On the generation of a regular multi-input multi-output technology using parametric output distance functions

Sergio Perelman [a] and Daniel Santin [b]

[a] CREPP, Université de Liège
[b] Universidad Complutense de Madrid

Abstract

Monte-Carlo experimentation is a well-known approach to test the performance of alternative methodologies under different hypothesis. In the frontier analysis framework, whatever parametric or non-parametric methods tested, most experiments have been developed up to now assuming single output multi-input production functions and data generated using a Cobb-Douglas technology. The aim of this paper is to show how reliable multi-output multi-input production data can be generated using a parametric output distance function approach. A flexible *translog* technology is used for this purpose that satisfies regularity conditions. Two meaningful outcomes of this analysis are the identification of a valid range of parameters values satisfying monotonicity and curvature restrictions and of a *rule of thumb* to be applied in empirical studies.

JEL Classification: C14, C15, C24
Keywords: Output distance function; technical efficiency, Monte-Carlo experiments.
Acknowledgments: The authors are grateful to Chris O'Donnell, David Roibás, Luis Orea and participants in the *IX European Workshop on Efficiency and Productivity Analysis* held in Brussels and in the II Congreso de Eficiencia y Productividad (EFIUCO) held in Córdoba for their helpful comments and suggestions.

Sergio Perelman, CREPP, Université de Liège
Bd. Du Rectorat 7 (B31), B-4000 Liège, Belgium
Phone: 32 43663098 – Fax: 32 43663106 - E-mail address: sergio.perelman@ulg.ac.be

Daniel Santín, Departamento de Economía Aplicada VI, Universidad Complutense de Madrid
Campus de Somosaguas, 28223 Pozuelo de Alarcón, Madrid, Spain
Phone: 34 913942377 – Fax: 34 913942431 - E-mail address: dsantin@ccee.ucm.es

1. Introduction

Most of real-world problems have to deal with multi-output multi-input process. This is especially observed in the public sector where policy makers usually face up to multidimensional decisions and budget trade-offs. In the field of production frontier analysis, a number of non-parametric and parametric methods have been proposed to build best practice frontiers and to measure technical efficiencies across decision making units (DMU). In order to shed light on the properties and on the potential advantages and disadvantages of competing techniques and methodologies to conduct a same task, Monte-Carlo experimentation appears as the *statistical referee* often selected.

Bowlin *et al.* (1985), Banker *et al.* (1987), Gong and Sickles (1992), Banker *et al*. (1993) and Thanassoulis (1993) initiated the tradition comparing non parametric, mainly DEA (Data Envelopment Analysis) *vs*. parametric frontier performances. In more recent years several Monte-Carlo experiment papers concerned DEA issues: Pedraja-Chaparro *et al*. (1997) study the benefits of weights restrictions, Ruggiero (1998) and Yu (1998) analyze the introduction of non-discretionary inputs, Zhang and Bartels (1998) investigate the effect of sample size on mean efficiency scores, Holland and Lee (2002) measure the influence of random noise and Steinmann and Simar (2003) assess the comparability of estimated inter-group mean efficiencies. These are only a few examples, the complete list of Monte-Carlo works and experimental designs also includes sensitivity analysis, random noise and inefficiency terms distributions, functional forms, number of replications, and so on.

Our goal is not to present here a complete survey of these studies, nor of their main conclusions, but to make the observation that in our knowledge without exception these studies were performed in a single output multi-input framework and most used a Cobb-Douglas technology to generate the data. Nevertheless, in a seminal paper, Lovell *et al*. (1994) introduced a methodology that allows the estimation of a parametric production function in a multi-output multi-input setting. For this purpose, they used an output distance function and a *translog* technology.

However, if authors performing Monte-Carlo experiments neglected parametric distance functions and, more generally, *translog* technologies in generating testable production data, the reason must be found in the difficulties encountered to impose behavioral regularity conditions, mainly monotonicity and convexity constrains on them. In a recent published paper, O'Donnell and Coelli (2005) addressed this issue and propose a Bayesian approach to impose curvature on distance functions in empirical studies. The aim of this paper is close related with them. We illustrate how reliable data for Monte-Carlo experiments can be generated using parametric distance functions. Using a flexible *translog* technology we derive the sufficient conditions in order to generate data in the case of a simple two-input two-output production function, which maybe straightforward generalized to higher dimensionality problems. Moreover, we identify a valid range of parameters values satisfying regularity conditions and a *rule of thumb*, a data treatment recommendation, to be applied in empirical estimations.

The sections of the paper are organized as follows. Section 2 presents the main properties and characteristics of parametric output distance functions. In Section 3 we derive the sufficient conditions for the monotonicity and convexity properties to be fulfil and Section 4 illustrates how these conditions apply in the two-output two-input setting. Section 5 presents the data generation process, step by step and the last section points out the main conclusions and the directions for further research.

2. The *translog* output distance function

Defining a vector of inputs $x = (x_1, \ldots, x_K) \in \Re^{K+}$ and a vector of outputs $y = (y_1, \ldots, y_M) \in \Re^{M+}$ the feasible multi-input multi-output production technology can be defined using de output possibility set $P(x)$ which can be produced using the input vector x:

$P(x) = \{y: x$ can produce $y)$ that is assumed to satisfy the set of axioms depicted in Färe and Primont (1995). This technology can also be defined as the output distance function proposed by Shephard (1970):

$$D_O(x, y) = inf\{\theta : \theta > 0, (x, y/\theta) \in P(x)\}$$

If $D_O(x, y) \le 1$ then (x, y) belongs to the production set $P(x)$. In addition $D_O(x, y) = 1$ if y is located on the outer boundary of the output possibility set. Regularity conditions assume that $P(x)$ is non-decreasing, linearly homogeneous and convex in outputs, and non-decreasing and quasi-convex in inputs.

In order to estimate the distance function in a parametric setting a *translog* functional form is assumed. According with Coelli and Perelman (1999) this specification fulfils a set of desirable characteristics: flexible, easy to derive and allowing the imposition of homogeneity. The *translog* distance function specification herein adopted for the case of K inputs and M outputs is:

$$ln\, D_{Oi}(x, y) = \alpha_0 + \sum_{m=1}^{M} \alpha_m\, ln\, y_{mi} + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} \alpha_{mn}\, ln\, y_{mi}\, ln\, y_{ni} + \sum_{k=1}^{K} \beta_k\, ln\, x_{ki} \quad (1)$$

$$+ \frac{1}{2}\sum_{k=1}^{K}\sum_{l=1}^{K} \beta_{kl}\, ln\, x_{ki}\, ln\, x_{li} + \sum_{k=1}^{K}\sum_{m=1}^{M} \delta_{km}\, ln\, x_{ki}\, ln\, y_{mi}\,, \quad i = 1,2,\ldots, N,$$

where i denotes the i^{th} unit (DMU) in the sample. In order to obtain the production frontier surface we set $D_O(x, y) = 1$ which implies $ln\, D_O(x, y) = 0$. The parameters of the above distance function must satisfy a number of restrictions. Symmetry requires:

$\alpha_{mn} = \alpha_{nm}$; $m, n = 1, 2,\ldots, M$, and

$\beta_{kl} = \beta_{lk}$; $k, l = 1, 2,\ldots, K$.

Moreover, linear homogeneity of degree + 1 in outputs can be imposed, in order to fulfil Euler's Theorem, in the following way:

$$\sum_{m=1}^{M} \alpha_m = 1,$$

$$\sum_{n=1}^{M} \alpha_{mn} = 0, \; m = 1, 2,\ldots, M, \text{ and}$$

$$\sum_{m=1}^{M} \delta_{km} = 0, \; k = 1, 2,\ldots, K.$$

This latter restriction indicates that distances with respect to the boundary of the production set are measured by radial expansions. Following Shephard (1970) homogeneity in outputs implies:

$D_O(x, \omega y) = \omega D_O(x, y)$, for any $\omega > 0$,

and according with Lovell *et al.* (1994) normalizing the output distance function by one of the outputs is equivalent to set $\omega = 1/y_M$ imposing homogeneity of degree +1, as follows:

$$D_O(x, y/y_M) = D_O(x, y)/y_M\,.$$

For unit i, we can rewrite the above expression as:

$$ln\left(D_{Oi}(x, y)/y_{Mi}\right) = TL(x_i, y_i/y_{Mi}, \alpha, \beta, \delta), \qquad i = 1, 2,\ldots,N,$$

where

$$TL(x_i, y_i / y_{Mi}, \alpha, \beta, \delta) = \alpha_0 + \sum_{m=1}^{M-1} \alpha_m \, ln\left(y_{mi} / y_{Mi}\right) + \frac{1}{2} \sum_{m=1}^{M-1} \sum_{n=1}^{M-1} \alpha_{mn} \, ln\left(y_{mi} / y_{Mi}\right) ln\left(y_{ni} / y_{Mi}\right)$$

$$+ \sum_{k=1}^{K} \beta_k \, ln\, x_{ki} + \frac{1}{2} \sum_{k=1}^{K} \sum_{l=1}^{K} \beta_{kl} \, ln\, x_{ki} \, ln\, x_{li} + \frac{1}{2} \sum_{k=1}^{K} \sum_{m=1}^{M-1} \delta_{km} \, ln\, x_{ki} \, ln\left(y_{mi} / y_{Mi}\right). \qquad (2)$$

And rearranging terms the function above can be rewritten as follows:

$$-\ln\left(y_{Mi}\right) = TL(x_i, y_i / y_{Mi}, \alpha, \beta, \delta) - \ln D_{Oi}\left(x, y\right), \quad i = 1, 2, \ldots, N,$$

where $-ln\, D_{Oi}\left(x, y\right)$ corresponds to the radial distance function from the boundary. Hence we can set $u = -ln\, D_{Oi}\left(x, y\right)$ and add up a term v_i capturing for noise to obtain the Battese and Coelli (1988) version of the traditional stochastic frontier model proposed by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977):

$$-ln\left(y_{Mi}\right) = TL(x_i, y_i / y_{Mi}, \alpha, \beta, \delta) + \varepsilon_i, \quad \varepsilon_i = v_i + u_i,$$

where $u = -ln\, D_{Oi}\left(x, y\right)$, the distance to the boundary set, is a negative random term assumed to be independently distributed as truncations at zero of the $N\left(\varphi, \sigma_u^2\right)$ distribution, and the v_i term is assumed to be a two-sided random (stochastic) disturbance designated to account for statistical noise and distributed *iid* $N\left(0, \sigma_v^2\right)$. Both terms are independently distributed $\sigma_{uv} = 0$.

3. Regularity conditions in a *well-behaved* output distance function: a review

One serious drawback in applied production studies is that most of times the true technology is completely unknown, especially in service activities like education or health. However, we learn from microeconomic foundations that a *well-behaved* production function must fulfil a number of *smooth* properties. Färe and Primont (1995) provide the general regularity properties for output distance functions: monotonicity (non-decreasing in inputs), convexity and homogeneity of degree +1 in outputs, and non-increasing and quasi-convexity in inputs[1]. These technological constraints rely on economics theory but real circumstances or a legal framework could relax or even do more restrictive these assumptions. Regardless whether or not these properties are always true in real production situations, they impose desirable assumptions for experimental data generation design.

Following O'Donnell and Coelli (2005), monotonicity and curvature conditions involve constraints on distance function partial derivatives on equation (1) with respect to inputs:

$$s_k = \frac{\partial\, ln\, D}{\partial\, ln\, x_k} = \beta_k + \sum_{l=1}^{K} \beta_{kl} \, lnx_l + \sum_{m=1}^{M} \delta_{km} \, ln\, y_m,$$

and with respect to outputs:

$$r_m = \frac{\partial\, ln\, D}{\partial\, ln\, y_m} = \alpha_m + \sum_{n=1}^{M} \alpha_{mn} \, lny_n + \sum_{k=1}^{K} \delta_{km} \, ln\, x_k.$$

Monotonicity implies two conditions on distance function partial elasticities. For D to be non-increasing in x it is required that:

[1] O'Donnell and Coelli (2005, footnote 1) contribute to clarify a typographical error in Färe and Primont (1995).

$$f_k = \frac{\partial D}{\partial x_k} = \frac{\partial \ln D}{\partial \ln x_k}\frac{D}{x_k} = s_k \frac{D}{x_k} \leq 0 \Leftrightarrow s_k \leq 0,$$

while for D to be non-decreasing in y it is required that:

$$h_m = \frac{\partial D}{\partial y_m} = \frac{\partial \ln D}{\partial \ln y_m}\frac{D}{y_m} = r_m \frac{D}{y_m} \geq 0 \Leftrightarrow r_m \geq 0.$$

For quasi-convexity in x, it is necessary to evaluate the corresponding bordered Hessian matrix on inputs:

$$F = \begin{bmatrix} 0 & f_1 & \cdots & f_k \\ f_1 & f_{11} & \cdots & f_{1k} \\ \vdots & \vdots & \cdots & \vdots \\ f_k & f_{1k} & \cdots & f_{kk} \end{bmatrix},$$

where $f_{kl} = \dfrac{\partial^2 D}{\partial x_k \partial x_l} = \dfrac{\partial f_k}{\partial x_l} = \dfrac{\partial\left(s_k \dfrac{D}{x_k}\right)}{\partial x_l} = \left(\beta_{kl} + s_k s_j - \delta_{kl} s_k\right)\left(\dfrac{D}{x_p x_j}\right),$

with $\delta_{kl} = 1$ if $p = j$ and 0 otherwise. For D to be quasi-convex on x over the nonnegative orthant (the n-dimensional analogue of the nonnegative quadrant) a sufficient condition implies that all principal minors of F must be negative.

Finally, for convexity in y we evaluate the Hessian matrix on outputs:

$$H = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1M} \\ h_{12} & h_{22} & \cdots & h_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ h_{1M} & h_{2M} & \cdots & h_{MM} \end{bmatrix},$$

where $h_{mn} = \dfrac{\partial^2 D}{\partial y_m \partial y_n} = \dfrac{\partial h_m}{\partial y_n} = \dfrac{\partial\left(r_m \dfrac{D}{y_m}\right)}{\partial y_n} = \left(\alpha_{mn} + r_m r_n - \delta_{mn} r_m\right)\left(\dfrac{D}{y_m y_n}\right).$

According with Lau (1978) the function D will be convex in y over the nonnegative orthant if and only if H is positive semi-definite. Thus, D will be convex in y if and only if all the principal minors of H are non-negative.

4. Sufficient conditions to generate regular data in a two-input two-output setting

In this section we are concerned with the generation of data in the simplest two-input two-output case, for which the output distance function can be defined as follows:

$$\ln D_O = \alpha_0 + \alpha_1 \ln y_1 + \alpha_2 \ln y_2 + \frac{1}{2}\alpha_{11}(\ln y_1)^2 + \frac{1}{2}\alpha_{22}(\ln y_2)^2 + \frac{1}{2}\alpha_{12}\ln y_1 \ln y_2 +$$
$$\frac{1}{2}\alpha_{21}\ln y_2 \ln y_1 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \frac{1}{2}\beta_{11}(\ln x_1)^2 + \frac{1}{2}\beta_{22}(\ln x_2)^2 + \frac{1}{2}\beta_{12}\ln x_1 \ln x_2 +$$
$$\frac{1}{2}\beta_{21}\ln x_2 \ln x_1 + \gamma_{11}\ln x_1 \ln y_1 + \gamma_{12}\ln x_1 \ln y_2 + \gamma_{21}\ln x_2 \ln y_1 + \gamma_{11}\ln x_2 \ln y_2 \qquad (3)$$

For the sake of simplicity we will assume separability between inputs and outputs, restricting all γ parameters to be zero[2]. Moreover, homogeneity of degree + 1 requires: $\alpha_1 + \alpha_2 = 1$, $\alpha_{11} + \alpha_{22} + \alpha_{12} + \alpha_{21} = 0$, while symmetry $\alpha_{12} = \alpha_{21}$ and $\beta_{12} = \beta_{21}$.

Then the monotonicity conditions on inputs can be written as follows:

$$s_1 = \frac{\partial \ln D}{\partial \ln x_1} = \beta_1 + \beta_{11} \ln x_1 + \beta_{12} \ln x_2 \leq 0,$$

$$s_2 = \frac{\partial \ln D}{\partial \ln x_2} = \beta_2 + \beta_{22} \ln x_2 + \beta_{12} \ln x_1 \leq 0,$$

and the monoticity conditions on outputs:

$$r_1 = \frac{\partial \ln D}{\partial \ln y_1} = \alpha_1 + \alpha_{11} \ln y_1 + \alpha_{12} \ln y_2 \geq 0,$$

$$r_2 = \frac{\partial \ln D}{\partial \ln y_2} = \alpha_2 + \alpha_{22} \ln y_2 + \alpha_{21} \ln y_1 \geq 0.$$

In order to verify $s_1 \leq 0$ and $s_2 \leq 0$ a sufficient condition is to impose the negativity of all inputs parameters $(\beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12} < 0)$. Moreover, input values distribution must be restricted to be $x_i \geq 1$ $(\ln x_i \geq 0)$, where i = 1, 2 denotes the inputs.

Monotonicity in outputs must be imposed to fulfil convexity at the same time. Convexity on outputs over the nonnegative orthant requires that the Hessian matrix in outputs to be positive semi-definite. The two-output Hessian matrix and its components are as follows:

$$H = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix},$$

$$h_{12} = \frac{\partial^2 D}{\partial y_1 \partial y_2} = \frac{\partial (r_1 D/y_1)}{\partial y_2} = (\alpha_{12} + r_1 r_2)(D/y_1 y_2),$$

$$h_{21} = \frac{\partial^2 D}{\partial y_2 \partial y_1} = \frac{\partial (r_2 D/y_2)}{\partial y_1} = (\alpha_{21} + r_2 r_1)(D/y_2 y_1),$$

$$h_{11} = \frac{\partial^2 D}{\partial y_1 \partial y_1} = \frac{\partial (r_1 D/y_1)}{\partial y_1} = (\alpha_{11} + r_1 r_1 - r_1)(D/y_1 y_1),$$

$$h_{22} = \frac{\partial^2 D}{\partial y_2 \partial y_2} = \frac{\partial (r_2 D/y_2)}{\partial y_2} = (\alpha_{22} + r_2 r_2 - r_2)(D/y_2 y_2).$$

Accordingly with O'Donnell and Coelli (2005, p. 501), in the case of 2 outputs the Hessian matrix will be positive semidefinite *if and only if* $\alpha_{11} \geq r_1 r_2 (\leq 0.25)$. To fulfil with these conditions we impose[3] $\alpha_{11} = 0.25$ and, given the homogeneity condition on outputs,

[2] This restriction can be easily relaxed imposing new sufficient conditions between inputs and outputs values.

[3] There exist infinite possibilities to impose curvature conditions on outputs through the parameters. We only adopt one of these possibilities at start point in order to illustrate how to perform the data generation process.

$\alpha_{22} = 0.25$ and $\alpha_{12} = -0.25$. Once outputs interactions parameters are imposed, we restrict $r_1; r_2 \geq 0$ dealing with the ratio of outputs and the value of α_1 and α_2:

$$\alpha_1 + \alpha_2 = 1,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0,\ r_1 r_2 \leq 0.25$$

$$r_1 = \alpha_1 + 0.25 \ln y_1 - 0.25 \ln y_2 \geq 0 \quad \Rightarrow \quad \alpha_1 - 0.25(\ln y_2 - \ln y_1) \geq 0,$$

$$r_2 = \alpha_2 + 0.25 \ln y_2 - 0.25 \ln y_1 \geq 0 \quad \Rightarrow \quad \alpha_2 + 0.25(\ln y_2 - \ln y_1) \geq 0.$$

These conditions impose on one hand that α_1 and α_2 must belong to the interval [0; 1]. Otherwise the monotonicity condition is broken. Therefore, in order to fulfil monotonicity and convexity the difference between the outputs logarithms is determined by α_1 and α_2. Whatever α_1 and α_2 values, the imposition of $\alpha_{11} = 0.25$ forces up that the absolute difference between the logarithms of outputs to be 4 as maximum. To obtain the plausible output interval it is required to assign values for α_1 and α_2 and to calculate the difference of the logarithms of outputs as it shown above. We can illustrate this result with several examples:

1. If $\alpha_1 = 0.50$ and $\alpha_2 = 0.50$, then $(\ln y_2 - \ln y_1) \in [-2;2]$.
2. If $\alpha_1 = 01.00$ and $\alpha_2 = 0.00$, then $(\ln y_2 - \ln y_1) \in [0;4]$.
3. If $\alpha_1 = 0.00$ and $\alpha_2 = 1.00$, then $(\ln y_2 - \ln y_1) \in [-4;0]$.
4. If $\alpha_1 = 0.25$ and $\alpha_2 = 0.75$ then $(\ln y_2 - \ln y_1) \in [-3;1]$.

This constraint is meaningful for data generation because this rule imposes that the ratio of outputs must be exogenously defined as $(\ln y_2 - \ln y_1) \in [a;b]$ being *a* the lowest value and *b* the upper one.

Moreover, it is possible to proceed in the other way around. To generate in a first step plausible exogenous output ratio in logarithms $(ln\, y_2 - ln\, y_1)$ and in the second step derive the range of valid parameters. For example a [–1.5; 1.5] ratio interval implies that α_1 and α_2 could take any value among $\alpha_1 \geq 0.375$ and $\alpha_2 \leq 0.625$, given $\alpha_1 + \alpha_2 = 1$. At the limit, $\ln(y_2/y_1) = 0$ corresponding to equal output values allows choosing any non-negative α_1 and α_2 parameter values, given $\alpha_1 + \alpha_2 = 1$.

Finally, for quasi-convexity in inputs, we must calculate the input bordered Hessian of equation (3):

$$F_D = \begin{bmatrix} 0 & f_1 & f_2 \\ f_1 & f_{11} & f_{12} \\ f_2 & f_{21} & f_{22} \end{bmatrix}$$

where,

$$f_1 = \frac{\partial D}{\partial x_1} = s_1 (D/x_1),$$

$$f_2 = \frac{\partial D}{\partial x_2} s_2 (D/x_2),$$

$$f_{11} = \frac{\partial^2 D}{\partial x_1 \partial x_1} = (\beta_{11} + s_1 s_1 - s_1)(D/x_1 x_1),$$

$$f_{12} = f_{21} = \frac{\partial^2 D}{\partial x_1 \partial x_2} = (\beta_{12} + s_1 s_2)(D/x_1\, x_2),$$

$$f_{22} = \frac{\partial^2 D}{\partial x_2 \partial x_2} = (\beta_{22} + s_2 s_2 - s_2)(D/x_2 x_2).$$

For D to be quasi-convex in x over the non-negative orthant a sufficient condition is that all principal minors of F were negative. This can be done restricting $s_1; s_2 \le 0$, which can be satisfied imposing all beta parameters to be negative. Therefore, the first principal minor $|F_1| = -f_1^2$ will be by construction always negative. And the second principal minor must be negative as well:

$$|F_2| = f_1 f_{12} f_2 - f_1 f_1 f_{22} + f_2 f_1 f_{21} - f_2 f_{11} f_2 = 2 f_1 f_2 f_{12} - f_1^2 f_{22} - f_2^2 f_{11} < 0,$$

which is equivalent to:

$$2 s_1 s_2 \frac{D^3}{x_1^2 x_2^2}(\beta_{12} + s_1 s_2) - s_1^2 \frac{D^3}{x_1^2 x_2^2}(\beta_{22} + s_2^2 - s_2) - s_2^2 \frac{D^3}{x_1^2 x_2^2}(\beta_{11} + s_1^2 - s_1) < 0.$$

Operating this expression and multiplying by -1 to change the sign of the inequality, and for better understanding, we obtain:

$$\beta_{22} s_1^2 - s_2 s_1^2 + \beta_{11} s_2^2 - s_1 s_2^2 - 2\beta_{12} s_1 s_2 > 0$$

In this expression all terms with a negative coefficient as well as the $-2\beta_{12} s_1 s_2$ term are positive. Therefore it is easy to show that a sufficient condition is that all inputs logarithms values were greater than zero (input values greater or equal to one). In such a case $s_2 < \beta_{22}$ and $s_1 < \beta_{11}$, so positive values always compensate negative ones, and second principal minor will satisfy the negativity constraint.

5. Steps in experimental design to generate regular data

To carry out the experimental design, and once we define $\alpha_{11} = 0.25$ the first step is the selection of the meaningful distribution ratio of outputs y_2/y_1, and its logarithm. As we have pointed out before this exogenous ratio must fulfils that $\ln(y_2/y_1) = (\ln y_2 - \ln y_1) \in [a;b]$, where $|a-b| \le 4$. An extreme difference of 4 will impose the range of valid α_1 and α_2 values. In second step, a distribution of distance function values has to be defined into the interval $[1;\infty]$. Efficient units will receive $D = 1$ $(\ln D = 0)$ and the remaining will receive a value according with a distributional assumption. Third step consist in generating a distribution for the random noise v, for example through a $N(0, \sigma_v^2)$ distribution. Fourth, we generate an input distribution with the only restriction that $x_i \ge 1$ $(\ln x_i \ge 0)$, where i = 1, 2 denotes the inputs. Moreover all parameters multiplying inputs must be negative. Now if we choose as numeraire $\ln y_1$ we can calculate $-\ln y_1$ through:

$$-ln(y_1) = \alpha_0 + \alpha_1\, ln\left(\frac{y_2}{y_1}\right) + \frac{1}{2}\alpha_{11}\left[ln\left(\frac{y_2}{y_1}\right)\right]^2 + \beta_1\, ln\, x_1 + \beta_2\, ln\, x_2 + \frac{1}{2}\beta_{11}[ln x_1]^2$$

$$+ \frac{1}{2}\beta_{22}[ln\, x_2]^2 + \beta_{12}\, ln\, x_1\, ln\, x_2 - ln\, D + v \qquad (4)$$

Where the value of α_0 must be imposed with the restriction that $-\ln y_1 - \alpha_0 < 0$, in order to avoid negative production values. Note that in the expression above $\alpha_{11} = 0.5$ in order to fulfil the curvature condition on outputs. Finally, once $-\ln y_1$ is calculated it is straightforward to compute $ln\, y_1$ and $ln\, y_2$ and data generation process will be concluded[4].

Evolving from this *well-behaved* production function we can extract the required number of samples to perform Monte-Carlo experimentation in a multi-input multi-output setting. The proposed methodology can be straightforward generalized to more dimensions. For example, in the case of three outputs it will be necessary to exogenously generate two ratios of outputs, say $\ln(y_2/y_1)$and $\ln(y_3/y_1)$, analogously to the two-input two-output case discussed here, that will impose the range of output parameters values, and so on.

Conclusions and further research

In the field of frontier analysis, several studies compared the performances of non-parametric and parametric approaches appealing to Monte-Carlo experiments but always used a single output and most of them Cobb-Douglas technology for data generation. The main explanation why authors did not considered more flexible technologies, *e.g. translog*, and multi-output multi-input distance functions must be found among the difficulties encountered to generate data satisfying regularity conditions, like monoticity and convexity. In a recent paper, O'Donnell and Coelli (2005) addressed these regularity issues and proposed a Bayesian approach to be used in empirical estimations. The aim of this paper is close related with them. We show how parametric output distance functions allow the generation of random data on production process characterized by multi-output multi-input dimensions. Moreover we derive the necessary conditions under which second order *translog* technologies fulfil regularity conditions. We think that this set of rules provides a valid tool to improve the conclusions of methodological studies performing Monte-Carlo experiments in the field of frontier analysis.

Furthermore, we show how a valid range of output ratios and parameters can be derived that satisfies the regularity conditions. These values will be useful for practitioners as a *rule of thumb* in empirical studies dealing with the estimation of parametric output distance functions technologies. More research is still necessary in order to measure the potential bias introduced in Monte-Carlo experimentation based exclusively on single output data.

References

Aigner, D. J. Lovell, C. A. K. and Schmidt, P. (1977). Formulation and estimation of stochastic production function models, *Journal of Econometrics*, 6, 21-37.

Banker, R., Charnes, A., Cooper, W. and Maindiratta, A. (1987), A comparison of data envelopment analysis and translog estimates of production frontiers using simulated observations from a known technology. In Applications in Modern Production Theory Inefficiency and Productivity. (A. Dogramaci and R. Fare, Eds.), 33-55. Kluwer Academic Publishers, Boston.

Banker, R., Gadh, V. and Gorr, W. (1993). A Monte Carlo comparison of two production frontier estimation methods: corrected ordinary least squares and data envelopment analysis, *European Journal of Operational Research*, 67, 332-343.

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.

[4] Although it is not common in the Monte-Carlo literature on production frontiers, an inmediate extension of this analysis for *T* time periods is possible adding up to equation (4) a time trend.

Bowlin, W., Charnes, A., Cooper, W. and Sherman, H. (1985). Data envelopment analysis and regression approaches to efficiency estimation and evaluation. *Annals of Operational Research*, 2, 113-138.

Coelli, T. and Perelman, S. (1999). Technical efficiency of European railways: a distance function approach. *Applied Economics*, 32, 1967-1976.

Färe, R. and Primont, D. (1995). Multi-output production and duality: theory and applications. Kluwer Academic Publishers, Boston.

Gong, B. H., and Sickles, R. (1992), Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data, *Journal of Econometrics* 51, 259-284.

Holland, D. S. and Lee, S. T. (2002), Impacts of random noise and specification on estimates of capacith derived from data envelopment analysis, *European Journal of Operational Research*, 137, 10-21.

Lau, L. J. (1978), Testing and imposing monotonicity, convexity and quasi-convexity constraints. In Fuss, M., McFadden, D. (Eds.), Production Economics: a dual approach to theory and applications. Amsterdam, North Holland.

Lovell, C. A. K., Richardson, S., Travers, P. and Wood, L. L. (1994). Resources and functionings: a new view of inequality in Australia, in Models and Measurement of Welfare and Inequality (Ed.) W. Eichhorn, Springer-Verlag, Berlin, 787-807.

Mas-Colell, A., Whinston, M., and Green, J. (1995), Microeconomic Theory, Oxford University Press.

Meeusen, W. and van den Broeck, J. (1977). Efficiency Estimation from Cobb-Douglas Production Functions with composed error, *International Economic Review*, 18, pp. 435-444.

O'Donnell, C. J. and Coelli, T. J. (2005), A Bayesian approach to imposing curvature on distance functions, *Journal of Econometrics*, 126, 493-523.

Pedraja-Chaparro, F., Salinas-Jimenez, J. y Smith, P. (1997). On the role of weight restrictions in Data Envelopment Analysis, *Journal of Productivity Analysis*, 8, 215-230.

Ruggiero, J. (1998). Non-discretionary inputs in data envelopment analysis, *European Journal of Operational Research*, 111, 461-469.

Shepard, R. W. (1970), Theory of Cost and Production Functions, Princeton University Press, Princeton, NY.

Steinmann, L. and Simar, L. (2003), On the comparability of efficiency scores in nonparametric frontier models, Discussion Paper 0318, Institut de Statistique, Université Catholique de Louvain.

Thanassoulis, E. (1993), A comparison of regression analysis and data envelopment analysis as alternative methods for assessing performance, Journal of Operational Research Society, 44, 1129-1145.

Yu, C. (1998), The effects of exogenous variables in efficiency measurement. A Monte Carlo study, European Journal of Operational Research, 105, 569-580.

Zhang, X. (1999), A Monte Carlo study on the finite sample properties of the Gibbs sampling method for a stochastic frontier model, Journal of Productivity Analysis, 14, 71-83.

Zhang, Y., Bartels, R., (1998). The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia, Sweden and New Zealand, Journal of Productivity Analysis, 9, 187-204.