

Bayesian estimation of the half-normal regression model with deterministic frontier

FRANCISCO J. ORTEGA

Avenida Ramón y Cajal 1, Sevilla, Spain.

Universidad de Sevilla, Sevilla, Spain

Phone: +34 954556970

e-mail: fjortega@us.es

JOSE M. GAVILAN

Avenida Ramón y Cajal 1, Sevilla, Spain

Universidad de Sevilla, Sevilla, Spain

Phone: +34 954556970

e-mail: gavi@us.es

Bayesian estimation of the half-normal regression model with deterministic frontier

Abstract A regression model with deterministic frontier is considered. This type of model has hardly been studied, partly owing to the difficulty in the application of maximum likelihood estimation since this is a non-regular model. As an alternative, the Bayesian methodology is proposed and analysed. Through the Gibbs algorithm, the inference of the parameters of the model and of the individual efficiencies are relatively straightforward. The results of the simulations indicate that the utilized method performs reasonably well.

Keywords *Deterministic Frontier, Bayesian estimation, Gibbs algorithm.*

JEL Classification *C11, C15, D24.*

1. INTRODUCTION

The origin of the analysis of the econometric models with frontier and the calculation of efficiency measures can be set in Farrel (1957), where the innovative idea of analysing the efficiency of a productive process is established in terms of the observed deviations between the actual production and an ideal frontier of maximum output. Aigner and Chu (1968), following the initiative of Farrel (1957), propose a model where a given output is expressed as a function of a series of inputs and unknown parameters plus a negative random perturbation; the deterministic part of the model corresponds to the production frontier or maximum attainable value of output given the inputs, and the random perturbation (difference between the actual production and its potential maximum value) represents the level of inefficiency in the productive process. This formulation of the model is known as the deterministic frontier model (DFM)

After the appearance of the stochastic frontier models (SFM), formulated for the first time by Aigner et al. (1977), Battese and Corra (1977), and Meeusen and van den Broeck (1977), the interest of the researchers in the DFM quickly declined in favour of the SFM. In the latter, a composite error term is introduced in the form of two random perturbations. The first is an inefficiency measure and the other represents all the possible sources of random variations.

In the specific literature, there are a number of negative results concerning the DFM. It has been argued that this type of modelization does not take into account any source errors nor does it consider any other kind of random variations and therefore they are assigned to inefficiency in the productive process. Another drawback of this modelization is that it poses a non-regular problem, and therefore the properties of the maximum likelihood estimator (MLE) are uncertain. In Greene (1980), this latter disadvantage is partially solved: naming ε the inefficiency, it is established that if the density of ε and its derivative at the point $\varepsilon = 0$ are equal to zero, then the DFM verifies the usual conditions of regularity. Several distributions, such as the log-normal distribution, and the gamma distribution with shape parameter greater than or equal to two, fulfil the aforementioned conditions, while other commonly used distributions, such as the exponential and the half-normal distributions, fail to fulfil the conditions.

Nor is the SFM free of drawbacks. In many cases, it is difficult to disentangle the random effect and the inefficiency of the composite error, and there is a significant proportion of samples that involve one or the other extreme (absence of inefficiency or absence of random errors), even for large sample sizes (Simar, 2007; Ortega and Gavilan, 2011). Therefore, consideration of the DFM (or at least the comparison of the result with the SFM) can be especially interesting when the estimation of the SFM suggests the absence of random errors

and, as a consequence, that 100% of the composite error is due to inefficiency. As a matter of fact, the software to estimate the SFM in this situation usually provides some kind of warning or error message. To be specific, using the version 1.1-0 of the package *frontier* in the software R (Coelli and Henningsen, 2013) the following warning message is obtained¹:

```
the parameter 'gamma' is close to the boundary of the parameter space [0,1]: this can cause convergence problems and can negatively affect the validity and reliability of statistical tests and might be caused by model misspecification
```

In the same way, version 9 of the software LIMDEP gives the following error message, stops the estimation, and no result is provided:

```
Error 143: Models - estimated variance matrix of estimates is singular
```

In Amsler et al. (2013), the DFM is analysed in a number of situations, concluding that in some simple cases it is possible to make an exact inference using analytic expressions for the MLE for finite samples. In models of a more complicated nature, the exact inference remains possible, although simulation processes are necessary to obtain the critical points. This property is interesting, since it cannot be forgotten that, in the SFM, the properties of the MLE and the inference process in general are asymptotic, and therefore imply limitations when the sample size is small or moderate.

In this paper, the Bayesian estimation of the DFM through the Gibbs algorithm (Gelfand and Smith, 1990) is presented as an alternative. The Bayesian approach has been widely developed and utilized in the SFM, particularly since the publication of the paper by Van de Broeck et al. (1994). In Griffin and Steel (2007), a number of formulations of the SFM and their implementation in the software of Bayesian estimation WinBUGS are presented. However, to the best of our knowledge, the Bayesian literature has paid little attention to the DFM.

The Bayesian approach to the DFM holds interesting advantages, such as the ease of adding constraints involving both the parameters and the observations, which make the model non-regular, or the possibility of obtaining exact inferences for finite samples, not only for the parameters of the model but also for the individual efficiencies (although simulation processes are required to obtain the estimated values).

The rest of the paper is organized as follows: in Section 2, the model, and the prior and the posterior distributions, are presented. In Section 3, the Gibbs algorithm for the considered model is developed. In Section 4, the properties of the estimators are analysed through

¹ The parameter γ mentioned is the proportion of the total variance of the composite error due to inefficiency. Therefore, if it is estimated to be 1, then the whole error is inefficiency, that is to say, the model does not have random noise and its frontier is deterministic.

simulation, and in Section 5, two well-known examples are revisited. Finally, in Section 6, the main conclusions of the paper are discussed.

2. THE MODEL: PRIOR AND POSTERIOR DISTRIBUTIONS

Let us consider the following production model:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} - \varepsilon_i, \quad i = 1, \dots, n \quad (1)$$

where y_i is the production, $\boldsymbol{\beta} \in \mathbb{R}^k$ is a vector of unknown parameters (hence, k is the number of parameters to estimate in the production function), \mathbf{x}_i is the vector of exogenous variables (with $x_{i1} = 1 \forall i$, that is, the model has an intercept corresponding to the parameter β_1), n is the sample size, and $\varepsilon_i > 0$ is a random perturbation that measures the inefficiency of the i -th observation. The hypothesis that ε_i follows a half-normal distribution $(0, \sigma^2)$ is assumed (Johnson et al., 1994; Pewsey, 2004), that is, the probability density function of ε_i is given by:

$$f(\varepsilon_i) = \frac{2}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\varepsilon_i^2}{2\sigma^2}\right\}, \quad \varepsilon_i > 0 \quad (2)$$

and therefore, the probability density function of y_i is:

$$f(y_i | \mathbf{x}_i, \boldsymbol{\beta}, \sigma) = \frac{2}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} (y_i - \mathbf{x}_i' \boldsymbol{\beta})^2\right\}, \quad y_i \leq \mathbf{x}_i' \boldsymbol{\beta} \quad (3)$$

The likelihood function of the model for a sample of size n is given by

$$L(\boldsymbol{\beta}, \sigma | \mathbf{y}, \mathbf{X}) \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_i (y_i - \mathbf{x}_i' \boldsymbol{\beta})^2\right\}, \quad y_i \leq \mathbf{x}_i' \boldsymbol{\beta} \quad \forall i = 1, \dots, n \quad (4)$$

where $\mathbf{y} = (y_1, \dots, y_n)' \in \mathbb{R}^n$ and $\mathbf{X}' = [\mathbf{x}_1' | \dots | \mathbf{x}_n'] \in M_{k \times n}$.

In order to simplify the notation, the set $B = \{\boldsymbol{\beta} \in \mathbb{R}^k / y_i \leq \mathbf{x}_i' \boldsymbol{\beta} \quad \forall i = 1, \dots, n\}$ is defined, that is, B is the parametric space or set of values of the parameter $\boldsymbol{\beta}$ that is compatible with the observed set of data according to the established model. As a consequence, the constraints can be expressed as $\boldsymbol{\beta} \in B$. The logical hypothesis that the set B is not empty is also assumed.

Let us point out that the stochastic frontier production model is formulated as:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} - \varepsilon_i + u_i, \quad i=1, \dots, n \quad (5)$$

where the hypothesis of the model (1) holds and it is also assumed that u_i follows a Normal distribution $(0, \sigma_U^2)$, and that there is independence between the perturbations ε_i and u_i . In this setting, a number of values y_i could be above $\mathbf{x}_i' \boldsymbol{\beta}$ due to the noise perturbation u_i . In this way, the constraints $y_i \leq \mathbf{x}_i' \boldsymbol{\beta}$, $i=1, \dots, n$ are removed, thus resulting in a likelihood without constraints.

As mentioned in Section 1, the interest is focused on the deterministic frontier model. The set of unknown parameters of the model (1) is $(\boldsymbol{\beta}, \sigma)$, where $\boldsymbol{\beta} \in B$ and $\sigma > 0$. The Bayesian inference is based on the determination of the joint posterior distribution of the parametric vector $(\boldsymbol{\beta}, \sigma)$ given the set of data, such posterior distribution is denoted by $\pi(\boldsymbol{\beta}, \sigma | \mathbf{y}, \mathbf{X})$. In order to obtain the posterior distribution it is necessary, first of all, to choose a joint prior distribution for the vector of parameters $(\boldsymbol{\beta}, \sigma)$; such prior distribution is denoted by $\pi(\boldsymbol{\beta}, \sigma)$.

In the setting of the Bayesian analysis of the linear regression model, a normal prior distribution for $\boldsymbol{\beta}$ and inverse gamma for σ^2 are frequently chosen; this Gaussian-inverse gamma distribution has the advantage of being conjugate, that is to say, the posterior distribution is also Gaussian-inverse gamma, which facilitates the inference process. However, this distribution is informative, because the hyperparameters must be chosen based on prior knowledge of the problem.

In this paper, a non-informative prior distribution has been chosen, which eliminates the need for choosing the hyperparameters and represents the cases in which “little is known about the values of $\boldsymbol{\beta}$ and σ^2 ”. A number of methods to obtain non-informative prior distributions have been proposed. In the linear regression model, the non-informative prior distribution most commonly used (Jeffreys, 1961, p.182; Box and Tiao, 1973, p.117) is:

$$\pi(\boldsymbol{\beta}, \sigma) \propto \sigma^{-1}. \quad (6)$$

In Zellner (1986), an excellent review of the linear regression model with various prior distributions is offered.

The kernel of the joint posterior distribution is calculated by multiplying the likelihood function (4) by the prior distribution (6), thereby obtaining:

$$\pi(\boldsymbol{\beta}, \sigma | \mathbf{y}, \mathbf{X}) \propto (\sigma^2)^{-(n+1)/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_i (y_i - \mathbf{x}_i' \boldsymbol{\beta})^2 \right\}, \quad \boldsymbol{\beta} \in B \quad (7)$$

which can be expressed as

$$\pi(\boldsymbol{\beta}, \sigma | \mathbf{y}, \mathbf{X}) \propto (\sigma^2)^{-(n+1)/2} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}, \boldsymbol{\beta} \in B. \quad (8)$$

As may be seen, the use of this non-informative prior distribution also leads to a Gaussian-inverse gamma kernel for the posterior distribution. However, the algebraic manipulation of this distribution is not easy, since the constraint $\boldsymbol{\beta} \in B$ makes it difficult to ascertain the integration limits, and forces the marginal distribution of each component β_j to be defined piecewise. In order to avoid this complication and to give an answer to the inference problem, the Gibbs algorithm is utilized (Gelfand and Smith, 1990).

3. FORMULATION OF THE GIBBS ALGORITHM

As pointed out in Section 2, the kernel of the posterior distribution is Gaussian-inverse gamma. Therefore, $\sigma^{-2} | \boldsymbol{\beta}, \mathbf{y}, \mathbf{X}$ follows a gamma distribution and $\boldsymbol{\beta} | \sigma, \mathbf{X}, \mathbf{y}$ follows a normal k -dimensional distribution (in this case truncated to the constraint $\boldsymbol{\beta} \in B$).

Indeed, by taking into consideration that the joint posterior distribution (8) can be expressed as

$$\pi(\boldsymbol{\beta}, \sigma | \mathbf{y}, \mathbf{X}) \propto (\sigma^{-2})^{\frac{n+1}{2}} \exp\left\{-\sigma^{-2} \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}, \boldsymbol{\beta} \in B \quad (9)$$

it is straightforward to deduce that $\sigma^{-2} | \boldsymbol{\beta}, \mathbf{y}, \mathbf{X}$ follows a gamma distribution, specifically:

$$\sigma^{-2} | \boldsymbol{\beta}, \mathbf{y}, \mathbf{X} \sim Ga\left(\frac{n+3}{2}, \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \quad (10)$$

In order to obtain the distribution of $\boldsymbol{\beta} | \sigma, \mathbf{X}, \mathbf{y}$, let us consider $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ (the estimator and the least-squares residuals of the unconstrained linear regression model). By taking into consideration that

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \hat{\mathbf{u}}'\hat{\mathbf{u}} + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \quad (11)$$

the joint posterior distribution (8) can be written as:

$$\pi(\boldsymbol{\beta}, \sigma | \mathbf{y}, \mathbf{X}) \propto (\sigma^2)^{-(n+1)/2} \exp\left\{-\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{2\sigma^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\}, \boldsymbol{\beta} \in B \quad (12)$$

from where it is obtained that

$$\boldsymbol{\beta} | \sigma, \mathbf{X}, \mathbf{y} \sim N_k \left(\hat{\boldsymbol{\beta}}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right), \boldsymbol{\beta} \in B, \quad (13)$$

that is, a k-dimensional normal distribution truncated to the subset $B \subseteq \mathbb{R}^k$.

In principle, a simple algorithm can be formulated through the following straightforward procedure:

Let m be the sample step of the algorithm. Given the values $\boldsymbol{\beta}^m$ and $(\sigma^2)^m$, obtained in the m -th step, the generated values in the next step are given by:

1. Generate a value v^{m+1} of the distribution $\sigma^{-2} | \boldsymbol{\beta}^m, \mathbf{y}, \mathbf{X}$. Take $(\sigma^2)^{m+1} = 1/v^{m+1}$.
2. a) Generate a value $\boldsymbol{\beta}^*$ from the distribution $\boldsymbol{\beta} | (\sigma^2)^{m+1}, \mathbf{y}, \mathbf{X}$, that is, generate a value

$$\boldsymbol{\beta}^* \sim N_k \left(\hat{\boldsymbol{\beta}}, (\sigma^2)^{m+1} (\mathbf{X}'\mathbf{X})^{-1} \right).$$

- b) If $\boldsymbol{\beta}^* \in B$, then take $\boldsymbol{\beta}^{m+1} = \boldsymbol{\beta}^*$

If $\boldsymbol{\beta}^* \notin B$, then return to step 2.a)

However, we have found that, in practice, the probability of rejection in this procedure is frequently so high that the method becomes unviable.

In order to overcome this major difficulty, one solution is to apply the Gibbs algorithm by obtaining samples from the one-dimensional conditional distributions, that is, from $\sigma | \boldsymbol{\beta}, \mathbf{y}, \mathbf{X}$ and $\beta_j | \boldsymbol{\beta}_{(j)}, \sigma, \mathbf{y}, \mathbf{X}$, $j=1, \dots, k$, where $\boldsymbol{\beta}_{(j)} = (\beta_1, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_k) \in \mathbb{R}^{k-1}$.

Therefore the Gibbs algorithm for this model is formulated as follows: Let m be the sample step of the algorithm. Given the values $\boldsymbol{\beta}^m$ and $(\sigma^2)^m$, obtained in the m -th step, the values generated in the next step are given by:

1. Generate a value v^{m+1} from the distribution $\sigma^{-2} | \boldsymbol{\beta}^m, \mathbf{y}, \mathbf{X}$. Take $(\sigma^2)^{m+1} = 1/v^{m+1}$.

2. FOR j=1 TO k

Generate a value β_j^{m+1} from the distribution

$$\beta_j | \beta_1^{m+1}, \dots, \beta_{j-1}^{m+1}, \beta_{j+1}^m, \dots, \beta_k^m, (\sigma^2)^{m+1}, \mathbf{y}, \mathbf{X}$$

NEXT j

3. Take $\boldsymbol{\beta}^{m+1} = (\beta_1^{m+1}, \dots, \beta_k^{m+1})'$.

In order to obtain simulated samples for $\beta_j | \boldsymbol{\beta}_{(j)}, \boldsymbol{\sigma}, \mathbf{y}, \mathbf{X}$, by rearranging the parameters, the following partition can be considered:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_j \\ \boldsymbol{\beta}_{(j)} \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_j^2 & \boldsymbol{\sigma}'_{j(j)} \\ \boldsymbol{\sigma}_{j(j)} & \boldsymbol{\Sigma}_{j(j)} \end{pmatrix}, \quad (14)$$

where $\boldsymbol{\Sigma} = \boldsymbol{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$. If there are no constraints in the model, it is known that

$$\beta_j | \boldsymbol{\beta}_{(j)}, \boldsymbol{\sigma}, \mathbf{y}, \mathbf{X} \sim N\left(\hat{\beta}_j + \boldsymbol{\Sigma}_{j(j)}^{-1} (\boldsymbol{\beta}_{(j)} - \hat{\boldsymbol{\beta}}_{(j)}), \sigma_j^2 - \boldsymbol{\sigma}'_{j(j)} \boldsymbol{\Sigma}_{j(j)}^{-1} \boldsymbol{\sigma}_{j(j)}\right). \quad (15)$$

For the sake of simplicity, let us assume initially that all the explanatory variables take non-negative values. Hence, the constraints $y_i \leq \mathbf{x}_i' \boldsymbol{\beta}$, $i = 1, \dots, n$ are equivalent to

$$\beta_j \geq \frac{y_i - \mathbf{x}_{i(j)}' \boldsymbol{\beta}_{(j)}}{x_{ij}}, i = 1, \dots, n \Leftrightarrow \beta_j \geq \max_i \left\{ \frac{y_i - \mathbf{x}_{i(j)}' \boldsymbol{\beta}_{(j)}}{x_{ij}} \right\} \Leftrightarrow \beta_j \geq b(\boldsymbol{\beta}_{(j)}, \mathbf{X}, \mathbf{y}), \quad (16)$$

where $b(\boldsymbol{\beta}_{(j)}, \mathbf{X}, \mathbf{y}) = \max_i \left\{ (y_i - \mathbf{x}_{i(j)}' \boldsymbol{\beta}_{(j)}) / x_{ij} \right\}$. (Note that if some x_{ij} are zero, then the i -th observation does not take part in the set of constraints that the parameter β_j must satisfy). In this way, it holds that

$$\beta_j | \boldsymbol{\beta}_{(j)}, \boldsymbol{\sigma}, \mathbf{y}, \mathbf{X} \sim N\left(\hat{\beta}_j + \boldsymbol{\Sigma}_{j(j)}^{-1} (\boldsymbol{\beta}_{(j)} - \hat{\boldsymbol{\beta}}_{(j)}), \sigma_j^2 - \boldsymbol{\sigma}'_{j(j)} \boldsymbol{\Sigma}_{j(j)}^{-1} \boldsymbol{\sigma}_{j(j)}\right), \beta_j \geq b(\boldsymbol{\beta}_{(j)}, \mathbf{X}, \mathbf{y}), \quad (17)$$

that is, the one-dimensional conditional distributions are normal truncated to the interval $[b(\boldsymbol{\beta}_{(j)}, \mathbf{X}, \mathbf{y}), +\infty)$. In practice, it is frequent for this interval to have a very small probability, which renders the simulation by the acceptance-rejection method unfeasible. However, the simulation of values from this distribution can easily be performed by following the algorithm by Marsaglia, which is explained in Devroye (1986), p.380.

Specifically, this algorithm generates values from a distribution $Z \sim N(0,1)$ constrained to the range $z > a$, where $a > 0$. The algorithm is as follows:

1. Generate U, V independent values from a distribution $U(0,1)$.
2. Calculate $Z = \sqrt{a^2 - 2\text{Log}(U)}$.
3. If $V \cdot Z > a$ return to 1. If $V \cdot Z \leq a$, take the value Z as result.

To apply this algorithm to our case, it is sufficient to take Z in each step as the typified distribution of $\beta_j | \beta_{(j)}, \sigma, \mathbf{y}, \mathbf{X}$, generate a value of Z , and then undo the typification.

In the framework of the production models, the variables are commonly measured in logarithmic terms and, as a consequence, negative values for the variables often appear.

If the hypothesis concerning the non-negativity of the explanatory variables is removed, then the conditional distribution is similar, but now the range of the parameter is given by $b_1(\beta_{(j)}, \mathbf{X}, \mathbf{y}) \leq \beta_j \leq b_2(\beta_{(j)}, \mathbf{X}, \mathbf{y})$, where

$$\begin{aligned} b_1(\beta_{(j)}, \mathbf{X}, \mathbf{y}) &= \max_{i | x_{ij} > 0} \left\{ (y_i - \mathbf{x}'_{i(j)} \beta_{(j)}) / x_{ij} \right\} \\ b_2(\beta_{(j)}, \mathbf{X}, \mathbf{y}) &= \min_{i | x_{ij} < 0} \left\{ (y_i - \mathbf{x}'_{i(j)} \beta_{(j)}) / x_{ij} \right\}. \end{aligned} \quad (18)$$

If the set of indices $\{i | x_{ij} > 0\}$ is empty, $b_1(\beta_{(j)}, \mathbf{X}, \mathbf{y}) = -\infty$ is taken (similarly, if $\{i | x_{ij} < 0\} = \emptyset$ then $b_2(\beta_{(j)}, \mathbf{X}, \mathbf{y}) = +\infty$). In this way, when generating the values of the one-dimensional conditional distributions $\beta_j | \beta_{(j)}, \sigma, \mathbf{y}, \mathbf{X}$, a number of different situations appear depending on the values of $b_1(\beta_{(j)}, \mathbf{X}, \mathbf{y})$ and $b_2(\beta_{(j)}, \mathbf{X}, \mathbf{y})$. This entails additional complexity when programming the Gibbs algorithm, which can be solved, according to the cases, by utilizing the Marsaglia algorithm or applying the acceptance-rejection method.

As initial value β^0 , the corrected least-squares estimator is used (Green, 1980; Ortega et al., 2009), which consists of modifying one of the components of the unconstrained least-squares estimator so that the obtained point belongs to the frontier of the feasible region. Specifically, the estimation of the intercept of the model is modified, by replacing the first component of $\hat{\beta}$ for $\hat{\beta}_1^* = \max_i \{y_i - \mathbf{x}'_{i(1)} \hat{\beta}_{(1)}\}$.

4. SOME SIMULATIONS

First of all, a number of simple simulated examples are carried out, which helps towards ascertaining the formulation of the model and offers a first evaluation of the proposed estimation method. Despite the fact that the analysed model presents some complications, the Bayesian estimation with the help of the Gibbs algorithm offer a solution that is relatively straightforward to put into practice and, as will be seen, in general provide satisfactory results.

Example 1. In order to better understand the nature of the problem and the complications of dealing with constraints involving both the parameters and the observations, we begin by considering a simple model, in which there is only one explanatory variable (plus an intercept), and a simulated sample of size $n=5$. Therefore, the regression model is $y_i = \beta_1 + \beta_2 x_i - \varepsilon_i$, $i=1, \dots, n$, where the hypotheses mentioned in Section 2 are assumed. The values $\beta_1 = \beta_2 = \sigma^2 = 1$ are chosen. For the variable \mathbf{x} , a sample of size 5 from a uniform model in $[0,1]$ is simulated. From these values, the corresponding sample for the variable \mathbf{y} is simulated. The resulting sample is shown in the Appendix.

The constraints that the considered model imposes are $y_i \leq \beta_1 + \beta_2 x_i$, $i=1, \dots, n$ or, equivalently, $\beta_1 \geq y_i - \beta_2 x_i$, $i=1, \dots, n$, where in general there will be redundant constraints.

For the obtained sample, the ordinary least-squares estimator is $\hat{\beta}_1 = 0.3106$ and $\hat{\beta}_2 = 0.7938$, which does not belong to the feasible region and that greatly departs from the actual value for the case of the intercept. As explained in Section 3, it holds that $\boldsymbol{\beta} | \sigma, \mathbf{X}, \mathbf{y} \sim N_2(\hat{\boldsymbol{\beta}}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$, $\boldsymbol{\beta} \in B$, where B is the set of points belonging to the feasible region. Therefore, the marginal posterior distribution of $\boldsymbol{\beta}$ concentrates a greater probability in the zone of the frontier of the feasible region nearer to the point given by $\hat{\boldsymbol{\beta}}$.

In Figure 1, a scatter plot corresponding to a simulated sample of size 2000 of the distribution of $\boldsymbol{\beta} | \sigma, \mathbf{X}, \mathbf{y}$ is offered, together with the straight lines that define the feasible region. Indeed, the points accumulate in the zone of greater probability. As mentioned earlier, some constraints can be redundant with some others (as is the case of the constraint in the bottom left-hand corner).

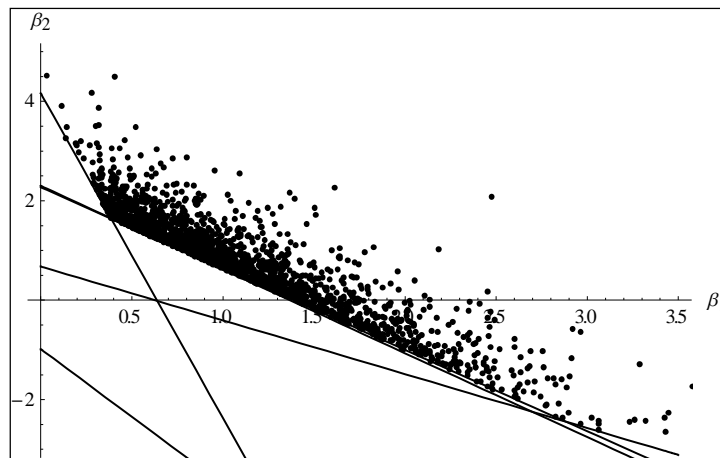


Fig. 1: Simulated sample of the posterior distribution (Example 1).

In order to estimate the values of the parameters, the criterion of the mathematical expectation of the posterior marginal distributions is used. Therefore, the average of each of the components of the simulated sample is taken (it is well-known that they converge towards the mathematical expectations as the size of the simulated sample tends towards infinity). In the same way, estimations of the standard errors are obtained from the standard deviations of the simulated sample and the symmetrical probabilistic intervals are obtained from the quantiles of order 0.025 and 0.975 of the simulated sample (Chen et al., 2000).

For the considered samples, the Bayesian estimation of the parameters is $\hat{\beta}_1^B = 1.0924$ and $\hat{\beta}_2^B = 0.8605$; the standard errors are $SE(\hat{\beta}_1^B) = 0.6036$ and $SE(\hat{\beta}_2^B) = 1.0433$, while the symmetric probabilistic (or credible) intervals are $SPI(\beta_1) = (0.3514, 2.5636)$ and $SPI(\beta_2) = (-1.5419, 2.5196)$. It should be emphasised that, in this case, the estimations are not very accurate because a very small sample size is considered.

As a consequence of the constraints involving the parameters and the observations, in this type of model the probability density function of the joint posterior distribution is not continuous at the frontier, and therefore certain graphical representations (such as the three-dimensional histogram and the contour plot) are not suitable, since they underestimate the true value of the density at the frontier. That is the reason why the scatter plot in Figure 1 has been shown. In order to have a better understanding of the posterior distribution, in Figure 2, the smoothed histograms of the marginal posterior distributions are presented. It is observed that the two distributions are skew and bell-shaped and have the highest density around the true values of the parameters.

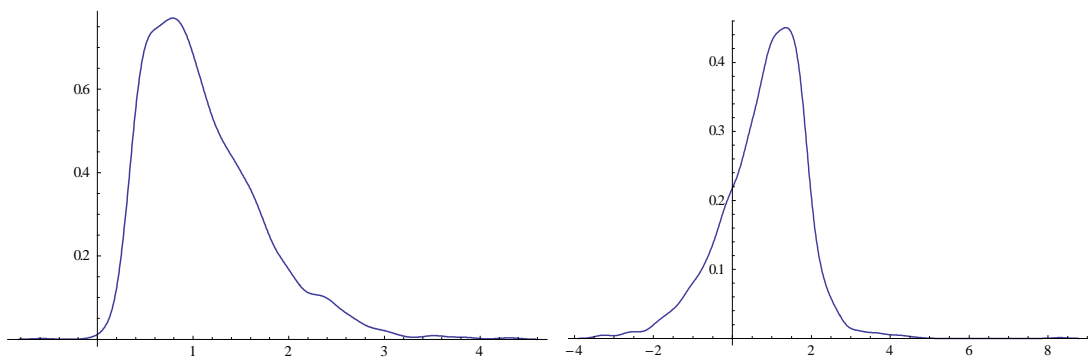


Fig. 2: Smooth Histogram of the posterior distribution of β_1 (left) and β_2 (right) of Example 1.

Example 2. A model with two explanatory variables and an intercept is now considered, for which a simulated sample of size $n = 50$ is obtained. Therefore, the regression model is $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} - \varepsilon_i$, $i = 1, \dots, n$, where the hypotheses mentioned in Section 2 are

assumed. Independent samples of size 50 from the uniform model in $[0,1]$ are simulated for the variables \mathbf{x}_2 and \mathbf{x}_3 . Taking $\beta_1 = \beta_2 = \beta_3 = \sigma^2 = 1$, the corresponding sample for the variable \mathbf{y} is simulated. The resulting sample is shown in the Appendix.

For the considered set of data, the ordinary least-squares estimator is the vector $\hat{\boldsymbol{\beta}} = (0.5632, 0.7837, 0.2729)'$, which does not belong to the feasible region and is far from the true value of the parameter, especially for the case of the intercept and the second slope. Through the Gibbs algorithm, a simulated sample of size 2000 for the distribution of $\boldsymbol{\beta} | \sigma, \mathbf{X}, \mathbf{y}$ is obtained, whose scatter plot is shown in Figure 3, where the zone of the feasible region with the highest probability can be observed.

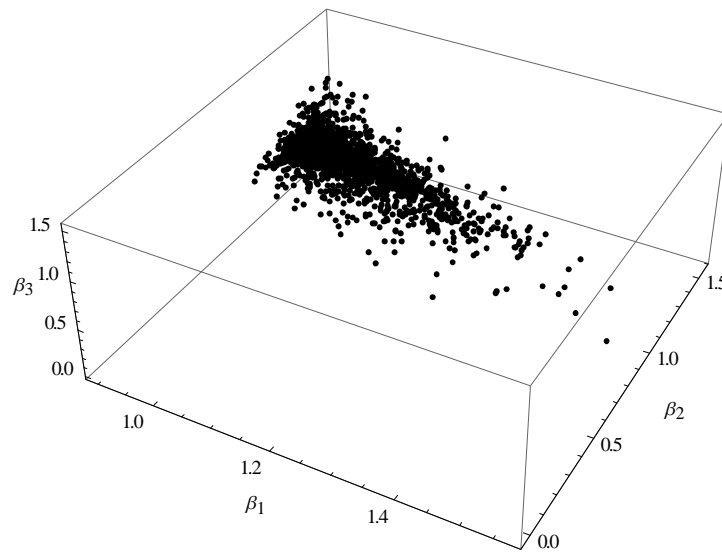


Fig. 3: Simulated sample of the posterior distribution (Example 2).

For the sample generated, the results obtained by applying the Bayesian DFM are shown in Table 1. As can be noted, for this sample of moderate size, the accuracy of the estimator is acceptable. In Figure 4, the smoothed histograms of the marginal posterior distributions are presented, where it is observed that the highest densities are located around the true values of the parameters. The results obtained by using maximum likelihood in the SFM are also shown in Table 1. Since the considered sample has been generated from a deterministic frontier model, an estimation of the parameter gamma equal (or near) to one is expected. However, the estimation obtained is, surprisingly, practically zero ($\hat{\gamma} = 0.000047$). Therefore, instead of identifying the deterministic frontier model, the SFM estimates, on the contrary, a stochastic frontier model without inefficiency, that is to say, all the composite error is assigned to noise. As a result of being unable to identify the true structure of the composite error of the model, the estimations of the parameters are very far from their true values. In this situation, all the

efficiencies are estimated to be practically 1 (the maximum attainable level of efficiency) in the SFM (all are above 0.995). For the case of the Bayesian DFM, the estimated average efficiency is 0.4587, which is near to the average of the true efficiencies, 0.4751. The average absolute of the differences between the true efficiencies and the efficiencies estimated through the Bayesian DFM is 0.018 (where the largest difference is 0.055), and their Pearson and Spearman rank coefficients of correlations are 0.9986 and 0.9973, respectively, again indicating reliable estimations. To conclude, in this example the SFM approach is unable to identify the true structure of the composite error and therefore provides unreliable estimations of the parameters and of the efficiencies. The Bayesian DFM approach, however, provides accurate estimations.

Table 1: Results of the Bayesian DFM and SFM estimations for the simulated sample of Example 2.

Parameter	Bayesian DFM			SFM	
	Estimation	Std. Error	SPI	Estimation	Std. Error
β_0	1.05907	0.090510	(0.948005, 1.29077)	0.566297	1.159324
β_1	1.04719	0.176417	(0.627511, 1.31936)	0.783662	0.267659
β_2	0.90686	0.194569	(0.529324, 1.31864)	0.272940	0.283047
σ^2	1.23846	0.272368	(0.822994, 1.86607)	0.311541	0.063564

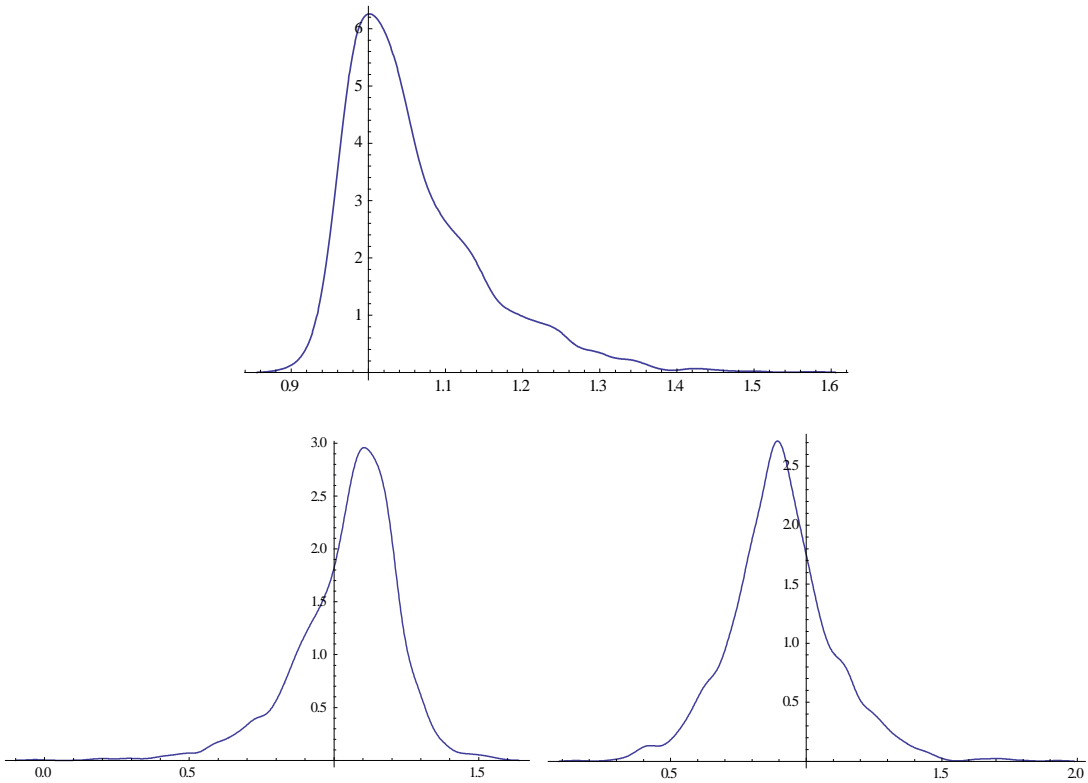


Fig. 4: Smooth Histogram of the posterior distribution of β_1 (upper), β_2 (bottom left) and β_3 (bottom right) of Example 2.

Similar examples, with a greater number of explanatory variables and with different starting values for the parameters, have been analysed. It has been observed that the accuracy of the estimators remains unaltered. Generally speaking, in the setting of the regression models with frontier (both deterministic and stochastic), the greatest difficulty is the estimation of the intercept; as a matter of fact, the ordinary least square estimators of the slopes are consistent (Green, 1980). The number of explanatory variables in the model and the values of the slopes are not determining factors in the statistical properties of the estimators of the intercept and the efficiencies. In this sense, in Coelli (1995), a simulation analysis for stochastic frontier models is carried out in which there is only one intercept and no explanatory variables, and it is argued that the main results remain unchanged when including explanatory variables. In Olsen et al. (1980), it is also observed that the inclusion of four explanatory variables in the stochastic frontier model provides very similar results to those of the model containing only the intercept.

The following is a Monte Carlo analysis, in which, for each of the chosen values of n , 1000 samples of size n are generated for the considered model. To these samples, the Gibbs algorithm with samples of size 2000 is applied, with the objective of studying the bias and the mean squared error (MSE) of the estimators, the length (Len) of the symmetrical intervals of probability 0.95 (SPI), and the proportion of success of such intervals (Succ), that is, the frequentist coverage probability of the symmetrical Bayesian intervals. For the case of the estimated efficiencies, their averages for each of the observations are obtained.

A model with one explanatory variable and an intercept is selected, together with the values $\beta_1 = \beta_2 = \sigma^2 = 1$. With the intention of analysing the possible influence of the dispersion in the set of data of the explanatory variables on the results, the procedure is carried out twice: the first time, the values for the variable \mathbf{x} are simulated from the model $U(0,1)$; while the second time, the model $U(0,5)$ is used instead. The results obtained are shown in Tables 2 and 3, respectively.

Table 2: Properties of the estimators and probabilistic intervals ($\beta_1 = \beta_2 = \sigma^2 = 1, x_i \sim U(0,1)$).

n	β_1					β_2				
	Bias	MSE	SPI	Len	Succ	Bias	MSE	SPI	Len	Succ
10	-0.0297	0.0746	(0.583,1.557)	0.974	0.912	0.0051	0.2736	(0.067,1.948)	1.881	0.904
20	-0.0074	0.0213	(0.779,1.337)	0.558	0.944	0.0047	0.0909	(0.429,1.621)	1.193	0.948
50	-0.0020	0.0060	(0.895,1.178)	0.283	0.948	0.0000	0.0173	(0.748,1.258)	0.509	0.944
100	0.0001	0.0019	(0.949,1.109)	0.159	0.948	-0.0003	0.0055	(0.865,1.134)	0.269	0.946
n	σ^2					Efficiencies (averages)				
	Bias	MSE	SPI	Len	Succ	Bias	MSE	SPI	Length	Succ
10	-0.1109	0.2383	(0.346,2.143)	1.797	0.856	0.0178	0.0113	(0.352,0.658)	0.306	0.908
20	-0.0529	0.1219	(0.489,1.768)	1.279	0.904	0.0041	0.0035	(0.410,0.599)	0.190	0.938

50	-0.0297	0.0419	(0.660,1.449)	0.790	0.928	0.0013	0.0008	(0.465,0.552)	0.088	0.951
100	-0.0099	0.0210	(0.737,1.297)	0.560	0.938	0.0008	0.0002	(0.495,0.541)	0.047	0.947

In relation to the bias, the values are, in general, close to 0 (with the exception of the parameter σ for $n=10$), and are smaller, in absolute values, as the sample size increases. Therefore, it can be concluded that the estimators have a good behaviour with respect to the estimation bias. In addition, it should be noted that the parameters β_1 and σ^2 present persistence of negative bias while β_2 and the efficiencies present positive bias in almost all the cases.

With regard to the MSE and the length of the intervals, reasonable values are obtained. In general, as expected, these two quantities decrease as the sample size increases.

On the other hand, it can be observed that the frequentist coverage probability of the Bayesian intervals is close to their nominal probability (0.95) for sample sizes of 20 or over except for the case of the parameter σ^2 , in which the values are maintained a little below the nominal value. For small sample sizes ($n=10$), the coverage probability is less than 0.95 in all cases.

The simulation analysis carried out also shows that the estimators of the parameters β_1 and β_2 , and of the efficiencies, have a better behaviour than the estimator of the parameter σ^2 .

Table 3: Properties of the estimators and probabilistic intervals ($\beta_1 = \beta_2 = \sigma^2 = 1$, $x_i \sim U(0,5)$)

n	β_1					β_2				
	Bias	MSE	SPI	Len	Succ	Bias	MSE	SPI	Len	Succ
10	-0.0289	0.0455	(0.553,1.292)	0.739	0.930	0.0005	0.0055	(0.807,1.081)	0.275	0.922
20	-0.0036	0.0283	(0.758,1.350)	0.591	0.934	-0.0015	0.0032	(0.888,1.097)	0.209	0.936
50	-0.0034	0.0054	(0.896,1.162)	0.266	0.952	-0.0001	0.0008	(0.952,1.050)	0.098	0.946
100	0.0005	0.0012	(0.951,1.081)	0.1306	0.950	-0.0004	0.0002	(0.973,1.026)	0.053	0.958
n	σ^2					Efficiencies (averages)				
	Bias	MSE	SPI	Len	Succ	Bias	MSE	SPI	Len	Succ
10	-0.1313	0.2406	(0.343,2.114)	1.771	0.842	0.0179	0.0106	(0.352,0.651)	0.299	0.926
20	-0.0832	0.1099	(0.488,1.725)	1.237	0.890	0.0052	0.0036	(0.410,0.597)	0.187	0.933
50	-0.0226	0.0431	(0.659,1.455)	0.7957	0.930	0.0022	0.0008	(0.466,0.554)	0.088	0.949
100	-0.0073	0.0197	(0.766,1.327)	0.561	0.948	0.0002	0.0002	(0.489,0.536)	0.047	0.947

The effect of the variability of \mathbf{x} is clearly manifest in the MSE and the length of the interval of the parameter β_2 ; as expected, these two quantities significantly decrease as the dispersion of the explanatory variable increases. It can also be observed that any increase in the variability of \mathbf{x} has no appreciable effect on the other parameters nor on the efficiencies.

Finally, in order to analyse the influence of the variance of the perturbation ε on the estimators, the simulation study has again been made selecting $\beta_1 = \beta_2 = 1$ and $\sigma^2 = 9$, and taking values for \mathbf{x} from distributions $U(0,1)$ and $U(0,5)$. The results for $\mathbf{x} \sim U(0,1)$ are shown in Table 4; similar results are obtained for $\mathbf{x} \sim U(0,5)$ and are therefore omitted.

Table 4: Properties of the estimators and probabilistic intervals ($\beta_1 = \beta_2 = \sigma^2 = 9$, $x_i \sim U(0,1)$).

n	β_1					β_2				
	Bias	MSE	SPI	Len	Succ	Bias	MSE	SPI	Len	Succ
10	-0.1318	0.8989	(-0.280,2.840)	3.120	0.906	0.034	2.764	(-1.820,3.876)	5.696	0.902
20	-0.0422	0.2653	(0.277,2.091)	1.814	0.924	0.0032	0.8939	(-0.693,2.721)	3.414	0.922
50	-0.0090	0.0528	(0.695,1.511)	0.816	0.930	0.0028	0.1559	(0.239,1.755)	1.516	0.942
100	-0.0030	0.0159	(0.844,1.267)	0.423	0.950	0.0049	0.0540	(0.607,1.396)	0.789	0.943
n	σ^2					Efficiencies (averages)				
	Bias	MSE	SPI	Len	Succ	Bias	MSE	SPI	Length	Succ
10	-1.3386	17.5966	(3.027,18.554)	15.527	0.824	0.0298	0.0226	(0.0824,0.4387)	0.3563	0.917
20	-0.7046	8.077	(4.365,15.547)	11.182	0.924	0.0108	0.0096	(0.1185,0.3487)	0.230	0.926
50	-0.1916	3.6670	(5.913,13.079)	7.166	0.928	0.0019	0.0023	(0.173,0.288)	0.115	0.947
100	-0.0889	1.7619	(6.732,11.772)	5.039	0.914	0.0001	0.0007	(0.202,0.265)	0.063	0.943

As expected, the increase of the variance of the perturbations has the effect of decreasing the accuracy of the estimators of the parameters β_1 , β_2 and σ^2 , and of making the bias, the MSE and the length of the probabilistic intervals larger. No influence on the coverage probabilities is observable.

For the case of the bias, since its absolute values are small, no clear factor of increase is visible. When analysing the parameters β_1 and β_2 , it can be appreciated that the length of the probabilistic intervals increases by an approximate factor of 3, while the MSE increases by an approximate factor of 9. These results are consistent with those for the ordinary least square estimator in the usual lineal regression model. The length of the probabilistic interval for σ^2 is approximately multiplied by 9 and its MSE by 81. In relation to the efficiencies, no significant variations are produced in the bias, nor in the MSE, nor in the length of the probabilistic intervals for the averages. Clearly, there is a decrease in the endpoints of such intervals, because when increasing σ , the average inefficiency increases and therefore the efficiency decreases.

5. TWO WELL-KNOWN EXAMPLES REVISTED

In this section, the estimation method analysed is applied to two well-known databases in the literature on the frontier production models. The first set of data, considered in Whiteman and Pearson (1993) and in Coelli et al. (1998), refers to telecommunication companies belonging to 21 countries in 1990. A single output y is considered, (which is an index based on incomes)

and two inputs: the capital factor x_1 (measured through the kilometres of the lines) and the labour factor x_2 (number of employees). The set of data is shown in the Appendix.

In this framework, we estimate, following the proposed methodology, a Cobb-Douglas production model, that is, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 - \varepsilon$, where the variables y , x_1 , x_2 are measured in logarithms and the perturbation ε follows a half-normal distribution $(0, \sigma^2)$, $\sigma > 0$.

The estimations of the parameters, together with their standard errors and symmetrical Bayesian intervals with probability 0.95, are shown in Table 5. As a starting point for the Gibbs algorithm, the unrestricted least-squares estimator is utilized, whose first component is modified so that it belongs to the frontier of the feasible region. A sample of size 10,000 from the joint posterior distribution is generated, from which the estimations offered in Table 5 are obtained. The results obtained in Coelli et al. (1998) are also shown in Table 5, where the parameters are estimated using maximum likelihood in a stochastic frontier production model. On page 198, they point out that “These results indicate that the vast majority of residual variation is due to the inefficiency effect, u_i , and that the random error, v_i , is approximately zero”, that is, that the frontier is actually deterministic. Therefore, the warning and error messages mentioned in the introduction are shown when using specific software to obtain the estimations of the model.

Table 5: Results of the estimation of the telecommunications model.

Parameter	Bayesian DFM			SFM	
	Estimation	Std. Error	Interval	Estimation	Std. Error
β_0	0.221224	0.088662	(0.098279, 0.429807)	0.170302	0.084008
β_1	0.923822	0.193276	(0.466299, 1.254850)	0.968427	0.216607
β_2	-0.075194	0.176745	(-0.402556, 0.322700)	-0.100169	0.197732
σ^2	0.376945	0.127413	(0.198190, 0.670106)	0.286561	0.077963

As can be observed, all the results achieved using SFM are similar to those obtained here by means of the Bayesian DFM. The largest discrepancy is observed in the parameter σ^2 , both in the estimation and in the standard error, where the absolute difference is approximately 0.09 for the estimation and 0.05 for the standard error. For the parametric vector β , the largest difference for the estimations is 0.05 (for the intercept) and for the standard error, this is 0.02 (for β_1).

In Table 6, the estimated efficiency is shown for each country, together with its standard errors and probabilistic intervals obtained using the Bayesian DFM. The estimated efficiencies calculated in Coelli et al. (1998) using the SFM are also shown. Only the estimated efficiencies are attained because, in principle, the maximum likelihood estimation of the SFM allows neither the standard errors nor the confidence intervals of the estimated efficiencies to be obtained.

Table 6: Efficiency of each of the countries.

Country	Bayesian DFM			SFM
	Efficiency	Std. Deviation	Interval	Efficiency
Australia	0.743777	0.072733	(0.575758, 0.859353)	0.784300
Austria	0.482098	0.026884	(0.410351, 0.511599)	0.510347
Belgium	0.610151	0.030610	(0.530624, 0.651875)	0.645905
Canada	0.685817	0.041381	(0.585526, 0.746132)	0.707590
Denmark	0.860641	0.047756	(0.742071, 0.924397)	0.915252
Finland	0.699179	0.047142	(0.594844, 0.773231)	0.747832
France	0.660550	0.061755	(0.536390, 0.790499)	0.668246
Germany	0.533938	0.043088	(0.426544, 0.606291)	0.543229
Iceland	0.702297	0.170060	(0.358880, 0.985397)	0.788019
Ireland	0.654088	0.106575	(0.429755, 0.825631)	0.715014
Italy	0.573716	0.051907	(0.474924, 0.684528)	0.582440
Japan	0.507945	0.063716	(0.395931, 0.651712)	0.504949
The Netherlands	0.796301	0.060814	(0.647231, 0.911924)	0.825264
New Zealand	0.664976	0.089423	(0.472417, 0.807720)	0.722120
Norway	0.784314	0.054093	(0.663300, 0.867537)	0.841189
Portugal	0.516446	0.051210	(0.404777, 0.597854)	0.555615
Spain	0.374454	0.022856	(0.321480, 0.413652)	0.386433
Sweden	0.876195	0.044082	(0.769552, 0.940367)	0.922833
Switzerland	0.947784	0.050564	(0.809997, 0.998708)	0.999471
Turkey	0.157315	0.009541	(0.132828, 0.172900)	0.163748
UK	0.914882	0.074745	(0.714045, 0.997120)	0.939083

With regard to the efficiencies, it is also observed that the results are very similar. The estimated average efficiency using the Bayesian DFM is 0.654613 while it is estimated to be 0.688994 when using the SFM. The estimated efficiencies using the Bayesian DFM are slightly lower than those obtained by utilising the SFM, however the average absolute difference is approximately 0.03, with the largest difference approximately 0.08 (for Iceland). Overall, the Pearson correlation coefficient for the estimations obtained for the two methodologies is 0.995364; and if solely the rank of the efficiencies is considered, then a Spearman rank correlation equal to 0.988312 is obtained. In both cases, the correlations are almost perfect.

The second set of data refers to 247 dairy farms located in Northern Spain, in the year 1997. The output variable y is given by the production of milk (in litres). Four inputs are considered: the total farm area x_1 (in hectares), the total labour, including hired and family labour x_2 (in worker equivalents), the total amount of feedstuffs fed to the dairy cows x_3 (in kilograms) and the average number of milking cows x_4 . Additional information about the variables can be found in Cuesta (2000) and Alvarez and Arias (2004).

In this setting, we estimate, according to the proposed methodology, a Cobb-Douglas production model, that is, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 - \varepsilon$, where the variables y, x_1, x_2, x_3, x_4 are measured in logarithms, and the perturbation ε follows a half-normal distribution $(0, \sigma^2)$, $\sigma > 0$.

The estimations of the parameters, together with their standard errors and symmetrical Bayesian intervals with probability 0.95 are shown in Table 7. The starting point for the Gibbs algorithm and the sample of size has been selected as in the previous example. The results obtained by using maximum likelihood in the SFM are also shown in Table 7.

Table 7: Results of the estimation of the Spanish dairy farm model.

Parameter	Bayesian DFM			SFM	
	Estimation	Std. Error	Interval	Estimation	Std. Error
β_0	5.59068	0.0992867	(5.31677, 5.71072)	5.0939209	0.1721667
β_1	0.0632154	0.0294959	(0.0226294, 0.12624)	0.0514371	0.0262519
β_2	0.0483018	0.0323639	(-0.0312353, 0.0962546)	0.0600841	0.0285625
β_3	0.413672	0.0157697	(0.393733, 0.447925)	0.4468382	0.0238105
β_4	0.55758	0.0157742	(0.522018, 0.581791)	0.5733713	0.0438262
σ^2	0.0884606	0.0083477	(0.0733932, 0.1055258)	0.0404724	0.0059082

As can be appreciated, the results are similar in the two approaches. However, the estimation of the parameter γ in the stochastic frontier setting is 0.8751, indicating that there is a stochastic frontier. This means that the Bayesian approach for the studied case of deterministic frontier is robust and performs reasonably well in the setting of a stochastic frontier with a high value of the parameter γ (the variance of the inefficiency term is much larger than the variance of the noise). For the estimations, the largest absolute difference between the two approaches is 0.4968 and is observed in the intercept; all the other absolute differences remain below 0.05. In relation to the standard errors, the largest absolute difference between the two approaches is 0.0729 and it is again observed in the intercept; all the other absolute differences remain below 0.03. With regard to the estimated efficiencies, the results are also similar. The estimated mean efficiency through DFM is 0.7756 and through SFM this is slightly higher, at 0.8657, due to the fact that, in the DFM framework, all the random error in the model is assigned to inefficiency while in the SFM case a little part of the error is random noise, and hence the global inefficiency is higher in the DFM framework, in other words, the global efficiency is lower. The average absolute difference between the estimated efficiencies in the two approaches is 0.0948, while the Pearson correlation is 0.8653 and the Spearman rank correlation is 0.8817, thereby indicating a high degree of concordance between the two approaches. In conclusion, in general, similar results are obtained, indicating the robustness of the Bayesian DFM approach. However,

as the frontier in this case is estimated to be stochastic, the differences between the two approaches are larger than in the previous example, in which the frontier is estimated to be deterministic.

6. CONCLUSIONS

The literature on frontier models has, in general, paid little attention to the DFM since the appearance of the SFM. However, the formulation using a deterministic frontier is simpler and enables exact results to be attained for the inference based on ML estimation in a number of cases (Amsler et al., 2013). The deterministic frontier model can be particularly useful in those situations where the effects of the inefficiency and the random noise in the SFM are complicated (and even “virtually impossible”) to disentangle, especially when the estimation of the SFM suggests that the entire variance of the composite error is due to inefficiency.

One of the main drawbacks of the DFM is that it poses a non-regular estimation problem (using the most common probability distributions for the inefficiency term), which renders the inference through maximum likelihood difficult. It has been shown that the formulation of the Gibbs algorithm for the estimation of the model using the Bayesian methodology is relatively straightforward and that it holds a number of advantages in comparison with other estimation methods, such as dealing with constraints involving both the parameters and the observations (which make the model non-regular), and the possibility of making inferences in a direct way concerning the individual efficiencies. Furthermore, as the kernel of the exact posterior distribution is used, the approximations to the exact results depend on the size of the simulated sample in the Gibbs algorithm, which can be as large as desired (in the maximum likelihood framework, the approximations depend on the sample size, which is fixed).

Both the simulation analysis and the examples considered show a good behaviour of the proposed methodology. For moderate sample sizes, the biases and the mean squared errors are considerably small and the frequentist coverage probability of the Bayesian intervals is practically identical to their nominal probability.

REFERENCES

- Aigner, DJ, Chu SF (1968) On estimating the industry production function. *Am Econ Rev* 58: 826-839.
- Aigner DJ, Lovell CA, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. *J Econom* 6: 21-37.
- Alvarez A, Arias C (2004) Technical efficiency and farm size: a conditional analysis. *Agr Econ* 30: 241-250.

- Amsler C, Leonard M, Schmidt P (2013) Estimation and inference in parametric deterministic frontier models. *J Prod Anal* 40: 293-305
- Battese GE, Corra GS (1977) Estimation of a production frontier model: With application to the Pastoral Zone of Eastern Australia. *Aust J Agr Econ* 21: 169-179.
- Box EP, Tiao GC (1973) *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Massachusetts.
- Chen MH, Shao QM, Ibrahim JG (2000) *Monte Carlo Methods in Bayesian Computation*, Springer-Verlag, New York.
- Coelli, T (1995) Estimators and Hypothesis Test for a Stochastic Frontier Function: A Monte Carlo Analysis, *J Prod Anal* 6: 247-268.
- Coelli T, Henningsen A (2013) *frontier: Stochastic Frontier Analysis*. R package version 1.1-0. <http://CRAN.R-Project.org/package=frontier> (last accessed 7 October 2015).
- Coelli T, Prasada Rao DS, Battese GE (1998) *An Introduction to Efficiency and Productivity Analysis*. Kluwer Academic Publishers, Boston.
- Cuesta, R (2000) A production model with firm-specific temporal variation in technical inefficiency: With application to Spanish dairy farms. *J Prod Anal* 13: 139-158.
- Devroye, L (1986) *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.
- Farrell MJ (1957) The measurement of productive efficiency. *J Roy Stat Soc A Sta* 120: 253-281.
- Gelfand AE, Smith AF (1990) Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc* 85: 398-409.
- Green WH (1980) Maximum Likelihood Estimations of econometric frontier functions. *J Econom* 13: 27-56.
- Griffin JE, Steel MFJ (2007) Bayesian stochastic frontier analysis using WinBUGS. *J Prod Anal* 27: 163-176.
- Jeffreys, H (1961) *Theory of Probability* (3rd ed). Oxford University Press, London.
- Johnson N, Kotz S, Balakrishnan N (1994) *Continuous Univariate Distributions* (2nd ed). Wiley, New York.
- Meeusen W, van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. *Int Econ Rev* 18: 435-444.
- Olsen JA, Schmidt P, Waldman DM (1980) A Monte Carlo Study of Estimators of the Stochastic Frontier Production Function. *J Econom* 13: 67-82
- Ortega, FJ, Basulto J, Camúñez JA (2009) Comparing Bayesian and corrected least-squares estimators in frontier production models. *Boletín de Estadística e Investigación Operativa* 25(2): 86-96.
- Ortega FJ, Gavilan JM (2011) Algunas observaciones acerca del uso de software en la estimación del modelo half-normal. *Rev Met Cuant* 11: 3-16.
- Pewsey A (2004) Improved likelihood based inference for the general half-normal distribution. *Commun Stat-Theor M* 33: 197-204.
- Simar L (2007) How to Improve the Performances of DEA/FDH Estimators in the Presence of Noise?. *J Prod Anal* 28: 183-201
- Van Den Broek J, Koop G, Osiewalski J, Steel MFJ (1994) Stochastic frontier models. A Bayesian perspective. *J Econom* 61: 273-303.
- Whiteman J, Pearson K (1993) *Benchmarking Telecommunications Using Data Envelopment Analysis*. *Aust Econ Pap* 12: 97-105.
- Zellner A (1986) On Assessing Prior Distributions and Bayesian Regression Analysis With g-Prior Distributions. Chapter 15 in *Bayesian Inference and Decision Techniques*, ed. By Goel P and Zellner A. Elsevier, Amsterdam.

APPENDIX: SETS OF DATA UTILIZED IN THE EXAMPLES

Simulated sample (Example 1)

y	1.389195	0.637083	-0.361800	1.365756	0.625770
x	0.611111	0.153082	0.367411	0.593594	0.923927

Simulated sample (Example 2)

y	{1.08249, 0.350275, 0.111249, 1.67043, 0.804949, 0.460018, 1.07821, 1.01255, 0.141129, 1.54748, 0.296863, 0.776595, 1.72269, 1.61751, 0.957871, 1.29077, 0.735434, 1.15074, 1.22122, 0.910826, 2.25897, 0.651734, 0.448724, 1.69737, 1.04723, 1.87568, 1.42077, 1.60624, \ 1.32974, 1.19284, 0.901887, 0.453288, 1.84041, 0.855025, 0.0904449, 1.05235, 1.58625, 1.49994, 0.544752, 2.65495, -0.155022, 1.82489, 1.24477, 0.711652, 1.66001, 0.570469, 0.21666, 1.00615, 2.26923, 1.40828}
x₂	{0.728113, 0.50603, 0.212999, 0.755097, 0.684416, 0.367649, 0.0441774, 0.934338, 0.340176, 0.572793, 0.350698, 0.893679, 0.195965, 0.876396, 0.80044, 0.680444, 0.0483495, 0.196176, 0.350175, 0.553006, 0.671351, 0.20208, 0.10168, 0.399454, 0.490328, 0.844086, 0.529784, 0.871228, 0.162341, 0.851186, 0.699436, 0.15938, 0.453054, 0.218776, 0.211959, 0.514752, 0.675255, 0.659948, 0.872841, 0.879089, 0.319521, 0.891303, 0.191057, 0.955342, 0.0475877, 0.149271, 0.117269, 0.822466, 0.83501, 0.201067}
x₃	{0.620443, 0.797728, 0.486157, 0.254354, 0.909706, 0.65939, 0.174122, 0.885507, 0.114798, 0.607201, 0.708227, 0.0685596, 0.613377, 0.802765, 0.40258, 0.0449246, 0.598802, 0.144266, 0.193644, 0.649677, 0.783038, 0.0725557, 0.662766, 0.4495, 0.202687, 0.614011, 0.990142, \ 0.370271, 0.771085, 0.0345378, 0.609937, 0.186131, 0.444118, 0.725772, 0.896521, 0.309074, 0.573408, 0.370633, 0.792472, 0.798635, 0.280654, 0.922334, 0.080111, 0.21696, 0.834997, 0.311379, 0.167264, 0.93436, 0.626333, 0.446917}

Set of data on the production and inputs in the example of telecommunications (Section 5).

Country	Output Index	Main lines (10 ⁶ km)	Number of Employees (10 ⁵)
Australia	0.74	0.7767	0.85
Austria	0.24	0.3223	0.18
Belgium	0.36	0.399	0.26
Canada	1.26	1.5296	1.05
Denmark	0.39	0.2911	0.18
Finland	0.29	0.267	0.2
France	2.06	2.8085	1.56
Germany	1.73	2.9981	2.12
Iceland	0.02	0.0126	0.02
Ireland	0.11	0.0983	0.13
Italy	1.48	2.2350	1.18
Japan	2.73	5.3236	2.77
The Netherlands	0.77	0.694	0.32
New Zealand	0.16	0.1473	0.17
Norway	0.27	0.2132	0.15
Portugal	0.19	0.2379	0.23

Spain	0.59	1.2603	0.75
Sweden	0.71	0.5849	0.42
Switzerland	0.56	0.3943	0.22
Turkey	0.15	0.6893	0.36
UK	2.53	2.5404	2.27