Flexible mixture modelling of stochastic frontiers

J. E. Griffin · M. F. J. Steel

Abstract This paper introduces new and flexible classes of inefficiency distributions for stochastic frontier models. We consider both generalized gamma distributions and mixtures of generalized gamma distributions. These classes cover many interesting cases and accommodate both positively and negatively skewed composed error distributions. Bayesian methods allow for useful inference with carefully chosen prior distributions. We recommend a two-component mixture model where a sensible amount of structure is imposed through the prior to distinguish the components, which are given an economic interpretation. This setting allows for efficiencies to depend on firm characteristics, through the probability of belonging to either component. Issues of label-switching and separate identification of both the measurement and inefficiency errors are also examined. Inference methods through MCMC with partial centring are outlined and used to analyse both simulated and real data. An illustration using hospital cost data is discussed in some detail.

Keywords Centring · Efficiency · Generalized gamma distribution · Prior elicitation · Skewness

JEL Classifications C11 · C23 · D24

1 Introduction

The use of stochastic frontier models in productivity analysis and efficiency measurement has seen a steady increase in recent years. The basic idea of a frontier is a characterization of best-practice technology in a particular sector: production frontiers indicate the maximum amount of outputs that can be produced by a certain technology with a given amount of inputs; cost frontiers give the lowest possible cost for the production of a certain amount of outputs with a given level of input prices. Stochastic frontier models refer to an unknown frontier, which has to be inferred from the data. In addition, some firms will not reach the best possible outcome given by the frontier, displaying a lack of efficiency. Thus, one can think of stochastic frontier models as models with unobserved heterogeneity in the location captured through "individual effects", specific to each economic unit and related to its efficiency. In a random effects analysis the distribution assumed for these (positive) effects becomes critical, especially since there are usually only a few observations available on every economic unit. In addition, these effects (and in particular their transformations to efficiencies) are typically the main purpose of the analysis, and inference on efficiencies often depends heavily on these distributional assumptions. In the literature a number of distributions have been proposed for the individual effects. 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 (Stevenson 1980) and the gamma distributions (Greene 1990). Griffin and Steel (2004) use a semiparametric modelling technique to find a distribution-free estimate of the efficiency distribution. They apply this model to the analysis of a cost frontier using a panel dataset of 382 US hospitals observed over 5 years, which will be described in more detail in the application in Sect. 8. The same data were previously analysed in Koop et al. (1997) through a parametric model with an exponential distribution on the individual effects. The results are very different, as shown

J. E. Griffin · M. F. J. Steel (✉)
Department of Statistics, University of Warwick,
Coventry CV4 7AL, UK
e-mail: M.F.Steel@stats.warwick.ac.uk

in Fig. 1, where the predictive[1] semiparametric and exponential parametric efficiency distributions are overplotted. Clearly, that sheds doubt on the exponential assumption for these data. Semiparametric models, however, do need a relatively large amount of data to give sensible results and are somewhat harder to deal with, both computationally and in terms of the interpretation of the results. Predictions are also typically less precise than those of a well-chosen parametric model. Consequently, we will view the semiparametric results as suggesting aspects of the efficiency distribution that need to be modelled. Currently used distributions may struggle to capture some useful shapes and, thus, a wider or alternative class needs to be considered. The resulting stochastic frontier models can be further analysed by either Bayesian or classical methods. It is important to avoid overfitting these particular data by "data-snooping" and, thus, a second analysis using a different dataset is performed (reported briefly in the concluding section), suggesting the general applicability of the methods.

In this paper we propose parametric models which are flexible enough to accommodate a wide range of efficiency distributions and which can capture the possible skewness and multimodality, such as displayed in Fig. 1. The generalized gamma class allows us to consider a wide range of possible shapes in a unified analysis. Many popular choices used in the literature turn out to be special cases and can simply be tested against the more general model using standard Bayesian or classical procedures. In addition, the use of a mixture framework is a structural way to allow for even greater flexibility and leads to a direct and meaningful classification of economic units into efficiency-related categories. Thus, this approach easily lends itself to the identification of outlying firms. Our current two-component mixture model will naturally highlight particularly good firms, but an extension to a three-component model could be used for identifying underperforming firms, e.g. in the context of monitoring clinical performance.

Prior information is critical for stochastic frontier models, where relatively little efficiency information is provided by the data. We provide prior elicitation rules that perform well with both simulated and real data and require only minimal user input. We also design robust and efficient Markov chain Monte Carlo (MCMC) algorithms for Bayesian inference with these models. Previous work (e.g. van den Broeck et al. 1994; Ritter and Simar 1997; Tsionas 2000; Greene 2003; Kozumi and Zhang 2005) has clearly illustrated the challenges of both maximum-likelihood and
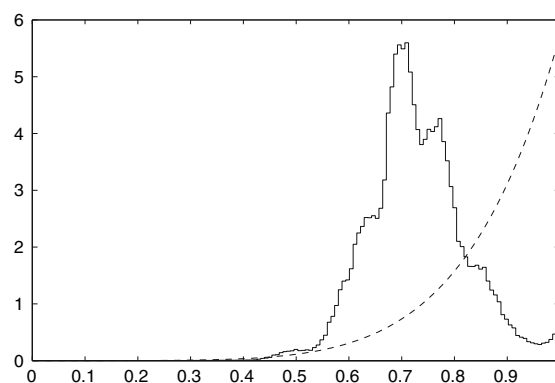

Fig. 1 Predictive efficiency densities for the hospital data. Semiparametric: solid line; parametric exponential: dashed line

Bayesian inference even with the much less flexible gamma inefficiency distribution. Given the large amount of flexibility of the inefficiency distributions suggested here, it is crucial to use priors that are moderately informative in the critical directions.

In the process, we analyse wider issues of imposing problem-specific structure through the prior on random effects models and how to use this structure to tackle the issue of label-switching in mixture models. We also examine the critical role of centring in designing efficient MCMC algorithms for these models.

Section 2 introduces the basic stochastic frontier model with panel data and discusses the lack of flexibility of commonly used distributions on the inefficiencies in capturing the skewness of the data. Section 3 proposes a class of generalized gamma inefficiency distributions as well as an even more flexible class of mixtures of generalized gamma distributions. The prior elicitation for the parameters in these distributions is discussed in Sect. 4. Section 5 presents a decomposition model and also provides a framework for the introduction of firm characteristics into the inefficiency distribution. Inference with these models is conducted through MCMC samplers, and Sect. 6 summarizes some of the main features, with details referred to the Appendix. Section 7 briefly discusses inference with simulated panel data. The hospital data mentioned above are analysed in Sect. 8. A final section concludes.

2 Stochastic frontier models

Let us call the economic units "firms" and focus on a panel context, where we denote by $i = 1, \ldots, n$ the firm index and use subscript $t = 1, \ldots, T_i$ for the time period. Note that the number of observations in time can differ per firm, accommodating unbalanced panels. The typical stochastic frontier model describes the logarithm of cost (or output, in

---

[1] Predictive results are obtained by integrating out the parameters (of the sampling model) using the posterior. In the case of Fig. 1, this leads to the distribution of efficiencies for an unobserved firm in the industry, given all the information in the data.

which case the positive inefficiencies u_i appear with a negative sign in (1)) for firm i at time t, denoted by y_{it}, as

$$y_{it} = \alpha + x'_{it}\beta + u_i + \varepsilon_{it}, \qquad (1)$$

where x_{it} is a vector of appropriate explanatory variables, usually involving logarithms of outputs and prices for cost frontiers and logarithms of inputs for production frontiers. Two commonly used parametric specifications for a cost (production) frontier are a Cobb-Douglas frontier, which is linear in the logs of the outputs and prices (inputs), and a translog frontier, which also includes squares and cross-products of the latter. In addition, (1) allows for measurement and specification error through an i.i.d. symmetric error term ε_{it} assumed to be Normally distributed

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

and the main feature of interest of the model are the one-sided "individual effects" u_i, which measure inefficiency (through the distance to the frontier) and are independently distributed according to some distribution on $\Re_+$. The fact that the inefficiencies u_i are assumed constant over time allows us to exploit the longitudinal or panel structure of the data. The latter assumption can be overly restrictive, especially if T_i is large, and various ways of making the efficiencies time-dependent have been proposed in the literature (see e.g. Cornwell et al. 1990, Kumbhakar 1990 or Battese and Coelli 1992). In this paper, we will not use time-varying inefficiency specifications, as we will allow for quite flexible inefficiency distributions, requiring substantial data information, and we only deal with data that were observed over short periods of time. However, given sufficient data, this framework can be extended to time-varying inefficiencies to accommodate longer panels. Throughout, we assume independence between ε_{it} and u_i.

Our preferred method of inference for these models will be Bayesian. The sampling model 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 with inefficiencies that are constant over time, the posterior distribution exists under the prior

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

in combination with any proper inefficiency distribution. Throughout this paper, we shall use the prior in (2). Without any additional complexity, we can multiply the prior above by an indicator function imposing economic regularity conditions on the frontier. These conditions impose economic theory constraints on β as e.g. we want to exclude production frontiers in which more inputs lead to less output. The prior (2) is invariant with respect to location and scale transformations of the data and is a convenient prior to use in the absence of strong prior information or as a "benchmark" prior.

3 The generalized gamma distribution

For the inefficiency distribution, we propose to use a generalized gamma distribution (Stacy 1962) which is generated by assuming a gamma distribution for powers of the inefficiency u_i, i.e.

$$u_i^c \sim Ga(\phi, \lambda),$$

where $Ga(a,b)$ denotes a gamma distribution with shape parameter a and precision parameter b (i.e. with mean a/b). This leads to the following inefficiency density function:

$$p(u_i | c, \phi, \lambda) = \frac{c\lambda^\phi}{\Gamma(\phi)} u_i^{c\phi-1} \exp\{-u_i^c \lambda\}. \qquad (3)$$

This three-parameter family includes a variety of simpler distributions. In particular, the gamma distribution arises when $c = 1$, the exponential distribution when $\phi = c = 1$, the Weibull distribution for $\phi = 1$ and the half-normal distribution for $c = 2$, $\phi = 1/2$. In general, if $\phi = 1/c$ we generate a class of "half-exponential power distributions", which are univariate versions of exponential power distributions (as in Box and Tiao 1973, Ch. 3) truncated to the positive real line. The distribution in (3) has a bell-shaped density function for $c\phi > 1$ and is reverse J-shaped otherwise. Johnson et al. (1994, p. 389) state that the distribution has positive skewness (as measured by the Pearson skewness measure) for values of $c < c(\phi)$ where $c(\phi)$ depends on ϕ and negative skewness for $c > c(\phi)$. The ability to generate negative skewness with this distribution contrasts with all the commonly used inefficiency distributions. Carree (2002) reports that in many empirical cases the estimated residuals of a stochastic production frontier are positively skewed, while all the usually adopted inefficiency distributions induce negative skewness of the "composed error term" (inefficiency and measurement error). For cost frontiers this problem arises when the residuals are negatively skewed. In particular, the composed error term $u_i + \varepsilon_{it}$ in (1) inherits a Pearson skewness coefficient

$$\gamma_1 = \frac{E[u_i - E(u_i)]^3}{[\sigma^2 + Var(u_i)]^{3/2}},$$

which implies that the skewness has the same sign as that of the inefficiencies and is, therefore, always positive for the usual choices of inefficiency distributions. Thus, in order to deal with both types of skewness in the estimated

residuals, a flexible inefficiency distribution should be able to account for both positive and negative skewness.

As a consequence of the logarithmic transformation of cost or output in (1), the efficiency of firm i is denoted by $r_i = \exp\{-u_i\}$. The distribution of the latter induced by (3) has density on (0,1) given by

$$p(r_i|c, \phi, \lambda) = \frac{c\lambda^\phi}{\Gamma(\phi)} \frac{1}{r_i} (-\ln r_i)^{c\phi-1} \exp\{-(-\ln r_i)^c \lambda\}.$$

(4)

Figure 2 illustrates the wide range of shapes of this implied efficiency distribution for various values of c and ϕ, while fixing $\lambda = 4$. Note that the row $\phi = 1$ corresponds to an efficiency distribution induced by a Weibull inefficiency, the column $c = 1$ presents the gamma case (exponential for $\phi = 1$) and the diagonal where $\phi = 1/c$ displays some half-exponential power cases (one of which corresponds to the half-Normal).

3.1 Mixtures of generalized gamma distributions

The generalized gamma distribution offers more flexibility than the gamma distribution and can capture both positive and negative skewness of the efficiency distribution but cannot capture any possible multimodality of this distribution. In Fig. 1 we observe that the semiparametric model leads to a multimodal efficiency distribution and such behaviour could be modelled using a mixture of generalized gamma distributions. Here, we will adopt a mixture of two components of generalized gamma form. The flexibility of these component distributions will allow this extension to accurately model many shapes. In principle, it would be simple to extend the methods to mixture models with more components, but restricting ourselves to two components has the advantage that it is then rather natural to give structural interpretations to the two possible modes: one mode could correspond to highly efficient firms, and
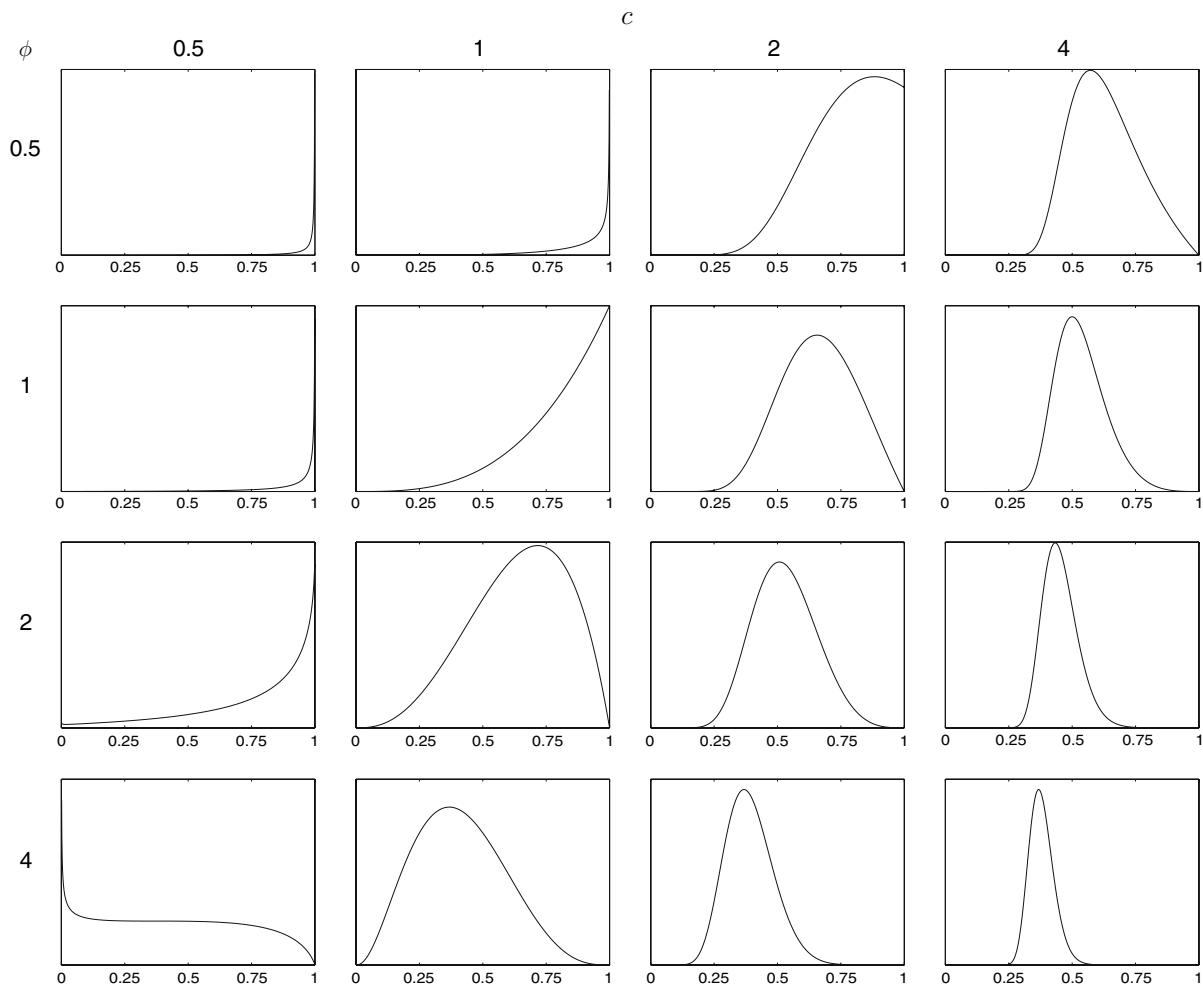

Fig. 2 Densities of the efficiency distribution for $\lambda = 4$ with various combinations of c and ϕ

the other to "mainstream" firms. The two groups could be associated with differences in how effectively a basically common technology is being used, for example by using different management structures. We would expect the mainstream firms to constitute the largest group by far: an industry with a few very efficient firms (and the others slowly catching up) is realistic, but in case there would only be a few laggards, the latter would quickly disappear from the market (at least in competitive situations).

If we denote by $p_{GG}(\cdot|c,\phi,\lambda)$ the generalized gamma density function in (3), the mixed inefficiency distribution is defined by

$$p(u_i|w,\theta) = w\,p_{GG}(u_i|c_1,\phi_1,\lambda_1) + (1-w)\,p_{GG}(u_i|c_2,\phi_2,\lambda_2),$$
(5)

where $w \in [0,1]$ is a weight parameter and $\theta = (c_1,\phi_1,\lambda_1,c_2,\phi_2,\lambda_2)$.

4 Prior distributions on the inefficiency parameters

Let us first focus on the single generalized gamma distribution in (3). As is clear from Fig. 2, the values of the parameters of the inefficiency distribution have a large effect on the shape of the efficiency distribution. Since the prior efficiency distribution is particularly critical for efficiency inference, we need to carefully elicit the priors on c, ϕ and λ. van den Broeck et al. (1994) specify an informative prior distribution for the precision parameter λ for an Erlang distribution (a gamma with integer shape parameter) through the prior predictive median efficiency, say, $r^\star$. This approach can be extended to the generalized gamma distribution proposed in Sect. 3. If we choose $\lambda|c,\phi \sim \mathrm{Ga}(\phi,\lambda_0)$, then after integrating out λ the variable u_i^c has a gamma-gamma distribution (see Bernardo and Smith 1994, p. 120): $u_i^c \sim \mathrm{Gg}(\phi,\lambda_0,\phi)$. If we then make the transformation to

$$h_i = \left(1 + \frac{u_i^c}{\lambda_0}\right)^{-1},$$

we can easily show that h_i has a symmetric Beta distribution, $h_i \sim \mathrm{Be}(\phi,\phi)$. Since the transformation from h_i to the efficiency $r_i = \exp\{-u_i\}$ is monotonic, we can deduce for any $a \in (0,1)$

$$\mathrm{P}(h_i < a) = I_a(\phi,\phi) = \mathrm{P}\!\left(r_i < \exp\left\{-\left[\lambda_0\frac{1-a}{a}\right]^{1/c}\right\}\right),$$
(6)

where $I_a(\phi,\phi)$ is the incomplete Beta function. By using the symmetry and choosing $a = 1/2$, we note that for any value of ϕ the prior median efficiency is $r^\star = \exp\{-\lambda_0^{1/c}\}$ and we can impose a prior median efficiency of $r^\star$ by

simply choosing $\lambda_0 = (-\ln r^\star)^c$ in the prior for λ, i.e. we take

$$\lambda|c,\phi \sim \mathrm{Ga}\big(\phi, (-\ln r^\star)^c\big).$$
(7)

In order to elicit reasonable priors for c and ϕ, we consider what happens for smaller values of a, say, $a = 1/10$ or $a = 1/5$. Given the symmetry of the prior distribution on h_i, and substituting the chosen value for λ_0, we can also write (6) as

$$I_a(\phi,\phi) = \mathrm{P}\!\left(r_i < r^{\star\, b(c,a)}\right) = \mathrm{P}\!\left(r_i > r^{\star\, 1/b(c,a)}\right),$$
$$\text{where } b(c,a) = \left(\frac{1-a}{a}\right)^{1/c}.$$

This makes it clear that prior independence between ϕ and c is not a very natural choice. In particular, for the statement in (6) to make sense, we need to associate large values of ϕ with small values for c. Some experimentation with "reasonable" values of $r^\star$ suggests that ϕ and c should roughly be inversely related. We will build that into our prior by adopting prior independence between c and $\psi \equiv \phi c$. In addition, we want a prior that allows for fairly large values of c and ϕ, while downweighing very small values, which lead to unrealistic efficiency distributions. In particular, we take the product prior

$$\psi \sim \mathrm{Ig}(d_1, d_1+1) \quad \text{and} \quad c \sim \mathrm{Ig}(d_2, d_2+1),$$
(8)

where $\mathrm{Ig}(a,b)$ denotes the inverted gamma distribution with mode $b/(a+1)$ (see Bernardo and Smith 1994, p. 119). Thus, the prior on ψ centres the prior over the half-exponential power case ($\phi = 1/c$), while that on c centres the prior over the gamma case. As a consequence, the prior in (8) is centred over the exponential case ($\phi = c = 1$), but allows for considerable deviations if d_1 and d_2 are not chosen to be too large. As $d_1 \to \infty$ we will restrict ourselves to the half-exponential power class and as $d_2 \to \infty$ the gamma class will be imposed. So we have a prior that is centred over commonly used inefficiency distributions and we only need to chose two hyperparameters that will control the prior spread around these standard classes of distributions. In the sequel, prior sensitivity analysis will be conducted. Figure 3 displays the resulting prior predictive efficiency distribution for $r^\star = 0.6$ and 0.8 and two very different values for d_1 and d_2. The prior predictive for $d_1 = d_2 = 10$ is not far from the predictive generated by an exponential inefficiency distribution, except for the mass close to zero. Clearly, as d_1 and d_2 get smaller, the prior distribution has more flexibility to deviate from its exponential centring distribution. That is also illustrated in Fig. 4 where the distribution of the prior predictive mass assigned to the efficiency interval (0.8, 0.9) is displayed (the right panel displays a truncated version to highlight the tail behaviour). As shown in Griffin and Steel (2004), this
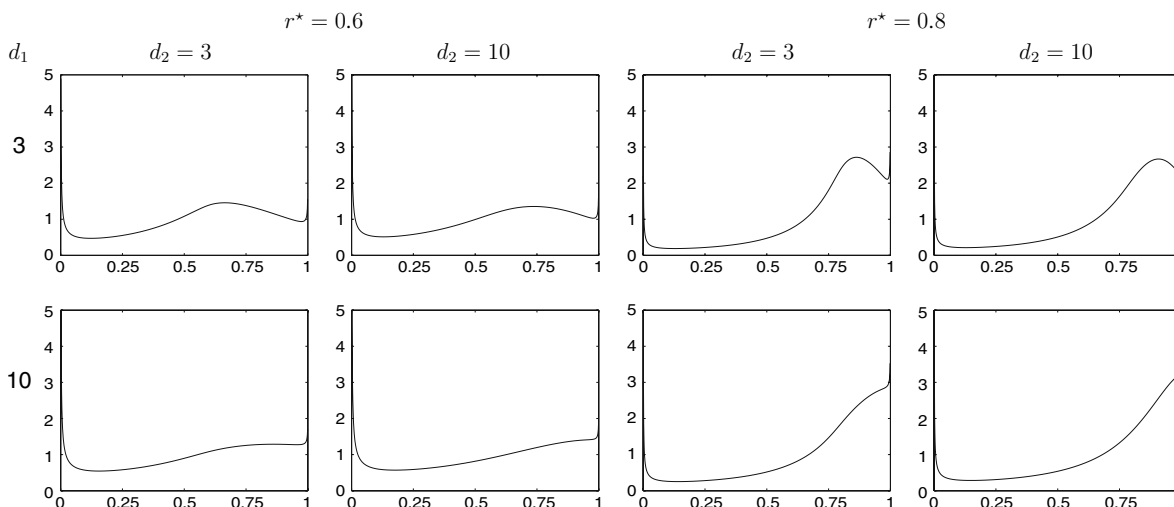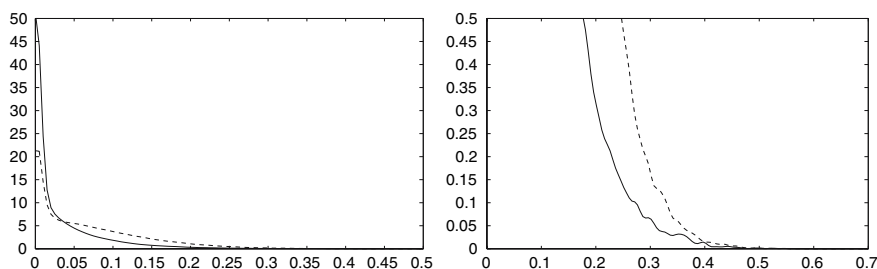
Fig. 3 Prior predictive efficiency density functions with $r^{\star} = 0.6$ (first two columns) and $r^{\star} = 0.8$ (last two columns)

Fig. 4 Prior predictive mass on the interval (0.8, 0.9) with $r^{\star} = 0.8$. Solid line: $d_1 = d_2 = 3$. Dashed line: $d_1 = d_2 = 10$. The right-hand panel displays a trimmed version

mass is truncated to be less than 0.26975 in the case of the exponential inefficiency distribution, which is the limiting case as d_1 and d_2 tend to infinity. For our generalized gamma model, there is no truncation of the mass on any interval. As the values of d_1 and d_2 become smaller, the flexibility of the efficiency distribution grows and more mass is assigned to small probabilities for that interval. The latter is hard to combine with a prior median of 0.8 if the shape is close to exponential.

The parameterization in terms of (ψ, c) is found to lead to better behaviour of the MCMC sampler, and the prior dependence between λ and c and between λ and ψ induced by the prior in (7) is quite small.

In summary, the full prior used here is given by the product of (2) and the prior in (7), (8).

In the mixture of generalized gamma distributions in (5), we could consider making the two components exchangeable by assigning the same prior distribution to their parameters, namely the one explained above. It would then also be most natural to use a symmetric prior on the weight parameter, w. In particular, we assume a beta distribution a priori: $w \sim \mathrm{Be}(w_0, w_1)$, with $w_0 = w_1$. The resulting prior predictive efficiency distribution is then, of course, the same as that for the one-component case.

However, this prior is not really in keeping with a more structural interpretation of these different components, and leaves the door open for the labelling problem that is associated with mixture modelling with symmetric components (see e.g. Celeux et al. 2000). Thus, we shall present a different, more structural, approach in the next section.

5 A decomposition model

Here we formulate a two-component model which is consistent with the structural interpretation of the two components provided in Subsect. 3.1. This, so-called, "decomposition model" combines the mixture model in (5) with a carefully chosen prior for the parameters. In particular, we use the prior from Sect. 4 with a moderate value for $r^{\star}$ on the parameters of the second component, which corresponds to mainstream firms. The first component then reflects the inefficiency distribution of the very efficient group of firms and for its prior we could take *e.g.* $r^{\star} = 0.975$, in combination with fairly large values of d_1 and d_2. The latter make this component behave somewhat like the exponential case, which puts a

substantial amount of mass close to full efficiency while keeping a bounded density function at full efficiency, as illustrated in Fig. 2. This brings this first component in line with the commonly held economic interpretation of the frontier behaviour of a group of (homogeneous) firms: the pioneering contributions in this literature, Aigner et al. (1977) and Meeusen and van den Broeck (1977) use half-Normal and exponential efficiency distributions, respectively. At the same time, this prior helps in distinguishing between the components and in resolving the identification issue mentioned in Sect. 6. This case would suggest using an asymmetric prior for w, which has a relatively large amount of mass close to zero, reflecting a prior belief that only a minority of firms are in the very efficient group. This decomposition prior implies two very different groups of firms and this will make it difficult for the model to fit data that were generated from a unimodal inefficiency distribution. In order to make sure that the model would not artificially introduce a bimodal structure driven merely by the prior, we also allow for the decomposition model to formally reduce to the one-component version by putting prior mass at $w = 0$. Thus, the resulting prior on the weight w for the decomposition model is $w \sim \text{Be}(w_0, w_1)$ with probability $1 - q$ and $w = 0$ with probability q, where we take $w_0 = 0.75$ and $w_1 = 6.75$ (corresponding to a prior mean of 0.1 and a prior standard deviation of 0.2, given the two-component model) and $q = 0.5$ is the prior mass on the one-component model.

Prior predictives for the decomposition mixture are displayed in Fig. 5 for two different values of $d = d_1 = d_2$ corresponding to the second component, while for the first component we fix $d_1 = d_2 = 10$. In addition, we use the value $r^\star = 0.975$ for the first component and $r^\star = 0.8$ for the second one, while for the weight w we take the prior described above. Both prior predictives are quite similar to the one obtained for the exponential inefficiency distribution in Fig. 2.

5.1 Efficiencies varying with firm characteristics

Often, it is reasonable to assume that certain characteristics of the firms can influence the efficiency distributions. This could be accommodated by allowing for groups of firms with similar characteristics to have their own inefficiency distributions, which are somehow related. A linear function of such firm characteristics was used as the mean of a truncated normal inefficiency distribution in Kumbhakar et al. (1991) and Fernández et al. (2002), while Koop et al. (1997) use such a linear function to parameterise the log of the mean of an exponential inefficiency distribution. In a semiparametric context, Griffin and Steel (2004) use a separate Dirichlet process to model the inefficiency distribution for each group of firms, where dependence is introduced through a common parametric model for the centring distributions, chosen as the inefficiency distribution of Koop et al. (1997), and a common mass parameter.

Here, we will explore the use of the mixture inefficiency distributions (Subsect. 3.1) and, in particular, the decomposition model in taking into account firm characteristics. Our decomposition model already has two groups of firms, one very efficient group and one mainstream group. Rather than letting firm characteristics affect the distributions for each of these groups, we will let the probability that a given firm belongs to either group, i.e. the weight w in (5), depend on its characteristics. As we will discuss in the context of our application, the main route through which learning about very high efficiency levels occurs in our models is through the weight. This makes it natural to let the weights depend on firm characteristics, rather than the component parameters.

If we denote by v_i the g-dimensional vector which groups the characteristics of the i-th firm, we model the weight $w(v_i)$ as $w(v_i) = f(v_i'\gamma)$, where $f(\cdot)$ is a monotonic function on $(0,1)$ and γ a parameter vector. A convenient choice for $f(\cdot)$ is the cdf of a standard normal distribution, leading to a probit model for the weights:

Fig. 5 Prior predictive efficiency density functions for the decomposition mixture model with for the first component: $r^\star = 0.975$, $d_1 = d_2 = 10$, and for the second component: $r^\star = 0.8$, $d_1 = d_2 = d$
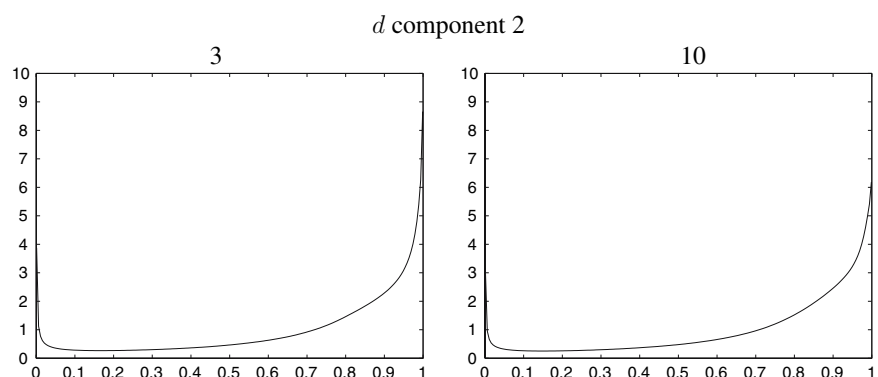
$$w(v_i) = \Phi(v_i'\gamma). \qquad (9)$$

The first element of v_i will be equal to unity, so that the case where $g = 1$ corresponds to the previous model, where firm characteristics are not taken into account. The continuous elements of v_i will be standardized to have zero sample mean and unit sample variance in order to facilitate prior elicitation on the parameter γ. The latter will be assigned a normal prior chosen so as to induce a reasonable prior on the weight for a practically relevant range of values of v_i. In particular, we take the elements of γ to be prior independent with $\gamma_1 \sim N(m_1, s_1^2)$, where m_1 and s_1^2 are chosen in such a way that the prior induced on $w(v_i)$ with $g = 1$ mimics the $Be(w_0, w_1)$ prior on w chosen for the model with constant weights. If $w_0 = w_1 = 1$, this can be done exactly by choosing $m_1 = 0$ and $s_1^2 = 1$. In other cases, we can match the median of $w \sim Be(w_0, w_1)$, which we denote by $Med(w)$, by adopting

$$m_1 = \Phi^{-1}[Med(w)],$$

irrespective of the choice for s_1^2. The latter can be chosen on the basis of the first quartile of w, $Q_1(w)$, which is matched by taking

$$s_1 = \frac{\Phi^{-1}[Med(w)] - \Phi^{-1}[Q_1(w)]}{0.6745},$$

leading to quite an accurate approximation. The prior on the other elements of γ will be $\gamma_i \sim N(0, s^2)$, $i = 2,...,g$, centred over the case where the characteristics have no influence and with a common variance s^2 which is chosen to be relatively small, implying that we need substantial sample information to conclude that a particular characteristic affects the weights.

We will now not include a prior point mass at zero, since the way in which we make the efficiency depend on firm characteristics presupposes two groups. We recommend that users first ascertain whether the two-component case is adequately supported by the data before using the model in this subsection.

Generally, the modeller has to choose whether to introduce covariates into the specification of the frontier or the inefficiency distribution.[2] This is a wider issue that touches upon the definition itself of the frontier: often frontiers are restricted to possess economic regularity conditions, whereas the influence of covariates in the inefficiency distribution can usually be in either direction. Typically, researchers will include covariates that can be assigned regularity conditions (such as actual inputs in a production process) in the frontier, but for other covariates

one has to decide whether they characterize the technology or determine the efficiency. We feel this choice is very much specific to each individual application. Of course, these decisions will affect the definition of the frontier (and, thus, of efficiency) and results should be interpreted with the appropriate definition in mind.

6 Inference

Inference in these models will be based on MCMC methods. MCMC samplers for stochastic frontier models with Erlang inefficiency distributions have been introduced in Koop et al. (1995, 1997), and we shall focus on the main differences with the samplers for these simpler models. For the basic generalized gamma model the sampler will be run over the parameters $(\alpha, \beta, \sigma^2, c, \phi, \lambda)$ augmented with the vector of inefficiencies $u = (u_1,...,u_n)'$.

For successful inference on stochastic frontier models with flexible inefficiency distributions we need to implement the idea of centring. Hierarchical centring was introduced by Gelfand et al. (1995) in the context of normal linear mixed effects models in order to improve the behaviour of maximisation and sampling-based (such as MCMC) algorithms. For (cost) frontiers, it involves a reparameterization from (α,u) to (α,z), where $z = (z_1,...,z_n)'$ and $z_i = \alpha + u_i$. As these models imply that the data are naturally informative on $\alpha + u_i$, identification of α and u_i separately relies on the distribution on u_i. If the latter is quite flexible and can accommodate situations with nearly all the mass far from zero, the "usual" uncentred parameterization can lead to very slow convergence. With the exception of Griffin and Steel (2004), all previous work in stochastic frontiers has used the uncentred parameterization, which can lead to problems even with gamma inefficiency distributions, as used in Tsionas (2000). Ritter and Simar (1997) discuss the problems caused by this identification problem in the context of maximum-likelihood estimation with gamma inefficiency distributions. We avoid these problems by using a hybrid sampler which randomly mixes updates from the centred and the uncentred parameterizations and we find that this dramatically improves the properties of the algorithm. More details on the MCMC sampling procedure used can be found in the Appendix.

As discussed in Sect. 3, the generalized gamma model encompasses a number of commonly used special cases for the inefficiency distribution. The standard Bayesian approach to model selection (and model averaging) is through the use of Bayes factors, defined as the ratio of marginal likelihoods and summarizing the strength of the data support for one model versus another. Typically, calculation of Bayes factors from MCMC output is

---

[2] Actually, from the point of view of statistical identification, they could appear in both places, although this would perhaps be hard to justify from an economic point of view (with the possible exception of time variables).

challenging, but they are easily computed in this case. In particular, as these more restricted classes of inefficiency distributions correspond to particular values for some of the model parameters, we can compute Bayes factors through the Savage-Dickey density ratio (the ratio of the posterior and the prior density values at the restriction; see Verdinelli and Wassermann, 1995). For example, the Bayes factor in favour of the exponential distribution over the generalized gamma distribution will be $p(\psi = 1, c = 1|\text{data})/p(\psi = 1, c = 1)$. Posterior odds between any two models are then immediately obtained by multiplying the prior odds with the appropriate Bayes factor.

For the applications, the results are based on a Markov chain of 330,000 drawings, thinned to every 30th, with a burn-in of 30,000 drawings.

7 Simulated data

As stated before, the prior on the inefficiency distribution is critical to conducting reasonable inference since a certain amount of structure is needed to guide the data. The decomposition case is the one that we considered most promising in the previous sections, but we need to guard against the prior driving the inference in the absence of data information. Thus, we now study the decomposition model in the context of simulated data with a unimodal efficiency distribution. We simulate data for $n = 382$ firms over $T = 5$ periods that are in line with the hospital application. In particular, we use the design matrix of that example as well as the posterior median value of σ

obtained with the one-component model, and adopt a gamma inefficiency distribution with suitably chosen characteristics, namely a Ga(5,15) distribution. Figure 6 displays the posterior distributions of the parameters of both components in the inefficiency distribution. Clearly, the data do not provide much information on the efficient component (component 1). Given the two-component model, the weight on this first component is concentrated around very small values (see Fig. 7(b)), while the weight also gets a point mass at zero of 0.68. Thus, the (correct) one-component model gets a higher posterior probability than the two-component model. The resulting predictive efficiency density in Fig. 7(a) is not far from the true inefficiency distribution. The posterior mode and the lower tail are very well reproduced. There is a bit more mass close to full efficiency, but there is no evidence of a substantial second mode there, driven entirely by the prior. If we wish to use the evidence on w for model choice, we would choose the one-component model, which will lead to a predictive that is even closer to the true distribution. Given the fairly small values of n and T and the challenge of calibrating a prior that should impose structure on the problem in a wide variety of situations, we are quite happy with the behaviour of our decomposition model for inference on these latent efficiencies.

8 Application to Hospital Cost Data

A translog cost frontier analysis for a balanced panel of $n = 382$ nonteaching US hospitals, observed over the years

Fig. 6 Prior and posterior distributions for c and ψ with decomposition prior and simulated data. Solid line: posterior. Dashed line: prior
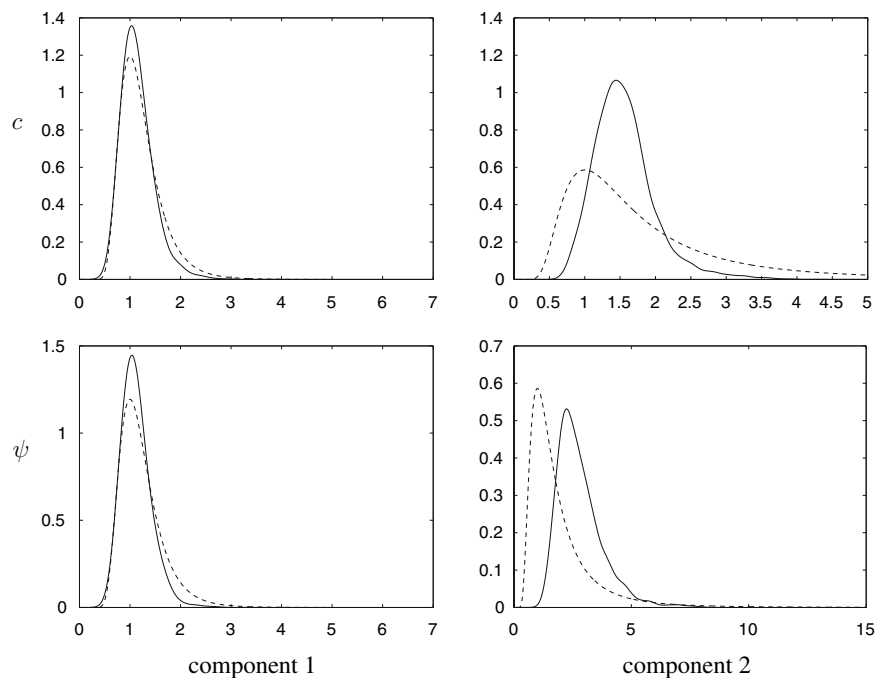
Fig. 7 Results with the
simulated data and the
decomposition prior. (**a**)
Predictive efficiency density;
solid line: predictive; dashed
line: true efficiency distribution.
(**b**) Prior and posterior density
of weight w in the two-
component model; solid line:
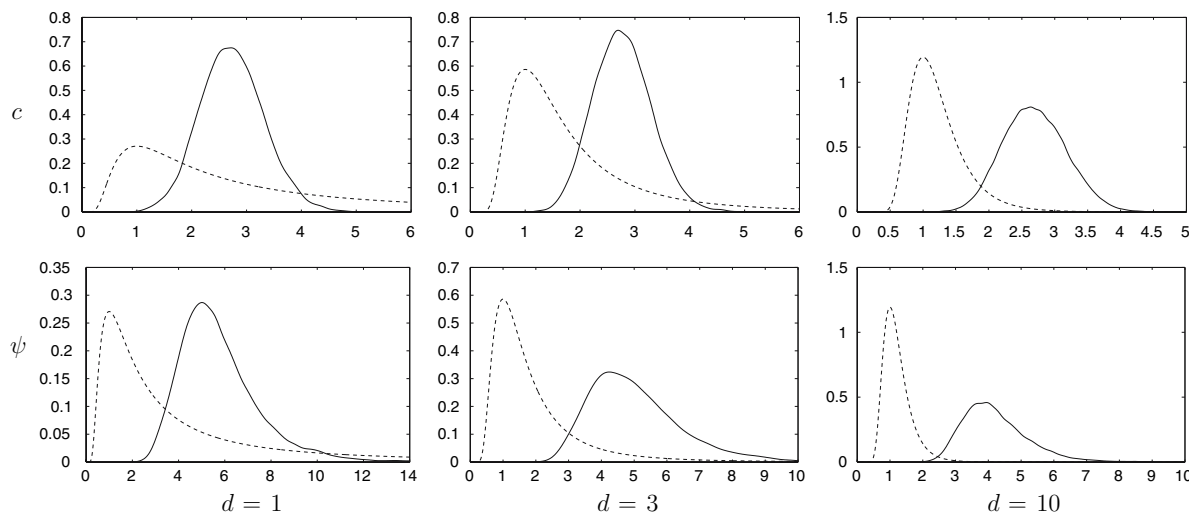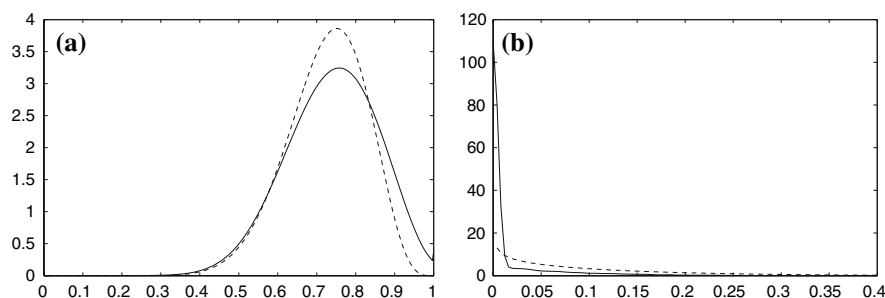posterior; dashed line: prior

Fig. 8 Prior and posterior densities for c and $\psi = c\phi$ with the hospital data; one-component model. Solid line: posterior. Dashed line: prior

1987–1991 ($T = 5$) was presented in Koop et al. (1997) on
the basis of an exponential inefficiency distribution. Details
on the data and general issues of hospital cost function
specification can be found in the latter paper. Griffin and
Steel (2004) use the same data with a nonparametric
inefficiency distribution, modelled through a Dirichlet
process prior. Figure 1 already presented the overall pre-
dictive efficiency distributions resulting from both
analyses.

The cost frontier used here and in the previous papers
involves five different outputs: number of cases, number of
inpatient days, number of beds, number of outpatient visits
and a case mix index. It also has a measure of capital stock,
an aggregate wage index, and a quadratic time trend[3]. A
translog specification with linear homogeneity in prices,
normalized with respect to the price of materials is chosen.

In this section, the main findings are that for the one-
component model only the generalized gamma and Wei-
bull specifications are supported by the data, but, as
expected, that model leads to very little mass close to full

efficiency. The two-component models get much closer to
the nonparametric efficiency distribution, especially the
decomposition model. Letting the weight in the latter
model depend on firm characteristics (ownership status and
staff ratio[4] with interactions) reveals that the probability of
being in the high efficiency group tends to be lowest for
non-profit hospitals with high staff ratios. The latter prob-
ability is highest for government-run hospitals with low
staff ratios.

8.1 The standard generalized gamma case

In the sequel, we shall denote by d a common value of d_1
and d_2 in (8). Throughout this subsection, we will use the
prior median efficiency $r^\star = 0.8$. Figure 8 displays the
prior and posterior densities of c and $\psi = c\phi$. Clearly, the
data do not support values in the neighbourhood of unity
for either of these variables, and are thus in conflict with
the exponential ($c = \psi = 1$) over which the prior is cen-
tred. In addition, we see that d equal to ten is too large, as it

---

[3] Actually, the data do not contain an explicit price for materials, so
we assume that material prices are constant across hospitals but not
over time. The latter dynamics is then approximated by a quadratic
time trend.

[4] This is the ratio of clinical personnel to average daily census,
transformed to a dummy variable.

Table 1 Hospital data: logarithms of the Bayes factors in favour of the restricted models against the one-component generalized gamma model

d	1	3	10
Gamma	−7.7	−22.5	−12.1
Half-exp.power	−28.5	−32.6	− 18.8
Weibull	2.1	−0.1	−6.5
Exponential	−164.7	−151.6	−118.1
Half-normal	−64.3	−53.8	−42.4

makes the prior too tight. Nevertheless, posterior results are only moderately affected (mainly in the right-hand tail) and they are very robust with respect to the choice of d. The case with $d = 3$ leads to a much more reasonable prior, which gives ample scope for the data to suggest departures from the exponential inefficiency distribution. Choosing $d = 1$ leads to an extremely diffuse prior, which was added here for the purpose of prior sensitivity analysis. Support from the data for the special cases of gamma, exponential, half-exponential power, half-Normal and Weibull inefficiencies can be assessed by Bayes factors. The latter are easily computed through Savage-Dickey density ratios (see Sect. 6), and are given in Table 1. Clearly, all special cases, except for the Weibull, are inappropriate for these data.
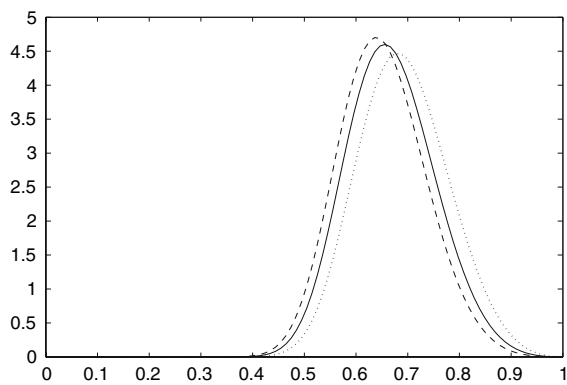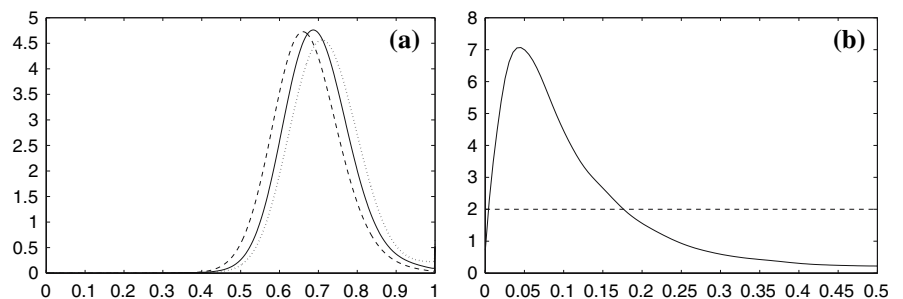

Fig. 9 Posterior predictive efficiency density with the hospital data; one-component model. Dashed line: $d = 1$. Solid line: $d = 3$. Dotted line: $d = 10$

The resulting predictive efficiency distribution is given in Fig. 9 for the three different values for d. As the value of d decreases, the mode is shifted from 0.68 to 0.64 and the shape of the distribution changes from slightly left-skewed to right-skewed. The left-skewness of the efficiency for $d = 10$ is in line with the right skewness of the inefficiency distribution which this tight prior is centred over. For $d = 1$ and 3, the generalized gamma manages to move away from the exponential and the gamma and generates some right skewness for the efficiency. In view of the relative tightness of the prior for $d = 10$, we assign more credibility to the result for smaller d. The latter suggest a slightly right-skewed efficiency distribution with most mass in between 0.45 and 0.9. The small mass assigned to efficiencies above 0.9 may well be considered suspicious. This feature is not in line with the semiparametric predictive in Fig. 1 and could be an artifact of the model with just one component in the efficiency distribution. Let us, therefore, consider the mixture model with two components in the next subsection.

8.2 The mixture case

8.2.1 Symmetric mixture

We first use the model in (5) with exchangeable priors and a Be(1,1), i.e. a uniform, prior distribution on the weight w. Posterior predictive results on the efficiencies are displayed in Fig. 10(a), using the same values for d as in the previous subsection.

From the raw posterior results on the weight w (not presented) it is clear that label-switching occurs. This is a common phenomenon in the case of exchangeable priors on the components, where both the likelihood and the prior are invariant to permutation of the labels (indices); see e.g. Celeux et al. (2000). Figure 10(a) is consistent with the assignment of only a minority of hospitals to the highly efficient group: for values of w smaller than 0.5, the first component corresponds to the efficient group, and for values larger than 0.5, it is the second component that captures the highly efficient firms. This can be inferred

Fig. 10 Posterior results with the hospital data; symmetric mixture. (**a**) Predictive efficiency density; dashed line: $d = 1$; solid line: $d = 3$; dotted line: $d = 10$. (**b**) Posterior density of weight w; solid line: posterior with $d = 3$; dashed line: prior

from the predictive efficiency distribution, which is not affected by the label switching, as (5) is totally invariant with respect to changing the labelling (as long as w is changed to $1 - w$). In order to conduct inference on component-specific parameters or the weight, we need to take the label-switching issue into account. We impose the identifiability constraint that the weight $(1 - w)$ of the mainstream group (corresponding to component 2) is larger than 0.5 (so that $w < 0.5$) and relabel the drawings accordingly. This identifiability constraint is a natural choice in our setting, where one would expect the "mainstream" group to constitute a majority of firms, while the first component refers to a (small) minority of very efficient firms. Richardson and Green (1997) and Stephens (2000) mention that the choice of the particular identifiability constraint can critically change the inference, but here the choice of the labelling convention is inspired by the context. The posterior distribution of the weight w presented in Fig. 10(b) is with this identifiability constraint imposed. In comparison with the one-component model, the resulting efficiency distributions in Fig. 10(a) have more mass close to one now, and bear a stronger resemblance to the semiparametric predictive in Fig. 1.

An interesting finding is that the efficiency distributions are slightly shifted to the right with respect to the one-component case. This also makes them closer to the semiparametric predictive. As the value of d ranges from 1 to 10, there is still a change in shape as before, but the modal values are higher (the internal mode now ranges from 0.67 to 0.71). This is connected to the fact that we now have sufficient mass close to one to "anchor" the

efficiency distribution. If we have little mass in the right-hand tail (as in the one-component case), it is very hard to pin down the value of the intercept α and the modal efficiency has a tendency to drift to lower values. This is linked to the identifiability issue mentioned in Sect. 6. We know from (11) in Appendix A.2 that $\alpha < \min_i\{z_i\}$, where the smallest z_i will correspond to values of the inefficiency u_i close to zero. This immediately implies that if the inefficiency distribution has little mass close to zero (i.e. close to full efficiency), small z_i values will correspond to an area of very low probability and will thus be very volatile. If, on the other hand, a reasonable amount of mass is allowed close to full efficiency, it will help to pin down $\min_i\{z_i\}$ and thus α. The one-component case without strong prior information centring it over a case with mass at full efficiency will be dominated by the main mass of firms and will have a tendency to let the frontier "drift away" and, thus, underestimate the efficiencies. This tendency is counteracted in the two-component case, where the frontier is pinned down by the efficient component, while still retaining the flexibility of accommodating a wide range of shapes and capturing the main mass of less efficient firms.

In these and similar models, Hall and Simar (2002) explore a classical nonparametric estimator of the frontier which performs best when the inefficiency distribution is flat close to zero. This is an alternative approach to dealing with this identifiability issue.

Figure 11 displays the posterior densities for c_i and $\psi_i = \phi_i c_i$, $i = 1,2$, after imposing the identifiability constraint explained above (i.e. that $w < 0.5$). The unimodal shapes suggest that this relabelling convention is quite a

Fig. 11 Hospital data: prior and posterior distributions for c and ψ with exchangeable components and $d = 3$. Solid line: posterior. Dashed line: prior
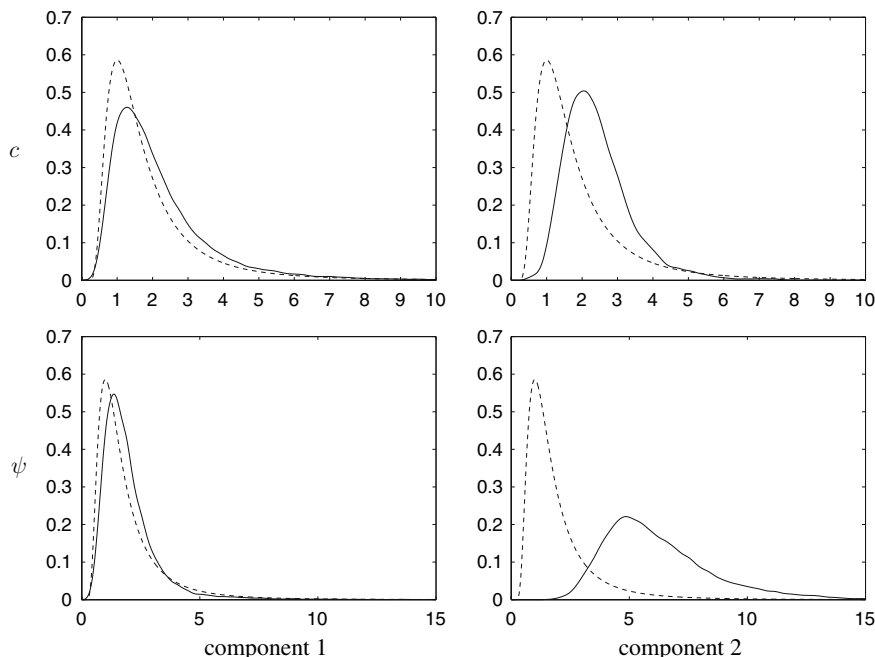
good choice for this problem. Whereas the data clearly are informative on the parameters of the second component (i.e. c_2 and particularly ψ_2), they are less so for those of the first component. Of course, the inference on the latter relies on a relatively small number of firms (the posterior modal value of w is under 0.05, as evidenced by Fig. 10(b)), and the posterior for these parameters is close to the prior. In combination with the efficiency distributions in Fig. 10(a) this suggest a clear posterior separation between the two components, with one corresponding to a typical mainstream hospital and another to a very efficient one, even without reflecting this in the prior.

8.2.2 Decomposition model

Rather than relying on identification through the weight in order to solve the label-switching, an alternative strategy is to avoid label-switching by making the problem sufficiently asymmetric. This is the decomposition idea mentioned in Sect. 5, which we implement by adopting different priors for the components. In particular, we take the prior on w to be a $Be(0.75, 6.75)$ in combination with $P(w = 0) = 0.5$ and we fix the prior median efficiency for the first (efficient) component at $r^\star = 0.975$, while taking $r^\star = 0.8$ for the second component. Even if we allow for substantial variation of the first component away from the exponential case, we achieve clear identification of the two components and no more label-switching occurs. Figure 12 shows the posterior results on the parameters in the case with $d = 10$ for component 1 and $d = 3$ for the second

component. Results are relatively similar to the results obtained with the symmetric prior, except that the posterior for the parameters of the first component is even closer to the (now tighter) prior and the right tails are now thinner throughout. The posterior probability attached to the one-component model is 0.22, so that, indeed, there is even less data information to drive the inference on c_1 and ψ_1. This also indicates that the data (moderately) favour the decomposition model over the one-component model. Figure 13(b) shows the continuous part of the posterior density of the weight, which is more concentrated close to zero than in the symmetric case (after relabelling), in line with the non-uniform prior. Figures 12 and 13(b) suggest that the learning about the highly efficient group in these models occurs through the weight. Generally, there is a trade-off between learning about the component properties and the weight, which is more pronounced for components with a small number of observations. The fairly tight prior distribution used on the parameters of the first component is in line with the basic economic concept underlying stochastic frontier models. In addition, it fits our prior beliefs and helps us to conduct inference on the weight w. The predictive efficiency distribution in Fig. 13(a) is now very close to its semiparametric counterpart in Fig. 1, suggesting that the parametric model we use here is sufficiently flexible to account for the main features of the efficiency distribution.

As indicated in Subsect. 5.1, we can allow for firm characteristics to affect the probability of belonging to the efficient group. Following Koop et al. (1997) and Griffin and Steel (2004), we explore the effect of hospital

Fig. 12 Hospital data: prior and posterior distributions for c and ψ with decomposition prior. Solid line: posterior. Dashed line: prior
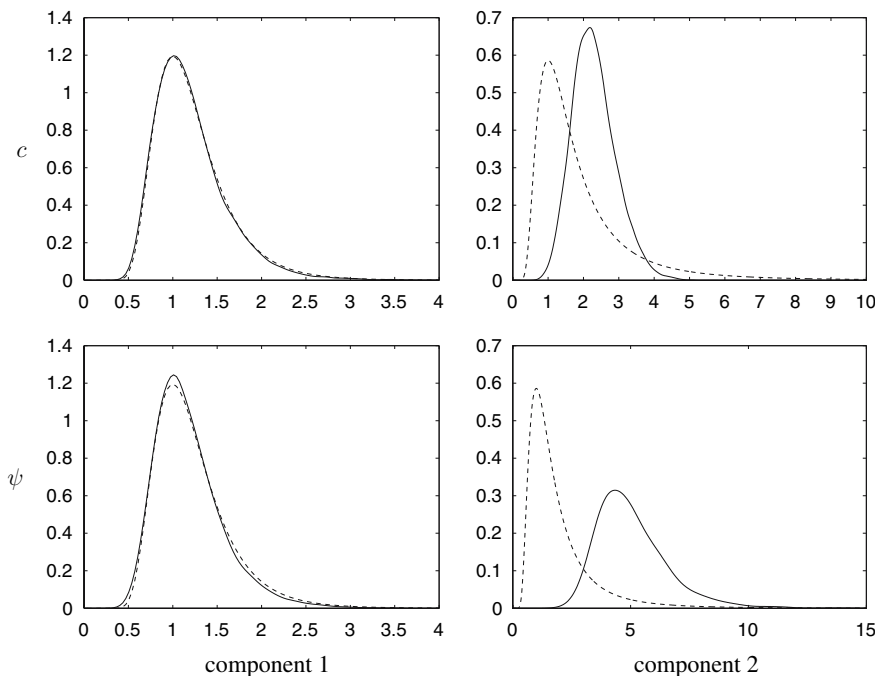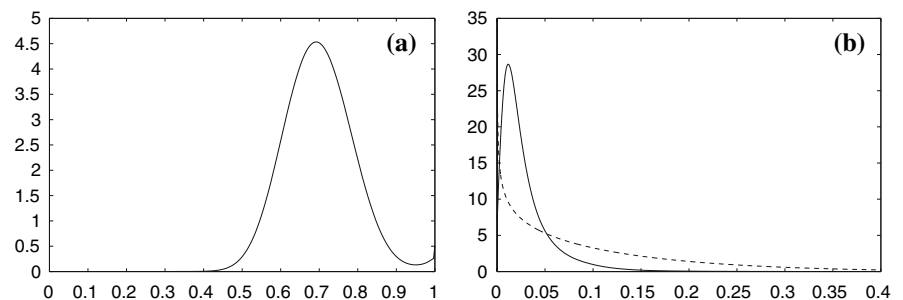
Fig. 13 Posterior results with the hospital data; mixture with decomposition prior. (**a**) Predictive efficiency density. (**b**) Prior and posterior density of weight w in the two-component model; solid line: posterior; dashed line: prior

ownership status (for-profit, non-profit or government-run) and the ratio of clinical personnel to average daily census. For ease of presentation, the latter has been discretized into a dichotomous variable "staff ratio", which takes the value one if this ratio is above average and zero otherwise. In contrast to Koop et al. (1997) we also use interactions of these factors. Table 2 presents some posterior results on the parameter γ in (9) as well as Bayes factors[5] against including each particular explanatory variable. Clearly, the data most support the inclusion of the staff ratio, the higher value of which leads to lower probability of high efficiency, in line with the results in Koop et al. (1997) and Griffin and Steel (2004). There is some evidence (although not clearly supported by the Bayes factor) that for-profit hospitals are slightly less efficient. This was also found in the previous studies, where it was suggested that such a lack of cost efficiency could be related to such hospitals hiring higher levels of staff. Table 3 sheds some light on this and generally puts the results in Table 2 in a more interpretable format by presenting the predictive probabilities of a typical firm being in the efficient group for each possible combination of ownership type and staff ratio. Indeed, as the for-profit firms seem to perform the worst with low staff ratios, they arguably do best when staff ratios are higher than average. Table 2 suggests that the probability of being in the high efficiency group is lowest for non-profit hospitals with high staff ratios and highest for government-run hospitals with low staff ratios. The overall picture here is in accordance with the predictive distributions of the probability that the (semi-parametric) efficiency is in the interval (0.95,1) presented in Figure 12 of Griffin and Steel (2004).

Finally, the sampler also immediately leads to probabilities that a particular firm is assigned to the efficient group, through the indicator variables s_i, $i = 1,\ldots,n$ mentioned in Appendix A.3. These probabilities are graphically displayed in Fig. 14 for all 382 hospitals. It is clear that there are very large differences between hospitals: while the majority belong almost certainly to the mainstream group, a few hospitals have high probabilities

Table 2 Hospital data: probit coefficients for explanatory variables in the weight and logarithms of the Bayes factors in favour of the model without each explanatory variable

	Median	95% credible interval	log Bayes factor
Intercept	−1.7	(−2.8, −0.8)	
For-profit	−0.3	(−2.1, 0.9)	0.3
Government	0.4	(−0.8, 1.4)	0.3
Staff ratio = 1	−1.1	(−2.4, 0.0)	−1.2
For-profit × staff ratio = 1	0.3	(−1.9, 1.9)	0.0
Government × staff ratio = 1	−0.4	(−2.1, 1.1)	0.1

Table 3 Hospital data: predictive medians and 95% credible intervals (in parentheses) for the probability of being in the highly efficient group

	Staff ratio = 0	Staff ratio = 1
Non-profit	0.04 (0.00, 0.22)	0.00 (0.00, 0.04)
For-profit	0.02 (0.00, 0.29)	0.00 (0.00, 0.11)
Government	0.09 (0.00, 0.41)	0.00 (0.00, 0.09)

of being in the efficient group. Interestingly, the effect of the inclusion of explanatory variables is quite marked: whereas the same firms are identified as efficient, the explanatory variables generally allow for higher probabilities of being in the efficient category. The latter suggests a better separation of the two groups, which is not surprising as we effectively use more information to help us in that task.

This approach can be compared with the probabilities that firms' efficiencies are in each of the quintiles of the predictive efficiency distribution, as presented in Griffin and Steel (2004) for the first 40 hospitals in the sample. This is a different way of classifying firms, based on a different model, but results are not dissimilar: for example, hospitals 28 and 31 were identified in Griffin and Steel (2004) as being almost always in the highest quintile, and these are indeed hospitals with unusually high probabilities of being in the efficient group. The probability of belonging to the efficient group can easily be used as a policy tool to identify highly efficient hospitals. In the context of applications where we are primarily interested in

---

[5] These Bayes factors are, again, computed through the Savage-Dickey density ratio, as explained in Sect. 6.

Fig. 14 Posterior probabilities of hospitals to be in the efficient group; mixture with decomposition prior. The hospitals are numbered from 1 to 382 on the horizontal axis. (**a**) Without explanatory variables. (**b**) With explanatory variables
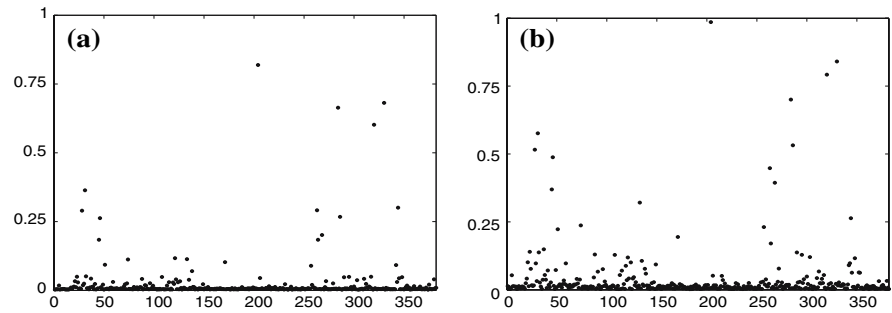

Table 4 Hospital data: medians and first and third quartiles (in parentheses) for the efficiency ranks of selected firms. Models compared are the one-component generalized gamma, the symmetric mixture model and the decomposition mixture model (the latter two also with covariates affecting the mixture weight)

	Gen. Gamma	Symm. Mixture	Symm. Mixt. Cov.	Decomp.	Decomp. Cov.
Min.	1 (1, 1)	1 (1, 2)	1 (1, 2)	1 (1, 2)	1 (1, 2)
12.5th perc.	49 (35, 68)	49 (34, 69)	53 (37, 72)	52 (36, 71)	51 (36, 69)
25th perc.	103 (83, 127)	102 (82, 126)	108 (87, 131)	103 (83, 128)	106 (85, 129)
37.5th perc.	140 (108, 173)	138 (107, 172)	146 (114, 179)	137 (107, 171)	135 (103, 168)
Median	191 (163, 219)	193 (166, 222)	189 (163, 217)	194 (167, 222)	195 (168, 222)
Hosp. A	232 (202, 263)	225 (193, 256)	234 (203, 262)	217 (185, 248)	215 (182, 247)
62.5th perc.	235 (209, 261)	237 (209, 260)	240 (213, 266)	235 (208, 259)	236 (209, 262)
Hosp. B	245 (213, 275)	239 (207, 272)	261 (228, 288)	240 (205, 271)	245 (211, 277)
75th perc.	284 (261, 303)	283 (259, 304)	281 (257, 301)	286 (263, 305)	283 (262, 304)
87.5th perc.	330 (316, 342)	331 (316, 343)	329 (314, 341)	331 (316, 343)	331 (316, 342)
Max.	382 (382, 382)	382 (382, 382)	382 (382, 382)	382 (382, 382)	382 (382, 382)

highlighting underperforming firms, such as in monitoring clinical performance (see Spiegelhalter et al. 2003, for discussion of the Bristol Royal Infirmary and the Harold Shipman Inquiries), we could add a third group for particularly low efficiency firms and use the probability of belonging to that group as an indicator of particularly bad performance.

In practice, we are often interested in ranking units in terms of efficiency, and Table 4 presents some results on the posterior distribution of hospital rankings for various models. Hospitals are selected according to percentiles of the ranking distribution for the symmetric mixture model, while Hospitals A and B reflect cases where the differences in rankings between the models are relatively large. What is clear from Table 4 is that inference on efficiency rankings is fairly precise, especially for extreme firms, and that rankings generally do not differ much between the models proposed here. The use of firm characteristics in the efficiency distribution (indicated in the table by the suffix "Cov.") does seem to have some effect on the ranking, especially for the symmetric mixture model (see Hospital B, in particular), and there are cases (for example, Hospital A) where the decomposition model leads to somewhat different results from the other models. However, as can be expected, the choice of efficiency distribution has much less of an effect on the relative ranking of firms than on the characteristics of predictive efficiency.

9 Concluding remarks

This paper examines stochastic frontier models, which are of particular empirical interest for efficiency measurement in a wide variety of applications. In addition, they are of theoretical interest as they constitute random effects models where the user has a real interest in the random effects and potentially a lot of prior information regarding these effects, which are interpretable as inefficiencies.

We investigate new, more general, classes of inefficiency distributions for these models. In particular, we consider generalized gamma distributions, which incorporate most previously suggested distributions as special cases. We also propose the use of finite mixtures of generalized gamma distributions to allow for more flexible multi-modal efficiency behaviour. The attraction of only having a few components in the mixture is that they are easily given an interpretation. We focus here on two components, where one corresponds to (a minority of) very

efficient firms and the other to the mainstream group of firms. This interpretation is reflected in the prior of our decomposition model. In order to guard against spuriously generating two quite different groups without much support from the data, we build in a prior probability of reducing to the one-component model. Within our mixture model, we also suggest a novel way of introducing firm characteristics into the efficiency distribution. The latter can be a critical policy instrument as it indicates how efficiencies depend on certain firm characteristics.

Throughout the paper, we focus on the importance of the prior, particularly on the parameters of the inefficiency distribution. Whereas we can use convenient improper priors on the other parameters, prior structure on the inefficiencies is critical in conducting sensible inference on the latent efficiencies given the quite small amounts of data information typically available. We believe the prior elicitation strategies outlined here to be sensible ways of extracting information with a minimum of elicitation effort. In particular, we recommend the use of the decomposition model. The latter model produces very reasonable results in the context of an application to hospital cost data. We have also applied this model in the context of a production frontier to an unbalanced panel of data on $n = 613$ highly specialized dairy farms in the Netherlands, observed for some or all of the years 1991–1994, which was previously analysed by Reinhard et al. (1999) and Fernández et al. (2002). Here we model a single output (an aggregate of milk and non-milk production) as in Reinhard et al. (1999) and use family labour, capital and a variable input (including hired labour) as inputs in a translog frontier specification. We use the same prior settings as for the hospital data. For the single-component model the generalized gamma inefficiency distribution again receives most support, with its Weibull special case now in a more distant second place. The data strongly favour a mixture model. The predictive

efficiency distribution resulting from the decomposition model (without the weights depending on firm characteristics) is shown in Fig. 15 and bears a striking similarity to the hospital results, which suggests the general applicability of these methods.

Random effects models have an inherent identification problem, in that the data are naturally informative on the composed error. This seriously complicates inference whenever the random effects are not centred over zero, as in stochastic frontier models or change-point analysis. For stochastic frontier models, these problems are well-documented in e.g. Ritter and Simar (1997) and have led part of the profession to believe there was little point in even attempting to fit models with more flexible inefficiency distributions. This issue is tackled here on two levels. At the modelling level, we recommend a model with two components, allowing for some mass close to the frontier. At the inference level, we use partial centring to improve the properties of the MCMC algorithm.

Mixture models are naturally prone to label-switching issues. The structural interpretation of our components allows for a natural way of relabelling the MCMC output. In the decomposition model the structure induced through the prior will typically introduce enough asymmetry to avoid label-switching altogether and allows effective inference on group membership. The probability of each firm belonging to the efficient group is a useful policy tool as it highlights particularly efficient firms. If the interest is in identifying underperforming firms (say, in the context of clinical performance) we could extend our model to three components, with the added component corresponding to unusually bad firms.

Appendix: Some Details on the MCMC Sampler

A.1 Drawing the inefficiencies

The full conditional for the inefficiencies is, of course, different from the one in the existing literature. The firm-specific inefficiencies u_i will be independent given all the other parameters and the observations, with density function:

$$p(u_i|\text{rest}) \propto u_i^{c\phi-1} \exp\left\{-\left[\frac{1}{2\sigma^2}(T_i u_i^2 + 2\mu_i u_i) + \lambda u_i^c\right]\right\},$$
(10)

where $\mu_i = T_i\alpha + \beta'X_i'\iota - y_i'\iota$ for a cost frontier and its negative for a production frontier, if we define
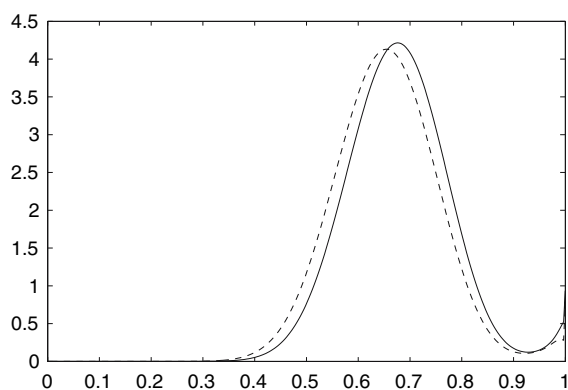

Fig. 15 Predictive efficiency density with the farm data; mixture with decomposition prior. Solid line: $d = 10$ for the second component; dashed line: $d = 3$ for the second component. For the first component we always take $d = 10$

$y_i = (y_{i1}, \ldots, y_{iT_i})'$, $X_i = (x_{i1}, \ldots, x_{iT_i})'$ and ι is a T_i-dimensional vector of ones. We can easily show that the conditional in (10) is log-concave if

$$u^c > \frac{1 - c\phi}{\lambda c(c-1)},$$

which is always satisfied if $c\phi > 1$ and $c > 1$. For this parameter combination, we use adaptive rejection sampling (see Gilks and Wild 1992) to sample directly from (10). In the other cases, we use random walk Metropolis-Hastings with a lognormal candidate generator. Random walk Metropolis-Hastings is also used to generate drawings for the parameters ϕ and c while λ can simply be drawn from a Gamma conditional. We use the reparameterization from (ϕ, c) to (ψ, c), where $\psi = \phi c$, since it leads to better mixing properties of the sampler.

A.2 Centring

As indicated in Sect. 6, we use a sampler which randomly mixes updates from the centred and the uncentred parameterizations. This basic idea is called "partial centring" in Papaspiliopoulos *et al.* (2003), who show that this generally leads to more robust sampling algorithms. We choose a centred update with probability 1/4, as we only need to "recentre" α once in a while. In addition, we found that centring works best if we integrate out λ while updating α (i.e. we effectively draw α and λ jointly). Thus, in the centred parameterization we use a random walk Metropolis-Hastings step to sample from the following conditional for $\alpha < \min_i\{z_i\}$:

$$p(\alpha|z, \beta, \sigma^2, c, \phi, y_1, \ldots, y_n) \propto \prod_{i=1}^{n}(z_i - \alpha)^{c\phi - 1}$$
$$\left[\sum_{i=1}^{n}(z_i - \alpha)^c + (-\ln r^\star)^c\right]^{-\phi(n+1)}, \quad (11)$$

which can be shown to be log-concave whenever $c\phi > 1$ and $c > 1$. We use adaptive rejection sampling in the latter case and a random walk Metropolis-Hastings step otherwise.

A.3 The mixture model

In the case of the mixture inefficiency distribution, as described in Sect. 3.1, we extend the MCMC sampler of the basic case by augmenting with an indicator variable $s_i, i = 1, \ldots, n$ which can take the values 0 or 1 and assigns firm i to one of the two efficiency groups (i.e. one of the two inefficiency components). For the mixture model the sampler will, thus, generate a chain on $(\alpha, \beta, \sigma^2, w, \theta, u, s)$, where $s = (s_1, \ldots, s_n)$. Inference about the parameters of

each component (c_j, ϕ_j, λ_j), $j = 1,2$ now only depends on those firms for which $s_i = j - 1$. The full conditional distribution for w will be

$$w|\text{rest} \sim \text{Be}\left(w_0 + n - \sum_{i=1}^{n} s_i, w_1 + \sum_{i=1}^{n} s_i\right),$$

and the values of s_i are updated through

$$P(s_i = 0|\text{rest})$$
$$= \frac{w p_{GG}(u_i|c_1, \phi_1, \lambda_1)}{w p_{GG}(u_i|c_1, \phi_1, \lambda_1) + (1-w) p_{GG}(u_i|c_2, \phi_2, \lambda_2)}.$$

For the decomposition case, where we have included a prior probability of $w = 0$, the full conditional distribution of w has the same form as above except when $\sum_{i=1}^{n} s_i = n$. In this case w is zero with probability $q^\star$ and $w \sim \text{Be}(w_0, w_1 + n)$ otherwise where

$$q^\star = \frac{q}{q + (1-q)\frac{\Gamma(w_0+w_1)\Gamma(w_1+n)}{\Gamma(w_1)\Gamma(w_0+w_1+n)}}.$$

Here we only considered switching to the one-component model when $\sum_{i=1}^{n} s_i = n$, which worked well. However, in applications with very large n it might be more efficient to also consider jumps between the models when $\sum_{i=1}^{n} s_i$ is relatively close to n.

In case we allow for the probability of being in the efficient group to depend on firm characteristics, we can use simple Gibbs sampling after data augmentation as in Albert and Chib (1993) to update the probit regression coefficients.

References

Aigner D, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. J Econometrics 6:21–37

Albert JH, Chib S (1993) Bayesian analysis of binary and polychotomous response data. J Am Stat Assoc 88:669–679

Battese GE, Coelli TJ (1992) Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. J Prod Anal 3:153–169

Bernardo JM, Smith AFM (1994) Bayesian theory. Wiley, Chichester

Box GEP, Tiao GC (1973) Bayesian inference in statistical analysis. Addison-Wesley, Reading

Carree MA (2002) Technological inefficiency and the skewness of the error component in stochastic frontier analysis. Econ Lett 77:101–107

Celeux G, Hurn M, Robert CP (2000) Computational and inferential difficulties with mixture posterior distributions. J Am Stat Assoc 95:957–970

Cornwell C, Schmidt P, Sickles R (1990) Production frontiers with cross-sectional and time-series variation in efficiency levels. J Econometrics 46:185–200

Fernández C, Koop G, Steel MFJ (2002) Multiple-output production with undesirable outputs: an application to nitrogen surplus in agriculture. J Am Stat Assoc 97:432–442

Fernández C, Osiewalski J, Steel MFJ (1997) On the use of panel data in stochastic frontier models with improper priors. J Econometrics 79:169–193

Gelfand AE, Sahu SK, Carlin BP (1995) Efficient parameterizations for normal linear mixed models. Biometrika 82:479–488

Gilks WR, Wild P (1992) Adaptive rejection sampling for Gibbs sampling. J R Stat Soc C 34:198–200

Greene WH (1990) A gamma-distributed stochastic frontier model. J Econometrics 46:141–163

Greene WH (2003) Simulated likelihood estimation of the normal-gamma stochastic frontier function. J Prod Anal 19:179–190

Griffin JE, Steel MFJ (2004) Semiparametric Bayesian inference for stochastic frontier models. J Econometrics 123:121–152

Hall P, Simar L (2002) Estimating a changepoint, boundary or frontier in the presence of observation error. J Am Stat Assoc 97:523–534

Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions. vol 1, 2nd edn. Wiley, New York

Koop G, Osiewalski J, Steel MFJ (1997) Bayesian efficiency analysis through individual effects: Hospital cost frontiers. J Econometrics 76:77–105

Koop G, Steel MFJ, Osiewalski J (1995) Posterior analysis of stochastic frontiers models using Gibbs sampling. Computational Stat 10:353–373

Kozumi H, Zhang XY (2005) Bayesian and non-Bayesian analysis of gamma stochastic frontier models by Markov chain Monte Carlo methods. Comput Stat 20:575–593

Kumbhakar SC (1990) Production frontiers, panel data and time-varying technical inefficiency. J Econometrics 46:201–212

Kumbhakar SC, Ghosh S, McGuckin JT (1991) A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms. J Bus Econ Stat 9:279–286

Meeusen W, van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed errors. Int Econ Rev 8:435–444

Papaspiliopoulos O, Roberts GO, Sköld M (2003) Non-centered parameterisations for hierarchical models and data augmentation. Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, Smith AFM, West M (eds) Bayesian statistics, vol 7. Clarendon Press, Oxford, pp 307–326

Reinhard S, Lovell CAK, Thijssen G (1999) Econometric application of technical and environmental efficiency: an application to dutch dairy farms. Am J Agr Econ 81:44–60

Richardson S, Green PJ (1997) On Bayesian analysis of mixtures with an unknown number of components (with discussion). J Roy Stat Soc B 59:731–792

Ritter C, Simar L (1997) Pitfalls of Normal-Gamma stochastic frontier models. J Prod Anal 8:167–182

Stacy EW (1962) A generalization of the gamma distribution. Ann Math Stat 33:1187–1192

Stephens M (2000) Dealing with label switching in mixture models. J R Stat Soc B 62:795–809

Stevenson RE (1980) Likelihood functions for generalized stochastic frontier estimation. J Econometrics 13:57–66

Spiegelhalter D, Grigg O, Kinsman R, Threasure T (2003) Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery. Int J Qual Health Care 15:7–13

Tsionas EG (2000) Full likelihood inference in Normal-gamma stochastic frontier models. J Productivity Anal 13:183–205

van den Broeck J, Koop G, Osiewalski J, Steel MFJ (1994) Stochastic frontier models: a Bayesian perspective. J Econometrics 61:273–303

Verdinelli I, Wasserman L (1995) Computing Bayes factors using a generalization of the Savage-Dickey density ratio. J Am Stat Assoc 90:614–618