J Prod Anal (2009) 31:101–112
DOI 10.1007/s11123-008-0122-6

Spatial stochastic frontier models: accounting for unobserved local determinants of inefficiency

Alexandra M. Schmidt · Ajax R. B. Moreira · Steven M. Helfand · Thais C. O. Fonseca

Published online: 25 December 2008

Abstract This paper analyzes the productivity of farms across 370 municipalities in the Center-West region of Brazil. A stochastic frontier model with a latent spatial structure is proposed to account for possible unknown geographical variation of the outputs. The paper compares versions of the model that include the latent spatial effect in the mean of output or as a variable that conditions the distribution of inefficiency, include or not observed municipal variables, and specify independent normal or conditional autoregressive priors for the spatial effects. The Bayesian paradigm is used to estimate the proposed models. As the resultant posterior distributions do not have a closed form, stochastic simulation techniques are used to obtain samples from them. Two model comparison criteria provide support for including the latent spatial effects, even after considering covariates at the municipal level. Models that ignore the latent spatial effects produce significantly different rankings of inefficiencies across agents.

Keywords Bayesian paradigm · Conditional autoregressive priors · Monte Carlo Markov chain · Stochastic frontier models · Spatial econometrics

JEL Classification C01 · C11

A. M. Schmidt is an Associate Professor of Statistics at the Federal University of Rio de Janeiro (UFRJ), Brazil; A. R. B. Moreira is a Research Fellow at the Applied Economics Research Institute (IPEA-RJ), Brazil; S. M. Helfand is an Associate Professor of Economics at the University of California, Riverside, USA; and, T. C. O. Fonseca is a PhD student in Statistics at the University of Warwick, UK.

A. M. Schmidt (✉)
Instituto de Matemática, Universidade Federal do Rio de Janeiro, Caixa Postal 68530, CEP 21945-970 Rio de Janeiro, RJ, Brazil
e-mail: alex@im.ufrj.br
URL: www.dme.ufrj.br/~alex

A. R. B. Moreira
IPEA: Instituto de Pesquisa Economica Aplicada Av Antonio Carlos, 51/17and., 20020100 Rio de Janeiro, RJ, Brazil

S. M. Helfand
Department of Economics, University of California, Riverside, CA 92521, USA

T. C. O. Fonseca
Department of Statistics, University of Warwick, Coventry CV4 7AL, UK

1 Introduction

In many real world examples of production, local conditions affect productivity. In the context of agriculture, which is the focus of the empirical application in this paper, differences across localities in transportation infrastructure, soils and climate, human capital, and other factors, can create systematic variation in the efficiency of agricultural production. Models with local fixed or random effects can be used if one is concerned with the bias that might arise due to omitted local variables, but is not interested in studying the impact of local conditions on productivity. Alternatively, when the impact of local conditions on productivity is of interest, variables that describe these conditions can be included directly in the production function. While interpretation is feasible in this case, the vector of determinants included in the model is unlikely to be complete. An additional local effect remains, and failure to account for this will lead to bias in the estimated coefficients. In this paper, we propose a solution to this problem. We construct a model that permits the inclusion of a vector of municipal level variables that have policy relevance and are of interest in their own right, while simultaneously capturing the effect of omitted local conditions.

In our approach, a municipal level latent effect is introduced in the inefficiency term of a stochastic frontier production function. Stochastic frontier production functions decompose output into three components: a deterministic component, a random disturbance, and the inefficiency of each producer (Aigner et al. 1977; Meeusen and van den Broeck 1977). In our model, the inefficiency of each producer depends on which municipality it is located in. Thus, the inefficiency u_{ij} of agent $j = 1,\ldots,n_i$ located in municipality $i = 1,\ldots,n$, is estimated conditioned on the unobserved local characteristics $\alpha_i, u_{ij} \sim p(\alpha_i)$, where $p(\alpha_i)$ is a positive and asymmetric probability density function that depends on α_i. The component α_i, which represents the unobserved local characteristics in each municipality, is a common component among the inefficiencies of all agents in each municipality. Therefore, the inefficiencies of agents that are in the same municipality are generated from the same distribution. These distributions, however, are allowed to differ across municipalities. The latent effects α_i could be independent of each other or spatially structured. The latter means that $\alpha_i - \alpha_k$, for all i not equal $k = 1,\ldots,n$ would tend to be smaller for municipalities that are closer together. We allow for this possibility in the model, and explore its relevance in the empirical section.

This article builds on contributions from three distinct literatures. First, it addresses issues of unobserved heterogeneity in panel data used to estimate a stochastic frontier production function. Tsionas (2002) incorporates heterogeneity in the context of a random coefficients stochastic frontier model, whereas Greene (2005) extends fixed and random effects models in the context of the stochastic frontier literature with a variety of approaches to incorporate firm specific heterogeneity. Greene discusses "true" fixed or random effects models in which there is an effect for each firm that is separate from the firm's inefficiency. In the context of Greene's panel these are firm level effects, whereas in our context these are municipal level effects.[1] We estimate a "true" effects model of the type proposed by Greene, and compare it to alternatives that allow for (a) the inclusion of covariates that would otherwise be perfectly collinear with the effects, (b) the unobserved heterogeneity to condition the inefficiencies, and (c) the unobserved heterogeneity to have a spatial structure. In contrast to Greene, the effects that we estimate are not simply incidental parameters. The sources of municipal level heterogeneity, both observed and unobserved, are of interest to us and have potential policy relevance.

Because we are concerned with the impact of local variables on efficiency, and the possible spatial correlation of these effects, this article is also related to the spatial econometrics literature. See Anselin (1988) for a review of spatial models, Gamerman and Moreira (2004) for a recent survey, and Druska and Horrace (2004) for an application to efficiency analysis in which productivity shocks are spatially correlated at each moment of time. Finally, we are not only interested in estimating efficiencies. We also seek to explain them. Helfand and Levine (2004) and Sampaio de Souza et al. (2005) provide two examples related to municipalities in Brazil—the focus of our empirical application. Both papers first use Data Envelopment Analysis (DEA) to estimate technical efficiency of agents in municipalities spread over a region of Brazil. Next, they use either a spatial SUR or a simultaneous autoregressive model, respectively, to investigate the determinants of those efficiencies. In contrast to our approach here, both papers perform the inference procedure in two steps, taking no explicit account of the uncertainty in the estimation of the efficiencies.

This article is part of a larger research project that uses microdata from the 1995/96 Agricultural Census in Brazil to study the determinants of agricultural productivity. The project is motivated by the observation that Brazil has one of the most unequal distributions of land holdings in the world, and an extremely high rate of rural poverty. Brazil is also among the largest producers and exporters in the world of many agricultural products. A deeper understanding of the determinants of productivity in Brazilian agriculture has important implications for reducing rural poverty among small farmers, and for relaxing constraints on macroeconomic growth by increasing foreign exchange earnings. This paper discusses methodological aspects of the larger project, and seeks to investigate if (i) the inefficiencies of the producers share a common municipal factor, (ii) this factor follows a spatial trend, and (iii) inclusion of this factor affects the inefficiency rankings of the producers.

This paper is organized as follows. Section 2 discusses the proposed model, and inference procedure. Section 3 presents the analysis of the data under the proposed model. As we explore different model specifications for the data, two model comparison criteria, one based on that suggested by Gelfand and Ghosh (1998) and the other on that by Spiegelhalter et al. (2002), are performed. Finally, Sect. 4 concludes the paper and describes avenues for future research.

2 Proposed model

Suppose that observations are available in the form of panel data. However, here, replications occur within each municipality rather than over time. Let y_{ij} be the logarithm of output of unit j in municipality i, $j = 1, 2,\ldots, n_i$,

[1] Greene considers a panel of firms (i) with repeated observations over time (t). We consider a panel of municipalities (i) with multiple observations on firms (j) in each municipality.

$i = 1, 2,\ldots, n$. We propose two different general model specifications, that is,

$$y_{ij} = g(\mathbf{r}_{ij}, \boldsymbol{\vartheta}) + \alpha_i - u_{ij} + \epsilon_{ij}, \quad \text{or} \tag{1}$$

$$y_{ij} = g(\mathbf{r}_{ij}, \boldsymbol{\vartheta}) - u_{ij}(\alpha_i) + \epsilon_{ij}, \tag{2}$$

where in both equations $g(\mathbf{r}_{ij}, \boldsymbol{\vartheta})$ represents the production function, $\mathbf{r}$ is a vector of traditional inputs and other covariates that might influence output, and $\boldsymbol{\vartheta}$ is a vector of parameters that describe the effect of each of the variables in $\mathbf{r}_{ij}$ on y_{ij}. The component $u_{ij}(.)$ models the inefficiency of unit j in location i. u_{ij} is assumed to be independent of ϵ_{ij} and to follow an asymmetric positive distribution. The random error, ϵ_{ij}, is normally distributed with mean zero and variance σ^2, that is, $\epsilon_{ij} \sim N(0, \sigma^2)$.

If there are no municipality level covariates, one might consider introducing α_i directly in the mean structure of y_{ij}. Equation 1 represents this specification. In this case, the inefficiencies are generated from a single distribution that does not vary across space, as u_{ij} does not depend on any local characteristic. But a local latent effect, α_i, is included in the mean structure of y_{ij}. This is Greene's (2005) "true" random effects model.

In contrast to the random effects model (1), the novelty of model (2) is that it considers each u_{ij} as a realization from a distribution that depends on a latent (unobserved) local effect α_i. This allows the distribution of the inefficiencies to vary across municipalities, while at the same time guaranteeing that all the units within a particular municipality have u_{ij} terms drawn from the same distribution. We allow for the possibility that α_i might represent a process that spreads through spatial contagion, such as with the diffusion of new technologies across space. These processes are usually represented by priors that vary smoothly across space. However, in our case, we only know the location of each municipality, not of each producer. Therefore, in several specifications, we assume that $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ follows a conditional autoregressive distribution which depends on its neighbors. Thus, a spatial structure is naturally imposed (Besag et al. 1991). Notice that in (2) the latent effects α_i are introduced in the second level of hierarchy, therefore they do not affect the y_{ij}'s directly. An advantage of this approach is that, due to the orthogonality between r and u, we are able to identify the municipal level effects (α_i) even when r contains municipal level covariates. In Sect. 3 we use several model comparison criteria to assess the performance of the specifications described in (1) and (2).

There are different proposals in the literature for the distribution of the inefficiency component: the exponential (Meeusen and van den Broeck 1977), the half-normal (Aigner et al. 1977), the truncated normal (Stevenson 1980), the gamma (Greene 1990), and the log-normal (Migon and Migon 2005). Here, we consider two distributions for u_{ij}, the truncated normal and the exponential. In the former, the variance is assumed to be common across units, and only the mean varies across municipalities. In the latter the mean and the variance are a function of the same parameter and these are allowed to vary across municipalities. More specifically, we assume

$$(u_{ij}|\alpha_i, \tau^2) \sim N^+(\alpha_i, \tau^2) \tag{3}$$

or

$$u_{ij}|\lambda_i \sim \exp(1/\lambda_i) \quad \text{with} \quad \log \lambda_i = \alpha_i, \tag{4}$$

where $N^+(a, b)$ denotes the normal distribution truncated at zero, whose associated normal has mean a and variance b, and $\exp(1/\lambda)$ denotes the exponential distribution with mean λ. The natural link function of the exponential distribution is used to model its mean as a function of α_i.

2.1 Inference procedure

As the inference procedure is made through the Bayesian Paradigm, an important issue is the effect of the prior distribution of the random effects α_i. Because the observations are being made across municipalities, we would expect these random effects to be spatially correlated. Due to the geographical characteristics of the data, it is natural to expect that the inefficiencies of units which are located in neighboring municipalities are of similar magnitude. That is, the magnitude of the inefficiencies might vary smoothly across the region. For this reason, following Besag et al. (1991), one possibility that we assume is a conditional autoregressive (CAR) prior for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)'$, that is

$$p(\boldsymbol{\alpha}|\psi^2) \propto \exp\left\{-\frac{1}{2\psi^2}\sum_{i=1}^{n}\sum_{j<i} W_{ij}(\alpha_i - \alpha_j)^2\right\}. \tag{5}$$

We denote this prior as $\boldsymbol{\alpha} \sim CAR(\psi^2)$. It is common practice to consider a $0 - 1$ neighborhood structure, such that $W_{ij} = 1$ if i and j share boundaries and 0 otherwise. We note that the distribution in Eq. 5 is an improper joint distribution for $\boldsymbol{\alpha}$, as we can add any constant to all α_i and (5) is unaffected (Banerjee et al. 2004). In order to remove posterior propriety concerns, we center the $\boldsymbol{\alpha}$ at each iteration of the algorithm described in the Appendix. As a result the posterior is guaranteed to be proper (see Besag and Kooperberg 1995; Gelfand and Sahu 1999 for more details). A simpler alternative is to assume a priori that the α_i are all independent among themselves, that is $\alpha_i \sim N(0, \psi^2), \forall i = 1, 2, \ldots, n$. We compare both specifications in Sect. 3.

Let $\mathbf{y} = (y_{11}, \ldots, y_{1n_1}, \ldots, y_{n1}, \ldots, y_{nn_n})'$ represent a random sample of the logarithm of the outputs of unit j, $j = 1, 2,\ldots, n_i$, in location $i = 1, 2,\ldots, n$. Under the specification in (3) or in (4), the likelihood is given by

$$p(\mathbf{y} \mid \mathbf{r}_{ij}, \boldsymbol{\vartheta}, u_{ij}, \tau^2, \sigma^2) \propto \left(\sigma^2\right)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left(y_{ij} - g(\mathbf{r}_{ij}, \boldsymbol{\vartheta}) + u_{ij}\right)^2 \right\}. \quad (6)$$

Assuming u_{ij} is distributed as in (3), Stevenson (1980) and Broeck et al. (1994) show that when integrating the likelihood above with respect to u_{ij} one obtains

$$p(\mathbf{y} \mid \mathbf{r}_{ij}, \boldsymbol{\vartheta}, \alpha_i, \sigma^2) \propto \prod_{i=1}^{n} \prod_{j=1}^{n_i} \frac{1}{\sqrt{\sigma^2 + \tau^2}} \Phi\left(\frac{y_{ij} - g(\mathbf{r}_{ij}, \boldsymbol{\vartheta}) + \alpha_i}{\sqrt{\sigma^2 + \tau^2}} \right) \times \Phi\left(\frac{\alpha_i \sqrt{\sigma^2 + \tau^2}}{\sigma\tau} - \frac{\tau(y_{ij} - g(\mathbf{r}_{ij}, \boldsymbol{\vartheta}) + \alpha_i)}{\sigma\sqrt{\sigma^2 + \tau^2}} \right) \Phi^{-1}\left(\frac{\alpha_i}{\tau}\right), \quad (7)$$

which is a skewed normal distribution as described in Domínguez-Molina et al. (2003). On the other hand, following the specification for u_{ij} as in (4), and marginalizing the likelihood with respect to u_{ij}, Stevenson (1980) and Broeck et al. (1994) show that

$$p(\mathbf{y} \mid \mathbf{r}_{ij}, \boldsymbol{\beta}, \alpha_i, \sigma^2) \propto \prod_{i=1}^{n} \prod_{j=1}^{n_i} \lambda_{ij}^{-1} \exp\left\{ \frac{-m_{ij}}{\lambda_{ij}} - \frac{\sigma^2}{2\lambda_{ij}^2} \right\} \Phi\left(\frac{m_{ij}}{\sigma}\right), \quad (8)$$

where $m_{ij} = g(\mathbf{r}_{ij}, \boldsymbol{\vartheta}) - y_{ij} - \frac{\sigma^2}{\lambda_{ij}}$, $\Phi(x)$ is the cumulative distribution and $\phi(x)$ is the probability density function of the standard normal distribution evaluated at point x. These marginalizations with respect to u_{ij} prove useful in the simulation methods which are described in the Appendix.

Following the Bayesian paradigm, the model specification is complete only after assigning a prior distribution to all unknowns (parameters) in the model. Thus far we have discussed the prior distribution of the latent random effects α_i, and of the one-sided disturbance term. Here we will discuss the prior distribution of the remaining parameters in the model. Let $\boldsymbol{\theta}$ comprise the parameter vector, such that $\boldsymbol{\theta} = (\boldsymbol{\vartheta}, \boldsymbol{\alpha}, \psi^2, \tau^2, \sigma^2)'$. Initially, we assume that all components of $\boldsymbol{\theta}$ are independent a priori. Therefore, the joint density prior distribution for $\boldsymbol{\theta}$ is given by

$$p(\boldsymbol{\theta}) = \prod_{i=0}^{q} [p(\vartheta_i)]\, p(\boldsymbol{\alpha} \mid \psi^2) p(\psi^2) p(\tau^2) p(\sigma^2).$$

Because the data we use in Sect. 3 have over 25,000 units, the information in the likelihood will dominate the information in the priors. For this reason, it is not problematic to use vague priors for most of the parameters in the model. Following standard procedures, for the coefficients $\vartheta_i, i = 1, \ldots, q$, we assign a zero mean normal prior distribution with variance (σ^2_{ϑ}) fixed at an arbitrarily large value, which in this case is $\sigma^2_{\vartheta} = 100$. For the variances τ^2 and σ^2 we assign inverse gamma prior distributions with both parameters equal to ε, i.e. $\tau^2 \sim IG(\varepsilon,\ \varepsilon)$, where ε is fixed at some small number, 0.1 in this case, to describe our prior ignorance about these parameters. Some care must be taken when assigning the prior distribution for ψ^2, the variance of the CAR distribution. This prior cannot be too uninformative (see Besag and Kooperberg 1995, p. 741 for details), as this is a non-identifiable parameter in the sense of Dawid (1979). We assign an inverse gamma prior, $IG(a_{\psi}, b_{\psi})$, whose mean is equal to the OLS variance estimate based on an independent stochastic frontier model, with a fixed variance. See, for example, Gelfand et al. (2004) and Banerjee et al. (2004), who adopt a similar strategy. Once the prior distribution has been assigned, following Bayes' theorem, the posterior distribution of $\boldsymbol{\theta}$ is proportional to the likelihood function times the prior. The resultant posterior does not have an analytical closed form. Therefore, simulation methods are needed to make inference about $p(\boldsymbol{\theta} \mid \mathbf{y})$. We make use of Markov chain Monte Carlo (MCMC) methods, in particular the Gibbs sampling algorithm (Gelfand and Smith 1990) with some steps of the Metropolis-Hastings (Hastings 1970) and the slice sampling (Neal 2003) algorithms. See the Appendix for more details on the implementation of the inference procedure.

3 Empirical analysis

The empirical analysis is based on a sample of farms in the Center-West region of Brazil. This region contained three states, 426 municipalities, and over 240,000 farms in 1996. In order to reduce forms of heterogeneity that are not the focus of this paper, we restricted the population to (a) owner operated farms, (b) family farms (defined as those that did not hire workers), and (c) farms that did not use inputs intensively (identified as farms in the lower half of the distribution of input expenditures per hectare). This group contains over 60,000 farms. From this, we drew a random sample of 25,494 spread over $n = 370$ municipalities such that we obtained at least 30 farms in each municipality.

Because we are interested in comparing the efficiency with which agricultural producers transform mulitple inputs into outputs, we follow standard practice of using an output quantity index as our dependent variable (see e.g. Coelli et al. 1998; Koop and Steel 2001). The index weights outputs with the same vector of average regional prices for all farms. By eliminating the spatial variation of prices, which is largely due to transportation costs, the output index eliminates one of the sources of spatial variation in the value of output.

We assume that the production frontier can be described by a constant returns to scale translog production function.[2] This is augmented by a number of variables that are relevant for explaining efficiency. More specifically, an output quantity index (y) is specified as a function of inputs (x), the conditions of each locality (z), fixed effects for farm size classes, and the existence of local public goods and institutions specific to each farm size class and locality (w), an asymmetric stochastic term $u_{ij}(.)$, and a random error (ε_{ij}).

The factors of production (x) include (a) area of the establishment, (b) the quantity of family labor, (c) the value of machines, (d) the value of other forms of capital (trees, stock of animals), and (e) expenditure on variable inputs such as fertilizer, pesticides, seeds, and hired labor. With the exception of the last variable, all inputs are determined prior to the production cycle and endogeneity is not a major concern. Variable inputs for each farm, in contrast, are clearly endogenous and were instrumented for by using the mean value from each locality and farm size group.

Farm size can affect productivity through a variety of channels. For example, the technology might have non-constant returns to scale, or there might exist deviations from the production frontier—inefficiency—that are common to size classes. The farm size dummies measure the non-linear effect of size on output, without identifying the source of these effects. Eight size classes are used here: 0–5 ha (the omitted class), 5–10, 10–20, 20–50, 50–150, 150–500 ha, 500–1000 and >1000 ha.

The productivity of an agent is also influenced by local conditions, represented by the existence of local public goods and institutions, soil conditions, climate, cost of transportation and distance to the consuming centers. These variables were divided into two groups according to how they were measured. The first group (w) is measured at the municipal level and farm size class of the unit. It includes access to electricity, technical assistance, cooperatives and credit, as well as the expected distance of the unit to the municipal seat. Other than distance, these variables were measured as a proportion of the agents who make use of the item. The second group (z) is measured only at the municipality level and includes (a) four indices of temperature and rainfall that capture the suitability of each location for temporary and/or permanent crop production, (b) five variables that capture soil, slope, and other physical attributes, (c) transport costs to the city of São Paulo, the dominant consumption center of Brazil, (d) distance to the sea, and to the relevant state capital, and (e) the illiteracy rate of the municipality.

[2] Deviations from constant returns to scale are captured in the farm size fixed effect described below. We adopt this specification in order to be consistent with the model used in our larger project on agricultural productivity, and because the restriction does not affect the spatial focus of the analysis here.

Here, we aim to investigate if the inclusion of municipal level covariates are able to capture the spatial structure present in the data, and how the inclusion or not of a latent effect influences the estimation of the efficiencies of the agents. For this reason, we entertain 16 different model specifications, considering two specifications for the distribution of the inefficiencies (truncated normal and exponential), two prior distributions for the latent effects α_i (CAR and independent $N(0, \psi^2)$) and the two specifications in Eqs. 1 and 2. For example, following the discussion above about the covariates we have available for this study, we initially propose

$$\mathbf{r}_{ij} = (\mathbf{x}_{ij}, \mathbf{w}_{ij}, \mathbf{z}_i)'. \tag{9}$$

The model that results from combining Eqs. 2 and 9 specifies that even after accounting for spatial variability using the information in $\mathbf{w}_{ij}$ and $\mathbf{z}_i$, there might be some *unobservable* spatial structure left which is captured by α_i. It is included in the mean of the inefficiencies due to the possible correlation with the $\mathbf{z}_i$'s, as already mentioned in Sect. 2. Another possibility would be to remove $\mathbf{z}_i$ from $g(.,.)$ and check if α_i would alone be able to capture all the spatial structure present in the data. In this case we assume in (2) that

$$\mathbf{r}_{ij} = (\mathbf{x}_{ij}, \mathbf{w}_{ij})'. \tag{10}$$

Rather than allow α_i to condition the inefficiencies, an alternative would be to use the municipal level random effects model described by Eq. 1. In this case $\mathbf{r}_{ij}$ is defined by (10), and u_{ij} in (1) is modelled either as $N^+(0, \tau^2)$ or as $\exp(1/\theta)$. With the random effects model in (1), it is clear that the inefficiencies are considered realizations from distributions which do not vary across municipalities. Table 1 summarizes the 16 models.

Because we are proposing different model specifications, we need some model choice techniques to compare the fitted models. From (7) and (8), the marginal likelihood will be difficult to compute. As we are using the CAR prior, implemented by centering $\boldsymbol{\alpha}$ at the end of each iteration to ensure a proper posterior (Besag et al. 1995), such marginalization is infeasible (Gelfand et al. 2006). Because of this, we compute two other criteria: the posterior predictive loss (EPD) which was introduced by Gelfand and Ghosh (1998) and the deviance information criterion (DIC), proposed by Spiegelhalter et al. (2002). Both criteria involve summing a goodness of fit term and a complexity penalty. Details are provided in the Appendix.

3.1 Results

Table 2 shows the values of EPD and DIC, and their respective components, under each of the sixteen fitted

Table 1 Specification of the 16 models fitted to the data (see Sect. 3 for details)

Models	Eqs. for y_{ij}	Dist. of Inef.	Dist. of α_i	Models	Eqs. for y_{ij}	Dist. of Inef.	Dist. of α_i
M1	2 and 9	$N^+(0, \tau^2)$	–	M9	2 and 9	$\exp(1/\tau^2)$	–
M2	2 and 9	$N^+(\alpha_i, \tau^2)$	$N(0, \psi^2)$	M10	2 and 9	$\exp(1/\lambda_i)$	$N(0, \psi^2)$
M3	2 and 9	$N^+(\alpha_i, \tau^2)$	$CAR(\psi^2)$	M11	2 and 9	$\exp(1/\lambda_i)$	$CAR(\psi^2)$
M4	2 and 10	$N^+(0, \tau^2)$	–	M12	2 and 10	$\exp(1/\tau^2)$	–
M5	2 and 10	$N^+(\alpha_i, \tau^2)$	$N(0, \psi^2)$	M13	2 and 10	$\exp(1/\lambda_i)$	$N(0, \psi^2)$
M6	2 and 10	$N^+(\alpha_i, \tau^2)$	$CAR(\psi^2)$	M14	2 and 10	$\exp(1/\lambda_i)$	$CAR(\psi^2)$
M7	1 and 10	$N^+(0, \tau^2)$	$N(0, \psi^2)$	M15	1 and 10	$\exp(1/\lambda_i)$	$N(0, \psi^2)$
M8	1 and 10	$N^+(0, \tau^2)$	$CAR(\psi^2)$	M16	1 and 10	$\exp(1/\lambda_i)$	$CAR(\psi^2)$

Table 2 Model comparison criteria for each of the sixteen model specifications fitted to the data (see text for details)

Models	EPD			DIC		
	P	G	D	p_D	$\overline{D}$	DIC
Models with municipality level covariates (z_i)						
M1	13556.87	3115.28	15114.51	15929.65	43814.68	59744.33
M2	12660.47	3070.08	14195.51	15295.18	42499.92	57795.10
M3	12529.30	2992.75	14025.68	15630.59	41975.33	57605.92
M9	14683.84	4872.77	17120.22	12604.94	47789.63	60394.58
M10	13286.06	3824.67	15198.40	13934.53	44391.28	58325.81
M11	13810.93	4476.45	16049.15	12830.83	46081.97	58912.80
Models without municipality level covariates (z_i)						
M4	13888.16	3243.67	15510.00	15795.33	44503.38	60298.70
M5	13322.49	3402.87	15023.93	15065.05	43878.90	58943.95
M6	13325.81	3437.99	15044.81	14979.33	43941.91	58921.24
M12	14938.05	5006.07	17441.09	12502.50	48291.04	60793.55
M13	13513.27	3835.90	15431.21	14029.91	44741.48	58771.38
M14	14135.30	4592.52	16431.56	12726.93	46630.87	59357.80
Models without municipality level covariates ($z)_i$ and with municipal random effects						
M7	12857.26	4455.67	15085.10	16482.82	42555.55	59038.36
M8	12858.75	5605.43	15661.46	16496.43	42526.44	59022.87
M15	13810.81	6028.05	16824.83	13081.53	46245.44	59326.97
M16	13809.80	7041.22	17330.41	13096.66	46214.57	59311.24

models defined in Table 1. Smaller values of EPD and DIC indicate better fit. Many different comparisons could be made based on this Table. For example, one could do a pairwise comparison of those models that have the local covariates z and those that do not. Another possible comparison is between the two distributions specified for the inefficiencies, the truncated normal and the exponential. Table 3 provides the sign of the difference between the respective values of EPD and DIC in these two settings. It is clear from these comparisons that the models with local level covariates (z) result in smaller values of both EPD and DIC. Table 3 also shows that the models with the truncated normal distribution generally perform better, as indicated by the smaller values of both EPD and DIC.

Table 3 Sign of the differences of EPD and DIC considering (i) models that were fitted with z and those that were not; (ii) the two specifications of the distribution of the inefficiencies

Pairwise comparison	EPD	DIC	Pairwise comparison	EPD	DIC
(i) *Models without z − models with z*					
M4 − M1	+	+	M12 − M9	+	+
M5 − M2	+	+	M13 − M10	+	+
M6 − M3	+	+	M14 − M11	+	+
(ii) *Exponential − truncated normal*					
M9 − M1	+	+	M12 − M4	+	+
M10 − M2	+	+	M13 − M5	+	−
M11 − M3	+	+	M14 − M6	+	+
M15 − M7	+	+	M16 − M8	+	+

We now turn our attention to the inclusion or not of a local latent effect, and the specification of this effect. Because the models with local covariates (z) performed better than the respective models without z, we restrict our attention to models with z. Notice, first, that the models that consider the latent local effect α_i perform better (smaller values of EPD and DIC) than those which do not. Compare, for example, the EPD of models 2 or 3 with model 1 (M2 = 14 195, M3 = 14 025, and M1 = 15 114). The results are ambiguous, however, about the specification of the prior for α_i. In the truncated normal case, which performed best in all cases, the spatial prior gives smaller values of EPD and DIC. This is not true when the exponential distribution is used.

In summary, the EPD and DIC criteria point to the following conclusions:

(1) Among models without local covariates, those models that allow the inefficiencies to depend on α_i perform better than the "true" random effects model discussed in Greene (2005). This is the comparison of M5 vs. M7 and of M13 vs. M15.
(2) Models with local covariates perform better than the respective models without local covariates.
(3) The latent local effect α_i brings information to the model even after considering local covariates. Therefore, there is some spatial structure left in the data which is being captured by α_i.
(4) The truncated normal distribution gives better results than the exponential distribution. This suggests that an extra parameter is needed to explain the variability of the inefficiencies, even though the variance is constant across municipalities.
(5) The best specification, according to the EPD and DIC criteria, is model 3. This model includes local covariates (z), conditions the distribution of the inefficiencies on α_i, and imposes a CAR structure on these latent effects.

In light of the observation above about model performance, in what follows we concentrate on the results for models M1, M2, M3, M9, M10, and M11.

In Table 4, we present the posterior summary of the sources of variability (σ^2, τ^2, the ratio $\kappa = \tau^2/(\tau^2 + \sigma^2)$ and ψ^2) for the different components of the models listed above. As described in Sect. 2, ψ^2, the variance of α_i, does not have a straight forward interpretation when the CAR prior is assumed, as it represents the variance of the *conditional* distribution of α_i given $\alpha_j \forall i \neq j$ and *not* of the marginal distribution of α_i (see Banerjee et al. (2004) for more details). We note that σ^2, the variance of the white noise, is the smallest for model M3, which is in agreement with EPD and DIC. We also note that κ is larger when α_i is included in the inefficiency term (models M2 and M3). This is a sign that α_i is capturing a source of variability that model M1 is not able to capture. It is worth mentioning that under model M9, τ^2, is the mean of the exponential distribution of u_{ij}. We do not show the summary of τ^2 under models M10 and M11 as in these cases, the mean of the exponential varies across municipalities.

One can also investigate how the inclusion of the local latent effect influences the coefficient estimates on the covariates in the model. Table 5 presents the posterior summary (median and 95% credible interval in brackets) of the coefficients for some of the most policy relevant covariates that are included in w and z. The comparison is performed with models M1, M2, and M3. Compared to M1, the inclusion of the latent effect in M2 and M3 frequently produces a considerable change in either the sign or magnitude of the coefficient. See, for example, how the coefficients on electricity and technical assistance change as we move from M1 to M3. When we compare models M2 and M3, in contrast, differences between the estimated coefficients and estimated posterior distributions tend to be smaller.

In contrast to the coefficients in Table 5, the impact of the farm size dummies on productivity is not sensitive to the specification of the model. Neither the distribution assumed for the inefficiencies nor the inclusion of the municipal level latent effect appears to have any significant impact on the posterior distributions of these coefficients. The results are not shown here due to space limitations. The economic implication of these results is that—

Table 4 Posterior summary, median and 95% credible intervals (in parentheses), for $\sigma^2, \tau^2, \kappa = \tau^2/(\tau^2 + \sigma^2)$ and ψ^2 for some of the fitted models

Models	σ^2	τ^2	κ	ψ^2
M1	0.3275 (0.3132; 0.3411)	1.9597 (1.9006; 2.030)	0.8567 (0.8486; 0.8651)	–
M2	0.3086 (0.2938; 0.3248)	1.9645 (1.8797; 2.0526)	0.8640 (0.8540; 0.8735)	1.0921 (0.8947; 1.3442)
M3	0.3050 (0.2887; 0.3214)	1.9607 (1.8556; 2.0678)	0.8654 (0.8547; 0.8761)	12.7935 (10.1537; 15.6325)
M9	0.3853 (0.3725; 0.3998)	1.2446 (1.2161; 1.2720)	–	–
M10	0.3373 (0.3260; 0.3483)	–	–	0.0576 (0.0444; 0.0745)
M11	0.3612 (0.3483; 0.3735)	–	–	2.0579 (1.6241; 2.613)

Table 5 Posterior summary, median and 95% credible intervals (in parentheses), for the coefficients of some local covariates under models M1, M2, and M3

Covariate in z_i	Models		
	M1	M2	M3
% With electricity	0.620 (0.53; 0.71)	0.402 (0.32; 0.499)	0.267 (0.14; 0.38)
% With techincal assistance	−0.254 (−0.31; −0.20)	0.050 (−0.03; 0.13)	0.097 (0.001; 0.19)
% In cooperatives	0.177 (0.09; 0.26)	−0.066 (−0.19; 0.06)	−0.158 (−0.29; −0.3)
% With access to credit	0.252 (0.14; 0.38)	0.417 (0.26; 0.59)	0.483 (0.27; 0.68)
Log of distance to municipal seat	−0.051 (−0.08; −0.03)	−0.012 (−0.04; 0.02)	0.028 (−0.005; 0.06)
Log of distance to São Paulo	0.014 (0.005; 0.02)	−0.007 (−0.02; 0.007)	−0.011 (−0.03; 0.006)
Log of distance to the sea	0.001 (−0.001; 0.003)	−0.006 (−0.01; −0.003)	−0.009 (−0.01; −0.005)
Log of distance to state capital	−0.027 (−0.05; −0.008)	−0.179 (−0.22; −0.13)	−0.237 (−0.28; −0.19)

independent of the model estimated—we observe an inverse relationship between farm size and productivity only up to about 1000 ha for this type of producer in this region of Brazil. This is broadly consistent with the results in Helfand and Levine (2004) and Moreira et al. (2007).

One of the aims of stochastic frontier models is to rank agents according to their efficiencies. Therefore, it is natural that we investigate the effect of the specification of the model on the ranking of agents. Because both model comparison criteria chose model M3 as best, we use the following approach to compare the other models with this one. First, recall that once we obtain a sample from the posterior distribution of the parameters in the model, we are able to obtain a sample of any function of these parameters, in particular of the inefficiencies u_{ij} and of the efficiencies $exp(-u_{ij})$. From these quantities we can also obtain posterior samples from the rankings of the efficiencies of the units. Let $rank(i, m)$ be the posterior median of the ranking of the efficiency of unit i under model m, $i = 1, 2,\ldots, 25494$. Then let $d(i,m,m') = |rank(i,m) - rank(i,m')|$, be the absolute difference between the estimated rankings of unit i under models m and m'. Now define $n_k(m,m') = \sharp\{i \,\text{such that}\, d(i,m,m') > n_k\}$ as the number of units for which the difference between the rankings is greater than percentile k. This measures the degree of distortion of the rankings under models m in relation to model m'. Normalizing this by the number of units n, (n_k/n) provides a measure of the relative degree of distortion between the models.

Table 6 presents the values of (n_k/n) for all of the models that include local covariates z in relation to the reference model M3. It is clear that the models which do not include α_i are those which result in the biggest differences. In this case, the difference in rankings is larger than 10%—or 2549 positions—for more than 25% of the units in the sample. When comparing with those models that assume the exponential distribution for the inefficiencies

Table 6 Pairwise comparison of the rankings of model M3 with the other models that include the municipal level covariates z

m = M3	Percentiles $n_k(m,m')/n$								
m'	0.005	0.01	0.03	0.05	0.10	0.15	0.20	0.25	0.30
M1	0.93	0.88	0.68	0.51	*0.25*	0.10	0.03	0.01	0.00
M9	0.94	0.88	0.71	0.57	*0.25*	0.09	0.03	0.01	0.00
M2	0.77	0.61	0.21	*0.07*	0.00	0.00	0.00	0.00	0.00
M10	0.91	0.83	0.56	*0.35*	0.08	0.02	0.00	0.00	0.00
M11	0.89	0.80	0.52	*0.29*	0.03	0.00	0.00	0.00	0.00

we have differences of more than 5% (1275 positions) for between 29% and 35% of the units. Finally, if we compare M3 with M2, thus checking for the effect of the prior of α_i on the difference of the rankings we notice that only 7% of the units differ by more than 5% of the positions. Although we do not show the results of the comparisons between all of the other models, we observed that the biggest differences among the rankings resulted from the inclusion or not of α_i.

Another way to investigate the impact of including α_i in the distribution of the inefficiencies would be to compute, for alternative models, the posterior mean of the inefficiencies in each municipality. Figure 1 summarizes this exercise for models M1 and M3. It shows the distribution of the means of the inefficiencies across municipalities for both models. It is clear that the models produce quite different results. Under M1, the municipal level inefficiencies are more concentrated in the middle of the distribution, leading to a bell-shaped histogram. Model M3, in contrast, which conditions the inefficiencies on α_i and imposes a CAR structure, produces a U-shaped histogram. Because M3 fit the data much better than M1, we conclude that M1 fails to capture a significant portion of the underlying heterogeneity that exists across municipalities in the Brazilian Center-West. Figure 2 shows a

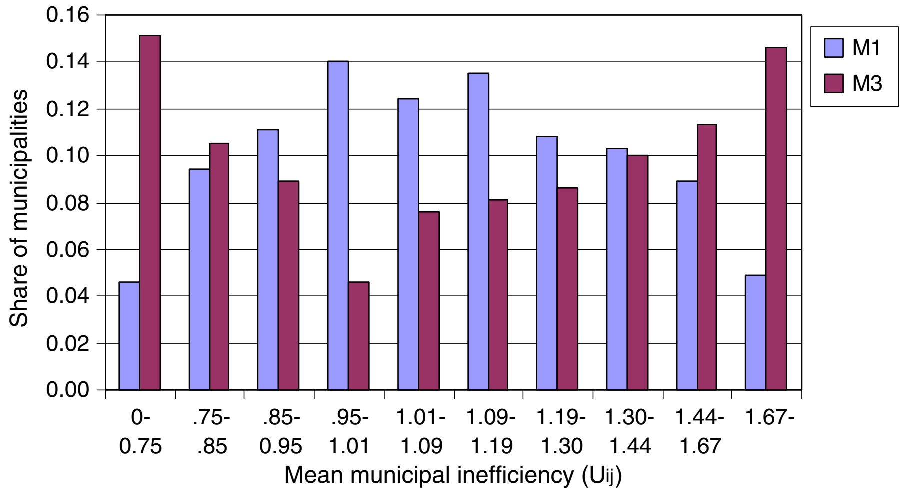

Fig. 1 Distribution of *u* across municipalities (see text for details)

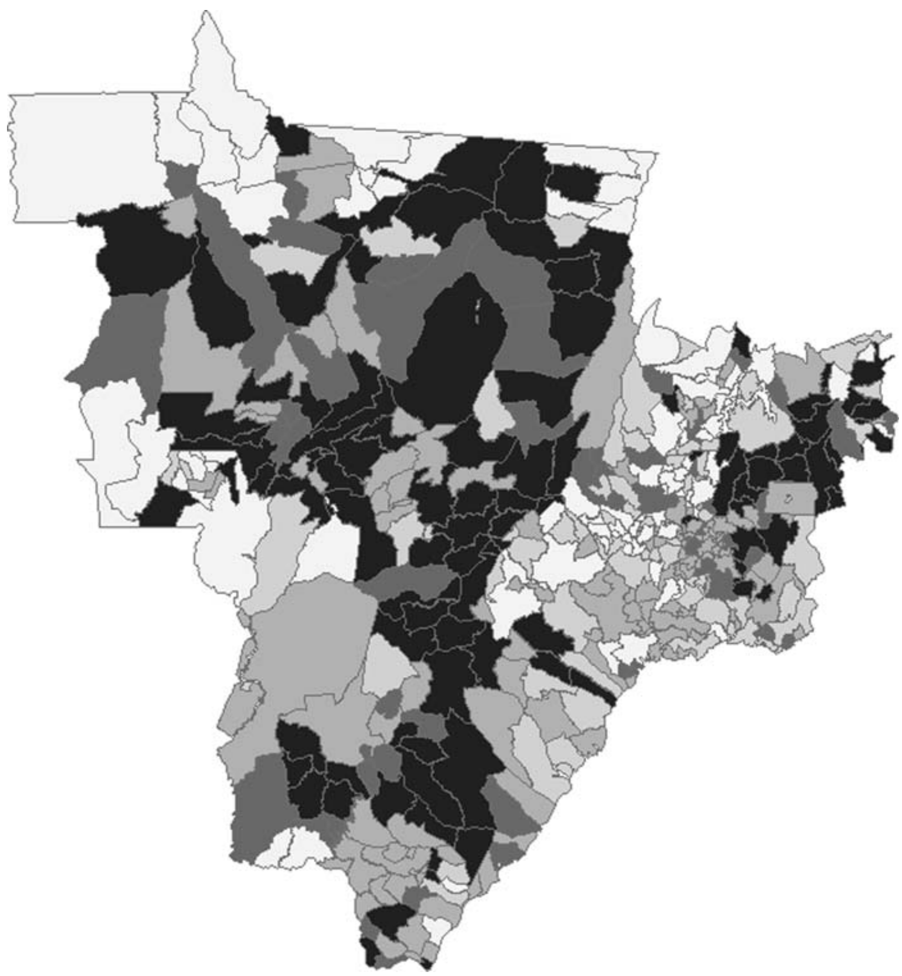

Fig. 2 Geographical display of the significance of the spatial effects α_i, $i = 1,\ldots,n$, under model M3. *Darker color* indicates positive significance (lower 2 sd credible limit above zero); *lighter color* negative significance (higher 2 sd credible limits below zero). The intermediate categories are: lower 1 sd credible limit above zero, credible interval including zero and higher 1 sd credible limit below zero

map of the latent spatial effects α_i for M3. These latent effects are what underlie the U-shaped histogram in Fig. 1. The Figure shows that most of the latent effects are significant, giving support to their presence in the model.

4 Concluding remarks and future work

This paper proposed a stochastic frontier model with a latent component in the one-sided disturbance term. The model is of particular interest for units which are observed across a geographical region. Inference was conducted under the Bayesian paradigm. Therefore, it was performed in a single framework, taking explicit account of the uncertainty involved in the estimation procedure of the parameters. We also explored the prior distributions one could consider for these latent effects. We proposed either the assignment of independent zero mean normal distributions or of a conditional autoregression model that imposes a spatial structure on the inefficiencies.

Data on inputs and outputs from a sample of farms in the Center-West region of Brazil was analyzed. Both the EPD and DIC model comparison criteria indicated the importance of including a municipal level latent effect in the inefficiency term of the model. This was the case even after the inclusion of municipal level covariates in the mean structure of y_{ij}. The best specification was the model that included local covariates (z), conditioned the distribution of the inefficiencies on the municipal level latent effects α_i, and imposed a CAR prior structure on these effects. This model performed better than models that only included latent effects in the mean of y_{ij}, such as in Greene (2005).

As an extension of the model proposed here, and following Ferreira and Schmidt (2006), one could investigate other spatial structures ($\mathbf{W}$) to model the $\boldsymbol{\alpha}$. Here we used a very simple one that is common in the literature. We treated municipalities that share common geographical boundaries as neighbors. But, based on some prior information, we could define a more complex structure of neighbors and explore its impact on a study of this type.

Acknowledgements This work benefitted from valuable discussions with Mark F. J. Steel, Helio S. Migon and Alan E. Gelfand. The research was supported by a grant from The United States Agency for International Development through a research program called BASIS/CRSP. Alexandra M. Schmidt was also supported by CNPq Grant No. 473312/2006-9. The authors wish to thank Liana O. Bernat for valuable research assistance and two anonymous referees and Professor Paul W. Wilson whose comments greatly improved the presentation of the paper.

Appendix: MCMC sampling scheme

Because of the marginalization in (7) or in (8), the full conditional distributions of the elements of $\boldsymbol{\vartheta}$ do not follow any known distribution. We make use of a Metropolis-Hastings random walk to sample from this full conditional. From our experience, the speed of convergence of the chains might be sensitive to the starting values of the elements of $\boldsymbol{\vartheta}$. We suggest to start the chain from their OLS estimates based on a multiple regression without the inefficiency component. In the case of the variances, σ^2 and τ^2, their full conditionals do not have an analytical closed form. Here we also suggest the use of a Metropolis-Hastings random walk step for $\log\sigma^2$ and $\log\tau^2$, respectively. On the other hand, ψ^2 has an inverse gamma posterior full conditional distribution regardless of whether an independent normal or a CAR prior is assigned to $\boldsymbol{\alpha}$.

Our main challenge when implementing the algorithm is sampling each α_i. Notice, that we have as many α_i as the number of municipalities in the sample. As the use of a random walk Metropolis-Hastings step requires the tuning of the variance of the proposal distribution, we prefer to use the slice sampling algorithm (Neal 2003) to sample from $\pi(\alpha_i \mid \boldsymbol{\theta}_{-\alpha_i}, \mathbf{y})$. The slice sampler is an auxiliary variable MCMC algorithm. It is based on the idea of slicing the target distribution (the full conditional of α_i in our case) horizontally at the contour level of a uniformly distributed variable $U(0, \pi(\alpha_i \mid .))$. Recall that the CAR density function is based on a pairwise difference. To guarantee propriety of the posterior we impose the constraint $\sum_i \alpha_i = 0$ after each iteration of the Gibbs sampler. See Besag et al. (1995) and Banerjee et al. (2004, p.164) for more details.

It is worth mentioning the computational advantage of marginalizing the likelihood with respect to u_{ij}. While running the MCMC, we avoid updating more than 25,000 parameters at each iteration of the algorithm. After convergence has been reached, we can use the samples from the elements of $\boldsymbol{\theta}$ to obtain posterior samples of the efficiencies, $\exp(-u_{ij}(.))$, through the posterior of u_{ij}.

The implementation of the MCMC algorithm was made in Ox[3]. Before fitting the proposed models to our dataset we performed a simulation study. This was done to check the efficiency of our sampling scheme and to ensure that the parameters of the model were being well estimated. In this simulation study we had more municipalities and less units per municipality. Quite vague priors were assigned to all parameters and the results showed that the summaries of posterior samples converged to the values used to create the data. Then, for the real dataset, for each of the 16 fitted models, we ran two parallel chains, each of length $L = 60{,}000$. We discarded the first 10,000 iterations and kept one at each 50th iteration, in order to avoid autocorrelation among the sampled values. Convergence was checked following standard procedures such as those described in Gamerman and Lopes (2006).

Model comparison criteria

Here we give more details about the two model comparison criteria used to compare the 16 models. We start by describing the posterior predictive loss criterion, then we describe the deviance information criterion.

[3] See http://www.doornik.com/ox/ for more details.

Posterior predictive loss

This measurement is based on replicates of the observed data, $Y_{i,j}^{rep}$, $i = 1,\ldots, n, j = 1,\ldots, n_i$, and the selected models are those that perform well under a loss function. This loss function penalizes actions both for departure from the corresponding observed value as well as for departure from what we expect the replicate to be (Banerjee et al. 2004). Assuming a normal model as in (6), and a squared loss function, the criterion is computed as

$$D_v = \frac{v}{v+1} G + P, \text{ where}$$
$$G = \sum_{i=1}^{n} \sum_{j=1}^{n_i} (\mu_{ij} - y_{ij}^{obs})^2 \quad \text{and} \quad P = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \sigma_{ij}^2.$$

In the equation above, $\mu_{ij} = E(Y_{ij}^{rep}|\mathbf{y})$ and $\sigma_{ij}^2 = Var(Y_{ij}^{rep}|\mathbf{y})$, the mean and variance of the predictive distribution of Y_{ij}^{rep} given the observed data $\mathbf{y}$. Banerjee et al. (2004) mention that ordering of models is typically insensitive to the choice of v, therefore we fix $v = 1$. Smaller values of D_v indicate better models. Notice that, after convergence has been reached, at each iteration of the MCMC we can obtain replicates of the observations given the sampled values of the parameters. Then we approximate the expected values above via Monte Carlo integration.

Deviance information criterion

Spiegelhalter et al. (2002) propose a generalization of the AIC based on the posterior distribution of the deviance, $D(\boldsymbol{\theta}) = -2\log p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta})$. The Deviance Information Criterion (DIC) is defined as

$$DIC = \overline{D} + p_D = 2\overline{D} - D(\overline{\boldsymbol{\theta}}),$$

where $\overline{D}$ defines the posterior expectation of the deviance, $\overline{D} = E_{\boldsymbol{\theta}|\mathbf{y}}(D)$, and p_D is the effective number of parameters, $p_D = \overline{D} - D(\overline{\boldsymbol{\theta}})$ and here $\overline{\boldsymbol{\theta}}$ represents the posterior mean of the parameters. Smaller values of DIC indicate a better fitting model. Notice that computation of DIC is easily achieved through MCMC methods.

There is no agreement in the Bayesian literature about which criterion works best. For example, Banerjee et al. (2004) do not recommend a choice between EPD and DIC. They claim that both involve summing a goodness of fit term and a complexity penalty. They go on to say that the fundamental difference is that the latter works in the parameter space with the likelihood, while the former works in the predictive space with posterior predictive distributions. They suggest that if the objective is to use the model for explanation, DIC should be preferred; whereas if the objective is prediction, EPD should be used.

References

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

Anselin L (1988) Spatial econometrics: methods and models. Kluwer Academic Publishers, Dordrecht

Banerjee S, Carlin B, Gelfand A (2004) Hierarchical modeling and analysis for spatial data. CRC Press, Chapman Hall

Besag J, Kooperberg C (1995) On conditional and intrinsic autoregression. Biometrika 82:733–746

Besag J, York J, Mollié A (1991) Bayesian image restoration, with two applications on spatial statistics (with discussion). Ann Inst Stat Math 43:1–59

Besag J, Green P, Higdon D, Mengersen K (1995) Bayesian computation and stochastic systems (with discussion). Stat Sci 10:3–66

Broeck Jv, Koop G, Osiewalski J, Steel MFJ (1994) Stochastic frontier models. J Econom 61:273–303

Coelli T, Rao DSP, Battese GE (1998) An introduction to efficiency and productivity analysis. Kluwer Academic Publihers, Norwell

Dawid A (1979) Conditional independence in statistical theory (with discussion). J R Stat Soc Ser B 41:1–31

Domínguez-Molina JA, González-Farías G, Ramos-Quiroga R (2003) Skew-normality in stochastic frontier analysis. Tech. Rep., PE/CIMAT, Mexico.

Druska V, Horrace WC (2004) Generalized moments estimation for spatial panel data: Indonesian rice farming. Am J Agric Econ 86:185–198

Ferreira GS, Schmidt AM (2006) Spatial modelling of the relative risk of dengue fever in rio de janeiro for the epidemic period between 2001 and 2002. Braz J Probab Stat 20:29–47

Gamerman D, Lopes HF (2006) Markov Chain Monte Carlo—stochastic simulation for Bayesian inference, 2nd edition. London, UK, Chapman & Hall

Gamerman D, Moreira ARB (2004) Multivariate spatial regression models. J Multivar Anal 91:262–281

Gelfand A, Ghosh S (1998) Model choice: a minimum posterior predictive loss approach. Biometrika 85:1–11

Gelfand AE, Sahu SK (1999) Identifiability, improper priors, and Gibbs sampling for generalized linear models. J Am Stat Assoc 94:247–253

Gelfand A, Smith A (1990) Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 85:398–409

Gelfand AE, Schmidt AM, Banerjee S, Sirmans CF (2004) Nonstationary multivariate process modeling through spatially varying coregionalization (with discussion). Test 13:1–50

Gelfand AE, Silander J, Wu S, Latimer A, Lewis PO, Rebelo AG, Holder M (2006) Explaining species distribution patterns through hierarchical modeling. Bayesian Anal 1:41–92

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

Greene WH (2005) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. J Econom 126:269–303

Hastings W (1970) Monte Carlo sampling methods using Markov Chains and their applications. Biometrika 57:97–109

Helfand SM, Levine ES (2004) Farm size and the determinants of productive efficiency in the Brazilian Center-West. Agric Econ 31:241–249

Koop G, Steel MFJ (2001) Bayesian analysis of stochastic frontier models. In: Baltagi B (eds) A companion to theoretical econometrics. Blackwell, Oxford, pp. 520–573

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

Migon HS, Migon MN (2005) Hierarchical bayesian models for stochastic frontier. Estadistica 57:27–52

Moreira ARB, Helfand S, Figueiredo AMR (2007) Explicando as diferenças na produtividade agrícola no Brasil (in Portuguese). Tech. rep., Instituto de Pesquisa Econômica Aplicada (IPEA-RJ), TD 1254, Rio de Janeiro, Brazil

Neal RM (2003) Slice sampling (with discussion). Ann Stat 31: 705–767

Sampaio de Souza MC, Cribari-Neto F, Stosic BD (2005) Explaining DEA technical efficiency scores in an outlier corrected environment: the case of public services in Brazilian municipalities. Brazilian Rev Econom 25:289–315

Spiegelhalter D, Best N, Carlin B, van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). J R Stat Soc Ser B 64:583–639

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

Tsionas EG (2002) Stochastic frontier models with random coefficients. J Appl Econom 17: 127–147