

## STOCHASTIC FRONTIER MODELS: A BAYESIAN PERSPECTIVE

Julien van den Broeck\* Gary Koop\*\*, Jacek Osiewalski\*\*\* and Mark F.J. Steel\*\*\*\*

### Abstract

---

A Bayesian approach to estimation, prediction and model comparison in composed error production models is presented. A broad range of distributions on the inefficiency term define the contending models, which can either be treated separately or pooled. Posterior results are derived for the individual efficiencies as well as for the parameters, and the differences with the usual sampling-theory approach are highlighted.

The required numerical integrations are handled by Monte Carlo methods with Importance Sampling, and an empirical example illustrates the procedures.

---

### Key words

Efficiency, composed error models, production frontier, prior elicitation

### Acknowledgements

We gratefully acknowledge stimulating discussions with Jacques H. Drèze, Jean-Pierre Florens, Julián de la Horra, Knox Lovell and Michel Mouchart without implicating them. The first author received support from the Belgian Federation for Scientific Research (NFWO, nr.S2/5-ND.E43), the second and third authors acknowledge the hospitality of the Center for Economics Research and of Tilburg University, whereas the fourth author had a fellowship from the Royal Netherlands Academy of Arts and Sciences (KNAW).

\* University of Antwerp, Belgium.

\*\* Department of Economics, Boston University, Boston, U.S.A.

\*\*\* Institute of Statistics, Econometrics and Informatics, Academy of Economics, Kraków, Poland.

\*\*\*\* Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Spain and Department of Econometrics, Tilburg University, The Netherlands.

## 1 Introduction

The use of frontier models, with the associated evaluation of efficiency of the economic units involved has a relatively long-standing tradition in econometrics. We refer the reader to Bauer (1990) for a survey of the literature. The production or cost frontier itself is not supposed to be known exactly to the econometrician and, generally, some symmetric error is attached to it. A second type of error is introduced to represent deviations of the individual units from the frontier (i.e. inefficiency), and is thus, by definition, of a one-sided nature.

This, so-called, composed error framework was first introduced in Meeusen and van den Broeck (1977) and Aigner, Lovell and Schmidt (1977) and, as Bauer (1990, p. 41) remarks "the basic set of econometric estimation techniques has changed relatively little in recent years". In fact, to the knowledge of the authors, the entire literature has been embedded in the sampling theory paradigm.

Here we reconsider the basic composed error model with different sampling distributions on the efficiency term, the exponential of Meeusen and van den Broeck (1977), the half-Normal of Aigner, Lovell and Schmidt (1977), the truncated Normal proposed by Stevenson (1980) and a gamma based on Greene (1990), in a formal Bayesian framework.

The costs associated with this shift of paradigm are mainly of a computational nature, but the gains are shown to be worth the effort. We formally take parameter uncertainty into account in deriving posterior densities for the efficiencies. In addition, we can easily treat our uncertainty concerning which sampling model to use by mixing over them with posterior model probabilities as weights. This pooling approach is quite natural in a Bayesian analysis. Given a particular model, we assume that all efficiencies are drawn from the same simple distribution, only we do not know which one. We avoid mixing different distributions for each individual efficiency within the sampling model, as this would render the analysis intractable. Rather, we deal with simple, tractable, models and mix over them at the final stage. In case we wish to choose a particular distribution for the efficiency term, we can use Bayes factors or posterior odds as a criterion for model selection.

We propose prior structures on the model-specific parameters  $\delta_j$  that do not affect the salient features of the different sampling models and derive a simple procedure for prior elicitation, based on only the prior median efficiency.

From an economic point of view, the need to survive in a competitive environment of most economic units induces a belief that many of them are close to the frontier, i.e. full efficiency. However, given the dynamic character of competition itself, strategic policies in the long run (secular inefficiency) could keep units away from their frontier. In many cases, this will be compounded with organisational inefficiency in the short run [see van den Broeck et al. (1991)]. Although these economic considerations guide us in forming our prior ideas concerning efficiency, they do not provide us with exact functional forms for the distribution of efficiencies. The latter uncertainty will here be resolved by mixing over a number of contending efficiency distributions.

An empirical application on the data set of Greene (1990) leads us to address some numerical issues, which might have influenced Greene's (1990) results. We analyse the gamma models under a diffuse improper prior on the parameter of the efficiency sampling distribution, and use our proper prior structures on  $\delta_j$  to compare them with the truncated Normal model and the model which imposes full efficiency.

## 2 Model framework

We consider a Bayesian approach to stochastic frontier models of the form

$$y_i = h(x_i, \beta) + v_i - z_i, \quad (1)$$

where  $y_i$  denotes the log of the output variable (or the negative of log cost) for firm  $i$  ( $i = 1, \dots, N$ ),  $x_i$  denotes the column vector of logs of exogenous variables,  $v_i$  is a symmetric disturbance capturing measurement error of the stochastic frontier

$$\tilde{y}_i = h(x_i, \beta) + v_i, \quad (2)$$

and  $z_i$  is a nonnegative disturbance modelling the level of inefficiency.  $v_i$  and  $z_i$  are independent of each other and across firms. This is the composed error model framework proposed by Meeusen and van den Broeck (1977) and Aigner, Lovell and Schmidt (1977). Any parametric Bayesian approach requires full specification of the likelihood function. In order to meet this

requirement we assume that  $h(\cdot, \cdot)$  is known and is a measurable function of  $\beta \in B \subseteq R^k$ , and that  $v_i$  is Normally distributed with mean 0 and variance  $\sigma^2$ .<sup>1</sup> However, as far as  $z_i$  is concerned, we shall consider several one-sided distributions, leading to different (competing) statistical models. By attaching prior probabilities to these models, our Bayesian approach naturally leads to their posterior probabilities, that tell us which of the competing models is most favoured by the data and also enable us to pool posterior inferences (about quantities of interest) based on different distributional assumptions about  $z_i$ . Rather than artificially imposing a specific form for the density function of  $z_i$ , we average out model uncertainty.

Since the exponential distribution of  $z_i$ , proposed by Meeusen and van den Broeck (1977), has a considerable theoretical appeal, we will assume it in our model 1,  $M_1$ . However, the zero mode of the exponential distribution (i.e. a gamma distribution with shape parameter 1 and unknown scale parameter  $\lambda$ ) seemed to be too strict a restriction for some authors. Stevenson (1980) considered gamma distributed  $z_i$  with shape parameter 2 or 3 as well, and Beckers and Hammond (1987) and Greene (1990) have recently presented maximum-likelihood estimation for the case when  $z_i$  is gamma distributed with unknown shape parameter  $P \in R_+$ . In his application Greene (1990) finds  $\hat{P} = 2.45 (\pm 1.10)$  as the ML estimate of  $P$ . In the case of non-integer  $P$ , the ML approach requires the evaluation of integrals which have no closed form solution and for which there are no polynomial approximations available.

Note that considering a gamma distribution with high  $P$  makes the shapes of the densities of  $v_i$  and  $z_i$  hardly distinguishable. Here, in order to clearly distinguish between  $z_i$  and  $v_i$ , and make computations relatively simple,<sup>2</sup> we adopt gamma distributions with fixed small integer values of the shape parameter as competing hypotheses. Therefore, under  $M_j$

$$p_j(z_i|\theta_j) = f_G(z_i|j, \lambda^{-1}) = \frac{\lambda^{-j}}{\Gamma(j)} z_i^{j-1} \exp(-\lambda^{-1} z_i) I(z_i \geq 0)$$

where  $\theta_j \in \Theta_j$  generally denotes the parameters in  $M_j$ ,  $I(\cdot)$  is the indicator function, and  $\lambda > 0$  is one of the parameters in  $\theta_j$ ,  $j = 1, 2, 3$ . Note that for  $j = 2$  the shape of the density function of  $z_i$  is already completely different from that in the exponential case ( $j = 1$ ), and thus we will restrict ourselves to  $j = 1, 2, 3$ . In the statistical literature, such distributions are

<sup>1</sup>As we show in appendix A. the assumption of Normality of  $v_i$ 's can be replaced by their joint sphericity, and all our results (except those relating to  $\sigma^2$ ) hold.

<sup>2</sup>Note that our Bayesian approach requires Monte Carlo integration over the whole parameter space even in the simplest case ( $M_1$ ).

referred to as Erlang distributions [see Johnson and Kotz (1970, p. 166)]. As an additional "baseline" model ( $M_0$ ) we will consider the standard ("average") model which does not allow for inefficiency considerations and formally corresponds to  $z_i \equiv 0$ . We will also consider, as  $M_4$ , a two-parameter truncated Normal distribution of  $z_i$ , proposed by Stevenson (1980).

The baseline model,  $M_0$ , is the simplest case without composed error, i.e. where all  $z_i$ 's are zero and thus all firms are on the production frontier, which is itself stochastic. We stress that this model is clearly not meant to describe reality, but we treat it as a benchmark against which the improvements of our various composed error models will be evaluated. The model will thus correspond to the following data density<sup>3</sup>

$$p_o(y_i|x_i, \theta_o) = f_N^1(y_i|h(x_i, \beta), \sigma^2),$$

which will be combined with a prior density  $p_o(\theta_o) = p_o(\beta, \sigma^2)$  in the usual fashion to obtain Bayesian posterior and predictive results.

Since we assume independence over firms, the likelihood function under  $M_j$  is generally given by

$$\ell_j(\theta_j|y, X) = p_j(y|X, \theta_j) = \prod_{i=1}^N p_j(y_i|x_i, \theta_j) \quad (3)$$

where  $y = (y_1, \dots, y_N)'$  and  $X$  is the matrix  $(x_1, \dots, x_N)'$ . Under a prior  $p_j(\theta_j)$ , the posterior density will be given by

$$p_j(\theta_j|y, X) = K_j^{-1} p_j(\theta_j) \ell_j(\theta_j|y, X) \quad (4)$$

where we have defined

$$K_j = p_j(y|X) = \int_{\Theta_j} p_j(\theta_j) \ell_j(\theta_j|y, X) d\theta_j. \quad (5)$$

In the special case of e.g. a Cobb-Douglas production technology, which implies a linear  $h(x_i, \beta)$ , and a natural-conjugate or Jeffreys' diffuse prior for  $\theta_o$ , we can analytically evaluate posterior and predictive densities under  $M_0$ .

This, however, will not be the case for the composed error models  $M_j$ ,  $j = 1, 2, 3, 4$ .

<sup>3</sup>Appendix B describes the density functions used in this paper.

### 3 Posterior and predictive results under Erlang inefficiency term

If  $M_j$ ,  $j \in \{1, 2, 3\}$ , is the model considered, then the joint distribution of  $y_i$  and  $z_i$ , given  $x_i$  and the parameters  $\theta_j$  is:

$$p_j(y_i, z_i | x_i, \theta_j) = f_N^1(y_i | h(x_i, \beta) - z_i, \sigma^2) f_G(z_i | j, \lambda^{-1}), \quad (6)$$

where the parameters are not model specific, i.e. for  $j = 1, 2, 3$ :  $\theta_j = \theta_* = (\beta, \sigma^2, \lambda) \in \Theta_* \subseteq R^k \times R_+ \times R_+$ . For the purpose of inference, the opposite factorization is of interest. It can be easily deduced from (6) that the conditional density of  $z_i$  given  $(y_i, x_i, \theta_j)$  is

$$p_j(z_i | y_i, x_i, \theta_j) = w_{ji}^{-1} \left[ \Phi \left( \frac{m_i}{\sigma} \right) \right]^{-1} z_i^{j-1} f_N^1(z_i | m_i, \sigma^2) I(z_i \geq 0), \quad (7)$$

where

$$m_i = m(y_i, x_i, \theta_*) = h(x_i, \beta) - y_i - \sigma^2/\lambda. \quad (8)$$

$\Phi(\cdot)$  denotes the distribution function of  $N(0, 1)$ , and  $w_{ji} = w_j(y_i, x_i, \theta_*)$  is the appropriate integrating constant of (7). Remark that if  $j = 1$ , then (7) is just a truncated Normal distribution, and  $w_{1i} = 1$ . This result was obtained by Jondrow et al. (1982). For  $j = 2, 3$ ,  $w_{2i}$  and  $w_{3i}$  are, respectively, the first and the second order moments of that truncated Normal distribution, and therefore

$$w_{2i} = m_i + \sigma \frac{f_N^1 \left( \frac{m_i}{\sigma} | 0, 1 \right)}{\Phi \left( \frac{m_i}{\sigma} \right)}, \quad (9)$$

$$w_{3i} = \sigma^2 + m_i w_{2i}. \quad (10)$$

Let  $c_{ji} = w_{ji} \Phi(m_i/\sigma)$ . Now, from (6) and (7), the sampling density of  $y_i$  can be written as

$$p_j(y_i | x_i, \theta_j) = \frac{\lambda^{-j}}{\Gamma(j)} \exp \left( -\frac{m_i}{\lambda} - \frac{\sigma^2}{2\lambda^2} \right) c_{ji}. \quad (11)$$

Note that (11) is equivalent to Greene's (1990) general formula (26). However, the correct forms of the data densities for the special cases with  $j = 2$  and  $j = 3$  have not been presented in the literature. Although in the exponential case ( $j = 1$ ) our (11) is equivalent to Stevenson's (1980) formula (20), his (21) and (22) for two other Erlang cases cannot be deduced from our (11) and (9)-(10).

The likelihood function under  $M_j$  ( $j = 1, 2, 3$ ) is a product of the densities in (11):

$$\ell_j(\theta_j|y, X) = \ell_j(\theta_*|y, X) = \left[ \frac{\lambda^{-j}}{\Gamma(j)} \right]^N \exp \left( -\frac{N\sigma^2}{2\lambda^2} - \frac{1}{\lambda} \sum_{i=1}^N m_i \right) \prod_{i=1}^N c_{ji}, \quad (12)$$

and under any prior density  $p_j(\theta_j) = p_j(\theta_*)$  the joint posterior density, obtained as in (4) and (5), is too complicated for analytical integration, and thus posterior moments and marginal densities will be calculated by  $(k + 2)$ -dimensional Monte Carlo integration with importance sampling.

The posterior distribution summarizes all the information about the parameters which is contained in the prior and in the observed data. In frontier analysis, however, the main interest is not in the parameters themselves, but in the individual efficiency (of the firm  $i$ ,  $i \in \{1, \dots, N\}$ ) measured by  $r_i = \exp(-z_i)$ .<sup>4</sup> The conditional posterior distribution of  $z_i$  given  $y_i$ ,  $x_i$  and the parameters, presented in (7), exactly corresponds to Greene's (1990) conditional distribution (27), and constitutes the common starting point for inference about individual (in)efficiencies in both sampling-theory and Bayesian approaches to stochastic frontier models. The difference lies in the treatment of the unknown parameters in (7). The sampling-theory approach of Jondrow et al. (1982) and Greene (1990) amounts to conditioning on  $\hat{\theta}_j$ , an estimate of  $\theta_j$ , whereas our Bayesian approach naturally averages out uncertainty about  $\theta_j$  by marginalizing (7) w.r.t. the posterior density of the parameters. This leads to the following posterior densities and moments

$$p_j(z_i|y, X) = \int_{\Theta_j} p_j(z_i|y_i, x_i, \theta_j) p_j(\theta_j|y, X) d\theta_j, \quad (13)$$

$$p_j(r_i|y, X) = r_i^{-1} p_j(z_i|y, X), \quad r_i \in (0, 1], \quad (14)$$

<sup>4</sup>Our model (1) is formulated in terms of logs of variables.

$$E_j[g(z_i)|y, X] = \int_{\Theta_j} E_j[g(z_i)|y_i, x_i, \theta_j] p_j(\theta_j|y, X) d\theta_j, \quad (15)$$

which summarize the evidence on the individual within-sample efficiency under  $M_j$ . In (14),  $p_j(z_i|y, X)$  is evaluated at  $z_i = -\ln r_i$ . In (15),  $g(z_i)$  is equal to  $z_i^q$  or  $r_i^q$  for obtaining posterior moments, and  $(k+3)$ -dimensional integration over  $\theta_j = (\beta, \sigma^2, \lambda)$  and  $z_i$  will be performed by Monte Carlo. Note that in the composed error case linearity of  $h(x_i, \beta)$  no longer leads to analytical results. The dimension of  $\beta$ ,  $k$ , is what really matters in our numerical integration.

Apart from the within-sample posterior analysis, some predictive analysis of an out-of-sample (maybe hypothetical) firm will usually be of interest. Assuming that (1) and (6) hold for some  $i > N$ , maintaining independence across all the firms under consideration, and using  $f$  to index the unobserved (forecasted) firm, we obtain

$$p_j(y_f|x_f, y, X) = \int_{\Theta_j} p_j(y_f|x_f, \theta_j) p_j(\theta_j|y, X) d\theta_j, \quad (16)$$

$$p_j(z_f|y, X) = \int_0^\infty f_G(z_f|j, \lambda^{-1}) p_j(\lambda|y, X) d\lambda, \quad (17)$$

as the predictive densities of the log of the actual output ( $y_f$ ) and its inefficiency component ( $z_f$ ), respectively. In (16),  $p_j(y_f|x_f, \theta_j)$  is the sampling density (of  $y_f$  given exogenous  $x_f$ ) of the same form as in (11) but with  $i = f$ , whereas in (17)  $p_j(\lambda|y, X)$  is the univariate marginal posterior density calculated from  $p_j(\theta_j|y, X)$ .

Note that (13) and (17) differ substantially in their construction and interpretation. (17) is posterior to the data on all the observed firms, but prior to the (yet unobserved) output of the firm  $f$ , and therefore can be treated as a Bayesian counterpart of the classical characteristics of "average" inefficiency (as opposed to individual inefficiency). Again, we marginalize w.r.t.  $\lambda$ , using its posterior density, whereas the standard non-Bayesian practice would be to replace  $\lambda$  by its estimate,  $\hat{\lambda}$ . Useful summary measures of the "average" efficiency are given by the first and second order moments of  $r_f = \exp(-z_f)$  given  $y, X$ . Since, for  $q > 0$ ,

$$\int_0^\infty \exp(-qz) f_G(z|j, \lambda^{-1}) dz = (1 + q\lambda)^{-j},$$

therefore, for  $j = 1, 2, 3$ ,



$$E_j(r_f^q|y, X) = \int_0^\infty (1 + q\lambda)^{-j} p_j(\lambda|y, X) d\lambda. \quad (18)$$

More directly interpretable posterior characteristics of both the individual and "average" efficiency will be given by the quantiles (e.g. median) of  $p_j(r_i|y, X)$ ,  $i \in \{1, \dots, N\}$ , and  $p_j(r_f|y, X)$ ,  $f \notin \{1, \dots, N\}$ , respectively.

#### 4 Truncated Normal distribution of the inefficiency term

Let us now present briefly the Bayesian analysis of the stochastic frontier model (1) with  $z_i$  distributed as  $N(\mu, \omega^2)$  truncated at 0, and with  $\mu \in R$  and  $\omega \in R_+$ , both unknown.

This model, denoted by  $M_4$  here, was proposed by Stevenson (1980) who generalized the original (half-Normal) model of Aigner et al. (1977) where  $z_i \sim |N(0, \omega^2)|$ .

Reparameterizing  $(\mu, \omega) \rightarrow (\psi, \omega)$  with  $\psi = \mu/\omega$  will prove useful.  $\psi$  indicates how many standard deviations the truncation point 0 is from the mean of the underlying Normal distribution.

Under  $M_4$  the joint distribution of the observed  $y_i$  and the unobservable  $z_i$ , given exogenous  $x_i$ 's and the parameters  $\theta_4 = (\beta, \sigma^2, \psi, \omega) \in \Theta_4 \subset R^{k+3}$ , takes the form

$$p_4(y_i, z_i|x_i, \theta_4) = f_N^1(y_i|h(x_i, \beta) - z_i, \sigma^2) \frac{f_N^1(z_i|\psi\omega, \omega^2)}{\Phi(\psi)} I(z_i \geq 0) \quad (19)$$

which can be rearranged as

$$p_4(y_i, z_i|x_i, \theta_4) = f_N^1(y_i|h_i - \psi\omega, \omega^2 + \sigma^2) f_N^1\left(z_i \left| \frac{\sigma^2\psi\omega - \omega^2(y_i - h_i)}{\omega^2 + \sigma^2}, \frac{\sigma^2\omega^2}{\omega^2 + \sigma^2} \right. \right) \frac{I(z_i \geq 0)}{\Phi(\psi)},$$

where  $h_i = h(x_i, \beta)$ . Integration over  $z_i \in [0, +\infty)$  leads to the sampling density of  $y_i$  under  $M_4$ , i.e.

$$p_4(y_i|x_i, \theta_4) = [\Phi(\psi)]^{-1} \Phi\left[\frac{\sigma^2\psi - \omega(y_i - h_i)}{\sigma\sqrt{\omega^2 + \sigma^2}}\right] f_N^1(y_i|h_i - \psi\omega, \omega^2 + \sigma^2), \quad (20)$$

which is in agreement with formula (5) of Stevenson (1980). Therefore the conditional density of  $z_i$ , given  $y_i$ ,  $x_i$  and the parameters is

$$p_4(z_i|y_i, x_i, \theta_4) = \left[ \Phi \left( \frac{\sigma^2 \psi - \omega(y_i - h_i)}{\sigma \sqrt{\omega^2 + \sigma^2}} \right) \right]^{-1} f_N^1 \left( z_i \left| \frac{\sigma^2 \psi \omega - \omega^2 (y_i - h_i)}{\omega^2 + \sigma^2}, \frac{\sigma^2 \omega^2}{\omega^2 + \sigma^2} \right. \right) I(z_i \geq 0), \quad (21)$$

a truncated Normal density, which for  $\psi = 0$ , i.e. for the half-Normal distribution of  $z_i$  given the parameters, reduces to the density obtained in Jondrow et al. (1982, Appendix).

Note that, since  $-z_i$  can be interpreted as the unknown negative mean of the Normal error  $y_i - h(x_i, \beta)$ , therefore, apart from truncation, the conditional posterior of  $z_i$  (given the parameters) shares the well-known properties of the posterior of the Normal mean under a Normal prior. This untruncated posterior of  $z_i$  is Normal, its precision, equal to  $\frac{\omega^2 + \sigma^2}{\sigma^2 \omega^2}$ , is a sum of the "sampling precision" ( $1/\sigma^2$ ) and the "prior precision" ( $1/\omega^2$ ), and its mean is a weighted average of the "prior mean" ( $\mu = \psi\omega$ ) and the "observation" ( $h_i - y_i$ ) with the weights equal to the proportions of the prior and sampling precisions in the total (posterior) precision.

The Bayesian analysis goes along the same lines as under  $M_1$ ,  $M_2$  or  $M_3$ , i.e. we use (3), (4) and (5) for  $j = 4$ .

Inferences about individual efficiencies of the observed firms will be based on the marginal posterior densities of  $r_i$ 's ( $i = 1, \dots, N$ ), i.e.

$$p_4(r_i|y, X) = \frac{1}{r_i} \int_{\Theta_4} p_4(z_i|y_i, x_i, \theta_4) p_4(\theta_4|y, X) d\theta_4,$$

where  $p_4(z_i|y_i, x_i, \theta_4)$ , the density in (21), should be evaluated at  $z_i = -\ln(r_i)$ . The posterior density and moments of the efficiency  $r_f$  of a yet unobserved (or "average") firm  $f$  will be calculated from

$$p_4(r_f|y, X) = r_f^{-1} \int_0^\infty \int_{-\infty}^{+\infty} [\Phi(\psi)]^{-1} f_N^1(-\ln r_f | \psi\omega, \omega^2) p_4(\psi, \omega|y, X) d\psi d\omega I(0 < r_f \leq 1),$$

and

$$E_4(r_f^q|y, X) = E_4(e^{-qz_f}|y, X) = \int_0^\infty \int_{-\infty}^\infty \frac{\Phi(\psi - q\omega)}{\Phi(\psi)} \exp \left[ q\omega \left( \frac{q\omega}{2} - \psi \right) \right] p_4(\psi, \omega|y, X) d\psi d\omega,$$

where  $p_4(\psi, \omega|y, X)$  is the posterior density of  $(\psi, \omega)$ , obtained as a marginal density of

$p_4(\theta_4|y, X)$ . Under  $M_4$ , the Bayesian analysis will require  $(k + 3)$ -dimensional Monte Carlo integration for calculating posterior results on the parameters, and  $(k + 4)$ -dimensional integration for obtaining both individual and "average" posterior characteristics of (in)efficiency.

## 5 Comparing models and pooling inferences

The Bayesian approach to comparing alternative models is often based on the posterior odds ratio, i.e.

$$\frac{P(M_j|y, X)}{P(M_\ell|y, X)} = \frac{P(M_j)}{P(M_\ell)} \cdot \frac{p(y|X, M_j)}{p(y|X, M_\ell)}, \quad (22)$$

where  $P(M_r|y, X)$  is the posterior probability of the  $r$ -th model,  $P(M_r)$  is its prior probability, and  $p(y|X, M_r) = p_r(y|X)$  is the marginal data density (under  $M_r$ ) obtained by integrating out the parameters. (22) can be formulated simply as follows: posterior odds are equal to prior odds times the Bayes factor. In the case of the comparison between any pair of the models considered previously, the Bayes factor is the ratio of the normalizing constants defined in (5), i.e.

$$B_{j\ell} = \frac{p_j(y|X)}{p_\ell(y|X)} = \frac{K_j}{K_\ell}.$$

The posterior probability of the  $j$ -th model is

$$P(M_j|y, X) = K_j P(M_j) / \sum_{\ell} K_\ell P(M_\ell). \quad (23)$$

Specification of the prior distributions of the parameters of the models being compared deserves much caution. For estimation purposes within individual models improper priors are often used, provided that they lead to well-defined posteriors (i.e.  $K_j < +\infty$ ).

However, the use of an improper prior results in the dependence of  $K_j = p(y|X, M_j)$  on an arbitrary positive constant. If the same improper prior is defined on the parameters which are common to all compared models, then the same constant appears in all  $K_j$ 's, and it cancels out when we calculate Bayes factors and posterior model probabilities. An improper prior defined over model-specific parameters, however, leaves posterior model probabilities undefined.

Thus, if we are going to compare only stochastic frontier models with different Erlang distributions of the inefficiency term, i.e. models  $M_1$ ,  $M_2$  and  $M_3$ , we can use  $p_j(\theta_j) = p(\theta_*)$ ,  $j = 1, 2, 3$ , with  $p(\theta_*)$  improper over all the parameters, since they are all common. If, however, we also consider  $M_0$  then  $\lambda$  becomes specific to  $M_j$  with  $j = 1, 2, 3$ , and in the prior structure  $p_j(\theta_*) = p(\beta, \sigma^2)p_j(\lambda|\beta, \sigma^2)$  the last factor needs to be a proper p.d.f. Usually the product structure  $p_j(\theta_*) = p(\beta, \sigma^2)p_j(\lambda)$  with proper  $p_j(\lambda)$  and possibly improper  $p(\beta, \sigma^2)$  will be assumed. Here we have implicitly assumed that  $p_j(\beta, \sigma^2) = p(\beta, \sigma^2)$ ,  $j = 0, 1, 2, 3$ , i.e. the prior on the stochastic frontier parameters does not vary over the models under consideration. Analogously, if we want to compare  $M_0$ ,  $M_1$ ,  $M_2$ ,  $M_3$  with  $M_4$ , we can assume  $p_4(\theta_4) = p(\beta, \sigma^2)p_4(\psi, \omega)$  with  $p(\beta, \sigma^2)$  the same as for the other models (and possibly improper), but  $p_4(\psi, \omega)$  necessarily proper.

Having calculated the posterior model probabilities, we do not need to choose among the competing models. For inference purposes we can average out specification uncertainty by mixing individual posterior densities of a quantity of interest, say  $\varphi$ , into its overall posterior density:

$$p(\varphi|y, x) = \sum_j P(M_j|y, X)p_j(\varphi|y, X). \quad (24)$$

Note that this pooling approach amounts to assuming that the conditional distribution of the inefficiency term,  $z_i$ , is simple (e.g. truncated Normal or exponential), only unknown. Rather than mixing different conditional distributions for each  $z_i$ , which would lead to a complicated sampling model, we mix over simple models differing in the distribution of all the inefficiency terms. Therefore, each model is in itself tractable, and the mixing is trivially performed at the final stage, as in (24).

## 6 Efficiency distributions and prior elicitation for model specific parameters

In the previous sections we have considered for each model  $M_j$  ( $j = 1, 2, 3, 4$ ) different conditional distributions of the inefficiency term ( $z_i$ ), which, after taking the Jacobian  $r_i^{-1} = \exp(z_i)$  into account, lead to the following distributions of the efficiency  $r_i = \exp(-z_i)$ :

1.  $p_j(r_i|\theta_j) = p_j(r_i|\delta_j)$ , the conditional prior distribution given the parameters, depending on the model specific parameters  $\delta_j \in \Delta_j$  only,
2.  $p_j(r_i|y, X, \theta_j) = p_j(r_i|y_i, x_i, \theta_j)$  for  $i \in \{1, \dots, N\}$ , the conditional posterior distribution,
3.  $p_j(r_i|y, X)$  for  $i \in \{1, \dots, N\}$ , the marginal posterior distribution,
4.  $p_j(r_f|y, X)$  for  $f \notin \{1, \dots, N\}$ , the predictive distribution (posterior distribution of the "representative firm" efficiency).

The last two distributions are of direct interest on the inference stage, since they combine all information on the individual or "average" efficiency which is contained both in the prior specification and the given set of data. The influence of the observed data can be assessed by the comparison between these two posterior distributions (unconditional w.r.t. the parameters) and the marginal prior distribution of the efficiency given by

$$p_j(r_i) = \int_{\Delta_j} p_j(r_i|\delta_j)p_j(\delta_j)d\delta_j,$$

where

$$p_j(r_i|\delta_j) = r_i^{-1}p_j(z_i|\delta_j)$$

and the last density is evaluated at  $z_i = -\ln(r_i)$ ,  $r_i \in (0, 1]$ .

Note that the marginal prior distribution of the efficiency, which should reflect the economic interpretation of  $r_i$  and possibly subjective beliefs of a researcher, is a useful starting point for eliciting the (hyperparameters of the) prior distribution of the model-specific parameters,  $\delta_j$  ( $j = 1, 2, 3, 4$ ). The classes of prior densities  $p_j(\delta_j)$  will be simple, allowing straightforward analytical derivation of the marginal densities  $p_j(z_i)$  and  $p_j(r_i)$ , and the elicitation of their hyperparameters will be based on quantiles. If there are two free hyperparameters in  $p_j(r_i)$ , or equivalently in  $p_j(z_i)$ , then two quantiles are sufficient, and the following questions could be asked: what is the prior probability that a firm (in the industry under consideration) is more than (e.g.) 90% efficient. and less than (e.g.) 50% efficient? Eliciting more quantiles would lead to more equations than hyperparameters, and therefore some optimization technique should be used. An alternative strategy, which will be followed here is to fix all hyperparameters but one at convenient values, and then elicit a value for the remaining hyperparameter through the

median only. For practitioners, the median seems an easily elicited quantity as a result of its intuitive definition and its appealing simplicity.

In the case of  $M_1$ ,  $M_2$  and  $M_3$ , the only model specific parameter is  $\lambda$  ( $\delta_j = \lambda$ ,  $j = 1, 2, 3$ ), the scale parameter of the Erlang distributed inefficiency term. Here an obvious choice is an inverted gamma prior distribution of  $\lambda$ , or equivalently a gamma prior of  $\lambda^{-1}$ , which is natural conjugate and gives the diffuse prior  $p(\lambda) \propto \lambda^{-1}$  as the limiting case. Combining  $p_j(z_i|\lambda) = f_G(z_i|j, \lambda^{-1})$  with  $p_j(\lambda^{-1}) = f_G\left(\lambda^{-1} \mid \frac{\zeta_0}{2}, \frac{\lambda_0}{2}\right)$  leads to the following three parameter inverted beta marginal prior of  $z_i$ :

$$p_j(z_i) = f_{IB}\left(z_i \mid \frac{\zeta_0}{2}, j, \frac{\lambda_0}{2}\right),$$

which, in turn, gives for  $r_i \in (0, 1]$

$$p_j(r_i) = r_i^{-1} f_{IB}\left(-\ln r_i \mid \frac{\zeta_0}{2}, j, \frac{\lambda_0}{2}\right)$$

as the prior density of the efficiency itself. This last density is not a standard one, but the quantile elicitation can be done in terms of  $u_i = \left(1 + \frac{2z_i}{\lambda_0}\right)^{-1} = \left(1 - \frac{2\ln r_i}{\lambda_0}\right)^{-1}$ , which is beta distributed with parameters  $\left(\frac{\zeta_0}{2}, j\right)$ . We have

$$\begin{aligned} Pr\{r_i < r^* | M_j\} &= Pr\{u_i < \left(1 - \frac{2\ln r^*}{\lambda_0}\right)^{-1} | M_j\} \\ &= \frac{1}{B\left(\frac{\zeta_0}{2}, j\right)} \int_0^{a_*} u^{\frac{\zeta_0}{2}-1} (1-u)^{j-1} du, \end{aligned}$$

where  $a_* = \left(1 - \frac{2\ln r^*}{\lambda_0}\right)^{-1}$ , and the integral is the incomplete beta function evaluated at  $a_*$ . In order to further simplify our prior elicitation we propose to take  $\zeta_0 = 2j$ . Then  $u_i$  has a symmetric beta distribution, and thus its median is  $\frac{1}{2}$ , i.e.  $Pr\{u_i < \frac{1}{2}\} = \frac{1}{2}$ . Now the whole elicitation procedure reduces to specifying the prior median of the efficiency, i.e. such a value  $r^* \in (0, 1)$  that  $Pr\{r_i < r^* | M_j\} = \frac{1}{2}$ . Since it corresponds to  $a_* = \frac{1}{2}$ , therefore  $\lambda_0$  is given by the equation  $\frac{1}{2} = \left(1 - \frac{2\ln r^*}{\lambda_0}\right)^{-1}$ , i.e.  $\lambda_0 = -2\ln r^*$ .

Summarizing, we propose as a prior on the model specific parameter  $\lambda^{-1}$  in  $M_j$  ( $j = 1, 2, 3$ )

$$p_j(\lambda^{-1}) = f_G(\lambda^{-1}|j, -\ln r^*), \quad (25)$$

where  $r^*$  is the prior median of the efficiency  $\tau_i$ . The implied marginal priors of  $z_i$  and  $r_i$  are

$$p_j(z_i) = f_{IB}(z_i|j, j, -\ln r^*) \quad (26)$$

and

$$p_j(r_j) = r_i^{-1} f_{IB}(-\ln r_i|j, j, -\ln r^*), \quad (27)$$

respectively. In practice, it seems preferable to choose  $r^*$  independently of the model entertained, i.e.  $Pr\{r_i < r^* | M_j\} = Pr\{r_i < r^*\} = \frac{1}{2}$ , so that the model dependence in the prior is confined to the value of  $\zeta_0 = 2j$ . This simple way of prior elicitation preserves in a natural way the differences that exist between the models conditionally upon the parameters. From a gamma (Erlang) distribution with shape parameter  $j$  for  $p_j(z_i|\lambda)$  we are led to a marginal prior  $p_j(z_i)$  of the inverted-beta form with essentially the same behaviour, but thicker tails, reflecting the prior uncertainty concerning  $\lambda$ . The proposed elicitation rule is thus not only very simple and easy to use in practice, but also avoids distorting the main characteristics of our different Erlang models.

In the case of  $M_4$ , there are two model specific parameters, namely  $\psi$  and  $\omega$ , the parameters of the truncated Normal distribution of  $z_i$  given  $\theta_4$ . Here, due to truncation, the usual Normal-inverted gamma prior specification for  $(\mu, \omega^2)$  (with  $\mu = \psi\omega$ ), natural-conjugate in the untruncated case, does not lead to an analytically tractable marginal prior of  $z_i$ . Therefore we propose the following modified version of the usual natural-conjugate prior

$$p_4(\mu|\omega^{-2}) = c_\omega f_N^1(\mu|0, a\omega^2)\Phi\left(\frac{\mu}{\omega}\right),$$

or, in terms of  $\psi$ ,

$$p_4(\psi|\omega^{-2}) = c_\omega f_N^1(\psi|0, a)\Phi(\psi), \quad (28)$$

and

$$p_4(\omega^{-2}) = f_G\left(\omega^{-2} \left| \frac{\nu_0}{2}, \frac{b}{2} \right.\right), \quad (29)$$

where  $c_\omega$  is the normalizing constant of the conditional priors of  $\mu$  and  $\psi$  given  $\omega^{-2}$ . The factor  $\Phi(\psi)$  makes the conditional prior in (28) asymmetric (positive values of  $\psi$  are more probable a priori than negative ones) and shifts its mean to the right of 0:

$$0 < E(\psi|\omega^{-2}) < c_\omega \sqrt{\frac{a}{2\pi}}.$$

Although this prior specification clearly favours (as conditional distributions of  $z_i$  given  $\omega^{-2}$  and  $\psi$ ) distributions with  $\psi > 0$ , i.e. with truncation below  $\mu = \psi\omega$ , there is enough prior mass in the negative tail of  $\psi$  to lead to the half-Normal distribution of  $z_i$  given  $\omega^{-2}$  and the half-Student marginal distribution of  $z_i$ :

$$p_4(z_i|\omega^{-2}) = 2f_N^1(z_i|0, (1+a)\omega^2) \quad I(z_i \geq 0),$$

$$p_4(z_i) = 2f_S^1\left(z_i \left| \nu_0, 0, \frac{\nu_0}{(1+a)b} \right.\right) \quad I(z_i \geq 0), \quad (30)$$

and it is obvious that  $c_\omega = 2$ . The latter fact implies that  $\psi$  and  $\omega^{-2}$  are a priori independent and thus

$$0 < E(\psi) < \sqrt{\frac{2a}{\pi}}.$$

From the density of  $p_4(z_i|\omega^{-2})$  we immediately deduce that the difference between the truncated Normal model with unknown  $\psi$  ( $a > 0$ ) and the half-Normal ( $a = 0$ ) is exactly the variance inflation. Note that in fact there are only two quantities we can elicit from (30), namely  $\nu_0$  and  $\tau_0^2 = \frac{b}{\nu_0}(1+a)$ , where  $a$  and  $b$  are indistinguishable, and the elicitation can be based on Student  $t$  quantiles. If  $u$  is distributed as the untruncated univariate Student  $t$  with  $\nu_0$  degrees of freedom, location 0 and scale 1, then, for  $r^* \in (0, 1]$

$$\begin{aligned} Pr\{\tau_i < r^* | M_4\} &= Pr\{z_i > -\ln r^* | M_4\} = 2Pr\left\{u > -\frac{\ln r^*}{\tau_0} | M_4\right\}. \\ &= 2[1 - Pr\{u \leq (-\ln r^*)/\tau_0 | M_4\}], \end{aligned}$$



and the median of  $\tau_i$  corresponds to the 0.75 quantile of  $u$ ,  $u_{0.75}(\nu_0)$ . This quantile is 1 for  $\nu_0 = 1$ , 0.816 for  $\nu_0 = 2$ , and tends to 0.674 as  $\nu_0$  approaches infinity. Eliciting  $r^*$  equal to the prior median of the efficiency leads to the equation  $u_{0.75}(\nu_0) = -\frac{\ln r^*}{\tau_0}$ , i.e.

$$\tau_0 = -\frac{\ln r^*}{u_{0.75}(\nu_0)}. \quad (31)$$

Since  $u_{0.75}(\nu_0)$  does not change much for  $\nu_0 > 5$ , we propose to fix  $\nu_0 = 10$  which gives  $\tau_0 = -\frac{10}{7}\ln r^*$  or  $\tau_0^2 \approx 2\ln^2 r^*$ . This corresponds to the following prior density of the model specific parameters of  $M_4$ :

$$p_4(\psi, \omega^{-2}) = 2\Phi(\psi)f_N^1(\psi|0, a)f_G(\omega^{-2}|5, \frac{10}{1+a}\ln^2 r^*), \quad (32)$$

and results in the following marginal prior densities of  $z_i$  and  $\tau_i$ :

$$p_4(z_i) = 2f_S^1(z_i|10, 0, (2\ln^2 r^*)^{-1})I(z_i \geq 0), \quad (33)$$

$$p_4(\tau_i) = \frac{2}{\tau_i}f_S^1(-\ln \tau_i|10, 0, (2\ln^2 r^*)^{-1})I(0 < \tau_i \leq 1), \quad (34)$$

which, as a result of our elicitation based on (30) no longer depend on  $a$ .

As in the Erlang models, the marginal inefficiency distribution has thicker tails than the one given  $\omega^2$ , and the influence of integrating out  $\psi$  with the prior in (32) is to relocate the mode at the origin and to inflate the variance. Again, prior elicitation rules are very simple and the prior densities do not affect the salient features of the model.

Although the marginal prior densities of  $z_i$  and  $\tau_i$  do not involve  $a$  any longer, posterior and predictive results will, of course, be affected by the value of  $a$ . Rather than complicate our elicitation procedure, we could conduct a sensitivity analysis with different values of  $a$ , one of which would be  $a = 0$ , corresponding to the half-Normal case of Aigner *et al.* (1977). As  $a$  grows, the relative importance of  $\psi$  in achieving the (fixed) stochastics of  $z_i$  in (33) becomes larger, and the relative influence of the stochastic nature of  $\omega^2$  decreases. Values for  $a$  can be calibrated against the fact that  $a$  approximates the prior variance of  $\psi$ .

## 7 Empirical implementation to the US electric utility industry

As an illustration, let us consider the data collected by Christensen and Greene (1976) for 123 electric utility companies in the US in 1970. The numbers are listed in the Appendix to Greene (1990). Both Christensen and Greene (1976) and Greene (1990) have fitted a cost function suggested by Nerlove (1963) and based on the Cobb-Douglas production function, but generalized to include a term in squared log of output  $Q$ , which permits returns to scale to vary with  $Q$ . There are three production factors, labour, capital and fuel with respective prices  $P_l$ ,  $P_k$  and  $P_f$ , and the specification of the cost function is:

$$y_i = -\beta_0 - \beta_1 \ln Q_i - \beta_2 \ln^2 Q_i - \beta_3 \ln\left(\frac{P_l}{P_f}\right)_i - \beta_4 \ln\left(\frac{P_k}{P_f}\right)_i + v_i - z_i, \quad (35)$$

where  $y_i = -\ln(\text{cost firm } i/P_{fi})$ .

### 7.1 Erlang Models under Diffuse Priors

In this subsection, we shall only report results on the Erlang models ( $j = 1, 2, 3$ ), which allows us to use independent diffuse prior densities on all parameters, since  $\lambda$  is common to all three models.

Thus, under the prior density

$$p_j(\theta_*) = p(\beta, \sigma^2, \lambda) = c\sigma^{-2}\lambda^{-1} \quad (36)$$

with  $c > 0$ , we obtain the posterior moments as stated in Table 1. These results have required seven dimensional integration using Monte Carlo with importance sampling. As importance functions we have used a Student  $t$  for  $\beta$  with 10 degrees of freedom, and inverted gamma densities on  $\sigma^2$  and  $\lambda$ . Starting from initial importance functions, based on e.g. the mode and the Hessian of the likelihood, we iteratively update the hyperparameters of the importance functions in preliminary runs.

Final results are based on 150,000 antithetic drawings for  $j = 1, 2$  and on 100,000 antithetic

replications for  $j = 3$ .

The last lines in Table 1 contain values for the relative importance of the symmetric component in the posterior out-of-sample error variance, calculated as  $VR_f = \text{var}(v_f|y, X)/\text{var}(v_f - z_f|y, X)$ , as well as for  $TV_f = \text{var}(v_f - z_f|y, X)$ .

Table 1. Posterior Moments of  $\theta_*$  with Diffuse Prior

	$j = 1$		$j = 2$		$j = 3$	
	mean	(s.dev.)	mean	(s.dev.)	mean	(s.dev.)
$\beta_0$	-7.467	(0.3388)	-7.442	(0.3427)	-7.361	(0.3647)
$\beta_1$	0.4272	(0.0424)	0.4071	(0.0421)	0.3968	(0.0389)
$\beta_2$	0.0295	(0.0028)	0.0304	(0.0028)	0.0311	(0.0027)
$\beta_3$	0.2480	(0.0643)	0.2572	(0.0684)	0.2479	(0.0678)
$\beta_4$	0.0489	(0.0610)	0.0603	(0.0638)	0.0698	(0.0587)
$\sigma^2$	0.0133	(0.0038)	0.0161	(0.0043)	0.0171	(0.0036)
$\lambda$	0.0912	(0.0246)	0.0500	(0.0168)	0.0345	(0.0116)
$\lambda^{-1}$	11.901	(3.733)	23.446	(11.368)	32.830	(12.246)
$VR_f$	0.5825		0.7059		0.7669	
$TV_f$	0.0228		0.0228		0.0223	

Just like in Nerlove (1963) (without composed error) and in Greene (1990), returns to scale tend to go down with output,<sup>5</sup> and results for  $\beta$  under this diffuse prior are not very different from Greene's (1990) results. Of course, his results are more comparable to the posterior mode than to the posterior mean, and both differ as a result of the fundamental lack of symmetry of the posterior density (and in particular the likelihood function).

As a result of the structure of the likelihood function in (12), it is extremely complicated to evaluate correctly for cases where some  $m_i$ 's in (8) are large negative numbers. Indeed, (12) involves a product over all firms of  $c_{ji}$ 's which then contain some very small numbers, emanating from the evaluation of the Normal distribution function for large negative arguments. In particular, we found the standard routine CDFN for cumulative Normal density functions in

<sup>5</sup>Returns to scale are given by  $1/(\beta_1 + 2\beta_2 \ln Q_i)$  and  $\beta_2$  has most of the posterior mass in the positive region.

GAUSS to be entirely useless<sup>6</sup> for evaluating  $\Phi(\cdot)$  with arguments less than -5, which do occur frequently in our computations, especially for  $j = 2, 3$ . Mechanical use of CDFN leads one very far astray in evaluating (12), as it appears that the likelihood shoots off to infinity for  $\sigma^2 \rightarrow 0$  (which blows up the arguments of the Normal distribution function). Extensive checking of such routines against the tables in Pearson and Hartley (1954) has led us to use CDFN for arguments  $\geq -5$ , and in other cases we use the continued fraction expansion (Abramowitz and Stegun, 1964, eq. 26.2.14) to the order 20. The latter gave us almost perfect accuracy up to values of the argument of -37 (i.e. for values of  $\Phi(\cdot)$  as small as  $10^{-300}$ ).

These numerical issues are especially critical for the results presented in Table 2, where the integrating constants, crucial for the posterior odds in (22), are tabulated [for  $c = 1$  in (36)]. At the posterior mode  $\bar{\theta}_j$ , we have also evaluated the posterior density function and the likelihood. Note that the latter value seems to agree with Greene's (1990) ML evaluation for  $j = 1$ , but then decreases slightly with  $j$ , which is very different from his results for a two-parameter gamma with  $j$  estimated at 2.45 with standard deviation 1.10. Both our results concerning the likelihood at the posterior mode and the integrating constant (i.e. the likelihood averaged out with the prior (36) where  $c$  is taken to be unity) favour the exponential model. Its posterior probability is 0.45, whereas for both  $j = 2$  and  $j = 3$  we get about 0.27. In contrast, Greene (1990) finds a log likelihood value of 112.72 for the gamma case with free  $j$ , which would imply very strong evidence against the exponential. Although we fix  $j$  and our results are thus not formally comparable, we find it quite unlikely that the (otherwise very flat) likelihood would shoot up to such a value in between our models  $M_2$  and  $M_3$ . In our view, Greene's (1990) findings might have been crucially influenced by the hard numerical problems that we have referred to in the previous paragraph.

Finally, at the cost of adding one dimension to the Monte Carlo integration, we present posterior results for the efficiency of firms within the sample,  $\tau_i$ , and the out-of-sample efficiency  $\tau_j$ . Posterior densities  $p_j(\tau_j|y, X)$  are graphed in Figure 1 for  $j = 1, 2, 3$ , from which we clearly see

<sup>6</sup>In fact, it gives exactly the same value to  $\Phi(x)$  for any  $x \leq -5$ . Whereas this is accurate for  $x = -5$ , it is of course far too large for  $x$  appreciably smaller than -5. Note that this seriously blows up values for the likelihood function in (12). See also Goldstein (1991). Neither CDFN nor any of the two polynomial approximations listed in Abramowitz and Stegun (1964) and briefly mentioned by Greene (1990, p. 148) were found to be of any use for this application.

Table 2. Evidence on Model Probabilities with Diffuse Prior

	$j = 1$	$j = 2$	$j = 3$
$\log K_j$	49.403	48.886	48.903
$\log p_j(\bar{\theta}_j y, X)$	60.258	59.029	58.370
$\log l_j(\bar{\theta}_j y, X)$	67.005	66.127	65.747
$P(M_j)$	1/3	1/3	1/3
$P(M_j y, X)$	0.4541	0.2706	0.2754

that for a greater value of  $j$  the modes are shifted to the left, in agreement with the properties of the sampling model. Using the posterior model probabilities of Table 2, the overall posterior density of out-of-sample efficiency for Erlang models is also plotted.

Posterior moments for  $r_f$  computed as in (18) and for individual efficiencies  $r_i$  ( $i = 1, \dots, 5$ ) of the first five firms in the sample<sup>7</sup> are presented in Table 3.

Table 3. Posterior Moments of Efficiencies with Diffuse Prior

	$j = 1$		$j = 2$		$j = 3$	
	mean	(s.dev.)	mean	(s.dev.)	mean	(s.dev.)
$r_f$	0.9169	(0.0807)	0.9077	(0.0694)	0.9038	(0.0618)
$r_1$	0.6169	(0.0777)	0.6645	(0.1246)	0.7435	(0.1007)
$r_2$	0.9695	(0.0301)	0.9530	(0.0329)	0.9392	(0.0360)
$r_3$	0.9302	(0.0609)	0.9212	(0.0550)	0.9058	(0.0591)
$r_4$	0.8901	(0.0878)	0.9052	(0.0660)	0.8960	(0.0654)
$r_5$	0.9582	(0.0400)	0.9419	(0.0407)	0.9277	(0.0432)
median $r_f$	0.94		0.93		0.915	

Just as the out-of-sample efficiency is more concentrated around the mean for larger  $j$ , the means of individual firm efficiencies are also less spread out as  $j$  goes up. Of course, for very large values of  $j$  the Erlang distribution tends to a Normal, which will become indistinguishable from the symmetric Normal on  $v_i$ . All the stochastics will then be attributed to determining the frontier itself, and all firms will tend to be equally efficient, although the level of efficiency will no longer be identified.

<sup>7</sup>These are, respectively, Cal. Pac. Utility, Montana Power, Upper Pen. Pwr., Mt. Carmel Pub. and Bangor Hydro.

Comparing Table 3 with Greene's (1990) results, we note some interesting differences, especially with his gamma model, which seems very far off (both from our results and his results using other distributions) for firms 1 and 2. Remark that these firms are, respectively, the least and the most efficient of the subset of five firms considered here.

## 7.2 Overall Analysis

In order to combine models with varying sets of model specific parameters  $\delta_j$  we shall use the prior elicitation rules derived in Section 6. For  $(\beta, \sigma^2)$  we continue to use the diffuse prior  $p(\beta, \sigma^2) = \sigma^{-2}$ . We now no longer compare sampling models under the same prior assumptions (as in the previous subsection), but we compare Bayesian models, i.e. combinations of a sampling model and a model-specific prior density. The proper priors in (25) and (28), (29) with  $\alpha = 1$ , were chosen to preserve the shape and characteristics of the corresponding sampling models, and are quite different. In order to isolate the model-specific elements, it is instructive to consider  $p_j(z_i) = \int_{\Delta} p_j(z_i|\delta_j)p_j(\delta_j)d\delta_j$ . Of course, the  $z_i$ 's are no longer independent when  $\delta_j$  is integrated out, so that their contribution to the  $j^{\text{th}}$  Bayesian model marginalized with respect to  $\delta_j$  is not exactly  $\prod_{i=1}^n p_j(z_i)$ , but the marginal priors of  $z_i$  or, rather, of  $r_i = -\ln(z_i)$  can nevertheless guide our intuition. Figures 2 and 3 plot the priors of  $r_i$  for  $j = 1, 2, 3, 4$ , using (27) and (34) with  $r^* = 0.875$  and  $0.5$ , respectively. Clearly, prior 1 is most conservative, since its shape is least affected by going from  $r^* = 0.875$  to  $r^* = 0.5$ , which are very different values from an economic point of view. From Figures 2 and 3 it seems that  $M_3$  should receive most of the posterior probability mass if prior and sample information are in accordance, i.e. if values of  $r_i$  close to the prior median  $r^*$  correspond to high values of the frontier likelihood  $p(y_i|z_i, x_i, \beta, \sigma^2)$ , which is not model-dependent.

On the other hand, if prior and sample information are in conflict, the posterior ranking should be relatively more in favour of  $M_1$ . Note, in passing, that the diffuse prior  $p(\lambda) \propto \lambda^{-1}$  used in Subsection 7.1 implies exactly the same marginal prior density  $p_j(r_i) \propto r_i^{-1}[-\ln(r_i)]^{-1}$  for all the Erlang models, irrespective of the value of  $j$ . In fact, this implies that the entire difference between the models in that case can be summarized by different degrees of dependence between the marginal priors on  $r_i$ . The graph denoted by "d" in Figure 2 describes (with arbitrary scaling) the prior of  $r_i$  under this diffuse assumption on  $\lambda$ .

Let us first present in Tables 4-6 the results for the prior median efficiency  $r^* = 0.875$ . In view of the posterior median efficiencies in the diffuse Erlang case (see Table 3), prior and likelihood each other should roughly be in accordance with each-other here. Indeed, Table 5 now attributes the bulk of the posterior probability mass to  $M_3$ , the Erlang model where the marginal prior on  $r_j$  is most concentrated around its mode. With the exception of  $M_2$ , which receives about 7% of the posterior probability, the rest of the models get little posterior probability, in particular  $M_0$ , the model which imposes full efficiency. Although the relative ranking of the Erlang models is very different from the diffuse case in Subsection 7.1, the posterior moments of  $\theta_*$  are hardly affected by the informative priors, as is obvious from comparing Tables 1 and 4. In fact, the posterior moments of  $\beta$  are close to the ones for the baseline model  $M_0$  under all the efficiency distributions considered. Of course, the posterior mean of  $\sigma^2$  decreases as soon as a composed error is assumed.

Table 4. Posterior Moments of  $\theta_j$  with Informative Priors on  $\delta_j$  ( $r^* = 0.875$ )

	$j = 0$		$j = 1$		$j = 2$		$j = 3$		$j = 4(a = 1)$	
	mean	(s.dev.)	mean	(s.dev.)	mean	(s.dev.)	mean	(s.dev.)	mean	(s.dev.)
$\beta_0$	-7.205	(0.3397)	-7.479	(0.3448)	-7.438	(0.3898)	-7.342	(0.3400)	-7.374	(0.3187)
$\beta_1$	0.3860	(0.0386)	0.4276	(0.0428)	0.4145	(0.0433)	0.3927	(0.0436)	0.3988	(0.0391)
$\beta_2$	0.0316	(0.0027)	0.0295	(0.0028)	0.0300	(0.0029)	0.0313	(0.0030)	0.0311	(0.0027)
$\beta_3$	0.2462	(0.0674)	0.2492	(0.0652)	0.2495	(0.0717)	0.2471	(0.0663)	0.2428	(0.0653)
$\beta_4$	0.0792	(0.0621)	0.0449	(0.0619)	0.0573	(0.0626)	0.0748	(0.0668)	0.0654	(0.0658)
$\sigma^2$	0.0213	(0.0028)	0.0129	(0.0036)	0.0150	(0.0034)	0.0171	(0.0033)	0.0131	(0.0033)
$\lambda$			0.0954	(0.0242)	0.0557	(0.0136)	0.0348	(0.0105)		
$\lambda^{-1}$			11.273	(3.312)	19.302	(5.907)	32.184	(12.429)		
$w$									0.4130	(0.5792)
$w^2$									0.0190	(0.0083)
$w^{-2}$									62.80	(23.89)
$VR_f$	1		0.5562		0.6721		0.7760			
$TV_f$	0.0213		0.0232		0.0223		0.0220			

Table 5. Model Probabilities with Informative Priors on  $\delta_j$  ( $r^* = 0.875$ )

	$j = 0$	$j = 1$	$j = 2$	$j = 3$	$j = 4(a = 1)$
$\log K_j$	48.070	50.294	52.066	54.654	50.134
$\log p_j(\hat{\theta}_j y, X)$	61.775	61.161	62.608	64.518	58.452
$P(M_j)$	1/5	1/5	1/5	1/5	1/5
$P(M_j y, X)$	0.0013	0.0116	0.0683	0.9089	0.0099

The total error variance  $TV_f$ , however, is quite stable across models. Comparison with the

Table 6. Posterior Moments of Efficiencies with Informative Priors on  $\delta_j$  ( $r^* = 0.875$ )

	$j = 0$	$j = 1$	$j = 2$	$j = 3$	$j = 4(a = 1)$
	mean (s.dev.)	mean (s.dev.)	mean (s.dev.)	mean (s.dev.)	mean (s.dev.)
$r_f$	1 (0)	0.91 (0.08)	0.90 (0.07)	0.90 (0.06)	0.83 (0.10)
$r_1$	1 (0)	0.55 (0.09)	0.66 (0.10)	0.68 (0.10)	0.70 (0.08)
$r_2$	1 (0)	0.97 (0.03)	0.95 (0.03)	0.94 (0.04)	0.96 (0.04)
$r_3$	1 (0)	0.93 (0.06)	0.92 (0.06)	0.90 (0.06)	0.87 (0.08)
$r_4$	1 (0)	0.89 (0.08)	0.89 (0.09)	0.89 (0.06)	0.84 (0.09)
$r_5$	1 (0)	0.96 (0.04)	0.94 (0.04)	0.93 (0.05)	0.94 (0.05)
median $r_f$	1	0.94	0.91	0.91	0.83

diffuse prior case in Table 1 shows that variance ratios  $VR_f$  are essentially unaffected by the added prior information, which does not contradict the data information here.

Results for the truncated Normal model are more different from Greene's (1990) *ML* estimates than for the Erlang models (barring the log likelihood value discussed in the previous subsection). Greene finds a large negative value for  $\mu$  (with an even much larger asymptotic standard error), whereas we find a positive mean of  $\psi$ , with a small sample posterior standard deviation of the same order of magnitude.

The resulting truncated Normal distribution of Greene (1990) is close to the exponential distribution (see his Figure 1), which seems at odds with his strong support for the gamma model with shape parameter  $P$  estimated at 2.45. Indeed, a positive posterior mean for  $\psi$  renders the truncated Normal model closer to the Erlang cases with  $j > 1$ , which receive most support in our findings. The posterior mean for  $\sigma^2$  seems in line with his estimate for the half-Normal, but is about twice his estimated value for the truncated Normal ( $a \neq 0$ ).<sup>8</sup> Furthermore, Greene's estimate of  $\omega^2$ , which is 0.1674, is roughly nine times as large as the posterior mean we find for the case with  $a = 1$ . The latter is, of course, linked to his large negative estimate for  $\mu$ . The likelihood function seems to be very flat in the direction of  $\mu$  and behaves much better when parameterized in terms of  $\psi = \mu/\omega$ . In addition, the prior on  $\psi$  in (28) helps in resolving problems of identification.

Posterior moments for within-sample and out-of-sample efficiencies are close to the diffuse case

<sup>8</sup>Table 1 in Greene (1990) is not easy to read. In his notation,  $\sigma^2$  is the variance of  $v = \sigma_v^2$  for the gamma models, but  $\sigma^2 = \sigma_u^2 + \sigma_v^2$  in the truncated Normal cases where  $\sigma_u^2$ , which corresponds to our  $\omega^2$ , is larger than  $\text{var}(u)$  due to truncation. In addition, the value presented for his  $\lambda = \sigma_u/\sigma_v$  does not seem in accordance with his estimates for  $\sigma_u^2$  and  $\sigma_v^2$  in the general truncated Normal model.



for the Erlang models, with the exception of firm 1, which displays a great sensitivity to prior and sampling model assumptions. As it is by far the least efficient of the firms considered and thus the furthest out in the tail of the distribution of  $z_i$ , it should, indeed, be most affected by different forms of tail behaviour of the efficiency distribution.

Although the individual firm efficiencies are not far from those of the Erlang models, the out-of-sample efficiency of the truncated Normal case is shifted away from full efficiency here (see Figure 4).

Comparing Table 6 to Greene's point estimates of inefficiency, we observe again the very high values he finds for firms 1 and 2 with the gamma model and the relatively close correspondence for the exponential model. In addition, his half-Normal model leads to estimates that are similar to our posterior means for  $j = 3$ , but the truncated Normal case of Greene produces a much larger spread of individual firm efficiencies than found in Table 6 for  $j = 4$ . The latter is, of course, to be expected as his truncated Normal model behaves like an exponential and induces a relatively flat efficiency distribution.

Let us now shift the prior median efficiency to a completely different value,  $r^* = 0.5$ , which is no longer in line with the data information. Tables 7-9 summarize the posterior findings where the probabilities of the Erlang models are now much closer, in line with our expectations based on Figures 2 and 3. The large conflict between the prior and the sample information for  $M_1 - M_4$  results in  $M_0$ , implicitly assuming full efficiency (which is much more in line with the data) taking most of the posterior probability mass. The disaccordance between prior and data information can be illustrated by computing Bayes factors for the Bayesian models with common sampling distribution and prior structure, but the different values of  $r^*$  used here, defined by  $B_j^* = p_j(y|X, r^* = 0.875)/p_j(y|X, r^* = 0.5)$ . These Bayes factors clearly favour  $r^* = 0.875$  and are given by 218, 1077, 4081, and 317.426 for  $j = 1, 2, 3$ , and 4, respectively. This also indicates that  $j = 1$  corresponds to the most conservative model.

Comparing Table 7 with Table 4 we note that less of the stochastics is attributed to the frontier in (2) and more to the inefficiency, as could be expected. This is evidenced by the drop in variance ratios, especially for higher  $j$ . Note that this fact reverses the ordering of  $VR_j$  among the Erlang models. We did not compute  $VR_j$  for  $M_4$ , but the more than ten-fold increase in the posterior mean of  $\omega^2$  combined with a decrease in the mean of  $\sigma^2$  definitely points in the

same direction. In addition, the now negative posterior mean of  $\mu^9$  requires large values of  $\omega^2$ .

Table 7. Posterior Moments of  $\theta_j$  with Informative Priors on  $\delta_j$  ( $r^* = 0.5$ )

	$j = 0$		$j = 1$		$j = 2$		$j = 3$		$j = 4(a = 1)$	
	mean	(s.dev.)	mean	(s.dev.)	mean	(s.dev.)	mean	(s.dev.)	mean	(s.dev.)
$\beta_0$	-7.205	(0.3397)	-7.526	(0.3295)	-7.532	(0.3193)	-7.529	(0.3197)	-7.478	(0.3333)
$\beta_1$	0.3860	(0.0386)	0.4398	(0.0418)	0.4324	(0.0385)	0.4311	(0.0405)	0.4412	(0.0405)
$\beta_2$	0.0316	(0.0027)	0.0289	(0.0028)	0.0294	(0.0026)	0.0294	(0.0028)	0.0290	(0.0027)
$\beta_3$	0.2462	(0.0674)	0.2447	(0.0641)	0.2411	(0.0632)	0.2316	(0.0654)	0.2270	(0.0657)
$\beta_4$	0.0792	(0.0621)	0.0362	(0.0603)	0.0385	(0.0599)	0.0432	(0.0620)	0.0326	(0.0631)
$\sigma^2$	0.0213	(0.0028)	0.0107	(0.0027)	0.0093	(0.0028)	0.0077	(0.0029)	0.0082	(0.0022)
$\lambda$			0.1199	(0.0202)	0.0870	(0.0128)	0.0751	(0.0104)		
$\lambda^{-1}$			8.586	(1.519)	11.762	(1.856)	13.573	(1.872)		
$\mu$									-0.9094	(0.2844)
$\omega^2$									0.2016	(0.0501)
$\omega^{-2}$									5.258	(1.270)
$VR_f$	1		0.4133		0.3658		0.2971			
$TV_f$	0.0213		0.0259		0.0254		0.0259			

Table 8. Model Probabilities with Informative Priors on  $\delta_j$  ( $r^* = 0.5$ )

	$j = 0$	$j = 1$	$j = 2$	$j = 3$	$j = 4(a = 1)$
$\log K_j$	48.070	44.910	45.084	46.340	37.466
$\log p_j(\hat{\theta}_j y, X)$	61.775	56.067	55.783	56.691	48.892
$P(M_j)$	1/5	1/5	1/5	1/5	1/5
$P(M_j y, X)$	0.7873	0.0334	0.0397	0.1396	0.00002

As the prior information is increasingly conservative for  $j = 4, 3, 2$  and 1, the out-of-sample efficiency  $r_f$  in Table 9 is not affected much for  $j = 1$ , but more and more as  $j$  increases. Figure 5 graphically displays the posterior densities of  $r_f$  for  $r^* = 0.5$ . Comparing Figures 4 and 5 the relative robustness of the exponential model with respect to large changes in  $r^*$  is clearly illustrated. Finally, note that for firm 1 the prior does not contradict the sample now, so that its individual efficiency is more concentrated than in the case with  $r^* = 0.875$ , and its mean is increasingly drawn towards the prior median as  $j$  grows (i.e. as more prior mass is concentrated around the median; see Figure 3).

<sup>9</sup>For  $M_4$  with  $r^* = 0.5$  the Monte Carlo integration performed better in terms of  $(\mu, \omega)$  than in  $(\psi, \omega)$ . A careful analysis of this issue is a topic of further research. Goldstein (1991) proposes to parameterize in terms of  $(\beta, \mu_2, \mu_3, \psi)$  where  $\mu_j$  is the  $j^{\text{th}}$  central moment of the composed error  $(v_i - z_i)$ .

Table 9. Posterior Moments of Efficiencies with Informative Priors on  $\delta_j$  ( $r^* = 0.5$ )

	$j = 0$	$j = 1$	$j = 2$	$j = 3$	$j = 4(a = 1)$
	mean (s.dev.)	mean (s.dev.)	mean (s.dev.)	mean (s.dev.)	mean (s.dev.)
$r_f$	1 (0)	0.89 (0.10)	0.85 (0.10)	0.81 (0.10)	0.68 (0.16)
$r_1$	1 (0)	0.60 (0.07)	0.55 (0.06)	0.50 (0.05)	0.52 (0.04)
$r_2$	1 (0)	0.97 (0.03)	0.95 (0.04)	0.92 (0.05)	0.97 (0.03)
$r_3$	1 (0)	0.92 (0.06)	0.88 (0.07)	0.83 (0.07)	0.90 (0.07)
$r_4$	1 (0)	0.87 (0.09)	0.81 (0.08)	0.77 (0.08)	0.81 (0.09)
$r_5$	1 (0)	0.96 (0.04)	0.93 (0.05)	0.89 (0.06)	0.95 (0.04)
median $r_f$	1	0.92	0.86	0.82	0.65

## 8 Conclusion

A Bayesian analysis of stochastic frontier models was shown to be both theoretically and practically feasible. A simple application of the rules of probability calculus leads to posterior densities of efficiencies, both within-sample (firm-specific) and out-of-sample (average), where parameter uncertainty is entirely taken care of. This paradigm thus allows direct posterior inference on firm-specific efficiencies, avoiding the much criticized two-step procedure of Jondrow et al. (1982). The difficult choice of a particular sampling model for the inefficiency error term is avoided by mixing over different models, reflecting the spectrum of distributions proposed in the literature. Inherent to this mixing and the associated computation of posterior model probabilities is the fact that proper priors are required on those parameters that are not common to all models considered. We propose a set of prior structures that preserve the main characteristics of the different sampling distributions. In addition, they allow for very simple prior elicitation rules, based only on one number, the prior median efficiency.

We apply these results to the analysis of 123 US electric utility companies in 1970, a data set used and listed in Greene (1990). We briefly treat some numerical issues that appear quite crucial for this application, and were partly overlooked in the previous literature. The empirical analysis relies on Monte Carlo integration with importance sampling for all parameters and the efficiency term. Although this is numerically not a trivial exercise,<sup>10</sup> it provides us with a

<sup>10</sup>A full implementation of our Gauss-386 programmes with the computation of a posterior density plot for  $r_f$  and first and second order moments for  $r_1$  through  $r_5$  ran at the rate of 10,000 antithetic drawings per hour on a 386-33 personal computer with coprocessor. Typical runs were of the order of 100,000 drawings.

wealth of information; the actual posterior densities of e.g. average efficiency can be plotted, taking into account all prior and sample information and with parameter and model uncertainty averaged out. We highlight the main differences with the sampling theoretical approach in Greene (1990) and perform the analysis both under diffuse and proper priors.

In the latter case, the choice of prior median efficiency,  $r^*$ , naturally affects our results. However, as expected, the exponential model is least affected by shifting from a prior median in accordance with the data to one in severe conflict with sample information. In this particular application, the truncated Normal model is found to be most sensitive to the choice of  $r^*$  and turns out to be the least favoured of all composed error models considered.

## Appendix A Joint sphericity of the disturbances of the stochastic frontier

Here we will show that, under the usual diffuse prior of  $\sigma^2$ , all the posterior results (on quantities other than the scale parameter  $\sigma$  of  $v_i$ 's) are perfectly robust w.r.t. departures from Normality within the class of joint spherical distributions of  $v = (v_1, \dots, v_N)'$ .

Assume that the vector  $v$  of  $N$  random disturbances of the stochastic frontiers (2) has a continuous spherical distribution with scale parameter  $\sigma$ , and is independent of all  $x_i$ 's and  $z_i$ 's. Therefore, under any  $M_j$  ( $j = 1, 2, 3, 4$ ), the density of  $v$  takes the form

$$p_j(v|X, z_1, \dots, z_N, \beta, \delta_j, \sigma) = p(v|\sigma) = \sigma^{-N} g_N(\sigma^{-2} v'v), \quad (\text{A.1})$$

where  $\delta_j$  groups the model specific parameters of  $M_j$ , and  $g(\cdot)$  is a known nonnegative function such that

$$\int_0^\infty u^{\frac{N}{2}-1} g_N(u) du = \Gamma\left(\frac{N}{2}\right) \pi^{-\frac{N}{2}}, \quad (\text{A.2})$$

which is both necessary and sufficient to make  $\sigma^{-N} g_N(\sigma^{-2} v'v)$  a proper, normalized density function, see e.g. Dickey and Chen (1985).

Assume that  $g_N(\cdot)$  is not indexed by  $\sigma$  and consider the following (improper) prior structure for the parameters of  $M_j$ :

$$p_j(\beta, \delta_j, \sigma) = p(\sigma) p_j(\beta, \delta_j) = \frac{c}{\sigma} p_j(\beta, \delta_j), \quad (\text{A.3})$$

where  $c$  is a positive constant and  $p_j(\beta, \delta_j)$  is functionally independent of  $\sigma$ . Thus the joint (improper) density of  $v$  and  $\sigma$  given  $x_i$ 's,  $z_i$ 's and  $(\beta, \delta_j)$  is given by

$$p_j(v, \sigma|X, z_1, \dots, z_N, \beta, \delta_j) = p(v|\sigma) p(\sigma) = c \sigma^{-N-1} g_N(\sigma^{-2} v'v).$$

Now the transformation from  $\sigma$  to  $u = \sigma^{-2} v'v$ , with the Jacobian  $\frac{1}{2u} \sqrt{\frac{v'v}{u}}$ , can be used to facilitate the integration over  $\sigma \in (0, +\infty)$ . Since this transformation does not affect  $g_N(\cdot)$ , which is not indexed by  $\sigma$ , we can use (A.2) in order to obtain the (improper) density of  $v$ :

$$\begin{aligned}
 p_j(v|X, z_1, \dots, z_N, \beta, \delta_j) = p(v) &= \frac{c}{2} (v'v)^{-\frac{N}{2}} \int_0^\infty u^{\frac{N}{2}-1} g_N(u) du \\
 &= \frac{c}{2} \Gamma\left(\frac{N}{2}\right) \pi^{-\frac{N}{2}} (v'v)^{-\frac{N}{2}}. \tag{A.4}
 \end{aligned}$$

Remark that the form of this density is not affected by the particular form of the function  $g_N(\cdot)$  and remains exactly the same for any  $g_N(\cdot)$  fulfilling (A.2). Since (A.4) can also be viewed as the density of the observation vector  $y$  given  $(X, z_1, \dots, z_N, \beta, \delta_j)$ , the joint density

$$p_j(y, z_1, \dots, z_N, \beta, \delta_j|X) = p_j(y|X, z_1, \dots, z_N, \beta, \delta_j) p_j(z_1, \dots, z_N|\delta_j) p_j(\beta, \delta_j)$$

and the posterior  $p_j(z_1, \dots, z_N, \beta, \delta_j|y, X)$  remain unaffected by the change from Normal to any other spherical distribution of  $v$ . If joint sphericity of  $(v' v_f)'$  is assumed, then, under (A.3), full robustness of the predictive density of  $y_f$  given  $(y, X, x_f)$  holds as well.

The result presented here was first proven in the general case of nonlinear regression with elliptical errors by Osiewalski and Steel (1992).

## Appendix B Probability density functions

A  $k$ -variate Normal density on  $x \in \mathbf{R}^k$  with mean vector  $b \in \mathbf{R}^k$  and  $PDS$   $k \times k$  covariance matrix  $C$ :

$$f_N^k(x|b, C) = [(2\pi)^k |C|]^{-\frac{1}{2}} \exp - \frac{1}{2}(x - b)'C^{-1}(x - b).$$

A  $k$ -variate Student  $t$  density on  $x \in \mathbf{R}^k$  with  $r > 0$  degrees of freedom, location vector  $b \in \mathbf{R}^k$  and  $PDS$   $k \times k$  precision matrix  $A$ :

$$f_S^k(x|r, b, A) = \frac{\Gamma(\frac{r+k}{2})}{\Gamma(\frac{r}{2})(r\pi)^{k/2}} |A|^{\frac{1}{2}} [1 + \frac{1}{r}(x - b)'A(x - b)]^{-\frac{r+k}{2}}.$$

A gamma density on  $z > 0$  with  $a, b > 0$ :

$$f_G(z|a, b) = b^a [\Gamma(a)]^{-1} z^{a-1} \exp(-bz).$$

A beta density on  $v \in (0, c)$  with  $a, b > 0$ :

$$f_B(v|a, b, c) = \frac{\Gamma(a+b)}{c\Gamma(a)\Gamma(b)} \left(\frac{v}{c}\right)^{a-1} \left(1 - \frac{v}{c}\right)^{b-1}.$$

A three-parameter inverted beta or beta prime density on  $w > 0$  with  $a, b, c > 0$  [see Zellner (1971, p. 376)]:

$$f_{IB}(w|a, b, c) = \frac{\Gamma(a+b)}{c\Gamma(a)\Gamma(b)} \left(\frac{w}{c}\right)^{b-1} \left(1 + \frac{w}{c}\right)^{-(a+b)}.$$

## References

- Abramowitz, M. and I.A. Stegun, 1964, *Handbook of Mathematical Functions* (Dover, New York).
- Aigner, D., C.A.K. Lovell, and P. Schmidt, 1977, Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* 6, 21-37.
- Bauer, P.W., 1990, Recent developments in the econometric estimation of frontiers, *Journal of Econometrics* 46, 39-56.
- Beckers, D.E. and C.J. Hammond, 1987, A tractable likelihood function for the normal-gamma stochastic frontier model, *Economics Letters* 24, 33-38.
- van den Broeck, J., F. Broeckx, and L. Kaufman, 1991, Decomposing stochastic frontier efficiency into secular and organisational efficiency, *Journal of Computational and Applied Mathematics* 37, 251-264.
- Dickey, P., and C.-H. Chen, 1985, Direct subjective-probability modelling using ellipsoidal distributions, in: J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds., *Bayesian Statistics 2* (North-Holland, Amsterdam).
- Christensen, L.R. and W.H. Greene, 1976, Economies of scale in U.S. electric power generation, *Journal of Political Economy* 84, 655-676.
- Goldstein, H.E., 1991, On the COLS moment estimation method for stochastic frontier production functions, mimeo, University of Oslo.
- Greene, W.H., 1990, A gamma-distributed stochastic frontier model, *Journal of Econometrics* 46, 141-163.
- Jondrow, J., C.A.K. Lovell, I.S. Materov, and P. Schmidt, 1982, On the estimation of technical inefficiency in the stochastic frontier production function model, *Journal of Econometrics* 19, 233-238.
- Johnson, N.L., and S. Kotz, 1970, *Distributions in Statistics: Continuous Univariate Distributions -1* (Houghton Mifflin, Boston).
- Meeusen, W., and J. van den Broeck, 1977, Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review* 8, 435-444.
- Nerlove, M., 1963, Returns to scale in electricity supply, in: C.F. Christ, ed., *Measurement in Economics* (Stanford University Press, Stanford).



- Osiewalski, J., and M.F.J. Steel, 1992, Robust Bayesian inference in elliptical regression models, *Journal of Econometrics*, forthcoming.
- Pearson, E.S. and H.O. Hartley, 1954, *Biometrika Tables for Statisticians*, Vol. 1 (Cambridge University Press, Cambridge).
- Stevenson, R.E., 1980, Likelihood functions for generalized stochastic frontier estimation, *Journal of Econometrics* 13, 57-66.
- Zellner, A., 1971, *An Introduction to Bayesian Inference in Econometrics* (Wiley, New York).

FIGURE 1: POSTERIOR EFFICIENCIES

out of sample' with diffuse prior

..... j=1    - - - - j=2    - - - - j=3    — Erlang  
mixed

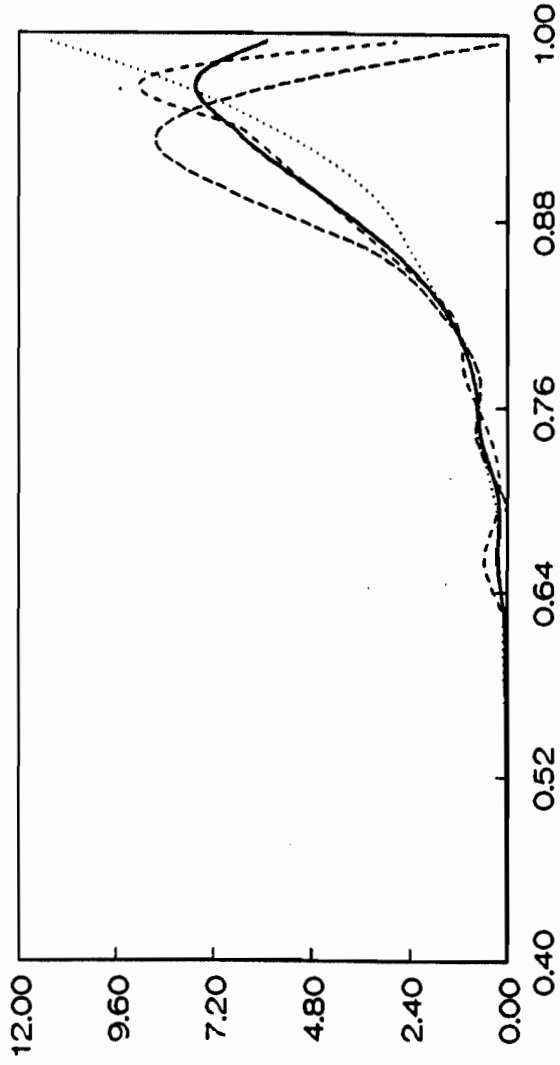


FIGURE 2: PRIOR EFFICIENCIES  
diffuse Erlang (d) and inf. for  $r=0.875$

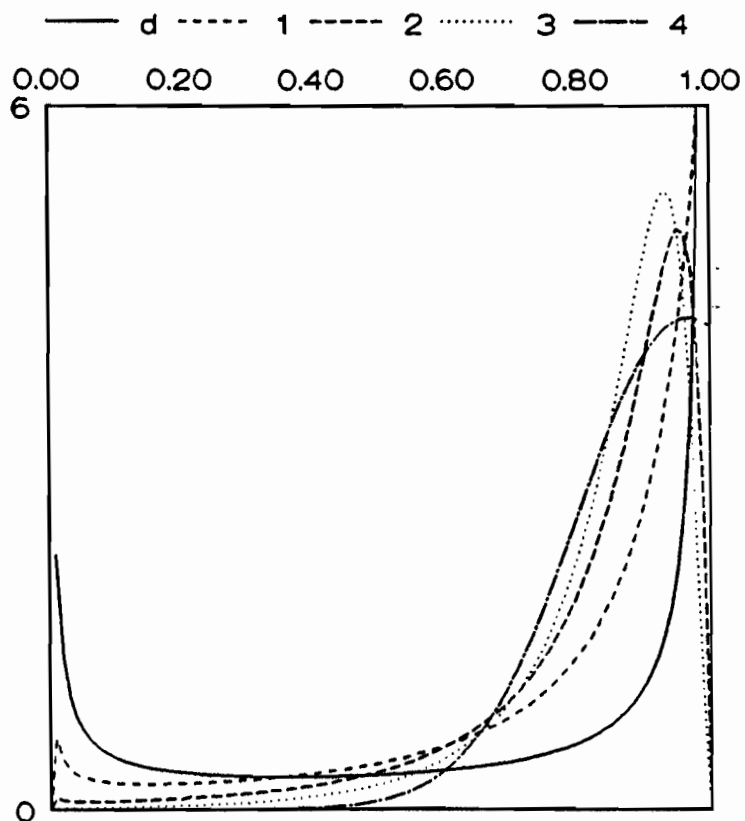


FIGURE 3: PRIOR EFFICIENCIES  
diffuse Erlang (d) and inf. for  $r=5$

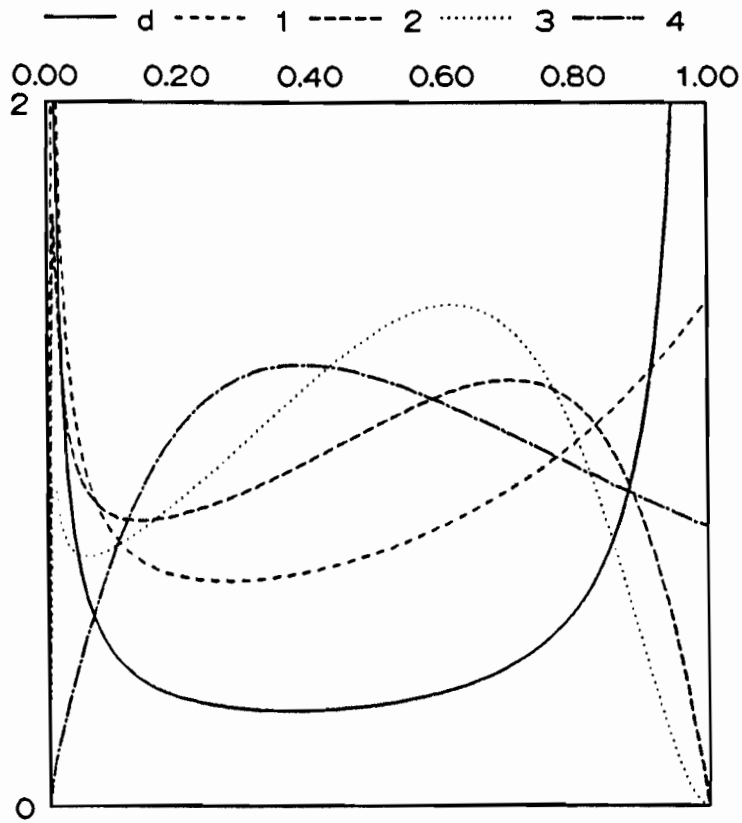


FIGURE 4: POSTERIOR EFFICIENCIES

out of sample with  $r^* = 0.875$

.....  $j=1$     .....  $j=2$     .....  $j=3$     .....  $j=4$

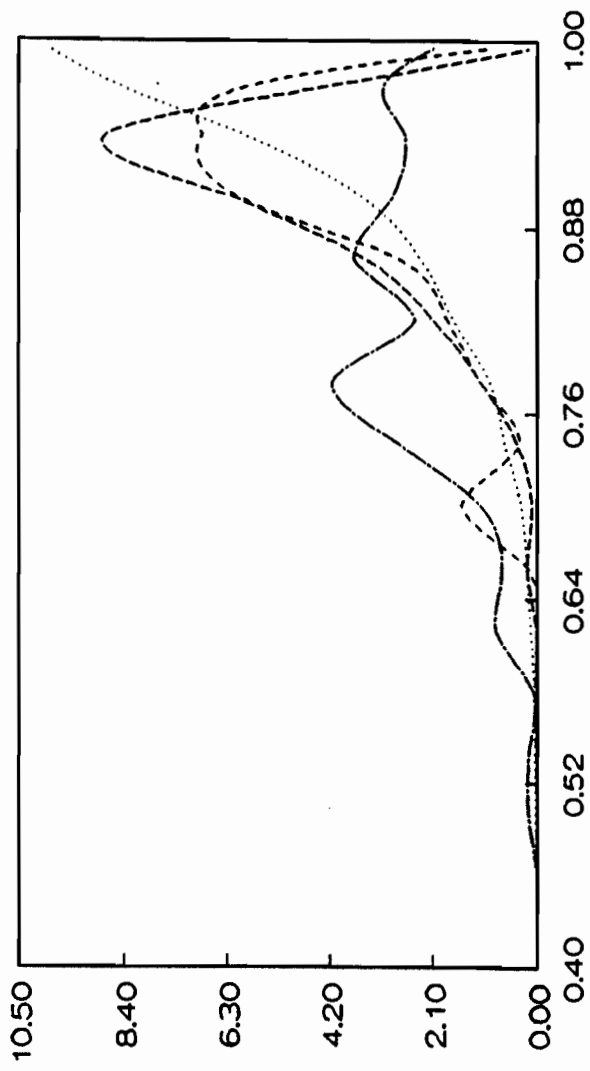


FIGURE 5: POSTERIOR EFFICIENCIES

out of sample with  $r^* = 0.5$

..... j=1    - - - - j=2    - - - - j=3    - - - - j=4

