

# Estimation and testing of stochastic frontier models using variational Bayes

**Gholamreza Hajargasht**

*Swinburne Business School*

*Swinburne University of Technology*

*Melbourne, Australia*

**William E. Griffiths**

*Department of Economics*

*University of Melbourne, Australia*

December 6, 2017

## Abstract

We show how a wide range of stochastic frontier models can be estimated relatively easily using *variational Bayes*. We derive approximate posterior distributions and point estimates for parameters and inefficiency effects for (a) time invariant models with several alternative inefficiency distributions, (b) models with time varying effects, (c) models incorporating environmental effects, and (d) models with more flexible forms for the regression function and error terms. Despite the abundance of stochastic frontier models, there have been few attempts to test the various models against each other, probably due to the difficulty of performing such tests. One advantage of the variational Bayes approximation is that it facilitates the computation of marginal likelihoods that can be used to compare models. We apply this idea to test stochastic frontier models with different inefficiency distributions. Estimation and testing is illustrated using three examples.

**Keywords** Technical efficiency · Marginal likelihood · Time-varying panel · Environmental effects · Mixture · Semiparametric model

**JEL Classification** C11 · C12 · C23 · D24

## Corresponding Author

William Griffiths

Economics Department,

University of Melbourne

Vic 3010, Australia

Email: [wegrif@unimelb.edu.au](mailto:wegrif@unimelb.edu.au)

Email for Hajargasht: [har@unimelb.edu.au](mailto:har@unimelb.edu.au)

## 1. Introduction

Variational Bayes is an approximate method for Bayesian inference. It is popular in machine learning (see e.g. Bishop 2006, Ch 10) and has recently gained attention in statistics (see e.g. Ormerod and Wand 2010 for a review and application to a wide range of models). It can be viewed as an alternative to Markov chain Monte Carlo (MCMC) where, in contrast to MCMC, posterior distributions and parameter estimates are obtained via an optimization procedure rather than by simulating draws from the posterior. Convergence of the optimization procedure is usually achieved within a small number of iterations, making it particularly useful for large models. It comes at a cost, however. It is based on an approximation whose error is hard to quantify. Also, it often tends to underestimate uncertainty in the posterior. The MCMC alternative is subject to sampling error, and is more prone to problems of non-convergence, but does not rely on an approximation. The objective of this article is to investigate the scope and merits of the application of variational Bayes machinery within the context of stochastic frontier models.

Since the seminal article of Aigner et al. (1977) who considered a Cobb-Douglas production function and a half-normal distribution for inefficiency effects, the basic stochastic frontier model has been extended in many ways, including models with a range of inefficiency distributions, extensions to time-invariant and time-varying panel data contexts, extensions that incorporate environmental factors that affect efficiencies, and to models with more flexible functional forms. Reviews can be found in Kumbhakar and Lovell (2000), Coelli et al (2005) and Parmeter and Kumbhakar (2014). In this article we show how several of these models can be estimated relatively easily using variational Bayes, and we compare the accuracy of the estimates with those obtained using MCMC.

Most models considered in this article can also be estimated by MCMC. However, we find several benefits from application of variational Bayes (VB). First, as we will see from

various examples, posterior distributions (in particular posterior means from VB) are good approximations to the posterior distributions from MCMC, while VB is considerably faster. Second, VB provides an analytical formula for a marginal likelihood lower bound which is useful to assess convergence and which provides some guidance for marginal likelihood values which can be notoriously difficult to estimate. Third, the VB posterior is a good candidate for (reciprocal) importance sampling. In particular, we show that the VB posterior can be used to obtain accurate estimates of marginal likelihoods which in turn can be used for model comparison. We apply this idea to discriminate between stochastic frontier models with different inefficiency distributions. Despite the abundance of alternative stochastic frontier models, few studies have tested the alternatives against each other. Fourth, there are situations in which stochastic frontier models come with a large number of parameters. In such cases it can be difficult to ensure MCMC convergence; VB is less likely to experience convergence problems.

The paper is organized as follows. In Section 2 we briefly review VB inference. In Section 3 we show how to use VB to derive posterior distributions and marginal likelihood lower bounds for panel data stochastic frontier models with exponential and gamma distributions. Derivations for three other commonly used distributions are provided in Appendix A. Using three well-known data sets, in Section 4 we estimate all the models discussed in Section 3 and Appendix A, and compare them using the model selection methodology developed in Section 3. The remainder of the paper considers various extensions of the basic model. In Section 5 we show how the model incorporating environmental variables can be estimated. Models with time-varying effects are considered in Section 6. In Section 7 we consider models with more flexible production/cost functions, more flexible distributions for inefficiency errors, and more robust error terms.

## 2. Bayesian inference using variational Bayes

In this section we briefly review a common version of VB known as mean-field variational Bayes. It is a faster and deterministic alternative to MCMC in which posterior distributions and parameter estimates converge within a limited number of iterations. It was developed as a Bayesian method of inference for machine learning in the late 1990s (see e.g., Attias 1999 and Jordan et al. 1999), and has been surveyed by Wainwright and Jordan (2008) and Bishop (2006, Ch 10), among others. More recently it has found its way into the statistical and econometric literature; e.g., McGrory and Titterton (2007), Ormerod and Wand (2010), Wand et al. (2011), Braun and McAulife (2010) and Nott et al. (2012). VB software such as Infer.NET (Minka et al. 2009) and STAN (Kucukelbir et al. 2015) has been developed with claims of being able to estimate a wide range of statistical models.

To introduce the VB framework, suppose we have a vector of observations  $\mathbf{y}$  and a model with likelihood  $L(\mathbf{y} | \boldsymbol{\theta})$ . Given a prior density  $p(\boldsymbol{\theta})$ , we wish to learn about the parameter vector  $\boldsymbol{\theta}$  conditional on the observations. We can use Bayes theorem to derive the posterior density as

$$p(\boldsymbol{\theta} | \mathbf{y}) = \frac{p(\mathbf{y}, \boldsymbol{\theta})}{p(\mathbf{y})} \propto L(\mathbf{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) \quad (2.1)$$

The notation  $p(\cdot)$  is used generically to denote a probability density function. Learning about  $\boldsymbol{\theta}$  (e.g., finding its posterior moments) usually involves integration of functions involving the posterior kernel  $L(\mathbf{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta})$ . For very simple models this can be done analytically or numerically. However, for complex models we need to use more elaborate methods, such as the very powerful and widely used MCMC methods which rely on techniques for drawing random numbers from complex densities. MCMC has been successfully applied to a wide range of models, but it can be time consuming and sometimes it is difficult to know whether the necessary stochastic convergence has been achieved. In VB the idea is to approximate the

posterior  $p(\boldsymbol{\theta} | \mathbf{y})$  with a simpler density  $q(\boldsymbol{\theta})$  which is more tractable. The optimal  $q(\boldsymbol{\theta})$  is obtained by minimizing the Kullback-Leibler distance between the true posterior and the simpler density as follows

$$\min_q KL(q || p) = \min_q \int q(\boldsymbol{\theta}) \log[q(\boldsymbol{\theta})/p(\boldsymbol{\theta} | \mathbf{y})] d\boldsymbol{\theta} \quad (2.2)$$

The word variational comes from the fact that this is a calculus of variations problem. To obtain a tractable approximation simplifying assumptions about  $q(\boldsymbol{\theta})$  are needed. One common VB assumption is

$$q(\boldsymbol{\theta}) = \prod_{k=1}^M q_k(\boldsymbol{\theta}_k) \quad (2.3)$$

where  $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_M\}$  is some partition of  $\boldsymbol{\theta}$  and the  $q_k$  s are probability density functions.

This factorized form corresponds to an approximation framework developed in physics known as *mean field theory* (Parisi, 1988). When used in this context, it is often referred to as mean-field variational Bayes (MFVB). Such a factorization might seem drastic. However, it does not assume complete independence between the  $\boldsymbol{\theta}_k$  s. Each approximate posterior density  $q_k(\boldsymbol{\theta}_k)$  depends on moments of the other components in  $\boldsymbol{\theta}$ . Using results from the calculus of variations, it can be shown that the optimal  $q_k$  are given by the following iterative procedure.

An initial  $q_1(\boldsymbol{\theta}_1)$  is found by taking the expectation of the log of the posterior kernel with respect to all other elements in  $\boldsymbol{\theta}$ . That is, we consider  $E_{\boldsymbol{\theta}_{-1}} \log[p(\mathbf{y}, \boldsymbol{\theta})]$  where the subscript  $(\boldsymbol{\theta}_{-1})$  denotes the set of all other components  $\{\boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \dots, \boldsymbol{\theta}_M\}$ . The various moments of  $\{\boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \dots, \boldsymbol{\theta}_M\}$  that appear in  $E_{\boldsymbol{\theta}_{-1}} \log[p(\mathbf{y}, \boldsymbol{\theta})]$  need to be assigned initial values. Then, the initial distribution for  $q_1(\boldsymbol{\theta}_1)$  is taken as  $q_1(\boldsymbol{\theta}_1) \propto \exp\{E_{\boldsymbol{\theta}_{-1}} \log[p(\mathbf{y}, \boldsymbol{\theta})]\}$ . The moments of  $q_1(\boldsymbol{\theta}_1)$  that will be required for the remaining  $q_j(\boldsymbol{\theta}_j)$  are then evaluated. Often the densities  $q_j(\boldsymbol{\theta}_j)$  are of a standard form and their moments are readily attainable, or, alternatively, they

are of dimension one or two in which case numerical integration is practical. Next, we take  $q_2(\boldsymbol{\theta}_2) \propto \exp\{E_{\boldsymbol{\theta}_{-2}} \log[p(\mathbf{y}, \boldsymbol{\theta})]\}$  where any required moments for  $\boldsymbol{\theta}_1$  come from  $q_1(\boldsymbol{\theta}_1)$  and those for other components of  $\boldsymbol{\theta}$  use the same initial values that were used in step one. The process continues for  $q_3(\boldsymbol{\theta}_3), \dots, q_M(\boldsymbol{\theta}_M)$ . When it returns to  $q_1(\boldsymbol{\theta}_1)$ , the moments in  $E_{\boldsymbol{\theta}_{-1}} \log[p(\mathbf{y}, \boldsymbol{\theta})]$  are now obtained from  $\{q_2(\boldsymbol{\theta}_2), \dots, q_M(\boldsymbol{\theta}_M)\}$ . Iteration continues until convergence with the moments in each  $E_{\boldsymbol{\theta}_{-i}} \log[p(\mathbf{y}, \boldsymbol{\theta})]$  being updated from each new  $q_j(\boldsymbol{\theta}_j)$ . If  $q_1(\boldsymbol{\theta}_1)^{old}, \dots, q_M(\boldsymbol{\theta}_M)^{old}$  are the densities at the end of an iteration, then the updating procedure can be described by the following representation

$$\begin{cases} q_1(\boldsymbol{\theta}_1)^{new} \leftarrow \frac{\exp\{E_{\boldsymbol{\theta}_{-1}} \log[p(\mathbf{y}, \boldsymbol{\theta})]\}}{\int \exp\{E_{\boldsymbol{\theta}_{-1}} \log[p(\mathbf{y}, \boldsymbol{\theta})]\} d\boldsymbol{\theta}_1} \\ \vdots \\ q_M(\boldsymbol{\theta}_M)^{new} \leftarrow \frac{\exp\{E_{\boldsymbol{\theta}_{-M}} \log[p(\mathbf{y}, \boldsymbol{\theta})]\}}{\int \exp\{E_{\boldsymbol{\theta}_{-M}} \log[p(\mathbf{y}, \boldsymbol{\theta})]\} d\boldsymbol{\theta}_M} \end{cases} \quad (2.4)$$

where  $E_{\boldsymbol{\theta}_{-i}} \log[p(\mathbf{y}, \boldsymbol{\theta})] = \int \log p(\mathbf{y}, \boldsymbol{\theta}) \prod_{j>i} q_j(\boldsymbol{\theta}_j)^{old} d\boldsymbol{\theta}_j \prod_{j<i} q_j(\boldsymbol{\theta}_j)^{new} d\boldsymbol{\theta}_j$ . The terms in the denominators are the normalizing constants. The iterative procedure stops when the increase in the lower bound of the marginal likelihood (defined below) is negligible.

If the  $q_j(\boldsymbol{\theta}_j)$  are standard densities such as the normal, gamma or Dirichlet, with moments that depend on the moments of the remaining parameters, it is sufficient to iterate over the moments extracted from the distributions in (2.4). This type of algorithm is known as a *coordinate ascent algorithm*. While cases that lend themselves to this algorithm are common, confining the analysis to such cases restricts considerably the type of models that can be handled via MFVB. Wand et al. (2011) discuss methods for overcoming obstacles created by not having densities which are standard forms. Those that we exploit for some of the models considered in this paper are the use of auxiliary variables (relevant for modelling a  $t$ -

distribution), and numerical integration. With numerical integration over an infinite or semi-infinite region, rather than a compact interval, strategies are needed to determine the effective support of the integrand for accurate computation. One also needs to guard against overflow and underflow. Where numerical integration is necessary for our illustrative examples, we follow the Wand et al. (2011) strategy for dealing with these issues.

As our examples demonstrate, there are a wide range of stochastic frontier models that lend themselves to estimation via MFVB. There are, however, more complex production models that are likely to be less tractable and difficult to fit within the standard MFVB framework, particularly when it does not lead to closed-form moments or univariate integrals. Examples are models with parameter constraints, endogeneity, and multiple outputs. If VB was to be used in such cases, stochastic algorithms or a combination of VB and MCMC methods would need to be explored.

## 2.1 Model selection

The Bayesian approach to comparing two models  $M_1$  and  $M_2$  with the same prior probabilities, is to use the Bayes factor defined as

$$BF = \frac{ML(\mathbf{y} | M_1)}{ML(\mathbf{y} | M_2)} \quad (2.5)$$

where  $ML(\mathbf{y} | M_j) = \int L(\mathbf{y} | \boldsymbol{\theta}_j, M_j) p(\boldsymbol{\theta}_j | M_j) d\boldsymbol{\theta}_j$  is the marginal likelihood. Under an MFVB framework, it can be shown that the log of marginal likelihood is given by

$$\ln ML = E_q \ln p(\mathbf{y}, \boldsymbol{\theta}) - E_q \ln q(\boldsymbol{\theta}) + KL(q \| p) \quad (2.6)$$

If we define  $\ln \underline{ML} = E_q \ln p(\mathbf{y}, \boldsymbol{\theta}) - E_q \ln q(\boldsymbol{\theta})$  and note that  $KL(q \| p)$  is always positive, but unknown, then  $\ln \underline{ML}$  provides a lower bound for the log of the marginal likelihood. This lower bound is used to monitor convergence of the iterative process in (2.4).<sup>1</sup>

In some instances  $\underline{ML}$  will provide a good approximation to the true marginal likelihood. However, it is limited in its usefulness for model choice because the difference  $ML - \underline{ML}$  can be different for different models. Precise estimation of marginal likelihoods can be a challenging task, even when using MCMC.<sup>2</sup> It can be greatly facilitated by using an MFVB posterior as a candidate distribution. Specifically, under a framework proposed by Gelfand and Dey (1994), a marginal likelihood can be estimated from

$$ML = \left[ \frac{1}{K} \sum_{k=1}^K \frac{g(\boldsymbol{\theta}^k)}{L(\mathbf{y} | \boldsymbol{\theta}^k) p(\boldsymbol{\theta}^k)} \right]^{-1} \quad (2.7)$$

where the  $\boldsymbol{\theta}^k$  s are MCMC draws from  $p(\boldsymbol{\theta} | \mathbf{y})$ , and  $g(\boldsymbol{\theta}^k)$  is a suitably chosen density function evaluated at  $\boldsymbol{\theta}^k$ . Our proposal is to set  $g(\boldsymbol{\theta}) = q(\boldsymbol{\theta})$ . The Gelfand-Dey method provides accurate estimates if the candidate density  $g(\boldsymbol{\theta})$  is a good approximation and has narrower tails than the true posterior, properties that we can typically expect from  $q(\boldsymbol{\theta})$ . Wang and Titterton (2005) have shown, and it is confirmed by our empirical examples, that MFVB posterior distributions normally have narrower spread than true posteriors. Thus, we expect them to have narrower tails, and they provide an ideal candidate for  $g(\boldsymbol{\theta})$ . In a Section 4 we use this methodology to compare several stochastic frontier models.

---

<sup>1</sup> Given a sample of MCMC draws  $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \dots, \boldsymbol{\theta}^K$ , it is also possible to estimate an upper bound for the log of the marginal likelihood using  $\ln \overline{ML} = \frac{1}{K} \sum_{k=1}^K \ln \left( \frac{L(\mathbf{y} | \boldsymbol{\theta}^k) p(\boldsymbol{\theta}^k)}{q(\boldsymbol{\theta}^k)} \right)$ . This may not be a tight bound, however and is

calculated by simulation which may be prone to simulation errors (Ji et al. 2010).

<sup>2</sup> See Fruhwirth-Schnatter (2006) for a review of alternative methods.

### 3. Inference for stochastic frontier models using MFVB

The basic stochastic frontier model with panel data can be written as

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} \mp u_i + v_{it} \quad (3.1)$$

where  $i=1,\dots,N$  indexes the firms and  $t=1,\dots,T$  indexes time,  $\mathbf{x}_{it}$  is  $k$ -dimensional row vector of regressors (e.g., logs of inputs or logs of input prices),  $y_{it}$  represents the logarithm of output or cost,  $\mathbf{x}_{it}\boldsymbol{\beta}$  is the log of the frontier production or cost function,  $\boldsymbol{\beta}$  is a vector of unknown parameters,  $u_i$  is a non-negative random error reflecting the inefficiency of firm  $i$ , and  $v_{it}$  represents noise. The negative sign before  $u_i$  is for a production frontier model and the plus sign is for the cost frontier case. Aigner et al. (1977) assumed that the  $u_i$ s are independent and identically distributed (i.i.d.) and follow a half-normal distribution,  $N^+(0, \sigma_u^2)$ ; the  $v_{it}$  s are usually assumed to be i.i.d. normal random variables with mean zero and constant variance  $\sigma^2$ .

Several studies have extended the basic model by considering other distributions such as exponential (Mueseen and van den Broeck 1977), truncated normal (Stevenson. 1980), gamma (Greene 1990), Weibull (Tsionas 2007), generalized gamma (Griffin and Steel 2008) and even mixtures of distributions (Griffin and Steel 2008). The Bayesian approach to the estimation of stochastic frontier models using Monte Carlo methods is popular and has been described in van den Broeck et al. (1994) and Koop and Steel (2001) among others. Griffin and Steel (2007) have shown how a wide range of stochastic frontier models can be estimated using the WINBUGS software.

In this section, we show how the stochastic frontier model with two alternative distributions can be estimated using MFVB and we refer to Appendix A where another three distributions are considered. We first consider a model with an exponential distribution where all the optimal MFVB posterior densities turn out to be standard forms. We then consider a

gamma distribution, a case where we have to use numerical integration. Details for the half-normal, truncated normal and lognormal distributions are provided in Appendix A.

### 3.1 Exponential inefficiency

Consider the stochastic frontier model in (3.1), with  $v_{it} \sim N(0, \sigma^2)$  and where  $u_i$  follows an exponential distribution with parameter  $\lambda$ ; i.e.,  $p(u_i | \lambda) = \lambda \exp(-\lambda u_i)$ . We assume the following priors

$$\boldsymbol{\beta} \sim N(\underline{\boldsymbol{\beta}}, \underline{\mathbf{V}}_{\boldsymbol{\beta}}) \quad \sigma^{-2} \sim G(\underline{A}_{\sigma}, \underline{B}_{\sigma}) \quad \lambda \sim G(\underline{A}_{\lambda}, \underline{B}_{\lambda})$$

with the gamma density defined as  $G(z; a, b) = b^a z^{a-1} e^{-bz} / \Gamma(a)$ . These priors are standard choices in the Bayesian stochastic frontier literature. In the empirical section we discuss how to choose values for the hyper-parameters. Using these priors and Bayes rule, we can write the log of the posterior kernel as

$$\begin{aligned} \ln p(\mathbf{y}, \boldsymbol{\beta}, \mathbf{u}, \lambda, \sigma^{-2}) = & C + \left( \frac{NT}{2} + \underline{A}_{\sigma} - 1 \right) \ln \sigma^{-2} - \frac{\sigma^{-2}}{2} \left[ \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \mathbf{x}_{it} \boldsymbol{\beta} \pm u_i)^2 + 2\underline{B}_{\sigma} \right] \\ & - \frac{1}{2} (\boldsymbol{\beta} - \underline{\boldsymbol{\beta}})' \underline{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} (\boldsymbol{\beta} - \underline{\boldsymbol{\beta}}) + (N + \underline{A}_{\lambda} - 1) \ln \lambda - \left( \sum_{i=1}^N u_i + \underline{B}_{\lambda} \right) \lambda \end{aligned} \quad (3.2)$$

where  $C = -((NT + k)/2) \ln 2\pi + \underline{A}_{\sigma} \ln \underline{B}_{\sigma} - \ln \Gamma(\underline{A}_{\sigma}) + \underline{A}_{\lambda} \ln \underline{B}_{\lambda} - \ln \Gamma(\underline{A}_{\lambda}) - (\ln |\underline{\mathbf{V}}_{\boldsymbol{\beta}}|)/2$ ;  $\mathbf{y}$  and  $\mathbf{u}$  are  $(NT \times 1)$  and  $(N \times 1)$  vectors containing the  $y_{it}$  and  $u_i$ , respectively. To use MFVB we need an appropriate factorized approximation to the posterior. We consider the following

$$q(\boldsymbol{\beta}, \sigma^{-2}, \mathbf{u}, \lambda) = q(\boldsymbol{\beta}) q(\sigma^{-2}) q(\mathbf{u}) q(\lambda) \quad (3.3)$$

For brevity we have suppressed the subscripts on the  $q$  densities. Using (3.2) and (2.4) we can derive the iterative procedure for the optimal densities as

$$\left\{ \begin{array}{l}
q(\boldsymbol{\beta}) \leftarrow N(\bar{\boldsymbol{\beta}}, \mathbf{V}_\beta) \quad \bar{\boldsymbol{\beta}} = \mathbf{V}_\beta \left( \mathbf{V}_\beta^{-1} \bar{\boldsymbol{\beta}} + \overline{\sigma^{-2}} \mathbf{x}' (\mathbf{y} \pm \bar{\mathbf{u}} \otimes \mathbf{i}_T) \right) \quad \mathbf{V}_\beta = \left( \overline{\sigma^{-2}} \mathbf{x}' \mathbf{x} + \mathbf{V}_\beta^{-1} \right)^{-1} \\
q(\sigma^{-2}) \leftarrow G(A_\sigma, B_\sigma) \quad A_\sigma = \underline{A}_\sigma + \frac{NT}{2} \quad B_\sigma = \underline{B}_\sigma + \frac{1}{2} \left( \sum_{i=1}^N \sum_{t=1}^T \left[ (y_{it} - \mathbf{x}_{it} \bar{\boldsymbol{\beta}} \pm \bar{u}_i)^2 + V[u_i] \right] + \text{tr}(\mathbf{x}' \mathbf{x} \mathbf{V}_\beta) \right) \\
q(\lambda) \leftarrow G(A_\lambda, B_\lambda) \quad A_\lambda = N + \underline{A}_\lambda \quad B_\lambda = \sum_{i=1}^N \bar{u}_i + \underline{B}_\lambda \\
q(u_i) \leftarrow TN(\mu_i, \nu^2) \quad \mu_i = \frac{1}{T} \left( \pm \sum_{t=1}^T (\mathbf{x}_{it} \bar{\boldsymbol{\beta}} - y_{it}) - \frac{\bar{\lambda}}{\sigma^{-2}} \right) \quad \nu^2 = \frac{1}{T \sigma^{-2}}
\end{array} \right. \quad (3.4)$$

In (3.4)  $\text{tr}(\cdot)$  and  $TN(\cdot, \cdot)$  denote the trace of a matrix and the truncated normal density function, respectively,  $\mathbf{x}$  is an  $(NT \times k)$  matrix with  $(i, t)$ -th row  $\mathbf{x}_{it}$ , and  $\mathbf{i}_T$  is a  $T$ -dimensional vector of ones. The quantities  $\bar{\boldsymbol{\beta}}$ ,  $\overline{\sigma^{-2}}$ ,  $\bar{u}_i$ ,  $\bar{\lambda}$ ,  $\mathbf{V}_\beta$  and  $V[u_i]$  are the relevant means and variances from the  $q$  densities in (3.4); e.g.,  $\overline{\sigma^{-2}} = E_{q(\sigma^{-2})}(\sigma^{-2})$ . They appear in the above distributions when we take expectations of the form  $E_{\boldsymbol{\theta}_{-i}}[\log p(\boldsymbol{\theta}, \mathbf{y})]$ . Using well-known results on expectations of normal, gamma, and truncated normal distributions, we can show that

$$\overline{\sigma^{-2}} = \frac{A_\sigma}{B_\sigma} \quad \bar{\lambda} = \frac{A_\lambda}{B_\lambda} \quad \bar{u}_i = \mu_i + \nu m(\alpha_i) \quad V[u_i] = \nu^2 \left\{ 1 - m(\alpha_i) [m(\alpha_i) + \alpha_i] \right\} \quad (3.5)$$

where  $\alpha_i = \mu_i / \nu$  and  $m(\cdot) = \phi(\cdot) / \Phi(\cdot)$ ;  $\phi(\cdot)$  and  $\Phi(\cdot)$  are the *pdf* and *cdf* of the standard normal distribution. With this information, we can establish a coordinate ascent algorithm where we use initial starting values for some of the moments and then iterate the values of  $\{\bar{\boldsymbol{\beta}}, \mathbf{V}_\beta, \overline{\sigma^{-2}}, \bar{\lambda}, \bar{u}_i, V[u_i]\}$  until the change in  $\ln \underline{ML}$  is negligible.

To find an expression for this lower bound of the marginal likelihood, we need to evaluate the two components of

$$\ln \underline{ML} = E_q \ln p(\mathbf{y}, \boldsymbol{\beta}, \mathbf{u}, \lambda, \sigma^{-2}) - E_q \ln q(\boldsymbol{\beta}, \mathbf{u}, \lambda, \sigma^{-2}) \quad (3.6)$$

It can be shown that the first component in (3.6) is given by

$$E_q \ln p(\mathbf{y}, \boldsymbol{\beta}, \mathbf{u}, \lambda, \sigma^{-2}) = C + (A_\sigma - 1) [\psi(B_\sigma) - \ln(A_\sigma)] - A_\sigma + (A_\lambda - 1) [\psi(B_\lambda) - \ln(A_\lambda)] - A_\lambda$$

where we have used expressions for the mean of the logarithm of a gamma distributed random variable and  $\psi(x) = d \log \Gamma(x) / dx$  is the digamma function. For the second component in (3.6) we have

$$E_q \ln q(\boldsymbol{\beta}, \mathbf{u}, \lambda, \sigma^{-2}) = E \ln q(\boldsymbol{\beta}) + E \ln q(\sigma^{-2}) + E \ln q(\lambda) + \sum_{i=1}^N E \ln q(u_i)$$

Using results on entropy of normal, gamma and truncated normal distributions, after some algebraic manipulation we obtain

$$\begin{aligned} \ln \underline{ML} = & \frac{(-NT + N) \ln 2\pi + N + k}{2} + \ln \left( \frac{\Gamma(A_\sigma) \Gamma(A_\lambda) \underline{B}_\sigma^{A_\sigma} \underline{B}_\lambda^{A_\lambda}}{\Gamma(\underline{A}_\sigma) \Gamma(\underline{A}_\lambda) \underline{B}_\lambda^{A_\lambda} \underline{B}_\sigma^{A_\sigma}} \right) + \frac{1}{2} \ln \left( \frac{|\mathbf{V}_\beta|}{|\underline{\mathbf{V}}_\beta|} \right) \\ & - \frac{(\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}})' \underline{\mathbf{V}}_\beta^{-1} (\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}) + \text{tr}(\underline{\mathbf{V}}_\beta^{-1} \underline{\mathbf{V}}_\beta)}{2} + N \ln \nu + \sum_{i=1}^N \left\{ \ln \Phi(\alpha_i) - \frac{\alpha_i \phi(\alpha_i)}{2\Phi(\alpha_i)} \right\} \end{aligned} \quad (3.7)$$

where  $k$  is the number of regressors.

### 3.2 Gamma inefficiency

The approximate posterior densities in (3.4) are all standard recognizable forms. With an assumption that the inefficiency error follows a half-normal distribution, i.e.,  $u_i \sim N^+(0, \lambda^2)$ , the resulting approximate posterior densities are again recognizable forms. Details are given in Appendix A. The other inefficiency error distributions that we consider, gamma, truncated normal, and lognormal, all have one or more approximate posterior densities that are not recognizable, making it necessary to use numerical integration to obtain some of the required moments. We now consider the gamma case as an example; results for the others are given in Appendix A.

Suppose  $u_i \sim G(\theta, \lambda)$ , and, following Griffin and Steel (2007), we use the priors  $(\lambda | \theta) \sim G(\theta, \underline{B}_\lambda)$  and  $\theta^{-1} \sim G(\underline{A}_\theta, \underline{B}_\theta)$ . For the other parameters, we retain the same priors as used for the exponential case. Then, the log of the posterior kernel can be obtained as

$$\begin{aligned} \ln p(\mathbf{y}, \boldsymbol{\beta}, \mathbf{u}, \lambda, \theta, \sigma^{-2}) = & C + \left( \frac{NT}{2} + \underline{A}_\sigma - 1 \right) \ln \sigma^{-2} - \frac{\sigma^{-2}}{2} \left( \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \mathbf{x}_{it} \boldsymbol{\beta} \pm u_i)^2 + 2\underline{B}_\sigma \right) \\ & - \frac{1}{2} (\boldsymbol{\beta} - \underline{\boldsymbol{\beta}})' \mathbf{V}_\beta^{-1} (\boldsymbol{\beta} - \underline{\boldsymbol{\beta}}) + (\theta - 1) \sum_{i=1}^N \ln u_i - \lambda \left( \sum_{i=1}^N u_i + \underline{B}_\lambda \right) + [(N+1)\theta - 1] \ln \lambda \quad (3.8) \\ & - (N+1) \ln \Gamma(\theta) + \theta \ln \underline{B}_\lambda - (\underline{A}_\theta - 1) \ln \theta - \underline{B}_\theta \theta^{-1} \end{aligned}$$

where  $C = -((NT + k)/2) \ln 2\pi + \underline{A}_\sigma \ln \underline{B}_\sigma - \ln \Gamma(\underline{A}_\sigma) + \underline{A}_\theta \ln \underline{B}_\theta - \ln \Gamma(\underline{A}_\theta) - (\ln |\mathbf{V}_\beta|)/2$ . From (3.8) and (2.4), and using a suitable factorization, it can be shown that the optimal densities are given by the following iterative process.

$$\left\{ \begin{array}{l} q(\boldsymbol{\beta}) \leftarrow N(\bar{\boldsymbol{\beta}}, \mathbf{V}_\beta) \quad \bar{\boldsymbol{\beta}} = \mathbf{V}_\beta \left( \mathbf{V}_\beta^{-1} \underline{\boldsymbol{\beta}} + \overline{\sigma^{-2}} \mathbf{x}' (\mathbf{y} \pm \bar{\mathbf{u}} \otimes \mathbf{i}_T) \right) \quad \mathbf{V}_\beta = \left( \overline{\sigma^{-2}} \mathbf{x}' \mathbf{x} + \mathbf{V}_\beta^{-1} \right)^{-1} \\ q(\sigma^{-2}) \leftarrow G(A_\sigma, B_\sigma) \quad A_\sigma = \underline{A}_\sigma + \frac{NT}{2} \quad B_\sigma = \underline{B}_\sigma + \frac{1}{2} \left( \sum_{i=1}^N \sum_{t=1}^T \left[ (y_{it} - \mathbf{x}_{it} \bar{\boldsymbol{\beta}} \pm \bar{u}_i)^2 + V[u_i] \right] + \text{tr}(\mathbf{x}' \mathbf{x} \mathbf{V}_\beta) \right) \\ q(\lambda) \leftarrow G(\bar{A}_\lambda, \bar{B}_\lambda) \quad \bar{A}_\lambda = (N+1)\bar{\theta} \quad \bar{B}_\lambda = \sum_{i=1}^N \bar{u}_i + \underline{B}_\lambda \\ q(\theta) \leftarrow C_1 \theta^{-\underline{A}_\theta + 1} \exp \left\{ \left( (N+1) \overline{\ln \lambda} + \ln \underline{B}_\lambda + \sum_{i=1}^N \overline{\ln u_i} \right) \theta - \underline{B}_\theta \theta^{-1} - (N+1) \ln \Gamma(\theta) \right\} \\ q(u_i) \leftarrow \frac{u_i^{\bar{\theta}-1} \exp \left\{ -\frac{\nu^2}{2} u_i^2 - \mu_i u_i \right\}}{\nu^{-\bar{\theta}} \exp \left\{ \mu_i^2 / 4\nu^2 \right\} \Gamma(\bar{\theta}) D_{-\bar{\theta}}(\mu_i/\nu)} \quad \nu^2 = T \overline{\sigma^{-2}} \quad \mu_i = \left( \pm \sum_{t=1}^T \overline{\sigma^{-2}} (y_{it} - \mathbf{x}_{it} \bar{\boldsymbol{\beta}}) - \bar{\lambda} \right) \end{array} \right. \quad (3.9)$$

The normalizing constant for  $q(\theta)$  has been left unspecified as  $C_1$ . The results in Beckers and Hammond (1987) have been used to find the normalizing constant for  $q(u_i)$ . In this constant  $D_{-\bar{\theta}}(\cdot)$  denotes the parabolic cylinder function. See Beckers and Hammond (1987, p.36) for details. The moments of  $u_i$  are given by

$$\overline{u_i^m} = \frac{\Gamma(\overline{\theta} + m) D_{-\overline{\theta}-m}(\mu_i/\nu)}{\nu^m \Gamma(\overline{\theta}) D_{-\overline{\theta}}(\mu_i/\nu)}$$

To set up a coordinate ascent algorithm, we have

$$\begin{aligned} \overline{\sigma^{-2}} &= \frac{A_\sigma}{B_\sigma} & \overline{\lambda} &= \frac{A_\lambda}{B_\lambda} & \frac{1}{C_1} &= \int_0^\infty q^*(\theta) d\theta & \overline{\theta} &= \int_0^\infty \theta q(\theta) d\theta & \overline{\ln \theta} &= \int_0^\infty \ln \theta q(\theta) d\theta \\ \overline{\ln q(\theta)} &= \int_0^\infty \ln q(\theta) q(\theta) d\theta & \overline{\ln u_i} &= \int_0^\infty \ln u_i q(u_i) du_i & V[u_i] &= \overline{u_i^2} - \overline{u_i}^2 \end{aligned} \quad (3.10)$$

where  $q^*(\theta) = q(\theta)/C_1$ . The optimal densities are obtained by iterating  $\overline{\boldsymbol{\beta}}$ ,  $\mathbf{V}_\beta$  and the quantities in (3.10) until convergence, with convergence assessed using changes in  $\ln \underline{ML}$ .

The integrals in (3.10) are all univariate integrals and usually straightforward to compute numerically. However, two problems may arise: First, sometimes the values of these integrals are extremely large or extremely small; in other words, we might face overflows or underflows. Second, the integrals are often taken over an infinite or semi-infinite region, rather than a compact interval, making it important to determine the effective support of the integrand for accurate computation. To avoid such problems that can arise in troublesome cases, in our empirical examples we calculate the integrals using the procedure described in Appendix B of Wand et al. (2011).

To calculate the lower bound for the marginal likelihood we first find expressions for  $E_q \ln p(\mathbf{y}, \boldsymbol{\beta}, \mathbf{u}, \lambda, \sigma^{-2}, \theta)$  and  $E_q \ln q(\boldsymbol{\beta}, \mathbf{u}, \lambda, \sigma^{-2}, \theta)$ . Then, after some algebra, it can be shown that

$$\begin{aligned} \ln \underline{ML} &= \frac{-NT \ln 2\pi + k}{2} + \ln \left( \frac{\Gamma(A_\sigma) \Gamma(A_\lambda) \underline{B}_\sigma^{A_\sigma} \underline{B}_\lambda^{A_\lambda} \underline{B}_\theta^{A_\theta}}{\Gamma(\underline{A}_\sigma) \Gamma(\underline{A}_\lambda) \underline{B}_\sigma^{A_\sigma} \underline{B}_\lambda^{A_\lambda}} \right) - \frac{(\overline{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}})' \mathbf{V}_\beta^{-1} (\overline{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}) + \text{tr}(\mathbf{V}_\beta^{-1} \mathbf{V}_\beta)}{2} \\ &+ \frac{1}{2} \ln \left( \frac{|\mathbf{V}_\beta|}{|\underline{\mathbf{V}}_\beta|} \right) + (\overline{\theta} - 1) \sum_{i=1}^N \overline{\ln u_i} - \sum_{i=1}^N \overline{\ln q(u_i)} - (N+1) \overline{\ln \Gamma(\overline{\theta})} - (\underline{A}_\theta - 1) \overline{\ln \theta} - \underline{B}_\theta \overline{\theta^{-1}} - \overline{\ln q(\overline{\theta})} \end{aligned} \quad (3.11)$$

## 4. Applications

Using three well-known data sets, in this section we estimate stochastic frontier models with exponential, half-normal, gamma, truncated normal and lognormal distributions using both MFVB and MCMC, and compare the estimated models using the proposed Bayesian model selection procedure described in Section 3.

### 4.1 Models and data

The first data set, collected by the International Rice Research Institute (IRRI) consists of a panel of 43 Philippine rice farms observed over 8 years from 1990 to 1997 — see Coelli, et al. (2005) for further information. Because we are assuming time-invariant inefficiencies, only the last 4 years are considered. The model estimated is

$$\ln y_{it} = \beta_0 + \beta_1 \ln(\text{land}_{it}) + \beta_2 \ln(\text{labor}_{it}) + \beta_3 \ln(\text{fert}_{it}) + \beta_4 \ln(\text{others}_{it}) - u_i + v_{it} \quad (4.1)$$

where  $y_{it}$ ,  $\text{land}_{it}$ ,  $\text{labor}_{it}$ ,  $\text{fert}_{it}$  and  $\text{others}_{it}$  represent output, land, hired labour, amount of fertilizer and other inputs, respectively.

In the second example six years of observations on 247 dairy farms are used to estimate a translog production frontier where milk production ( $y$ ) depends on the inputs feed ( $x_1$ ), land ( $x_2$ ), labor ( $x_3$ ) and cows ( $x_4$ ), and a trend ( $t$ ) for technological change — see Alvarez et al. (2005) for further information. The function is

$$\begin{aligned} y_{it} = & \beta_0 + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \beta_3 x_{3,it} + \beta_4 x_{4,it} + \beta_{11} x_{1,it}^2 + \beta_{22} x_{2,it}^2 + \beta_{33} x_{3,it}^2 + \beta_{44} x_{4,it}^2 + \beta_{12} x_{1,it} x_{2,it} \\ & + \beta_{13} x_{1,it} x_{3,it} + \beta_{14} x_{1,it} x_{4,it} + \beta_{23} x_{2,it} x_{3,it} + \beta_{24} x_{2,it} x_{4,it} + \beta_{34} x_{3,it} x_{4,it} + \gamma_1 t + \gamma_{11} t^2 - u_i + v_{it} \end{aligned}$$

The third example uses a cross-sectional data set of  $N = 122$  U.S. electric utilities in 1970. The data were originally analysed by Christensen and Greene (1976) and then by Greene (1990) and by Griffin and Steel (2007). Following them, we consider the following frontier cost function

$$\ln(C_i / pf_i) = \beta_0 + \beta_1 \ln(Q_i) + \beta_2 \ln(Pl_i / Pf_i) + \beta_3 \ln(Pk_i / Pf_i) + \beta_4 \ln(Q_i^2) + u_i + v_i \quad (4.2)$$

where output ( $Q$ ) is produced at cost ( $C$ ) using three factors: labor, capital, and fuel, whose respective factor prices are  $Pl$ ,  $Pk$  and  $Pf$ .

## 4.2 Prior specifications

The general forms for the priors were those given earlier for the exponential and gamma models and those given in Appendix A for the others. For all models, we consider the vague priors  $\boldsymbol{\beta} \sim N(\mathbf{0}, 0.001^2 \mathbf{I}_k)$  and  $\sigma_v^{-2} \sim G(0.001, 0.001)$ . Hyper-parameters for the  $u_i$  distributions are specified as  $a_\lambda = 1$ ,  $b_\lambda = \ln(0.875)$  for the exponential,  $a_\lambda = 1$ ,  $b_\lambda = 1/37.5$  for the half-normal,  $b_\lambda = \ln(0.875)$ ,  $a_\theta = 3$ ,  $b_\theta = 4$  for the gamma,  $a_\mu = 0$ ,  $b_\mu = 4$ ,  $a_\lambda = 1$ ,  $b_\lambda = 1/37.5$  for the truncated normal, and  $a_\mu = -2$ ,  $b_\mu = 4$ ,  $a_\lambda = 1$ ,  $b_\lambda = 1$  for the lognormal model. For exponential, gamma and half-normal, these are values that have been suggested in the literature (see e.g., van den Broeck et al. 1994 or Griffin and Steel 2007). For the truncated normal and lognormal we have specified values in such a way that the implied marginal prior for  $u_i$  resembles the implied prior for the gamma model. After using simulation to integrate out their hyper-parameters, the implied marginal priors for  $u_i$  all have similar shapes. See Figure 1.

## 4.3 Parameter estimates

Tables 1, 2 and 3 contain the posterior means of the parameters for the various models and the three data sets using both MFVB and MCMC. We wrote R code for MFVB and for MCMC we used WINBUGS, following the instructions in Griffin and Steel (2007). Convergence of the MFVB algorithms was achieved in less than 500 iterations in all cases except for the gamma model applied to the electricity data; it took 1000 iterations. We tried a number of different starting values but always reached the same solution. In all MCMC cases we considered 60000 iterations, discarding the first 50000 as the burn-in period.

The signs and magnitudes of the production elasticity estimates for *land*, *labour*, *fert* and *others* given in Table 1 are all consistent with expectations. They are also almost identical

across the models and estimation methods. However, the inefficiency parameter estimates are somewhat different across the two estimation methods. In the most extreme case, the parameter  $\lambda$  for the truncated normal model is estimated as 17.689 from MCMC versus 13.096 from MFVB. The probability intervals for these parameters are generally very wide, however. For example, the 95% probability interval for  $\lambda$  from MCMC is (5.210, 47.902), and, as we see later, the estimated inefficiency scores, which are often the main item of interest, are reasonably close despite such differences. Similar conclusions can be drawn from Tables 2 and 3 that contain the results for the dairy and electricity data examples. The estimates for the technology parameters are almost identical across the models and estimation methods. Parameters of inefficiency distributions differ slightly across estimations methods but are again well within 95% probability intervals.

#### **4.4 Inefficiency estimates**

In stochastic frontier analysis, estimation of inefficiency effects is often of primary importance and frequently the main objective of the study. It is important, therefore, to investigate the accuracy of estimated inefficiency scores from MFVB. In Figures 2, 3 and 4 MFVB and MCMC estimates of the inefficiency effects are compared for all models and data sets. In each graph, the horizontal axis represents estimated effects from MCMC and the vertical axis represents estimates from MFVB. We have also depicted the 45 degree line; if MFVB and MCMC estimates are identical they will lie on the 45 degree line. For the rice data and the dairy data in Figures 2 and 3, they are extremely close. For the electricity data in Figure 4, there are some visible differences, particularly for the half-normal and truncated normal distributions. However, even in these cases the ranking of firms is generally preserved. For the case of the half-normal model where there is the largest discrepancy, we have also depicted the (25%, 75%) probability bands from MCMC. The MFVB estimates are well within the interval given

by these bands. It is also worth noting that the electricity data are not a panel. Without repeated observations on the  $u_i$ , we expect their estimation to be less precise.

Using the rice data as an example, other features of the posterior distributions are compared in Table 4 and Figure 5<sup>3</sup>. Table 4 contains posterior standard deviations for all the parameters. Consistent with other studies such as Wang and Titterton (2005), the standard deviations from MFVB are smaller than those from MCMC for all parameters. For the technology parameters, the inefficiency effects, and  $\sigma^{-2}$ , the difference is not substantial, but for the inefficiency parameters there are substantial differences. Marginal posterior densities for selected inefficiency effects are compared in Figure 5. Those from the MCMC sample were obtained by averaging the conditional distributions of the  $u_i$ s over the MCMC draws of the other parameters. Again we find the differences are not substantial, especially for the effects which are smaller in magnitude. Relative to the rice data example, the MFVB and MCMC marginal posteriors are closer for the dairy data example and further apart for the electricity data.

#### 4.5 Model comparison

Despite the wide variety of distributions suggested in the literature for efficiency effects, there have been few studies formally comparing these models. The Bayesian exceptions that we are aware of are Griffin and Steel (2008) who compare alternative distributions using a generalization of the Savage-Dickey method and Ehlers (2011) who uses the DIC statistics produced by WINBUGS<sup>4</sup>. Here we use the methodology described earlier to compute Bayes factors which are generally recognized as the preferred Bayesian criterion for comparing models. The five distributional assumptions are compared for each of the three data sets.

---

<sup>3</sup> The same general conclusions can be drawn from the other data sets. Comparable tables and figures are available from the authors on request.

<sup>4</sup> The DIC (deviance information criterion) is a goodness-of-fit with-penalty measure similar in nature to the Akaike and Bayesian information criteria. Because it uses averages rather than maxima as estimates, it is readily computed from MCMC output.

In Table 5 we report the lower-bounds for the logs of the marginal likelihoods from MFVB, the MCMC-estimated logs of the marginal likelihoods, computed using our proposal where the MFVB posterior is used in conjunction with the Gelfand-Dey method, and the expected mean-squared errors (MSE) of the log-marginal-likelihood estimates<sup>5</sup>. For all data sets, the largest marginal likelihoods are those for the gamma model; the differences between the marginal likelihoods for the other four models are much less pronounced. The MSE values suggest the log-marginal likelihoods have been estimated with a high degree of accuracy. As expected, the MFVB lower bounds for the marginal likelihoods are smaller than the corresponding estimates from MCMC. Because these differences are not uniform, the Bayes' factors computed from the lower bound are not necessarily a good guide to the more accurate comparison provided by the Bayes' factors computed from the MCMC estimates. In Table 6 we report the logs of the Bayes factors against the half-normal model. In each case the first and second ranked models are the same irrespective of whether the MFVB lower bounds or the MCMC estimates are used, but there is some divergence after that. The gamma model is clearly favoured for all data sets. For the rice and electricity data sets the exponential model is ranked second, and for the dairy data the truncated normal model is ranked second.

## 5. Extensions: models with environmental variables

It is often of interest to identify and estimate the effects of factors influencing technical efficiencies. A standard approach to do this is to make the mean of the inefficiency distribution a function of the factors. We follow this tradition and describe MFVB estimation of two popular models incorporating environmental factors. MFVB estimation of one model where the exogenous environmental variables take the form of 0-1 dummy variables is discussed in Appendix B. A second model, the one that we discuss in this section, incorporates general

---

<sup>5</sup> Methods for computing the MSE are described in Chen et al. (2000, p.133-145) and Frühwirth-Schnatter (2006, p.152-154).

exogenous variables along the lines proposed by, for example, Battese and Coelli (1995) and Alvarez et al. (2006). Specifically, we consider the following model with a lognormal distribution for the effects<sup>6</sup>

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} \mp u_{it} + v_{it} \quad (5.1)$$

The  $u_{it}$ s are assumed to be independently but not identically distributed and are drawn from  $u_{it} \sim LN(m_{it}, \lambda^2)$  where  $m_{it} = \mathbf{z}_{it}\boldsymbol{\mu}$ ;  $\mathbf{z}_{it}$  is a  $1 \times M$  vector of environmental variables; and  $\boldsymbol{\mu}$  is an  $M \times 1$  vector of unknown parameters to be estimated. From the lognormal assumption, we can write  $u_{it} = \exp(\mathbf{z}_{it}\boldsymbol{\mu})u_{it}^0$  where  $u_{it}^0 \sim LN(1, \lambda^2)$  does not depend on  $\mathbf{z}_{it}$ . Alvarez et al. (2006) have argued in favour of this scaling property in the context of models with environmental variables. This specification has also the convenient feature that  $\mu_j = \partial \ln E(u) / \partial z_j$  implying  $100 \times \mu_j$  can be interpreted as the percentage change in  $E(u)$  from a one unit change in  $z_j$ .

The priors specified for  $(\boldsymbol{\beta}, \sigma^2, \lambda^{-2})$  are those used for the lognormal distribution in Appendix A, namely  $\boldsymbol{\beta} \sim N(\underline{\boldsymbol{\beta}}, \underline{\mathbf{V}}_{\boldsymbol{\beta}})$ ,  $\sigma^{-2} \sim G(\underline{A}_{\sigma}, \underline{B}_{\sigma})$ , and  $\lambda^{-2} \sim G(\underline{A}_{\lambda}, \underline{B}_{\lambda})$ . For  $\boldsymbol{\mu}$  we use  $\boldsymbol{\mu} \sim N(\underline{\boldsymbol{\mu}}, \underline{\mathbf{V}}_{\boldsymbol{\mu}})$ . Using a suitable factorization we can derive the optimal densities as

---

<sup>6</sup> It is possible to consider other distributions such as truncated normal. We consider a lognormal distribution because it is more convenient for obtaining conditional distributions for the parameters  $\boldsymbol{\mu}$  that appear in the inefficiency distribution.

$$\left\{ \begin{array}{l}
q(\boldsymbol{\beta}) \leftarrow N(\bar{\boldsymbol{\beta}}, \mathbf{V}_\beta) \quad \bar{\boldsymbol{\beta}} = \mathbf{V}_\beta \left( \mathbf{V}_\beta^{-1} \bar{\boldsymbol{\beta}} + \overline{\sigma^{-2} \mathbf{x}' (\mathbf{y} \pm \bar{\mathbf{u}} \otimes \mathbf{i}_T)} \right) \quad \mathbf{V}_\beta = \left( \overline{\sigma^{-2} \mathbf{x}' \mathbf{x}} + \mathbf{V}_\beta^{-1} \right)^{-1} \\
q(\sigma^{-2}) \leftarrow G(A_\sigma, B_\sigma) \quad A_\sigma = \underline{A}_\sigma + \frac{NT}{2} \quad B_\sigma = \underline{B}_\sigma + \frac{1}{2} \left( \sum_{i=1}^N \sum_{t=1}^T \left[ (y_{it} - \mathbf{x}_{it} \bar{\boldsymbol{\beta}} \pm \bar{u}_i)^2 + V[u_i] \right] + \text{tr}(\mathbf{x}' \mathbf{x} \mathbf{V}_\beta) \right) \\
q(\boldsymbol{\mu}) \leftarrow N(\bar{\boldsymbol{\mu}}, \mathbf{V}_\mu) \quad \bar{\boldsymbol{\mu}} = \left( \overline{\lambda^{-2} \mathbf{z}' \mathbf{z}} + \mathbf{V}_\mu \right)^{-1} \left( \overline{\lambda^{-2} \mathbf{z}' \ln \mathbf{u}} + \mathbf{V}_\mu \bar{\boldsymbol{\mu}} \right) \quad \mathbf{V}_\mu = \left( \overline{\lambda^{-2} \mathbf{z}' \mathbf{z}} + \mathbf{V}_\mu \right)^{-1} \\
q(\lambda^{-2}) \leftarrow G(A_\lambda, B_\lambda) \quad A_\lambda = \frac{NT}{2} + \underline{A}_\lambda \quad B_\lambda = \frac{1}{2} \left( \sum_{i=1}^N \sum_{t=1}^T \left[ \overline{(\ln u_{it})^2} - 2(\mathbf{z}_{it} \bar{\boldsymbol{\mu}}) \overline{\ln u_{it}} \right] + \text{tr} \left[ (\mathbf{z}' \mathbf{z}) \overline{\boldsymbol{\mu} \boldsymbol{\mu}'} \right] \right) + \underline{B}_\lambda \\
q(u_{it}) \leftarrow C_{1,it} \exp \left\{ \left( \overline{\lambda^{-2} \mathbf{z}_{it} \bar{\boldsymbol{\mu}} - 1} \right) \ln u_{it} - 0.5 \overline{\lambda^{-2}} (\ln u_{it})^2 - 0.5 \overline{\sigma^{-2}} (u_{it}^2 \mp 2u_{it} (\mathbf{x}_{it} \bar{\boldsymbol{\beta}} - y_{it})) \right\}
\end{array} \right. \quad (5.2)$$

where  $C_{1,it}$  is a normalizing constant,  $\mathbf{z}$  is an  $(NT \times M)$  matrix containing the  $\mathbf{z}_{it}$ , and  $\bar{\mathbf{u}}$  is now of dimension  $NT$ . All the densities in (5.2) are of a standard form except that for  $u_{it}$  whose various moments need to be obtained by numerical integration. The coordinate ascent algorithm can be set up by iterating over the moments

$$\left\{ \bar{\boldsymbol{\beta}}, \mathbf{V}_\beta, \overline{\sigma^{-2}}, \bar{\boldsymbol{\mu}}, \mathbf{V}_\mu, \overline{\lambda^{-2}}, \bar{u}_i, \overline{\ln u_{it}}, \overline{(\ln u_{it})^2}, V[u_i], \overline{\boldsymbol{\mu} \boldsymbol{\mu}'} \right\}$$

until the change in the following lower bound for the log of marginal likelihood is negligible.

$$\begin{aligned}
\ln \underline{ML} = & \frac{-2NT \ln 2\pi + k + M}{2} + \ln \left( \frac{\Gamma(A_\sigma) \underline{B}_\sigma^{A_\sigma} \Gamma(A_\lambda) \underline{B}_\lambda^{A_\lambda}}{\Gamma(\underline{A}_\sigma) \underline{B}_\sigma^{A_\sigma} \Gamma(\underline{A}_\lambda) \underline{B}_\lambda^{A_\lambda}} \right) - \frac{(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{V}_\beta^{-1} (\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \text{tr}(\mathbf{V}_\beta^{-1} \mathbf{V}_\beta)}{2} \\
& + \frac{1}{2} \ln \left( \frac{|\mathbf{V}_\beta|}{|\underline{\mathbf{V}}_\beta|} \right) + \frac{1}{2} \ln \left( \frac{|\mathbf{V}_\mu|}{|\underline{\mathbf{V}}_\mu|} \right) - \frac{(\bar{\boldsymbol{\mu}} - \boldsymbol{\mu})' \mathbf{V}_\mu^{-1} (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}) + \text{tr}(\mathbf{V}_\mu^{-1} \mathbf{V}_\mu)}{2} - \sum_{i=1}^N \sum_{t=1}^T (\overline{\ln u_{it}} + \ln q(u_{it}))
\end{aligned} \quad (5.3)$$

To illustrate estimation of this model, we apply it to the rice data with the specification

$$\ln y_{it} = \beta_0 + \beta_1 \ln(\text{land}_{it}) + \beta_2 \ln(\text{labor}_{it}) + \beta_3 \ln(\text{fert}_{it}) + \beta_4 \ln(\text{others}_{it}) - u_{it} + v_{it}$$

$$u_{it} \sim LN(\mu_0 + \mu_1 \text{edu}_{it} + \mu_2 \text{ban}_{it}, \lambda^2)$$

where *edu* and *ban* represent education of the household head in years and the proportion of the area classified as bantog (upland) fields, respectively. For a prior on  $\boldsymbol{\mu}$  we specify

$\mu_0 \sim N(-2, 0.25)$  and  $\mu_i \sim N(0, 10)$ ,  $i = 1, 2$ , the latter being relatively vague priors centred over zero. The off-diagonal elements in  $\mathbf{V}_\mu$  were set to zero. Settings for the other hyperparameters were the same as those used previously. Posterior means and standard deviations for the technology and inefficiency distribution parameters are reported in Table 7. The estimate  $\hat{\mu}_1 = -0.076$  suggests an extra year of education leads to a 7.6% percent decrease in mean inefficiency, while  $\hat{\mu}_2 = -1.061$  suggests an increase in the proportion of bantog fields by one percentage point reduces mean inefficiency by 1.061%. The estimates from MFVB and MCMC are similar in magnitude. A similar conclusion can be reached for the estimated efficiency effects from both methods compared in Figure 6.

## 6. Extensions: a time varying panel data model

Several authors, for example, Cornwell, et al. (1990), Kumbhakar (1990), Battese and Coelli (1992), Lee and Schmidt (1993) and Sickles (2005), have suggested ways of relaxing a strong assumption made in model (3.1), that the inefficiency effects are invariant over time. In the previous section we considered one of those ways: assume the mean of the inefficiency effects depends on environmental factors. In this section we consider another time varying inefficiency model: the one proposed by Lee and Schmidt (1993). Using the model  $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} \mp u_{it} + v_{it}$ , they specify the inefficiency effects as the product of the traditional time invariant component  $u_i$  and a firm invariant component  $\alpha_t$ . That is,  $u_{it} = \alpha_t u_i$ , where  $\alpha_1 = 1$  and  $\alpha_t > 0$ . This specification does not require data on the factors causing inefficiency to change over time, but it does assume the impact of such factors is the same for all firms in a given time period. A value  $\alpha_t < 1$  implies firms in that period are more efficient relative to the base period and vice versa. We assume exponential distributions for both components and gamma priors on the parameters of the exponential distributions. That is,  $u_i \sim \exp(\lambda)$ ,  $\alpha_t \sim \exp(\lambda_\alpha)$ ,  $\lambda \sim G(\underline{A}_\lambda, \underline{B}_\lambda)$

, and  $\lambda_\alpha \sim G(\underline{A}_\alpha, \underline{B}_\alpha)$ . Let  $\mathbf{u}$  and  $\boldsymbol{\alpha}$  be vectors containing the components  $u_i$  and  $\alpha_i$ , respectively. This specification results in the following optimal densities all of which are standard forms.

$$\left\{ \begin{array}{l} q(\boldsymbol{\beta}) \leftarrow N(\bar{\boldsymbol{\beta}}, \mathbf{V}_\beta) \quad \bar{\boldsymbol{\beta}} = \mathbf{V}_\beta \left( \mathbf{V}_\beta^{-1} \underline{\boldsymbol{\beta}} + \overline{\sigma^{-2}} \mathbf{x}' (\mathbf{y} \pm \bar{\mathbf{u}} \otimes \mathbf{i}_T) \right) \quad \mathbf{V}_\beta = \left( \overline{\sigma^{-2}} \mathbf{x}' \mathbf{x} + \mathbf{V}_\beta^{-1} \right)^{-1} \\ q(\sigma^{-2}) \leftarrow G(A_\sigma, B_\sigma) \quad A_\sigma = \underline{A}_\sigma + \frac{NT}{2} \quad B_\sigma = \underline{B}_\sigma + \frac{1}{2} \left( \sum_{i=1}^N \sum_{t=1}^T \left[ (y_{it} - \mathbf{x}_{it}' \bar{\boldsymbol{\beta}} \pm \bar{\alpha}_i \bar{u}_i)^2 + V[u_i] V[\alpha_i] \right] + \text{tr}(\mathbf{x}' \mathbf{x} \mathbf{V}_\beta) \right) \\ q(\lambda) \leftarrow G(A_\lambda, B_\lambda) \quad A_\lambda = N + \underline{A}_\lambda \quad B_\lambda = \sum_{i=1}^N \bar{u}_i + \underline{B}_\lambda \\ q(\lambda_\alpha) \leftarrow G(A_\alpha, B_\alpha) \quad A_\alpha = T - 1 + \underline{A}_\alpha \quad B_\alpha = \sum_{t=2}^T \bar{\alpha}_t + \underline{B}_\alpha \\ q(\alpha_t) \leftarrow TN(\mu_{\alpha_t}, \nu_{\alpha_t}^2) \quad \mu_{\alpha_t} = \frac{1}{\sum_{i=1}^N \bar{u}_i^2} \left( \pm \sum_{i=1}^N (\mathbf{x}_{it}' \bar{\boldsymbol{\beta}} - y_{it}) \bar{u}_i - \frac{\bar{\lambda}_\alpha}{\sigma^{-2}} \right) \quad \nu_{\alpha_t}^2 = \frac{1}{\sigma^{-2} \sum_{i=1}^N \bar{u}_i^2} \quad t = 2, \dots, T \\ q(u_i) \leftarrow TN(\mu_i, \nu_i^2) \quad \mu_i = \frac{1}{\sum_{t=1}^T \bar{\alpha}_t^2} \left( \pm \sum_{t=1}^T (\mathbf{x}_{it}' \bar{\boldsymbol{\beta}} - y_{it}) \bar{\alpha}_t - \frac{\bar{\lambda}}{\sigma^{-2}} \right) \quad \nu_i^2 = \frac{1}{\sum_{t=1}^T \bar{\alpha}_t^2 \sigma^{-2}} \end{array} \right.$$

A coordinate ascent algorithm is established by iterating over

$$\left\{ \bar{\boldsymbol{\beta}}, \overline{\sigma^{-2}}, \bar{\lambda}, \bar{\lambda}_\alpha, \bar{\mathbf{u}}, \bar{\boldsymbol{\alpha}}, \bar{\mathbf{u}}^2, \bar{\boldsymbol{\alpha}}^2, \mathbf{V}_\beta, V[u_i], V[\alpha_i] \right\}$$

The log of the lower bound of the marginal likelihood used to assess convergence is

$$\begin{aligned} \ln \underline{ML} = & \frac{-(NT + k) \ln 2\pi + (N + T - 1 + k) (\ln 2\pi + 1)}{2} + \ln \left( \frac{\Gamma(\underline{A}_\sigma) \underline{B}_\sigma^{\underline{A}_\sigma} \Gamma(\underline{A}_\lambda) \underline{B}_\lambda^{\underline{A}_\lambda} \Gamma(\underline{A}_\alpha) \underline{B}_\alpha^{\underline{A}_\alpha}}{\Gamma(\underline{A}_\sigma) \underline{B}_\sigma^{\underline{A}_\sigma} \Gamma(\underline{A}_\lambda) \underline{B}_\lambda^{\underline{A}_\lambda} \Gamma(\underline{A}_\alpha) \underline{B}_\alpha^{\underline{A}_\alpha}} \right) \\ & - \frac{(\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}})' \mathbf{V}_\beta^{-1} (\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}) + \text{tr}(\mathbf{V}_\beta^{-1} \mathbf{V}_\beta)}{2} + \frac{1}{2} \ln \left( \frac{|\mathbf{V}_\beta|}{|\underline{\mathbf{V}}_\beta|} \right) + \ln \nu^N \nu_\alpha^{T-1} + \sum_{i=1}^N \left\{ \ln \Phi(\mu_i / \nu) - \frac{\mu_i \phi(\mu_i / \nu)}{2\nu \Phi(\mu_i / \nu)} \right\} \\ & + \sum_{t=2}^T \left\{ \ln \Phi(\mu_{\alpha_t} / \nu_\alpha) - \frac{\mu_{\alpha_t} \phi(\mu_{\alpha_t} / \nu_\alpha)}{2\nu_\alpha \Phi(\mu_{\alpha_t} / \nu_\alpha)} \right\} \end{aligned}$$

To illustrate this model, we apply it to the dairy data. For brevity, in Table 8 we report only estimates for the parameters of the inefficiency distribution and the  $\alpha_i$  s. The estimates  $\hat{\boldsymbol{\alpha}}' = [1, 0.93, 0.88, 0.87, 0.85, 0.82]$  suggest a gradual increase in efficiency over time. The posterior means from MFVB and MCMC almost coincide and the posterior standard deviations

are close but smaller for MFVB. The estimated  $u_i$  s for MCMC and MFVB are plotted in Figure 7; again, they almost coincide.

## 7. Extensions: More Robust Models

The basic model in (3.1) has a linear-in-parameters technology, its efficiency effects are drawn from one- or two-parameter distributions, and the error term follows a normal distribution. In this section we consider more flexible models under less stringent assumptions and show how they can be estimated under a MFVB framework. We first extend the model to a case where we have a mixture of distributions for the efficiency effects. Next we introduce nonparametric elements into the technology part of the model, and finally we consider more flexible distributions for the random error terms.

### 7.1 A model with a mixture of distributions for inefficiency effects

There are a variety of distributions that can be used to create greater flexibility in the modelling of the inefficiency effects. Two examples that can be estimated by MFVB using auxiliary variable methods are the generalised beta or truncated- $t$  distributions. Another alternative, and the one we focus on in this section, is a mixture of distributions. We outline an estimation procedure for a general mixture for  $u_i$ , but give specific formulas for a mixture of exponential distributions. Other possibilities such as a mixture of gamma densities could be considered. For example, Griffin and Steel (2008) consider a mixture of generalized gamma distributions.

The density function for  $u_i$  for a general mixture with  $K$  components can be written as

$$p(u_i) = \sum_{j=1}^K \pi_j f(u_i | \boldsymbol{\theta}_j) \quad (7.1)$$

where  $f(u_i | \boldsymbol{\theta}_j)$  is a particular distribution that, for the moment, we leave unspecified,  $\pi_j$  is a component weight, and  $\boldsymbol{\theta}_j$  is a vector of parameters for the  $j$ -th component. It is convenient

to introduce an indicator variable  $z_{ij}$  equal to 1 if  $u_i$  belongs to the  $j$ -th mixture component, and 0 otherwise. Then, for  $\mathbf{z}'_i = (z_{i1}, z_{i2}, \dots, z_{iK})$ , and  $\boldsymbol{\theta} = \{\boldsymbol{\theta}_j, j = 1, 2, \dots, K\}$ , we can write

$$p(u_i | \mathbf{z}_i, \boldsymbol{\theta}) = \prod_{j=1}^K f(u_i | \boldsymbol{\theta}_j)^{z_{ij}} \quad (7.2)$$

$$p(\mathbf{z}_i | \boldsymbol{\pi}) = \prod_{j=1}^K \pi_j^{z_{ij}} \quad (7.3)$$

Using the Dirichlet prior  $p(\boldsymbol{\pi}) \propto \prod_{j=1}^K \pi_j^{\alpha_j - 1}$ , the priors  $\sigma^{-2} \sim G(\underline{A}_\sigma, \underline{B}_\sigma)$  and  $\boldsymbol{\beta} \sim N(\underline{\boldsymbol{\beta}}, \underline{\mathbf{V}}_\beta)$ , and the unspecified prior  $p(\boldsymbol{\theta}_j | \underline{\mathbf{A}}_j)$ , the log of the posterior kernel can be written as

$$\begin{aligned} \ln p(\mathbf{y}, \boldsymbol{\beta}, \mathbf{u}, \mathbf{z}, \boldsymbol{\pi}, \boldsymbol{\theta}, \sigma^{-2}) = & C + \left( \frac{NT}{2} + \underline{A}_\sigma - 1 \right) \ln \sigma^{-2} - \frac{\sigma^{-2}}{2} \left[ \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \mathbf{x}_{it} \boldsymbol{\beta} \pm u_i)^2 + 2\underline{B}_\sigma \right] \\ & + \sum_{i=1}^N \prod_{j=1}^K z_{ij} \ln f(u_i | \boldsymbol{\theta}_j) + \sum_{j=1}^K \ln p(\boldsymbol{\theta}_j | \underline{\mathbf{A}}_j) + \sum_{j=1}^K \left( \sum_{i=1}^N z_{ij} + \alpha_j - 1 \right) \ln \pi_j - \frac{1}{2} (\boldsymbol{\beta} - \underline{\boldsymbol{\beta}})' \underline{\mathbf{V}}_\beta^{-1} (\boldsymbol{\beta} - \underline{\boldsymbol{\beta}}) \end{aligned} \quad (7.4)$$

It can be shown that (7.4), along with an appropriate factorization, leads to the following optimal densities.

$$\left\{ \begin{array}{l} q(\boldsymbol{\beta}) \leftarrow N(\bar{\boldsymbol{\beta}}, \mathbf{V}_\beta) \quad \bar{\boldsymbol{\beta}} = \mathbf{V}_\beta \left( \underline{\mathbf{V}}_\beta^{-1} \underline{\boldsymbol{\beta}} + \sigma^{-2} \mathbf{x}'(\mathbf{y} \pm \bar{\mathbf{u}} \otimes \mathbf{i}_T) \right) \quad \mathbf{V}_\beta = \left( \sigma^{-2} \mathbf{x}' \mathbf{x} + \underline{\mathbf{V}}_\beta^{-1} \right)^{-1} \\ q(\sigma^{-2}) \leftarrow G(A_\sigma, B_\sigma) \quad A_\sigma = \underline{A}_\sigma + \frac{NT}{2} \quad B_\sigma = \underline{B}_\sigma + \frac{1}{2} \left( \sum_{i=1}^N \sum_{t=1}^T \left[ (y_{it} - \mathbf{x}_{it} \bar{\boldsymbol{\beta}} \pm \bar{u}_i)^2 + V[u_i] \right] + \text{tr}(\mathbf{x}' \mathbf{x} \underline{\mathbf{V}}_\beta) \right) \\ q(\mathbf{z}_i) \leftarrow \text{Mult} \left( \exp \left\{ \overline{\ln f(u_i | \boldsymbol{\theta}_1)} + \overline{\ln \pi_1} \right\}, \dots, \exp \left\{ \overline{\ln f(u_i | \boldsymbol{\theta}_K)} + \overline{\ln \pi_K} \right\} \right) \\ q(\boldsymbol{\pi}) = \text{Dir} \left( \underline{\boldsymbol{\alpha}}_1 + \sum_{i=1}^N \bar{z}_{i1}, \dots, \underline{\boldsymbol{\alpha}}_K + \sum_{i=1}^N \bar{z}_{iK} \right) \\ q(u_i) \ \& \ q(\boldsymbol{\theta}) \text{ will depend on the distribution for the mixture components } f(u_i | \boldsymbol{\theta}_j), \\ \text{and the prior distribution } p(\boldsymbol{\theta}_j | \underline{\mathbf{A}}_j). \end{array} \right. \quad (7.5)$$

In (7.5),  $\text{Mult}(\cdot)$  refers to the multinomial distribution with probabilities proportional to its arguments, and  $\text{Dir}(\cdot)$  refers to the Dirichlet distribution. The approximate posterior distributions for  $u_i$  and  $\boldsymbol{\theta}$ , as well as the quantities  $\overline{\ln f(u_i | \boldsymbol{\theta}_j)}$ , depend on what further assumptions are made about  $u_i$  and  $\boldsymbol{\theta}$ . For example, for a mixture of exponential distributions where  $f(u_i | \boldsymbol{\theta}_j) = \lambda_j e^{-\lambda_j u_i}$ , with prior  $\lambda_j \sim G(\underline{A}_j, \underline{B}_j)$ , we have

$$\left\{ \begin{array}{l} \overline{\ln f(u_i | \boldsymbol{\theta}_j)} = -\bar{\lambda}_j \bar{u}_j - \overline{\ln \lambda_j} \\ q(\lambda) \leftarrow G(A_\lambda, B_\lambda) \quad A_j = \sum_{i=1}^N \bar{z}_{ij} + \underline{A}_j \quad B_j = \sum_{i=1}^N \bar{z}_{ij} \bar{u}_i + \underline{B}_j \\ q(u_i) \leftarrow TN(\mu_i, \nu^2) \quad \mu_i = \frac{1}{T} \left[ \pm \sum_{t=1}^T (\mathbf{x}_{it} \bar{\boldsymbol{\beta}} - y_{it}) - \left( \sum_{j=1}^K \bar{z}_{ij} \bar{\lambda}_j \right) / \overline{\sigma^{-2}} \right] \quad \nu^2 = \frac{1}{T \overline{\sigma^{-2}}} \end{array} \right.$$

We can again set up a coordinate ascent algorithm by iterating over appropriate expectations from (7.5) until convergence is achieved.

## 7.2 A semiparametric stochastic frontier model

There are several approaches to nonparametric and semiparametric stochastic frontier modelling – see for example Griffin and Steel (2004), and Parmeter and Kumbhakar (2014) for a review. One approach which has a nice Bayesian counterpart and which can be easily set up within an MFVB framework is one that utilizes spline modelling of the regression function. To introduce this approach, we consider the following stochastic frontier model

$$y_{it} = f(x_{it}) \mp u_i + v_{it} \quad (7.6)$$

where  $x_{it}$  is a univariate regressor whose impact is to be modelled non-parametrically. Extensions to various multivariate settings such as partially linear, additive and fully nonparametric models are relatively straightforward (see e.g., Ruppert, et al. 2003). Using polynomial splines to approximate  $f(x_{it})$ , we can write

$$f(x_{it}) \approx \beta_0 + \beta_1 x_{it} + \dots + \beta_p x_{it}^p + \sum_{k=1}^K w_k (x_{it} - \kappa_k)_+^p \quad (7.7)$$

where  $\{\kappa_1, \kappa_2, \dots, \kappa_K\}$  are points known as knots, chosen within the range of  $x$ ;  $(x_{it} - \kappa_k)_+^p$  is a polynomial term equal to zero when  $x_{it} < \kappa_k$ . Combining (7.6) and (7.7), leads to

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} \mp \mathbf{u} \otimes \mathbf{i}_T + \mathbf{v} \quad (7.8)$$

where  $\mathbf{X}$  is a matrix with  $\mathbf{x}_{it} = \left[ 1, x_{it}, \dots, x_{it}^p, (x_{it} - \kappa_1)_+^p, \dots, (x_{it} - \kappa_K)_+^p \right]$  in its  $(i, t)$ -th row, and  $\boldsymbol{\beta}' = (\beta_0, \dots, \beta_p, w_1, \dots, w_K)$ . In the penalized spline approach, a large number of knots are chosen – for example, 20 knots at equidistant intervals – but to avoid over-fitting a constraint on the variability of the regression function is imposed, often of the form  $\sum_{k=1}^K w_k^2 \leq C$ . An equivalent way of imposing this penalty under a Bayesian framework is to assume the following prior (Ruppert et al. 2003)

$$\boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{K}) \text{ where } \mathbf{K} = \begin{pmatrix} \tau^2 \mathbf{I}_{p+1} & \mathbf{0} \\ \mathbf{0} & \tau^2 \mathbf{I}_K \end{pmatrix} \quad (7.9)$$

From (7.8) and (7.9) the model becomes a linear-in-the-parameters stochastic frontier model with  $\boldsymbol{\beta}$  having the normal prior in (7.9). Assuming  $u_i \sim \exp(\lambda)$ ,  $\lambda \sim G(\underline{A}_\lambda, \underline{B}_\lambda)$  and  $\tau \sim G(\underline{A}_\tau, \underline{B}_\tau)$  we can derive the optimal densities as

$$\left\{ \begin{array}{l} q(\boldsymbol{\beta}) \leftarrow N(\bar{\boldsymbol{\beta}}, \mathbf{V}_\beta) \quad \bar{\boldsymbol{\beta}} = \mathbf{V}_\beta \left( \overline{\sigma^{-2}} \mathbf{x}' (\mathbf{y} \pm \bar{\mathbf{u}} \otimes \mathbf{i}_T) \right) \quad \mathbf{V}_\beta = \left( \overline{\sigma^{-2}} \mathbf{x}' \mathbf{x} + \overline{\mathbf{K}^{-1}} \right)^{-1} \\ q(\sigma^{-2}) \leftarrow G(A_\sigma, B_\sigma) \quad A_\sigma = \underline{A}_\sigma + NT/2, \quad B_\sigma = \underline{B}_\sigma + \frac{1}{2} \left( \sum_{i=1}^N \sum_{t=1}^T \left[ (y_{it} - \mathbf{x}_{it} \bar{\boldsymbol{\beta}} \pm \bar{u}_i)^2 + V[u_i] \right] + \text{tr}(\mathbf{X}' \mathbf{X} \mathbf{V}_\beta) \right) \\ q(\tau^{-2}) \leftarrow G(A_\tau, B_\tau) \quad A_\tau = \underline{A}_\tau + K/2, \quad B_\tau = \underline{B}_\tau + [\bar{\mathbf{w}}' \bar{\mathbf{w}} + \text{tr}(\mathbf{V}_w)]/2 \\ q(\lambda) \leftarrow G(A_\lambda, B_\lambda) \quad A_\lambda = N + \underline{A}_\lambda, \quad B_\lambda = \sum_{i=1}^N \bar{u}_i + \underline{B}_\lambda \\ q(u_i) \leftarrow TN(\mu_i, \nu^2) \quad \mu_i = \frac{1}{T} \left( \pm \sum_{t=1}^T (\mathbf{x}_{it} \bar{\boldsymbol{\beta}} - y_{it}) - \frac{\bar{\lambda}}{\overline{\sigma^{-2}}} \right), \quad \nu^2 = \frac{1}{T \overline{\sigma^{-2}}} \end{array} \right. \quad (7.10)$$

where  $\overline{\mathbf{K}^{-1}} = \begin{pmatrix} \tau^{-2} \mathbf{I}_{p+1} & \mathbf{0} \\ \mathbf{0} & \tau^{-2} \mathbf{I}_K \end{pmatrix}$  and  $\bar{\mathbf{w}}$  is the relevant subvector of  $\bar{\boldsymbol{\beta}}$ . A coordinate ascent algorithm iterates over the moments  $\left\{ \bar{\boldsymbol{\beta}}, \overline{\sigma^{-2}}, \overline{\tau^{-2}}, \overline{\ln \tau^{-2}}, \bar{u}_i, \bar{\lambda}, \mathbf{V}_\beta, V[u_i] \right\}$ . Setting  $\alpha_i = \mu_i / \nu$ , the lower bound of the log of the marginal likelihood is

$$\ln \underline{ML} = \frac{(-NT + N) \ln 2\pi + N + K + (p+1)(1 + \ln \underline{\tau}^{-2})}{2} + \ln \left( \frac{\Gamma(\underline{A}_\sigma) \underline{B}_\sigma^{\underline{A}_\sigma} \Gamma(\underline{A}_\lambda) \underline{B}_\lambda^{\underline{A}_\lambda} \Gamma(\underline{A}_\tau) \underline{B}_\tau^{\underline{A}_\tau}}{\Gamma(\underline{A}_\sigma) \underline{B}_\sigma^{\underline{A}_\sigma} \Gamma(\underline{A}_\lambda) \underline{B}_\lambda^{\underline{A}_\lambda} \Gamma(\underline{A}_\tau) \underline{B}_\tau^{\underline{A}_\tau}} \right) \\ - \frac{\overline{\boldsymbol{\beta}' \mathbf{K}^{-1} \boldsymbol{\beta}} + \text{tr}(\overline{\mathbf{K}^{-1} \mathbf{V}_\beta})}{2} + \frac{1}{2} \ln(|\mathbf{V}_\beta|) + \frac{K}{2} \ln \tau^{-2} + N \ln \nu + \sum_{i=1}^N \left\{ \ln \Phi(\alpha_i) - \frac{\alpha_i \phi(\alpha_i)}{2\Phi(\alpha_i)} \right\}$$

In what follows a limited Monte Carlo experiment with 100 replications is conducted to investigate the performance of the MFVB semiparametric estimator relative to other estimators. We consider a bivariate stochastic frontier model of the form  $y_{it} = f(x_{1,it}, x_{2,it}) - u_i + v_{it}$ . The errors  $v_{it}$  and inefficiencies  $u_i$  are generated from  $N(0, 0.1^2)$  and  $\exp(6)$  distributions, respectively. To have data sets close to real applications, values for  $x_{1,it}$  and  $x_{2,it}$  are taken as the logs of area and labor in the Philippine rice data set. We chose 20 knots and use a space filling algorithm (Nychka et al., 1998) for the location of the knots. For  $f$  we consider the following two cases depicted in Figures 8(a) and 8(b).

$$\text{Translog model:} \quad f(x_1, x_2) = x_1 - 0.1x_1^2 + 1.5x_2 - 0.25x_2^2$$

$$\text{Arctan model:} \quad f(x_1, x_2) = x_1 - 0.1x_1^2 + \arctan(2x_2 - 0.5) + 0.25x_2$$

The first case is a translog function commonly used in production economics; the second function has been chosen so that it is not of a translog form with respect to  $x_2$ , but is still a reasonable candidate for a production function. Four estimators were considered: translog-MFVB, translog-MCMC, semiparametric-MFVB and semiparametric-MCMC. In Figures 9 and 10 the posterior means for the  $u_i$  and for the input elasticities (derivatives with respect to  $x_1$  and  $x_2$ ) for each of the replications, and using each of the estimators, are compared with their true values. The true values for the  $u_i$  and for the derivatives at all data points are on the horizontal axes; their corresponding estimated values are on the vertical axes. Observations scattered evenly around the 45 degree lines drawn in red suggest no bias. A concentration of observations above or below a 45 degree line depicts bias.

The results for the translog model in Figure 9 show that all estimators perform well. The posterior means from MFVB and MCMC are very similar. The semiparametric estimator is slightly less efficient than the parametric estimator applied to the true functional form and, as is typical with nonparametric spline estimators, for the elasticities it performs poorly at the boundary points. For the *arctan* model in Figure 10, the posterior means from MFVB and MCMC are again very similar. However, choosing the incorrect translog functional form has led to significant biases in the elasticities, and has increased the variance of the inefficiency estimates, relative to the results from the semiparametric estimator.

### 7.3 Robust distribution for error term

All the models considered so far have assumed that the distribution of the error term is  $v_{it} \sim N(0, \sigma^2)$ . Using an auxiliary variable method, the analysis can be extended to more robust distributions. Estimation of stochastic frontier models with a  $t$ -distributed error term has been discussed in Griffin and Steel (2007) who used WinBUGS software, and Tsionas (2012) who used a fast Fourier transform method. MFVB estimation for a regression model with not only  $t$ -errors, but also asymmetric Laplace errors has been developed in Wand et al. (2011). Extension to stochastic frontier models for all of these distributions is straightforward. Here we discuss a model with a  $t$ -distributed error. Consider

$$(y_{it} | u_i, \boldsymbol{\beta}, \sigma^2, \nu) \sim t(\mathbf{x}_{it}\boldsymbol{\beta} \mp u_i, \sigma^2, \nu) \quad (7.11)$$

To obtain a posterior that is easy to work with we introduce an auxiliary variable " $a$ " and use the following result (see e.g., Wand et al. 2011)

$$(z | a) \sim N(\mu, a\sigma^2) \quad \text{and} \quad a^{-1} \sim G(\nu/2, \nu/2) \quad \Rightarrow \quad z \sim t(\mu, \sigma^2, \nu) \quad (7.12)$$

Therefore, we can equivalently write (7.11) as

$$(y_{it} | u_i, \boldsymbol{\beta}, \sigma^2, a_{it}) \sim N(\mathbf{x}_{it}\boldsymbol{\beta} \mp u_i, a_{it}\sigma^2) \quad \text{with} \quad (a_{it}^{-1} | \nu) \sim G(\nu/2, \nu/2) \quad (7.13)$$

Letting  $u_i \sim \exp(\lambda)$ , and assuming priors  $\boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{V}_\beta)$ ,  $\sigma^{-2} \sim G(\underline{A}_\sigma, \underline{B}_\sigma)$ ,  $\lambda \sim G(\underline{A}_\lambda, \underline{B}_\lambda)$ , and  $p(\nu) = \underline{\lambda}_\nu \exp(-\underline{\lambda}_\nu \nu)$ , we can derive the optimal densities as

$$\left\{ \begin{array}{l} q(\boldsymbol{\beta}) \leftarrow N(\bar{\boldsymbol{\beta}}, \mathbf{V}_\beta) \quad \bar{\boldsymbol{\beta}} = \mathbf{V}_\beta \left( \mathbf{V}_\beta^{-1} \underline{\boldsymbol{\beta}} + \overline{\sigma^{-2}} \mathbf{x}' \mathbf{D}(\overline{a_{it}^{-1}}) (\mathbf{y} \pm \bar{\mathbf{u}} \otimes \mathbf{i}_T) \right) \quad \mathbf{V}_\beta = \left( \overline{\sigma^{-2}} \mathbf{x}' \mathbf{D}(\overline{a_{it}^{-1}}) \mathbf{x} + \mathbf{V}_\beta^{-1} \right)^{-1} \\ q(\sigma^{-2}) \leftarrow G(A_\sigma, B_\sigma) \quad A_\sigma = \underline{A}_\sigma + NT/2 \quad B_\sigma = \underline{B}_\sigma + \frac{1}{2} \left( \sum_{i=1}^N \sum_{t=1}^T \overline{a_{it}^{-1}} \left[ (y_{it} - \mathbf{x}_{it}' \bar{\boldsymbol{\beta}} \pm \bar{u}_i)^2 + V[u_i] \right] + \text{tr} \left( \mathbf{x}' \mathbf{D}(\overline{a_{it}^{-1}}) \mathbf{x} \mathbf{V}_\beta \right) \right) \\ q(\lambda^{-2}) \leftarrow G(A_\lambda, B_\lambda) \quad A_\lambda = N + \underline{A}_\lambda \quad B_\lambda = \sum_{i=1}^N \bar{u}_i + \underline{B}_\lambda \\ q(u_i) \leftarrow TN(\mu_i, \nu_i) \quad \mu_i = \frac{1}{\sum_{t=1}^T \overline{a_{it}^{-1}}} \left( \pm \sum_{t=1}^T \overline{a_{it}^{-1}} (\mathbf{x}_{it}' \bar{\boldsymbol{\beta}} - y_{it}) - \frac{\bar{\lambda}}{\sigma^{-2}} \right) \quad \nu_i = \frac{1}{\sigma^{-2} \sum_{t=1}^T \overline{a_{it}^{-1}}} \\ q(a_{it}^{-1}) \leftarrow G(A_{a_{it}}, B_{a_{it}}) \quad A_{a_{it}} = \frac{\bar{\nu} + 1}{2} \quad B_{a_{it}} = \frac{\bar{\nu}}{2} + \frac{\overline{\sigma^{-2}}}{2} \left( \left[ (y_{it} - \mathbf{x}_{it}' \bar{\boldsymbol{\beta}} \pm \bar{u}_i)^2 + V[u_i] \right] + \mathbf{x}_{it}' \mathbf{V}_\beta \mathbf{x}_{it}' \right) \\ q(\nu) \leftarrow C_1 \exp \left\{ \frac{1}{2} \left( \sum_{i=1}^N \sum_{t=1}^T \overline{\ln a_{it}^{-1}} - \sum_{i=1}^N \sum_{t=1}^T \overline{a_{it}^{-1}} - 2\underline{\lambda}_\nu \right) \nu + NT \left[ \frac{\nu}{2} \ln \left( \frac{\nu}{2} \right) - \ln \Gamma \left( \frac{\nu}{2} \right) \right] \right\} \end{array} \right. \quad (7.14)$$

where  $\mathbf{D}(\overline{a_{it}^{-1}})$  is a diagonal matrix with the  $\overline{a_{it}^{-1}}$  as its diagonal elements. Using the notation

$\overline{\mathbf{a}^{-1}}$  and  $\overline{\ln \mathbf{a}^{-1}}$  to denote vectors containing the  $\overline{a_{it}^{-1}}$  and  $\overline{\ln a_{it}^{-1}}$ , respectively, a coordinate ascent algorithm can be established by iterating over  $\left\{ \bar{\boldsymbol{\beta}}, \overline{\sigma^{-2}}, \overline{\mathbf{a}^{-1}}, \overline{\ln \mathbf{a}^{-1}}, \bar{\mathbf{u}}, \bar{\nu}, \bar{\lambda}, \mathbf{V}_\beta, V[u_i] \right\}$ .

Setting  $\alpha_i = \mu_i / \nu_i$ , the lower bound of the log of the likelihood function is given by

$$\begin{aligned} \ln \underline{ML} &= \frac{(-NT + N) \ln 2\pi + N + k}{2} + \ln \left( \frac{\Gamma(A_\sigma) \underline{B}_\sigma^{A_\sigma} \Gamma(A_\lambda) \underline{B}_\lambda^{A_\lambda} [\Gamma(A_a)]^{NT} (\nu/2)^{NT\nu/2}}{\Gamma(\underline{A}_\sigma) \underline{B}_\sigma^{A_\sigma} \Gamma(\underline{A}_\lambda) \underline{B}_\lambda^{A_\lambda} [\Gamma(\underline{A}_a)]^{NT} \prod_i \prod_t B_{a_{it}}^{A_a}} \right) \\ &\quad - \frac{(\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}})' \mathbf{V}_\beta^{-1} (\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}) + \text{tr}(\mathbf{V}_\beta^{-1} \mathbf{V}_\beta)}{2} + \frac{1}{2} \ln \left( \frac{|\mathbf{V}_\beta|}{|\underline{\mathbf{V}}_\beta|} \right) + N \ln \nu \\ &\quad + \sum_{i=1}^N \left\{ \ln \Phi(\alpha_i) - \frac{\alpha_i \phi(\alpha_i)}{2\Phi(\alpha_i)} \right\} - \overline{\ln q(\nu)} \end{aligned}$$

## 8. Conclusion

In this article we investigated the scope for application of variational Bayes to the stochastic frontier context. Our findings can be summarized as follows: Firstly, as is evident by the large number of examples that we considered, MFVB can be employed to estimate many of the

existing stochastic frontier models in the literature. Secondly, by comparing posterior means from MFVB and MCMC through three empirical examples, we found that the approximation errors from MFVB could be considered negligible. This result is particularly useful for instances when MCMC convergence is in doubt. Thirdly, despite the variety of stochastic frontier models, there have been few studies comparing alternative models, probably due to difficulty of performing the required tests. We showed that MFVB posterior is a convenient density for employing the Gelfand-Dey marginal likelihood estimation framework, producing accurate estimates of the marginal likelihoods. Thus, MFVB is a useful device for facilitating model comparison even when one prefers MCMC estimates to those from MFVB. Fourthly, our results based on 3 empirical examples favour the gamma distribution over 4 other commonly used distributions. Finally, although it was not explored in this article, MFVB posteriors, or modifications of them with fatter tails, are likely to be good potential candidates for direct importance sampling.

## Appendix A: Models with other inefficiency distributions

### Half Normal Inefficiency

*Model* 
$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} \mp u_i + v_{it}, \quad u_i \sim N^+(0, \lambda^2)$$

*Priors* 
$$\boldsymbol{\beta} \sim N(\underline{\boldsymbol{\beta}}, \underline{\mathbf{V}}_{\boldsymbol{\beta}}) \quad \sigma^{-2} \sim G(\underline{A}_{\sigma}, \underline{B}_{\sigma}) \quad \lambda^{-2} \sim G(\underline{A}_{\lambda}, \underline{B}_{\lambda})$$

### *Optimal Densities*

$$\left\{ \begin{array}{l} q(\boldsymbol{\beta}) \leftarrow N(\bar{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}}) \quad \bar{\boldsymbol{\beta}} = \mathbf{V}_{\boldsymbol{\beta}} \left( \underline{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} \underline{\boldsymbol{\beta}} + \overline{\sigma^{-2}} \mathbf{x}' (\mathbf{y} \pm \bar{\mathbf{u}} \otimes \mathbf{i}_T) \right) \quad \mathbf{V}_{\boldsymbol{\beta}} = \left( \overline{\sigma^{-2}} \mathbf{x}' \mathbf{x} + \underline{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} \right)^{-1} \\ q(\sigma^{-2}) \leftarrow G(A_{\sigma}, B_{\sigma}) \quad A_{\sigma} = \underline{A}_{\sigma} + \frac{NT}{2} \quad B_{\sigma} = \underline{B}_{\sigma} + \frac{1}{2} \left( \sum_{i=1}^N \sum_{t=1}^T \left[ (y_{it} - \mathbf{x}_{it} \bar{\boldsymbol{\beta}} \pm \bar{u}_i)^2 + V[u_i] \right] + \text{tr}(\mathbf{x}' \mathbf{x} \mathbf{V}_{\boldsymbol{\beta}}) \right) \\ q(\lambda^{-2}) \leftarrow G(A_{\lambda}, B_{\lambda}) \quad A_{\lambda} = \frac{N}{2} + \underline{A}_{\lambda} \quad B_{\lambda} = \frac{1}{2} \sum_{i=1}^N \bar{u}_i^2 + \underline{B}_{\lambda} \\ q(u_i) \leftarrow TN(\mu_i, \nu^2) \quad \mu_i = \pm \frac{\overline{\sigma^{-2}} \sum_t (\mathbf{x}_{it} \bar{\boldsymbol{\beta}} - y_{it})}{T \overline{\sigma^{-2}} + \lambda^{-2}} \quad \nu^2 = \frac{1}{T \overline{\sigma^{-2}} + \lambda^{-2}} \end{array} \right.$$

### *Coordinate ascent algorithm*

Iterate the following quantities until the change in  $\ln \underline{ML}$  is negligible

$$\bar{\boldsymbol{\beta}} \quad \mathbf{V}_{\boldsymbol{\beta}} \quad \overline{\sigma^{-2}} = \frac{A_{\sigma}}{B_{\sigma}} \quad \overline{\lambda^{-2}} = \frac{A_{\lambda}}{B_{\lambda}} \quad \bar{u}_i = \mu_i + \nu m(\alpha_i) \quad V[u_i] = \nu^2 \left\{ 1 - m(\alpha_i) [m(\alpha_i) + \alpha_i] \right\}$$

where  $\alpha_i = \mu_i / \nu$  and  $m(\cdot) = \phi(\cdot) / \Phi(\cdot)$ .

### *Lower bound for marginal likelihood*

$$\begin{aligned} \ln \underline{ML} = & \frac{(-NT) \ln 2\pi + k + N + 2N \ln 2}{2} + \ln \left( \frac{\Gamma(A_{\sigma}) \Gamma(A_{\lambda}) \underline{B}_{\sigma}^{A_{\sigma}} \underline{B}_{\lambda}^{A_{\lambda}}}{\Gamma(\underline{A}_{\sigma}) \Gamma(\underline{A}_{\lambda}) \underline{B}_{\sigma}^{A_{\lambda}} \underline{B}_{\lambda}^{A_{\sigma}}} \right) \\ & - \frac{(\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}})' \underline{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} (\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}) + \text{tr}(\underline{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} \mathbf{V}_{\boldsymbol{\beta}})}{2} + \frac{1}{2} \ln \left( \frac{|\mathbf{V}_{\boldsymbol{\beta}}|}{|\underline{\mathbf{V}}_{\boldsymbol{\beta}}|} \right) + N \ln \nu + \sum_{i=1}^N \left\{ \ln \Phi(\alpha_i) - \frac{\alpha_i \phi(\alpha_i)}{2\Phi(\alpha_i)} \right\} \end{aligned}$$

### Truncated-Normal Inefficiency

*Model* 
$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} \mp u_i + v_{it}, \quad u_i \sim N^+(\lambda\mu, \lambda^2)$$

*Priors* 
$$\boldsymbol{\beta} \sim N(\underline{\boldsymbol{\beta}}, \underline{\mathbf{V}}_{\boldsymbol{\beta}}) \quad \sigma^{-2} \sim G(\underline{A}_{\sigma}, \underline{B}_{\sigma}) \quad \mu \sim N(\underline{A}_{\mu}, \underline{B}_{\mu}^2) \quad \lambda^{-1} \sim G(\underline{A}_{\lambda}, \underline{B}_{\lambda})$$

*Optimal densities*

$$\left\{ \begin{array}{l} q(\boldsymbol{\beta}) \leftarrow N(\bar{\boldsymbol{\beta}}, \mathbf{V}_\beta) \quad \bar{\boldsymbol{\beta}} = \mathbf{V}_\beta^{-1} \left( \mathbf{V}_\beta^{-1} \underline{\boldsymbol{\beta}} + \overline{\boldsymbol{\sigma}^{-2}} \mathbf{y} \pm \bar{\mathbf{u}} \otimes \mathbf{i}_T \right) \quad \mathbf{V}_\beta = \left( \overline{\boldsymbol{\sigma}^{-2}} \mathbf{x}' \mathbf{x} + \mathbf{V}_\beta^{-1} \right)^{-1} \\ q(\boldsymbol{\sigma}^{-2}) \leftarrow G(A_\sigma, B_\sigma) \quad A_\sigma = \underline{A}_\sigma + \frac{NT}{2} \quad B_\sigma = \underline{B}_\sigma + \frac{1}{2} \left( \sum_{i=1}^N \sum_{t=1}^T \left[ (y_{it} - \mathbf{x}_{it} \bar{\boldsymbol{\beta}} \pm \bar{u}_i)^2 + V[u_i] \right] + \text{tr}(\mathbf{x}' \mathbf{x} \mathbf{V}_\beta) \right) \\ q(\mu) \leftarrow C \exp \left\{ \left( \overline{\lambda^{-1}} \sum_{i=1}^N \bar{u}_i + \underline{A}_\mu \underline{B}_\mu^{-2} \right) \mu - \left( \frac{N + \underline{B}_\mu^{-2}}{2} \right) \mu^2 - N \ln \Phi(\mu) \right\} \\ q(\lambda^{-1}) \leftarrow \frac{\lambda^{-(N+A_\lambda-1)} \exp\{-s^2 \lambda^{-2}/2 - \gamma \lambda^{-1}\}}{s^{-\bar{A}_\lambda} \exp(\gamma^2/4s^2) \Gamma(A_\lambda) D_{-A_\lambda}(\gamma/s)} \quad A_\lambda = N + \underline{A}_\lambda \quad s^2 = \sum_{i=1}^N \bar{u}_i^2 \quad \gamma = - \left( \bar{\mu} \sum_{i=1}^N \bar{u}_i + \underline{B}_\lambda \right) \\ q(u_i) \leftarrow TN(\mu_i, \nu^2) \quad \mu_i = \frac{\pm \overline{\boldsymbol{\sigma}^{-2}} \sum_t (\mathbf{x}_{it} \bar{\boldsymbol{\beta}} - y_{it}) + \overline{\lambda^{-1}} \bar{\mu}}{T \overline{\boldsymbol{\sigma}^{-2}} + \overline{\lambda^{-2}}} \quad \nu^2 = \frac{1}{T \overline{\boldsymbol{\sigma}^{-2}} + \overline{\lambda^{-2}}} \end{array} \right.$$

*Coordinate ascent algorithm*

Iterate the following quantities until the change in  $\ln \underline{ML}$  is negligible.

$$\begin{aligned} \bar{\boldsymbol{\beta}} \quad \mathbf{V}_\beta \quad \overline{\boldsymbol{\sigma}^{-2}} = \frac{A_\sigma}{B_\sigma} \quad \frac{1}{C} = \int_{-\infty}^{\infty} q^*(\mu) d\mu \quad \bar{\mu} = \int_0^{\infty} \mu q(\mu) d\mu \quad \overline{\lambda^{-1}} = \frac{A_\lambda D_{-(A_\lambda+1)}(\gamma/s)}{s D_{-A_\lambda}(\gamma/s)} \\ \overline{\lambda^{-2}} = \frac{A_\lambda (A_\lambda + 1) D_{-(A_\lambda+2)}(\gamma/s)}{s^2 D_{-A_\lambda}(\gamma/s)} \quad \bar{u}_i = \mu_i + \nu m(\alpha_i) \quad V[u_i] = \nu^2 \{1 - m(\alpha_i)[m(\alpha_i) + \alpha_i]\} \end{aligned}$$

where  $q^*(\mu) = q(\mu)/C$ ,  $\alpha_i = \mu_i/\nu$  and  $m(\cdot) = \phi(\cdot)/\Phi(\cdot)$ .

*Lower bound for marginal likelihood*

$$\begin{aligned} \ln \underline{ML} = \frac{(-NT - 1) \ln 2\pi + k + N}{2} + \ln \left( \frac{\Gamma(A_\sigma) \underline{B}_\sigma^{A_\sigma} \underline{B}_\lambda^{A_\lambda}}{\Gamma(\underline{A}_\sigma) \Gamma(\underline{A}_\lambda) \underline{B}_\sigma^{A_\sigma}} \right) - \frac{\underline{A}_\mu \underline{B}_\mu^{-2}}{2} + \frac{\ln \underline{B}_\mu^{-2}}{2} + \frac{1}{2} \ln \left( \frac{|\mathbf{V}_\beta|}{|\underline{\mathbf{V}}_\beta|} \right) \\ - \frac{(\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}})' \underline{\mathbf{V}}_\beta^{-1} (\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}) + \text{tr}(\underline{\mathbf{V}}_\beta^{-1} \mathbf{V}_\beta)}{2} + N \ln \nu + \sum_{i=1}^N \left\{ \ln \Phi(\alpha_i) - \frac{\alpha_i \phi(\alpha_i)}{2\Phi(\alpha_i)} \right\} + (N + \underline{A}_\lambda - 1) \overline{\ln \lambda^{-1}} \\ - \frac{\overline{\lambda^{-2}}}{2} \sum_{i=1}^N \bar{u}_i^2 - \frac{(N + \underline{B}_\mu^{-2}) \overline{\mu}^2}{2} + \overline{\lambda^{-1}} \left( \sum_{i=1}^N \bar{u}_i \bar{\mu} + \underline{B}_\lambda \right) + \underline{A}_\mu \underline{B}_\mu^{-2} \bar{\mu} - N \overline{\ln \Phi(\mu)} - \overline{\ln q(\mu)} - \overline{\ln q(\lambda^{-1})} \end{aligned}$$

Lognormal Inefficiency

*Model*  $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} \mp u_i + v_{it}, \quad u_i \sim LN(\mu, \lambda^2)$

*Priors*  $\boldsymbol{\beta} \sim N(\underline{\boldsymbol{\beta}}, \underline{\mathbf{V}}_{\boldsymbol{\beta}}) \quad \sigma^{-2} \sim G(\underline{A}_{\sigma}, \underline{B}_{\sigma}) \quad \mu \sim N(\underline{\mu}, \underline{V}_{\mu}) \quad \lambda^{-2} \sim G(\underline{A}_{\lambda}, \underline{B}_{\lambda})$

*Optimal densities*

$$\left\{ \begin{array}{l} q(\boldsymbol{\beta}) \leftarrow N(\bar{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}}) \quad \bar{\boldsymbol{\beta}} = \mathbf{V}_{\boldsymbol{\beta}} \left( \underline{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} \bar{\boldsymbol{\beta}} + \overline{\sigma^{-2}} \mathbf{x}' (\mathbf{y} \pm \bar{\mathbf{u}} \otimes \mathbf{i}_T) \right) \quad \mathbf{V}_{\boldsymbol{\beta}} = \left( \overline{\sigma^{-2}} \mathbf{x}' \mathbf{x} + \underline{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} \right)^{-1} \\ q(\sigma^{-2}) \leftarrow G(A_{\sigma}, B_{\sigma}) \quad A_{\sigma} = \underline{A}_{\sigma} + \frac{NT}{2} \quad B_{\sigma} = \underline{B}_{\sigma} + \frac{1}{2} \left( \sum_{i=1}^N \sum_{t=1}^T \left[ (y_{it} - \mathbf{x}_{it} \bar{\boldsymbol{\beta}} \pm \bar{u}_i)^2 + V[u_i] \right] + \text{tr}(\mathbf{x}' \mathbf{x} \mathbf{V}_{\boldsymbol{\beta}}) \right) \\ q(\mu) \leftarrow N(\bar{\mu}, V_{\mu}) \quad \bar{\mu} = \frac{\overline{\lambda^{-2}} \sum_{i=1}^N \overline{\ln u_i} + \underline{\mu} \underline{V}_{\mu}^{-1}}{N \overline{\lambda^{-2}} + \underline{V}_{\mu}^{-1}} \quad V_{\mu} = \frac{1}{N \overline{\lambda^{-2}} + \underline{V}_{\mu}^{-1}} \\ q(\lambda^{-2}) \leftarrow G(A_{\lambda}, B_{\lambda}) \quad A_{\lambda} = \frac{N}{2} + \underline{A}_{\lambda} \quad B_{\lambda} = \frac{1}{2} \left( \sum_{i=1}^N \overline{(\ln u_i)^2} + N \bar{\mu}^2 - 2 \bar{\mu} \sum_{i=1}^N \overline{\ln u_i} \right) + \underline{B}_{\lambda} \\ q(u_i) \leftarrow C_{li} \exp \left\{ \left( \overline{\lambda^{-2}} \bar{\mu} - 1 \right) \ln u_i - 0.5 \overline{\lambda^{-2}} (\ln u_i)^2 - 0.5 \overline{\sigma^{-2}} \left[ T u_i^2 \mp 2 u_i \sum_{t=1}^T (\mathbf{x}_{it} \bar{\boldsymbol{\beta}} - y_{it}) \right] \right\} \end{array} \right.$$

*Coordinate ascent algorithm*

Iterate the following quantities until the change in  $\ln \underline{ML}$  is negligible.

$$\begin{aligned} \bar{\boldsymbol{\beta}} \quad \mathbf{V}_{\boldsymbol{\beta}} \quad \overline{\sigma^{-2}} = \frac{A_{\sigma}}{B_{\sigma}} \quad \overline{\lambda^{-2}} = \frac{A_{\lambda}}{B_{\lambda}} \quad \bar{\mu} \quad \overline{\mu^2} = \bar{\mu}^2 + V_{\mu} \quad \frac{1}{C_{li}} = \int_0^{\infty} q^*(u_i) du_i \quad \bar{u}_i = \int_0^{\infty} u_i^2 q(u_i) du_i \\ V[u_i] = \int_0^{\infty} u_i^2 q(u_i) du_i - \bar{u}_i^2 \quad \overline{\ln u_i} = \int_0^{\infty} \ln u_i q(u_i) du_i \quad \overline{(\ln u_i)^2} = \int_0^{\infty} (\ln u_i)^2 q(u_i) du_i \end{aligned}$$

*Lower bound for marginal likelihood*

$$\begin{aligned} \ln \underline{ML} = \frac{(-NT - N) \ln 2\pi + k + 1}{2} + \ln \left( \frac{\Gamma(A_{\sigma}) \Gamma(A_{\lambda}) \underline{B}_{\sigma}^{A_{\sigma}} \underline{B}_{\lambda}^{A_{\lambda}}}{\Gamma(\underline{A}_{\sigma}) \Gamma(\underline{A}_{\lambda}) \underline{B}_{\sigma}^{A_{\lambda}} \underline{B}_{\lambda}^{A_{\sigma}}} \right) - \frac{(\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}})' \underline{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} (\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}) + \text{tr}(\underline{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} \mathbf{V}_{\boldsymbol{\beta}})}{2} \\ + \frac{1}{2} \ln \left( \frac{|\mathbf{V}_{\boldsymbol{\beta}}|}{|\underline{\mathbf{V}}_{\boldsymbol{\beta}}|} \right) + \frac{1}{2} \ln \left( \frac{V_{\mu}}{\underline{V}_{\mu}} \right) - \frac{\left[ (\bar{\mu} - \underline{\mu})^2 + V_{\mu} \right] \underline{V}_{\mu}^{-1}}{2} - \sum_{i=1}^N \overline{\ln u_i} - \sum_{i=1}^N \overline{\ln q(u_i)} \end{aligned}$$

## Appendix B: Model with zero-one environmental variables

Here we return to model (3.1),  $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} \mp u_i + v_{it}$ , and, following Koop et al. (1997), we assume  $u_i$  has the exponential distribution  $p(u_i | \lambda_i) = \lambda_i \exp(-\lambda_i u_i)$  with  $\lambda_i = \prod_{j=1}^m \phi_j^{z_{ij}}$ ; the  $z_{ij}$  are 0-1 dummy variables with the exception of the first, where  $z_{i1} = 1$ . The dummy variables describe the presence or absence of some environmental factors. For example, some firms might be privately owned and some publicly owned and we might be interested in whether one category of firms is more efficient. A value  $\phi_j > 1$  ( $\phi_j < 1$ ) implies that firms with  $z_{ij} = 1$  come from a distribution with lower (higher) mean inefficiency. If  $\phi_j = 1$  for  $j = 2, \dots, m$ , then the model reduces to the standard stochastic frontier model with exponential inefficiencies. Using the priors for  $(\boldsymbol{\beta}, \sigma^2)$  given in Section 3, and assuming  $\phi_j \sim G(\underline{A}_j, \underline{B}_j)$  as suggested by Koop et al. (1997), we can derive the following optimal densities:

$$\left\{ \begin{array}{l} q(\boldsymbol{\beta}) \leftarrow N(\bar{\boldsymbol{\beta}}, \mathbf{V}_\beta) \quad \bar{\boldsymbol{\beta}} = \mathbf{V}_\beta \left( \mathbf{V}_\beta^{-1} \bar{\boldsymbol{\beta}} + \sigma^{-2} \mathbf{x}'(\mathbf{y} \pm \bar{\mathbf{u}} \otimes \mathbf{i}_T) \right) \quad \mathbf{V}_\beta = \left( \sigma^{-2} \mathbf{x}'\mathbf{x} + \mathbf{V}_\beta^{-1} \right)^{-1} \\ q(\sigma^{-2}) \leftarrow G(A_\sigma, B_\sigma) \quad A_\sigma = \underline{A}_\sigma + \frac{NT}{2} \quad B_\sigma = \underline{B}_\sigma + \frac{1}{2} \left( \sum_{i=1}^N \sum_{t=1}^T \left[ (y_{it} - \mathbf{x}_{it} \bar{\boldsymbol{\beta}} \pm \bar{u}_i)^2 + V[u_i] \right] + \text{tr}(\mathbf{x}'\mathbf{x} \mathbf{V}_\beta) \right) \\ q(\phi_j) \leftarrow G(A_j, B_j) \quad A_j = \sum_{i=1}^N z_{ij} + \underline{A}_j \quad B_j = \underline{B}_j + \sum_{i=1}^N z_{ij} \bar{u}_i \sum_{s \neq j}^m \bar{\phi}_s^{z_{is}} \\ q(u_i) \leftarrow TN(\mu_i, \nu^2) \quad \mu_i = \frac{1}{T} \left( \pm \sum_{t=1}^T (\mathbf{x}_{it} \bar{\boldsymbol{\beta}} - y_{it}) - \prod_{j=1}^m \bar{\phi}_j^{z_{ij}} / \sigma^{-2} \right) \quad \nu^2 = \frac{1}{T \sigma^{-2}} \end{array} \right.$$

Interestingly, all these densities are standard forms; we therefore can easily set-up a coordinate ascent algorithm by iterating over

$$\left\{ \bar{\boldsymbol{\beta}}, \mathbf{V}_\beta, \sigma^{-2} = \frac{A_\sigma}{B_\sigma}, \bar{\phi}_j = \frac{A_j}{B_j}, \bar{u}_i = \mu_i + \nu m(\alpha_i), V[u_i] = \nu^2 (1 - m(\alpha_i)) [m(\alpha_i) + \alpha_i] \right\}$$

where  $\alpha_i = \bar{u}_i / \nu$ , and  $m(\cdot) = \phi(\cdot) / \Phi(\cdot)$ . The marginal likelihood lower bound used to assess convergence is

$$\ln \underline{ML} = \frac{(-NT + N) \ln 2\pi + N + k}{2} + \ln \left( \frac{\Gamma(A_\sigma) \underline{B}_\sigma^{A_\sigma} \prod_{j=1}^m \Gamma(A_j) \underline{B}_j^{A_j}}{\Gamma(\underline{A}_\sigma) B_\sigma^{A_\sigma} \prod_{j=1}^m \Gamma(\underline{A}_j) B_j^{A_j}} \right) + \frac{1}{2} \ln \left( \frac{|\mathbf{V}_\beta|}{|\underline{\mathbf{V}}_\beta|} \right) - \frac{(\bar{\beta} - \underline{\beta})' \underline{\mathbf{V}}_\beta^{-1} (\bar{\beta} - \underline{\beta}) + \text{tr}(\underline{\mathbf{V}}_\beta^{-1} \mathbf{V}_\beta)}{2} + N \ln \nu + \sum_{i=1}^N \left\{ \ln \Phi(\alpha_i) - \frac{\alpha_i \phi(\alpha_i)}{2\Phi(\alpha_i)} \right\}$$

## References

- Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of econometrics*, 6(1), 21-37.
- Alvarez, A., Amsler, C., Orea, L., & Schmidt, P. (2006). Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. *Journal of Productivity Analysis*, 25(3), 201-212.
- Attias, H. (1999, July). Inferring parameters and structure of latent variable models by variational Bayes. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence* (pp. 21-30). Morgan Kaufmann Publishers Inc.
- Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. In *International applications of productivity and efficiency analysis* (pp. 149-165). Springer Netherlands.
- Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical economics*, 20(2), 325-332.
- Beckers, D. E., & Hammond, C. J. (1987). A tractable likelihood function for the normal-gamma stochastic frontier model. *Economics Letters*, 24(1), 33-38.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Braun, M., & McAuliffe, J. (2010). Variational inference for large-scale models of discrete choice. *Journal of the American Statistical Association*, 105(489), 324-335.
- Chen, M. H., Shao, Q. M., & Ibrahim, J. G. (2012). *Monte Carlo methods in Bayesian computation*. Springer Science & Business Media.
- Christensen, L. R., & Greene, W. H. (1976). Economies of scale in US electric power generation. *Journal of political Economy*, 84(4, Part 1), 655-676.
- Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer Science & Business Media.

- Cornwell, C., Schmidt, P., & Sickles, R. C. (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of econometrics*, 46(1-2), 185-200.
- Ehlers, R. S. (2011). Comparison of Bayesian models for production efficiency. *Journal of Applied Statistics*, 38(11), 2433-2443.
- Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media.
- Gelfand, A. E., & Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 501-514.
- Greene, W. H. (1990). A gamma-distributed stochastic frontier model. *Journal of econometrics*, 46(1-2), 141-163.
- Griffin, J. E., & Steel, M. F. (2004). Semiparametric Bayesian inference for stochastic frontier models. *Journal of econometrics*, 123(1), 121-152.
- Griffin, J. E., & Steel, M. F. (2007). Bayesian stochastic frontier analysis using WinBUGS. *Journal of Productivity Analysis*, 27(3), 163-176.
- Griffin, J. E., & Steel, M. F. (2008). Flexible mixture modelling of stochastic frontiers. *Journal of Productivity Analysis*, 29(1), 33-50.
- Ji, C., Shen, H., & West, M. (2010). Bounded approximations for marginal likelihoods. *Technical report*.
- Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1998). An introduction to variational methods for graphical models. *NATO ASI SERIES D BEHAVIOURAL AND SOCIAL SCIENCES*, 89, 105-162.
- Koop, G., Osiewalski, J., & Steel, M. F. (1997). Bayesian efficiency analysis through individual effects: Hospital cost frontiers. *Journal of econometrics*, 76(1), 77-105.
- Koop, G., & Steel, M. F. (2001). Bayesian analysis of stochastic frontier models. *A companion to theoretical econometrics*, Blackwell, Oxford, 520-573.
- Kucukelbir, A., Ranganath, R., Gelman, A., & Blei, D. (2015). Automatic variational inference in Stan. In *Advances in neural information processing systems* (pp. 568-576).
- Kumbhakar, S. C. (1990). Production frontiers, panel data, and time-varying technical inefficiency. *Journal of econometrics*, 46(1-2), 201-211.

- Kumbhakar, S. L., & Lovell, K. (2000). *Stochastic frontier analysis*, Cambridge: Cambridge University Press.
- Lee, Y. H., & Schmidt, P. (1993). A production frontier model with flexible temporal variation in technical efficiency. *The measurement of productive efficiency: Techniques and applications*, 237-255.
- McGrory, C. A., & Titterton, D. M. (2007). Variational approximations in Bayesian model selection for finite mixture distributions. *Computational Statistics & Data Analysis*, 51(11), 5352-5367.
- Minka, T., Winn, J., Guiver, G., & Kannan, A. (2010). *Infer.Net 2.3*, Cambridge, U.K.: Microsoft Research Cambridge.
- Nott, D. J., Tan, S. L., Villani, M., & Kohn, R. (2012). Regression density estimation with variational methods and stochastic approximation. *Journal of Computational and Graphical Statistics*, 21(3), 797-820.
- Nychka, D., Haaland, P. D., O'Connell, M. A., & Ellner, S. (1998). Appendix A: FUNFITS, data analysis and statistical tools for estimating functions. *Case studies in environmental statistics*, 132, 159.
- Ormerod, J. T., & Wand, M. P. (2010). Explaining variational approximations. *The American Statistician*, 64(2), 140-153.
- Parisi, G. (1988). *Statistical field theory*. Addison-Wesley.
- Parmeter, C. F., & Kumbhakar, S. C. (2014). Efficiency analysis: a primer on recent advances. *Foundations and Trends® in Econometrics*, 7(3-4), 191-385.
- Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression* (No. 12). Cambridge university press.
- Sickles, R. C. (2005). Panel estimators and the identification of firm-specific efficiency levels in parametric, semiparametric and nonparametric settings. *Journal of econometrics*, 126(2), 305-334.
- Stevenson, R. E. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of econometrics*, 13(1), 57-66.
- Tsionas, E. G. (2007). Efficiency measurement with the Weibull stochastic frontier. *Oxford Bulletin of Economics and Statistics*, 69(5), 693-706.

- Tsionas, E. G. (2012). Maximum likelihood estimation of stochastic frontier models by the Fourier transform. *Journal of Econometrics*, 170(1), 234-248.
- Van den Broeck, J., Koop, G., Osiewalski, J., & Steel, M. F. (1994). Stochastic frontier models: A Bayesian perspective. *Journal of Econometrics*, 61(2), 273-303.
- Wainwright, M. J., & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2), 1-305.
- Wand, M. P., Ormerod, J. T., Padoan, S. A., & Fuhrwirth, R. (2011). Mean field variational Bayes for elaborate distributions. *Bayesian Analysis*, 6(4), 847-900.
- Wang, B., & Titterton, D. M. (2005, January). Inadequacy of interval estimates corresponding to variational Bayesian approximations. In *AISTATS*.

Figure 1- Implied Marginal Prior for  $u_i$

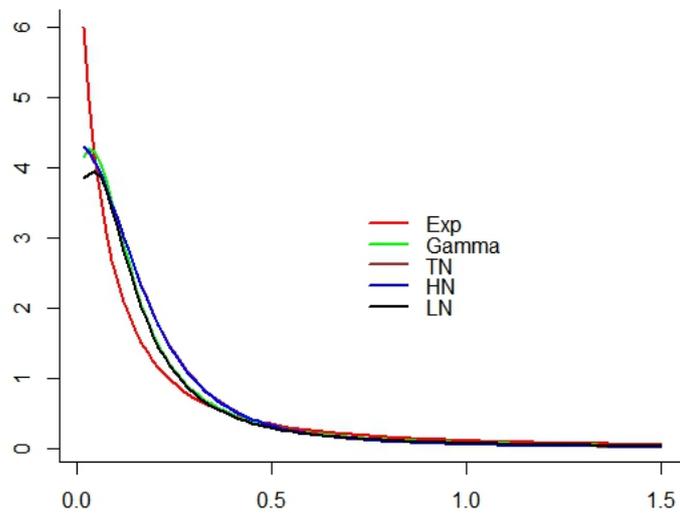


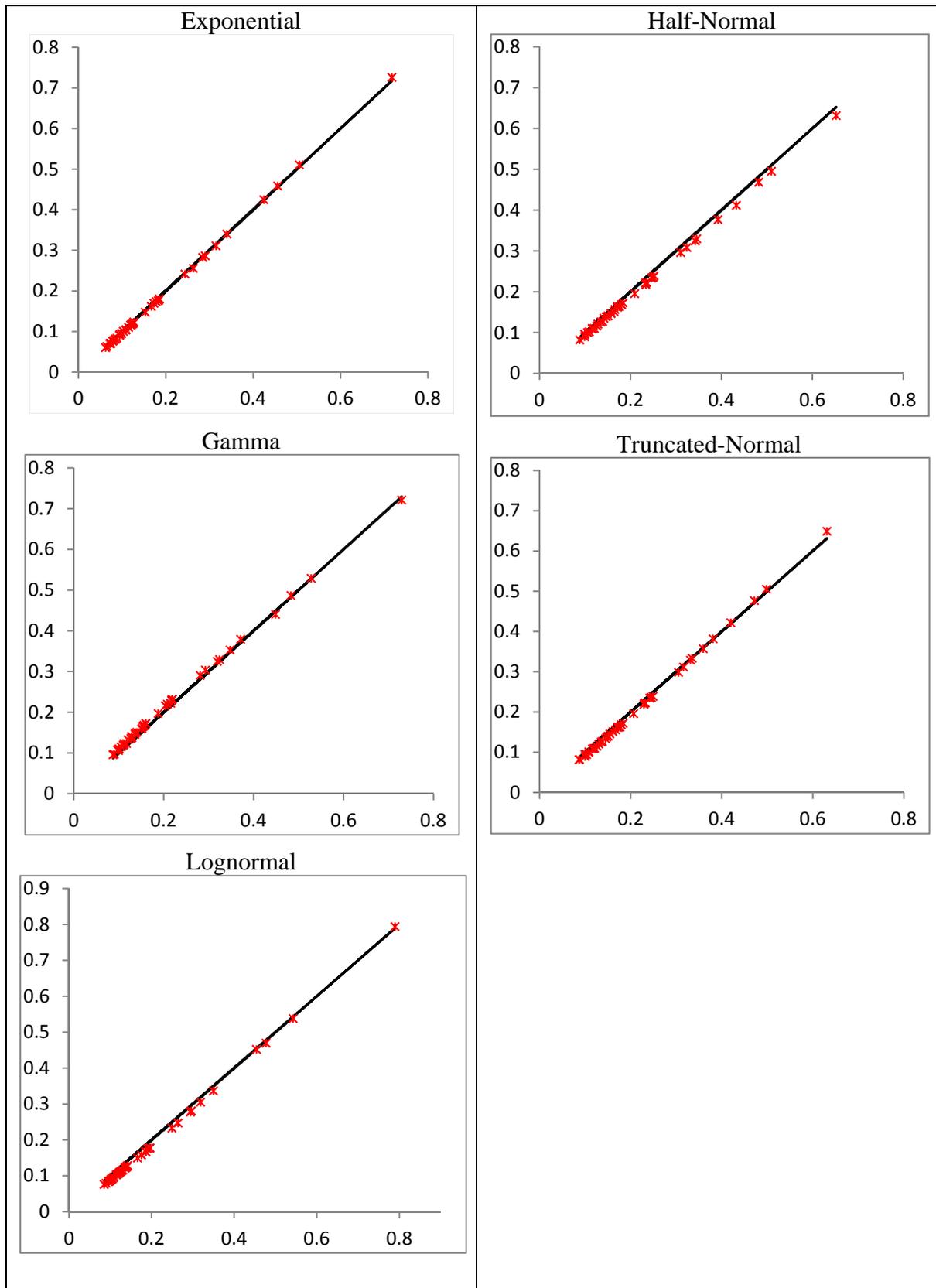
Figure 2 Estimated Efficiency Effects ( $\hat{u}_i$ ) for Rice Data - MFVB vs MCMC

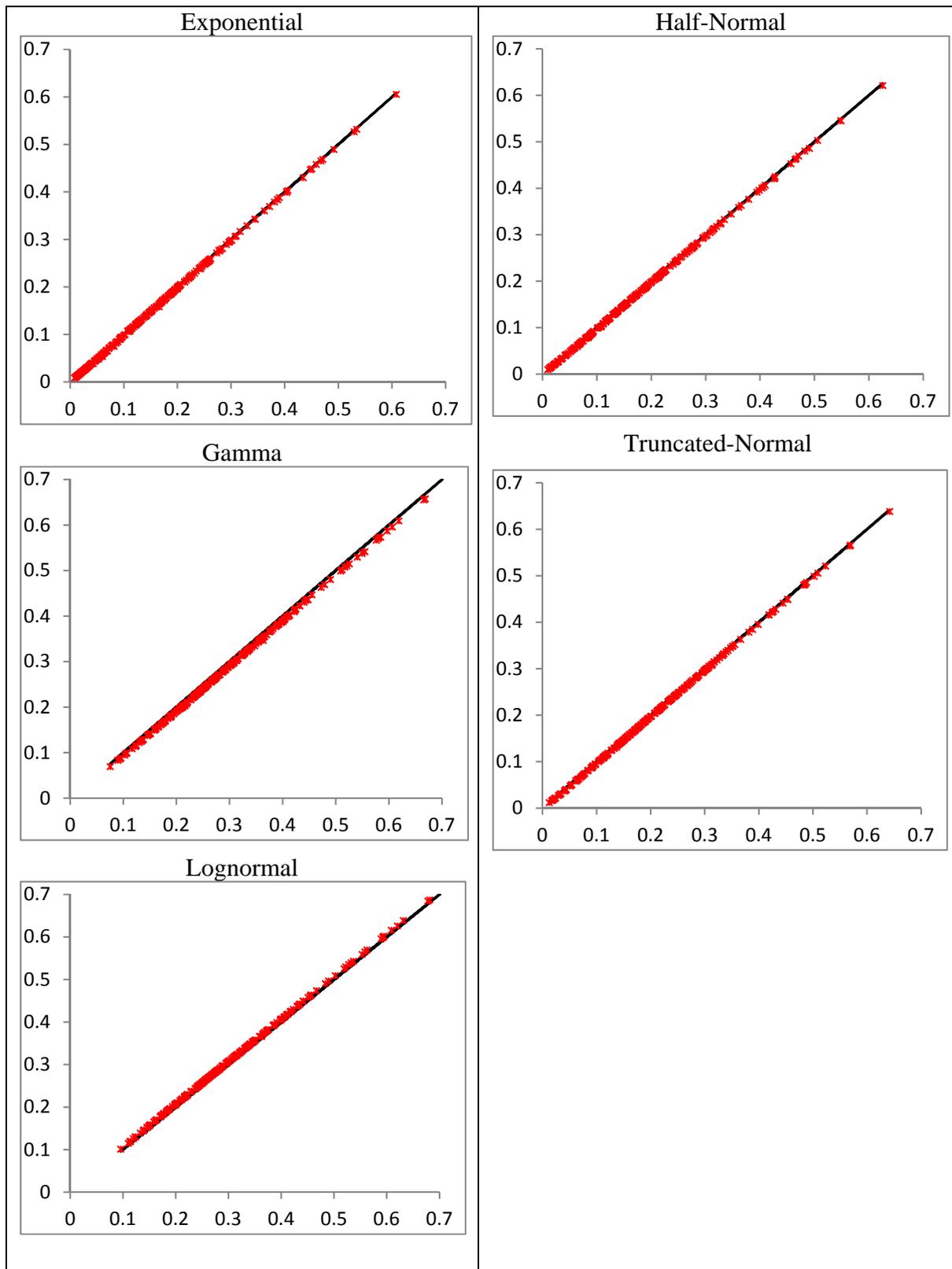
Figure 3 Estimated Efficiency Effects ( $\hat{u}_i$ ) for Dairy Data - MFVB vs MCMC

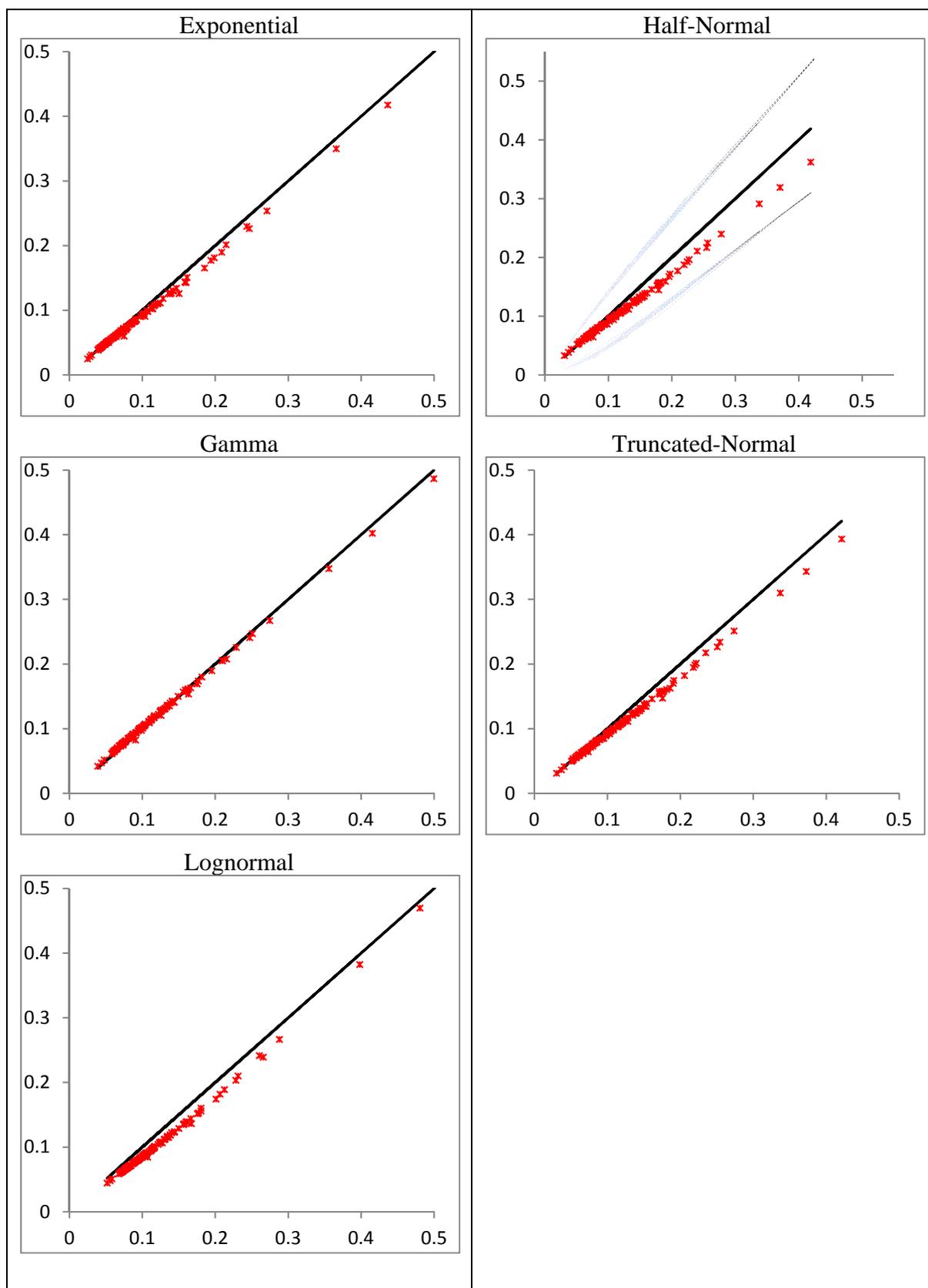
Figure 4 Estimated Efficiency Effects ( $\hat{u}_i$ ) for Electricity Data - MFVB vs MCMC

Figure 5 MFVB vs MCMC Marginal Posterior Distributions of Selected  $\hat{u}_i$  s  
Based on Exponential Distribution for Rice Data example

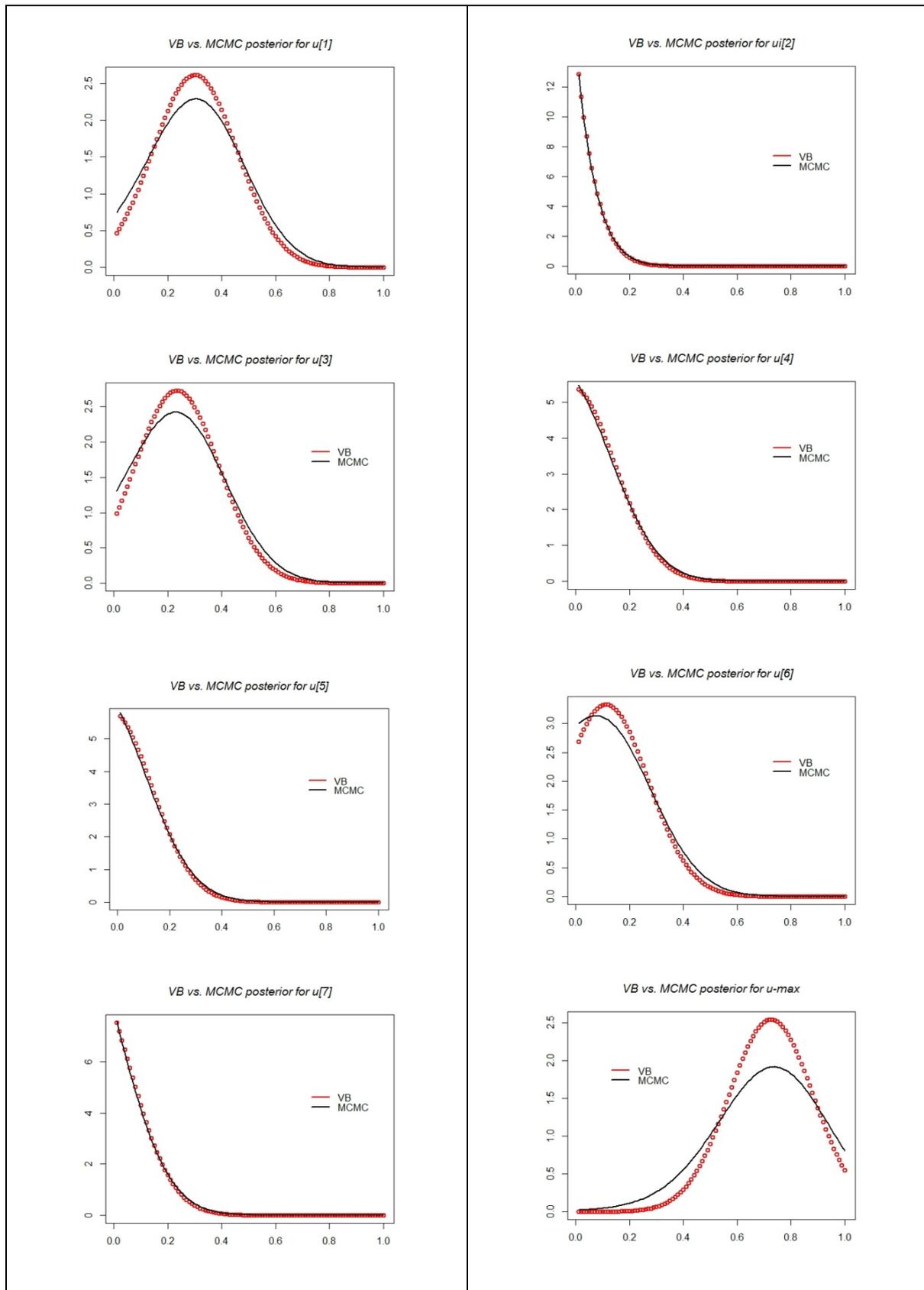


Figure 6 Estimated Efficiency Effects based on Lognormal Battese-Coelli model for Rice Data: MFVB vs MCMC

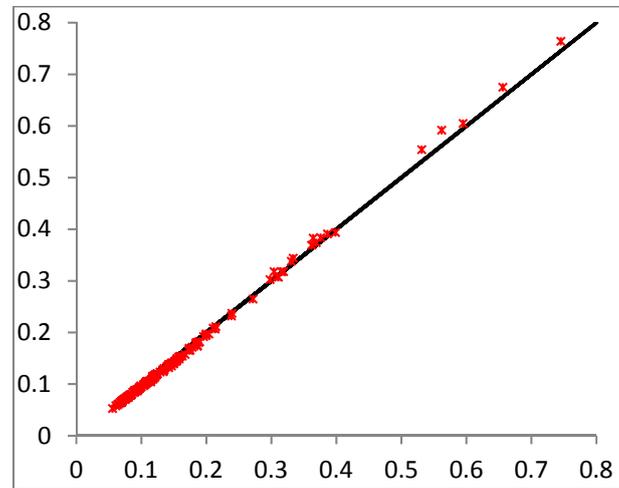
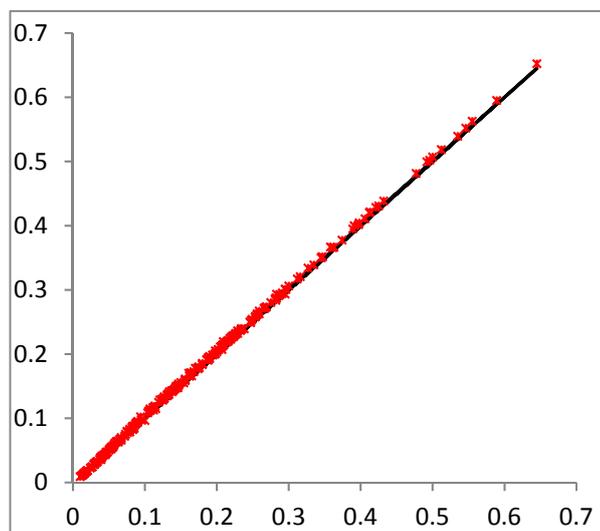


Figure 7 Estimated Efficiency Effects based on Time-varying Model  
for Dairy Data: MFVB vs MCMC



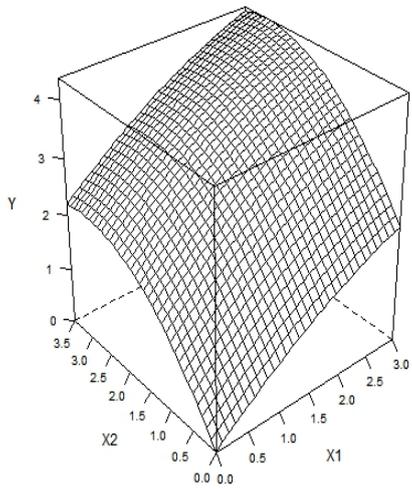
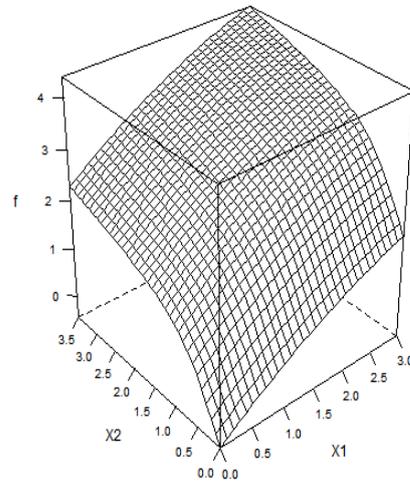
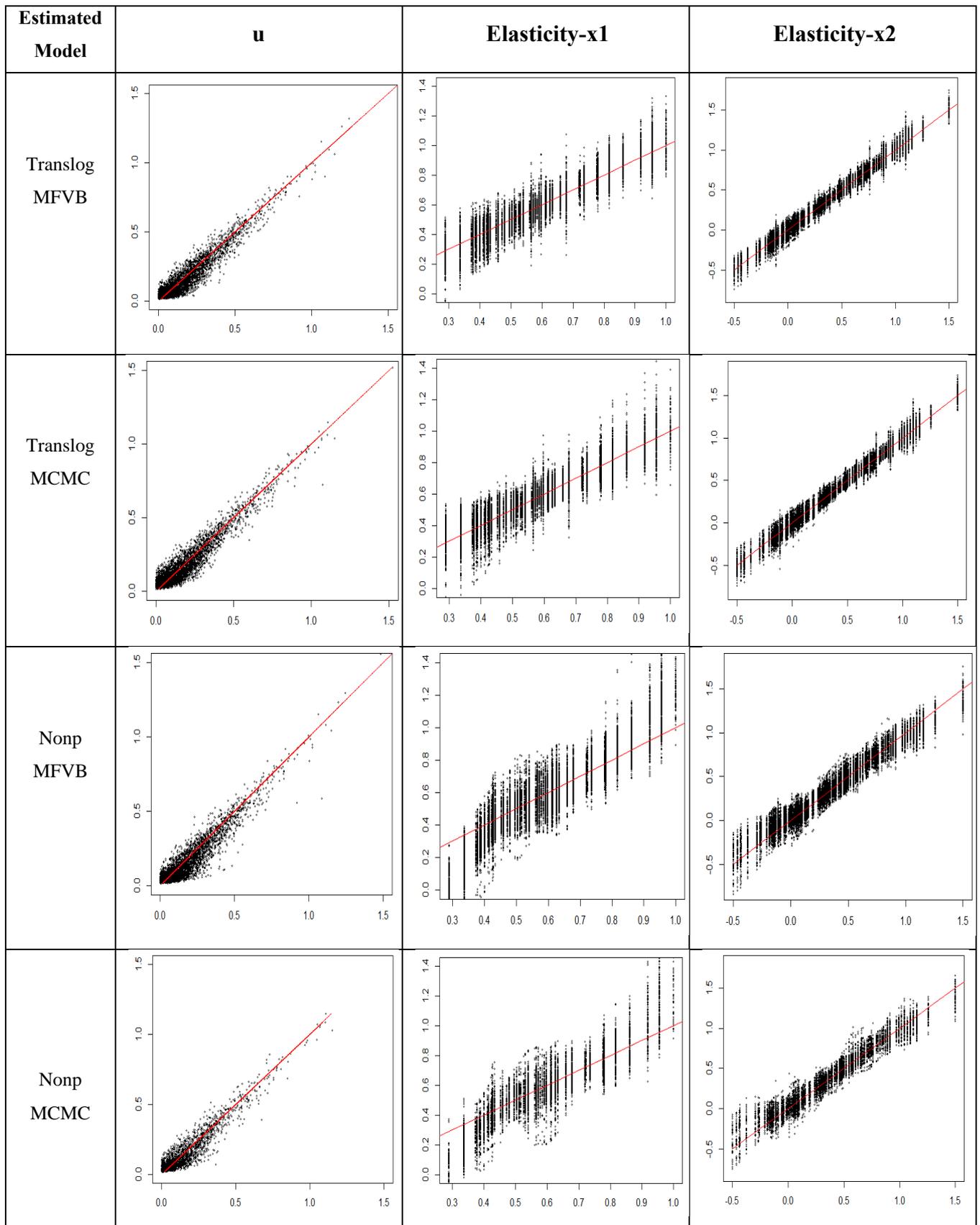


Figure 8 (a) Translog function



8 (b) Arctan function

Figure 9 Posterior Means versus True Values from Translog model<sup>a</sup>

<sup>a</sup>True values are on the horizontal axes; posterior means are on the vertical axes.

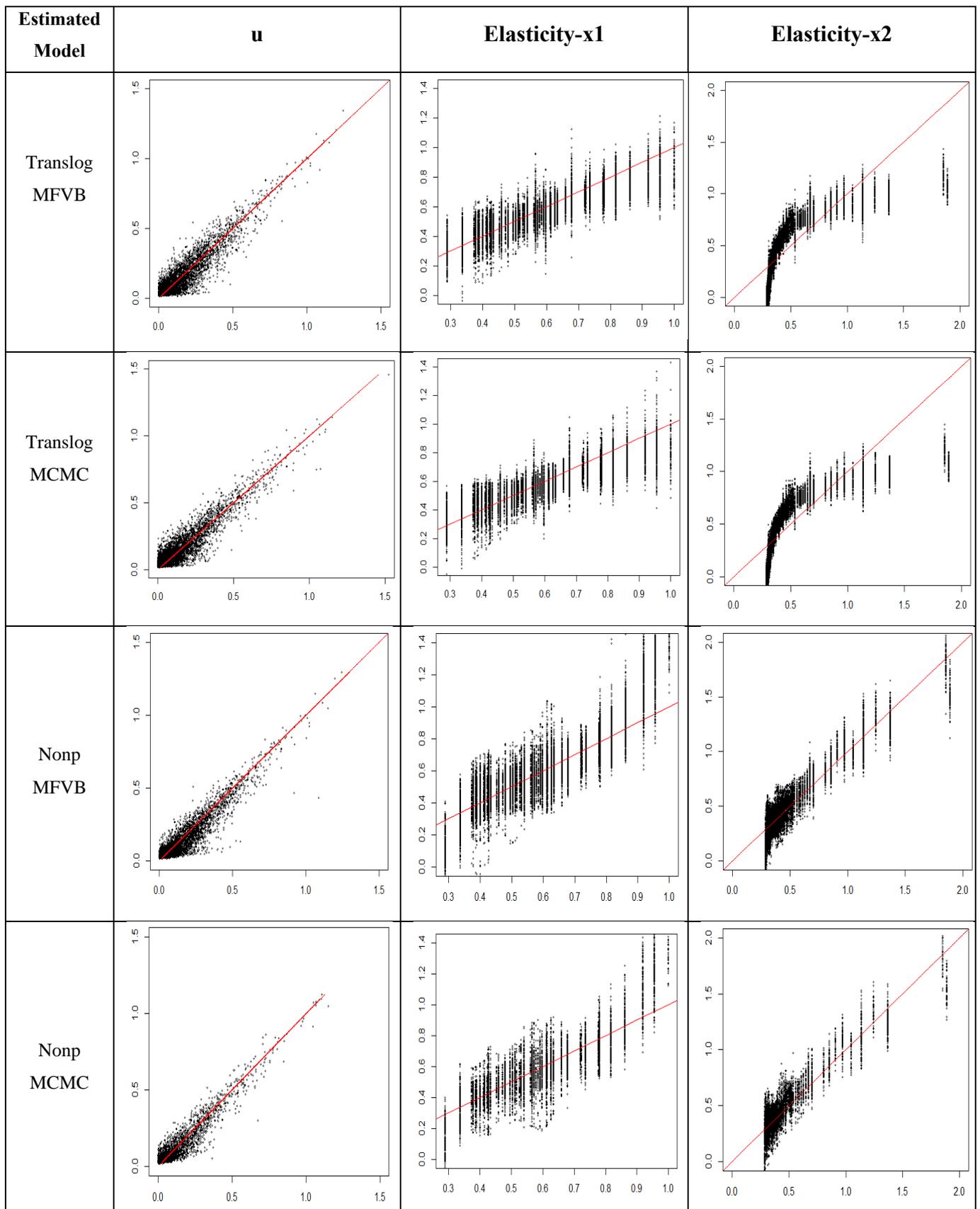
Figure 10 Posterior Means versus True Values from Arctan Model<sup>a</sup><sup>a</sup>True values are on the horizontal axes; posterior means are on the vertical axes.

Table 1 Posterior Means for Parameters of Various Models – Rice Data

Par	EXP		HN		Gamma		TN		LN	
	MCMC	VB								
<b>C</b>	-0.945	-0.947	-0.975	-0.981	-0.922	-0.934	-0.980	-0.972	-0.911	-0.910
<i>Land</i>	0.383	0.383	0.384	0.387	0.386	0.384	0.386	0.387	0.379	0.380
<i>Labor</i>	0.293	0.293	0.306	0.306	0.296	0.298	0.307	0.304	0.290	0.287
<i>Fert</i>	0.196	0.196	0.205	0.204	0.196	0.200	0.205	0.203	0.194	0.192
<i>Other</i>	0.058	0.057	0.051	0.049	0.056	0.054	0.049	0.051	0.059	0.060
$\sigma^{-2}$	10.183	10.183	10.122	10.068	10.177	10.125	10.063	10.114	10.147	10.155
$\lambda$	6.061	5.797	15.288	14.825	7.386	8.073	17.689	13.096	1.393	1.138
$\mu$ or $\theta$					1.510	1.721	0.015	-0.119	-2.058	-2.104

Table 2- Posterior Means for Parameters of Various Models – Dairy Data

Par	EXP		HN		Gamma		TN		LN	
	MCMC	VB								
<b>C</b>	11.6673	11.6679	11.6854	11.6836	11.7981	11.7842	11.7065	11.7037	11.8094	11.8164
<b>X1</b>	0.6573	0.6573	0.6592	0.6587	0.6634	0.6627	0.6611	0.6606	0.6613	0.6615
<b>X2</b>	0.0425	0.0434	0.0421	0.0416	0.0421	0.0421	0.0414	0.0408	0.0432	0.0433
<b>X3</b>	0.0472	0.0474	0.0494	0.0495	0.0461	0.0466	0.0507	0.0505	0.0451	0.0456
<b>X4</b>	0.3613	0.3603	0.3563	0.3573	0.3514	0.3527	0.3536	0.3543	0.3531	0.3532
<b>X11</b>	0.3371	0.3377	0.3187	0.3191	0.2710	0.2724	0.3000	0.3003	0.2773	0.2761
<b>X22</b>	-0.1318	-0.1277	-0.1082	-0.1101	-0.0808	-0.0854	-0.0876	-0.0873	-0.0918	-0.0935
<b>X33</b>	-0.1240	-0.1196	-0.1204	-0.1235	-0.1453	-0.1444	-0.1230	-0.1269	-0.1468	-0.1462
<b>X44</b>	0.1247	0.1258	0.1182	0.1179	0.1116	0.1110	0.1136	0.1128	0.1135	0.1133
<b>X12</b>	-0.0685	-0.0663	-0.0632	-0.0645	-0.0408	-0.0420	-0.0551	-0.0572	-0.0431	-0.0424
<b>X13</b>	0.0993	0.0984	0.0939	0.0925	0.0929	0.0927	0.0878	0.0876	0.0956	0.0951
<b>X14</b>	-0.1598	-0.1603	-0.1475	-0.1477	-0.1290	-0.1292	-0.1388	-0.1383	-0.1321	-0.1321
<b>X23</b>	0.0146	0.0149	0.0083	0.0091	0.0038	0.0045	0.0036	0.0029	0.0078	0.0072
<b>X24</b>	0.0264	0.0252	0.0224	0.0235	0.0103	0.0113	0.0183	0.0189	0.0120	0.0118
<b>X34</b>	-0.0213	-0.0217	-0.0172	-0.0173	-0.0148	-0.0148	-0.0147	-0.0140	-0.0157	-0.0153
<b>t</b>	0.0304	0.0307	0.0312	0.0311	0.0320	0.0318	0.0317	0.0316	0.0317	0.0317
$t^2$	-0.0028	-0.0028	-0.0029	-0.0028	-0.0029	-0.0029	-0.0029	-0.0029	-0.0029	-0.0029
$\sigma^{-2}$	158.866	158.103	162.387	161.671	165.695	165.019	164.486	164.063	164.845	164.557
$\lambda$	6.2581	6.3281	21.2656	21.7130	18.7577	18.2824	38.6811	38.9029	5.7477	6.0920
$\mu$ or $\theta$					5.6250	5.1282	0.9205	0.9158	-1.2721	-1.2401

Table 3- Posterior Means for Parameters of Various Models – Electricity Data

Par	EXP		HN		Gamma		TN		LN	
	MCMC	VB								
$C$	-7.543	-7.550	-7.409	-7.411	-7.523	-7.525	-7.426	-7.429	-7.633	-7.636
$Q$	0.462	0.459	0.442	0.439	0.457	0.453	0.442	0.442	0.467	0.468
$pl/pf$	0.235	0.240	0.220	0.225	0.231	0.235	0.223	0.227	0.243	0.247
$pk/pf$	0.040	0.042	0.061	0.065	0.049	0.051	0.060	0.061	0.035	0.034
$Q2$	0.028	0.028	0.029	0.029	0.028	0.028	0.028	0.028	0.027	0.027
$\sigma^{-2}$	93.055	83.810	95.364	76.142	89.288	82.391	92.320	79.345	91.644	83.202
$\lambda$	10.486	10.770	48.686	51.020	14.932	15.390	46.732	39.771	2.118	1.666
$\mu$ or $\theta$					1.707	1.840	-0.167	-0.288	-2.371	-2.504

Table 4 Posterior Standard Deviations for Parameters – Rice Data

	Exp		HN		Gamma		TN		LN	
	MCMC	VB								
$\beta_0$	0.403	0.345	0.405	0.347	0.406	0.346	0.410	0.346	0.403	0.345
$\beta_1$	0.106	0.091	0.105	0.091	0.105	0.091	0.107	0.091	0.105	0.091
$\beta_2$	0.102	0.092	0.103	0.093	0.103	0.093	0.103	0.093	0.103	0.092
$\beta_3$	0.063	0.058	0.063	0.058	0.064	0.058	0.066	0.058	0.063	0.058
$\beta_4$	0.035	0.028	0.035	0.028	0.036	0.028	0.035	0.028	0.035	0.028
$\sigma^{-2}$	1.279	1.098	1.275	1.086	1.277	1.091	1.289	1.091	1.269	1.095
u[1]	0.163	0.146	0.157	0.132	0.168	0.139	0.156	0.134	0.170	0.151
u[2]	0.057	0.055	0.072	0.068	0.071	0.064	0.073	0.067	0.066	0.051
u[3]	0.153	0.138	0.151	0.130	0.159	0.132	0.150	0.132	0.158	0.137
u[4]	0.097	0.092	0.113	0.103	0.112	0.097	0.111	0.104	0.102	0.085
u[5]	0.098	0.090	0.115	0.104	0.110	0.096	0.115	0.104	0.099	0.081
u[6]	0.128	0.116	0.135	0.118	0.137	0.114	0.132	0.119	0.133	0.111
u[7]	0.083	0.078	0.104	0.094	0.096	0.086	0.101	0.094	0.087	0.071
u[8]	0.128	0.117	0.136	0.121	0.136	0.117	0.134	0.122	0.129	0.111
u[9]	0.097	0.091	0.114	0.104	0.110	0.097	0.114	0.104	0.100	0.083
u[10]	0.100	0.093	0.115	0.107	0.112	0.099	0.118	0.107	0.103	0.084
u[11]	0.159	0.143	0.152	0.131	0.165	0.136	0.153	0.133	0.164	0.146
u[12]	0.071	0.062	0.083	0.072	0.081	0.070	0.082	0.072	0.074	0.058
u[13]	0.123	0.114	0.132	0.119	0.131	0.114	0.131	0.120	0.124	0.107
u[14]	0.081	0.077	0.096	0.090	0.097	0.084	0.098	0.090	0.087	0.070
u[15]	0.197	0.154	0.174	0.134	0.194	0.145	0.177	0.136	0.205	0.167
u[16]	0.059	0.057	0.073	0.067	0.072	0.065	0.074	0.067	0.068	0.053
u[17]	0.067	0.062	0.085	0.078	0.081	0.072	0.084	0.077	0.072	0.057
u[18]	0.075	0.069	0.095	0.086	0.090	0.078	0.097	0.085	0.081	0.062
u[19]	0.094	0.087	0.111	0.102	0.106	0.094	0.112	0.102	0.097	0.078
u[20]	0.091	0.083	0.109	0.098	0.103	0.090	0.108	0.098	0.096	0.075
u[21]	0.071	0.067	0.089	0.083	0.085	0.076	0.089	0.082	0.076	0.061
u[22]	0.146	0.135	0.145	0.129	0.152	0.130	0.146	0.130	0.150	0.133
u[23]	0.128	0.115	0.137	0.122	0.135	0.116	0.137	0.123	0.128	0.106
u[24]	0.176	0.155	0.165	0.134	0.181	0.148	0.166	0.136	0.187	0.167
u[25]	0.119	0.111	0.132	0.119	0.131	0.113	0.132	0.120	0.122	0.102
u[26]	0.084	0.078	0.100	0.090	0.098	0.085	0.099	0.091	0.090	0.072
u[27]	0.094	0.090	0.113	0.104	0.111	0.096	0.113	0.104	0.100	0.081
u[28]	0.067	0.061	0.082	0.075	0.079	0.070	0.082	0.075	0.072	0.057
u[29]	0.164	0.149	0.154	0.133	0.167	0.141	0.158	0.135	0.171	0.157
u[30]	0.211	0.157	0.197	0.135	0.216	0.152	0.199	0.137	0.213	0.164
u[31]	0.071	0.066	0.085	0.078	0.084	0.074	0.085	0.078	0.077	0.061
u[32]	0.091	0.084	0.106	0.096	0.103	0.090	0.105	0.096	0.094	0.077
u[33]	0.080	0.075	0.097	0.088	0.092	0.082	0.095	0.088	0.085	0.069
u[34]	0.186	0.156	0.169	0.135	0.189	0.146	0.169	0.137	0.190	0.168
u[35]	0.102	0.094	0.117	0.105	0.113	0.098	0.117	0.106	0.105	0.085
u[36]	0.125	0.117	0.133	0.121	0.134	0.117	0.134	0.122	0.127	0.110
u[37]	0.088	0.081	0.104	0.095	0.100	0.088	0.104	0.095	0.091	0.074
u[38]	0.074	0.069	0.090	0.082	0.087	0.077	0.089	0.082	0.079	0.064
u[39]	0.099	0.090	0.111	0.100	0.109	0.094	0.112	0.100	0.101	0.083
u[40]	0.159	0.142	0.153	0.131	0.163	0.135	0.154	0.133	0.163	0.145
u[41]	0.078	0.071	0.093	0.083	0.091	0.078	0.094	0.083	0.082	0.065
u[42]	0.115	0.105	0.127	0.113	0.125	0.107	0.126	0.114	0.114	0.097
u[43]	0.064	0.062	0.081	0.075	0.080	0.071	0.080	0.075	0.072	0.057
$\lambda$	1.840	0.874	8.160	3.125	3.018	0.934	9.941	3.598	0.773	0.240
$\theta / \mu$					0.366	0.140	0.467	0.230	0.384	0.137

Table 5 Logs of Marginal Likelihoods

	Rice Data			Dairy Data			Electricity Data		
	MFVB	MCMC	MSE	MFVB	MCMC	MSE	MFVB	MCMC	MSE
<b>EXP</b>	-118.73	-117.09	0.0015	1122.95	1133.45	0.234	2.118	4.619	0.0033
<b>HN</b>	-119.17	-117.39	0.0039	1139.86	1152.67	0.132	1.100	4.010	0.0067
<b>Gamma</b>	-116.66	-116.16	0.0091	1157.04	1163.24	0.663	3.832	5.990	0.0512
<b>TN</b>	-119.78	-117.27	0.0045	1142.60	1155.86	0.717	0.305	4.188	0.0143
<b>LN</b>	-119.96	-117.31	0.0036	1140.93	1149.62	0.599	0.273	4.238	0.0237

Table 6 Logs of Bayes-Factors against Half-Normal Model

	Rice Data		Dairy Data		Electricity Data	
	MFVB	Bayes	MFVB	Bayes	MFVB	Bayes
<b>EXP</b>	0.447	0.302	-16.907	-19.221	1.018	0.609
<b>HN</b>	0.000	0.000	0.000	0.000	0.000	0.000
<b>Gamma</b>	2.515	1.232	17.183	10.565	2.732	1.981
<b>TN</b>	-0.606	0.116	2.741	3.186	-0.795	0.178
<b>LN</b>	-0.786	0.077	1.068	-3.051	-0.827	0.228

Table 7 MFVB and MCMC Posterior Means and Standard Deviations for

Lognormal Battese-Coelli Model – Rice Data

	MFVB		MCMC	
	Mean	SD	Mean	SD
<b>C</b>	-0.687	0.290	-0.709	0.398
<i>Land</i>	0.468	0.076	0.462	0.100
<i>Labor</i>	0.264	0.078	0.267	0.098
<i>Fert</i>	0.174	0.049	0.178	0.060
<i>Other</i>	0.041	0.024	0.040	0.028
$\sigma^{-2}$	14.408	1.554	14.485	3.269
$\lambda$	0.849	0.091	0.965	0.450
$\mu_0$	-2.144	0.303	-2.123	0.438
$\mu_1$	-0.076	0.037	-0.069	0.066
$\mu_2$	-1.061	0.272	-1.275	0.753

Table 8 MFVB and MCMC Posterior Means and Standard Deviations:

## Time-Varying Model– Dairy Data

	MCMC		MFVB	
	Mean	SD	Mean	SD
$\lambda$	5.7440	0.4670	5.6101	0.3562
$\lambda_\alpha$	1.1460	0.5080	1.1506	0.5095
$\alpha_2$	0.9290	0.0330	0.9266	0.0217
$\alpha_3$	0.8760	0.0320	0.8742	0.0217
$\alpha_4$	0.8710	0.0330	0.8679	0.0217
$\alpha_5$	0.8530	0.0330	0.8505	0.0217
$\alpha_6$	0.8170	0.0330	0.8133	0.0217



Minerva Access is the Institutional Repository of The University of Melbourne

**Author/s:**

Hajargasht, G; Griffiths, WE

**Title:**

Estimation and testing of stochastic frontier models using variational Bayes

**Date:**

2018-10

**Citation:**

Hajargasht, G. & Griffiths, W. E. (2018). Estimation and testing of stochastic frontier models using variational Bayes. *Journal of Productivity Analysis*, 50 (1-2), pp.1-24.  
<https://doi.org/10.1007/s11123-018-0531-0>.

**Persistent Link:**

<http://hdl.handle.net/11343/214680>

**File Description:**

Accepted version