Semiparametric Bayesian Inference for Stochastic Frontier Models

J.E. Griffin and M.F.J. Steel*

Institute of Mathematics and Statistics
University of Kent at Canterbury, CT2 7NF, U.K.

Abstract

In this paper we propose a semiparametric Bayesian framework for the analysis of stochastic frontiers and efficiency measurement. The distribution of inefficiencies is modelled nonparametrically through a Dirichlet process prior. We suggest prior distributions and implement a Bayesian analysis through an efficient Markov chain Monte Carlo sampler, which allows us to deal with practically relevant sample sizes. We also allow for the efficiency distribution to vary with firm characteristics. The methodology is applied to a cost frontier, estimated from a panel data set on 382 U.S. hospitals.

Keywords: Dirichlet process, Efficiency measurement, Hospital cost frontiers, Markov chain Monte Carlo

JEL classification: C11; C14; C23

1 Introduction

Stochastic frontier models have been commonly used in the empirical study of firm efficiency and productivity, since the seminal work of Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977). A production frontier represents the maximum amount of output that can be obtained from a given level of inputs. Similarly, cost frontiers describe the minimum level of cost given a certain output level and certain input prices. In practice, the actual output of a firm will typically fall below the maximum that is technically possible. The deviation of actual from maximum output is a measure of inefficiency and is the focus of interest in many applications. These models, thus, typically combine

*Corresponding author: Mark Steel, Institute of Mathematics and Statistics, University of Kent at Canterbury, CT2 7NF, U.K. Tel.: +44-1227-82 7352; Fax: +44-1227-82 7932; Email: M.F.Steel@ukc.ac.uk.

two stochastic elements in the specification of the sampling model: one is a symmetric error term, corresponding to the usual measurement error, and another is the one-sided inefficiency. From the very beginning, the distribution to be used for the inefficiency error has been a source of contention. For example, Aigner *et al.* (1977) assume a half-Normal distribution, while Meeusen and van den Broeck (1977) adopt an exponential. Later proposals include the truncated Normal and the gamma distributions. Whereas van den Broeck, Koop, Osiewalski and Steel (1994) consider Bayes factors between these models in a Bayesian parametric analysis, we will here allow for nonparametric inefficiency distributions in a Bayesian framework. Given that the measurement of inefficiencies (or their transformations to efficiencies) is often a key objective in this area, the specification of the inefficiency distribution is a much more critical issue than in most other models with unobserved heterogeneity in the location (random effects), such as the usual Normal linear mixed models, where the random effects are often not of particular interest.

The main contribution of this paper is to define a Bayesian semiparametric model, based on the Dirichlet process, in the context of stochastic frontiers. We pay particular attention to the centring distribution of the Dirichlet process and the prior specification of its mass parameter. We use an improper prior on the frontier parameters and the measurement error variance, which facilitates the task of prior elicitation. In order to allow the efficiency distribution to vary with certain firm characteristics, we also consider a case where the inefficiency distributions are separately specified for groups of firms with the same characteristics. Links between these distributions are provided through the centring distribution (where a parametric dependence on firm characteristics is implemented) and the common mass parameter. We highlight some theoretical restrictions of parametric models, and design an efficient Markov chain Monte Carlo (MCMC) sampler, using a mixture of a centred and a non-centred parameterization. Finally, we conduct an application to a hospital cost frontier on the same data as used by Koop, Osiewalski and Steel (1997) in a parametric context.

The existing literature on semiparametric inference in the context of frontiers is entirely couched in a classical statistical framework. A first approach is that of Fan, Li and Weersink (1996) and Huang and Fu (1999), who advocate a nonparametric specification of the frontier, and adopt a parametric inefficiency distribution. The problem setting of Park and Simar (1994) is closer to the present paper, since they assume a parametric frontier and are nonparametric on the inefficiency distribution. A panel or longitudinal structure of the data is critical to their approach. Their results on estimators for efficiencies are asymptotic and require T, the number of time periods covered in the sample, to go to infinity. Their setup is extended in Park, Sickles and Simar (1998) to allow for dependence between inefficiencies and regressors. Applications of this approach in a multi-output setting can be found in Adams, Berger and Sickles (1999) and Sickles, Good and Getachew (2002).

Section 2 describes the class of models that we focus on while Section 3 provides details of the Dirichlet process used to implement a nonparametric inefficiency distribution. Some restrictions inherent in parametric inefficiency models are explained in Section 4, while Section 5 describes the MCMC sampler used for conducting inference. The sixth section applies the methodology developed in this paper to a set of hospital cost data, and Section 7 concludes.

Code (in C++) to implement the methods introduced in this paper can be obtained from the website http://www.ukc.ac.uk/IMS/statistics/people/M.F.Steel/software.html.

2 Stochastic Frontier Models

If we denote by $i = 1, \dots, n$ the firm index and use subscript $t = 1, \dots, T$ for the time period, the typical stochastic frontier model describes the logarithm of cost (or output[1]) y_{it} as

$$y_{it} = \alpha + x_{it}'\beta + u_i + v_{it}, \tag{1}$$

where x_{it} is a vector of appropriate explanatory variables[2], we have an i.i.d. symmetric error term v_{it} reflecting measurement and specification error

$$v_{it} \sim \mathrm{N}(0, \sigma^2),$$

and the main distinguishing feature of the model is the one-sided disturbance u_i, which measures inefficiency and is independently distributed as

$$u_i \sim F. \tag{2}$$

The sampling model further specifies independence between v_{it} and u_i. In order to exploit the longitudinal or panel structure of the data, we assume that the inefficiencies u_i are constant over time[3].

Whereas F is usually chosen to be some parametric family of distributions on $\Re_+$, we shall take F to be a random probability measure, generated by a Dirichlet process, as explained in the next section.

The sampling model above is parameterized by an intercept α, frontier regression parameters in the vector β and the variance of the measurement error σ^2. Fernández *et al.* (1997) examine the use of improper prior distributions for α, β and σ^2. They show that for panel data ($T > 1$) with inefficiencies that are constant over time, the posterior distribution exists under the prior

$$p(\alpha, \beta, \sigma^2) \propto \sigma^{-2}, \tag{3}$$

in combination with any proper inefficiency distribution. We shall adopt (3) for our semiparametric model and, thus, the posterior distribution will exist. Without any additional complexity, we can multiply the prior above by an indicator function imposing economic regularity conditions on the frontier. In case of additional prior information, the prior in (3) can also easily be replaced by a (conditionally) natural-conjugate Normal-inverted gamma prior, but the prior used here has the advantage of being invariant with respect to location and scale transformations of the data and does not require any prior elicitation effort. Conditionally on the u_i's, the prior adopted is the independence Jeffreys' prior. In summary, (3) is a convenient prior to use in the absence of strong prior information or as a "benchmark" prior.

For cross-section data ($T = 1$) or cases where inefficiencies are allowed to vary over time, Fernández *et al.* (1997) show that (3) does not lead to posterior existence and propose a slightly different improper prior that penalizes very small values of σ^2.

[1]In the case of an output frontier, the positive inefficiencies u_i appear with a negative sign in (1).

[2]The vector x_{it} will usually involve logarithms of outputs and prices for cost frontiers and logarithms of inputs for production frontiers.

[3]See Fernández, Osiewalski and Steel (1997) for a discussion of this issue and its implications for posterior existence.

3 The Dirichlet Process

The Dirichlet process is a commonly used nonparametric prior distribution which was introduced by Ferguson (1973). It is defined for a space Θ and a σ-field $\mathcal{B}$ of subsets of Θ. The process is parameterised in terms of a probability measure H on $(\Theta, \mathcal{B})$ and a positive scalar M. A random probability measure, F, on $(\Theta, \mathcal{B})$ follows a Dirichlet process DP(MH) if, for any finite measurable partition, $B_1, \ldots, B_k$ the vector $(F(B_1), \ldots, F(B_k))$ follows a Dirichlet distribution with parameters $(MH(B_1), \ldots, MH(B_k))$. This results in the following moment measures for $B \in \mathcal{B}$

$$\mathrm{E}\{F(B)\} = H(B), \tag{4}$$

$$\mathrm{Var}\{F(B)\} = \frac{H(B)\{1 - H(B)\}}{M+1}. \tag{5}$$

It is clear that H centres the process and M can be interpreted as a precision parameter. Another important property established by Ferguson (1973) is that the realisations of F are discrete distributions. Sethuraman (1994) introduced an alternative constructive representation of the Dirichlet process

$$F \stackrel{d}{=} \sum_{i=1}^{\infty} p_i \delta_{\theta_i}$$

where δ_x is the Dirac measure which places measure 1 on the point x, while $\theta_1, \theta_2, \ldots$ are i.i.d. realisations of H and $p_i = V_i \prod_{j<i}(1 - V_j)$ where V_i are i.i.d. Beta distributed with parameters $(1, M)$, *i.e.* $V_i \sim$ Be$(1, M)$. If we assume $y_1, \ldots, y_{n+1} \stackrel{i.i.d.}{\sim} F$ and F has a Dirichlet process prior then the predictive distribution is

$$F(y_{n+1}|y_1, \ldots, y_n) = \frac{1}{M+n} \sum_{i=1}^{n} \delta_{y_i} + \frac{M}{M+n} H, \tag{6}$$

as derived by Blackwell and MacQueen (1973).

The Dirichlet process prior is often incorporated into semiparametric models using the hierarchical framework

$$y_i \sim g(y_i|\theta)$$

$$\theta \sim F$$

$$F \sim \mathrm{DP}(MH), \tag{7}$$

where g is a probability density function. This model was introduced by Antoniak (1974) to produce continuous nonparametric distributions on the observables and is often referred to as a Mixture of Dirichlet Processes. The marginal distribution for y_i is a mixture of the g distribution. This basic model can be extended: the density g or the centring distribution H can be (further) parameterised and inference can be made about these parameters. In addition, inference can be made about the mass parameter M. The stochastic frontier model described in Section 2 fits into this general semiparametric framework by completing the specification of the model with this Dirichlet process prior distribution for F in (2).

In our context, possible choices for the centring distribution of H are gamma distributions. We take the class of Gamma distributions with fixed integer shape parameter[4] j and random scale (precision) parameter λ, *i.e* with mean j/λ, denoted by $\text{Ga}(j,\lambda)$. In the application, we shall focus in particular on the case where $j = 1$, *i.e.* H is exponential with mean $1/\lambda$, a very commonly used inefficiency distribution for parametric stochastic frontier modelling.

Finally, we extend the model specification to include prior distributions for the parameters of the centring distribution and M. In the following analyses the parameter λ is given an informative prior distribution which is elicited through the prior median efficiency, as explained in van den Broeck *et al.* (1994). In particular, if we define efficiency as $r_i = \exp(-u_i)$, and adopt the prior distribution

$$\lambda \sim \text{Ga}(j, -\ln r^\star), \tag{8}$$

then $r^\star$ is the implied prior median efficiency.

3.1 Prior distribution for M

There have been several methods suggested for choosing a prior distribution for M. Escobar and West (1995) suggest a gamma distribution for M where the parameters are elicited by considering the distribution of the number of distinct elements in the first n draws from the Dirichlet process. Walker and Mallick (1997) use the formula $M = \text{E}(\omega^2)/\text{Var}(\mu)$ where μ and ω^2 are the mean and variance of the unknown distribution. Eliciting these moments can be used to define the prior distribution. An alternative approach, followed here, is to interpret the parameter M of the Dirichlet process as a "prior sample size". This idea arises from the form of the predictive distribution in (6), where $M/(M+n)$ is the mass given to the centring distribution in the posterior predictive. Carota and Parmigiani (2002), for example, specify an informative prior distribution on this quantity. We define $\zeta \equiv M/(M+n_0)$ where n_0 is a hyperparameter. If we assume $\zeta \sim \text{Be}(a,b)$ then we obtain

$$p(M) = \frac{n_0^b\,\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{M^{a-1}}{(M+n_0)^{a+b}} \tag{9}$$

i.e. M follows an Inverted Beta distribution (see Zellner, 1971, p. 375) *a priori*. This prior implies

$$\text{E}(M) = \frac{n_0\,a}{b-1}, \qquad b > 1$$

$$\text{Var}(M) = \frac{n_0^2\,a(a+b-1)}{(b-1)^2(b-2)}, \qquad b > 2$$

$$\text{Mode}(M) = \frac{n_0(a-1)}{b+1}, \qquad a > 1.$$

[4] Such distributions are sometimes called "Erlang distributions".

If we choose $a = b = \eta$ then the prior median for M is n_0. From the expression of the variance it is clear that the latter (if it exists) is decreasing in η. This suggests a simple elicitation method where M ("prior sample size") is centred around n_0 and the prior distribution for M will become more concentrated as η increases.

To help in our choice of the hyperparameters, we can further explore the effect of M by considering the properties of the implied prior efficiency distribution. After all, the efficiencies (defined as $r_i = \exp(-u_i)$) are often the quantities of critical interest in this context. Under an exponential H and the prior in (8) with $r^\star = 0.8$, Figure 1 shows the prior predictive efficiency distribution. This is the prior mean efficiency distribution with the randomness in F and λ integrated out with their respective priors. Since the mean of F is always H, the value of M (or indeed, its prior) does not matter for Figure 1. Figure 2 displays the distributions of the prior measure assigned to certain efficiency intervals, for various values of M. We also include the parametric case where the prior inefficiency distribution equals the centring distribution H (we can view this as the case with $M = \infty$). It is important to stress that for finite M we are considering a random measure, so that the probabilities assigned to certain intervals are inherently random. An added source of randomness is the fact that we have assigned a prior to the parameter in H, namely λ. If we were to fix λ, the probabilities of intervals corresponding to the parametric case would reduce to fixed numbers, but that would not happen for the semiparametric cases. The probabilities for a given interval are all centred around the same quantity (corresponding to H, see (4)), but the variation clearly decreases with M in line with (5).

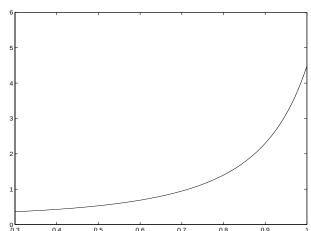
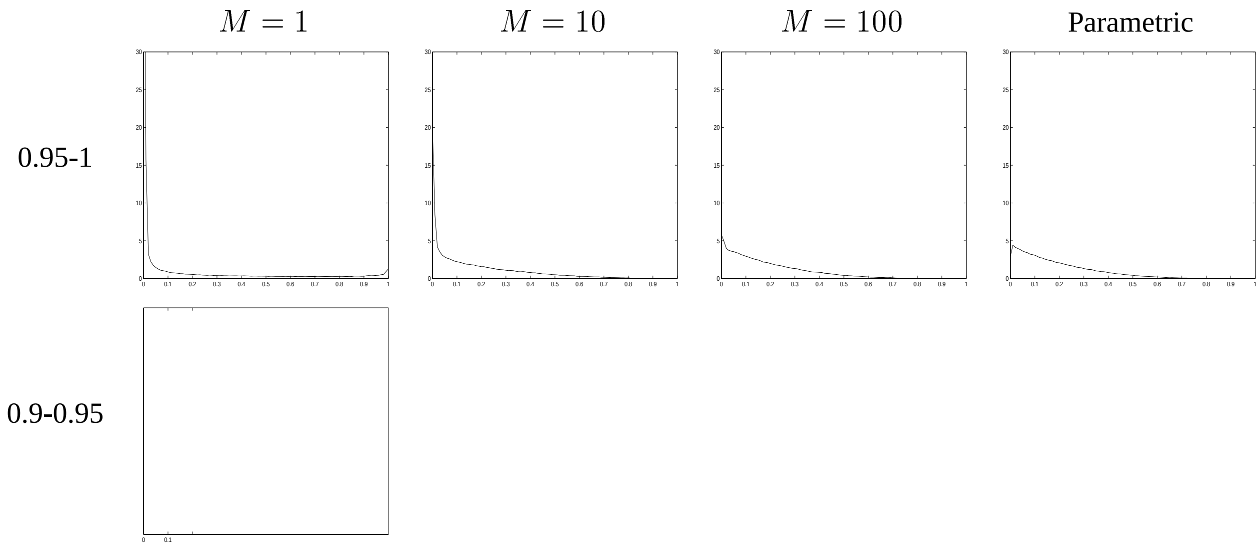


Figure 1: Prior predictive density function of efficiency

3.2 Accounting for firm characteristics

In many situations, we may be interested in making the inefficiency distribution, F, depend on certain firm characteristics. It can be quite reasonable to assume that groups of similar firms, *e.g.* defined through their size or ownership structure, have similar efficiencies but that the efficiency distribution varies between groups.

We describe the characteristics of the i-th firm by a vector w_i and we use the latter to define a family of distribution F_w. We restrict attention to the case where w_i can take a (small) finite number of possible values. The approach we follow here is similar to the one described in Carota and Parmigiani (2002), based on the idea of Products of Dirichlet Processes (Cifarelli and Regazzini, 1978). The random probability measure, F_w, is modelled using separate Dirichlet processes for each value of w_i. Dependence is



M = 1
M = 10
M = 100
Parametric
0.95-1
0.9-0.95

introduced through the centring distribution which is now assumed to be a function of the firm characteristics, H_w. A parametric model is used to specify H_w. The frontier parameters and the mass parameter M are assumed to be the same for all groups of firms. The use of a common parameter M allows borrowing of strength between groups and avoids problems in groups with small sample sizes.

Koop *et al.* (1997) introduce a parametric model for the mean of the inefficiency distribution by replacing the common parameter λ with the expression $\exp(w_i'\gamma)$ where the parameter vector γ indicates how the mean of the distribution changes with the firm characteristics in w_i. They use this idea in the context of an exponential distribution, but it can be immediately extended to a Gamma form for H. Thus, the inefficiency centring distribution for the i-th firm is now a Gamma distribution with mean $j\exp(-w_i'\gamma)$. It is useful to parameterize in terms of $\phi_l = \exp(\gamma_l)$, and we shall adopt the prior $p(\phi_l) = \text{Ga}(a_l, g_l)$, taking $a_1 = j$, $g_1 = -\ln r^\star$ and $a_l = g_l = 1, l > 1$ which leads to prior efficiency results close to that of the model with common efficiency distribution.[5]

4 Restrictions of the parametric model

Besides the obvious lack of flexibility associated with a fully parametric model, Figure 2 also illustrates another interesting fact. The probabilities of areas under the parametric efficiency distribution (implied by an exponential distribution H on the inefficiency) are restricted. In particular, the probabilities are always strictly less than one if the intervals do not have 0 or 1 as one of the endpoints. It is easy to derive that the probability of the interval (r_l, r_u) where $0 < r_l < r_u < 1$ is maximized by taking

$$\lambda_{max} = \frac{\ln\left(\frac{\ln r_l}{\ln r_u}\right)}{\ln r_u - \ln r_l}$$

and $P(r_l < r < r_u) \leq r_u^{\lambda_{max}} - r_l^{\lambda_{max}}$. For example, the upper bound on the prior efficiency probability in the interval $(0.8, 0.9)$ is 0.26975 in the parametric case (which is illustrated in Figure 2). This shows an important lack of flexibility of the parametric model with an exponential inefficiency distribution.

Thus, one could argue in favour of the general Gamma (or Erlang) inefficiency distribution, where such restrictions do not occur. This, however, can lead to a different type of problem. If we assume that $u_i \sim \text{Ga}(j, \lambda)$, the resulting efficiency distribution has an internal mode on $(0, 1)$ if $j, \lambda > 1$, given by

$$\text{Mode}(r_i) = \exp\left\{\frac{1-j}{\lambda - 1}\right\}$$

which links j and λ once the efficiency mode is fixed. In the application, the general shape for the efficiency distribution provided by the semiparametric model indicates a mode around 0.7 and a moderately narrow spread. Capturing these features requires values like $j = 8$ and $\lambda = 20.63$. However, a parametric model with $u_i \sim \text{Ga}(8, \lambda)$ seems inestimable as the latter inefficiency distribution is virtually indistinguishable from a Normal. Thus, the model would not allow us to separately identify the measurement error and the inefficiencies: see Ritter and Simar (1997). Similar problems would apply to all

[5] A similar prior was adopted for the exponential case $j = 1$ in Koop *et al.* (1997).

other parametric models that have been used in the literature: for example, a half-Normal inefficiency distribution would be restrictive in assigning mass to efficiency intervals, whereas a truncated Normal would lead to identification problems in the context of our hospital application.

5 The MCMC Sampler

The model is fitted using an MCMC algorithm. There are two generic methods for fitting mixtures of Dirichlet processes. In general, sampling is not straightforward because F is an infinite dimensional parameters. Ishwaran and James (2001) suggest truncating F. An alternative methods is due to Escobar and West (1995). They marginalise over F and use the predictive distribution which is available analytically as in (6). The method implemented here is based on a refinement of the latter approach described in MacEachern (1994). The sampler uses a data augmentation scheme. Each observation is associated with an element of the Dirichlet process using the latent variable s_i for the i-th observation. We denote the number of observations that are assigned to the k-th element (or "cluster") by $n_k, k = 1, \dots, K$ and the K-dimensional vector of distinct inefficiencies by u. The sampler will thus be run on $(s, u, M, \alpha, \beta, \sigma^2, \lambda|y)$, where y groups all NT observations and we do not explicitly indicate conditioning on the explanatory variables x_{it} and w_i. Following MacEachern (1994) we integrated u from the model when sampling the latent variables s. This technique is used in MCMC schemes for mixtures of Dirichlet processes in order to improve the mixing of the chain. The inefficiencies in u are then sampled conditionally on s.

The full conditional distribution of the latent allocations s_i is then given by

$$p(s_i = k|\alpha, \beta, \sigma^2, \lambda, s_{-i}, y) = \frac{n_k}{n+M} \int p(y_i|\alpha, \beta, \sigma^2, u_k)\, p(u_k|s_{-i}, \alpha, \beta, \sigma^2, \lambda, y_{-i})\, du_k,\ 1 \le k \le K \tag{10}$$

$$p(s_i = K+1|\alpha, \beta, \sigma^2, \lambda, s_{-i}, y) = \frac{M}{n+M} \int p(y_i|\alpha, \beta, \sigma^2, u_i)\, h(u_i|\lambda)\, du_i, \tag{11}$$

where y_i denotes the T observations for firm i, s_{-i} denotes all elements of s except for s_i, y_{-i} is similarly defined and $h(\cdot)$ is the density function corresponding to H. The integrals in equations (10) and (11) can both be calculated analytically for the class of distributions we consider for u_i. The integral in (11) is given by

$$p(y_i|\alpha, \beta, \sigma^2, \lambda) = \frac{\lambda^j}{\Gamma(j)} \frac{T^{-1/2}}{(2\pi\sigma^2)^{(T-1)/2}} \Phi\left(\frac{m_i}{\sqrt{T}\sigma}\right) c_{ji} \exp\left\{-\frac{1}{2}\sigma^{-2}\left[\sum_{t=1}^{T}(y_{it} - \alpha - x_{it}'\beta)^2 - \frac{1}{T}m_i^2\right]\right\},$$

where

$$m_i = \sum_{t=1}^{T}(y_{it} - \alpha - x_{it}'\beta) - \lambda\sigma^2,$$

j is the integer shape parameter of H and $c_{1i} = 1$, $c_{2i} = \frac{m_i}{T} + \frac{\sigma}{\sqrt{2\pi T}}\exp\left(-\frac{1}{2\sigma^2}\frac{m_i^2}{T}\right)/\Phi(m_i/\sqrt{T}\sigma)$ and $c_{3i} = \frac{\sigma^2}{T} + \frac{m_i}{T}c_{2i}$.

The integral in (10) is taken with respect to the full conditional distribution for u_k which can be expressed as follows. Let $y^{(k)}$ and $x^{(k)}$ be the responses and covariates for the firms allocated to the k-th cluster and define

$$m_k' = \sum_{i=1}^{n_k} \sum_{t=1}^{T} \left(y_{it}^{(k)} - \alpha - x_{it}^{(k)'} \beta \right) - \lambda \sigma^2.$$

The required distribution for u is then the product of

$$p(u_k|s, \alpha, \beta, \sigma^2, \lambda, y) = {c_{jk}'}^{-1} \left[\Phi \left(\frac{m_k'}{\sqrt{n_k T} \sigma} \right) \right]^{-1} u_k^{j-1} f_N \left(u_k \left| \frac{m_k'}{n_k T}, \frac{\sigma^2}{n_k T} \right. \right) I(u_k \geq 0),$$

where c_{jk}' is as defined above replacing m_i by m_k'/n_k and σ^2 by σ^2/n_k. This expression simplifies to a truncated Normal for $j = 1$ (an exponential centring distribution for the inefficiencies). The integral in (10) is given by

$$\begin{aligned} p(y_i|\alpha, \beta, \sigma^2, \lambda, y_{-i}, s_{-i}) = & \frac{c_{jk}''}{c_{jk}'} \left[\Phi \left(\frac{m_k'}{\sqrt{n_k T} \sigma} \right) \right]^{-1} \left[\Phi \left(\frac{m_k''}{\sqrt{(n_k + 1) T} \sigma} \right) \right] \sqrt{\frac{n_k}{n_k + 1}} \left(\frac{1}{2 \pi \sigma^2} \right)^{T/2} \\ & \times \exp \left\{ -\frac{1}{2} \sigma^{-2} \left[\sum_{t=1}^{T} (y_{it} - \alpha - x_{it}' \beta)^2 + \frac{{m_k'}^2}{n_k T} - \frac{{m_k''}^2}{(n_k + 1) T} \right] \right\}, \end{aligned}$$

where

$$m_k'' = m_k' + \sum_{t=1}^{T} \left(y_{it} - \alpha - x_{it}' \beta \right)$$

and c_{jk}'' is defined as above using $m_k''/(n_k + 1)$ and $\sigma^2/(n_k + 1)$.

The parameter M is sampled using a Metropolis-Hastings random walk step within the Gibbs sampler. The full conditional distribution for M is given by

$$p(M|s) \propto \frac{M^K}{\prod_{i=1}^{n} M + i - 1} p(M),$$

where $p(M)$ is given by (9) and K is the number of distinct values (clusters) in the Dirichlet process.

The full conditional distributions for λ and σ^2 are simply given by

$$\lambda | u \sim \text{Ga} \left((n + 1) j, \sum_{i=1}^{n} u_{s_i} - \ln r^\star \right),$$

and

$$\sigma^{-2} | s, u, \alpha, \beta, y \sim \text{Ga} \left(nT/2, \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - \alpha - x_{it}' \beta - u_{s_i})^2 / 2 \right),$$

leading to standard Gibbs steps.

To update the sampler in α we use a mixture of a centred and a non-centred sampler. Centring in normal linear models was introduced by Gelfand, Sahu and Carlin (1995). The issue of updating α

is particularly critical in stochastic frontier models, since both α and u_i have an additive effect: the model is naturally informative on the sum of both and identification has to come from the distributional assumptions on u_i. This is always a problem but even more so in the semiparametric case where the inefficiency distribution has no rigid parametric form and supports distributions with nearly all their mass a long way from zero. The local moves of the Gibbs sampler can be slow to adjust α and to move much of the mass "close" to zero with a non-centred parameterization. Thus, we use a hybrid sampler by mixing updates from the centred and the non-centred parameterizations in order to improve the properties of the algorithm. The centred parameterisation in the stochastic frontier model is $z_i = \alpha + u_i$. In the centred parameterization we use Gibbs step updates $\alpha|z, \lambda$ and $\beta|\alpha, u, \sigma^2, y$. The latter full conditional is the usual linear model update, while the full conditional distribution for $\alpha|z, \lambda$ has the form

$$\prod_{i=1}^{n}(z_i - \alpha)^{j-1}\exp(n\lambda\alpha), \qquad \alpha < \min(z_i).$$

If $j = 1$, then $\min(z_i) - \alpha$ has an exponential distribution with mean $1/n\lambda$, which allows simple updating of α. In the non-centred (original) parameterisation α and β can easily be updated through a Normal full conditional as in the usual linear model (see *e.g* (A.7) in Koop *et al.*, 1997).

Model fitting for the case with varying inefficiency distributions uses a slight modification of the MCMC sampler described above. Throughout, we need to change λ to $\exp(w_i'\gamma)$ and we replace the step for λ by a sampler for $\phi_l = \exp(\gamma_l)$. If all w_{il} only take the values 0 and 1 then the full conditional for ϕ_l is

$$p(\phi_l|\phi_{-l}, u, s) = \text{Ga}\left(\phi_l \,\middle|\, a_l + j\sum_{i=1}^{n} w_{il}, g_l + \sum_{i=1}^{n} w_{il}u_i \prod_{k\neq l}\phi_k^{w_{ik}}\right).$$

In addition, we change n to $n^{(m)}$, the sample size for the m-th group of firms (characterised by the same values for w_i) in the samplers for s and u and repeat these steps for each group. For the mass parameter M, the conditional posterior becomes

$$p(M|s) \propto p(M)\prod_m \frac{M^{K^{(m)}}}{\prod_{i=1}^{n^{(m)}} M + i - 1},$$

where $K^{(m)}$ is the cluster size for group m. Finally, for the centred parameterisation, the full conditional distribution becomes

$$\prod_m \prod_{i=1}^{n^{(m)}} \left(z_i^{(m)} - \alpha\right)^{j-1} \exp\left(n^{(m)}\prod_l \phi_l^{w_{il}}\alpha\right), \qquad \alpha < \min_{i,m}\left(z_i^{(m)}\right)$$

where $z_i^{(m)}$ is the equivalent of z_i for group m.

We are interested in the posterior distribution of functions of the unknown distribution F and in particular the probability mass over particular ranges. Gelfand and Kottas (2002) describe a method for drawing approximate realisations of F using a finite truncation of the representation described in

Sethuraman (1994). Conditionally on s, F follows an updated Dirichlet process, DP$(M^\star F^\star)$ where $M^\star = M + n$ and

$$F^\star \stackrel{d}{=} \sum_{i=1}^{K} \frac{n_k}{n+M}\delta_{u_i} + \frac{M}{n+M}H.$$

An approximate realisation, F_{draw}, from this updated Dirichlet process can be sampled using Sethuraman's representation truncated to have a atoms. Ishwaran and James (2001) describe potential methods for choosing a suitably large value for a. The realisation, F_{draw}, is

$$F_{\text{draw}} \stackrel{d}{=} \sum_{i=1}^{a} p_i \delta_{\theta_i}$$

where $\theta_i \stackrel{i.i.d.}{\sim} F^\star$ and $p_i = V_i \prod_{j<i}(1 - V_j)$ where $V_i \stackrel{i.i.d.}{\sim}$ Be$(1, M^\star)$. This immediately leads to the predictive distribution of efficiencies corresponding to unobserved firms. The predictive of the observables, $p(y_{n+1}, \ldots, y_{n+m}|y_1, \ldots, y_n)$, can also be calculated through a Rao-Blackwellised scheme (Gelfand and Smith, 1990) using the approximate draw, F_{draw}. Conditional on F, α, β and σ^2, the future observations are independent and have an analytically known distributional form.

6 Application to a Hospital Cost Frontier

We use the same data as in Koop *et al.* (1997), where a parametric Bayesian analysis was conducted and we refer to the latter paper for further details and background of hospital cost estimation, the data and the particular frontier used. We now use our semiparametric modelling approach centred over the parametric models favoured by Koop *et al.* (1997). The data correspond to $n = 382$ nonteaching U.S. hospitals over the years 1987-1991 ($T = 5$), selected so as to constitute a relatively homogeneous sample. The frontier describing cost (C) involves five different outputs $Y_1, \ldots, Y_5$: number of cases, number of inpatient days, number of beds, number of outpatient visits and a case mix index. We also include a measure of capital stock, K, an aggregate wage index, P and a time trend t to capture any missing dynamics. We choose a flexible translog specification and impose linear homogeneity in prices, which allows us to normalize with respect to the price of materials. Thus, in the notation of (1) and dropping observational subscripts for ease of notation, $y = \ln C$ and $x'\beta$ becomes:

$$x'\beta = \sum_{i=1}^{5} \beta_i \ln Y_i + \beta_6 \ln P + \sum_{i=1}^{5}\sum_{j=i}^{5} \psi_{ij} \ln Y_i \ln Y_j + \beta_7 (\ln P)^2 + \sum_{i=1}^{5} \beta_{7+i} \ln Y_i \ln P +$$

$$\beta_{13} \ln K + \sum_{i=1}^{5} \beta_{13+i} \ln Y_i \ln K + \beta_{19} \ln P \ln K + \beta_{20} (\ln K)^2 + \beta_{21} t + \beta_{22} t^2,$$

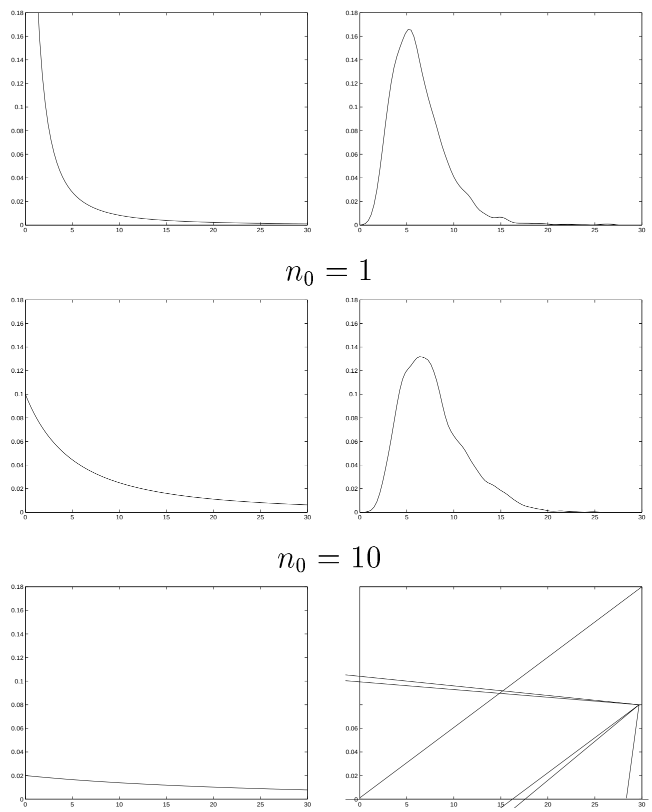
where $\psi_{ij}, i \leq j$ provide the remaining 15 elements of β.

Throughout, we choose an exponential centring distribution H ($j = 1$) which implies a prior median efficiency of $r^\star = 0.8$.

All results are based on runs of length 75 000, where we record every 5-th drawing after a burn-in of 10 000 draws, which proved enough for convergence.

6.1 The case with a common efficiency distribution

We first examine the case where the nonparametric inefficiency distribution is common to all hospitals. In order to explore the robustness of the posterior results with respect to the choice of n_0 in the prior on M in (9), Figure 3 contrasts prior and posterior density functions for the mass parameter M. Throughout, we take $\eta = 1$, reflecting a large amount of prior uncertainty. Clearly, the choice of n_0, although leading to very different priors, has virtually no effect on the posterior for M, which indicates that the data convey quite a bit of information. It is also noteworthy that the posteriors are concentrated on fairly small values of M, leading to quite small values of the weight $M/(M+n)$ (where $n = 328$) assigned to the centring distribution in the predictive and thus indicating that a significant departure from the centring distribution is to be expected.



MCMC output.

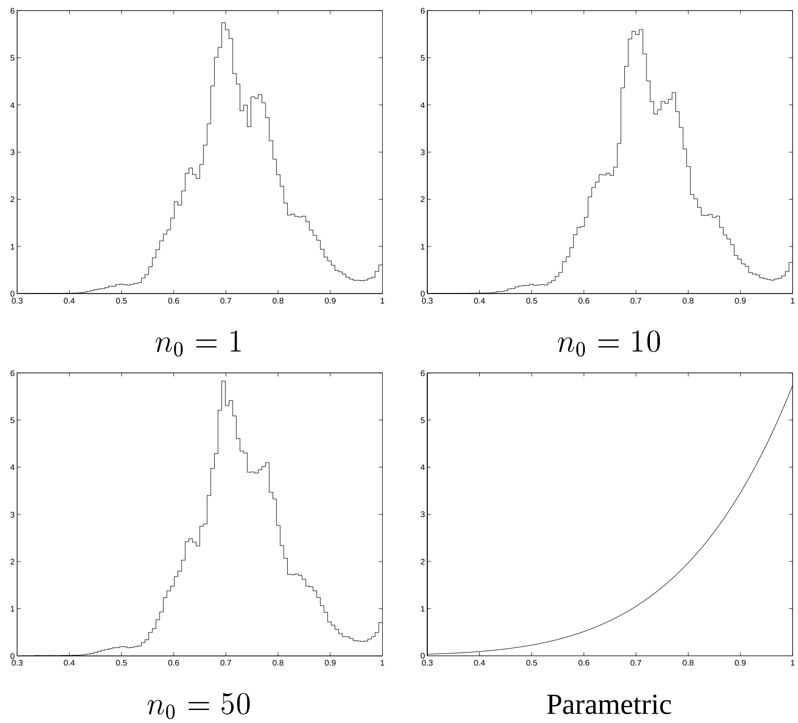


Figure 4: Posterior predictive efficiency distribution

Note that the distributions in Figure 4 are also very robust with respect to the choice of n_0. In the sequel, we shall therefore limit ourselves to showing results for $n_0 = 10$ only. The corresponding prior predictive distribution is found in Figure 1. In Figure 4, we also display the predictive efficiency distribution of the parametric model[6]. It is clear that the latter is far too restrictive to capture the data information. In particular, there is a lot of mass in the area (0.6, 0.8), which can not be accommodated by the parametric model (based on an exponential inefficiency distribution).

In order to get a better picture of the posterior mass on certain efficiency intervals, the graphs in Figure 5 indicate the distribution of mass in selected intervals. The randomness of these probabilities corresponds to variation in F (for the semiparametric model) and λ, according to their posterior distributions. Comparison with the equivalent prior measures in Figure 2 shows that the data information generally leads to a much tighter distribution. The main point of Figure 5, however, is the comparison between the semiparametric and the parametric models. It is clear that the parametric model is far too restrictive, giving the illusion of very precise inference on these interval probabilities. In particular, the parametric model greatly underestimates the uncertainty associated with the mass assigned to the intervals in $(0.6, 0.9)$: it puts all the mass close to the maximum value derived in Section 4 for $(0.8, 0.9)$, and dramatically falls short of the probability in the intervals $(0.7, 0.8)$ and $(0.6, 0.7)$. As a consequence, it puts far too much mass in the intervals to the right of 0.9.

[6]For the parametric model, this is very much like the prior predictive in Figure 1, except that this is now averaged with the posterior distribution of λ rather than with the prior.

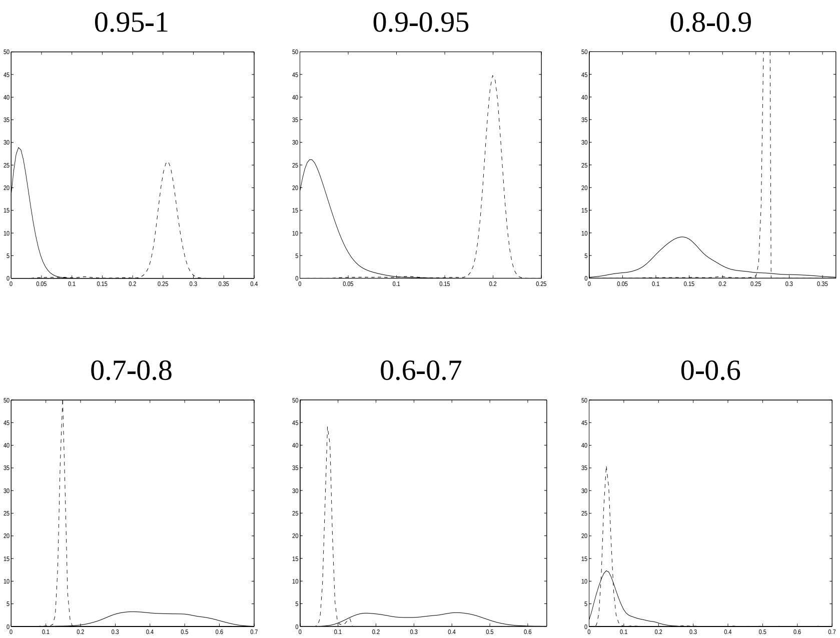


Figure 5: Posterior distribution of selected efficiency interval probabilities. Semiparametric: drawn lines. Parametric: dashed lines

So far, we have considered the efficiency distribution corresponding to an unobserved hospital in this industry, but we can also be interested in the efficiency of firms that we have actually observed. Such firm-specific posterior efficiency distributions are displayed in Figure 6 for a number of firms: the hospitals corresponding to the minimum, maximum and quartiles of the efficiency distribution, as measured by the posterior mean efficiencies in the parametric model.

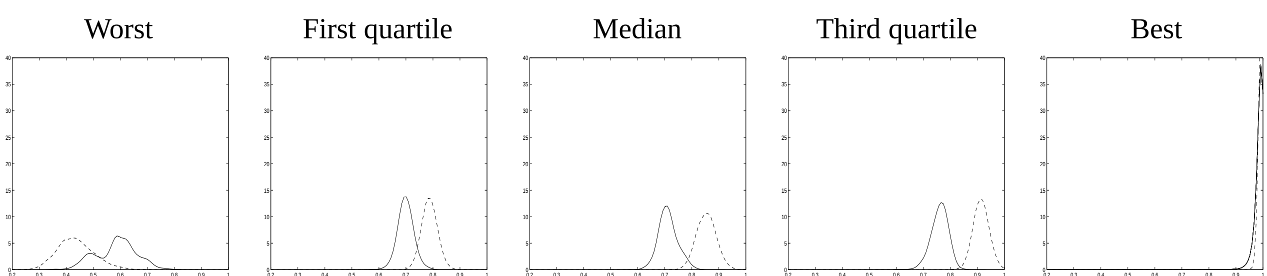


Figure 6: Posterior efficiency distributions for five selected firms. Semiparametric: drawn lines. Parametric: dashed lines

Even though the spread is not as large as for the predictive efficiency distribution in Figure 4 it is clear from Figure 6 that considerable uncertainty exists about the firm efficiencies for hospitals that we have actually observed. In line with the shape of the predictive efficiency distribution, the parametric model leads to larger differences between the firms, especially for the three central firms, but even the best and the worst firms are more extreme. Clearly, there is a substantial difference between the efficiency inference resulting from both models. For example, the efficiency of the third quartile hospital, for which the parametric model assigns most mass to the interval $(0.84, 1)$, is found to be much smaller if we use the less restrictive semiparametric model. In fact, the posterior efficiency distributions corresponding to both models hardly overlap! A similar massive overestimation occurs for the median firm. Except for the very worst hospital, all firms considered here see their efficiency estimates artificially boosted by imposing the functional form of the parametric model. This is consistent with the fact that the parametric model is obliged to locate the frontier close to the bulk of the data, whereas the semiparametric model can allow for a few firms doing much better than the rest. Although for some of the hospitals the semiparametric efficiency distribution displays a slightly larger spread, the extra flexibility of the semiparametric model is not reflected in unreasonable amounts of uncertainty. Thus, with the present dataset of only moderate size, efficiency inference is quite feasible through our semiparametric framework.

The uncertainty inherent in firm efficiencies makes a clear ranking in terms of efficiency less than straightforward. What we can do, however, is to compute the probability that one hospital is more efficient than another. Tables 1 and 2 state these probabilities for the five hospitals mentioned above. The numbers confirm the fact that these hospitals are assigned more clearly separated efficiency distributions in the parametric model. Table 1 also illustrates the discrete nature of the Dirichlet Process: entries (i, j) and (j, i) do not necessarily add up to unity, and the difference is the probability that the two firms are equally efficient (*i.e.* they belong to the same cluster).

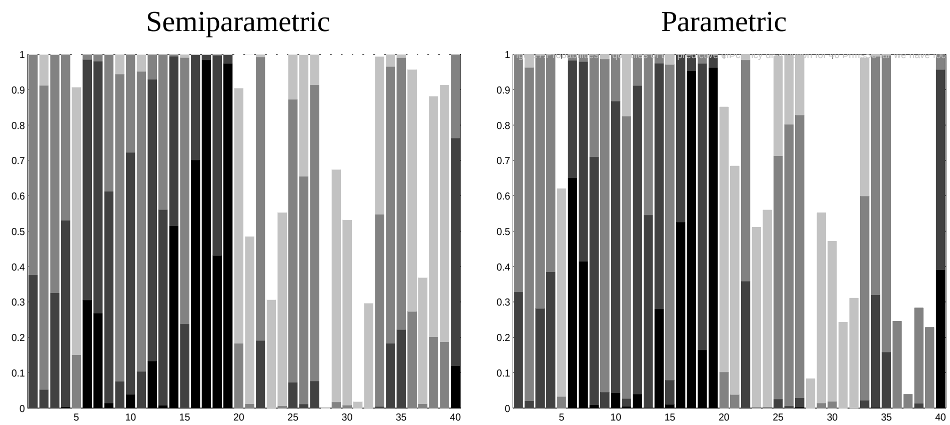
	Worst	Q1	Median	Q3	Best
Worst	0	0.0398	0.0273	0.0030	0
Q1	0.8960	0	0.2065	0.0158	0
Median	0.9215	0.4365	0	0.0460	0
Q3	0.9852	0.9340	0.8313	0	0
Best	1.0000	1.0000	1.0000	1.0000	0

Table 1: Entry i, j is the probability that hospital i is more efficient than hospital j; Semiparametric

	Worst	Q1	Median	Q3	Best
Worst	0	0	0	0	0
Q1	1.0000	0	0.0725	0.0015	0
Median	1.0000	0.9275	0	0.0825	0
Q3	1.0000	0.9985	0.9175	0	0.0020
Best	1.0000	1.0000	1.0000	0.9980	0

Table 2: Entry i, j is the probability that hospital i is more efficient than hospital j; Parametric

A quick characterization of the relative position of a firm in the predictive efficiency distribution can be obtained by considering the probabilities that the firm's efficiency falls in each of the quintiles of the predictive efficiency distribution. A graphical display as in Figure 7 might be a good way to quickly assess the relative efficiencies of a number of firms. This figure plots the probabilities assigned to each quintile for the first 40 firms in the sample in a stacked bar chart where darker shading indicates lower efficiency. Most firms have substantial probabilities of being in two quintiles, but some hopitals are mostly in the lower or the upper quintile: for example, hospitals 17 and 19 are mostly in the lowest quintile, while hospitals 28 and 31 (especially for the semiparametric model) have high probability on the highest quintile. This is a quick visual aid to screen for outlying hospitals (in either direction).



interest. As an illustration, we can consider the wage elasticity of cost, which is given by $\varepsilon_P = \beta_6 + 2\beta_7 \ln P + \sum_{i=1}^{5} \beta_{7+i} \ln Y_i + \beta_{19} \ln K$. This is a property of the frontier, which we can evaluate at the values for $P, K, Y_i, i = 1, \ldots, 5$ observed in the sample. Figure 8 displays the posterior densities of ε_P for the observed levels of prices, capital and output (in the mid-sample year 1989) corresponding to the five firms selected earlier. The spread of wage elasticities is similar for both models, but location does vary substantially, especially for the firms with extreme mean efficiencies, where the semiparametric model tends to shift the elasticity towards more reasonable values. The worst firm has an unusually low wage, which is an area where the frontier does not allow for precise inference on the wage elasticity. To get a better idea of how wage elasticities vary over the range of observed values, Figure 9(a) plots the posterior means of ε_P corresponding to each of the nT observations for both models against each-other. It is clear that there is substantial variation, but the within-sample spread is larger for the parametric model than for the semiparametric one. In order to assess the properties of both frontier surfaces, we did not impose the regularity condition $\varepsilon_P > 0$[8], and we notice from Figure 9(a) that the parametric frontier leads to a few violations, while the semiparametric frontier leads to virtually no violations of this condition, suggesting that it is better behaved than the parametric frontier. Interestingly, if we consider the dynamics of ε_P over the sample period, it has almost uniformly decreased. This is illustrated by Figures 9(b) and 9(c) which plot the posterior means of the wage elasticity in 1991 versus that in 1987 for the same firm. Both models indicate a substantial decline in the relative effect of wages on best-practice (fully efficient) levels of cost.

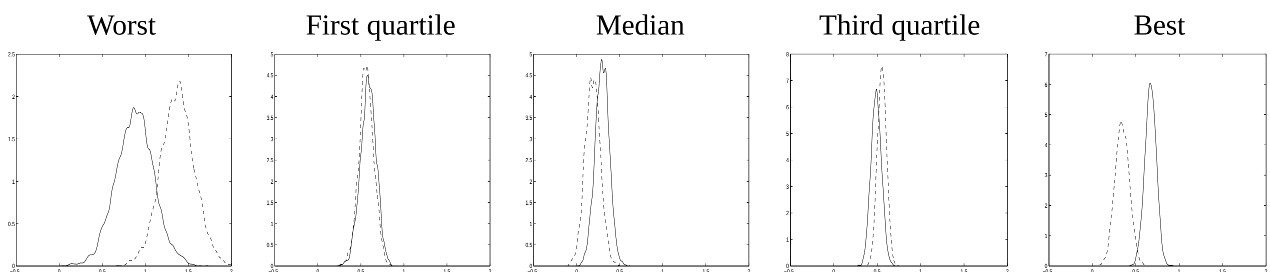
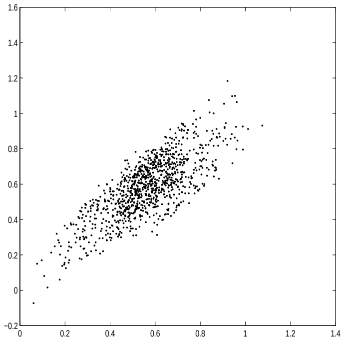


Figure 8: Posterior wage elasticity distributions for five selected firms. Semiparametric: drawn lines. Parametric: dashed lines

Finally, we compare the quality of the models by how well they predict out of sample. In order to assess this, we consider predicting the observable y_f given the corresponding values of the regressors, grouped in a vector x_f. We shall now randomly split the sample into n hospitals on which we base our posterior inference and q hospitals which we retain in order to check the predictive accuracy of the model. As a formal criterion, we use the log predictive score, introduced by Good (1952). It is a strictly proper scoring rule (*i.e.*, it induces the forecaster to be honest in divulging his predictive distribution) and is associated with the well-known Kullback-Leibler divergence between the actual sampling density and the predictive density. For $f = n+1, \ldots, n+q$ (*i.e.*, for each hospital in the prediction sample) we base our measure on the joint qT-dimensional predictive cost distribution evaluated in these retained

[8]This can be done trivially, but would make very little difference, even for the parametric case, as is obvious from Figure 9(a).

(a) (b) (c)



6.2 The varying efficiency distribution case

We now examine the model explained in Subsection 3.2 where we specify a separate inefficiency distribution for each group of hospitals, where groups are characterized by ownership structure (for-profit, non-profit or government-run) and a dummy variable, called “staff ratio”, based on the number of clinical workers per patient[10]. Different combinations of these characteristics lead to six groups of firms, with sample sizes $n^{(m)}$ ranging from 20 (government-run, low staff ratio) to 141 (non-profit, low staff ratio). Throughout this Subsection, we shall use $n_0 = 10$ in the prior for M. Figure 10 plots the posterior distribution of M, which puts even more mass on small values than in the common efficiency distribution case. However, note that sample sizes $n^{(m)}$ can be substantially smaller than the full sample (with $n = 382$).

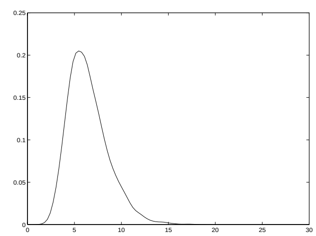


Figure 10: Posterior density for M; Varying efficiency case

The posterior predictive efficiency distributions for unobserved firms in each of these groups are displayed in Figure 11, where the superimposed dashed lines reflect the predictives from the parametric model (with a mean that changes with w_i as explained in Subsection 3.2). Importantly, taking into account these characteristics reduces the predictive uncertainty, as is evidenced by the fact that the predictive distributions have smaller spread than in the common efficiency case. Clearly, the differences between the semiparametric and parametric models are huge, as before, but we also note that the semiparametric efficiency distributions vary a lot between hospital groups. Modal efficiency values tend to be around 0.8 for non-profit and government hospital, and around 0.7 for for-profit hospitals[11], but these numbers hide important differences in distribution. In order to further illustrate some of these differences, Figure 12 indicates the probability distributions of the posterior mass on the interval $(0.95, 1)$. It is clear that the hospital characteristics have an important effect on these distributions. The main influence of the ratio of clinical staff to patients seems to be to lower the probability of very high efficiencies in the interval $(0.95, 1)$, except for the for-profit hospitals where it sharply reduces the mass around $r_{n+1} = 0.8$ (see Figure 11). This is an interesting finding, which corroborates the results of Koop *et al.* (1997), who find that in the parametric model the higher staff ratio tends to decrease efficiency. We have followed the setup of Koop *et al.* (1997) in parameterizing the dependence of the mean inefficiency on w_i by including

[10]This variable equals one if the ratio (averaged over years) of clinical workers to average daily census is above the mean computed over all hospitals.

[11]This is roughly in line with our parametric results and the results in Koop *et al.* (1997), who use an exponential parametric model and find that predictive efficiencies are similar for non-profit and government-run hospitals and tend to be larger than those for for-profit hospitals. However, the parametric model tends to overestimate efficiencies.

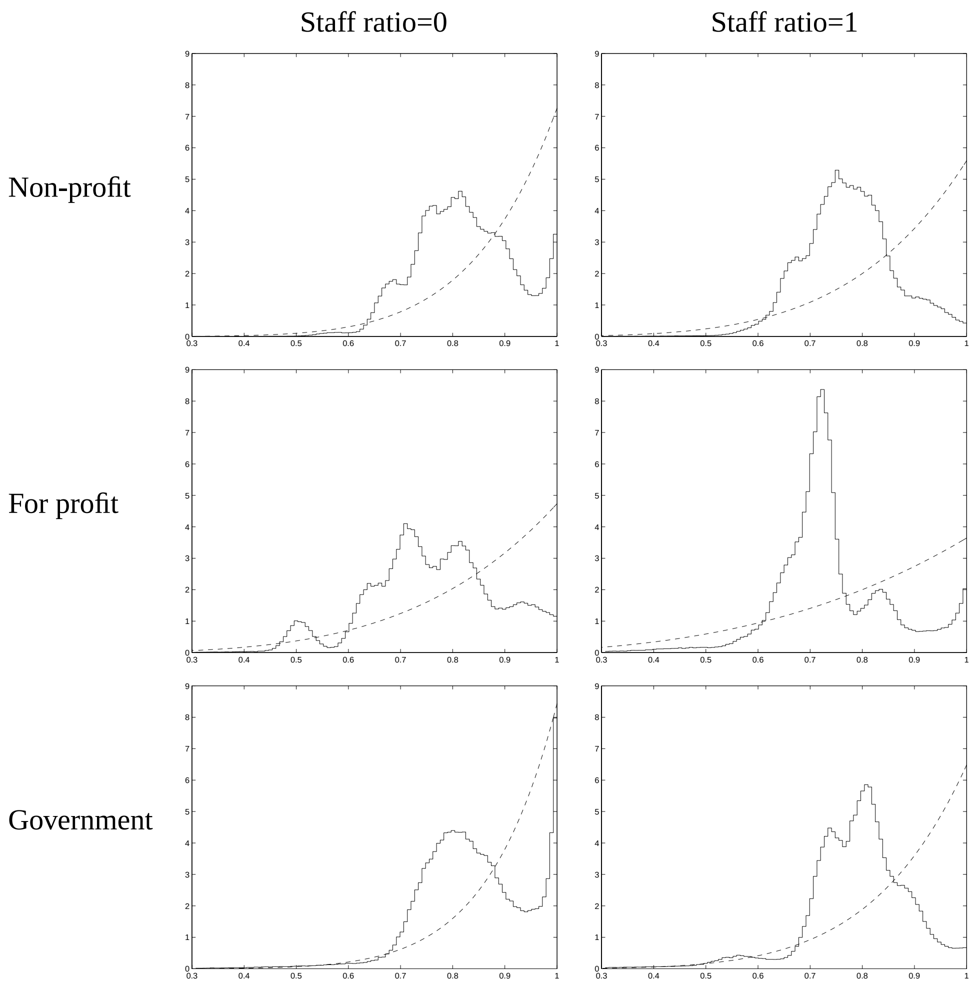


Figure 11: Predictive densities; Varying efficiency case. Semiparametric: drawn lines. Parametric: dashed lines

only the main effects and no interactions. This means that the effect of staff ratio is constrained to be the same for all ownership types. In view of our semiparametric results, this seems in contrast with the data evidence. At the cost of adding two more parameters to γ, we could try and capture these interaction effects in the parametric model. The semiparametric model, however, still allows a much more detailed investigation. One would perhaps not expect to find for-profit hospitals to be generally less efficient. One possible explanation for this fact is that for-profit hospitals provide an important quality aspect, which is not captured in our measured outputs. One way to offer higher quality is to assign more clinical staff per patient. This would indeed tend to reduce cost efficiency. However, whereas Figure 11 indicates that this sharply reduces the probability of finding a very efficient non-profit or government-run hospitals, it has a rather different effect on for-profit hospitals. Thus, this suggests that the very efficient non-profit and government-run hospitals indeed have less than average staff per patient, but this is not the case for the for-profit hospitals: whereas the "third-quartile" hospitals in this group will tend to have less than average staff, the very efficient ones are not particularly characterized by their staff ratio. Interestingly, the for-profit hospitals with high staff ratio (34 out of 64 for-profit hospitals) have a very tight predictive efficiency distribution (around 0.7), indicating that this is quite a homogeneous group.

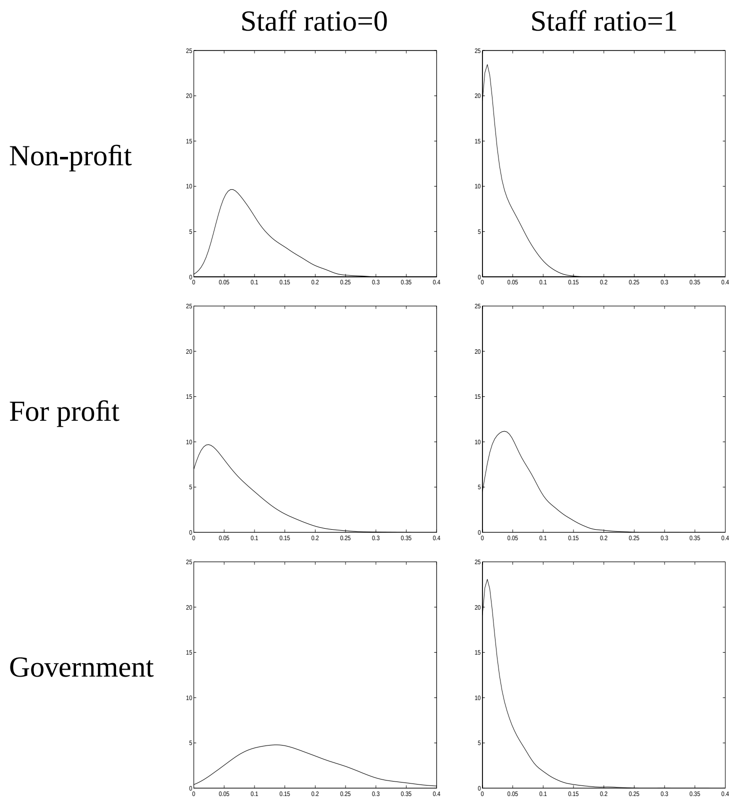


Figure 12: Posterior distribution of efficiency probability in $(0.95, 1)$

These Figures also illustrate that the common efficiency distribution (as displayed in Figure 5) is not simply a weighted average of the varying efficiency distributions for the different groups. For example, in every group the probability of the interval $(0.95, 1)$ has more mass away from zero than that for the common efficiency case. In some sense, this is really a small-sample effect, namely that in very small samples the centring distribution is given a larger weight, given the common M, and thus more weight is allocated to regions close to full efficiency. Whereas this issue will undoubtedly play a part, it has to be stressed that the effect also occurs for the largest groups and the second smallest group displays one of the smallest differences with the common efficiency case. Thus, the varying efficiency case leads to efficiency distributions that are somewhat closer to the prior (the parametric model) throughout due to smaller sample sizes, but for some groups the observations are more strongly in conflict with the exponential prior than for others. As an example, consider government-run hospitals where both groups are of similar (small) size, and the distribution for hospitals with low staff ratio is much closer to the prior (certainly in areas close to full efficiency) than for their high staff ratio counterparts.

Figure 13 displays the posterior efficiency distributions for the five selected firms mentioned earlier. A comparison with the common efficiency distribution results in Figure 6 reveals that the parametric model leads to virtually identical inference as in the common efficiency case. The semiparametric results, however, are now somewhat closer to the outcome of the parametric model. This also points to the fact

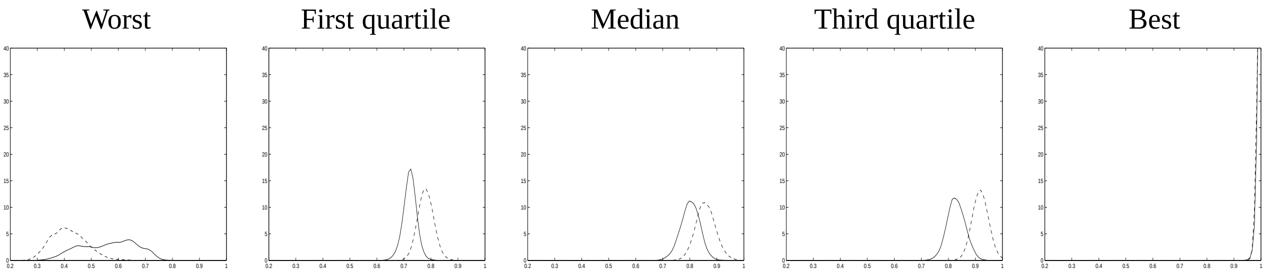


Figure 13: Posterior efficiency distributions for five selected firms; Varying efficiency case. Semiparametric: drawn lines. Parametric: dashed lines

that the centring distribution gets a bit more importance once we start dividing the sample into groups. In line with the move towards the parametric model, the ability to rank the five hospitals in terms of their efficiency is also increased, as Table 5 indicates. Another difference with the common efficiency case is that now all but the first two hospitals belong to different groups, and thus the probability of a tie is zero for most efficiency comparisons. Despite the smaller sample sizes in each group, the posterior efficiency distributions in Figure 13 are actually slightly less dispersed than in the common efficiency case. This is a consequence of the fact that firms in a particular group tend to more homogeneous.

	Worst	Q1	Median	Q3	Best
Worst	0	0.0243	0.0035	0.0022	0
Q1	0.9120	0	0.0173	0.0003	0
Median	0.9965	0.9828	0	0.2095	0
Q3	0.9978	0.9998	0.7905	0	0
Best	1.0000	1.0000	1.0000	1.0000	0

Table 5: Entry i, j entry is the probability that hospital i is more efficient than hospital j - Semiparametric with varying efficiency distribution

Thus, taking into account firm characteristics affects both predictive efficiency distributions and the posterior efficiency distributions of individual hospitals. However, if we combine the two and construct a comparison of firms in terms of the probabilities of quintiles of the predictive efficiency distribution, displayed in Figure 14, the result is quite similar to the case with a common efficiency distribution as in Figure 7. The issue of whether or not we use separate efficiency distributions for groups of firms does not, therefore, seem to matter much if we wish to identify hospitals that have performed particularly well or poorly. It does play an important part, however, if we wish to predict the efficiency of an unobserved hospital, and if we want to draw conclusions on the effect of certain characteristics on efficiency.

Out-of-sample predictive ability of the model proved comparable to that of the semiparametric model with common efficiency distribution.

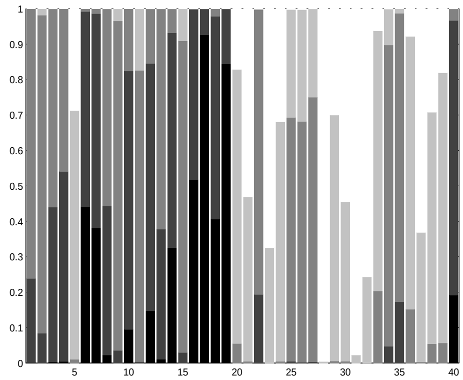


Figure 14: Probabilities of quintiles of the predictive efficiency distribution for 40 firms; Semiparametric with varying efficiency distribution

7 Concluding Remarks

In this paper, we have developed a Bayesian semiparametric modelling framework for stochastic frontier models. The nonparametric part of the model is the inefficiency distribution, for which we use a random measure F based on a Dirichlet process. We centre F over a class of Gamma distributions with unknown scale parameter, with a prior that is elicited through the prior median efficiency. We recommend the use of an Inverted Beta prior on the mass parameter of the process, and we calibrate the hyperparameter choice and investigate the robustness properties. A modification of the model, where we allow for a separate nonparametric inefficiency distribution for different groups of firms, is also proposed. Dependence is then introduced through the centring distribution, which is also allowed to depend on firm characteristics, and a common mass parameter. In addition, of course, the frontier is common to all groups. We develop and implement an efficient MCMC algorithm and apply our methodology to a panel data set of 382 hospitals, for which we estimate a cost frontier.

The empirical results illustrate the feasibility of practically useful inference on efficiencies or properties of the frontier with data sets of moderate size using both our modelling frameworks. The findings in the case of a common efficiency distribution also clearly indicate the limitations of the parametric model that is often used in this literature (and was used on the same data by Koop *et al.* 1997). Predicting the efficiency for an unobserved firm on the basis of the parametric model is shown to be totally misleading. Even the efficiency inference on observed hospitals is substantially different from what we conclude on the basis of the semiparametric model. Since the semiparametric results indicate more clustering of the predictive efficiency distribution there is more overlap in the efficiency distributions of the quartile firms. Nevertheless, identification of where individual hospitals are situated with respect to the predictive efficiency distribution is not fundamentally changed by assuming a nonparametric distribution for inefficiencies. Inference on wage elasticity of cost is mainly affected in that outlying firms are shifted towards more reasonable values, suggesting that the frontier surface estimated by the semiparametric model is better behaved.

Allowing for varying efficiency distributions, according to firm characteristics, again leads to large differences in predictive and posterior efficiency inference between the parametric and semiparametric models. Interestingly, it also indicates quite substantial differences between the efficiency distributions corresponding to the various groups. Since each group has its own specific nonparametric inefficiency distribution (linked as indicated above), the semiparametric model allows for detailed analysis of how these distributions change with hospital characteristics. There is a cost, however, in that sample sizes per group can become quite small, which limits the amount of inference that can be drawn from the data.

Model validation on the basis of out-of-sample predictive ability strongly favours the semiparametric models.

References

Adams, R.M., Berger, A.N. and Sickles, R.C. (1999): “Semiparametric approaches to stochastic panel frontiers with applications in the banking industry,” *Journal of Business and Economic Statistics*, 17, 349-358.

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

Antoniak, C. E. (1974): “Mixtures of Dirichlet processes with applications to non-parametric problems,” *Journal of the American Statistical Society*, 2, 1152-74.

Blackwell, D. and MacQueen, J.B. (1973): “Ferguson distributions via Pólya urn schemes,” *Annals of Statistics*, 1, 353-355.

Carota, C. and Parmigiani, G. (2002): “Semiparametric regression for count data,” *Biometrika*, 89, 265-281.

Cifarelli, D.M. and Regazzini, E. (1978): “Nonparametric statistical problems under partial exchangeability. The use of associative means,” (in Italian) *Annali del'Instituto di Matematica Finianziara dell'Università di Torino, Serie III*, 12, 1-36.

Escobar, M. D. and West, M. (1995): “Bayesian density-estimation and inference using mixtures,” *Journal of the American Statistical Association*, 90, 577-588 .

Fan, Y.Q., Li, Q. and Weersink, A. (1996): “Semiparametric estimation of stochastic production frontier models,” *Journal of Business and Economic Statistics*, 14, 460-468.

Ferguson, T. S. (1973): “A Bayesian analysis of some nonparametric problems,” *Annals of Statistics*, 1, 209-230.

Fernández, C., Osiewalski, J. and Steel, M.F.J. (1997): “On the use of panel data in stochastic frontier models with improper priors,” *Journal of Econometrics*, 79, 169-193.

Gelfand, A. E., and Kottas, A. (2002): “A Computational Approach for Full Nonparametric Bayesian Inference under Dirichlet Process Mixture Models,” *Journal of Computational and Graphical Statistics*, 11, 289-305.

Gelfand, A. E., Sahu, S.K. and Carlin, B.P. (1995): “Efficient parameterizations for normal linear mixed models,” *Biometrika*, 82, 479-488.

Gelfand, A. E. and Smith, A. F. M. (1990): “Sampling-based approaches to calculating marginal densities,” *Journal of the American Statistical Association*, 85, 398-409.

Good, I.J. (1952): “Rational Decisions,” *Journal of the Royal Statistical Society*, B, 14, 107-114.

Huang, C.J. and Fu, T.T. (1999): “An average derivative estimation of stochastic frontiers,” *Journal of Productivity Analysis*, 12, 45-53.

Ishwaran, H. and James, L. (2001): “Gibbs Sampling Methods for Stick-Breaking Priors,” *Journal of the American Statistical Society*, 96, 161-73.

Koop, G., Osiewalski, J. and Steel, M.F.J. (1997): “Bayesian efficiency analysis through individual effects: Hospital cost frontiers,” *Journal of Econometrics*, 76, 77-105.

MacEachern, S.N. (1994): “Estimating Normal means with a conjugate style Dirichlet process prior,” *Communications in Statistics, B*, 23, 727-741.

Meeusen, W. and van den Broeck, J. (1977): “Efficiency estimation from Cobb-Douglas production functions with composed errors,” *International Economic Review*, 8, 435-444.

Park, B.U., Sickles, R.C. and Simar, L. (1998): “Stochastic panel frontiers: A semiparametric approach,” *Journal of Econometrics*, 84, 273-301.

Park, B.U. and Simar, L. (1994): “Efficient semiparametric estimation in a stochastic frontier model,” *Journal of the American Statistical Association*, 89, 929-936.

Ritter, C., Simar, L. (1997): “Pitfalls of Normal-Gamma stochastic frontier models,” *Journal of Productivity Analysis*, 8, 167-182.

Sethuraman, J. (1994): “A constructive definition of Dirichlet priors,” *Statistica Sinica*, 4, 639-50.

Sickles, R.C., Good, D.H. and Getachew, L. (2002): “Specification of distance functions using semi- and nonparametric methods with an application to the dynamic performance of Eastern and Western European air carriers,” *Journal of Productivity Analysis*, 17, 133-155.

van den Broeck, J., Koop, G., Osiewalski, J. and Steel, M.F.J. (1994): “Stochastic Frontier Models: A Bayesian perspective,” *Journal of Econometrics*, 61, 273-303.

Walker, S. and Mallick, B.K. (1997): “A note on the scale parameter of the Dirichlet process,” *Canadian Journal of Statistics*, 25, 473-479.

Zellner, A. (1971): *An Introduction to Bayesian Inference in Econometrics*, New York, Wiley.