This is a postprint version of the following published document:

Veiga, M.H., Galán, J. y Wiper, M. (2014) Bayesian estimation of inefficiency heterogeneity in stochastic frontier models. Journal of Productivity Analysis, v. 42, n. 1, pp. 85-101. Avalaible in:
http://dx.doi.org/10.1007/s11123-013-0377-4

Bayesian estimation of inefficiency heterogeneity in stochastic frontier models

Jorge E. Galán · Helena Veiga · Michael P. Wiper

Abstract Estimation of the one sided error component in stochastic frontier models may erroneously attribute firm characteristics to inefficiency if heterogeneity is unaccounted for. However, unobserved inefficiency heterogeneity has been little explored. In this work, we propose to capture it through a random parameter which may affect the location, scale, or both parameters of a truncated normal inefficiency distribution using a Bayesian approach. Our findings using two real data sets, suggest that the inclusion of a random parameter in the inefficiency distribution is able to capture latent heterogeneity and can be used to validate the suitability of observed covariates to distinguish heterogeneity from inefficiency. Relevant effects are also found on separating and shrinking individual posterior efficiency distributions when heterogeneity affects the location and scale parameters of the one-sided error distribution, and consequently affecting the estimated mean efficiency scores and rankings. In particular, including heterogeneity simultaneously in both parameters of the inefficiency distribution in models that satisfy the scaling property leads to a decrease in the uncertainty around the mean scores and less overlapping of the posterior efficiency distributions, which provides both more reliable efficiency scores and rankings.

J. E. Galán (✉) · H. Veiga · M. P. Wiper
Department of Statistics and Instituto Flores de Lemus, Universidad Carlos III de Madrid, C/Madrid 126, 28903 Getafe, Spain
e-mail: jegalan@est-econ.uc3m.es

H. Veiga · M. P. Wiper
BRU/UNIDE, Avenida das Forças Armadas, 1600-083 Lisboa, Portugal

Keywords Stochastic frontier models · Efficiency · Unobserved heterogeneity · Bayesian inference

JEL Classification C11 · C23 · C51 · D24

1 Introduction

Stochastic frontier models, first introduced in Aigner et al. (1977) and Meeusen and van den Broeck (1977), are important tools for efficiency measurement. These models require the specification of an economic, functional form based on a production or cost function which includes a composite error term. This error term can be decomposed into two parts, firstly a two-sided, idiosyncratic error and secondly, a non-negative inefficiency component. Measures of efficiency are obtained from this one-sided error, which is typically assumed to follow some specific distribution. The most common distributions for the one-sided error are the half-normal (Aigner et al. 1977), exponential (Meeusen and van den Broeck 1977), truncated normal (Stevenson 1980), and gamma (Greene 1990).

However, the estimated inefficiency component often includes some firm characteristics other than outputs, inputs, or prices defined from the production or cost function, which should not be attributed to inefficiency. These are exogenous variables (e.g. type of ownership, GDP level in the country of operation) that have an effect on the technology used by firms or directly on their inefficiency. If these variables are not taken into account in the model specification, this may affect the estimation of the inefficiencies or of the frontier significantly.

Firm characteristics can be modeled in the frontier if they imply heterogenous technologies or in the one-sided error component if they affect the inefficiency. In the former case, covariates are directly included in the functional form and the main interest is to model unobserved heterogeneity (see Greene 2005). In the case of heterogeneity in the inefficiency, covariates are usually included in the parameters of the one-sided error distribution (see Huang and Liu 1994).

Heterogeneity in stochastic frontier models has also been studied in the Bayesian context. The Bayesian approach to stochastic frontiers introduced by van den Broeck et al. (1994) presents advantages in terms of formally deriving posterior densities for individual efficiencies, incorporating economic restrictions, and in the easy modeling of random parameters through hierarchical structures. Hierarchical models have been used to capture heterogeneous technologies (see Tsionas 2002) and heterogeneity in the inefficiency has been considered through covariates in the distribution of the non-negative error component (see Koop et al. 1997). Modeling observed heterogeneity using non parametric and flexible mixtures of inefficiency distributions are other interesting recent contributions (see Griffin and Steel 2004, 2008).

On the other hand, unobserved heterogeneity in the non-negative error component has been very little explored in the literature from a frequentist or a Bayesian approach. However, ignoring its existence means that heterogeneity which is not captured by observed covariates is wrongly attributed to inefficiency and consequently leads to bad efficiency estimates.

In this work, we propose, within a Bayesian framework, the inclusion of a random parameter in the distribution of the inefficiency with the aim of capturing unobserved heterogeneity. This parameter has three characteristics. It can be allowed to be time-varying, it can be included simultaneously with observed covariates in the inefficiency distribution in order to distinguish observed from unobserved heterogeneity and it can indicate whether or not observed covariates do a good job in capturing the existing heterogeneity.

Regarding the one-sided error, we use a truncated normal distribution, which is one of the most used distributions in studies involving observed heterogeneity in the inefficiency. In particular, covariates are often included in the location parameter of this distribution following the Battese and Coelli (1995) model. However, it is not clear in which parameter of the inefficiency distribution heterogeneity should be included. Wang (2002) proposed modeling the covariates simultaneously in the location and scale parameters of the truncated distribution. Alvarez et al. (2006) analyze a particular specification of truncated normal distributed inefficiencies that has the property of preserving the shape while changing the scale of the inefficiency, and also estimate a model where heterogeneity is captured only by the scale parameter of this distribution. We think that at an individual level, the moments of the distributions affected have different effects on the posterior efficiency distributions of each firm. Since this is possible to be studied from a Bayesian context, a second aim of this work is to analyze the effects on the posterior efficiency distributions of including both observed and unobserved heterogeneity in the location, scale or both parameters of the truncated normal distribution. For the latter case, we extend to the Bayesian framework the scaling property model proposed by Alvarez et al. (2006). This allows us to think of the inefficiency as being composed of two parts, one component capturing natural managerial skills and other component which depends on observed and unobserved firm characteristics.[1]

For illustration, we use two data sets which have been previously analyzed only in the frequentist context. The first data set is from a controversial report by the World Health Organization (WHO) on the efficiency of national health systems (see WHO 2000), while the second evaluates the economic efficiency of US domestic airlines. These two applications allow us to explore our models in different directions. In particular, in the WHO application, since the observed covariates are inefficiency related and time invariant, we include them in different parameters of the inefficiency distribution together with a time invariant random parameter. On the other hand, in the second application observed heterogeneity variables are time-varying and frontier drivers, so the unobserved heterogeneity component is allowed to change over time and its effects in the posterior efficiency distributions are evaluated when it is included in the location, scale or both parameters of the one-sided error distribution.

Our proposal of using a random parameter is successful in capturing unobserved inefficiency heterogeneity whether its is modeled alone or together with observed covariates. Moreover, we find that capturing heterogeneity using models that preserve the scaling property leads to less uncertainty around mean efficiency scores and less overlapping of posterior efficiency distributions.

The rest of this paper is organized as follows. Section 2 presents a brief literature review on heterogeneity in stochastic frontier models and the proposed model. Section 3 presents the Bayesian inference and model selection criteria. Section 4 reports the applications to the WHO and the US domestic airlines data sets. Finally, in Section 5 we provide conclusions and consider some possible extensions of our approach.

[1] We also studied these effects using models that follow half-normal and exponential distributions for the inefficiency. These results are available from the authors upon request.

2 Inefficiency heterogeneity in stochastic frontier models

2.1 A brief literature review

The original stochastic frontier model introduced by Aigner et al. (1977) and Meeusen and van den Broeck (1977) has the following form:

$$y_{it} = \mathbf{x_{it}}\boldsymbol{\beta} + v_{it} - u_{it} \tag{1}$$

where y_{it} represents the output of firm i at time t, $\mathbf{x_{it}}$ is a vector that contains the input quantities used in the production process, v_{it} is an idiosyncratic error that is typically assumed to follow a normal distribution and u_{it} is the one-sided component representing the inefficiency and follows some non-negative distribution.

Firm specific heterogeneity not specified in (1) can be mistaken for inefficiency if it is not identified. Heterogeneity can either shift the efficiency frontier or change the location and scale of the inefficiency estimations (see Kumbhakar and Lovell 2000; Greene 2008, for complete reviews). In general, when external factors are supposed to capture technological differences and these are out of the firms' control, heterogeneity should be specified in the frontier. In this case, the main interest is capturing unobserved effects. In the classical context, this has been modeled through fixed and random effects or models with random parameters (see Greene 2005). Bayesian approaches have been based on frontier models with hierarchical structures (see Tsionas 2002; Huang 2004).

When heterogeneity is more related to efficiency and thus more likely to be under firms' control, then this should affect directly the one-sided error term. In the parametric context, inefficiency heterogeneity is often included in the location or scale parameters of the inefficiency distribution. For example, covariates shift the underlying mean of inefficiency in Kumbhakar et al. (1991), Huang and Liu (1994) and Battese and Coelli (1995). A reduced form of these models assumes that the location parameter of the distribution of u_{it} depends on vectors of covariates $\mathbf{z_{it}}$ and parameters $\boldsymbol{\delta}$ as follows:

$$\begin{aligned} u_{it} &\sim N^+(\mu_{it}, \sigma_u^2) \\ \mu_{it} &= \mathbf{z_{it}}\boldsymbol{\delta}. \end{aligned} \tag{2}$$

The scale parameter of the one-sided error component has also been modeled as a function of firm characteristics. Reifschnieder and Stevenson (1991) provided one of the first linear specifications where this parameter varies across firms. A similar model was proposed by Caudill et al. (1995) with the aim of treating heteroscedasticity in frontier models. These authors found biased inefficiency estimations when heteroscedasticity was not accounted for.[2] The proposed model specifies the variance of a half-normal distributed inefficiency as an exponential function of time invariant covariates:

$$\begin{aligned} u_i &\sim N^+(0, \sigma_{u_i}^2) \\ \sigma_{u_i} &= \sigma_u \cdot \exp(\mathbf{z_i}\boldsymbol{\gamma}). \end{aligned} \tag{3}$$

Although the original proposal in (3) was presented in a cross sectional framework, it can be easily extended to include time-varying covariates and inefficiencies (see Hadri et al. 2003a, b, for an extension to panel data). It is also possible to define $u_{it} = u_i \cdot g(t)$ where $g(t)$ is a function of time (e.g. the parametric funtion introduced by Battese and Coelli 1992). The specification in (3) has the characteristic of changing the scale of the inefficiency distribution while preserving its shape and is referred in the literature as the scaling property (see Wang and Schmidt 2002; Alvarez et al. 2006). In general, this property allows us to think about inefficiency as being composed of two parts: $u_{it} = u_{it}^* \cdot f(\mathbf{z_{it}}, \boldsymbol{\delta})$. The first component is a base inefficiency, which is not affected by firm characteristics and captures random managerial skills, while the second component is a function of heterogeneity variables determining how well management is performed under these conditions. Another important feature of this property is that the interpretation of the effects of covariates on the inefficiency is direct and independent of the inefficiency distribution. The scaling property also holds when the inefficiency is exponentially distributed (see Simar et al. 1994), or in a particular case of truncated normal inefficiency where both parameters are an exponential function of firm characteristics as follows (see Wang and Schmidt 2002; Alvarez et al. 2006):

$$\begin{aligned} u_{it} &\sim N^+(\mu_{it}, \sigma_{u_{it}}^2) \\ \mu_{it} &= \mu \cdot exp(\mathbf{z_{it}}\boldsymbol{\delta}) \\ \sigma_{u_{it}} &= \sigma_u \cdot \exp(\mathbf{z_{it}}\boldsymbol{\delta}). \end{aligned} \tag{4}$$

Specification (4) for the inefficiency is a variation of a previous proposal by Wang (2002) where both the mean and the variance of truncated normal inefficiencies are simultaneously affected by the same covariates but with different coefficients. Other authors have also proposed heterogeneity specifications that include firm characteristics in the variance of the idiosyncratic error with the aim of treating heteroscedasticity in frontier models (see Hadri, 1999).

In the Bayesian context, Koop et al. (1997) presented different structures for the mean of the inefficiency component as Bayesian counterparts to the classical fixed and random effects models. One of these specifications is the

[2] In a previous study, Caudill and Ford (1993) also found biased estimates of the frontier parameters.

varying efficiency distribution model, which includes firm specific covariates in the parameter of an exponential distribution. These covariates link the firm effects and only the inefficiencies of firms sharing common characteristics are drawn from the same distribution. The distribution below presents a time invariant inefficiency that depends on vectors of binary covariates $\mathbf{z_i}$ and parameters γ:

$$u_i \sim Ex(\lambda_i^{-1})$$
$$\lambda_i = \exp(\mathbf{z_i}\gamma). \tag{5}$$

Since this model is intended to be a counterpart of a frequentist random effects model, it is specified to obtain time invariant inefficiencies. However, as in the case of (3), it is possible to define $u_{it} = u_i \cdot g(t)$ or to include time-varying covariates. Also, it would be possible to draw inefficiencies for every firm and period of time from the distribution with a firm specific parameter.

The literature on modeling unobserved firm characteristics in the inefficiency is still scarce. In the frequentist context, Greene (2005) proposed a model where the coefficients of the observed covariates are allowed to be firm specific and vary randomly. In the Bayesian framework, Koop et al. (1997) propose a model that may capture unobserved inefficiency heterogeneity. In this case, the inefficiency is assumed to be exponentially distributed with firm specific mean and independent priors.

2.2 The model

In this section, we present a general stochastic frontier model for panel data that allows the modeling of both observed and unobserved inefficiency heterogeneity. For the one-sided error we use an exponential specification of a truncated normal distribution where the location, scale, or both parameters can model firm heterogeneity. The general model in the case of a production function is:

$$y_{it} = \mathbf{x_{it}}\boldsymbol{\beta} + \mathbf{z^*_{it}}\boldsymbol{\delta} + v_{it} - u_{it}; v_{it} \sim N(0, \sigma_v^2)$$
$$u_{it} \sim N^+(\mu \cdot \exp(\mathbf{z_{it}}\gamma I_1 + \tau_{it} I_2), \sigma_u^2 \cdot (\exp(\mathbf{z_{it}}\gamma I_3 + \tau_{it} I_4))^2), \tag{6}$$

where y_{it} is the output of firm i at time t, $\mathbf{x_{it}}$ is the row vector of input quantities, $\mathbf{z^*_{it}}$ is a row vector of the observed heterogeneity variables that affect the technology; $\mathbf{z_{it}}$ is a row vector of observed covariates with effects in the inefficiency; τ_{it} is a random parameter that captures time-varying unobserved firm effects in the inefficiency; and, $\boldsymbol{\beta}, \boldsymbol{\delta}$, and γ are the corresponding parameter column vectors. I_1 to I_4 are indicator variables taking the value of 1 when either observed covariates or unobserved heterogeneity are accounted for in the location or scale parameters, respectively, and 0 otherwise.

This model nests other specifications in the literature that capture only observed heterogeneity. When I_3 and I_4 are equal to zero, the model reduces to an exponential specification of the Battese and Coelli (1995) model in (2). If I_1 and I_2 are equal to 0, the model allows only the scale parameter to include heterogeneity. This specification has only been studied before by Alvarez et al. (2006) in the framework of testing the scaling property. If additionally the location parameter μ is set to zero, our model becomes an extension of the half-normal model proposed by Caudill et al. (1995) in (3). Finally, if both parameters are allowed to include simultaneously the same type of heterogeneity (I_1,$I_3 = 1$ or/and I_2,$I_4 = 1$) our proposal becomes an extension of the scaled Stevenson model in (4). In case heterogeneity is considered time invariant, the vector of observed covariates $\mathbf{z_{it}}$ and the unobserved heterogeneity parameter τ_{it} can be set to vary only across firms.

It is easy to extend this specification to a hierarchical model which also allows for additional, unobserved, firm effects in the technology. However, in practical applications, mean posterior efficiencies are found to be very close to 1 for almost all firms (see Huang 2004; Tsionas 2002, for similar results). From our point of view, these results are inconclusive as they do not allow us to get reliable efficiency rankings.

3 Bayesian inference

The use of Bayesian methods in stochastic frontier analysis was introduced by van den Broeck et al. (1994) and has become very common in recent applications. Bayesian approaches have various attractive properties and, in particular, restrictions such as regularity conditions are easily incorporated and parameter uncertainty is formally considered in deriving posterior densities for individual efficiencies.

All the models derived from the general specification in (6) are fitted by Bayesian methods. In order to do this, we first need to introduce prior distributions for the model parameters. We assume proper but relatively disperse prior distributions throughout. In particular, the distributions assumed for the parameters in the frontier function are as follows: $\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{\Sigma_\beta}), \boldsymbol{\delta} \sim N(\mathbf{0}, \Sigma_\delta)$ with diffuse, inverse gamma priors for the variances. Finally, the variance of the idiosyncratic error term is inverse gamma, that is equivalent to $\sigma_v^{-2} \sim G(a_{\sigma_v^{-2}}, b_{\sigma_v^{-2}})$ with low values for the shape and scale parameters.

Regarding observed inefficiency heterogeneity, the distribution of the one-sided error component for the truncated normal model is: $u_{it}|\gamma, \mathbf{z_{it}} \sim N^+(\mu \cdot \exp(\mathbf{z_{it}}\gamma), \sigma_u^2 \cdot (\exp(\mathbf{z_{it}}\gamma))^2)$, where μ and σ_u^2 are defined as in

Griffin and Steel (2007). When models include heterogeneity in the inefficiency $\boldsymbol{\gamma}$ is $N(\mathbf{0}, \boldsymbol{\Sigma}_\gamma)$ distributed with a diffuse prior for the covariance matrix.

In the case of unobserved heterogeneity in the inefficiency, the unknown parameter is specified to have a hierarchical structure: $\tau_{it} \sim N(\overline{\tau}, \sigma_\tau^2)$, where $\overline{\tau} \sim N(0, 10)$ and $\sigma_\tau^{-2} \sim G(0.5, 0.5)$. The random parameter τ_{it} can be defined to be either time-varying or not.

The complexity of these models makes it necessary to use numerical integration methods such as Markov Chain Monte Carlo (MCMC), and in particular the Gibbs sampling algorithm with data augmentation as introduced by Koop et al. (1995). For our models, implementation was carried out using the WinBUGS package following the general procedure outlined in Griffin and Steel (2007). For models not considering unobserved heterogeneity in the inefficiency, the MCMC algorithm involved 50,000 MCMC iterations where the first 10,000 were discarded in a burn-in phase. On the other hand, for models including our proposal to capture unobserved heterogeneity, hyperparameters $\overline{\tau}$ and σ_τ^{-2} presented slow convergence and high autocorrelation. In particular, if initial values are set far from the posterior mean, convergence is observed only after 50,000 iterations and autocorrelations of order around 20 are identified. Therefore, for these models 550,000 iterations were used for the MCMC, thinning every 25 iterations and discarding the first 50,000. Finally, although we do not display the details here, sensitivity analysis of our results to changes in other prior parameters was also carried out. Results showed that the posterior inference was relatively insensitive to small changes in these parameters.

3.1 Model selection

The different models are evaluated in terms of three criteria, the DIC_3, which is a variant of the Deviance Information Criterion (DIC), the Log Predictive Score (LPS) and the Mean Square Error (MSE) of predictions.

The standard choice for comparing competing models in Bayesian statistics is to use the Bayes factor, that is the ratio of the posterior odds to the prior odds in favour of the first model. However, the accurate calculation of the Bayes factor is very difficult in complex models which need MCMC techniques for parameter estimation such as those we examine here. Therefore, we prefer to use an alternative Bayesian model choice criterion based on the DIC_3. This is a variant of the DIC which is a within sample measure of fit introduced by Spiegelhalter et al. (2002) commonly used in Bayesian analysis.

Defining the deviance of a model with parameters θ as $D(\theta) = -2 \log f(\mathbf{y}|\theta)$, where $\mathbf{y}$ are the data, then $DIC = 2\overline{D(\theta)} - D(\bar{\theta})$ where $\bar{\theta}$ represent some mean posterior parameter estimates. However, the DIC is well known to possess a number of stability problems in certain cases such as random effects models and mixture models (see Celeux et al. 2006). In particular, we can note here that the representation we use for the parameters of the inefficiency term is a type of random effects model in the cases where we include an unobserved heterogeneity term. Furthermore and more recently, Li et al. (2012) also remark on the lack of robustness of the original *DIC* in models with data augmentation such as those we examine here. For such cases, Celeux et al. (2006) recommend the use of the DIC_3 criterion as one of the best choices among various alternatives to the DIC. The formulation for this criterion is:

$$DIC_3 = -4E_\theta[\log f(\mathbf{y}|\theta)|\mathbf{y}] + 2\log \widehat{f}(\mathbf{y}).$$

This criterion is based on the expected deviance and an estimate of the predictive density $\hat{f}(\cdot)$ which are both easy and stable to calculate from the MCMC output provided by WinBugs.

We also compare the models in terms of their predictive performance. In order to do this, we calculate the LPS and the MSE of predictions. The LPS is a proper scoring rule developed in Good (1952) that assesses the post-sample behaviour of the models associated with the Kullback-Leibler divergence between the actual sampling density and the predictive density (see Griffin and Steel 2004; Ferreira and Steel 2007, for previous applications of LPS in stochastic frontier models).[3] In general, LPS examines how well a model performs when its implied predictive distribution is compared with observations not used in the inference sample. The procedure consists of partitioning the sample into two sets. The first, is a training data set used to fit the model and the second is a prediction set used to evaluate the predictive performance of the first set. In our implementation for the panel data models, the training data set contains the observations up to the penultimate time period at which data are observed for each firm. Then, if t_i represents the index of the last time point when data are observed for firm i, the predictive set contains the set of observations y_{1,t_1} to y_{k,t_k} for the k firms in the sample. The average of the log predictive density functions evaluated at observed out-of-sample values are calculated and the formulation is the following:

$$LPS = \frac{-1}{k}\sum_{i=1}^{k} \log f(y_{i,t_i}|\text{previous data})$$

Finally, the calculation of the predictive MSE involves again the partition of the sample into two parts as earlier. The models are fitted using the training sample and their

[3] More details on this criterion and an approximate lower bound for the LPS are described in Fernandez et al. (2001).

estimated parameters are used to predict the data for the last observation of every firm. The MSE is calculated as follows:

$$MSE = \frac{1}{k}\sum_{i=1}^{k}\left(y_{i,t_i} - E\left[(\beta' x_{i,t_i} - \overline{u}_{i,t_i})|\text{previous data }\right]\right)^2,$$

where k is the number of firms as earlier and $\overline{u}_{i,t_i}$ is the mean of the inefficiency component, which is different depending on the distribution and varies with the firm for models with heterogeneity in the inefficiency.

4 Empirical applications

In this section, we analyze two data sets, estimate the models presented in Sect. 2 and interpret the results.

4.1 Application to WHO data set

Evans et al. (2000) estimated the technical efficiency of 191 countries in the provision of health by using a classical fixed effects stochastic frontier model for an unbalanced panel. The original data set covers 5 years from 1993 to 1997 and the production function model proposed was the following:

$$\begin{aligned}\ln(DALE_{it}) = \alpha_i + \beta_1 \ln(HExp_{it}) + \beta_2 \ln(Educ_{it}) \\ + \beta_3 \frac{1}{2}\ln^2(Educ_{it}) + v_{it},\end{aligned}$$

where $DALE$ is the disability adjusted life expectancy, a measure that considers mortality and illness and represents health output. Input amounts are measured by $HExp$ and $Educ$, which are health expenditure and the average years of education, respectively.

Their results were reported by the WHO and suffered from several criticisms since the authors did not consider the effects of heterogeneity in their study, even though the sample included countries with very different characteristics such as Switzerland, China, or Zimbabwe. This led to unexpected country health system performance rankings.

Greene (2004) proposed to capture differences among countries in this sample by including eight exogenous variables: *Tropics*, *PopDen*, *GEff*, *Voice*, *Gini*, *GDP*, *PubFin*, and *OECD*. *Tropics* is a binary variable that takes the value 1 if the country is located in the tropic and 0 otherwise. This is out of the control of the countries and distinguishes them by the type of diseases found in this region. *PopDen* is the country population density, which may capture effects of dispersion but also congestion in the provision of health. These two variables are characteristics of the health provision in each country and then they are included as covariates in the production function following Greene (2004). Regarding the other variables, *GEff* is an indicator of government efficiency; *Voice* is a measure of political democratization and freedom; *Gini* is the income inequality coefficient; *GDP* is the per capita country gross domestic product; *PubFin* is the proportion of health care financed with public resources, and *OECD* is a binary variable that takes the value 1 if the country belongs to the organization and 0 otherwise. These variables are policy related and more likely to be drivers of the efficiency in the sense that income, inequality and government characteristics may affect the way health services are managed. However, in this field there is no theory on where these variables should be placed at (see Greene 2004).[4]

For this application the general model is:

$$\begin{aligned}\ln(DALE_{it}) = \alpha + \beta_1 \ln(HExp_{it}) + \beta_2 \ln(Educ_{it}) \\ + \beta_3 \frac{1}{2}\ln^2(Educ_{it}) + \beta_4 Tropics_i \\ + \beta_5 \ln(PopDen_i) + \mathbf{z_i}\boldsymbol{\delta} + v_{it} - u_{it}; v_{it} \sim N(0,\sigma_v^2) \\ u_{it} \sim N^+(\mu \cdot exp(\mathbf{z_i}\gamma I_1 + \tau_i I_2), \sigma_u^2 \\ \cdot (exp(\mathbf{z_i}\gamma I_3 + \tau_i I_4))^2).\end{aligned} \tag{7}$$

We begin our analysis by estimating models where unobserved heterogeneity is not considered, that is I_2,$I_4 = 0$. Model I is the heterogeneity free base model where I_1,I_3, and $\boldsymbol{\delta}$ are also equal to zero. Model II includes the covariates in the frontier as technology heterogeneity variables but not in the inefficiency (I_1,$I_3 = 0$). Models III to V consider observed heterogeneity in the inefficiency distribution and not in the production function ($\boldsymbol{\delta} = \mathbf{0}$). In particular, Model III does it only through the location parameter, that is $I_1 = 1$ and $I_3 = 0$. Model IV includes the observed covariates through the scale parameter ($I_1 = 0$, $I_3 = 1$). Finally, Model V preserves the scaling property since both parameters of the inefficiency distribution includes the same covariates and coefficients (I_1,$I_3 = 1$).

Table 1 reports the estimation results. They show that models considering observed heterogeneity improve from the base model in terms of fit and predictive performance. In particular, models including heterogeneity in the inefficiency distribution exhibit the lowest values for the three model comparison criteria. This suggests that covariates in $\mathbf{z_i}$ are inefficiency related. Regarding the estimated frontier coefficients, we observe decreasing returns to scale in health provision for all models and countries. This implies that efforts of countries in terms of increasing health expenditure or education are reflected in less than proportional life expectancy improvements. Results for the

[4] After performing some tests Greene (2004) chose a model that includes Gini and GDP in the inefficiency and the rest of covariates in the production function.

Table 1 Posterior means of the parameter distributions

Parameters	Model I	Model II	Model III	Model IV	Model V
Production function					
α	3.5741	3.4786	3.8447	3.7107	3.7688
β_1	0.0613	0.0255	0.0239	0.0639	0.0413
β_2	0.2262	0.2359	0.2497	0.2483	0.1601
β_3	−0.0396	−0.0488	−0.0612	−0.0462	−0.0327
β_4	−0.0168	−0.0143	−0.0054	−0.0433	−0.0088
β_5	0.0009	−0.0023	0.0005	0.0013	0.0009
$\delta_1(Gini)$		−0.1469			
$\delta_2(\ln GDP)$		0.0617			
$\delta_3(GEff)$		−0.0142			
$\delta_4(Voice)$		0.0178			
$\delta_5(OECD)$		−0.0261			
$\delta_6(\ln PubFin)$		−0.0364			
Inefficiency					
$\gamma_1(Gini)$			3.7799	8.2122	5.0537
$\gamma_2(\ln GDP)$			−0.2661	−0.2798	−0.6618
$\gamma_3(GEff)$			−0.0431	−0.1324	−0.0539
$\gamma_4(Voice)$			0.0774	0.1592	0.0300
$\gamma_5(OECD)$			−0.0923	−3.3892	−1.0498
$\gamma_6(\ln PubFin)$			0.0618	0.3762	0.0760
μ	−1.5837	−1.4106	−0.6204	−1.4226	−0.3720
σ_u^2	0.2382	0.2141	0.4056	0.0537	0.0581
Pred. eff. mean	0.8779	0.8773	0.9076	0.7853	0.9144
Pred. eff. SD	0.1037	0.1033	0.1375	0.0813	0.0715
DIC_3	−2,517.2820	−2,809.5814	−3,015.7270	−2,989.3070	−3,094.4026
LPS	−122.8900	−130.4520	−180.5074	−169.2150	0.0869
MSE	0.1387	0.1051	0.1028	0.0933	−185.9830

inefficiency covariates suggest that higher equality, income, government efficiency or pertaining to the OECD increase the efficiency of health provision. However, higher levels of democracy and public finance of health services lead to lower efficiency.

Focusing on models III to V which are those including inefficiency heterogeneity, we observe that the best fit and predictive performance is obtained by the scaling property model (Model V). Results for the predictive efficiency distribution suggest that including covariates in the location parameter of the inefficiency increases its mean, while including them in the scale parameter decreases its dispersion. In particular, the scaling property model which includes covariates in both parameters of the one-sided error distribution presents the highest mean and the lowest dispersion of the predictive efficiency distribution among all models.

The most clarifying insights come from the efficiency rankings since they allow country comparisons. Figure 1 shows efficiency rankings' scatter plots comparing the base model against the other four models. For Model II, which includes the covariates in the frontier, most countries preserve a similar position except for small changes in the middle rankings. Spearman's rank correlation with the base model is 0.92. In contrast, models III to V differ widely from the base model in the top and middle positions and the Spearman's rank correlations with the base model are 0.76, 0.77 and 0.75, respectively.[5] However, badly performing countries are always roughly the same regardless of the model used. This latter group is composed mainly of central African countries (e.g. Zambia, Botswana, Zimbabwe), which share some characteristics related to low income, tropical diseases, etc.

In order to observe in detail the changes that occur in the top ranked countries under the different models, Table 2 shows the top 20 most efficient countries under all five models. Although there are differences, the ranking is quite stable when we consider the first two models. They include countries such as Oman, Yemen and Cape Verde and other developing countries from Middle East, Asia, North of

[5] Among models with inefficiency heterogeneity, rank correlation is very high (0.99).

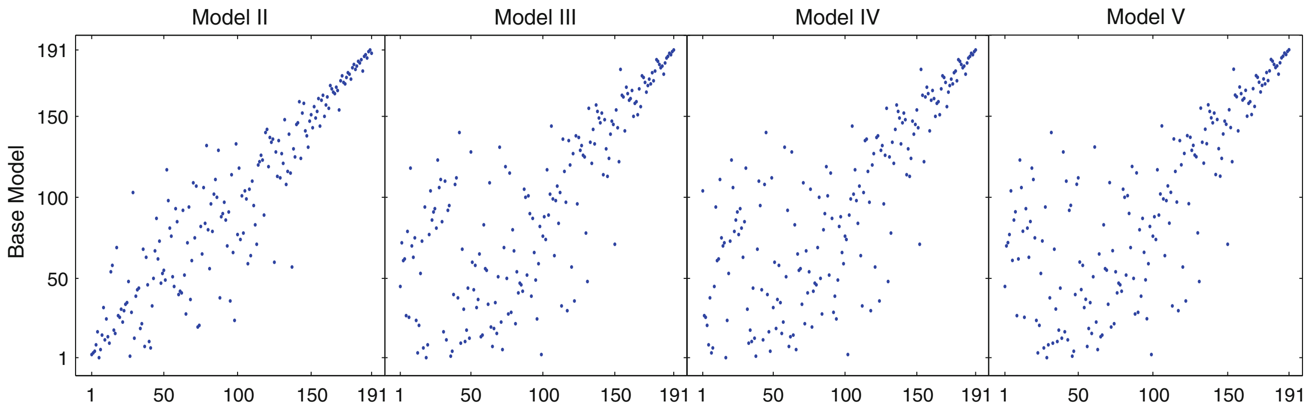


Fig. 1 Efficiency rankings - Base model vs. heterogeneity models

Table 2 Top 20 most efficient countries

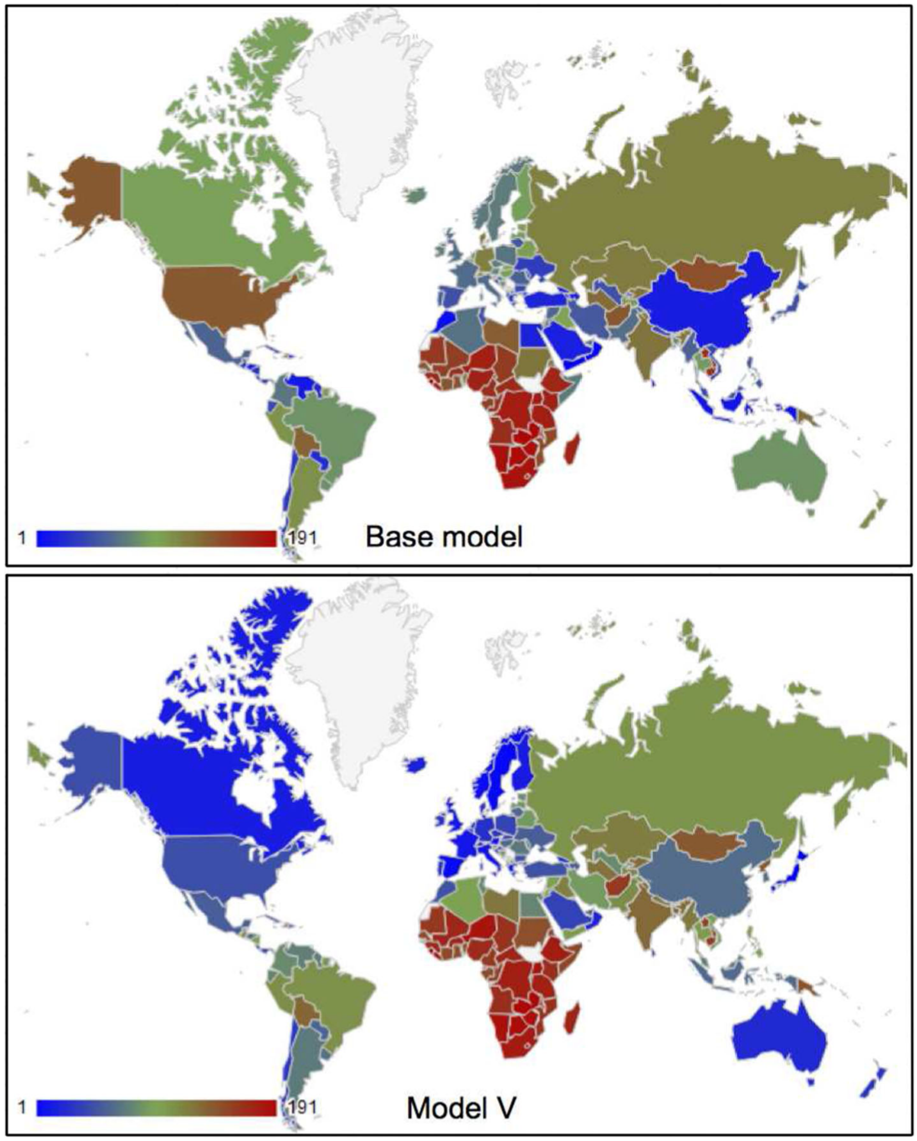
Model I	Model II	Model III	Model IV	Model V
1. Oman	1. Yemen	1. Japan	1. Luxembourg	1. Japan
2. Solomon Islands	2. Jamaica	2. Sweden	2. Spain	2. Norway
3. Yemen	3. Morocco	3. Italy	3. Greece	3. Sweden
4. Jamaica	4. Armenia	4. France	4. Malta	4. Austria
5. Morocco	5. Turkey	5. Spain	5. Armenia	5. Luxembourg
6. Cape Verde	6. Oman	6. Iceland	6. Cyprus	6. Italy
7. Georgia	7. Cape Verde	7. Greece	7. Jamaica	7. Belgium
8. Indonesia	8. Honduras	8. Germany	8. Georgia	8. Finland
9. Armenia	9. Cuba	9. Norway	9. Japan	9. Spain
10. Sri Lanka	10. China	10. United Kingdom	10. Slovakia	10. France
11. Venezuela	11. Nicaragua	11. Ireland	11. Italy	11. Denmark
12. China	12. El Salvador	12. Singapore	12. France	12. Switzerland
13. Saudi Arabia	13. Sri Lanka	13. Jamaica	13. New Zealand	13. Iceland
14. El Salvador	14. Moldova	14. Malta	14. Ireland	14. Greece
15. Honduras	15. Mexico	15. Portugal	15. Norway	15. Canada
16. Azerbaijan	16. Costa Rica	16. Czech Republic	16. Sweden	16. Netherlands
17. Turkey	17. Azerbaijan	17. Georgia	17. Oman	17. United Kingdom
18. Costa Rica	18. Colombia	18. Slovakia	18. Singapore	18. Australia
19. Dominican Rep.	19. Spain	19. Oman	19. Portugal	19. Germany
20. Egypt	20. Greece	20. Armenia	20. Czech Republic	20. New Zealand

Africa and Latin America in the top positions. However, this changes completely when observed heterogeneity affects the inefficiency. In models III to V, developed countries rank in the first positions, as might be intuitively expected, and for the scaling property model all top 20 countries are from this group. Differences are important compared to the base model. For example, Japan, Norway and Sweden which are the top 3 countries under Model V, rank in positions 45, 70 and 72, respectively, under the base model.

In fact, using a scaling property model with heterogeneity in both parameters of the inefficiency distribution has an important effect over the ranking. Figure 2 shows that while most of the African countries continue to exhibit low efficiency; there is a significant change in the positions of the top and middle ranked observations. The best performing countries, in particular, the developed countries are very sensitive to the inclusion of relevant covariates such as income and inequality that distinguish them from developing countries.

The main evidence is that models that include inefficiency heterogeneity lead to important moves and shrinkages of the individual posterior efficiency distributions changing the estimated mean efficiency scores and rankings. Figure 3 shows the posterior 90 % credible intervals

Fig. 2 Heat map of efficiency rankings—Base model versus Model V



of efficiencies for some selected countries. It can be seen that when covariates affect the location parameter (Model III), the gap between the worst and the best performing countries increases, which leads to a separating effect on the posterior distributions. On the other hand, the intervals are narrower when the observed heterogeneity affects the scale parameter of the inefficiency (Model IV), which implies that estimation uncertainty diminishes. For the scaling property model (Model V) both effects are observed. This leads to less dispersion and overlapping of posterior efficiency distributions, which allow for more reliable conclusions about efficiency scores and rankings.[6]

[6] Similar results were obtained from other scaling-type models following half-normal and exponential distributions but they performed a bit worse in terms of fit and predictive performance. Results are available from authors upon request.

As mentioned previously, one of the advantages of preserving the scaling property is the decomposition of the one-sided error term into a base and a heterogeneity component. In particular, for Model V, $u_{it} = u^{*}_{it} \cdot \exp(\mathbf{z_i}\gamma)$ where $u^{*}_{it} \sim N^{+}(\mu, \sigma^2_u)$. Table 3 presents this decomposition in terms of efficiency for countries in Fig. 3. We observe that countries such as Yemen and Brazil present higher base efficiency but lower total efficiency than developed countries. This may indicate that these countries present good managerial skills in health provision but under their specific characteristics, they exploit their management abilities to a lesser extent than the developed countries. One of the countries taking great advantage of environmental characteristics is the USA, where efficiency in health provision is highly dependent in their particular attributes. These results are in line with those obtained by contrasting the base model and Model V. Other group of

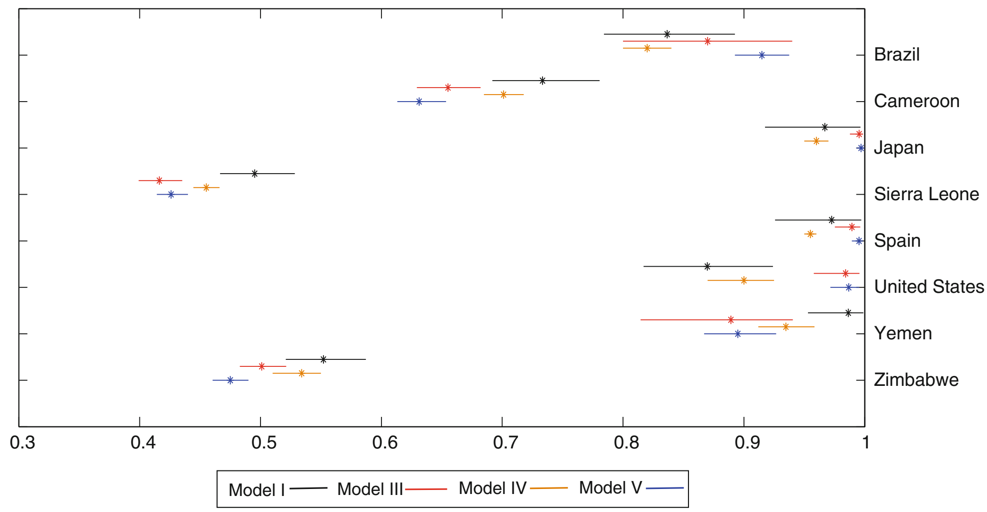


Fig. 3 90 % credible intervals of the posterior efficiency distributions for selected countries with half-normal inefficiencies

countries, mainly from Africa exhibit low base and low total efficiency. This may indicate both, poor natural managerial abilities, and inability to perform well under their relative bad conditions. Consequently, these countries present very bad performance under all models whether heterogeneity is considered or not.

Overall, we observe that observed heterogeneity variables are inefficiency related and their inclusion in the parameters of the one sided error component distribution has a large impact on the countries' efficiency ranking. Moreover, allowing observed heterogeneity to affect simultaneously both the location and scale parameters of the one-sided error distribution in a way such that the scaling property is preserved has relevant effects on shrinking and separating the distributions of posterior individual efficiencies.

4.1.1 Unobserved inefficiency heterogeneity

Results obtained above allows us to test our proposal to capture latent heterogeneity through a random parameter.

Table 3 Posterior mean of base and total efficiency for selected countries

Country	Base efficiency	Total efficiency
Brazil	0.6716	0.9149
Cameroon	0.2543	0.6313
Japan	0.6371	0.9970
Sierra Leone	0.2808	0.4260
Spain	0.6579	0.9953
United States	0.3702	0.9867
Yemen	0.7312	0.8950
Zimbabwe	0.2491	0.4750

Since previous results favor the scaling property model, we analyze unobserved heterogeneity in models that satisfy this property.

First, we estimate Model A where we assume no information about observed heterogeneity variables in $\mathbf{z_i}$. That is, we impose $I_2, I_4 = 1$ and $I_1, I_3 = 0$ in Eq. (7). Notice that these covariates are time invariant, so for this application the random parameter capturing unobserved effects is defined to be firm specific and constant over time, as well.

We propose to estimate two additional models, where observed covariates are also considered to affect inefficiency. In these cases all indicator variables in Eq. (7) are equal to 1. This allows us to analyze the efficacy of the parameter τ_i to capture information from omitted covariates and to identify those which are relevant. Model B considers the variables *Gini* and *GDP* in addition to the random parameter. These two variables capture the most relevant aspects of inequality and income distinguishing countries and were also found to be the most inefficiency related by Greene (2004) after performing a frequentist based test. Finally, we estimate Model C where τ_i is estimated along with all the covariates in $\mathbf{z_i}$.

Results are presented in Table 4. In general, we observe that all model comparison criteria improve compared to models I and II when the unobserved component is included in the inefficiency distribution. This implies that the random component captures part of the heterogeneity identified by covariates in $\mathbf{z_i}$ and therefore, it is a good alternative when no observed heterogeneity variables are available.

A second finding is that when τ_i is included simultaneously with observed variables in the inefficiency distribution, this parameter can be used as an indicator of the

Table 4 Posterior means of the parameter distributions for unobserved heterogeneity models

Parameters	Model A	Model B	Model C
Production function			
α	3.8459	3.7533	3.7322
β_1	0.0257	0.0242	0.0245
β_2	0.2121	0.3724	0.4133
β_3	−0.0361	−0.0854	−0.0994
β_4	−0.0045	−0.0031	−0.0081
β_5	−0.002	−0.0049	−0.0057
Inefficiency			
γ_1 (*Gini*)		1.9501	1.2605
γ_2(ln *GDP*)		−0.5424	−0.3633
γ_3 (*GEff*)			−0.0699
γ_4 (*Voice*)			0.0244
γ_5 (*OECD*)			−0.7456
γ_6(ln *Pubfin*)			0.0826
$\bar{\tau}$	−4.6167	−0.8029	−0.7449
σ_τ^{-2}	1.0396	2.1922	1.9758
μ	−1.6546	−1.4855	−0.3916
σ_u^2	0.0791	0.0990	0.0566
Pred. eff. mean	0.8325	0.8767	0.9145
Pred. eff. SD.	0.0904	0.0995	0.0712
DIC_3	−2957.82	−3017.61	−3085.19
LPS	−146.771	−152.95	−180.4269
MSE	0.1037	0.1014	0.0882

suitability of the observed covariates to capture inefficiency heterogeneity. In fact, it is observed that Model B, which includes only two covariates in $\mathbf{z_i}$ besides the random parameter, improves in terms of fit and predictive performance in comparison to Model A but it is not as good as Model V that include six covariates. This would mean that *Gini* and *GDP* are relevant heterogeneity variables but they are not able to capture all the inefficiency heterogeneity. On the other hand, Model C that includes all observed covariates plus the parameter τ_i performs a little worse than Model V (see model comparison criteria in Tables 1, 4). This would imply that the six covariates in $\mathbf{z_i}$ capture all the relevant inefficiency heterogeneity.

These conclusions are the same when we compare the posterior predictive efficiencies of models including the unobserved component to those of models I and V (see Fig. 4). It can be seen that the predictive efficiency distribution becomes less disperse to the extent inefficiency heterogeneity is better identified by the random parameter, observed covariates or a combination of both. Also, it is observed that the predictive efficiency distribution of Model C is very close to that of Model V, which suggests that the parameter τ_i is irrelevant when the observed covariates are able to capture most of the inefficiency heterogeneity.

4.2 Application to airlines

The airline industry is an interesting sector where performance and efficiency have been studied in the literature using parametric and non-parametric methods. Usually, production functions are employed to evaluate technical efficiency and environmental covariates are often included in the frontier as exogenous variables (see Coelli et al. 1999).

In this application we use a Cobb-Douglas cost function with an output quadratic term to evaluate economic efficiency of the airline industry. The model in (6) can be easily extended to a cost function and as in the previous application we consider individual characteristics to capture firms heterogeneity. We use a data set of 24 US domestic airlines over 15 years, from 1970 to 1984, with a

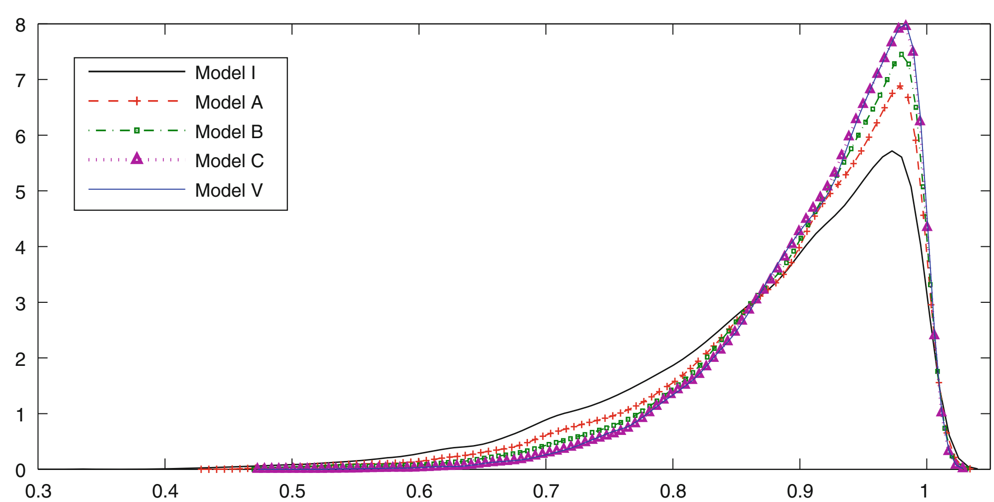


Fig. 4 Kernel densities of posterior efficiency distributions

total of 246 observations. This is a revised sample obtained from a data set used by Greene (2008).[7]

The general model for this application is the following:

$$\ln C_{it} = \alpha + \beta_1 \ln Pm_{it} + \beta_2 \ln Pf_{it} + \beta_3 \ln Pl_{it} + \beta_4 \ln Pe_{it} + \beta_5 \ln(y_{it}) + \beta_6 \frac{1}{2}\ln^2(y_{it}) + \beta_7 t + \beta_8 t^2 + \mathbf{z_{it}}\boldsymbol{\delta} + v_{it} + u_{it} v_{it} \sim N(0, \sigma_v^2) u_{it} \sim N^+(\mu \cdot exp(\mathbf{z_{it}}\gamma I_1 + \tau_{it} I_2), \sigma_u^2 \cdot (exp(\mathbf{z_{it}}\gamma I_3 + \tau_{it} I_4))^2), \quad (8)$$

where C_{it} is the total cost supported by airline i at time t in the output production, and Pm_{it}, Pf_{it}, Pl_{it}, Pe_{it} are the input prices of material, fuel, labor and equipment, respectively. Cost and prices are normalized by the property price. y_{it} is the output of airline i at time t and it is an index that aggregates regular passenger, mail, charter, and other freight services. In order to capture possible technological changes over the 15 years covered by the sample we include a trend and its square into the model.

Regarding heterogeneity, $\mathbf{z_{it}}$ is a vector containing information of three observed covariates (load factor, average stage length and points served); while τ_{it} is the unobserved heterogeneity random parameter for firm i at time t. Load factor is the effective performed tonne-passenger per kilometer by the airline as a proportion of the total available tonne-passenger per kilometer. Stage length is the ratio of total performed kilometers to the total number of departures. And, points served is the number of destinations.

Variables in $\mathbf{z_{it}}$, as well as other variables of size, are commonly used in productivity and efficiency analysis of the airlines sector but their behavior as drivers of either the frontier or the inefficiency is an open issue. Coelli et al. (1999) present a review on studies using environmental variables in both cases and note that variables in $\mathbf{z_{it}}$ may be argued to have effects on costs and inefficiency.[8] In particular, airlines face high fix but low variable costs, thus we would expect airlines with high load factor to incur in lower costs to transport the same outputs than airlines with a low value for this variable. Its effect on inefficiency would also be negative since a higher load factor implies a higher capital utilization ratio. Airlines operating with high stage length would incur in lower takeoff, landing, parking and other airport costs. Also, they are expected to be more efficient since their aircrafts are being productive for longer time periods. Finally, points served are expected to have a positive effect on total costs since a larger network requires more resources but also more managerial skills which may result on higher or lower inefficiency depending on the routes optimization carried out.

Similarly to the WHO application, first we estimate models not considering unobserved inefficiency heterogeneity (I_2 , $I_4 = 0$), and then we analyze the effects of the unobserved component in a subsection. The base model (Model I) does not consider any type of heterogeneity; therefore, $\boldsymbol{\delta} = \mathbf{0}$ and I_1,$I_3 = 0$. Model II considers only frontier heterogeneity by including the observed covariates in the cost function. Models III to V consider covariates in $\mathbf{z_{it}}$ as determinants of the inefficiency and include them in the location, scale or both parameters of the one-sided error distribution, respectively.[9]

Table 5 reports the estimation results. We observe that Model II which includes the observed heterogeneity variables in the cost function present the best fit and predictive performance, suggesting variables in $\mathbf{z_{it}}$ to be drivers of the frontier.[10] Nevertheless, models with inefficiency covariates also improve results from the base model. Among these models, the one that includes covariates in both parameters of the inefficiency distribution and preserves the scaling property (Model V) presents the best values in terms of DIC_3 and LPS. However, differences are narrower than in the previous application, in particular compared to Model III, which exhibits the lowest value of MSE. As in the WHO application, models including observed heterogeneity in the scale parameter of the inefficiency exhibit lower dispersion of the predictive efficiency distribution. Regarding the estimated coefficients, we identify increasing returns to scale and expected effects of covariates on costs and inefficiency as discussed above. From the estimation results obtained for Model II we conclude that load factor and stage length affect negatively costs, while the network size has the opposite effect. Overall, considering heterogeneity has effects on the estimations of posterior mean efficiencies with respect to the base model, as we observe in Fig. 5.

4.2.1 *Unobserved inefficiency heterogeneity*

Since the observed covariates are related to frontier heterogeneity, our benchmark is Model II. We assume that it

[7] The original data set includes 256 observations, ten years of observations for an extra airline company. We excluded this firm since we do not have data for the exogenous variables of this airline.

[8] Coelli et al. (1999) evaluate both alternatives for a technical efficiency analysis and conclude statistically in favor of a model including them in the inefficiency term.

[9] For all models, monotonicity conditions were found to be not satisfied because of negative signs obtained for prices coefficients. This result was also obtained by Greene (2008). Therefore, we impose regularity conditions by requiring the cost function to have positive elasticities on prices ($\partial c_{it}/\partial p_{it} > 0$). We follow the procedure described in Griffin and Steel (2007) by restricting coefficients β_1 to β_4 to be positive through truncated normal prior distributions for these parameters.

[10] In fact, most of the efficiency studies applied to airlines have treated size and network environment variables as frontier drivers (see Coelli et al. 1999).

Table 5 Posterior means of the parameter distributions

Parameters	Model I	Model II	Model III	Model IV	Model V
Cost function					
α	1.7774	2.4628	1.3411	0.7570	1.6229
$\beta_1(\ln Pm)$	0.3595	0.1483	0.2699	0.1119	0.2890
$\beta_2(\ln Pf)$	0.1755	0.1952	0.1970	0.1948	0.2243
$\beta_3(\ln Pl)$	0.2361	0.4844	0.2992	0.4496	0.2170
$\beta_4(\ln Pe)$	0.0520	0.1890	0.1161	0.1365	0.1372
$\beta_5(\ln y)$	0.9421	0.9598	0.8941	0.8606	0.9654
$\beta_6(\frac{1}{2}\ln^2 y)$	0.0884	0.0385	0.0424	0.0382	0.0442
$\beta_7(t)$	−0.0286	−0.0379	−0.0197	−0.0228	−0.0368
$\beta_8(t^2)$	0.0006	0.0005	−0.0007	−0.0007	0.0001
$\delta_1(Load)$	–	−0.9135	–	–	–
$\delta_2(\ln Stage)$	–	−0.2173	–	–	–
$\delta_3(\ln Points)$	–	0.1498	–	–	–
Inefficiency					
$\gamma_1(Load)$	–	–	−0.6252	−0.8716	−0.8045
$\gamma_2(\ln Stage)$	–	–	−0.2061	−0.3663	−0.4924
$\gamma_3(\ln Points)$	–	–	0.2520	0.2504	0.3058
μ	0.0214	0.2092	0.3513	0.2840	0.3514
σ_u^2	0.1843	0.1245	0.1520	0.1227	0.1272
Pred. eff. mean	0.8687	0.7862	0.6812	0.7535	0.7095
Pred. eff. s.d.	0.1007	0.1275	0.1769	0.0991	0.0873
DIC_3	−605.2869	−815.3940	−697.8570	−674.0670	−704.8459
LPS	−13.7340	−33.6520	−19.7923	−18.6471	−21.6690
MSE	0.0257	0.0096	0.0134	0.0195	0.0178

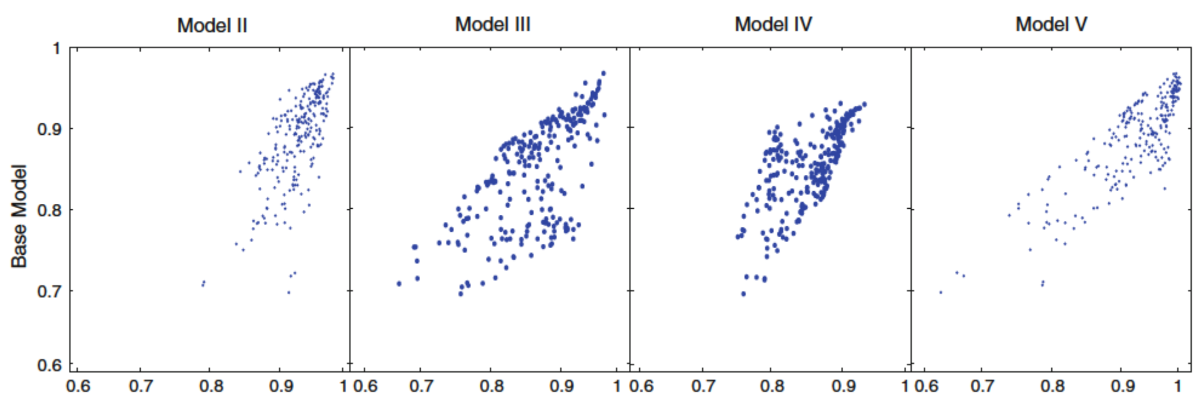


Fig. 5 Posterior mean efficiencies—base model versus heterogeneity models

may exist inefficiency heterogeneity in the sector related to other factors not considered by variables in $\mathbf{z}_{it}$. Therefore, we evaluate the inclusion of a time-varying random parameter in the distribution of the inefficiency when it is specified in the location, scale, or both parameters of the one-sided error distribution.

Results for the three estimated models that includes τ_{it} in the location (Model A), scale (Model B) and both (Model C) parameters of the inefficiency distribution are presented in Table 6. It can be observed that all three models improve their fit and predictive performance in comparison to Model II. In particular, models A and C exhibit the best values for the three criteria. However, when the random parameter is included in the scale parameter of the inefficiency distribution (models B and C), a decrease in the dispersion of the predictive efficiency distribution is observed.

The effects on the individual posterior efficiencies using the random parameter are similar to those found in the previous application using observed covariates. That is, when τ_{it} is considered in the location parameter of the one-sided error distribution, the posterior efficiencies of different airlines are more separated from each other, and when it is included in the scale parameter, we observe a shrinking effect and consequently a decrease in the dispersion of the posterior efficiency distributions. Figure 6 shows these effects for some selected airlines. We can observe that Model C, which includes the random parameter in both parameters of the inefficiency distribution and satisfies the scaling property, separates and shrinks the individual posterior efficiency distributions providing both more reliable efficiency scores and rankings.

Table 6 Posterior means of the parameter distributions for unobserved heterogeneity models

Parameters	Model A	Model B	Model C
Cost function			
α	1.6660	0.4699	2.9144
$\beta_1(\ln Pm)$	0.4359	0.3034	0.1534
$\beta_2(\ln Pf)$	0.1937	0.1970	0.2373
$\beta_3(\ln Pl)$	0.1553	0.2362	0.3300
$\beta_4(\ln Pe)$	0.1469	0.1541	0.2036
$\beta_5(\ln y)$	0.8707	0.8782	0.9761
$\beta_6(\frac{1}{2}\ln^2 y)$	0.0447	0.0264	0.0431
$\beta_7(t)$	−0.0323	−0.0127	−0.0270
$\beta_8(t^2)$	0.0007	−0.0008	−0.0006
$\delta_1(Load)$	−1.0958	−1.1420	−0.8560
$\delta_2(\ln Stage)$	−0.2472	−0.2351	−0.2047
$\delta_3(\ln Points)$	0.1063	0.0705	0.1354
Inefficiency			
$\bar{\tau}$	−3.4905	−4.2213	−3.5143
σ_τ^{-2}	1.7290	0.8951	1.2549
μ	0.6105	0.3428	0.3206
σ_u^2	0.1047	0.0519	0.0757
Pred. eff. mean	0.7739	0.8349	0.7969
Pred. eff. SD	0.0889	0.0187	0.0469
DIC_3	−971.7110	−938.8550	−984.3692
LPS	−40.5279	−36.7801	−39.6470
MSE	0.0089	0.0092	0.0086

Preserving the scaling property makes it possible to decompose inefficiency for Model C. In this case, $u_{it} = u_{it}^* \cdot \exp(\tau_{it})$ where $u_{it}^* \sim N^+(\mu, \sigma_u^2)$. Table 7 exhibits the decomposition in terms of efficiency for the airlines plotted above. The difference between the base and total efficiency allows us to distinguish the way unobserved firm effects are handled by airlines managers. For instance, airline 12 presents lower base efficiency but higher total efficiency than airline 17, suggesting that the former handles their specific characteristics better.

Finally, using the results of Model C, in Fig. 7 we plot the probabilities of being the most efficient airline in the sample period for some selected firms. This can be easy

Table 7 Posterior mean of base and total efficiency for selected airlines

Airline ID	Base efficiency	Total efficiency
1	0.4837	0.8245
2	0.3052	0.7669
5	0.4017	0.7614
8	0.6238	0.8092
12	0.3571	0.8970
17	0.5466	0.7194
18	0.5824	0.8352
19	0.3920	0.7317

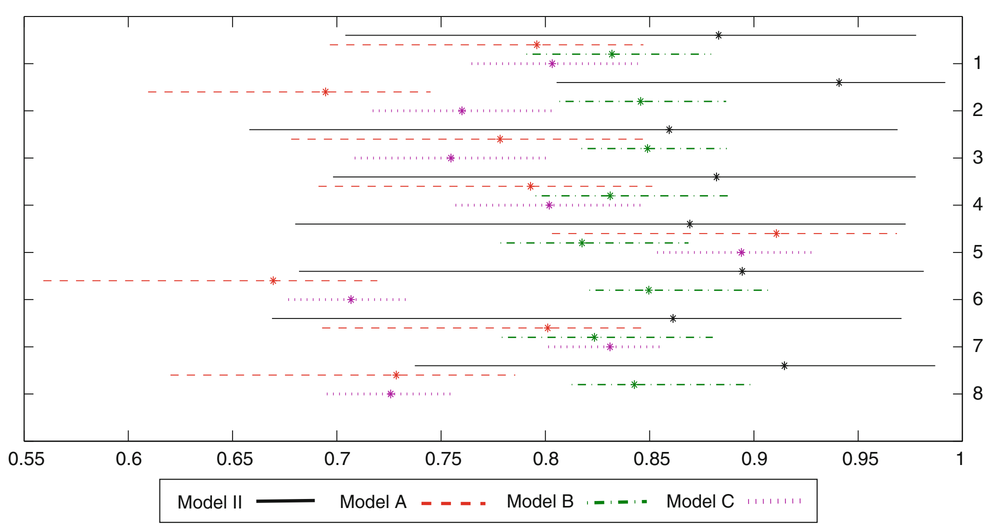


Fig. 6 90 % credible intervals of the posterior efficiency distributions for selected airlines

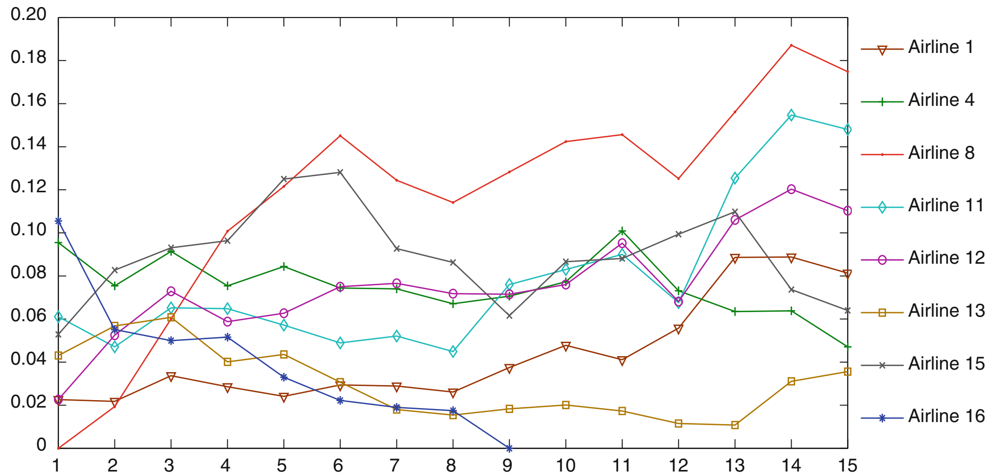


Fig. 7 Probability of being the most efficient firm in the sample period

calculated in the Bayesian context from the posterior individual distributions of efficiencies and might be very useful in empirical studies. We observe that for the last 10 years of the sample period, airline 8 is the most likely to be the benchmark firm. Also, it is possible to see improvements and declines in the airlines' performance along time. For instance, airline 11 presents a high relative improvement of its performance especially in the last 3 years, while airline 16 starts being the most likely benchmark firm and decreases very fast its probability up to being zero in year 9.

Summing up, the performance indicators suggest that firm characteristics such as the distance between destinations, the capacity offered, and the size of the network differentiate the airlines in terms of the cost frontier they face. However, there is still latent inefficiency heterogeneity related to unobserved factors. This is captured through a time varying random parameter that improves fit and predictive performance. The way this parameter is included in the inefficiency has different effects in terms of separating and shrinking the individual posterior efficiency distributions. The most desirable effects are obtained when the unobserved heterogeneity component is included both in the location and scale parameters of the inefficiency distribution in models that satisfy the scaling property.

5 Conclusions and extensions

In stochastic frontier analysis the inefficiency component may be erroneously estimated when firm characteristics are not taken into account. These firm characteristics induce heterogeneity that might result in different firm frontiers, or may have an impact directly on the inefficiencies. This issue has been widely studied before. However, unobserved inefficiency heterogeneity has been little explored.

In this work we have put forward the modeling of heterogeneity in a Bayesian context by capturing both the observed and unobserved heterogeneity in the inefficiency component distribution. We have proposed to capture latent heterogeneity through a random parameter which can be allowed to be time-varying depending on the application. Also, the effects of including both types of heterogeneity in different parameters of a truncated normal distributed inefficiency were studied. The models were fitted to two data sets previously studied only in the frequentist context and the results were compared to those obtained with models that ignore heterogeneity or include it in the frontier.

Our findings suggest that unobserved inefficiency heterogeneity can be properly captured by a random parameter. Models including this parameter whether alone or simultaneously with observed covariates improve in terms of fit and predictive performance as long as latent heterogeneity remains unidentified. In this sense, it can be used to distinguish unobserved heterogeneity from inefficiency and to validate the suitability of observed covariates to capture it.

Differences in efficiency rankings and mean scores were observed when inefficiency heterogeneity was included in different parameters of the one-sided error distribution. This was found to be related to effects in the posterior efficiency distributions. In particular, considering firms' heterogeneity in the location parameter of the inefficiency has an effect on separating the firm specific posterior efficiency distributions from each other, which leads to more reliable rankings. On the other hand, when heterogeneity affects only the scale parameter of the inefficiency, an important shrinking effect is observed on the individual posterior efficiency distributions. This results in less uncertainty around mean individual efficiency scores. Finally, including the heterogeneity in both parameters of the inefficiency distribution in models that preserve the scaling property leads to both separating and shrinking effects. This allows less overlapping of the posterior efficiency distributions and provide both more reliable efficiency scores and rankings. These results are consistent whether we use observed covariates or our proposal to model unobserved heterogeneity.

Preserving the scaling property was also found to lead to better fit and predictive performance indicators. Models with this property were extended to the Bayesian context and can be used with our proposal to capture unobserved inefficiency heterogeneity. This allows to decompose inefficiency into a base component measuring natural managerial skills and other measuring the effect of latent factors causing unobserved heterogeneity.

In this paper, we propose a intuitive procedure to capture unobserved inefficiency heterogeneity that can be easily extended in the future to different specifications and distributions of the one-sided error.

Acknowledgments The authors would like to thank Mark Steel and Jim Griffin for their comments and suggestions as well as the participants of the 33rd National Congress on Statistics and Operations Research and the Permanent Seminar on Efficiency and Productivity of Universidad de Oviedo. Financial support from the Spanish Ministry of Education and Science, research projects ECO2012-3401, MTM2010-17323 and SEJ2007-64500 is also gratefully acknowledged.

References

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

Alvarez A, Amsler C, Orea L, Schmidt P (2006) Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. J Prod Anal 25:201–212

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

Battese G, Coelli T (1995) A model for technical inefficiency effects in a stochastic frontier production model for panel data. Empir Econ 20:325–332

Caudill S, Ford J (1993) Biases in frontier estimation due to heteroscedasticity. Econ Lett 41:17–20

Caudill S, Ford J, Gropper D (1995) Frontier estimation and firm-specific inefficiency measures in the presence of heteroskedasticity. J Bus Econom Stat 13:105–111

Celeux G, Forbes F, Robert C, Titterington D (2006) Deviance information criteria for missing data models. Bayesian Anal 4:651–674

Coelli T, Perelman S, Romano E (1999) Accounting for environmental influences in stochastic frontier models: with application to international airlines. J Prod Anal 11:251–273

Evans D, Tandon A, Murray C, Lauer J (2000) The comparative efficiency of national health systems in producing health: an analysis of 191 countries. Discussion paper no. 29, World Health Organization, EIP/GPE/EQC

Fernandez C, Ley E, Steel M (2001) Benchmark priors for Bayesian model averaging. J Econom 100:381–427

Ferreira J, Steel M (2007) Model comparison of coordinate-free multivariate skewed distributions with an application to stochastic frontiers. J Econom 137:641–673

Good I (1952) Rational decisions. J R Stat Soc B 14:107–114

Greene W (1990) A gamma-distributed stochastic frontier model. J Econom 46:141–164

Greene W (2004) Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems. Health Econ 13:959–980

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

Greene W (2008) The econometric approach to efficiency analysis. The measurement of productive efficiency and productivity growth, chap 2. Oxford University Press, Inc., New York, pp 959–980

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

Griffin J, Steel M (2007) Bayesian stochastic frontier analysis using WinBUGS. J Prod Anal 27:163–176

Griffin J, Steel M (2008) Flexible mixture modelling of stochastic frontiers. J Prod Anal 29:33–50

Hadri K (1999) Estimation of a doubly heteroscedastic stochastic frontier cost function. J Prod Anal 17:359–363

Hadri K, Guermat C, Whittaker J (2003a) Estimating farm efficiency in the presence of double heteroscedasticity using panel data. J Appl Econ 2:255–268

Hadri K, Guermat C, Whittaker J (2003b) Estimation of technical inefficiency effects using panel data and doubly heteroscedastic stochastic production frontiers. Empir Econ 28:203–222

Huang H (2004) Estimation of technical inefficiencies with heterogeneous technologies. J Prod Anal 21:277–296

Huang H, Liu J (1994) Estimation of a non-neutral stochastic frontier production function. J Prod Anal 5:171–180

Koop G, Osiewalski J, Steel M (1997) Bayesian efficiency analysis through individual effects: hospital cost frontiers. J Econom 76:77–106

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

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

Kumbhakar S, Lovell C (2000) Stochastic frontier analysis. Cambridge University Press, New York

Li Y, Zeng T, Yu J (2012) Robust deviance information criterion for latent variable models. Singapore Management University, Research Collection School of Economics (Open Access). Paper 1403

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

Reifschnieder D, Stevenson R (1991) Systematic departures from the frontier: a framework for the analysis of firm inefficiency. Int Econ Rev 32:715–723

Richardson S (2002) Discussion of spiegelhalter et al. J R Stat Soc B 64:626–627

Simar L, Lovell C, van den Eeckaut P (1994). Stochastic frontiers incorporating exogenous influences on efficiency. Discussion paper no. 9403, Institut de Statistique, Universit Catholique de Louvain

Spiegelhalter D, Best N, Carlin B, van der Linde A (2002) Bayesian measures of model complexity and fit. J R Stat Soc 64(4):583–639

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

Tsionas E (2002) Stochastic frontier models with random coefficients. J Appl Econom 17:127–147 van den Broeck J, Koop G, Osiewalski J, Steel M (1994) Stochastic frontier models: a bayesian perspective. J Econom 61:273–303

Wang H (2002) Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. J Prod Anal 18:241–253

Wang H, Schmidt P (2002) One step and two step estimation of the effects of exogenous variables on technical efficiency levels. J Prod Anal 18:129–144

WHO (2000) Health systems: improving performance. The World Health report. World Health Organization, Geneva