Document de travail du LEM / Discussion paper LEM
2019-09

Semi-parametric Estimation of Efficiency and Technology, with an Application to French Bus Transportation

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/

Semi-parametric Estimation of Efficiency and Technology, with an Application to French Bus Transportation

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-lille1.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

April 2019

Abstract: In this paper, we introduce a statistical semi-parametric method to estimate production frontier based on Aigner and Chu frontier estimation method. This procedure has been shown to produce maximum likelihood estimators of the parameters. However, no other statistical properties have been derived. Using Wald's consistency theorem for maximum likelihood estimators, we show that the estimator is consistent by verifying the required conditions. Inference is based on re-sampling methods. Because the parameters to estimate define the frontier, standard bootstrap procedures were not available. Sub-sampling offers a valuable alternative that works in this case. We have used this procedure to estimate the efficiency of the bus transportation industry in France in the presence of infrastructure. The procedure is shown to be robust.

Keywords: Efficiency, Semi-parametric methods, Bootstrap, Production technology, Bus transportation.

JEL Codes: C51, C61, D24, L25

1. Introduction

The measurement of efficiency of a group of decision units roughly consists in enveloping the data and then measuring the distance between the realized performance of the unit and this estimated frontier. This distance between the unit and the frontier is what we call inefficiency. In fact, the true technology defines the frontier between feasible and infeasible productions. Behind this approach there is a considerable volume of economic theory. For instance, the frontier itself is of importance as it contains all the relevant information about marginal product, elasticity of substitution and returns to scale. The characterization of the frontier rests on an axiomatic that provides some indication on what we should expect from the frontier. Still, this frontier remains not directly observable and must be estimated. This paper proposes to use a semi-parametric approach, based on the method developed by Aigner and Chu (1968), to estimate the frontier. This method is usually presented as fully deterministic, but can be viewed as a maximum likelihood estimator. The properties of this estimator have been considered to be limited and the approach was abandoned in favor of Stochastic Frontier Analysis (SFA) for which we have clear statistical properties. In this paper, we show that the Aigner and Chu estimator is better behaved than initially thought and that it is possible to conduct some inference under this framework.

There are actually two major approaches to estimate frontiers. The first approach uses operational research tools to envelop the data; the main methods are data envelopment analysis (DEA) and free disposable hull (FDH). The biggest advantage of these methods is that they do not make assumptions about the functional form of the frontier. A nonparametric envelop of the data is constructed and all observation inside the frontier of the envelope are deemed inefficient. This approach is mainly deterministic, although pioneering work by Simar and Wilson (1998) has been done to correct this defect, allowing for some inference. The second approach is rooted in econometric stochastic methods. The starting point is identical as Aigner and Chu. The input-output combinations must belong to the production set, the frontier is unknown and a parametric structure is fitted in order to contain the inefficient units. Contrary to the Aigner and Chu approach where the frontier is "landed" over the data, in the case of SFA it is argued that the frontier does not have to envelop the data and in fact, some observations can be outside because of measurement errors and the likes. Inefficiency becomes just one of the stochastic components added to the frontier (that can be either parametric or not), the other being the unobserved errors. Stochastic frontiers were introduced in articles by Aigner, Lovell and Schmidt (1977), Battese and Corra (1977), Meeusen and Broeck (1977) and Greene (1980a). Green (2008) offers a good survey of the approach.

The structure of SFA is organized around three components. The first one is the frontier itself, the second is the random component containing all the non-measurable factors that are exogenous to the decision process of the unit and the final component is inefficiency, assumed to be a stochastic component. The approach we follow here is to neglect the first purely random component and consider that only the stochastic component of the model is inefficiency. The idea consists in enveloping the data with a pre-specified functional form, more sort of landing it on the data in a way that minimizes a given distance between all the points in the data set and the frontier. This idea is from Aigner and Chu (1968). Applications of Aigner and Chu's method are not that numerous. Various standard applications are Førsund and Hjalmarsson (1979), Albriktsen and Førsund (1990), Førsund (1992). It has also been used from a cost point of view (Førsund and Jansen, 1977) and Malmquist index calculation of productivity growth (Nishimizu and Page, 1982).

Flirting with the stochastic approach in the framework of Aigner and Chu is not new. Timmer (1971) has changed the inequality constraints of the model to allow for output observations to lie above the frontier. However, this probabilistic approach amount to arbitrarily leave outside a share of the efficient units. Afriat (1972) proposal lies in between SFA and a stochastic Aigner and Chu model. He has postulated a deterministic frontier with a stochastic inefficiency. Inefficiency is modeled as a gamma distribution, an idea we find in Greene (1980b). This leads to the corrected least squares (Richmond, 1974). In fact, Schmidt (1975) has shown that the Aigner and Chu approach generates maximum likelihood estimators and Greene (1980b) has shown that standard tools cannot be used to derive the asymptotic properties of these estimators. He shows however that when the inefficiency is gamma distributed, the estimators of the frontier are well-behaved. The likelihood so obtained does not correspond to any known optimization problem and definitely not to the Aigner and Chu linear and quadratic optimization problems.

We first present the semi-parametric model in the primal space (this is by no means restrictive as the model can be applied in the dual space, *e.g.* Ouellette *et al.* (2016)). Then, we look at the consistency of the estimator. We show that under both the linear and quadratic models, the estimation procedure is consistent. Consequently, it is possible to construct an inference procedure for the parameters and measures of the technology. For this we use some sub-sampling procedures and we construct the appropriate confidence intervals. Finally, an application to the French bus transportation industry is presented to illustrate how the procedure works.

2. The Model

In this section, we present our version of the Aigner and Chu model. The framework is as follows. Let us suppose that the output and all inputs are continuous variables and that inputs can be either variable (under the control of the decision maker or discretionary) or quasi-fixed (not under the control of the decision maker at decision time, in other words non-discretionary). That is, a production activity uses a set of variable inputs $\boldsymbol{z} = \{z_i, i = 1, \ldots, p\}$ and quasi-fixed inputs $\boldsymbol{q} = \{q_j, j = 1, \ldots, s\}$ to produce the output $y \in \mathrm{R}$. Let $\mathrm{x} = (\mathrm{z}, \mathrm{q})$ and define the possible production as follows:

$$y \leq f(\mathrm{x}), \tag{2.1}$$

where f is a production function using the input vector $\boldsymbol{x}$ to produce y. The efficiency measure in the sense of Debreu-Farrell is given by:

$$TE(y, \mathrm{x}) = \frac{y}{f(\mathrm{x})} \leq 1. \tag{2.2}$$

Suppose now that the production function is parametric and that we keep the Debreu-Farrell framework, so that inefficiency enters the model multiplicatively. Then, we can write:

$$y = f(\mathrm{x}, \beta)\, TE, \tag{2.3}$$

with $0 \leq TE \leq 1$, β is the production function parameter vector. The logarithmic transformation of equation (2.3) gives:

$$\ln y = \ln f(\boldsymbol{x}, \beta) + \ln TE = \ln f(\boldsymbol{x}, \beta) + \varepsilon, \tag{2.4}$$

where $\varepsilon = \ln TE \leq 0$ is a measure of technical efficiency and it satisfies $-\varepsilon = -\ln TE \approx 1 - TE$ when small. To measure the technical efficiency and to characterize the technology we need to estimate the parameter vector β. This requires being more specific about the ε_i s.

To do so, we apply Aigner and Chu's method and build a common frontier for all firms at the same time. This imposes a constraint on the ε_i s. The procedure consists in minimizing the sum of the distance between the observed output and the efficient output of all firms. The estimated frontier clearly depends on the functional form we use. To illustrate how the procedure works, suppose that there are n firms, indexed $i = 1, \ldots, n$ and that the production function is Cobb-Douglas with k inputs:

$$y_i = e^{b_0} \prod_{j=1}^{k} x_{ji}^{b_j} TE_{ji} . \tag{2.5}$$

Taking logarithms on both sides of this equation, we get:

$$\ln y_i = \beta_0 + \sum_{j=1}^{k} \beta_j \ln x_{ji} + \varepsilon_i = \beta_0 + \beta^T \ln \boldsymbol{x}_i + \varepsilon_i . \tag{2.6}$$

The inefficiency is entirely included in ε_i and nothing prevents us from making it stochastic, but we do not do it here, and for this reason we call this model a deterministic frontier. However, it is possible to add some structure on this term. One natural thing to do is to force the ε_i s to be non-positive. Under this constraint, Aigner and Chu (1968) suggest two methods to compute the parameters. For the first procedure, the estimated parameters are the arguments minimizing the following linear program:

$$\begin{aligned} &\min_{\alpha,\beta} \left\{ -\sum_{i=1}^{n} \varepsilon_i \right\} \\ &s.t : \varepsilon_i = \ln y_i - \beta_{0i} - \beta^T \ln \boldsymbol{x}_i \leq 0 \ \forall i = 1,...,n \\ &\beta_0 \geq 0, \beta^T \geq 0 \end{aligned} \tag{2.7}$$

The alternate procedure consists in solving the following quadratic program:

$$\begin{aligned} &\min_{\alpha,\beta} \sum_{i=1}^{n} \varepsilon_i^2 \\ &s.c : \varepsilon_i = \ln y_i - \beta_0 - \beta^T \ln \boldsymbol{x}_i \leq 0 \ \forall i = 1,...,n \\ &\beta_0 \geq 0, \beta^T \geq 0. \end{aligned} \tag{2.8}$$

In both models, the slacks are measures of ε_i s, and so, we directly obtain a measure of the inefficiency score for each Decision Making Unit (DMU):

$$\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i . \tag{2.9}$$

There is no reason to restrict the production function to a Cobb-Douglas functional form. The only requirement is that the function to be estimated be linear in the parameters.

In particular, it means that $\ln f\left(\boldsymbol{x}, b\right)$ can be a translog functional form. In the next section, we derive the statistical properties of the Aigner and Chu model for such general linear in the parameters functional forms.

3. The Statistical Model

The objective of this section is to derive a procedure to obtain confidence intervals for the parameter estimates returned by the Aigner and Chu method. The first step would be to characterize the type of estimators we obtain using this procedure, and then if the estimator is valid, we need to know where it converges when the sample goes to infinity. Then only, we would be able to propose a methodology to characterize the dispersion of the estimates.

It can be shown that both, the linear and the quadratic estimators proposed by Aigner and Chu are maximum likelihood estimators. What is less obvious is that these estimators are both consistent. Consistency is the minimal requirement for statistical inference, but due to some properties of the estimator, the standard asymptotic tools cannot be used to derive the estimators' distribution. Because of this, we rely on simulation based inference methods, subsampling to be precise, to assess the dispersion of the estimates of the frontier parameters.

3.1 Maximum Likelihood Estimation

For all units *i*, with *i=1,…,n*, we suppose that the frontier is linear in the parameters, that is $\alpha_0 + \boldsymbol{X}_i\beta_0$ with true parameter vector $\left[\alpha_0 \quad \beta_0\right]$ and that all observed outputs, y_i, are on or under the frontier. Note that y_i can be a transformed output and X_i can be made of transformed inputs, cross-products and the likes (i.e. $\ln z_p$, $\ln k_s \ln z_p$, etc.), so it is compatible with a translog functional form. The frontier model is:

$$y_i = \alpha_0 + \boldsymbol{X}_i\beta_0 + \varepsilon_i, \tag{3.1}$$

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 $\alpha_0 \in \mathbb{R}$ and the $(k\times 1)$ vector $\beta_0 = \begin{bmatrix} \beta_{01} & \cdots & \beta_{0k} \end{bmatrix}^T$ are the (unknown) parameters and the scalar ε_i is the inefficiency term, clearly non positive. Stacking all n observations, we get:

$$\boldsymbol{y} = \alpha \mathbf{1}_n + \boldsymbol{X}\beta + \varepsilon, \tag{3.2}$$

where $\boldsymbol{y}$ is a $(n\times 1)$ vector of observed outputs, $\mathbf{1}_n$ is a $(n\times 1)$ vector of ones, ε is the $(n\times 1)$ vector of inefficiency and $\boldsymbol{X}$ is the $(n\times k)$ input matrix obtained by stacking the $\boldsymbol{X}_i = \begin{bmatrix} X_{i1} & \cdots & X_{ik} \end{bmatrix}$.

Suppose now that the inefficiency term is exponentially distributed. That is, $f(\varepsilon,\sigma) = (1/\sigma)\exp(\varepsilon/\sigma)$ where $-\varepsilon > 0$ and $\sigma_0 > 0$. The log-likelihood is:

$$\ell(\sigma,\alpha,\beta \mid y) = \ln L(\sigma,\alpha,\beta \mid y) = -n\ln\sigma + \frac{1}{\sigma}\sum_{i=1}^{n}(y_i - \alpha - \boldsymbol{X}_i\beta), \tag{3.4}$$

where $y_i - \alpha - \boldsymbol{X}_i\beta \le 0, \forall i = 1,\ldots,n$. Solving the first-order conditions for $\hat{\sigma}_n$ and substituting back into equation (3.4) gives the concentrated likelihood function:

$$\ell^c(\alpha,\beta \mid y) = -n\ln\hat{\sigma}_n - \frac{n}{\hat{\sigma}_n}\hat{\sigma}_n = -n\ln\left\{\frac{1}{n}\sum_{i=1}^{n}(y_i - \alpha - \boldsymbol{X}_i\beta)\right\} - n. \tag{3.5}$$

Maximizing $\ell^c(\alpha,\beta \mid y)$ consists in minimizing the first term under the negativity constraint on the ε_i. That is:

$$\begin{aligned} &\min_{\alpha,\beta}\sum_{i=1}^{n}\left(y_i - \alpha - \boldsymbol{X}_i\beta\right) \\ &s.t. \quad y_i \le \alpha + \boldsymbol{X}_i\beta, \forall i = 1,\ldots,n. \end{aligned} \tag{3.6}$$

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 (3.6) 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 (1980a) the expectation of the score is not null. That is, let the $\left(1\times(2+k)\right)$vector $S_n(\beta) = n^{-1}\partial \ln L/\partial \beta$ be the score of the likelihood, then:

$$\mathrm{E}\left[S(\boldsymbol{\theta})\right] = \begin{bmatrix} 0 & -\left(\dfrac{1}{\sigma}\right) & -\left(\dfrac{1}{\sigma}\right)\bar{\boldsymbol{X}} \end{bmatrix} \neq 0. \tag{3.7}$$

where $\boldsymbol{\theta} = \left(\sigma, \alpha, \boldsymbol{\beta}^T\right)$ and 0 and $-1/\sigma$ are scalars, and $\left(-1/\sigma\right)\bar{\boldsymbol{X}}$ is $\left(1\times k\right)$. The Hessian matrix is also nonsingular. That is:

$$-\mathrm{E}\left[\frac{1}{n}\frac{\partial^2 \ell}{\partial \theta \partial \theta^T}\right] = -\mathrm{E}\begin{bmatrix} \ell_{\sigma\sigma^T} & \ell_{\alpha\sigma^T} & \ell_{\beta\sigma^T} \\ \ell_{\sigma\alpha^T} & \ell_{\alpha\alpha^T} & \ell_{\beta\alpha^T} \\ \ell_{\sigma\beta^T} & \ell_{\alpha\beta^T} & \ell_{\beta\beta^T} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sigma^2} & -\dfrac{1}{\sigma^2} & -\dfrac{1}{\sigma^2}\bar{\boldsymbol{X}} \\ -\dfrac{1}{\sigma^2} & 0 & \boldsymbol{0}_k^T \\ -\dfrac{1}{\sigma^2}\bar{\boldsymbol{X}}^T & \boldsymbol{0}_k & \boldsymbol{0}_{(k\times k)} \end{bmatrix}. \tag{3.8}$$

Consequently, traditional methods for inference cannot be used. However, if we can establish the convergence of the estimator, then we can turn to re-sampling methods to construct confidence intervals for the estimator.

We show here that it is possible to apply Wald (1949) consistency result to the Aigner and Chu's estimator. The fundamental characteristic of the result is that it does not use the derivatives of the likelihood function. This avoids dealing with the non-zero score expectation and the singularity of the hessian matrix. There are eight assumptions we need to verify for consistency of the maximum likelihood estimator.

In Wald's theorem, there are no restrictions on the parameter space except that it is closed, i.e. the parameter space Θ is a closed set. In our model, when σ goes to zero the density degenerate and the problem collapses. Then we suppose that $\sigma \in [\underline{\sigma},\infty)$ where $\underline{\sigma} > 0$ is an arbitrarily small constant bounded away from zero. We also require the frontier parameters to be bounded from below for exactly the same reason. The economic problem is relevant for positive productions.

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

This assumption forces the exponential distribution to be well-defined for admissible ε_i (they must be all negative). 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. Now, we are in a position to state the main result.

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

$$
\begin{aligned}
\hat{\theta}_n &= \arg\max_{\theta}\left\{\ell(\sigma,\alpha,\beta \mid y) \mid y_i \le \alpha + \boldsymbol{X}_i\beta\right\} \\
&= \arg\max_{\alpha,\beta,\sigma}\left\{-n\ln\sigma + \frac{1}{\sigma}\sum_{i=1}^{n}\left(y_i - \alpha - \boldsymbol{X}_i\beta\right) \mid y_i \le \alpha + \boldsymbol{X}_i\beta\right\}
\end{aligned}
$$

Then, under Assumption 4.1, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$ is consistent, $\operatorname*{plim}_{n\to\infty}\hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_0$.

Proof. The proof is in Appendix A.

Now, we consider Aigner and Chu's quadratic model. Suppose that ε_i is distributed half-normal:

$$f\left(\varepsilon,\sigma^2\right)=\sqrt{\frac{2}{\pi\sigma^2}}\exp\left\{-\frac{\varepsilon^2}{2\sigma^2}\right\}, \tag{3.9}$$

where $\varepsilon \leq 0$. The first two moments of this distribution are $m_e = (2s^2/p)^{1/2} = 0.7978s$ and $s_e^2 = s^2\left[\left(p - 2\right)/p\right] = 0.3653s^2$. Given a sample of *n* firms, the log-likelihood is

$$\ell\left(\sigma^2,\alpha,\boldsymbol{\beta} \mid y\right)=\frac{n}{2}\ln\left(\frac{2}{\pi}\right)-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-\alpha-\boldsymbol{X}_i\beta\right)^2, \tag{3.10}$$

where $y_i - a - \boldsymbol{X}_i b$ £ 0. As above, using the first-order conditions to solve for σ^2, we find $\hat{\sigma}_n^2$ that we substitute back in likelihood function to get the concentrated log-likelihood:

$$\ell^C\left(\sigma^2,\alpha,\boldsymbol{\beta} \mid y\right)=\frac{n}{2}\left(\ln\left(\frac{2}{\pi}\right)-1\right)-\frac{n}{2}\ln\left(\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\alpha-\boldsymbol{X}_i\beta\right)^2\right). \tag{3.11}$$

It is immediate to conclude that $\ell^C\left(\sigma^2,\alpha,\boldsymbol{\beta} \mid y\right)$ is maximized when the last term in equation (3.11) is minimized under the negativity constraint of the inefficiency terms, *i.e.* $\varepsilon \leq 0$. Since the logarithm is a monotone transform, this is exactly:

$$\begin{aligned}&\min_{\alpha,\beta}\sum_{i=1}^{n}\left(y_i-\alpha-\mathrm{X}_\mathrm{i}\beta\right)^2\\&s.t.\quad y_i \leq \alpha+\mathrm{X}_\mathrm{i}\beta,\forall i=1,\ldots,n\end{aligned} \tag{3.12}$$

This is Aigner and Chu quadratic program. As in Schmidt (1975), this is summarized as follows:

Proposition 3.3. *The solution to equation (3.12) is a maximum likelihood estimator when the inefficiency terms follow a half-normal distribution.*

Again, we have the same problem we had with the exponential distribution, the consistency and asymptotic normality of the estimator cannot be proved using the standard method. The expectation of the score is not equal to zero, that is:

$$\mathrm{E}\left[n^{-1}\boldsymbol{\ell}_q\right]=\left[\begin{array}{ccc} 0 & -\frac{2}{s\sqrt{2p}} & -\frac{2}{s\sqrt{2p}}\bar{\boldsymbol{X}} \end{array}\right]\neq 0, \tag{3.12}$$

where $\boldsymbol{\theta}=\left[\sigma^2,\alpha,\boldsymbol{\beta}\right]$, and the Hessian matrix:

$$-\mathrm{E}\left[n^{-1}\frac{\partial^2\ell}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}\right]=-\mathrm{E}\left[\begin{array}{ccc} \ell_{\sigma\sigma^T} & \ell_{\alpha\sigma^T} & \ell_{\beta\sigma^T} \\ \ell_{\sigma\alpha^T} & \ell_{\alpha\alpha^T} & \ell_{\beta\alpha^T} \\ \ell_{\sigma\beta^T} & \ell_{\alpha\beta^T} & \ell_{\beta\beta^T} \end{array}\right]=\left[\begin{array}{ccc} \frac{1}{2\sigma^4} & -\frac{2}{\sqrt{2\pi}\sigma^3} & -\frac{2}{\sqrt{2\pi}\sigma^3}\bar{\boldsymbol{X}} \\ -\frac{2}{\sqrt{2\pi}\sigma^3} & \frac{1}{\sigma^2} & \frac{1}{\sigma^2}\bar{\boldsymbol{X}} \\ -\frac{2}{\sqrt{2\pi}\sigma^3}\bar{\boldsymbol{X}}^T & \frac{1}{\sigma^2}\bar{\boldsymbol{X}}^T & \frac{1}{\sigma^2}\overline{\boldsymbol{X}^T\boldsymbol{X}} \end{array}\right] \tag{3.13}$$

is clearly singular, as all second-order minors are null.

Now, we turn to the consistency of the estimator. As with the exponential distribution model, the estimator can be shown to be consistent using Wald's theorem. As before, denote by $F(y,\boldsymbol{\theta})$ the cumulative distribution of the half normal density, $f(y,q)$ where $F\left(y,q\right)=\int_{-\infty}^{y} f\left(s,q\right)ds$. Of course, $F(y,q)$ does not exist in closed form. There are $k+2$ parameters to be estimated, a,b, and σ^2. Again, we suppose that the parameter space is a closed subset of the $k+2$-Euclidean space, $\Theta\subset\mathbb{R}^{(k+2)}$. However, we have to restrict this space. As before, if there is no inefficiency the distribution collapses leading to a problem that is not well-defined. For this reason, we impose that the distribution is not degenerate by constraining the variance to be strictly positive, *i.e.* $\sigma^2\geq\underline{\sigma}^2>0$. For the problem to have an economic sense, we bound from below again the intercept and the slope parameters. For this, we need the quantity $a+\boldsymbol{X}b$ to be bounded from below. Since all inputs are finite, this is achieved if we assume that $\alpha\geq M>-\infty$ and $\beta_i\geq M>-\infty$ for all $i=1,...,k$.

Assumption 3.2. *The parameter space,* Θ, *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. In fact, to be efficient is a zero-probability event. Note also that this assumption, as before, gives one of Wald's conditions. Now, here is the second consistency result:

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

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

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

Proof. The proof is in Appendix A.

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. As we have seen in Section 2, the deterministic frontier models are not based on statistical assumptions. However, we have shown above that the Aigner and Chu model produces consistent maximum likelihood estimators. Now, we have to use the Data Generating Process (DGP) on which the estimator rests to conduct some inference for the parameters. The missing ingredient is the distribution of the estimator and since our estimator does not produce standard errors, the problem is compounded. One way out of this problem is to simulate the DGP to deduce the 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 parameter set. Simar and Wilson (1998, 1999) and Andrews (2000) document this circumstance. Following Politis and Romano (1992, 1994), we use sub-sampling, a procedure that contrasts with Efron's bootstrap method that draws pseudo-samples with replacement the size of the original sample. Bickel *et al.* (1997) and Andrews (2000) detail the properties of sub-sampling.

The 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, $\hat{b} = b\left(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\right)$. 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 $\hat{b}^b = b\left(\boldsymbol{X}_1^b, \boldsymbol{X}_2^b, \ldots, \boldsymbol{X}_m^b\right)$ with m not necessarily equal to n. In the bootstrap world, $\hat{b}^b - \hat{b}$ would have the same distribution as $\hat{b} - b$ so that replicating B times the estimation on pseudo-samples can give us a set of values that approximate the true distribution.

There is a caveat here however. The traditional bootstrap is not suited to approximate the distribution of parameters that are on the frontier of the parameter set. In fact, the observations are bounded by the parameters we are estimating. It makes it impossible to obtain observations on the « other » side of the frontier. A solution is the smooth bootstrap, proposed by Simar and Wilson (1998), but it is virtually impossible to implement in our case. An alternative is to move the noise we need around the frontier inside the frontier and then suppose that this noise can be translated on the frontier. This is roughly the sub-sampling principle. In the procedure above, we use 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 procedures 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 biased in small samples. Since the bias is given by $\text{Bias}\left[\hat{\theta}_i\right] = \text{E}(\hat{\theta}_i) - \theta$ we can correct it using sub-sampling as follows:

$$Bias_B\left[\hat{\theta}_i\right] = B^{-1}\sum_{b=1}^{B}\hat{\theta}_i^b - \hat{\theta}_i, \tag{3.14}$$

and the bias corrected estimate is:

$$\hat{\hat{\theta}}_i = \hat{\theta}_i - Bias\left[\hat{\theta}_i\right] = 2\hat{\theta}_i - B^{-1}\sum_{b=1}^{B}\hat{\theta}_{bi}^{*}. \tag{3.15}$$

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 $\hat{\theta}(x,y) - \theta$. Given the size of the interval, α, this consists in finding a_α and b_α such that:

$$\Pr\left(-b_\alpha \le \hat{\theta}(x,y) - \theta \le -a_\alpha\right) = 1 - \alpha. \tag{3.16}$$

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

$$\left[\hat{\theta}(x,y) - \theta\right] \| P \sim \left[\hat{\theta}^b(x,y) - \hat{\theta}(x,y)\right] \| P. \tag{3.17}$$

So we are justified to use the empirical distribution $\hat{\theta}_i^b(x,y) - \hat{\theta}_i(x,y)$ for b = 1, 2,…, B and i=0,…,k to approximate the distribution of $\hat{\theta}(x,y) - \theta$. Therefore, for each parameter, sort the $\hat{\theta}_i^b(x,y)-\hat{\theta}_i(x,y)$, $b=1,...,B$ in ascending order and eliminate $(\alpha/2)\times 100\,\%$ elements at each end of the distribution. The corresponding values of $\hat{\theta}_i^{\pm(\alpha/2)\times 100\,\%}(x,y)-\hat{\theta}_i(x,y)$ are the empirical values of $-a_\alpha^*$ and $-b_\alpha^*$. Now substitute in equation (3.16) a_α^* and b_α^* for a_α and b_α, and $\left[\hat{\theta}^*(x,y)-\hat{\theta}(x,y)\right]$ for $\left[\hat{\theta}(x,y)-\theta\right]$, to obtain:

$$\Pr(-b_\alpha^* \le \hat{\theta}(x,y)-\theta \le -a_\alpha^*) \approx 1-\alpha\,. \tag{3.18}$$

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

$$\hat{\theta}(x,y)+a_\alpha^* \le \theta \le \hat{\theta}(x,y)+b_\alpha^*. \tag{3.19}$$

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

4. An Application of the Semi-Parametric Model

In this section, we illustrate how to apply the method presented above. Since Aschauer (1988), public infrastructures, an input supplied by the government, have been at the center of attention in trying to determine if they were impacting returns to scale. Our application focuses on the bus transportation industry in France. We exclude all public transit companies to focus on the inter-city transportation of passengers. The data are for the period 2000-2004. After homogenization of the data, we obtain 2554 DMUs, roughly 500 DMU per year. The production process is divided into two types of inputs, variable (including labor, maintenance, energy and materials) and quasi-fixed (capital and infrastructure). The output is the number of kilometers travelled by the vehicles (including empty trips). Table 4.1 presents the exact description of the variables used, while Table 4.2 presents the descriptive statistics.

[INSERT TABLE 4.1 HERE]

[INSERT TABLE 4.2 HERE]

The functional form we used to estimate the frontier is a translog. That is,

$$\begin{aligned}\ln y = \alpha_o + \sum_{p=1}^{P} \alpha_p \ln z_p + \sum_{s=1}^{S} \alpha_s \ln k_s \\ + \frac{1}{2}\sum_{p=1}^{P}\sum_{p'=1}^{P} \beta_{pp'} \ln z_p \ln z_{p'} + \frac{1}{2}\sum_{s=1}^{S}\sum_{s'=1}^{S} \beta_{ss'} \ln k_s \ln k_{s'} + \sum_{p=1}^{P}\sum_{s=1}^{S} \beta_{ps} \ln z_p \ln k_s,\end{aligned} \tag{4.1}$$

where α_o is the constant, $z^T = \begin{bmatrix} L & E & ENT & M \end{bmatrix}$ is the vector of variable inputs (labor, L, energy, E, maintenance, ENT, miscellaneous materials, M) and $\boldsymbol{k}^T = \begin{bmatrix} V & Inf \end{bmatrix}$ is the vector of quasi-fixed inputs (the number of vehicles, V, and the infrastructure indicator, Inf). The main advantage of the semi-parametric approach is that we can directly measure the main characteristics of the technology directly from the functional form. Note that because our sample span a short period, we have neglected technical change.

Let $\hat{y}_i$ be the estimated output (the exponential of the output right-hand side of equation 4.1), then define the output oriented technical efficiency as $\varphi_i = y_i^{obs} / \hat{y}_i$. The average efficiency of the industry is given by:

$$\varphi = \sum_{i=1}^{n} \varphi_i \times \left(y_i^{obs} \Big/ \sum_{j=1}^{n} y_j^{obs} \right), \tag{4.2}$$

where φ is the average technical efficiency measure.

The variable input elasticity of output (E_p), quasi-fixed input elasticity of output (H_s) and returns to scale (η) are given by:

$$E_p = \frac{\partial \ln y}{\partial \ln z_p} = \alpha_p + \sum_{p'} \beta_{pp'} \ln z_{p'} + \sum_{s} \beta_{ps} \ln k_s , \tag{4.3}$$

$$H_s = \frac{\partial \ln y}{\partial \ln k_s} = \alpha_s + \sum_{s'} \beta_{ss'} \ln k_{s'} + \sum_{s} \beta_{ps} \ln z_p , \tag{4.4}$$

$$\eta = \sum_{p=1}^{P}\left[\alpha_p + \sum_{p'}\beta_{pp'}\ln z_{p'} + \sum_{s}\beta_{ps}\ln k_s\right] + \sum_{s=1}^{S}\left[\alpha_s + \sum_{s'}\beta_{ss'}\ln k_{s'} + \sum_{p}\beta_{ps}\ln z_p\right]$$
$$= \sum_{p=1}^{P}E_p + \sum_{s=1}^{S}H_s, \qquad (4.5)$$

where $p = (L, E, ENT, M)$ and $s = (V, Inf)$.

We have solved the linear and quadratic programs. We have also computed the confidence intervals for the parameters and returns to scale, following the sub-sampling method presented in the previous section. All calculations are made using a *SAS* program written by the authors. For the sampling procedure, we have used 3000 replications, B = 3000, with sub-sampling size of 80% (m = 0.8). Robustness checks have been made, and the results appear to be not very sensitive to the size of the sub-samples.

Tables 4.3 and 4.4 contain the estimates for the linear and quadratic program respectively. The first column contains the parameter estimates, the second the bias and the third the corrected estimates. The last six columns present the upper and lower bounds of size 90%, 95% and 99% confidence intervals, respectively. The results are consistent with what is usually found in translog estimation, as many parameters are not significant. However, for each input there are always parameters that are significant. The most intriguing results are for capital (the number of buses adjusted for capacity), as not that many parameters were significant. In fact, the own effect parameters α_V and β_{VV}, and some cross effect terms, $\beta_{L\_V}, \beta_{ENT\_V}, \beta_{V\_Inf}$ are not significant at all size. Nonetheless, the crossed terms β_{E_V}, β_{M_V} are significant at size 95%. Consequently, we cannot conclude that capital does not contribute to the frontier.

[INSERT TABLE 4.3 HERE]

[INSERT TABLE 4.4 HERE]

It is interesting to note that the optimization criterion is not making a large difference, as the results obtained from the linear program are very similar to those obtained from the quadratic optimization program. The choice of the functional form is not neutral, however. We have also estimated these models using a Cobb-Douglas production function, and the results are without any ambiguity different.

We report in Table 4.5 the input elasticities of output. Essentially, they give us the impact of each input on production. As it can be noticed, they are all positive and smaller than one. Note that again, there is virtually no difference between the linear program and the quadratic program results.

Table 4.5: Input Elasticity of Output

	E_L	E_E	E_{ENT}	E_M	H_V	H_{Inf}
Translog PL	0.613	0.187	0.052	0.024	0.014	0.496
Translog PQ	0.609	0.188	0.057	0.025	0.015	0.442

$E_p = \partial \ln y / \partial \ln z_p$, $p = L, E, ENT, M$ are the variable input elasticities of output while $H_s = \partial \ln y / \partial \ln k_s$, $s = V, Inf$ are the quasi-fixed input elasticities of output.

Now, note that the effect the marginal effect of an input on its marginal productivity can be computed as follows:

$$E_{pp} = \frac{\partial^2 y}{\partial z_p \partial z_p} = \frac{\partial}{\partial z_p}\left(E_p \times \frac{y}{z_p} \right) = \frac{\beta_{pp}}{z_p} \times \frac{y}{z_p} - E_p \times \frac{y}{z_p^2} = (\beta_{pp} - E_p)\frac{y}{z_p^2}, \quad \forall p, \qquad (4.6)$$

and similarly for the quasi-fixed inputs. The sign of E_{pp} and H_{ss} are entirely determined by $(\beta_{pp} - E_p)$ and $(\beta_{ss} - H_s)$. From Table 4.3, 4.4 and 4.5 we note that $\beta_{pp} < E_p$ and $\beta_{ss} < H_s$ so that these marginal effects are always negative. In other words, all inputs exhibit decreasing marginal productivity.

The labor elasticity is always the largest, so among the factor controlled by the firm, labor is the most important. The elasticity is equal to 0.613 for PL (0.609 for PQ) followed by energy, maintenance and materials. The relatively low vehicle elasticity, H_V = 0.0145, is probably due to the fact that vehicles are rarely full. So increasing the number of vehicles has a small effect on output. Consequently, it is probably more how the vehicles are used that has an impact. Increasing the number of drivers or the load factor may have more impact. The impact of the infrastructure is striking. Its effect is less than labor, but very important in the production process, the elasticity is equal to 0.496 for PL (0.442 for PQ). It shows that the size of the infrastructure is a key element of the production process in this industry. More infrastructures allows the firm to produce more for the same level of the other inputs, as less time is spent in traffic jams and the likes.

Now, let us turn to returns to scale measurement. To measure the returns to scale, we have projected every firm on the production frontier and measured there the returns to scale (recall that returns to scale make sense only for points on the frontier). Table 4.6 present the descriptive statistics (median, minimum, maximum, and the first and last quartiles). Confidence intervals for these quantities are also reported. Figures 1 and 2 present the distribution of the returns to scale for the linear and quadratic problems respectively.

Since we are interested by the impact of the infrastructure, we can calculate the returns to scale with and without the infrastructure. The returns to scale for the inputs is under the control of the firms are given by $\eta_{NoInf} = E_L + E_E + E_{ENT} + E_M + H_V$, leaving the variable infrastructure out of the calculation, while the long run returns to scale (including infrastructures) are given by $\eta_{Inf} = \eta_{NoInf} + H_{inf}$. When the variable infrastructure is not included in the calculation of the returns to scale, only 16.9% of the firms operate under increasing returns to scale for the linear program, and 16.2% for the quadratic program. It means that when infrastructures are ignored, most firms exhibit decreasing returns to scale. When infrastructures are included, virtually all firms operate under increasing returns to scale for both programs, linear and quadratic.

The histograms in Figure 1 and 2 make this feature obvious. At size 5% and below, it is not possible to reject the hypothesis that the firms operate under constant returns to scale when the variable infrastructure is not included, as the confidence intervals contain one. So, for what is under their control, firms operate at the optimal size. This is another story for the long run returns to scale as in both programs the null hypothesis that the firms operate under constant returns to scale is clearly rejected. It means that the size of the firms is not optimal and that some restructuring of the industry is desirable. We can interpret these results by saying that either that the infrastructures are not big enough (recall that $H_{\inf}$ is downward slopping), or there are too many small firms or a combination of both. So firms must merge and the government must invest in the infrastructure.

The solution to both programs has allowed us to compute the technical efficiency, φ_i, of all individual firms *i*. Table 4.7 contains the weighted average of the output oriented efficiency as shown in equation (4.2) for both programs. The average is also reported for the small and large firms.

Table 4.7. Technical Efficiency

	Linear	Quadratic
Total	0.853	0.859
Smallest 5%	0.749	0.755
Largest 5%	0.858	0.864

Linear and Quadratic indicate the nature of the error minimization program, φ is the weighted efficiency.

There is no difference between the two criteria used for the minimization of the errors again. The results show that it is possible to increase production by roughly 14% without changing the quantity of inputs used. The French bus transportation industry seems relatively more efficient than some of its European counterpart (Hirschhausen and Cullmann (2010), and Odeck and Alkadi (2001)) despite being subsidized by the State through some local monopolies. As expected, the smallest firms are the least efficient, while the largest are the most efficient.

Conclusion

In this paper, we have introduced a statistical semi-parametric method to estimate production frontier. The method is based on Aigner and Chu (1968). Although it has been shown that the estimators proposed is a maximum likelihood estimator, it has been believed that it was not possible to conduct statistical inference because traditional methods to show consistency and asymptotic normality did not apply. However, using Wald (1949) proof of consistency, a proof that did not require differentiability, we have shown that the estimator was consistent. Inference was based on re-sampling methods. Because the parameters to estimate define the frontier, standard bootstrap procedures were not available. Sub-sampling offers a valuable alternative that works in this case. We have just moved by translation the problem inside the data set and assumed that it worked at the frontier.

We have used this procedure to estimate the efficiency of the bus transportation industry in France. The procedure is shown to be robust. The bus transportation industry is also shown to be operating under increasing returns to scale, while if we do not take into account the infrastructure, the industry operates under constant returns to scale. These results are significant at the 95% confidence level.

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: Gains, Losses, and Remedies for Losses". *Statistica Sinica*, 7(1):1–31.

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.

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*, 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.

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., Vigeant, S., and Zhang, L. (2013). *Cost Efficiency in the French Bus Transportation Industry*. 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.

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

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.

Appendix A. Proofs

This appendix contains the proofs of the two consistency results.

A.1 The exponential case

We begin with the consistency of the ML estimator under the assumption that the inefficiency is exponentially distributed. To do this, we need to introduce some notation. Let $F(y,\theta)$ be the cumulative function of the exponential density $f(y,\theta)$. That is $F(y,\theta)=\int_{-\infty}^{y} f(s,\theta)ds$. Recall that the exponential density in our problem is defined as:

$$f(y,\theta)=f(y,\alpha,\beta,\sigma)=\frac{1}{\sigma}\exp\left\{\frac{1}{\sigma}\left(y-\alpha-\boldsymbol{X}_i\beta\right)\right\} \text{ where } \theta=(\alpha,\beta,\sigma). \tag{A.1}$$

Now, given $\rho>0$ and all θ' such that $\left\|\theta-\theta'\right\|\le\rho$, define

$$f(y,\theta\,|\,\rho):=\sup_{\theta'} f(y,\theta')=\sup_{\alpha',\beta',\sigma'}\frac{1}{\sigma'}\exp\left\{\frac{1}{\sigma'}\left(y-\alpha'-\boldsymbol{X}_i\beta'\right)\right\} \tag{A.2}$$

where $\left\|\theta-\theta'\right\|=\left[\left(\alpha-\alpha'\right)^2+\left(\beta-\beta'\right)^2+\left(\sigma-\sigma'\right)^2\right]^{1/2}$ and

$$f^*(y,\theta\,|\,\rho)=\begin{cases} f(y,\theta\,|\,\rho) & \text{if } f(y,\theta\,|\,\rho)>1 \\ 1 & \text{if } f(y,\theta\,|\,\rho)\le 1\end{cases} \tag{A.3}$$

Similarly, given $r>0$, define

$$\varphi(y,r):=\sup_{\theta} f(y,\theta)=\sup_{\alpha,\beta,\sigma}\frac{1}{\sigma}\exp\left\{\frac{1}{\sigma}\left(y-\alpha-\boldsymbol{X}_i\beta\right)\right\}, \tag{A.4}$$

for all $\left\|\theta\right\|\ge r$, and

$$\varphi^*(y,r)=\begin{cases}\varphi(y,r) & \text{si } \varphi(y,r)>1 \\ 1 & \text{si } \varphi(y,r)\le 1\end{cases}. \tag{A.5}$$

Lemma A.1. *The cumulative function* $F(y,\boldsymbol{\theta})$ *is absolutely continuous.*

Proof. This follows from Royden (1988), Proposition 4.14 since f is continuous for all admissible values of $\boldsymbol{\theta}$.

QED

This gives the first condition of Wald's theorem. The following six lemmas establish the remaining conditions to be satisfied. They essentially show that the likelihood function is well-behaved for extreme values of the parameter vectors and in the neighborhood of a given vector.

Lemma A.2. *Let* $\theta_0 = (\alpha_0, \beta_0, \sigma_0)$ *be the true parameter vector and given sufficiently small* ρ *and sufficiently large* r*, then the expectations*

$$\int_{-\infty}^{\alpha_0 + X\beta_0} \ln f^*\left(y,\theta \mid r\right) dF(y,\theta_0) \text{ and } \int_{-\infty}^{\alpha_0 + X\beta_0} \ln \varphi^*(y,r) dF(y,\theta_0)$$

are finite when evaluated with respect to the true probability measure, i.e. $F\left(y,\theta_0\right)$.

Proof. We begin by showing that $\int_{-\infty}^{\alpha_0 + X\beta_0} \ln f^*\left(y,\theta \mid r\right) dF(y,\theta_0)$ is finite. Note that since $\ln 1 = 0$ we have $\ln f^* \geq 0$. For a given admissible parameter vector, $\left(\alpha,\beta,\sigma\right) = \theta$ with $\sigma > 0$, the exponential density is monotonically increasing in ε (recall that ε is negative). Then, for all α and β such that $y \leq \alpha + \boldsymbol{X}_i\beta$ the function f is monotonically increasing in y up to the upper bound, $y = \alpha + \boldsymbol{X}_i\beta$, and $f(y,\theta) = 0$ for $\alpha + \boldsymbol{X}_i\beta \leq y \leq \alpha_0 + \boldsymbol{X}_i\beta_0$. Now, for σ sufficiently small, suppose that the function assumes values larger than one on the interval $\left[\underline{y}, \bar{y}\right]$ with $\bar{y} < \alpha_0 + \mathbf{X}_i\beta_0$.

Now, let $N_\rho(\boldsymbol{\theta})$ be a closed neighborhood of $\boldsymbol{\theta}$, satisfying $\|\boldsymbol{\theta} - \boldsymbol{\theta}'\| \leq \rho$. Clearly, this set is compact and since f is continuous in $\boldsymbol{\theta}$ the function f reaches its maximum inside this set. Denote the maximum as $\max f = \sup f \leq e^M$ with $M \leq \infty$. This holds for all $y \in \left[\underline{y}, \bar{y}\right]$ since f is monotonically increasing.

Let the indicator function for $f > 1$ be denoted $I(f > 1)$, then using $dF = fdy$, we have:

$$\begin{aligned}\int_{-\infty}^{\alpha_0 + X\phi_0} \ln f^*\left(y, \theta \mid \rho\right) dF\left(y, \boldsymbol{\theta}_0\right) &= \int_{-\infty}^{\alpha_0 + X\beta_0} \ln f^*\left(y, \theta \mid \rho\right) f(y, \theta_0) dy \\ &= \int_{-\infty}^{\alpha_0 + X\beta_0} \ln[f\left(y, \theta \mid \rho\right)] I(f > 1) f(y, \theta_0) dy \qquad \text{(A.6)} \\ &\leq \int_{-\infty}^{\alpha_0 + X\beta_0} \ln e^M f(y, \theta_0) dy = M \int_{-\infty}^{\alpha_0 + X\beta_0} f\left(y, \theta_0\right) dy = M.\end{aligned}$$

This completes the proof of the first part.

Now, let us consider the second integral. As above, it is only relevant if f assumes values larger than one for some $\boldsymbol{X}_i$ and y. Otherwise, the result trivially holds. Given $\boldsymbol{X}_i$ and for all values of y, the function $s^{-1}\exp\left\{s^{-1}(y - a - \boldsymbol{X}_i b)\right\}$ tends to its supremum when α et b are chosen such that $y - a - \boldsymbol{X}_i b \to 0$ and $\sigma \to \underline{\sigma}$. Let us denote the supremum as $\exp\{M\}$. Then, for given $\boldsymbol{X}_i$ and y, we need to choose α and β to find the supremum of f knowing that σ will ne such that $\sigma \to \underline{\sigma}$. Note also that since $\sigma \geq \underline{\sigma} > 0$, the density function is always bounded for all α and β such that $a + \boldsymbol{X}_i b \geq y$ (*i.e.* y is admissible).

It means that α and b should be such that $a + \boldsymbol{X}_i b \to y$ (in fact, $y = a + \boldsymbol{X}_i b$) with $\|\boldsymbol{\theta}\| > r$ and $\sigma = \underline{\sigma}$ if possible or $\sigma \to \underline{\sigma}$. Note also that if $y - \bar{a} - \boldsymbol{X}\bar{b} < 0$ for $\left\|\left(\bar{a}, \bar{b}\right)\right\|^2 - \underline{s}^2 = r^2$ then $f\left(y, q\right) < f\left(\bar{a} + \boldsymbol{X}_i \bar{b}, \bar{q}\right)$ with $y < \bar{a} + \boldsymbol{X}_i \bar{b}$ and $\sigma = \underline{\sigma}$.

Therefore sup φ^* is reached for $\sigma = \underline{\sigma}$, since $f(y,\theta)$ is bounded between zero and M (f reaches its maximum in $\sigma = \underline{\sigma}$ and $y = \alpha + X_i\beta$ for all y).

Now, since f is monotone in y, if there is no y such that $f \geq 1, \forall \|\theta\| > r$, then $\int_0^{\alpha_0 + X\beta_0} \log \varphi(y,r)dF = 0$. Thus, let us consider y such that $f(y,\theta) \geq 1$ for some $\|\theta\| > r$. We have:

$$\int_{y}^{\alpha_0+X\beta_0} \ln\left\{\sup_{\|\theta\|>r} f\left(y,\theta\right)\right\} dF\left(y,\theta_0\right) \leq \int_{y}^{\alpha_0+X\beta_0} \ln\left\{e^M\right\} dF\left(y,\theta_0\right) \leq M \int_{-\infty}^{\alpha_0+X\beta_0} dF\left(y,\theta_0\right) = M. \tag{A.7}$$

QED

Lemma A.3. *For all convergent sequence* $\{\theta_n\}$, $\lim_{n\to\infty}\theta_n = \theta$, *we have for all* y, $\lim_{n\to\infty} f(y,\theta_n) = f(y,\theta)$.

Proof. Since θ belongs to a closed set and since f is continuous in y and its parameters for all $\boldsymbol{\theta} \in \Theta$, then all convergent sequence in Θ, *i.e.* $\lim_{n\to\infty}\theta_n = \theta$, implies that $\lim_{n\to\infty} f(y,\theta_n) = f(y,\theta)$. In fact, this is just restating that *f* is continuous.

QED

The following lemma identifies the parameters.

Lemma A.4. *Let* θ_1 *and* θ_2 *be two distinct parameter vectors,* $\theta_1 \neq \theta_2$. *Then* $F(y,\theta_1) \neq F(y,\theta_2)$ *for at least one y.*

Proof. For two distinct parameter vectors, the exponential density function has at most finitely many points such that $f\left(y,\theta_1\right) = f\left(y,\theta_2\right)$. Consequently, except for a finite number of *y*s,

$$f\left(y,\theta_1\right)=\frac{1}{\sigma_1}\exp\left\{\frac{1}{\sigma_1}\left(y-\alpha_1-\boldsymbol{X}_i\beta_1\right)\right\}\neq\frac{1}{\sigma_2}\exp\left\{\frac{1}{\sigma_2}\left(y-\alpha_2-\boldsymbol{X}_i\beta_2\right)\right\}=f\left(y,\theta_2\right).$$

(A.8)

Now, suppose that $F(\tilde{y},\theta_1)=F(\tilde{y},\theta_2)$ for some $\tilde{y}$, we need to show that there exists a y such that $F\left(y,q_1\right)$ ¹ $F(y,q_2)$. Now, by assumption for $\theta_1,\theta_2\in\Theta$ such that $\theta_1\neq\theta_2$ and $\tilde{y}$ we have:

$$F\left(\tilde{y},\theta_1\right)=\int_{\tilde{y}}^{\alpha_1+X\beta_1}\ln f\left(s,\beta_1\right)ds=\int_{\tilde{y}}^{\alpha_2+X\beta_2}\ln f\left(s,\beta_2\right)ds=F(\tilde{y},\beta_2). \quad \text{(A.9)}$$

For an arbitrarily small Δy define $y=\tilde{y}+\Delta y$. Then we have:

$$\begin{aligned}F\left(y,\theta_1\right)&=\int_{y}^{\alpha_1+X\beta_1}f\left(s,\theta_1\right)ds=\int_{\tilde{y}}^{\alpha_1+X\beta_1}f\left(s,\theta_1\right)ds+\int_{y}^{\tilde{y}}f\left(s,\theta_1\right)ds\\&\approx F\left(\tilde{y},\theta_1\right)+f\left(y,\theta_1\right)\Delta y.\end{aligned} \quad \text{(A.10)}$$

Similarly, we have:

$$F\left(y,\theta_2\right)=\int_{y}^{\alpha_0+X\beta_0}f\left(s,\beta_2\right)ds\approx F\left(\tilde{y},\beta_2\right)+f\left(y,\beta_2\right)\Delta y. \quad \text{(A.11)}$$

Then

$$F\left(y,\theta_1\right)-F\left(y,\theta_2\right)\cong\left(f\left(y,\theta_1\right)-f\left(y,\theta_2\right)\right)\Delta y\,. \quad \text{(A.12)}$$

Since $f\left(y,\theta_1\right)-f\left(y,\theta_2\right)\neq 0$ (except for a finite number of points, in which case we choose $\bar{y}\neq y$) we have $F\left(y,\boldsymbol{\theta}_1\right)\neq F\left(y,\boldsymbol{\theta}_2\right)$.

QED

Lemma A.5. *When* $\lim_{n\to\infty}\left\|q_n\right\|=\infty$, *then* $\lim_{n\to\infty}f\left(y,q_n\right)=0$ *for all* y, (*except possibly on a set of measure zero with respect to* $F\left(y,q_0\right)$).

Proof. Since there is a lower bound on σ, α and β we only have to consider $\lim_{n\to\infty}\|\theta_n\| = \infty$ for positive values of the parameters. We have two cases to consider. First suppose that $\sigma_i \to \infty$ at the same rate or faster than α and β. Then $\sigma^{-1}(y-\alpha-X_i\beta)$ is bounded or goes to zero. But, then:

$$\lim_{i\to\infty}\frac{1}{\sigma_n}\exp\left\{\frac{y-X_i\beta_n-\alpha_n}{\sigma_n}\right\}\to 0 \quad \forall y. \tag{A.13}$$

Otherwise, if $\alpha,\beta\to\infty$, we also have $\exp\left\{\sigma^{-1}(y-\alpha-X_i\beta)\right\}\to 0$ and the result is trivial. Now, $\alpha+X_i\beta \geq y$ and $\|\theta\|\to\infty$ are sufficient to handle the possible negative values of α and β because in this specific case we must have $\sigma\to\infty$ at a faster rate than $\|\alpha,\beta\|\to\infty$.

QED

Lemma A.6. $\int_{-\infty}^{\infty}|\ln f(y,\theta_0)|\,dF(y,\theta_0) \leq M < \infty$ *for the true* θ_0.

Proof. Using the continuity of the exponential distribution, we have $dF(y,\theta_0)=f(y,\theta_0)dy$ and by definition we also have $y \leq \alpha_0 + X_i\beta_0$, so

$$\begin{aligned}\int_{-\infty}^{\infty}\left|\ln f(y,\theta_0)\right|dF(y,\theta_0) &= \int_{-\infty}^{\alpha_0+X\beta_0}\left|-\ln\sigma_0+\frac{1}{\sigma_0}\left(y-X_i\beta_0-\alpha_0\right)\right|f(y,\theta_0)dy\\ &\leq \int_{-\infty}^{\alpha_0+X\beta_0}\left|-\ln(\sigma_0)\right|f(y,\theta_0)dy+\int_{-\infty}^{\alpha_0+X\beta_0}\left|\frac{1}{\sigma_0}\left(y-X_i\beta_0-\alpha_0\right)\right|f(y,\theta_0)dy\\ &\leq \int_{-\infty}^{\alpha_0+X\beta_0}\left|-\ln(\sigma_0)\right|dF(y,\theta_0)+\int_{-\infty}^{\alpha_0+X\beta_0}\left|\frac{1}{\sigma_0}\left(y-X_i\beta_0-\alpha_0\right)\right|dF(y,\theta_0).\end{aligned} \tag{A.14}$$

The first term is clearly bounded. Note also that

$$\lim_{y\to-\infty}\frac{1}{\sigma_0}\left(y-X_i\beta_0-\alpha_0\right)\cdot\exp\left\{\frac{1}{\sigma_0}\left(y-X_i\beta_0-\alpha_0\right)\right\}\to 0. \tag{A.15}$$

Thus we can rewrite the last term as follows:

$$\int_{-\infty}^{\alpha_0+X\beta_0}\left|\frac{1}{\sigma_0}\left(y-\boldsymbol{X}_i\beta_0-\alpha_0\right)\right|dF\left(y,\boldsymbol{\theta}_0\right)$$
$$=\int_{0}^{\alpha_0+X\beta_0}\left|\frac{1}{\sigma_0}\left(y-\boldsymbol{X}_i\beta_0-\alpha_0\right)\right|dF\left(y,\boldsymbol{\theta}_0\right)+\int_{-\infty}^{0}\left|\frac{1}{\sigma_0}\left(y-\boldsymbol{X}_i\beta_0-\alpha_0\right)\right|dF\left(y,\boldsymbol{\theta}_0\right). \tag{A.16}$$

The first term is clearly finite, while the second is easily shown to be finite by simple integration.

QED

Lemma A.7. *$f\left(y,\boldsymbol{\theta}\,|\,\rho\right)$ is a measurable function of y for all $\boldsymbol{\theta}$ and ρ.*

Proof. Note that since Θ is closed then for $N_\rho\left(\boldsymbol{\theta}\right)=\left\|\boldsymbol{\theta}-\boldsymbol{\theta}'\right\|\leq\rho$, either $N_\rho\left(\boldsymbol{\theta}\right)\subseteq\Theta$ or $N_\rho\left(\boldsymbol{\theta}\right)\cap\Theta$ is entirely in the parameter space and clearly both are compact. From the definition of $f(y,\boldsymbol{\theta})$, we know that $f\left(y,\boldsymbol{\theta}\,|\,\rho\right)$ is well defined and, using the compactness of $N_\rho\left(\boldsymbol{\theta}\right)$, this function assumes its maximum for the smallest σ in $N_\rho\left(\boldsymbol{\theta}\right)$ (or $N_\rho\left(\boldsymbol{\theta}\right)\cap\theta$). Let us denote this value as $\sigma_{\min}=\min\sigma\in N_\rho\left(\boldsymbol{\theta}\right)$. Since $N_\rho\left(\boldsymbol{\theta}\right)$ is compact, $\sigma_{\min}$ is a limit point of this set. Then, for such a $\sigma_{\min}$ the *supremum* of

$$f\left(y,\boldsymbol{\theta}'\right)=\sup_{\alpha,\boldsymbol{\beta}}\frac{1}{\sigma_{\min}}\exp\left\{\frac{1}{\sigma_{\min}}\left(y-\alpha-\boldsymbol{X}_i\beta\right)\right\} \tag{A.17}$$

is reached for the largest $y-\alpha-\boldsymbol{X}_i\beta$ possible. When possible, we choose α and β, such that $y-\alpha-\boldsymbol{X}_i\beta=0$, with $\alpha,\boldsymbol{\beta}\in N_\rho\left(\boldsymbol{\theta}\right)$ (or $N_\rho\left(\boldsymbol{\theta}\right)\cap\theta$). For any other y we have by definition that $y-\alpha-\boldsymbol{X}\beta\leq 0$. Now, if $y-\alpha-\boldsymbol{X}_i\beta<0$ for all $\underline{\sigma},\alpha,\beta\in N_\rho\left(\boldsymbol{\theta}\right)$, then we let $\alpha=\underline{\alpha}$ and $\beta=\underline{\beta}$, the largest value of $y-\underline{\alpha}-\boldsymbol{X}_i\underline{\beta}$. Thus, $f\left(y,\boldsymbol{\theta}\,|\,\rho\right)$ is either constant or decreasing and in both case continuous except possibly for finitely many points. Consequently, $f\left(y,\boldsymbol{\theta}\,|\,\rho\right)$ is measurable.

QED

Proof of Proposition 3.2. By Lemma 1, the exponential specification ensures that Assumption 1 of Wald (1949) is satisfied since the cumulative distribution is absolutely continuous. Since our parameter space is closed, Wald's Assumption 7 is satisfied. Lemmas A.2 à A.7 verifies the other assumptions. Consequently, by Wald's Theorem 2, we have the consistency of our estimator.

QED

A.2 The half-normal case

Recall that the half-normal density in our problem is defined as:

$$f\left(y,\theta\right)=f\left(y,\alpha,\beta,\sigma^2\right)=\sqrt{\left(2/\pi\sigma^2\right)}\exp\left\{\left(-1/2\sigma^2\right)(y-\alpha-\boldsymbol{X}\beta)^2\right\} \qquad \text{(A.18)}$$

where $\theta=\left(\alpha,\beta,\sigma^2\right)$. Using the same definition as in the previous subsection, given $\rho>0$ and all θ' such that $\left\|\theta-\theta'\right\|\leq r$, define

$$f\left(y,\theta\,|\,\rho\right):=\sup_{\theta'} f\left(y,\theta'\right)=\sup_{\alpha',\beta',\sigma'}\frac{2}{\sqrt{2\pi\sigma'}}\exp\left\{-\frac{1}{2(\sigma^2)'}\left(y-\alpha'-\boldsymbol{X}\beta'\right)^2\right\} \qquad \text{(A.19)}$$

and let $f^*\left(y,\theta\,|\,r\right)$ be as in equation (A.3). Given $r>0$, define

$$\varphi\left(y,r\right):=\sup_{\theta} f\,(y,)=\sup_{\alpha,\beta,\sigma}\frac{2}{\sqrt{2\pi\sigma}}\left\{-\frac{1}{2\sigma^2}(y-\alpha-\boldsymbol{X}\beta)^2\right\}, \qquad \text{(A.20)}$$

And let $\varphi^*\left(y,r\right)$ be as in equation (A.5).

Lemma A.8. *The cumulative distribution* $F\left(y,\theta\right)$ *is absolutely continuous.*

Proof. Identical to the proof of Lemma A.1.

QED

Lemma A.9. *Let* $\theta_0=(\alpha_0,\beta_0,\sigma_0)$ *be the true parameter vector and given sufficiently small* ρ *and sufficiently large* r*, then the expectations*

$$\int_{-\infty}^{\alpha_0+X\beta_0}\ln f^*\left(y,\theta\,|\,r\right)dF(y,\theta_0)\ \ and\ \ \int_{-\infty}^{\alpha_0+X\beta_0}\ln\varphi^*(y,r)dF(y,\theta_0)$$

are finite when evaluated with respect to the true probability measure, i.e. $F\left(y,\theta_0\right)$.

Proof. Requires only minor adjustment to the proof of Lemma A.2.

QED

Lemma A.10. *For all convergent sequence* $\{\theta_i\}$, $\lim_{i\to\infty}\theta_i=\theta$, *we have for all* y, $\lim_{i\to\infty} f(y,\theta_i)=f(y,\theta)$.

Proof. Identical to the proof of Lemma A.3.

QED

Lemma A.11. *Let* θ_1 *and* θ_2, *be two different parameter vectors,* $\theta_1 \neq \theta_2$. *Then* $F(y,\theta_1)\neq F(y,\theta_2)$ *for at least one* y.

Proof. Identical to the proof of Lemma A.4.

QED

Lemma A.12. *If* $\lim_{i\to\infty}\|\theta_i\|=\infty$, *then* $\lim_{i\to\infty} f(y,\theta_i)=0$ *for all* y.

Proof. Since $f(y,\theta)=\sqrt{2/\pi\sigma^2}\exp\left\{-(y-\alpha-X\beta)^2/2\sigma^2\right\}$ and the fact that $\left(y-\alpha-X\beta/\sigma\right)^2$ is bounded above for all admissible combination of (α,β,σ^2), then $\exp\left\{-\left(y-\alpha-X\beta\right)^2/2\sigma^2\right\}$ is bounded above when $\sigma_i^2\to\infty$ or when $\left\|(\alpha_i,\beta_i,\sigma_i^2)\right\|\to\infty$. Consequently:

$$\lim_{\|(\alpha_i,\beta_i,\sigma_i)\|\to\infty}\sqrt{\frac{2}{\pi\sigma_i^2}}\exp\left\{\frac{-1}{2\sigma_i^2}(y-\alpha_i-X\beta_i)^2\right\}\to 0 \quad \forall\, y. \qquad \text{(A.21)}$$

QED

Lemma A.13. *For the true parameter vector,* θ_0, *we have that* $\int_{-\infty}^{\infty}|\ln f(y,\theta_0)|\,dF(y,\theta_0)\leq M<\infty$.

Proof. It requires only minor modifications to the proof of Lemma A.6.

QED

Lemma A.14. $f(y,\theta\,|\,\rho)$ *is a measurable function of* y *for all* θ *and* ρ.

Proof. It requires only minor modifications to the proof of Lemma A.7.

QED

Proof of Proposition 3.4. Replace Lemmas A2 to A7 by Lemmas A8 to A14 in the proof of Proposition 3.2.

QED

Table 4.1: Description of the variables

	Variable	Symbol	Data	Description	Sources
Output	Kilometer	Y	*KMTURV10*	Kilometers, including empty trips (total passenger activity)	EAE
Variable Inputs	Labor	Quantity (L)	*TOTEFF*	Total employment (full time equivalent)	EAE
		Price (w_L)	D_L/L	Cost of labor for the employer	Calculated
		Expenditures (D_L)	*RCH31+RCH32+RCH26*	Sum of wages, taxes and external labor subcontracting	EAE
	Fuel	Quantity (E)	D_E/W_E	Total fuel.	Calculated
		Price (w_E)	Price index	CPI – All households continental France – Fuel.	INSEE
		Expenditures (D_E)	*ACHACARB*	Fuel total expenses	EAE
	Maintenance	Quantity (ENT)	D_{ENT}/W_{ENT}	Maintenance	Calculated
		Price (w_{ENT})	Price index	CPI- All households continental France – Car repair	INSEE
		Expenditures (D_{ENT})	*RCH28S2(REPARTOT)*	Maintenance total expenses	EAE
	Material	Quantity (M)	D_M/W_M	Materials	Calculated
		Price (w_M)	Laspeyres Price index	Laspeyres price index computed from the Input-Ouput matrix	Calculated
		Expenditures (D_M)	*RCH701+RCH71+RCH25+RCH28S4-ACHACARB-RCH26*	Material total expenses	EAE
Quasi-Fixed Inputs	Number of bus	K_B	*BUS+BUSA*	Number of buses used	EAE
	Number of bus	K_{C1}	*CAR10*	Buses, 9 seats or less	EAE
		K_{C2}	*CAR20*	Buses, 10 to 29 seats	EAE
		K_{C3}	*CAR30*	Buses, 30 to 49 seats	EAE
		K_{C4}	*CAR40*	Buses, 50 to 59 seats	EAE
		K_{C5}	*CAR50*	Buses, 60 seats or more	EAE

Infrastructures	Average speed	Vt_i	$Vt_i = N^{-1}\sum_{r=1}^{N} Vt_{ir}$	Average speed in employment zone i	Computed
	Total length of highways	L_i	$L_i = \sum_{r=1}^{N} L_{ir}$	Total road length in employment zone i	CGDD
	Surface	ST_i	*SUP*	Area of employment zone i	INSEE
	Distance	d_{ij}	Mapinfo Network Map	Distance between the center of zone i and j	Computed
	Congestion time	t_{con}	$TIME_j$	Time wasted because of traffic jams r	CGDD
	Goods flow	F_{ij}	F_{ij}	Flow of goods in 2002 between zone i and j	CGDD
	Infrastructures	Inf_i	$Inf_i = \left(\frac{L_i V_i}{ST_i}\right) * \sum_j L_j \, exp\{-\nu d_{ij} - \gamma t_{conij}\}$	Accessibility index for employment zone i	Computed

Note: EAE is the *Enquête annuelle des entreprises* (French Annual survey of firms), INSEE is the *Institut national de la statistique et des études économiques* (the French National Agency for Statistics and Economic Studies), and CGDD is the *Commissariat Général au Développement Durable* (the Sustainable Development Agency of the French Governement).

Table 4.2: Descriptive statistics

	Variable	Symbol	N	Maximum	Minimum	Standard error	Mean	Median
Output	Kilometer	Y	2,554	45,040,389	10,886	3,060,574.48	2,097,411.79	1,183,424.50
	Variable expenditures	D_{COST}	2,554	144,424	16.86	7,531.98	3,349.87	1,594
Variable Inputs	Labor	L	2,554	1,957	2	121.37	67.89	36
		w_L	2,554	80.86	0.50	7.42	26.19	26.13
		D_L	2,554	94,433	1	4,444.43	1,931.42	916.50
	Fuel	E	2,554	8,968.19	1.05	687.30	451.83	248.32
		w_E	2,554	102.02	92.34	3.71	96.87	94.92
		D_E	2,554	9,149	1	667.07	437.87	240
	Maintenance	ENT	2,554	3,997.19	1.78	251.84	141.56	69.47
		w_{ENT}	2,554	112.28	100	4.72	105.10	103.77
		D_{ENT}	2,554	4,488	2	267.11	149.16	73
	Material	M	2,554	47,847.68	2.82	2,420.73	802.80	314.67
		w_M	2,554	106.49	100	2.37	103.46	104
		D_M	2,554	48,635	3	2,501.16	831.42	323
Quasi-Fixed Inputs	Number of bus	K_B	2,554	716	0	31.52	7.74	0
	Number of bus	K_{car_p}	2,554	43,020	0	3,341.86	2,393.96	1,395
		K_{C1}	2,554	65	0	4.31	1.13	0
		K_{C2}	2,554	127	0	6.20	2.05	0
		K_{C3}	2,554	127	0	9.65	2.99	0
		K_{C4}	2,554	682	0	37.74	14.65	0
		K_{C5}	2,554	171	0	6.73	1.34	0
	Number of seats in vehicles	KV	2,554	67,270	5	4,088.05	2,742.05	1,575
Infrastructures	Infrastructures	Inf_i	337	15.31	7.81	1.01	12.82	12.85
	Congestion time	t_{con}	337	120.43	5.85	11.40	20.02	17.42
	Average speed	Vt_i	337	116.41	37.74	12.54	69.01	68.67

	Length	L_i	337	856.20	11.37	152.27	285.14	255.71
	Distance	d_{ij}	115,940	1,044.28	8.61	191.81	395.60	386.20
	Surface	ST_i	337	6,256.05	46.36	998.52	1,591.49	1,448.19
	Goods flow	F_{ij}	116,281	4,280,771.20	1	29,019.03	2,318.28	79.50

Table 4.3: Parameter estimates and their confidence intervals – Linear program

	Estimated parameter	Bias	Corrected estimates	$I_{90\%}$		$I_{95\%}$		$I_{99\%}$	
				Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
α_L	0.903	-0.068	0.970	0.441	1.576	0.354	1.671	0.186	1.758
α_E	0.012	0.059	-0.047	-0.326	0.146	-0.388	0.151	-0.507	0.157
α_{ENT}	-0.038	0.040	-0.077	-0.225	0.014	-0.250	0030	-0.322	0.140
α_M	-0.042	0.039	-0.081	-0.375	0.056	-0.439	0.083	-0.559	0.237
α_V	-0.021	-0.038	0.017	-0.042	0.203	-0.042	0256	-0.281	0.393
α_{Inf}	2.264	-0.324	2.588	2.010	3.379	1.902	3.466	1.738	3.618
β_{LL}	0.251	-0.026	0.277	0.198	0.419	0.191	0.454	0.167	0.520
β_{EE}	0.063	-0.016	0.079	0.041	0.136	0.033	0.145	0.018	0.163
β_{ENTENT}	0.003	-0.003	0.006	-0.002	0.028	-0.028	0.034	-0.036	0.048
β_{MM}	-0.002	0.000	-0.002	-0.012	0.014	-0.014	0.018	-0.018	0.026
β_{VV}	0.007	0.001	0.005	-0.009	0.014	-0.014	0.016	-0.024	0.018
β_{InfInf}	-0406	0.065	-0.472	-0.626	-0.227	-0.669	-0.202	-0.751	-0.177
β_{L_E}	-0.094	0.013	-0.107	-0.170	-0.067	-0.181	-0.057	-0.191	-0.044
β_{L_ENT}	-0.036	0.001	-0.037	-0.072	-0.008	-0.072	0.019	-0.075	0.038
β_{L_M}	-0.017	0.000	-0.017	-0.035	0.016	-0.040	0.020	-0.058	0.032
β_{L_V}	0.005	0.000	0.005	-0.013	0.018	-0.021	0.026	-0.044	0.036
β_{L_Inf}	-0.086	0.016	-0.102	-0.174	-0.045	-0.201	-0.031	-0.248	0.016
β_{E_ENT}	0.015	0.002	0.012	-0.007	0.029	-0.010	0.030	-0.015	0.030
β_{E_M}	0.006	0.000	0.006	-0.005	0.013	-0.007	0.014	-0.012	0.021
β_{E_V}	-0.009	0.006	-0.015	-0.026	-0.007	-0.029	-0.005	-0.035	-0.001
β_{E_INF}	0.028	-0.022	0.050	0.020	0.103	0.014	0.117	-0.005	0.143
β_{ENT_M}	0.005	-0.001	0.005	-0.004	0.009	-0.007	0.009	-0.013	0.009
β_{ENT_V}	-0.006	-0.002	-0.004	-0.012	0.016	-0.012	0.019	-0.013	0.031
β_{ENT_Inf}	0.028	-0.003	0.031	0.007	0.055	0.003	0.056	-0.013	0.061
β_{M_V}	0.002	-0.001	0.003	0.001	0.018	0.000	0.022	-0.003	0.035
β_{M_Inf}	0.014	-0.001	0.015	-0.017	0.040	-0.023	0.050	-0.044	0.072
β_{V_Inf}	0.007	0.007	0.000	-0.042	0.014	-0.052	0.014	-0.084	0.070

The first column contains the original estimates, $bias_B\left[\hat{\alpha}_i\right]$ is the bootstrap estimate of the bias, the third column contains the corrected estimates, and finally the $I_{\alpha\%}$ are the bounds of the confidence intervals of size α for $\alpha = 90\%, 95\%, 99\%$.

Table 4.4: Parameter estimates and their confidence intervals – Quadratic program

	Parameter Estimates	Bias	Corrected estimates	$I_{90\%}$		$I_{95\%}$		$I_{99\%}$	
				Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
α_L	0.901	0.035	0.866	0.389	1.373	0.333	1.460	0.237	1.594
α_E	0.007	0.003	0.004	-0.399	0.208	-0.495	0.232	-0.607	0.269
α_{ENT}	-0.009	0.027	-0.035	-0.161	0.091	-0.186	0.176	-0.227	0.233
α_M	-0.042	0.058	-0.100	-0.396	0.066	-0.433	0.114	-0.490	0.202
α_V	-0.023	0.041	-0.063	-0.045	0.260	-0.088	0.303	-0.342	0.380
α_{Inf}	2.241	-0.142	2.384	1.807	2.578	1.679	2.709	1.475	2.901
β_{LL}	0.251	-0.072	0.323	0.245	0.446	0.240	0.482	0.218	0.547
β_{EE}	0.063	-0.008	0.071	0.036	0.139	0.032	0.145	0.018	0.160
β_{ENTENT}	0.004	-0.004	0.008	-0.008	0.024	-0.025	0.034	-0.030	0.042
β_{MM}	-0.002	0.000	-0.002	-0.013	0.014	-0.015	0.016	-0.019	0.020
β_{VV}	0.007	0.006	0.001	-0.013	0.014	-0.016	0.014	-0.023	0.014
β_{InfInf}	-0.398	0.013	-0.411	-0.583	-0.209	-0.616	-0.164	-0.668	-0.132
β_{L_E}	-0.094	0.006	-0.100	-0.164	-0.061	-0.176	-0.053	-0.196	-0.033
β_{L_ENT}	-0.036	0.007	-0.043	-0.073	0.002	-0.075	0.016	-0.078	0.034
β_{L_M}	-0.017	0.004	-0.021	-0.043	0.021	-0.048	0.023	-0.055	0.031
β_{L_V}	0.005	0.001	0.004	-0.021	0.023	-0.028	0.027	-0.045	0.035
β_{L_Inf}	-0.085	0.013	-0.098	-0.203	-0.034	-0.234	-0.021	-0.283	0.002
β_{E_ENT}	0.016	0.002	0.014	-0.003	0.031	-0.006	0.033	-0.014	0.034
β_{E_M}	0.006	0.000	0.006	-0.008	0.014	-0.010	0.016	-0.015	0.023
β_{E_V}	-0.010	0.008	-0.017	-0.028	-0.007	-0.031	-0.005	-0.045	0.000
β_{E_INF}	0.029	-0.020	0.049	-0.003	0.090	-0.032	0.119	-0.040	0.122
β_{ENT_M}	0.005	-0.001	0.005	-0.002	0.009	-0.003	0.009	-0.007	0.012
β_{ENT_V}	-0.006	-0.006	0.000	-0.012	0.019	-0.012	0.023	-0.015	0.030
β_{ENT_Inf}	0.021	0.003	0.018	-0.005	0.043	-0.014	0.049	-0.031	0.053
β_{M_V}	0.003	-0.006	0.008	0.001	0.026	0.000	0.028	-0.002	0.035
β_{M_Inf}	0.014	-0.002	0.017	-0.018	0.048	-0.026	0.055	-0.040	0.064
β_{V_Inf}	0.007	0.018	-0.011	-0.059	0.015	-0.066	0.030	-0.086	0.080

The first column contains the original estimates, $bias_B\left[\hat{\alpha}_i\right]$ is the bootstrap estimate 0f the bias, the third column contains the corrected estimates, and finally the $I_{\alpha\%}$ are the bounds of the confidence intervals of size α for $\alpha = 90\%, 95\%, 99\%$.

Table 4.6: Confidence intervals for the returns to scale

Translog		Estimates	Bias	Corrected estimates	$I_{90\%}$		$I_{95\%}$		$I_{99\%}$	
					Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
η_{Inf} Linear program	Median	1.258	-0.014	1.272	1.223	1.348	1.212	1.360	1.181	1.385
	Min	1.038	-0.042	1.080	1.026	1.110	1.018	1.179	0.977	1.220
	Max	1.562	0.119	1.443	1.234	1.663	1.223	1.700	1.216	1.752
	P25	1.193	-0.021	1.214	1.163	1.286	1.146	1.302	1.120	1.328
	P75	1.332	-0.015	1.347	1.257	1.447	1.239	1.461	1.190	1.479
η_{NoInf} Linear program	Median	0.932	0.007	0.925	0.847	0.989	0.843	1.026	0.827	1.080
	Min	0.707	-0.061	0.768	0.598	0.939	0.579	0.947	0.528	0.956
	Max	1.156	0.044	1.112	0.813	1.278	0.801	1.321	0.752	1.351
	P25	0.890	0.002	0.888	0.808	0.955	0.804	0.969	0.792	1.033
	P75	0.977	0.013	0.964	0.830	1.069	0.828	1.090	0.811	1.124
η_{Inf} Quadratic program	Median	1.261	-0.019	1.280	1.233	1.353	1.228	1.360	1.207	1.380
	Min	1.033	-0.067	1.100	1.025	1.098	1.016	1.100	0.989	1.101
	Max	1.560	0.185	1.375	1.223	1.628	1.217	1.662	1.212	1.702
	P25	1.195	-0.041	1.236	1.178	1.294	1.170	1.302	1.146	1.315
	P75	1.337	-0.009	1.346	1.273	1.446	1.264	1.454	1.233	1.467
η_{NoInf} Quadratic program	Median	0.931	-0.019	0.950	0.888	1.019	0.882	1.028	0.867	1.063
	Min	0.709	-0.012	0.721	0.559	0.851	0.545	0.873	0.529	0.910
	Max	1.150	-0.072	1.202	1.074	1.301	1.035	1.324	0.921	1.351
	P25	0.890	-0.009	0.899	0.822	0.963	0.817	0.977	0.805	1.010
	P75	0.975	-0.031	1.006	0.940	1.077	0.924	1.095	0.887	1.120

The first column contains the original estimates, $bias_B\left[\hat{\alpha}_i\right]$ is the bootstrap estimate of the bias, the third column contains the corrected estimates, and finally the $I_{\alpha\%}$ are the bounds of the confidence intervals of size α for $\alpha = 90\%, 95\%, 99\%$.

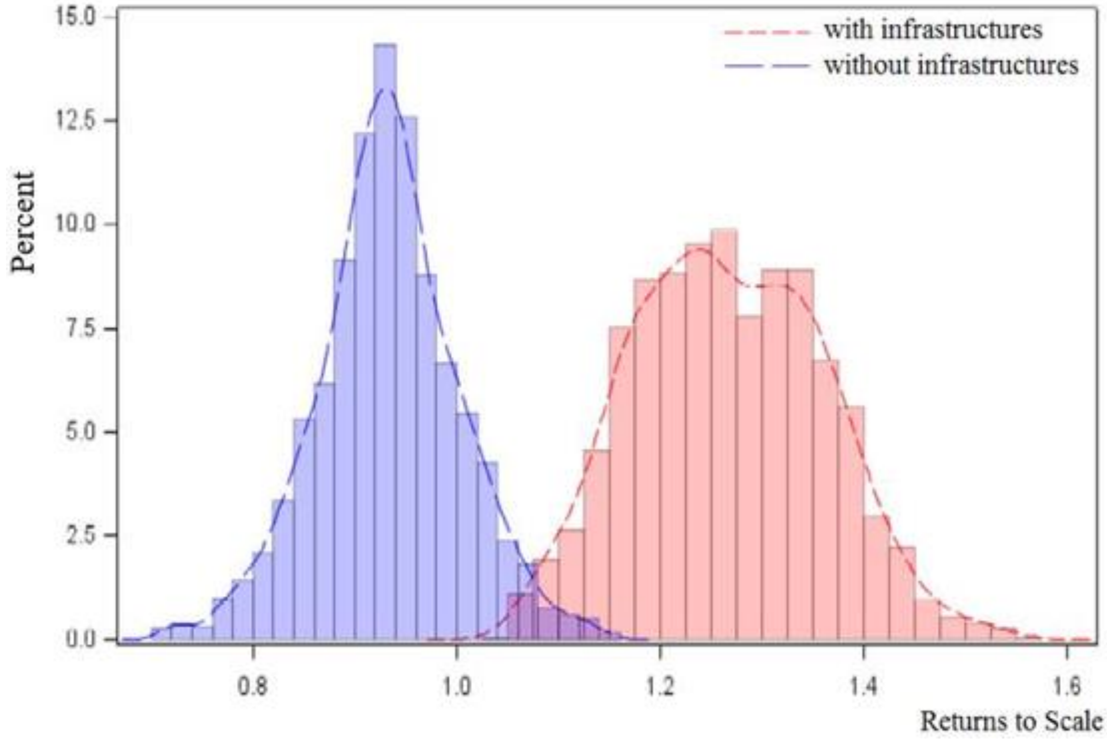

FIGURE 1. Returns to Scale Histogram, with and without infrastructures, Linear Programming

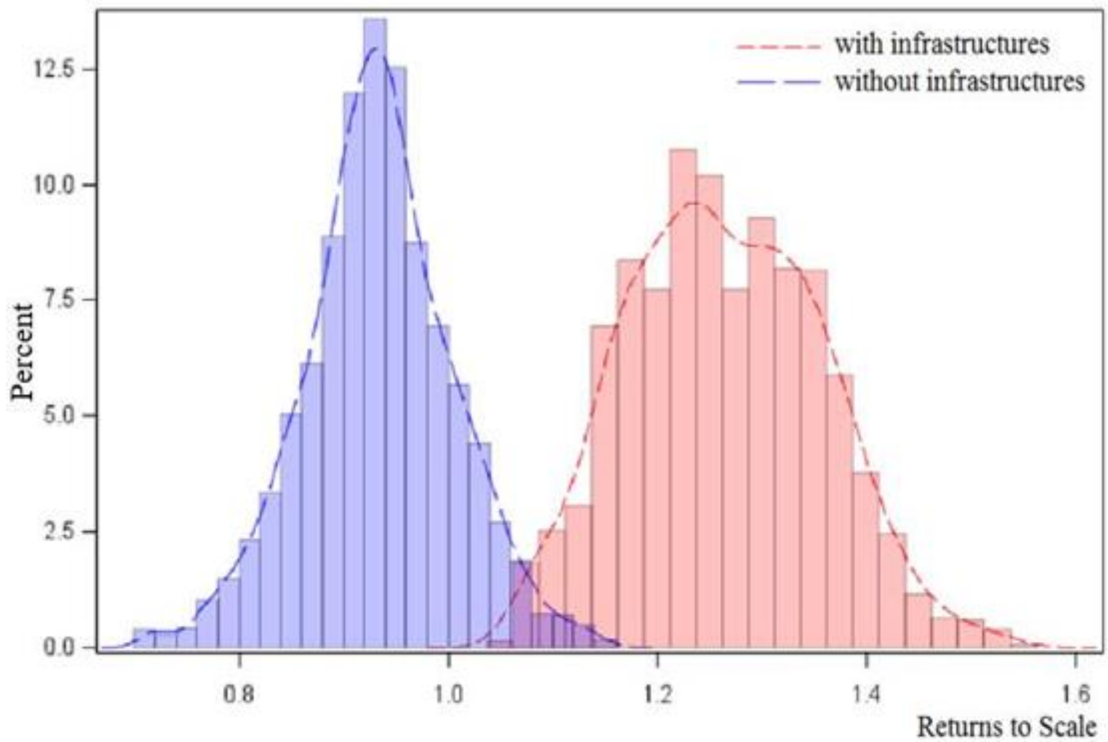

FIGURE 2. Returns to Scale Histogram, with and without infrastructures, Quadratic Programming