w_p, x_{ip}, k_{iq}, and y_i are the true values of the variable and quasi-fixed inputs and the output, respectively, ν_{w_p} (resp. $\nu_{x_{ip}}, \nu_{q_{iq}}, \nu_{y_i}$) is the measurement error of variable w_p (resp. x_{ip}, k_{iq}, y_i), and ε_{x_p} is the optimization error on variable input x_p. We assume there are no optimization errors on the prices ($\varepsilon_{w_p} = 0, \forall p$) as we assume the input market to be competitive. Since the output and quasi-fixed factors are given at decision time, they are not subject to optimization errors. Using these definitions, we can deduce the gap between the observed cost and the variable cost:

$$C(\boldsymbol{w},\boldsymbol{k},y) \equiv \boldsymbol{w}^T \boldsymbol{x}(\boldsymbol{w},\boldsymbol{k},y) = C^{obs} - (\boldsymbol{w}^{obs})^T (\varepsilon_{\boldsymbol{x}} + \nu_{\boldsymbol{x}}) - (\boldsymbol{x}^{obs})^T \nu_{\boldsymbol{w}} + \nu_{\boldsymbol{w}}{}^T (\varepsilon_{\boldsymbol{x}} + \nu_{\boldsymbol{x}}) \tag{2.8}$$

or, more conveniently:

$$C^{obs} = C(\boldsymbol{w},\boldsymbol{k},y) + \left[(\boldsymbol{w}^{obs})^T \nu_{\boldsymbol{x}} + (\boldsymbol{x}^{obs})^T \nu_{\boldsymbol{w}} - \nu_{\boldsymbol{w}}{}^T \nu_{\boldsymbol{x}} \right] + (\boldsymbol{w}^{obs} - \nu_{\boldsymbol{w}})^T \varepsilon_{\boldsymbol{x}}. \tag{2.9}$$

[1] The model we develop here follows Ouellette and Petit (2010). In this paper they also develop a multiplicative error model.

The observed cost is decomposed into three components; the minimal variable cost $C(\boldsymbol{w},\boldsymbol{k},y)$, the measurement errors $\nu_c = (\boldsymbol{w}^{obs})^T \nu_{\boldsymbol{x}} + (\boldsymbol{x}^{obs})^T \nu_{\boldsymbol{w}} - \nu_{\boldsymbol{w}}{}^T \nu_{\boldsymbol{x}}$, and the optimization errors, $\varepsilon_c = (\mathrm{w}^{obs} - \nu_{\mathrm{w}})^T \varepsilon_{\mathrm{x}}$. The deviation from the optimal costs is denoted $\mu = \varepsilon_c + \nu_c$.

Now, we have the elements to adapt the Aigner and Chu (1968) (hereafter A-C) production frontier estimation model to variable cost frontier. From (2.9) we have:

$$C^{obs} - C\left(\boldsymbol{w}^{obs} - \nu_w, \boldsymbol{k}^{obs} - \nu_k, y^{obs} - \nu_y\right) - \varepsilon_c - \nu_c = 0. \tag{2.10}$$

To implement A-C estimation procedure, we need a functional form for the cost function. Suppose it is linear in the parameters (a Cobb-Douglas or a translog functional form are good candidates, but let us keep it in its simplest form for the moment) so that we can rewrite equation (2.10) as follows:

$$C^{obs} - (\boldsymbol{w}^{obs} - \nu_{\boldsymbol{w}})^T \beta_{\boldsymbol{w}} - \left(y^{obs} - \nu_y\right)^T \beta_y - \left(\boldsymbol{k}^{obs} - \nu_{\boldsymbol{k}}\right)^T \beta_{\boldsymbol{k}} - \varepsilon_c - \nu_c = 0. \tag{2.11}$$

As in A-C, suppose that measurement errors are null for all variables, i.e., $\nu_{\boldsymbol{w}} = 0, \nu_{\boldsymbol{x}} = 0$, $\nu_{\boldsymbol{k}} = 0$ and $\nu_y = 0$ then we have $\nu_c = (\boldsymbol{w}^{obs})^T \nu_{\boldsymbol{x}} + (\boldsymbol{x}^{obs})^T \nu_{\boldsymbol{w}} - \nu_{\boldsymbol{w}}{}^T \nu_{\boldsymbol{x}} = 0$ and $\nu_c = (\boldsymbol{w}^{obs} - \nu_x)^T \varepsilon_{\boldsymbol{x}} = \nu_c = (\boldsymbol{w}^{obs})^T \varepsilon_x$. This leaves us with:

$$C^{obs} - (\boldsymbol{w}^{obs})^T \beta_{\boldsymbol{w}} - \left(y^{obs}\right)^T \beta_y - \left(\boldsymbol{k}^{obs}\right)^T \beta_{\boldsymbol{k}} - (\boldsymbol{w}^{obs})^T \varepsilon_x = 0 . \tag{2.12}$$

Denote the optimization error term (or inefficiency) as $\varepsilon = (\boldsymbol{w}^{obs})^T \varepsilon_x$ then impose the non-negativity constraint on this term as in A-C, to get:

$$C^{obs} - (\boldsymbol{w}^{obs})^T \beta_{\boldsymbol{w}} - \left(y^{obs}\right)^T \beta_y - \left(\boldsymbol{k}^{obs}\right)^T \beta_{\boldsymbol{k}} = \varepsilon \geq 0. \tag{2.13}$$

Now, the estimation procedure will consists in minimizing an aggregate of the optimization errors, for example $\sum_{i=1}^{n} \varepsilon_i$, by choice of β_w, β_k and β_y under the constraint that $\varepsilon_i \geq 0$ for all $i = 1,...,n$.

To illustrate the procedure, suppose that the variable cost function for *P* variable inputs, *Q* quasi-fixed inputs and *S* outputs is of the Cobb-Douglas type. That is:

$$C_i = e^{\alpha_o} \prod_{p=1}^{P} w_{pi}^{\beta_p} \prod_{q=1}^{Q} k_{qi}^{-\beta_s} \prod_{s=1}^{S} y_{si}^{\beta_q} U_i , \tag{2.14}$$

where $U_i \geq 1$ is the cost efficiency of firm *i*. Log-linearization of (2.14) gives:

$$\ln C_i = \alpha_o + \sum_{p=1}^{P} \beta_p \ln w_{pi} - \sum_{q=1}^{Q} \beta_q \ln k_{qi} + \sum_{s=1}^{S} \beta_s \ln y_{si} + \varepsilon_i \,, \tag{2.15}$$

where $\varepsilon_i = \ln U_i$ is the non-negative optimization error term.

To solve for the parameters, we need an optimization criterion, as we mentioned above. A-C suggests two possible criteria. The first criterion is to minimize the sum of the errors, $\sum_{i=1}^{n} \varepsilon_i$, by choice of β_w, β_k and β_y under the constraint that $\varepsilon_i \geq 0$ for all $i = 1,...,n$ and the non-negativity of the parameters and the input price homogeneity of the cost function. This essentially leads to the following linear programming optimization problem:

$$\begin{aligned}
&\min_{(\alpha,\beta)\geq 0} \sum_{i=1}^{n} \varepsilon_i = \min_{(\alpha,\beta)\geq 0} \sum_{i=1}^{n} \left(\ln C_i - \left(\alpha_o + \sum_{p=1}^{P} \beta_r \ln w_{ri} - \sum_{q=1}^{Q} \beta_q \ln k_{qi} + \sum_{s=1}^{S} \beta_s \ln Y_{si} \right) \right) \\
&s.t. \\
&\ln C_i - \left(\alpha_o + \sum_{p=1}^{P} \beta_r \ln w_{ri} - \sum_{q=1}^{Q} \beta_q \ln k_{qi} + \sum_{s=1}^{S} \beta_s \ln Y_{si} \right) \geq 0 \;\; \forall\, i \\
&\sum_{p=1}^{P} \beta_r = 1.
\end{aligned} \tag{2.16}$$

The other criterion proposed by A-C is to minimize the sum of the square of the optimization errors, under the same constraints as in the linear program. This essentially leads to a quadratic programming problem. That is:

$$\begin{aligned}
&\min_{(\alpha,\beta)\geq 0} \sum_{i=1}^{n} \varepsilon_i^2 = \min_{(\alpha,\beta)\geq 0} \sum_{i-1}^{n} \left(\ln C_i - \left(\alpha_o + \sum_{p=1}^{P} \beta_r \ln w_{ri} - \sum_{q=1}^{Q} \beta_q \ln k_{qi} + \sum_{s=1}^{S} \beta_s \ln Y_{si} \right) \right)^2 \\
&s.t. \\
&\ln C_i - \left(\alpha_o + \sum_{p=1}^{P} \beta_r \ln w_{ri} - \sum_{q=1}^{Q} \beta_q \ln k_{qi} + \sum_{s=1}^{S} \beta_s \ln Y_{si} \right) \geq 0 \;\; \forall\, i \\
&\sum_{p=1}^{P} \beta_r = 1.
\end{aligned} \tag{2.17}$$

We are not restricted to use a Cobb-Douglas functional form, the only restriction we have so far, is that it has to be linear in the parameters. In our application we will use a translog functional form. Note also, that once we have the cost function, it is possible to estimate the characteristics of the technology such as the returns to scale, the input and output flexibility, elasticity of substitution, etc.

3. The Statistical Model

The problem with the procedure proposed by A-C is that it is purely deterministic and consequently it is not possible to conduct any inference on the estimated parameters and the measured quantities that characterizes the technology of the firms. It has been shown by Schmidt (1975) that both, the linear programming estimator and the quadratic programming estimator, are maximum likelihood estimators. However, it has been shown by Greene (1980) that standard asymptotic methods cannot be used to characterize the estimators. Ouellette *et al*. (2016) have shown however that the estimator is consistent and, in spite of the fact the asymptotic distribution cannot be ascertain, they have suggested a simulation procedure to obtain an empirical distribution of the parameters. Consequently, those estimators can be used to for statistical inference. The section present an estimator adapted from and the simulation procedure we used for statistical inference.

3.1 Maximum Likelihood Estimation

For all units *i*, with $i=1,\ldots,n$, we suppose that the frontier is linear in the parameters, and we denote it $\alpha + \boldsymbol{X}_i\boldsymbol{\beta}$, that all observed costs, C_i, are on or above the frontier, and as in the previous section, ε_i is the cost inefficiency. This gives:

$$C_i = \alpha + \boldsymbol{X}_i\beta + \varepsilon_i, \tag{3.1}$$

where $\boldsymbol{X}_i = \begin{bmatrix} X_{i1} & \cdots & X_{ik} \end{bmatrix}$ is a $(1\times k)$ vector of inputs for unit i, the scalar $\alpha \in \mathbb{R}$ and the $(k\times 1)$ vector $\beta = \begin{bmatrix} \beta_1 & \cdots & \beta_k \end{bmatrix}^T$ are the parameters and the scalar ε_i is non negative. Keeping in mind that $\varepsilon_i \geq 0$, it follows that $C_i \geq \alpha + \boldsymbol{X}_i\beta$. The functional form is linear in the parameters, but not necessarily in the variables and is therefore compatible with a translog functional form. [2]

Suppose now that the inefficiency term is exponentially distributed. That is, $f(\varepsilon,\sigma) = (1/\sigma)\exp(-\varepsilon/\sigma)$ where $\varepsilon > 0$ and $\sigma > 0$. The true parameter vector is $[\alpha_0 \quad \beta_0]$ so that $\varepsilon_i = C_i - \alpha_0 - \boldsymbol{X}_i\beta_0$. The log-likelihood is:

[2] That is, X and C are not restricted to be in level. When the functional form is a translog, we use the logarithm of the cost and of the prices, output and quasi-fixed inputs included in X.

$$\ell\left(\sigma,\alpha,\beta \mid y\right)=\ln L\left(\sigma,\alpha,\beta \mid y\right)=-n\ln\sigma-\frac{1}{\sigma}\sum_{i=1}^{n}(C_i-\alpha-\boldsymbol{X}_i\beta), \tag{3.2}$$

where $C_i-\alpha-\boldsymbol{X}_i\beta\geq 0, \forall\, i=1,\ldots,n$. Solving the first order conditions for $\hat{\sigma}_n$ and substituting back into equation (3.2) gives the concentrated likelihood function:

$$\ell^{c}\left(\alpha,\beta \mid y\right)=-n\ln\hat{\sigma}_n-\frac{n}{\hat{\sigma}_n}\hat{\sigma}_n=-n\ln\left\{\frac{1}{n}\sum_{i=1}^{n}(C_i-\alpha-\boldsymbol{X}_i\beta)\right\}-n. \tag{3.3}$$

Maximizing $\ell^{c}\left(\alpha,\beta \mid y\right)$ consists in minimizing the first term under the non-negativity constraint on the ε_i. But, this is equivalent to:

$$\begin{aligned}&\min_{\alpha,\boldsymbol{\beta}}\sum_{i=1}^{n}\left(C_i-\alpha-\boldsymbol{X}_i\beta\right)\\ &s.t.\quad C_i\geq\alpha+\boldsymbol{X}_i\beta, \forall i=1,\ldots,n.\end{aligned} \tag{3.4}$$

This is exactly Aigner and Chu's linear program. As in Schmidt (1975) we have the following result:

Proposition 3.1. *The solution to equation (3.4) is a maximum likelihood estimator when the inefficiency terms follow an exponential distribution.*

This is an interesting result, but it is relevant only if we can use it to construct some inference for the estimated parameters. Unfortunately, the standard proofs for consistency and asymptotic normality do not work here. As was shown by Greene (1980) and Ouellette *et al*. (2016) the expectation of the score is not null and the Hessian matrix is singular. Consequently, showing consistency and asymptotic normality is not a simple endeavor. However, if we can show the consistency of the estimator, it might be possible to use re-sampling methods to construct confidence intervals for the estimator, despite the fact that we do not have the asymptotic distribution.

Ouellette *et al*. (2016) have shown that the estimator is consistent using Wald's (1949) consistency proof of the maximum likelihood estimator. In order to use this proof, assume that:

Assumption 3.1. *The parameter space, Θ, is closed and is a subset of $\mathbb{R}^{k+2}$ such that for a positive scalar we have $\Theta\subseteq\left[-M,\infty\right)^{k+1}\times\left[\underline{\sigma},\infty\right)\subseteq\mathbb{R}^{k+2}$ with $\underline{\sigma}>0$.*

This assumption forces the exponential distribution to be well-defined for admissible ε_i (they are all positive). It also implies that firms are inefficient on average and they cannot be all efficient. In fact, efficiency is a probability zero event in this model. The consistency result is given by the following proposition.

Proposition 3.2. *Let* $\theta = (\sigma, \alpha, \beta)$ *and define the maximum likelihood estimator,* $\hat{\theta}_n$*, as the argument that maximizes* $-n\ln\sigma - (1/\sigma)\sum_{i=1}^{n}\left(C_i - \alpha - \boldsymbol{X}_i\beta\right)$ *subject to* $C_i > \alpha + \boldsymbol{X}_i\beta$*. That is:*

$$\hat{\theta}_n = \arg\max_{\alpha,\beta,\sigma}\left\{-n\ln\sigma - \frac{1}{\sigma}\sum_{i=1}^{n}\left(C_i - \alpha - \boldsymbol{X}_i\beta\right) \mid C_i \geq \alpha + \boldsymbol{X}_i\beta\right\}.$$

Then, under Assumption 3.1, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$ *is consistent,* $\operatorname{plim}_{n\to\infty}\hat{\theta}_n = \theta_0$.

Proof. Ouellette *et al*. (2016).

Let us consider the quadratic model of Aigner and Chu. Suppose that ε_i follows a half-normal distribution:

$$f\left(\varepsilon, \sigma^2\right) = \sqrt{\frac{2}{\pi\sigma^2}}\exp\left\{-\frac{\varepsilon^2}{2\sigma^2}\right\}, \tag{3.5}$$

where $\varepsilon \geq 0$. Given a sample of *n* firms, the log-likelihood is

$$\ell\left(\sigma^2, \alpha, \boldsymbol{\beta} \mid y\right) = \frac{n}{2}\ln\left(\frac{2}{\pi}\right) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(C_i - \alpha - \boldsymbol{X}_i\beta\right)^2 \tag{3.6}$$

subject to the constraint that $C_i - \alpha - \boldsymbol{X}_i\beta \geq 0$. Borrowing the technic we used with the exponential distribution, it is easy to conclude that:

Proposition 3.3. *The solution to* $\min_{\alpha,\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left(y_i - \alpha - \boldsymbol{X}_i\beta\right)^2 : C_i \geq \alpha + \boldsymbol{X}_i\beta\right\}$ *is a maximum likelihood estimator when the cost-inefficiency is half-normal distributed.*

Again, we face the same problem we had with the exponential distribution; the consistency and asymptotic normality of the estimator cannot be proved using the standard method. But we can use again Ouellette *et al*. (2016) result that shows that the estimator is consistent. Suppose that the following condition holds:

Assumption 3.2. *The parameter space,* Θ, *is closed and is a subset of* $\mathbb{R}^{k+2}$ *such that for a positive scalar* M *we have* $\Theta \subseteq \left[-M,\infty\right)^{k+1} \times \left[\underline{\sigma}^2,\infty\right) \subseteq \mathbb{R}^{k+2}$ *with* $\underline{\sigma}^2 > 0$.

We have to allow for negative parameters because we might want to specify a translog production function. This assumption imposes inefficiency for some units; this is due to the fact that the distribution cannot degenerate at zero. As above, efficiency is a probability zero event. Consequently:

Proposition 3.4. *Let* $\theta = (\sigma, \alpha, \beta)$ *and define the maximum likelihood estimator,* $\hat{\theta}_n$, *as the argument that maximizes the likelihood function:*

$$\ell\left(\sigma^2,\alpha,\beta \mid y\right) = \frac{n}{2}\ln\left(\frac{2}{\pi}\right) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(C_i - \alpha - \boldsymbol{X}_i\beta\right)^2$$

under the constraint $C_i \geq \alpha + \boldsymbol{X}_i\beta$. *Under Assumption 3.2, the maximum likelihood estimator,* $\hat{\theta}_n$, *is consistent, i.e.* $\text{plim}_{n\to\infty}\hat{\theta}_n = \theta_0$.

Proof. Ouellette *et al*. (2016).

3.2 Inference in the Semi-Parametric Model

In this sub-section we present an approach for statistical inference in the Aigner and Chu model. We do not have the distribution of the parameters, but we know at least that they are consistent. This feature allows us to simulate the DGP in order to deduce some statistical properties of the estimator.

The crucial step is to generate pseudo-samples that are consistent with the DGP. There are many methods, but the problem at stake here drives the procedure to be adopted. We have to deal with the fact that the parameters of interest are on the boundary of the support of the random variable (the inefficiency). Simar and Wilson (1998, 1999) and Andrews (2000) document this problem and it is well known that in these circumstances naïve bootstrap does not return consistent results. Consequently, we follow Politis and Romano (1992, 1994) and use sub-sampling. Bickel *et al.* (1997) and Andrews (2000) detail the properties of sub-sampling.

The re-sampling principle is as follows. Suppose we have a sample of n *i.i.d.* random vectors, $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$, and an estimator, $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}\left(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\right)$. We wish to characterize the precision of that estimator. Since the empirical cumulative distribution F_N is a consistent maximum likelihood estimator of the true distribution, then it can be used to generate pseudo-samples to estimate $\hat{\boldsymbol{\beta}}^b = \boldsymbol{\beta}\left(\boldsymbol{X}_1^b, \boldsymbol{X}_2^b, \ldots, \boldsymbol{X}_m^b\right)$ with *m* not necessarily equal to *n*. In the bootstrap world, $\hat{\boldsymbol{\beta}}^b - \hat{\boldsymbol{\beta}}$ would have the same distribution as $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ so that replicating B times the estimation on pseudo-samples can give us a set of values that approximate the true distribution. In the case of sub-sampling, we choose pseudo-sample of size *m* strictly smaller than the size of the original sample, *n*. For the convergence of the procedure we need that *m* satisfies $m \to \infty$ and $m/n \to 0$ when $n \to \infty$. The pseudo-samples need not be drawn with replacement. Politis, Romano and Wolf (1999) have shown that sub-sampling procedure converges under weaker conditions than the standard bootstrap. Simar and Wilson (2011) show that a sub-sampling method would generate a consistent inference even if the frontier depends on the estimated parameters. This is the procedure we apply.

Before constructing the confidence interval, note that, in spite of the fact that the Aigner and Chu estimator is consistent, it is also biased in small samples. Since the bias is given by $\mathrm{Bias}\left[\hat{\beta}_i\right] = \mathrm{E}(\hat{\beta}_i) - \beta$ we can correct it using sub-sampling as follows:

$$Bias_B\left[\hat{\beta}_i\right] = B^{-1}\sum_{b=1}^{B}\hat{\beta}_i^b - \hat{\beta}_i, \tag{3.7}$$

where *B* is the number of sub-samples generated. The bias corrected estimate is:

$$\hat{\hat{\beta}}_i = \hat{\beta}_i - Bias_B\left[\hat{\beta}_i\right] = 2\hat{\beta}_i - B^{-1}\sum_{b=1}^{B}\hat{\beta}_i^b. \tag{3.8}$$

To construct the confidence intervals we follow the procedure proposed by Hall (1992a) and Efron and Tibshirani (1993). We wish to construct a confidence interval for $\hat{\beta} - \beta$. Given the size of the interval, α, this consists in finding a_α and b_α such that:

$$\Pr\left(-b_\alpha \leq \hat{\beta} - \beta \leq -a_\alpha\right) = 1 - \alpha. \tag{3.9}$$

The problem is that we do not know the probability distribution. However, we know that when *B* tends to infinity, the following condition holds:

$$\left[\hat{\beta} - \beta\right]\Big| P \sim \left[\hat{\beta}^b - \hat{\beta}\right]\Big| P. \tag{3.10}$$

So we are justified to use the empirical distribution $\hat{\beta}^b - \hat{\beta}$ for b = 1,…,B to approximate the distribution of $\hat{\beta} - \beta$. Therefore, for each parameter $i=0,\dots,k$, sort the $\hat{\beta}_i^b(x,y) - \hat{\beta}_i(x,y)$, b = 1,…,B in ascending order and eliminate $(\alpha/2)\times 100\,\%$ elements at each end of the distribution. The corresponding values of $\hat{\beta}_i^{\pm(\alpha/2)\times 100\,\%} - \hat{\beta}_i$ are the empirical values of $-a_\alpha^*$ and $-b_\alpha^*$. In equation (3.9) replace a_α^* and b_α^* by a_α and b_α, and $\left[\hat{\beta} - \beta\right]$ by $\left[\hat{\beta}^*(x,y) - \hat{\beta}\right]$, to obtain:

$$\Pr(-b_\alpha^* \le \hat{\beta} - \beta_0 \le -a_\alpha^* \mid x) \approx 1 - \alpha. \tag{3.11}$$

Finally, a $(1-\alpha)$-confidence interval is given by:

$$\hat{\beta} + a_\alpha^* \le \beta_0 \le \hat{\beta} + b_\alpha^*. \tag{3.12}$$

In the following section we present an application of this methodology.

4. Data

We apply in this paper the methodology presented above on the French bus transportation industry. We have excluded all public transit companies to focus on the inter-city transportation of passengers. The data are for the period 2000-2004 and after homogenization of the data we obtain 2554 DMU, roughly 500 DMU per year. Note that it is not a panel, as firms entered, exited and restructured over that period, making it difficult to follow them over the entire period. Excluding firms to obtain a balanced panel would have shrunk the sample considerably.

We constructed our database mainly from two data sources. Firstly, we obtained company data from the *Annual Survey of Companies in the Transport Sector* (*ASCT*), which is provided by the *National Institute of Statistics and Economic Studies* (INSEE). *ASCT* is the only national survey on firms, and it was made by INSEE in partnership with the statistical offices of all departments in France. It provides a range of information on the companies whose transport is the main activity. Most of the information relates to the usual accounting results (turnover and its decomposition, account balances, investment, etc.). It also provides information on the specific charges to transport such as kilometers driving, maintenance and repairs, insurance premiums, fuel purchases, etc. The company database was completed with price indexes from INSEE. Secondly, the data used to construct the road infrastructure index are obtained from *French General Commission for Sustainable Development* (*GCSD*). *GCSD* provided us with a database that included the length of the highway network, the average speed of vehicles, congestion time, the distance between each study area (region, county, area of employment, municipalities, etc.), and the flow of goods between these areas.

We have divided in two subsections the description of the database used. The first subsection describes the source and method to construct the output, inputs, price and cost variables, which is the standard part of the database. The next subsection focuses on a less standard variable, the infrastructure index used to capture the firm specific access to the road network.

4.1 Prices, Quasi-fixed Inputs and Output

The production process used in this paper is divided into two types of inputs, variable inputs (which can be adjusted by firms instantly) and quasi-fixed inputs (for which the quantity cannot be adjusted in short term). For the variable inputs, we require prices variables while for the quasi-fixed we need quantities. Given data availability, we have used two variable inputs, labor and a material aggregate, two quasi-fixed inputs, capital and infrastructures, and an output.

The definition of the inputs closely follows the tradition in the transportation literature. The firms' variable inputs are labor and an aggregate of different component coming from the operating budget. Variable cost is then the sum of the labor and material expenditure. Consequently, we need to have two of the following variables: price, quantity and expenditure for each variable. The labor price is traditionally measured by the wage, but this information is not directly available in our database. However, we have the total amount of payroll and the number of full-time equivalent employees. Therefore, the ratio of these two variables gives us a reasonable approximate for the price of labor: $w_l = Payroll/L$. The database includes expenditures for fuel (*F*), maintenance and repair (*R&M*), and materials and supplies (*M&S*), and no quantities. Prices for these variables can be obtained from the INSEE and it is therefore possible construct a complete database. However, the prices would be the same for all firms in every years in ours sample. This causes some problems for the estimation procedure (in fact the optimization algorithm failed to find a solution most of the time), so we have decided to aggregate these three components of the cost and construct a firm specific price as a weighted sum of the price indexes, using cost shares as weights. This procedure gives us a form specific materials price which is used in the production process for a company:

$$w_M = w_F S_F + w_{R\&M} S_{R\&M} + w_{M\&S} S_{M\&S} \tag{4.1}$$

where $S_i = E_i / \left(E_F + E_{R\&M} + E_{M\&S}\right), i = F, R\&M, M\&S$, is the cost share of variable input i in the sum of "other" variable cost. The quasi-fixed inputs are the stock of capital and the infrastructure used by the companies in the production process. The capital stock of a bus transportation company is mainly the buses they use in production (although it may include some buildings and parking spaces for buses). We use for capital the transportation potential. That is, capital must be related to the number of buses and their number of seats. One way of doing this, it to compute the number of seats available. To do so, we use a weighted sum the number of buses using the number of seat as weight to reflect their size. This represents the total number of vehicle seats for each company. This new variable is then used as an approximation of capital in our cost function. Since we focus on road infrastructure, variables on the road network characteristics are required. To capture the service provided by road infrastructures, we have decided to use an accessibility indicator in our analysis. The computation of this index is discussed in the next section. Finally, the variable cost is given by the sum of the payroll, the fuel expenditures, the repair and maintenance expenditures and the material and supplies expenditures.

The last component of the cost function is the output. There are many desirable alternatives, like the number of passengers per kilometer or the total passenger kilometers. Unfortunately, our database remains silent on these options, so we have used the total number of kilometers the vehicles have travelled, loaded or not. This has the virtue of capturing the management of the routes and services supplied to customers, in particular on charter routes.

4.2 Infrastructure as an accessibility index

For a bus transportation company, it is not so much the infrastructure availability that is required as much as its accessibility. A congested road network of a given size does not offer the same service as a non-congested network of the same size. Consequently, an infrastructure input must be able to reflect this accessibility to the network. The literature is not shy on indexes that can the benefits to companies of transport infrastructure at the regional level. One of the most popular indicators in transportation economics is the market potential, which is a gravity-based measure. It has been extensively used (see, for example, Keeble *et al*. (1988), Bruinsma and Rietveld (1993), Schürman *et al*. (1997), López *et al*. (2008)) in the transportation literature, but never in the productivity analysis. The market potential index relates positively the economic benefit of the origin node, denoted *i*, to the 'mass' of destination node *j* and is inversely proportional to the distance or travel time between these nodes. That is:

$$A_i = \sum_j Op_j f\left(d_{ij}\right) \tag{4.2}$$

where A_i is the economic benefit (services provided by the road infrastructure in our case) for the companies in the zone *i*, Op_j represents the opportunities offered by destination *j* (quantity and quality of road infrastructures) and *f* is a decreasing resistance function of the distance between origin *i* and destination *j*.

Our purpose here is to build an infrastructure index that represents the service generated by road infrastructure relative to a point of origin. For our index we use employment zone, this is the smallest partition of France with all the required information to construct the index. This index should be the combination of the amount of infrastructure in the reference zone and the access to potential interconnections between different zones. Consequently, we should take into consideration some important elements. Firstly, although it should include the quantity of infrastructure such as road length, the size of the zone should also be considered: the larger the zone is, the less infrastructure service the firm can get from a given network size. Secondly, when we measure the potential accessibility the distance between the origin and the destination provides the service but it is counter-balanced by congestion time. In other words, for a given distance, the time spent on the road gives the quantity of service one can get from the infrastructure. If we put together these two components, we get an indicator that takes into account the characteristics of the infrastructure in the origin, R_{i1}, and a resistance function between the origin *i* and destination *j* including time spent traveling, and quantity of infrastructure in destination R_{i2}. Our indicator is measured as

$$Inf_i = R_{i1}R_{i2} = \frac{Len_i Sp_i}{Ar_i} \sum_j Len_j \exp\{-\nu d_{ij} - \gamma t_{ij}^{con}\} \tag{4.3}$$

with $R_{i1} = Len_i Sp_i / Ar_i$ and $R_{i2} = \sum_j Len_j \exp\{-\nu d_{ij} - \gamma t_{ij}^{con}\}$, where, L_i is the total road length in *i*, Sp_i represents the average road speed, Ar_i is the total land area of zone *i*, d_{ij} is the distance between *i* and *j*, t_{ij}^{con} is the average time lost due to congestion, and ν and γ are parameters to be estimated.

To estimate the parameters ν and γ, we use a double-constrained gravity model. In this type of model, the only variables used are the distance and congestion time between two zones.

The flows of goods in and out of the origins are assumed equal and constitute the two constraints in the model. The model formulation is as follows:

$$F_{ij} = a_i \cdot O_i \cdot b_j \cdot D_j \cdot C_{ij} \tag{4.4}$$

where F_{ij} is the flow of goods between i and j, O_i is the total observed goods out of zone i, $O_i = \sum_j F_{ij}$; D_j is the total in flow of goods at destination j, $D_j = \sum_i F_{ij}$. $a_i = 1/\sum_i b_j D_j C_{ij}$ and $b_j = 1/\sum_j a_i O_i C_{ij}$ are the adjustment coefficients associated to O_i and D_j, respectively, and finally $C_{ij} = d_{ij}^{-\nu} \left(t_{ij}^{con}\right)^{-\gamma}$ denotes the resistance function that depends on the distance and congestion time. As it is standard in the literatures (Flowerdew and Aitkin (1982)**,** D'Aubigny *et al.*, (2000), Griffith and Fischer (2013), the parameters are estimated using a Poisson regression. The advantage of the Poisson regression in our case is to duly weight small flows whose role is often exaggerated by the methods of log-linear adjustment. The most important feature of the Poisson regression is that it ensures that the in-flow and out-flow constraints are met for all zones.

Table 4.1 presents the exact description of the variables used for the infrastructure index, while Table 4.2 presents the descriptive statistics for the entire database.

[INSERT TABLE 4.1 HERE]

[INSERT TABLE 4.2 HERE]

5. Results

In this section we present the results of an application of our A-C cost frontier estimation procedure to the bus passenger transportation in France. The biggest advantage of this procedure over a nonparametric method resides in the fact that we are using a parametric frontier, allowing us to deduce and compute easily the technology characteristics. Consequently, we start by defining the quantities we will put our focus on and then we present the results. Nonetheless, our main attention is on returns to scale, as it allows us to look at potential improvement of the industrial structure.

The cost frontier is assumed to be a translog functional form.

$$\ln C = \varphi_o + \varphi_Y \ln y + \frac{1}{2}\gamma_{YY}(\ln y)^2 + \sum_{p=1}^{P} \varphi_{w_p} \ln w_p + \sum_{s=1}^{S} \varphi_{k_s} \ln k_s + \frac{1}{2}\sum_{p=1}^{P}\sum_{p'=1}^{P} \gamma_{pp'} \ln w_p \ln w_{p'}$$
$$+\frac{1}{2}\sum_{q=1}^{Q}\sum_{q'=1}^{Q} \gamma_{qq'} \ln k_q \ln k_{q'} + \sum_{p=1}^{P}\sum_{q=1}^{Q} \gamma_{pq} \ln w_p \ln k_q + \sum_{p=1}^{P} \gamma_{py} \ln w_p \ln y + \sum_{q=1}^{Q} \gamma_{qy} \ln k_q \ln y \qquad (5.1)$$

where φ_0 is the constant, $\boldsymbol{w}_p = \left(w_L, w_M\right)$ are the variable input prices (Labor and Materials) $\boldsymbol{k}_s = (V, Inf)$ are the quasi-fixed inputs (vehicles and infrastructures) and the parameters satisfy the symmetry requirement: $\gamma_{pp'} = \gamma_{p'p}$, $\gamma_{qq'} = \gamma_{q'q}$ and $\gamma_{pq} = \gamma_{qp}$. The estimates of the parameters are obtained from the solution of the optimization problems given by equation (2.16) and (2.17) by replacing the cost function by Equation (5.1). We also impose some other regularity conditions, such as the price homogeneity:

$$\sum_{p=1}^{P} \varphi_p = 1; \sum_{p=1}^{P} \varphi_{pi} = 0, i = 1,..., P; \sum_{p=1}^{P} \varphi_{py} = \sum_{p=1}^{P} \varphi_{pq} = 0, q = 1, \ldots, Q \qquad (5.2)$$

The variable input shares and the cost elasticity of output must be positive:

$$S_p = \frac{\partial \ln C}{\partial \ln w_p} = \varphi_p + \sum_{p'}^{P} \gamma_{pp'} \ln w_{p'} + \sum_{q=1}^{Q} \gamma_{pq} \ln k_q + \gamma_{py} \ln y > 0. \qquad (5.3)$$

$$\xi_{Cy} = \frac{\partial \ln C}{\partial \ln y} = \varphi_y + \sum_{p=1}^{P} \gamma_{py} \ln w_p + \gamma_{yy} \ln y + \sum_{q=1}^{Q} \gamma_{qy} \ln k_q > 0. \qquad (5.4)$$

The cost elasticity of the quasi-fixed factors must be negative, that is:

$$\xi_{Ck_q} = \frac{\partial \ln C}{\partial \ln k_q} = \varphi_{k_q} + \sum_{p=1}^{P} \gamma_{pq} \ln w_p + \sum_{q=1}^{Q} \gamma_{qq'} \ln k_{q'} + \gamma_{qy} \ln y < 0. \qquad (5.5)$$

5.1 The technology measures.

The cost frontier allows us to compute for each firm its efficient cost, and obviously it is the observed cost projection on the frontier. Consequently we have $\phi_i = C_i^{obs} / \hat{C}_i$ where $\hat{C}_i$ is the projected cost on the frontier given the prices, the quasi-fixed inputs (including infrastructure) and the output. That is, $\hat{C}_i$ is the efficient cost. The aggregate cost inefficiency is given by:

$$\phi = \sum_{i=1}^{N} \phi_i \times \left(C_i^{obs} \Big/ \sum_{j=1}^{n} C_j^{obs} \right) \text{ where } \phi_i = \frac{C_i^{obs}}{\hat{C}_i}, \qquad (5.6)$$

where ϕ is the average efficiency for the industry. Using (5.3) we can easily deduce that the variable input share elasticity of capital is given by:

$$\tau_{S_p k_s} = \frac{\partial \ln S_p}{\partial \ln k_s} = \frac{\gamma_{ps}}{S_p}, \forall (p,s). \tag{5.7}$$

The quasi-fixed inputs introduce a positive bias for input p when $\tau_{S_p k_s} > 0$, negative when $\tau_{S_p k_s} < 0$ and neutral when $\tau_{S_p k_s} = 0$. This is of particular interest to understand the effect of the infrastructure. The effect of a quasi-fixed factor change on output is given by:

$$\xi_{yk_q} = \frac{\partial \ln y}{\partial \ln k_q} = -\frac{\xi_{Ck_q}}{\xi_{Cy}}, \forall q. \tag{5.8}$$

The impact of a change in vehicles of infrastructure on the (conditional) demand of variable input p is given by:[3]

$$\xi_{pq} = \frac{\partial \ln x_p}{\partial \ln k_q} = \frac{\partial \ln S_p}{\partial \ln k_q} + \frac{\partial \ln C}{\partial \ln k_q} = \frac{\gamma_{pq}}{S_p} + \xi_{Ck_q} = \tau_{S_p k_q} + \xi_{Ck_q}, \ \forall p, \tag{5.9}$$

Short-run and long-run returns to scale are respectively given by:

$$\eta_{ct} = \left(\frac{\partial \ln C}{\partial \ln y} \right)^{-1} \tag{5.10}$$

$$\eta_{lt} = \left(\frac{\partial \ln C}{\partial \ln y} \right)^{-1} \left(1 - \sum_{s=1}^{S} \xi_{Ck_s} \right) = \eta_{ct} \cdot \left(1 - \sum_{s=1}^{S} \xi_{Ck_s} \right). \tag{5.11}$$

5.2 Estimation results

The estimated parameters are presented in Table 5.1. The table includes the estimated value, the bias, the corrected estimates and finally confidence intervals of size 90%, 95% and 99%. The confidence intervals contain all the information necessary to determine when a parameter is significant. Note that the procedure does not return standard errors, but they are not necessary since we have the full distribution of the estimates (that can even be skewed). All input parameters are significant at 1%, including capital and infrastructure. As mentioned above, some regularity conditions have been imposed on the estimation, so the cost function is homogenous of degree one in the input prices, increasing in the output and the input prices and decreasing in the quasi-fixed inputs. The two methods (the linear and quadratic minimization programs) return parameter estimates that are fairly similar, so the optimization

[3] Nadiri and Mamouneas (1994) first used this formula to measure this impact.

program is not consequential on the results, a desirable property. It is worth noting that the functional form plays a crucial role (we have estimated Cobb-Douglas as well as the Translog cost functions and the difference is striking).

Table 5.1 presents the cost elasticities. That is, the two variable input shares, the cost output elasticity, the quasi-fixed input elasticities of cost and the infrastructure elasticity of output as defined by Equation (5.3)-(5.5) and (5.8). The numbers are average over the sample of firms. The optimization criterion is again a non-factor as the results a virtually identical for all elasticities. Labor cost dominates the cost of the firms. From the estimate of ξ_{Cy} we have that a 1% increase of the output translates into a cost increase between 1.19% and 1.29%. This shows that the short-run constraint in the adjustment of the quasi-fixed factors plays an important role. From ξ_{CV} and ξ_{CInf} we see that increasing the number of bus reduces the cost by 0.33% while the impact of an increase of 1% of the infrastructure is of an order of 0.54%. The last column of Table 5.2 gives the effect of an increase of 1% of the infrastructure on the output. The effect is positive and non negligible (a magnitude of slightly more than 0.4%). This conforms the role played by the infrastructure in production process of the firm.

Table 5.2: Cost elasticities.

	S_L	S_K	ξ_{Cy}	ξ_{CV}	ξ_{CInf}	ξ_{yInf}
Linear program	0,777	0,223	1,188	-0,330	-0,552	0,465
Quadratic program	0,795	0,205	1,288	-0,329	-0,537	0,417

As suggested above, the infrastructure play a non-negligible role, suggesting that it might be interesting to investigate it more deeply. Using Equation (5.8) and (5.9) we can investigate the impact of increasing the public infrastructure on the variable inputs and quasi-fixed input (the vehicles). We say the infrastructure are substitutable, independent and complementary to the variable x_p if ξ_{p_Inf} is respectively smaller, equal or larger than zero. Table 5.3 presents these calculations. The results show that materials are complementary to the infrastructure while labor is substitutable. The result is not that clear cut for the vehicles. Under the linear program it exhibits substitutability while under the quadratic program it is a complementarity, so the optimization criterion does play a role here. This might be explained by the fact that this relationship depends on second order terms that are estimated with a weaker precision. If we conclude that the effect of infrastructure is neutral on vehicles (the values under each program are of the same magnitude but of opposite sign) it is possible to conclude that better infrastructure reduce congestion time and thus requires less labor, but because it has a positive effect on output, more gas and maintenance is required for the given stock of vehicles.

Table 5.3: Infrastructure external input elasticities

	τ_{L_Inf}	τ_{M_Inf}	ξ_{L_Inf}	ξ_{M_Inf}	ξ_{V_Inf}
Linear program	-0.647	2.256	-1.199	1.704	0.406
Quadratic program	-0.313	1.215	-0.850	0.678	-0.400

Returns to scale are computed using Equations (5.10) and (5.11). Table 5.4 presents the estimated results while the distribution of the returns to scale are displayed on Figure 5.1 and 5.2. Among the main features of the returns to scale, we note that the short run returns to scale are dominant as only 18.3% (17.4%) of the firms exhibit increasing returns to scale under the linear program (quadratic program). So, in the short run, most of the firms do operate under decreasing returns to scale. Note that the third quartile is equal to 0.953 under the linear program and 0.899 under the quadratic program. Finally, the median is significantly different than one at 95% when using the linear program, however this is not true with the quadratic program. In that latter case, it seems that firms operate for a large number of them under an optimal scale (the returns to scale are not significantly different than one).

The story is totally different when one considers the long run returns to scale. That is, when vehicles and infrastructure are included in the calculation, almost all firms operate under increasing returns to scale (95.5 % for the linear program and 99.6 % in the quadratic case). Even the first quartile is significantly different than one (and clearly, greater than one), confirming that most firms operate under increasing returns to scale.

To summarize our analysis, when we take into account the presence of the infrastructure in the production process of the French bus transportation firms, the industry is largely characterized by increasing returns to scale, for all criteria for the optimization program and all test size. It means that it is more costly to supply the same quantity using two firms than using one. That is, one firm would produce at a lower cost the same quantity produced by two firms. It is then desirable or inevitable that the concentration must increase in this industry so that economies of scale can be exploited. Therefore, the industry would benefit from a restructuring by reducing the number of firms.

To conclude this section, let us look at efficiency. By construction, the efficiency scores are larger than one because we envelope the data from below giving us a lower envelope of the data. Consequently, all cost inefficient observations are located above the frontier. The cost inefficiency scores and the average cost efficiency are defined by Equation (5.6). The efficiency analysis is summarized in Table 5.5. The main result is that, no matter the optimization criterion, firms can lower their cost on average by 11.9% while keeping output constant. This confirms the relative efficiency of this industry. It is also interesting to note that the performance of the firms changes with their size. Our results show that larger firms tend to be more efficient than smaller one. While the largest firms are virtually efficient, the smallest are more than 20% less efficient, confirming that size matters in this industry. To illustrate that fate, we have formed groups of 127 firms (representing 5% of the sample) and drawn the relationship between their size (measured by the turnover) and the efficiency score. These are presented on Figure 5.3 and 5.4. The negative slope is clear, as firms increase in size, they become more efficient.

Table 5.5: Cost efficiency

		Translog	
		Linear Program	Quadratic Program
	$\boldsymbol{\theta}$	1.122	1.120
	$\dot{\boldsymbol{N}}$	49	
$\boldsymbol{\theta}$	Smallest 5 %	1.236	1.233
	Largest 5 %	1.056	1.060

Note: θ is the weighted average efficiency score, $\dot{N}$ is the absolute value of the difference between the efficiency ranking of the firms based on the criteria used to solve for the efficiency.

6. Conclusion

In this paper we have used a statistical Aigner and Chu model applied to cost data to estimate a cost frontier. We have included quasi-fixed factors to account for the fact that capital (buses) and infrastructures are either slow to adjust or simply not under the control of the firm. This allowed us to recover the main characteristic of the technology.

We have applied this methodology to inter-city bus transportation firms in France. We where able to account for the infrastructure and build an index that captured firm specific access to the road network, accounting average speed, congestion and the likes.

Our results show that firms are mostly cost-efficient, but operate under increasing returns to scale. This shows that the industry is not at maturity and some restructuring has to take place in order to produce at a lower cost on average. The nature of this inefficiency is not directly obvious but it seems that a large number of firms benefits from public contracts and those are regulated under somewhat specific conditions that prevent firms from exploiting all their opportunities. This still has to be explored, but there is clearly a link to be explored between the State intervention in the industry and this inefficient operation scale.

References

Afriat, S.N. (1972). "Efficiency Estimation of Production Functions". *International Economic Review*, 13(3), 568-98.

Aigner, D.J. and Chu, S. F. (1968). "On Estimating the Industry Production Function". *The American Economic Review*, 58(4), 826-839.

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

Albriktsen, R. and Førsund, F.R. (1990), "A Productivity Study of the Norwegian Building Industry". *Journal of Productivity Analysis*, 2(1), 53-66.

Andrews, D.W.K. (2000). "Inconsistency of the Bootstrap When a Parameter Is on the Boundary of the Parameter Space". *Econometrica*, 68, 399-405.

Aschauer, D.A. (1989). "Is public expenditure productive?". *Journal of Monetary Economics*, 23, 177-200.

Battese, G.E and Corra, G.S. (1977). "Estimation of a Production Frontier Model with Application to the Pastoral Zone of Eastern Australia". *Australian Journal of Agricultural Economics*, 21, 169-179.

Bickel, P., Götze, F. and Van Zwet, W. (1997). "Resampling Fewer than *n* Observations:

Bruinsma, F.R., Rietveld, P. (1998). "The accessibility of European cities: theoretical framework and comparison approaches". *Environment and Planning*, 30, 449– 521.

Griffith, D., and Fischer, M. (2013). "Constrained variants of the gravity model and spatial dependence: model specification and estimation issues". J*ournal of Geographical Systems*, Springer, vol. 15(3), 291-317.

D'Aubigny, G., Calzada, C., Grasland, C. and Robert, D., (2000). "Approche poissonnienne des modèles d'interaction spatiale". Cybergéo, 126

Efron, B. and Tibshirani, R. (1993). "*An Introduction to Bootstrap*". New York, Chapman & Hall.

Farrell, M.J. (1957). "The Measurement of Productive Efficiency". *Journal of Royal Statistical Society, Series A, General*, 120, 253-81.

Flowerdew, R., And Aitkin, M. (1982). ″A method of fitting the gravity model based on the Poisson distribution″. *Journal of Regional Science*, 22(2), 191-202.

Førsund, F.R. (1992). "A Comparison of Parametric and Nonparametric Efficiency Measures: The Case of Norwegian Ferries ", *Journal of Productivity Analysis*, 3, 25-44.

Førsund, F.R. and Hjalmarsson, L. (1979). "Frontier Production Functions and Technical Progress: A Study of General Milk Processing in Swedish Dairy Plants", *Econometrica, Econometric Society*, 47(4), 883-900.

Førsund, F.R. and Jansen, E.S. (1977). "On Estimating Average and Best Practice Homothetic Production Functions via Cost Functions". *International Economic Review,* 18(2), 463-76.

Greene, W. H. (1980a). "Maximum likelihood estimation of econometric frontier functions". *Journal of Econometrics* , 13(1), 27-56.

Greene, W. H. (1980b). " On the estimation of a flexible frontier production model ". *Journal of Econometrics* , 13(1), 101-115

Greene, W. H. (2008). "The Econometric Approach to Efficiency Analysis". Chapter 2 in Fried, H.O., Lovell, C.A.K et Schmidt, S. (Eds.) "*The Measurement of Productive Efficiency: Technics and Applications*". New York, Oxford University Press.

Hall, P. (1992a). "*The Bootstrap and Edgeworth Expansion*". New York: Springer-Verlag.

Hirschhausen, C. and Cullmann, A. (2008). "Next Stop: Restructuring?: A Nonparametric Efficiency Analysis of German Public Transport Companies". *Transportation Research Part E*: *Logistics and Transportation Review,* 46(3), 436-445.

Keeble, D., Offord, J., Walker, S. (1988.) "*Peripheral regions in a community of twelve*". Office for Official Publications of the European Communities, Brussels/ Luxembourg.

López, E., Gutiérrez, J., Gómez, G., (2008). "Measuring regional cohesion effects of large-scale transport infrastructure investment: an accessibility approach". *European Planning Studies*, 16 (2), 277–301.

Malmquist, S. (1953). "Index Numbers and Indifference Curves". *Trabajos de Estadistica*, 4, 209-242.

Meeusen,W. and Van Den Broeck. (1977). "Efficiency Estimation from Cobb-Douglass Production Functions with Composed Errors". *International Economic Review*, 18, 435-444.

Nishimizu, M. and Page, J. M. (1982). "Total Factor Productivity Growth, Technological Progress and Technical Efficiency Change: Dimensions of productivity Change in Yugoslavia, 1965-1978". *Economic Journal*, 92(368), 920-936.

Odeck, J. and Alkadi, A. (2001). "Evaluating Efficiency in the Norwegian Bus Industry using Data Envelopment Analysis". *Transportation*, 28, 211-32

Ouellette, P. and Petit, P. (2010) "*Mesure de l'Efficience des Établissements de Santé : Revue et Synthèse Méthodologique*" Centre sur la productivité et la prospérité, HEC Montréal.

Ouellette, P., Vigeant, S., Zhang, L. (2016). "*Semi-parametric Estimation of Efficiency and Technology, with an Application to French Bus Transportation*." Mimeo, Université de Lille: Sciences et Technologies.

Politis, D. N. and Romano, J. P. (1992). "A Circular Block-Resampling Procedure for Stationary Data". In LePage, R., Billard, L. (Eds.) "*Exploring the Limits of Bootstrap*". New York: John Wiley, 263-270.

Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling*. New York: Springer.

Porter, M.E. (1979). "The Structure within Industries and Companies' Performance". *The Review of Economics and Statistics*, 61, 214-27.

Richmond, J. (1974). "Estimating the Efficiency of Production". *International Economic Review*, 15, 515-521.

Royden, H. L. (1988). "*Real Analysis. 3rd Ed* ". Prentice Hall, Englewood Cliffs, New Jersey, Pages, 444.

Schmidt, P. (1975). "On the Statistical Estimation of Parametric Frontier Production Functions", *Review of Economics and Statistics*, 58, 238-239.

Schürman, C., Spiekermann, K., Wegener, M., (1997). "*Accessibility indicators*". Berichte aus dem Institut für Raumplanung, 39, IRPUD, Dortmund.

Simar, L. and Wilson, P.W. (1998). "A General Methodology for Bootstrapping in Nonparametric Frontier Models". *Papers* 9811, Catholique de Louvain - Institut de Statistique.

Simar, L. and Wilson, P.W. (1999). "Some Problems with the Ferrier/Hirschberg Bootstrap Idea". *Journal of productivity Analysis*, 11, 67-80.

Simar, L. and Wilson, P.W. (2011). "Inference by the m out of n bootstrap in nonparametric frontier models", *Journal of Productivity Analysis*, 36, 33–53.

Wald, A. (1949). "Note on the consistency of the maximum likelihood estimate". *The Annals of Mathematical Statistics*, 20, 595–601.

Figures.

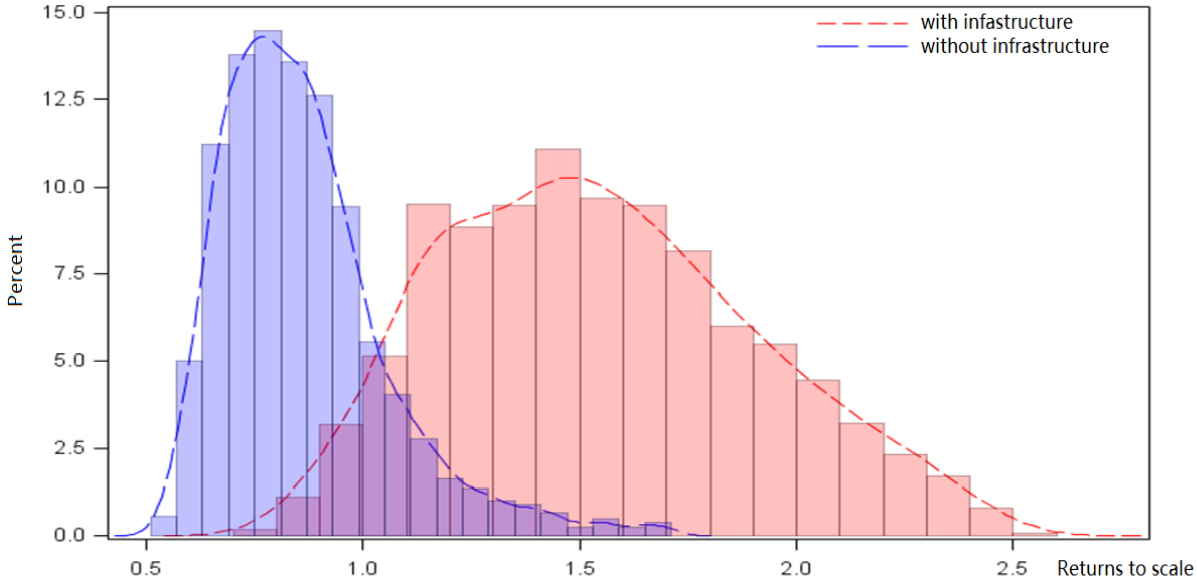

Figure 5.1 Returns to Scale with and without infrastructure, Linear Program

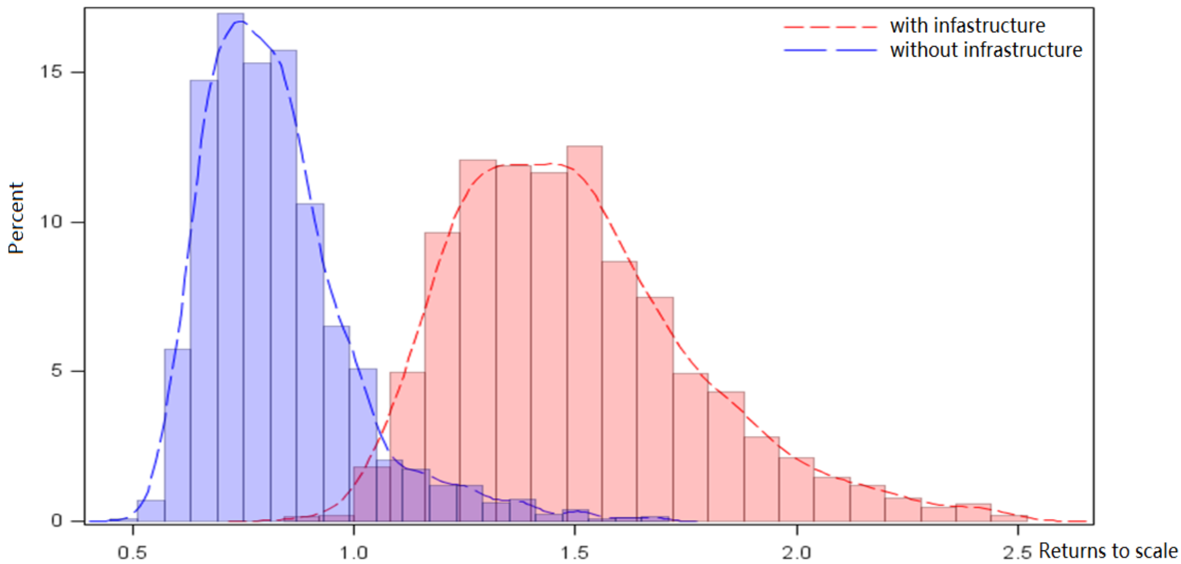

Figure 5.2 Returns to Scale with and without infrastructure, Quadratic Program

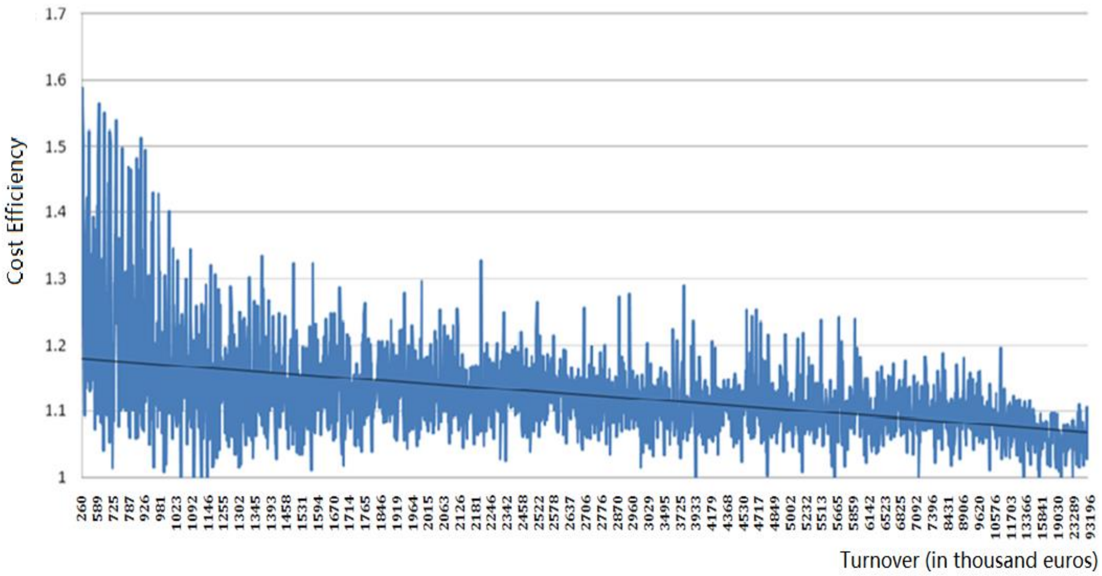

Figure 5.3 Relation between cost efficiency and turnover, Linear Program

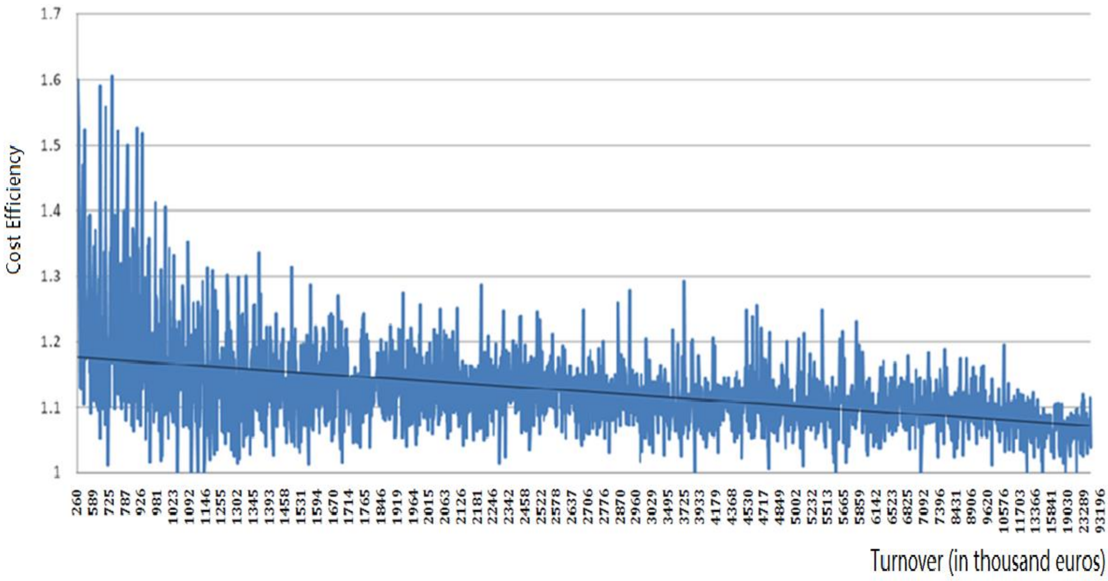

Figure 5.4 Relation between cost efficiency and turnover, Quadratic Program

Tables.

Table 4.1: Description of the variables

Variable	Symbole	Données	Description	Sources
Average speed	Sp_i	$Sp_i = N^{-1} \sum_{r=1}^{N} Sp_{ir}$	Average speed in employment zone *i*	Computed
Total highway lengths	Len_i	$Len_i = \sum_{r=1}^{N} Len_{ir}$	Total road and highway length in employment zone *i*	CGDD
Area	Ar_i	*SUP*	Area of employment zone *i*	INSEE
Distance	d_{ij}	Mapinfo network map	Distance the centroid of employment zones *i* and *j*	Computed
Congestion time	t_{con}	$TEMPS_j$	Time lost due to congestion on road *r*	CGDD
Goods flow	F_{ij}	F_{ij}	Annual good flow in 2002 between zone *i* and *j*	CGDD
Infrastructure index	Inf_i	$\left(\frac{Len_i Sp_i}{Ar_i}\right) * \sum_j Len_j\, exp\{-\nu d_{ij} - \gamma t_{conij}\}$	Accessibility index for employment zone *i*	Computed

Table 4.2 : Descriptive statistics

	Variable	Symbol	Maximum	Minimum	Standard error	Mean	Median
Output and cost	Kilometers	Y	45 040 389	10 886	3 060 574,48	2 097 411,79	1 183 424,50
	Variable cost	D_{COUT}	144 424	16,86	7 531,98	3 349,87	1594

Variable Input	Labor	L	1 957	2	121,37	67,89	36
		w_L	80,86	0,50	7,42	26,19	26,13
		E_L	94433	1	4 444,43	1 931,42	916,50
	Fuel	F	8 968,19	1,05	687,30	451,83	248,32
		w_E	102,02	92,34	3,71	96,87	94,92
		E_F	9149	1	667,07	437,87	240
	Repair and maintenance	$R\&M$	3 997,19	1,78	251,84	141,56	69,47
		$W_{R\&M}$	112,28	100	4,72	105,10	103,77
		$E_{R\&M}$	4 488	2	267,11	149,16	73
	Material and supplies	$M\&S$	47 847,68	2,82	2 420,73	802,80	314,67
		$W_{M\&S}$	106,49	100	2,37	103,46	104
		$E_{M\&S}$	4 8635	3	2 501,16	831,42	323
Quasi-fixed Inputs	Number of city bus	K_B	716	0	31,52	7,74	0
	Number of bus	K_{car_p}	43 020	0	3 341,86	2 393,96	1 395
		K_{C1}	65	0	4,31	1,13	0
		K_{C2}	127	0	6,20	2,05	0
		K_{C3}	127	0	9,65	2,99	0
		K_{C4}	682	0	37,74	14,65	0
		K_{C5}	171	0	6,73	1,34	0
	Number of seats	KV	67 270	5	4 088,05	2 742,05	1 575
Infrastructures	Infrastructure	Inf_i	15,31	7,81	1,01	12,82	12,85
	Congestion time	t_{con}	120,43	5,85	11,40	20,02	17,42
	Average speed	Sp_i	116,41	37,74	12,54	69,01	68,67

	Length	Len_i	856,20	11,37	152,27	285,14	255,71
	Distance	d_{ij}	1 044,28	8,61	191,81	395,60	386,20
	Area of zone *i*	$Area_i$	6 256,05	46,36	998,52	1 591,49	1 448,19
	Flow of goods	F_{ij}	4 280 771,20	1	29019,03	2318,28	79,50

Table 5.1: Parameter estimates and statistical inference – linear and quadratic program

	Estimates	Bias	Corrected estimates	90% Interval		95% Interval		99% Interval	
				Inf	Sup	Inf	Sup	Inf	Sup
				Linear Program					
φ_{W_L}	0,607	-0,264	0,871	0,483	0,962	0,472	0,980	0,433	1,000
φ_{W_K}	0,393	-0,036	0,429	0,038	0,517	0,020	0,528	0,000	0,567
φ_V	1,732	-0,300	2,032	1,482	2,315	1,463	2,463	1,305	2,582
φ_{Inf}	-1,325	-0,456	-0,869	-1,5192	-0,6288	-1,649	-0,574	-1,762	-0,310
φ_y	-2,217	-0,277	-1,940	-2,4472	-1,819	-2,684	-1,516	-2,910	-1,302
γ_{WL^2}	-0,261	-0,076	-0,185	-0,532	-0,236	-0,665	-0,163	-0,825	-0,032
γ_{WK^2}	-0,261	-0,076	-0,185	-0,532	-0,236	-0,665	-0,163	-0,825	-0,032
γ_{V^2}	0,079	-0,029	0,108	0,020	0,095	0,016	0,158	0,012	0,221
γ_{INF^2}	1,203	-0,192	1,395	1,142	1,518	1,082	1,696	0,811	1,959
γ_{WL_WK}	0,261	0,076	0,185	0,163	0,532	0,136	0,665	0,032	0,825
γ_{WL_V}	-0,002	0,002	-0,004	-0,011	0,005	-0,018	0,008	-0,023	0,019
γ_{WL_Inf}	-0,503	0,171	-0,674	-0,731	-0,475	-0,974	-0,339	-1,208	-0,203
γ_{WK_V}	0,002	-0,002	0,004	-0,05	0,011	-0,08	0,014	-0,19	0,003
γ_{WK_Inf}	0,503	-0,171	0,674	0,475	0,731	0,339	0,974	0,203	1,208

γ_{V_Inf}	-0,316	0,016	-0,331	-0,505	-0,157	-0,631	-0,131	-0,782	-0,079
γ_{y^2}	0,205	-0,194	0,399	0,137	0,281	0,125	0,302	0,119	0,373
γ_{WL_y}	0,094	0,000	0,094	0,009	0,217	0,008	0,260	0,007	0,304
γ_{V_y}	-0,083	-0,007	-0,076	-0,125	-0,064	-0,166	-0,046	-0,206	-0,028
γ_{Inf_Y}	-0,134	0,131	-0,265	-0,178	-0,075	-0,182	-0,063	-0,221	-0,038
Quadratic program									
φ_{WL}	0,453	-0,309	0,762	0,147	0,692	0,113	0,961	0,089	1,00
φ_{WK}	0,547	-0,112	0,659	0,308	0,853	0,039	0,887	0	0,911
φ_{V}	1,257	-0,415	1,672	1,088	1,440	1,014	1,839	0,855	2,012
φ_{INF}	-1,625	-0,156	-1,468	-1,714	-1,530	-1,822	-1,166	-1,969	-0,913
φ_{y}	-1,750	-0,418	-1,332	-1,867	-1,196	-1,884	-0,978	-1,955	-0,636
γ_{WL^2}	0,119	-0,185	0,304	0,077	0,400	0,059	0,511	0,052	0,631
γ_{WK^2}	0,119	-0,185	0,304	0,077	0,400	0,059	0,511	0,052	0,631
γ_{V^2}	0,082	0,094	-0,012	-0,063	0,103	-0,074	0,142	-0,096	0,195
γ_{Inf^2}	0,939	-0,228	1,168	0,819	0,989	0,756	1,173	0,632	1,288
γ_{WL_WK}	-0,119	0,185	-0,304	-0,400	-0,077	-0,511	-0,059	-0,631	-0,052
γ_{WL_V}	0,008	0,043	-0,035	-0,181	0,163	-0,212	0,207	-0,301	0,239
γ_{WL_Inf}	-0,249	-0,040	-0,209	-0,736	-0,146	-0,987	-0,101	-1,402	-0,079
γ_{WK_V}	-0,008	-0,043	0,035	-0,163	0,181	-0,207	0,212	-0,239	0,301

γ_{WK_Inf}	0,249	0,040	0,209	0,146	0,736	0,101	0,987	0,079	1,402
γ_{V_Inf}	-0,045	-0,012	-0,033	-0,111	-0,007	-0,130	-0,006	-0,163	-0,004
γ_{y^2}	0,208	0,100	0,108	0,046	0,236	0,037	0,270	0,030	0,330
γ_{WL_y}	0,039	0,042	-0,003	0,026	0,145	0,015	0,208	0,009	0,296
γ_{V_y}	-0,103	0,083	-0,111	-0,233	-0,097	-0,253	-0,089	-0,288	-0,057
γ_{Inf_y}	-0,063	0,043	-0,106	-0,166	-0,058	-0,233	-0,046	-0,306	-0,028

Tableau 5.4: Returns to scale

		Estimate	Bias	Corrected estimates	90% Interval		95% Interval		99% Interval	
					Inf	Sup	Inf	Sup	Inf	Sup
Linear program $h^E_{without}$	Median	0,833	0,012	0,821	0,760	0,900	0,747	0,912	0,713	1,093
	Min	0,527	0,054	0,473	0,331	0,607	0,303	0,634	0,287	0,653
	Max	1,700	0,047	1,653	1,718	1,789	1,550	1,808	1,474	1,866
	P25	0,728	0,075	0,653	0,583	0,789	0,578	0,826	0,559	0,857
	P75	0,953	-0,11	1,063	0,900	1,071	0,893	1,128	0,863	1,176
Linear program h^E_{with}	Median	1,514	-0,021	1,535	1,491	1,579	1,463	1,641	1,336	1,681
	Min	0,765	0,031	0,734	0,671	0,929	0,660	0,971	0,569	1,015
	Max	2,57	0,134	2,436	2,469	2,592	2,258	2,631	2,142	2,712
	P25	1,262	0,145	1,117	1,055	1,297	1,036	1,323	0,989	1,351
	P75	1,789	-0,209	1,998	1,704	1,793	1,690	1,868	1,634	1,947
Quadratic program $h^E_{without}$	Median	0,796	0,016	0,780	0,727	0,920	0,677	1,096	0,642	1,154
	Min	0,485	0,044	0,441	0,426	0,838	0,403	0,872	0,374	0,924
	Max	1,692	0,047	1,645	1,461	1,750	1,439	1,955	1,407	1,967
	P25	0,706	0,082	0,624	0,610	0,730	0,600	0,816	0,587	0,822
	P75	0,899	-0,152	1,051	0,776	0,930	0,704	1,139	0,677	1,185
Quadratic program h^E_{with}	Median	1,465	-0,018	1,483	1,267	1,548	1,257	1,631	1,215	1,701
	Min	0,869	0,025	0,844	0,600	1,508	0,331	1,542	0,107	1,676
	Max	2,501	0,11	2,391	2,159	2,587	2,127	2,890	2,079	2,908
	P25	1,294	0,165	1,129	1,117	1,339	1,100	1,495	1,076	1,505
	P75	1,655	-0,200	1,855	1,429	1,712	1,407	1,912	1,376	1,924