Efthymios G. Tsionas, Kien C. Tran, Panayotis G. Michaelides

Bayesian inference in threshold stochastic frontier models

Article (Accepted version) (Refereed)

Original citation: Tsionas, Efthymios G. and Tran, Kien C. and Michaelides, Panayotis G. (2017) *Bayesian inference in threshold stochastic frontier models.* Empirical Economics. ISSN 0377-7332
DOI: 10.1007/s00181-017-1364-9

Bayesian Inference in Threshold Stochastic Frontier Models

Efthymios G. Tsionas
Lancaster University Management School, LA1 4YW, UK &
Athens University of Economics and Business, Athens, Greece

Kien C. Tran*
Department of Economics
University of Lethbridge
4401 University Drive W
Lethbridge, Alberta
T1K 3M4 Canada

and

Panayotis G. Michaelides
Systemic Risk Centre
London School of Economics, London, UK
&
School of Applied Mathematics and Physics
National Technical University of Athens
Athens, Greece

Abstract

In this paper, we generalize the stochastic frontier model to allow for heterogeneous technologies and inefficiencies in a structured way that allows for learning and adapting. We propose a general model and various special cases, organized around the idea that there is switching or transition from one technology to the other(s), and construct threshold stochastic frontier models. We suggest Bayesian inferences for the general model proposed here and its special cases using Gibbs sampling with data augmentation. The new techniques are applied, with very satisfactory results, to a panel of world production functions using, as switching or transition variables, human capital, age of capital stock (representing input quality), as well as a time trend to capture structural switching.

JEL Codes: C11, C13.

Keywords: Stochastic frontier, regime switching, efficiency measurement, Bayesian inference, Markov Chain Monte Carlo.

*Corresponding author. Emails: m.tsionas@lancaster.ac.uk, kien.tran@uleth.ca and pmichael@central.ntua.gr. A first version of this paper was presented at the CEA/CAES joint Annual Meetings, May 25-28, 2006 in Montreal, Canada. We would like to thank John Geweke for valuable comments and for pointing out some computational errors in the previous version of the paper. Also, many thanks go to Subal Kumbhakar for helpful suggestions. Finally, we would like to thank an Associate Editor and the two anonymous referees for constructive comments. The usual caveat applies.

1. Introduction

A common but *ad hoc* approach in analyzing the relationship between firm size and efficiency is to split the sample of firms into sub-group based on some measures that related to the size of the firms (see for example, Mbaga et al. 2003). However, some decision must be made concerning what is the appropriate threshold (i.e., how big must a firm be to be categorized as “large”) at which to split the sample. When this value is unknown, some method must be employed in its selection. This type of problem can be rectified by employing threshold stochastic frontier regression.

The stochastic frontier regression can also be useful in examining the heterogeneity in production across sectors of a given industry or across countries. For instant, capital stock of different age / quality / productivity and / or human capital of different quality is often used (in an aggregate manner) in production functions (see, for example, Limam and Miller 2004; Koop, Osiewaski and Steel 1999). This, effectively, creates differences in the technological possibilities and gives, in that way, rise to heterogeneity in production. Moreover, in any given sector of an industry, or more so in different countries, different technologies are used because the costs of adopting new technology (or at least better technology) differ across countries or sectors and the rates of innovation also differ substantially.

In this paper, we propose a general class of threshold stochastic frontier models that allow for sample splitting or transition, adoption and implementation of new technologies based on the class of threshold models. In particular, we model the transition to the different technology using another perspective. We allow the transition to depend on certain exogenous variables such as human capital and the age of capital stock that represent input quality, and the time trend that allows modeling structural change, i.e., the models proposed here allow for single or multiple covariates in the transition process. In other words, the paper considers a set of threshold SF models. These are essentially switching regression models in which the switching mechanism is a Probit model, and in which the regimes can differ in their coefficients, or in the variance of statistical noise, or in the variance of inefficiency.

To estimate the parameters of the proposed models, we use Bayesian inference procedures that organized around Gibbs sampling with data augmentations. The new techniques are then applied to a panel of world production functions using as switching or transition variables, human capital, the age of the capital stock (representing input quality), and a time trend to capture structural switching or structural transition.

The paper is organized as follows. Section 2 briefly reviews the standard stochastic frontier model. Section 3 proposes a general threshold stochastic frontier models and discusses various special cases via parameter restrictions. Bayesian inferences for the proposed model and its special cases are detailed in Section 4. Section 5 discusses model comparisons. Section 6 extends the models discussed in Section 3 to the multiple threshold case. An empirical application is presented in Section 7. Section 8 concludes the paper. Details on the numerical methods for Bayesian inference and marginal likelihood considerations are given in the Appendices.

2. The standard stochastic frontier model

The basic production stochastic frontier model that we use as a starting point and basis for comparison is[1]

$$y_{it} = \beta' x_{it} + v_{it} - u_{it},\ i = 1,...,n,\ t = 1,...,T,$$

where y_{it} denotes logarithm of output, x_{it} is a $k \times 1$ vector of explanatory variables (typically, logarithms of inputs like labor and capital), β is a $k \times 1$ vector of parameters, v_{it} is a two-sided random error term representing factors that are beyond the firms control, and $u_{it} \geq 0$ represents technical inefficiency. Following the standard practice in stochastic frontier literature, we assume that v_{it} are *i.i.d.* $N(0,\sigma_v^2)$ and u_{it} are *i.i.d.* $N_+(0,\sigma_u^2)$, where $N_+(0,\sigma_u^2)$ denotes the half-normal distribution with density $p(u_{it}) = (\pi\sigma_u^2)^{-1/2}\exp(-u_{it}^2/(2\sigma_u^2))$. Furthermore, we assume that (x_{it}, u_{it}, v_{it}) are mutually independent. The probability distribution function of the dependent variable is given by

$$p(y_{it} \mid x_{it}, \theta) = \frac{2}{\sigma}\varphi\left(\frac{e_{it}}{\sigma}\right)\Phi\left(-\lambda\frac{e_{it}}{\sigma}\right)$$

[1] Cost frontiers can be accommodated by reverse the sign of u_{it}.

where θ represents the parameter vector $(\beta, \sigma_v, \sigma_u)'$, $\sigma^2 = \sigma_v^2 + \sigma_u^2$, $\lambda = \sigma_u / \sigma_v$, $e_{it} = y_{it} - \beta' x_{it}$, $\varphi(.)$ and $\Phi(.)$ are probability density function and cumulative distribution function of a standard normal variate, respectively. Given this density, and the independence assumptions, it is easy to formulate the likelihood function and use the maximum likelihood method to estimate the parameters. For an excellent introduction to stochastic frontier analysis, see Greene (1993), and Kumbhakar and Lovell (2000). Bayesian analysis of the model proceeds using Markov Chain Monte Carlo methods, especially Gibbs sampling with data augmentation. More specifically, we augment the parameter vector θ with the latent technical inefficiencies u. Given a prior $p(\theta)$ for the structural parameters, and the "prior" $p(u \mid \theta) = (\pi\sigma_u^2)^{-nT/2} \exp[-u'u / (2\sigma_u^2)]$, application of Bayes' theorem yields immediately the posterior distribution

$$p(\theta, u \mid y, X) \propto \sigma_v^{-nT} \sigma_u^{-nT} \exp\left[-\frac{(y + u - \beta' X)'(y + u - \beta' X)}{2\sigma_v^2} - \frac{u^2}{2\sigma_u^2}\right] p(\theta) .$$

where y and u are $nT \times 1$ vectors and X is an $nT \times K$ matrix. Gibbs sampling requires drawing random numbers from the conditional posterior distributions and it is well known that these distributions are in standard families, so implementation of Gibbs sampling with data augmentation is straightforward, provided the prior $p(\theta)$ results in conditionally conjugate posterior distributions - this usually requires conditional prior that are special cases of the normal-gamma family.

It is well understood that for a large number of applications, assuming homogeneous technology is almost invariably an inappropriate assumption and several studies have proposed alternative models. The simplest way to introduce technological heterogeneity is to place "fixed effects" in the model by including the appropriate dummy variables in the regressor matrix X. Another way is to assume random coefficients (Tsionas (2002)), latent class frontier models (Greene (2001, 2004) and Orea and Kumbhakar, 2004), Markov switching model (Tsionas and Kumbhakar (2004)). In what follows, we propose a general model that extends and reinforces the heterogeneity issue, and can be implemented using Bayesian inference techniques and practical simulation methods.

3. General threshold stochastic frontier model

To extend the standard stochastic frontier model that allows for technology heterogeneity, we consider the following general threshold stochastic frontier model.

$$y_{it} = \beta' x_{it} + \delta' \tilde{x}_{it} I(q_{it}' \bar{\gamma} - \varepsilon_{it} \leq 0) + v_{it} - u_{it} \quad (1)$$

where $I(.)$ is the indicator function, q_{it} is a $m \times 1$ vector representing threshold variables and ε_{it} are random errors assumed to be i.i.d. $N(0,\sigma_{\varepsilon}^2)$. The x_{it}, $\tilde{x}_{it}$ and q_{it} may have common variables. A leading case is where $\tilde{x}_{it} = x_{it}$ but $\tilde{x}_{it}$ can be a strict proper subset of x_{it}. Let q_{1it} be the first element of q_{it} and q_{2it} the other elements of q_{it}. We assume that the first element of q_{it} is the constant 1 and the first element of $\bar{\gamma}$ is normalized to 1 while the others are denoted by γ, so that $q_{it}' \bar{\gamma} = q_{1it} + q_{2it}' \gamma$. We assume that the one-sided error term $u_{it} \sim i.i.d.\ N_{+}(0,\sigma_{u_0}^2)$ if $q_{1it} + q_{2it}' \gamma - \varepsilon_{it} \leq 0$, and $u_{it} \sim i.i.d.\ N_{+}(0,\sigma_{u_1}^2)$, otherwise; similarly, $v_{it} \sim i.i.d.\ N(0,\sigma_{v_0}^2)$ if $q_{1it} + q_{2it}' \gamma - \varepsilon_{it} \leq 0$, and $v_{it} \sim i.i.d.\ N(0,\sigma_{v_1}^2)$, otherwise. Furthermore, we assume that $(x_{it}, v_{it}, u_{it}, \varepsilon_{it})$ are mutually independent[2].

In model (1) observations are divided into two regimes, and this model allows for the frontier parameters to differ depending on the threshold function $q_{it}' \bar{\gamma} - \varepsilon_{it}$, and hence introduce heterogeneity in the technology component of the model. It also allows for all the frontier parameters to switch between regimes, but this is not essential for the analysis that follows. Model (1) is different from the Markov switching stochastic frontier proposed by Tsionas and Kumbhakar (2005) in that the switching variable is observable. The Markov switching model posits that regime switches are exogenous. No attempt is made to explain the reason why regime changes occur and no attempt is made to explain the timing of such changes. The threshold effect has found applications in macro and in cross-section growth regressions (see Hansen (2000) for discussion), and to the best of our knowledge, model (1) is the first application in the stochastic frontier literature.

---

[2] Other alternative distributions assumption for u_{it} such as truncated normal, exponential and Gamma are available and can be adapted for this model.

Model (1) also differs from the latent class model proposed by Greene (2001, 2004) and Orea and Kumbhakar (2004) in the sense that, in the latent class models, the regime change may be permanent implying that there is some persistence in the movement from one regime to another. In fact, the latent class models do not model the transition at all and assume instead that once adopted, a technique remains in effect forever.

4. Bayesian Inference

In principle, the parameters in model (1) or any of its special case can be estimated using a direct profiled maximum likelihood (ML). However, due to the high degree of nonlinearity, the computation of the profiled ML is numerically intensive and prohibitively expensive, especially when the sample size is large. Furthermore, if there are only a few observations in one regime, numerical problems will arise. In this paper we suggest alternative estimation algorithms based on Bayesian inference.

First, note that model (1) generalizes the simple threshold framework to allow for the threshold variable to be combination of the regressors and/or other variables such as firm's size validating the use of discontinuous variables as well as continuous variables for sample splitting. Second, various models can be deduced from model (1) via various parameters restrictions. For instant, when $\sigma_{u_o}^2 = \sigma_{u_1}^2$ model (1) reduces to the Latent Class (LC) model of Greene (2001, 2004). For convenience and later analysis, we will denote this model as "Model 2." When $\sigma_{\varepsilon}^2 = 0$ and q_{2it} is only a constant, model (1) reduces to a threshold stochastic frontier model with a single threshold. We call this model as "Model 3." Finally, when $\sigma_{\varepsilon}^2 = 0, \sigma_{v_0}^2 = \sigma_{v_1}^2, \sigma_{u_0}^2 = \sigma_{u_1}^2$ and q_{2it} is only a constant, model (1) collapses to the simplest threshold stochastic frontier model which we label as "Model 4."

Finally, due to the similarity in the specifications of the priors for the slopes, variance parameters, and the kernel posteriors between the main model (model 1) and various special cases, we will present the Bayesian analysis of the simplest threshold model first (model 4) and gradually extend the analysis to other models that eventually lead to our main model. In this way, our analysis provides the readers with an intuitive and logical way to conduct Bayesian inference. Finally, for purpose of discussion, we present the case where $x_{it} = \tilde{x}_{it}$.

4.1. Model 4: Simple Threshold Stochastic Frontier ($\sigma_{\varepsilon}^2 = 0$, $\sigma_{v_0}^2 = \sigma_{v_1}^2$, $\sigma_{u_0}^2 = \sigma_{u_1}^2$ and $q_{2it} = \alpha$)

Under these restrictions, our general model (1) can be conveniently rewritten as

$$y_{it} = \delta' w_{it}(\alpha) + v_{it} - u_{it}$$

where $w_{it}(\alpha) = \begin{pmatrix} x_{it} \\ x_{it} I(q_{it} \le \alpha) \end{pmatrix}$ and $\delta = \begin{pmatrix} \beta_2 \\ \beta_2 - \beta_1 \end{pmatrix}$. The density of the dependent variable y_{it} is given by

$$p(y_{it} \mid x_{it}, q_{it}, \theta) = \frac{2}{\sigma} \phi\left(\frac{e_{it}}{\sigma}\right) \Phi\left(-\lambda \frac{e_{it}}{\sigma}\right),$$

where $e_{it} = y_{it} - \delta' w_{it}(\alpha)$, $\sigma^2 = \sigma_v^2 + \sigma_u^2$, $\lambda = \sigma_u / \sigma_v$, θ represents the model parameter vector $(\delta, \sigma_v, \sigma_u)'$, ϕ and Φ represent the standard normal density and distribution function, respectively. Based on the above density, implementation of ML is easy conditional on the parameter α. Searching over the parameter value that maximizes the log-likelihood function provides estimates of all parameters[3]. To implement the Bayesian techniques, we make the following assumptions about the prior distribution. The priors of δ, σ_v and σ_u are assumed to be independent of each other, and given the nonnegative prior hyperparameters $\bar{Q}_v > 0$ and $\bar{Q}_u > 0$,

$$\delta \sim N(\bar{\delta}, \quad \bar{V}^{-1}), \ \frac{\bar{Q}_v}{\sigma_v^2} \sim \chi^2(\bar{n}_v), \ \frac{\bar{Q}_u}{\sigma_u^2} \sim \chi^2(\bar{n}_u).$$

where $\bar{n}_v, \bar{n}_u > 0$. The prior for δ is normal while the priors for the scale coefficients are inverted *gamma*. Indeed, $\frac{Q}{\sigma^2} \sim \chi^2(\nu)$ implies $\sigma^2 \sim Gam^{-1}(\frac{\nu}{2}, \frac{Q}{2})$, $\nu, Q > 0$. To be more specific, the prior for the scale parameters we adopt, imply that from a fictitious sample that is

[3] The parameter α is not different in principle from the other parameters. Asymptotic variances come from the information matrix, estimated using first or second derivatives, and these derivatives are well-defined regardless of the mechanism by which the likelihood was maximized. Hence, the asymptotic variances for the other parameters, conditional on the value of α, are not correct. Finally, it has to be noted that we maximized over α.

related to N, we get a sum of squares which is Q. The choices $N=1$ and $Q=0.01$ result in relatively "uninformative" priors. The choice $N=0$ results in a proper posterior, although the prior itself is no longer proper (however, one can set $N=0.1$). Furthermore, we know that improper prior densities can, but do not necessarily, lead to proper posterior distributions (see, e.g. Gelman (2006, p. 517)).

We leave the conditional prior of α, $p(\alpha \mid \delta, \sigma_v, \sigma_u)$ unspecified for the moment, and we assume that $(\delta, \sigma_v, \sigma_u)$ are mutually independent. Given the prior $p(\theta)$, the kernel posterior, augmented with the latent inefficiency vector u, is

$$p(\delta,\sigma_v,\sigma_u,\alpha,u \mid y,X,q) \propto$$
$$\sigma_v^{-(nT+\bar{n}_v-1)}\sigma_u^{-(nT+\bar{n}_u-1)} \exp\left[-\frac{\bar{Q}_v + (y-\delta' W(\alpha)-u)'(y-\delta' W(\alpha)-u)}{2\sigma_v^2} - \frac{\bar{Q}_u + u'u}{2\sigma_u^2} - \tfrac{1}{2}(\delta-\bar{\delta})'\bar{V}^{-1}(\delta-\bar{\delta})\right] p(\alpha)$$

where y and u are $nT \times 1$ vectors, $W(\alpha)$ is an $nT \times k$ stack matrix whose elements are $w_{it}^{'}$. The posterior conditional distributions that required for implementation of Gibbs sampling with data augmentation are as follows. The conditional posterior of the regression coefficients is

$$\delta \mid \sigma_v, \sigma_u, \alpha, u, y, W, q \sim N(\hat{\delta},\ \hat{V})$$

where

$$\hat{\delta} = [W(\alpha)' W(\alpha) + \sigma_v^2 \bar{V}^{-1}]^{-1}[W(\alpha)'(y+u) + \sigma_v^2 \bar{V}^{-1}\bar{\delta}],$$
$$\hat{V} = \sigma_v^2 [W(\alpha)' W(\alpha) + \sigma_v^2 \bar{V}^{-1}]^{-1}.$$

The conditional posterior of the two-sided error variance is

$$\frac{\bar{Q}_v + (y-\delta' W(\alpha)-u)'(y-\delta' W(\alpha)-u)}{\sigma_v^2} \mid \delta, \sigma_u, u, y, W, q \sim \chi^2(nT + \bar{n}_v)$$

The conditional posterior distribution of the one-sided error variance is

$$\frac{\bar{Q}_v + u'u}{\sigma_u^2} \mid \delta, \sigma_v, u, y, W, q \sim \chi^2(nT + \bar{n}_u)$$

The conditional posterior distribution of latent technical inefficiencies is

$$u_{it} \mid \delta, \sigma_v, \sigma_u, \alpha, y, W, q \sim N_+(-e_{it}\sigma_u^2 / (\sigma_v^2 + \sigma_u^2), \quad \sigma_v^2\sigma_u^2 / (\sigma_v^2 + \sigma_u^2))$$

These distributions are amenable to fast and efficient random number generation. The troublesome parameter in this context is α. The conditional kernel posterior distribution is

$$p(\alpha \mid \delta, \sigma_v, \sigma_u, u, y, W, q) \propto \exp\left[-\frac{(y - \delta' W(\alpha) - u)'(y - \delta' W(\alpha) - u)}{2\sigma_v^2}\right] p(\alpha)$$

Since the likelihood can be integrated analytically with respect to the latent variables u, an alternative marginalized conditional kernel posterior distribution is given by

$$p(\alpha \mid \delta, \sigma_v, \sigma_u, y, W, q) \propto \prod_{i=1}^{n}\prod_{t=1}^{T}\left\{\varphi\left(\frac{y_{it} - \delta' w_{it}(\alpha)}{\sigma}\right)\Phi\left(-\lambda\frac{y_{it} - \delta' w_{it}(\alpha)}{\sigma}\right)\right\} p(\alpha)$$

Here we employ a simple random walk Metropolis-Hastings algorithm to draw the above conditional posterior distribution, instead of a griddy Gibbs sampling, due to its easily tuned by the acceptance rate and arguably is more exact.

4.2. Model 3: $\sigma_\varepsilon^2 = 0$ and $q_{2it} = \alpha$.

Under these restrictions, this model is similar to model 4 discussed above with the exception that it relaxes the assumption that the composed errors have the same structures in both regimes. Thus, Bayesian inference for this model requires some modifications. First, the probability density of the dependent variable y_{it} is given by

$$p(y_{it} \mid x_{it}, q_{2it}, u_{it}, \theta) = \left\{\frac{2}{\sigma_0}\varphi\left(\frac{e_{it}}{\sigma_0}\right)\Phi\left(-\lambda_0\frac{e_{it}}{\sigma_0}\right)\right\}^{I(q_{it} \le \alpha)} \left\{\frac{2}{\sigma_1}\varphi\left(\frac{e_{it}}{\sigma_1}\right)\Phi\left(-\lambda_1\frac{e_{it}}{\sigma_1}\right)\right\}^{I(q_{it} > \alpha)}$$

where $\sigma_j^2 = \sigma_{vj}^2 + \sigma_{uj}^2$ and $\lambda_j = \sigma_{uj} / \sigma_{vj}$, $j = 0,1$. Second, the modification for the prior distributions are as follows.

$$\delta \sim N(\bar{\delta}, \bar{V}^{-1}), \ \frac{\bar{Q}_{vj}}{\sigma_{vj}^{2}} \sim \chi^{2}(\bar{n}_{vj}), \ \frac{\bar{Q}_{uj}}{\sigma_{uj}^{2}} \sim \chi^{2}(\bar{n}_{uj}), \ j = 0,1 \tag{3}$$

Third, let

$$X_1(\alpha) = [x_{it} : I(q_{it} \geq \alpha) = 1], \ X_2(\alpha) = [x_{it} : I(q_{it} \geq \alpha) = 0],$$

$$y_1(\alpha) = [y_{it} : I(q_{it} \geq \alpha) = 1], \ y_2(\alpha) = [y_{it} : I(q_{it} \geq \alpha) = 0],$$

$$u_1(\alpha) = [u_{it} : I(q_{it} \geq \alpha) = 1], \ u_2(\alpha) = [u_{it} : I(q_{it} \geq \alpha) = 0].$$

then kernel posterior distribution is given by

$$p(\delta, \sigma_{v0}, \sigma 1, \sigma_{u0}, \sigma 1, \alpha, u \mid y, X, q) \propto \sigma_{v0}^{-(nT+\bar{n}_{v0}-1)} \sigma_{v1}^{-(nT+\bar{n}_{v1}-1)} \sigma_{u0}^{-(nT+\bar{n}_{u}-1)} \sigma_{u1}^{-(nT+\bar{n}_{u1}-1)}$$

$$\exp\left[-\frac{\bar{Q}_{v0} + (y_0(\alpha) - \beta_0^{'} X_0(\alpha) - u_0(\alpha))^{'}(y_0(\alpha) - \beta_0^{'} X_0(\alpha) - u_0(\alpha))}{2\sigma_{v0}^{2}} - \frac{\bar{Q}_{u0} + u_0^{'} u_0}{2\sigma_{u0}^{2}}\right]$$

$$\exp\left[-\frac{\bar{Q}_{v1} + (y_1(\alpha) - \beta_1^{'} X_1(\alpha) - u_1(\alpha))^{'}(y_1(\alpha) - \beta_1^{'} X_1(\alpha) - u_1(\alpha))}{2\sigma_{v1}^{2}} - \frac{\bar{Q}_{u1} + u_1^{'} u_1}{2\sigma_{u1}^{2}} - \tfrac{1}{2}(\delta - \bar{\delta})^{'} \bar{V}^{-1}(\delta - \bar{\delta})\right] p(\alpha) \tag{4}$$

Posterior conditional distributions for implementing Gibbs sampling with data augmentation are straightforward generalizations of those corresponding previous subsection 4.1. More specifically, we obtain the following results. For $j = 0,1,$ the conditional posterior of the regression coefficients is,

$$\beta_j \mid \sigma_{v0}, \sigma_{u0}, \sigma_{v1}, \sigma_{u1}, \alpha, u, y, X, q \sim N(\hat{\beta}_j, \ \hat{V}_j) \tag{5}$$

where

$$\hat{\beta}_j = [X_j(\alpha)^{'} X_j(\alpha) + \sigma_{vj}^{2} \bar{V}_j^{-1}]^{-1} [X_j(\alpha)^{'}(y_j + u_j) + \sigma_{vj}^{2} \bar{V}_j^{-1} \bar{\beta}_j]$$

$$\hat{V}_j = \sigma_{vj}^{2} [X_j(\alpha)^{'} X_j(\alpha) + \sigma_{vj}^{2} \bar{V}_j^{-1}]^{-1}$$

The conditional posterior of the two-sided error variances are

$$\frac{\bar{Q}_{vj} + (y_j - \delta' X_j(\alpha) - u_j(\alpha))'(y_j - \delta' X_j(\alpha) - u_j(\alpha))}{\sigma_{vj}^2} \mid \delta, \sigma_{u0}, \sigma_{u1}, u, y, X, q \sim \chi^2(nT + \bar{n}_v) \qquad (6)$$

The conditional posterior distribution of the one-sided error variance is

$$\frac{\bar{Q}_{uj} + u_j' u_j}{\sigma_{uj}^2} \mid \delta, \sigma_{v0}, \sigma_{v1}, u, y, X, q \sim \chi^2(nT + \bar{n}_{uj}) \qquad (7)$$

The conditional posterior distribution of latent technical inefficiencies is

$$u_{it} \mid \delta, \sigma_v, \sigma_u, \alpha, y, X, q \sim N_+(-e_{it}\sigma_{uj}^2 / (\sigma_{vj}^2 + \sigma_{uj}^2), \quad \sigma_{vj}^2\sigma_{uj}^2 / (\sigma_{vj}^2 + \sigma_{uj}^2)) \qquad (8)$$

where $e_{it} = y_{it} - \beta_j' x_{it}(\alpha)$. Finally, the conditional kernel posterior distribution for α is

$$p(\alpha \mid \delta, \sigma_{v0}, \sigma_{u0}, \sigma_{v1}, \sigma_{u1}, u, y, X, q) \propto \prod_{i=1}^{n}\prod_{t=1}^{T}\left\{\varphi\left(\frac{y_{it} - \beta_j' x_{it}(\alpha)}{\sigma_0}\right)\Phi\left(-\lambda_0 \frac{y_{it} - \beta_j' x_{it}(\alpha)}{\sigma_0}\right)\right\}^{I(q_{it} \le \alpha)} \left\{\varphi\left(\frac{y_{it} - \beta_1' x_{it}}{\sigma_1}\right)\Phi\left(-\lambda_1 \frac{y_{it} - \beta_1' x_{it}(\alpha)}{\sigma_1}\right)\right\}^{I(q_{it} > \alpha)} p(\alpha)$$

A simple random walk Metropolis-Hastings algorithm is used with the same implementation as previous model to provide a draw from this conditional posterior distribution.

4.3. Model 2: Latent Class Model ($\sigma_{u_0}^2 = \sigma_{u_1}^2$)

For identification of γ, we normalized $\sigma_\varepsilon^2 = 1$. This is necessary because there is no information about the scaling of the regime split. Under these parameters restrictions, model (1) is the same as the Latent Class model of Greene (2001, 2004) and it can be estimated using the classical ML approach. Under the Bayesian framework, the probability density function of the dependent variable y_{it} is given by:

$$p(y_{it} \mid x_{it}, q_{2it}, u_{it}, \theta) = \left\{\frac{2}{\bar{\sigma}_0}\varphi\left(\frac{e_{it}}{\bar{\sigma}_0}\right)\Phi\left(-\bar{\lambda}_0 \frac{e_{it}}{\bar{\sigma}_0}\right)\right\}\Phi(q_{2it}'\gamma) + \left\{\frac{2}{\bar{\sigma}_1}\varphi\left(\frac{e_{it}}{\bar{\sigma}_1}\right)\Phi\left(-\bar{\lambda}_1 \frac{e_{it}}{\bar{\sigma}_1}\right)\right\}\left[1 - \Phi(q_{2it}'\gamma)\right]$$

where $e_{it} = y_{it} - \beta' x_{it} - \delta' x_{it} I(q_{1it} + q_{2it}'\gamma - \varepsilon_{it} \le 0)$, $\bar{\sigma}_j^2 = \sigma_{vj}^2 + \sigma_u^2$ and $\bar{\lambda}_j = \sigma_u / \sigma_{vj}$, $j = 0,1$. The Bayesian treatment of this model is more complicated than the previous two simpler models due to more complex structure of the threshold index. To facilitate the computation, let

$$I_{it} = \begin{cases} 1 & if \quad q_{1it} + q_{2it}'\gamma - \varepsilon_{it} \le 0, \\ 0 & if \quad q_{1it} + q_{2it}'\gamma - \varepsilon_{it} > 0 \end{cases}$$

The prior distributions for the slope and variance parameters are the same as in (3) with a small modification of the last term where $\frac{\bar{Q}_u}{\sigma_u^2} \sim \chi^2(\bar{n}_u)$. The prior of γ is $\gamma \sim N(\bar{\gamma},\ \bar{V}_\gamma)$, independently of the latent indicator variables I. The "prior" of I is already provided by the model specification as $P(I_{it} = 1) = \Phi(q_{2it}'\gamma)$ and $P(I_{it} = 0) = 1 - \Phi(q_{2it}'\gamma) = \Phi(-q_{2it}'\gamma)$. The same is true for q_{it} whose prior is simply $q_{it} \sim N(\gamma, I_{nT})$.

By augmenting the parameter vectors with latent variables u, q_1 and I, the kernel posterior is then given by

$$\begin{aligned} & p(\delta, \sigma_{v_1}, \sigma_{v_2}, \sigma_u, \gamma, u, I, q_1 \mid y, X, q_2) \propto \\ & \prod_{i=1}^{n} \prod_{t=1}^{T} (2\pi\sigma_{vI_{it}}^2)^{-1/2} (\pi\sigma_u^2)^{-1/2} \exp\left(-\frac{(y_{it} + u_{it} - \delta_{I_{it}}' x_{it})^2}{2\sigma_{vI_{it}}^2} - \frac{u_{it}^2}{2\sigma_u^2} \right) p(\delta, \sigma_{v_1}, \sigma_{v_2}, \sigma_u, \gamma, I) \end{aligned}$$

Bayesian analysis using Gibbs sampling with data augmentation is conducted as in previous case. Given the vector of latent indicators, we redefine

$$y_j = [y_{it} : I_{it} = j],\ X_j = [x_{it} : I_{it} = j],\ u_j = [u_{it} : I_{it} = j],\ j = 0,1$$

which represents a partition of the data and the latent inefficiencies in terms of the regime. Then the conditional posterior distributions for the slope parameters, two-sided and one-sided conditional variances and the latent technical inefficiency are followed similarly (with $\sigma_{u_0}^2 = \sigma_{u_1}^2$) to those given in (5)-(8), respectively.

For the parameter vector γ, due to the probit structure of the latent indicator variables, we have

$$p(\gamma \mid \delta, \sigma_{v_1}, \sigma_{v_2}, \sigma_u, \gamma, u, I, q_1, q_2, y, X) \propto \prod_{i=1}^{n}\prod_{t=1}^{T} \Phi(q_{2it}^{'}\gamma)^{I_{it}} \Phi(-q_{2it}^{'}\gamma)^{1-I_{it}} \exp\left[-\frac{1}{2}(\gamma-\bar{\gamma})^{'}\bar{V}_{\gamma}^{-1}(\gamma-\bar{\gamma})\right]$$

We have used the random walk Metropolis algorithm to generate random numbers from this distribution. Given the current state $\gamma_{(1)}$, we generate a candidate draw $\gamma \sim N(\gamma_{(1)}, h \cdot C)$, where C is the covariance matrix, and h is a tuning parameter which is set to maintain a reasonable acceptance rate, which in our case we choose to be close to 25%. The candidate is accepted with probability $\min\{1, Q(\gamma) / Q(\gamma_{(1)})\}$, where $Q(\gamma) \equiv p(\gamma \mid \delta, \sigma_{v_1}, \sigma_{v_2}, \sigma_u, \gamma, u, I, q_1, q_2, y, X)$ is the conditional posterior kernel. The overall algorithm performed quite well and convergence was fast. It should be noted here that any prior for the parameters γ could have been accommodated since the random walk Metropolis algorithm is quite general.

4.4. Main model (Model 1):

Most of the Bayesian analysis of the main model follows similarly as in model 2 with the exception that now the assumption of $\sigma_{u_0}^2 = \sigma_{u_1}^2$ is relaxed. To accommodate for this, Bayesian inference can be done similarly to those as in model 3 and no other modifications are needed for this model using Gibbs sampling.

5. Model comparison

It is important to determine whether the threshold effect is quantitatively important. Under the null hypothesis of no threshold effect, model (1) reduces to a standard stochastic frontier model (e.g. $H_0 : \beta_1 = \beta_2$) implying that the threshold parameters are not identified under the null hypothesis. Hence, the parameters of the switching equation are not identified when the two regimes are the same, and the parameters of one of the two regimes become unidentified when the parameters of the switching equation imply zero or close to zero probabilities of one of the regimes. Actually, this case arises, for instance, even in some regime switching model when the values of the corresponding intercepts are close to each other. Even though in such circumstances the standard tests might seem to be appropriate, their application for typically available samples could lead to dramatic size distortions. Hence, the usual asymptotic theory breaks down and standard tests may exhibit significant size distortions. In the context of linear models with weak instruments see Staiger and Stock (1997), and for nonlinear models estimated by GMM, see Stock and Wright (2000).

In brief, when there is a non-identified parameter under the null hypothesis, the classical tests yield misleading results, and the situation is sharply different. Hence, the properties of these tests are only asymptotic and difficult to derive. Furthermore, the finite sample performances of these tests are not well understood.

In fact, one has to apply non-standard tests. Various tests of specification that involve nuisance parameters which are not identified under the null hypothesis are proposed in the literature (see, *inter alia*, Davies (1987), Andrews and Ploberger (1994), Hansen (1996), and Anatolyev (2004)). For instance, Davies (1987) tested a simple hypothesis against a family of alternatives indexed by a one-dimensional parameter, θ when the tests' distribution is chi-squared. The results were applied to the detection of a discrete frequency component of unknown frequency in a time series. Next, Andrews and Ploberger (1994), in a seminal paper, derived asymptotically optimal tests for testing problems in which a nuisance parameter exists under the alternative hypothesis but not under the null. The paper is particularly interesting, because the problem considered is non-standard and the classical asymptotic optimality results do not apply. A weighted average power criterion is used by the authors to generate optimal tests. In the non-standard cases, which are of particular importance, new optimal tests are obtained.

Furthermore, Hansen (1996) studied the asymptotic distribution theory for tests which involve nuisance parameters which are not identified under the null hypotheses. The asymptotic distributions of standard test statistics are described as functionals of chi-square processes. In general, the distributions depend upon a large number of unknown parameters. It is shown that a transformation based upon a conditional probability measure yields an asymptotic distribution free of nuisance parameters, and that this transformation can be easily approximated via simulation. The theory is applied to threshold model and Monte Carlo methods are used to assess the finite sample distributions. Moreover, threshold regression methods are constructed in Hansen (1999), and non-standard asymptotic theory of inference is developed which allows construction of confidence intervals and testing of hypotheses.

Also, Anatolyev (2004) provided asymptotic approximations under a drifting parameter DGP for distributions of classical tests and of those designed for the case of complete nonidentification. His simulations showed that the usual asymptotic theory does fail, although actual sizes of the classical LR test display surprising robustness to the degree of identification.

From a Bayesian perspective, different models (including no threshold effect model) can be compared via the computation of marginal likelihood, posterior odds ratios and Bayes factor. However, the main complication for model comparison in Bayesian framework is the sensitivity of the choice of priors for the unidentified parameters. Consequently, sensitivity check need be done in conducting model comparison. In addition, the priors on model-specific parameter have to be proper.

To construct the posterior odds ratios and Bayes factor, let M_0 and M_1 denote the model under the null and the alternative hypothesis, respectively. Also, let $p(y \mid M_i)$ be the marginal likelihood for model i and $p(M_i)$ be the prior model i probability for $i = 0{,}1$. Then the posterior odd ratio and Bayes factor are given by:

$$PO_{ij} = \frac{p(y \mid M_i)p(M_i)}{p(y \mid M_j)p(M_j)}$$

and

$$BF_{ij} = \frac{p(y \mid M_i)}{p(y \mid M_j)}$$

respectively, so that $p(M_i) = p(M_j)$, and the Bayes factor is simply the ratio of the two marginal likelihoods. Thus, in comparing different models, computation of the marginal likelihood for each model is needed. Appendix A provides detailed discussion on the marginal likelihood considerations for the model proposed.

Also, it might be of interest to determine the appropriate model under the parameter restrictions discussed in Section 2. As in the case for threshold effect, posterior odds ratios or Bayes factors can be implemented directly here.

6. Extension to multiple threshold case

The proposed models in Section 3 have only a single threshold. In some applications, there may be multiple thresholds. To simplify the analysis and for exposition purposes, we will confine the discussion to the simplest threshold model (model 4) with the double threshold only. For more than two thresholds, Bayesian analysis of this model are given in Appendix B. Extension of other cases, including model (1), to multiple threshold case follows similarly and are available from the authors upon request.

The simplest double threshold stochastic frontier model takes the form:

$$y_{it} = \beta_1' x_{it} I(q_{it} \le \alpha_1) + \beta_2' x_{it} I(\alpha_1 < q_{it} \le \alpha_2) + \beta_3' x_{it} I(q_{it} > \alpha_2) + v_{it} - u_{it}$$

where the thresholds are ordered so that $\alpha_1 < \alpha_2$. We will focus on the double threshold case since the methods extend in a straightforward manner to higher-order threshold cases. Conditional on the threshold parameters α_1 and α_2, posterior simulation for the other parameters, and latent technical inefficiency proceeds using the principles set forth in Section 4. In particular, given the threshold parameters, the observations can be categorized to one of the three regimes and parameters can be obtained using simple Gibbs updates on a regime-specific basis. Therefore, we can write the model as

$$y_{it} = \beta'_{S_{it}(\alpha)} x_{it,S_{it}(\alpha)} + v_{it,S_{it}(\alpha)} - u_{it,S_{it}(\alpha)}, \tag{6}$$

where $\alpha = (\alpha_1, \alpha_2)$, $S_{it}(\alpha) = 1$ if $q_{it} \le \alpha_1$, $S_{it}(\alpha) = 2$ if $\alpha_1 < q_{it} \le \alpha_2$ and $S_{it}(\alpha) = 3$ if $q_{it} > \alpha_2$, $x_{it,s} = \{x_{it} : S_{it}(\alpha)\} = s$, for $s = 1,2,3$, and similarly for the error terms. In vector notation (6) may be written as $y = W_\alpha \beta + v - u$, where W_α is the matrix consisting of all observations $x_{it,S(\alpha)}$, and β is the vector of all regression coefficients.

The posterior conditional distribution of the threshold parameters is

$$p(\alpha \mid \beta, \sigma_v, \sigma_u, y, X, u) \propto \exp\left[-\tfrac{1}{2}(y + u - \beta' W_\alpha)' \Sigma^{-1} (y + u - \beta' W_\alpha)\right] p(\alpha),$$

where Σ is the $nT \times nT$ diagonal matrix whose diagonal elements are equal to $\sigma^2_{v,S_{it}(\alpha)}$, and $p(\alpha)$ represents the prior on the threshold parameters. Since the latent technical inefficiency variables can be explicitly integrated out of the posterior, a simpler form obtains:

$$p\left(\alpha \mid \beta, \sigma_v, \sigma_u, y, X, u\right) \propto \left[\prod_{i,t} \frac{2}{\sigma_{S_{it}(\alpha)}} \phi\left(\frac{\varepsilon_{it}}{\sigma_{S_{it}(\alpha)}}\right) \Phi\left(\frac{-\lambda_{S_{it}(\alpha)}\varepsilon_{it}}{\sigma_{S_{it}(\alpha)}}\right)\right] p(\alpha)$$

where $\varepsilon_{it} = y_{it} - \beta'_{S_{it}(\alpha)} x_{it,S_{it}(\alpha)}$, $\lambda_{S_{it}(\alpha)} = \sigma_{u,S_{it}(\alpha)} / \sigma_{v,S_{it}(\alpha)}$, and $\sigma^2_{S_{it}(\alpha)} = \sigma^2_{v,S_{it}(\alpha)} + \sigma^2_{u,S_{it}(\alpha)}$. To generate random drawings from this distribution we consider the distributions $p\left(\alpha_1 \mid \alpha_2, \beta, \sigma_v, \sigma_u, y, X, u\right)$ and $p\left(\alpha_2 \mid \alpha_1, \beta, \sigma_v, \sigma_u, y, X, u\right)$, and we use a Metropolis algorithm for each, with a uniform proposal distribution. The range of the proposal distribution is adjusted during the burn-in phase to produce acceptance rates close to 25%. In generating draws from these distributions we have to account for the constraints $\alpha_1 < \alpha_2$ and $q_{\min} \le \alpha_1, \alpha_2 \le q_{\max}$, where $q_{\min}$ and $q_{\max}$ represent the minimum and maximum value of the threshold variable in the sample. In practice we set them equal to the 1% and 99% percentiles of the threshold variable, and we enforce $\alpha_1 < \alpha_2$ using a rejection technique.

Finally, to determine the number thresholds in a particular model, we propose to use marginal likelihood comparison and this approach is similar to that of the model selections in frequentist approach. Appendix A provides details discussion how to evaluate the marginal likelihood. Thus, the appropriate number of thresholds is chosen with the highest marginal likelihood.

7. Empirical Application

7.1. Data

Limam and Miller (2004) examined cross-country patterns of economic growth by estimating a stochastic frontier production function for several developed and developing countries. In addition, they incorporated the quality of inputs in analyzing output growth, where the productivity of capital depends on its average age, while the productivity of labor depends on its average level of education. The rationale is that the older the physical capital, the less new technology is embedded in the capital stocks, and the less productive the capital. Moreover, the productivity of labor increases with the level of education. In this model, output growth can be decomposed into efficiency change, technological change, and input change.

They assumed a standard Cobb-Douglas production function, where aggregate output is produced using the aggregate physical capital stock and labor. Because older capital incorporates less new technology, one expects that the higher the average age, the less productive the capital stock. Similarly, the more educated workers are, the higher the productivity of labor.

The sample contains 80 countries over the period 1960-89. To introduce the effects of geographical location, we consider five subgroups: Africa (23 countries), Latin America (18 countries), East Asia (9 countries), South Asia (7 countries) and the West (23 countries). The dependent variable used in this study is the GDP per capita; the inputs are capital and labor; and the variables that are used in the construction of threshold index are the average age of capital stock, the average education attainment and the time trend. Details about construction of the data and complete list of the countries used in the study are given in Limam and Miller (2004). Finally, all variables are in logarithm except for the trend.

7.2. Results:

Priors

All regressions coefficients are assumed to follow multivariate normal distributions of the form $\beta \sim N(0, 100 I_{\dim(\beta)})$. All scale parameters σ_v and σ_u have relatively non-informative inverted gamma priors, $0.01 / \sigma^2 \sim \chi^2(1)$. For the models 3 and 4 model the threshold parameter we assume a log normal prior of the type, $\log \alpha \sim N(1, 0.3^2)$, implying that α is roughly between 1 and 7 with prior probability 5%. For model 1 and 2, the coefficients γ are assumed to have[4] $\gamma \sim N(0.1, 1.0 I_{\dim(\beta)})$. Our benchmark prior is informative but quite diffuse. We use this prior to see whether meaningful results can be obtained despite the fact that we do not use "sample split" information which is precise enough. If this prior provides reasonable results, then we can address the issue of prior sensitivity and robustness. For the two thresholds models, the (truncated) prior for both threshold parameters is uniform in the interval $(q_{\min}, q_{\max})$. Another possibility is to use the lognormal prior.

As mentioned above, model selection and testing could be sensitive to the choice of the priors of the threshold parameters. Thus, it is important to check for sensitivity of the results to reasonable changes in the priors for the threshold parameters. To do this, we have adopted three

---

[4] We also used the prior $\gamma \sim N(0.1, 0.01 I_{\dim(\beta)})$ to see if a strong prior dominates the data. This was not the case so the data is quite informative in this case.

other priors: (i) $\ln\alpha_j \sim N(0,\ 0.5^2)$, (ii) $\ln\alpha_j \sim N(0,\ 2^2)$; and (iii) $\ln\alpha_j \sim N(2,\ 1)$ for $j = 1,2$. Finally, the following variables are used as thresholds for all models: age of capital, education and time trend.

Bayesian analysis is implemented using Markov Chain Monte Carlo simulation organized around Gibbs sampling with data augmentation. For Gibbs sampling, see Geweke (1999) and the references therein. For Bayesian analysis applied to stochastic frontier models see van den Broeck, Koop, Osiewalski and Steel (1994), Koop, Osiewalski and Steel (2000a,b) and Koop and Steel (2001). Prior elicitation has been considered by these authors in detail and, therefore, we do not repeat it here[5]. Prior elicitation for α is non-trivial but we think the prior selected here should be adequate for most practical purposes. Finally, we have selected the scale parameter of the prior for σ_u (with 1 degree of freedom) so that prior median efficiency is 0.5, 0.7 or 0.9. Our results were robust to this choice.

Gibbs sampling has been implemented using 60,000 iterations, the first 20,000 of which are discarded to mitigate the impact of startup effects. Convergence is monitored using Geweke's (1992) convergence diagnostic and is reported in Table 1 for a single threshold only. Convergence results for zero and double thresholds are similar and hence omitted here. Note that all t-statistics from Geweke's diagnostics were less than 1.7, and the smallest relative numerical efficiency was 0.4 (which is relatively low). Moreover, we take 110,000 draws after an initial 500,000 from different initial conditions have been computed. The results were not sensitive to the initial conditions, which were drawn at random (10 sets in total). We have obtained convergence in all models, except Model 1-age.

First, we determine the number of thresholds for each model. Each model is estimated with none, one and two thresholds, and then the marginal likelihoods are used to facilitate the inference on the number of thresholds. For conservation of space, we do not present all the estimation results here but they are available from the authors upon request. For each model, we found that irrespective of the choice of threshold variables, as well as the choice of priors, the marginal likelihood of a single threshold is always higher than that of zero and two thresholds. Thus, our findings suggest that there is strong evidence of a single threshold in each of the model considered. Consequently, for the remainder of the discussion, we will focus mainly on single

[5] We have tried to use non-informative priors for location and scale parameters as well.

threshold models.

Posterior statistics for all single threshold models with different threshold variable are presented in Tables 2 through 4. Sensitivity of the results to the change in priors were conducted and our result indicated that our results are not excessively sensitive to change in the priors[6]. Thus, the reported results are based on the original prior. Examine the results from Tables 2-4 reveal that all the parameter estimates and more importantly, the estimates of firm specific efficiencies (FSE) are particularly sensitive the model specification as well as the choice of variable that induce the threshold. In particular, the sensitivity of FSE estimates to the form of the model is not something new in applied studies, and such estimates are often sensitive to model specification and distributional assumptions about the two-sided or one-sided error terms. Moreover, in nonlinear models like the ones analyzed in this paper, FSE is expected to be different across alternative models that make radically different assumptions about the functional form, the nature of switching or the covariates. Clearly, the choice between different models is an empirical issue, and with our approach, marginal likelihood provides a natural way to do that.

To this end, by comparing the values of the log marginal likelihood[7] (reported as LML in last row of Tables 2-4), show that the most prefer model is *Model 3* when the threshold variable is the logarithm of education. This suggests that a *probabilistic* mixture (Model 1 and 2) is highly unlikely in the light of the data, and heterogeneity is best captured by *deterministic* separation of the sample in terms of human capital[8]. Thus, for the remainder of this section, we will confine our attention on the results of Model 3 with log of education as a threshold variable.

Focusing on the results reported in the third column of Table 3, we see that the first regime-which is characterized by education values *below* the threshold-has lower labor elasticity, lower capital elasticity, and technical progress averaging 0.6% per year relative to the second regime, where technical progress averages 0.1% per year with a very small posterior standard deviation. The posterior distributions of labor and capital elasticities, technical change and threshold parameter are displayed in Figure 1. The value at which regime switching is 1.723 with very small posterior standard deviation, suggesting regime switching at about exp $(1.723+0.5\times 0.002^2)$ = 5.6 years of education. Years of education in the sample average about 9.4. These results

[6] Sensitivity analyses in the form of figures are available from the authors upon request.
[7] We have used the Bartlett adjustment to compute the Laplace approximation of LML. This practice is also favored by some Monte Carlo results reported in Appendix B of this paper. For the Laplace approximation and various other adjustments see also Geweke, McCausland and Stevens (2003).
[8] Model 3 is also preferred to a simple half-normal stochastic frontier model where the value of LML is -512.46.

emphasize the importance of human capital for productivity-materialized here in the form of higher input elasticities in the second regime. Firm specific technical efficiency[9] averages 0.925 with standard error 0.033 and ranges from 0.688 to 0.984. To get a better understanding on the performance of FSE in each regime, its density plot is presented in Figure 2. From Figure 2, we observe that FSE for the low human capital regime (regime 1) are rather tightly concentrated at high values whereas for the high human capital regime (regime 2), the distribution of efficiencies is more spread and efficiency can be as low as about 0.70. In the low human capital regime, technical efficiency ranges from 0.904 to 0.965, averaging 0.938 and its standard deviation is 0.010. In the high human capital regime, it ranges from 0.688 to 0.984, averaging 0.917 and the standard deviation is 0.039. These results imply that technical efficiency is much more variable in the high human capital regime and although human capital may affect input productivity, it does not seem to be very relevant for improvements in technical efficiency of production. In that sense, it is productivity rather than efficiency that provides the natural playground for human capital and its effect on production. From another point of view, other institutional factors may be responsible for the larger variation of efficiency among countries with a high level of human capital stock, whereas the same factors can be thought of as approximately similar in countries with a lower level of human capital. Further analysis is needed to better understand and examine the differences in efficiency among countries with a higher level of human capital. Since this is not the subject matter of this paper and we do not pursue it here but we believe it is an interesting issue for further applied research.

8. Concluding Remarks

The purpose of this paper was to propose a class of threshold stochastic frontier models that allow for learning and adapting to the "best" technology. We introduced the main model and various special cases organized around the idea that there is a switching from one technology to the other and constructed threshold stochastic frontier models. Bayesian inferences using Gibbs sampling with data augmentation are provided for the analysis of the proposed models. We applied our new models and techniques to a panel of world production frontiers using the

[9] See Koop and Steel (2001) for details. The sampling-theory concept is the familiar Jondrow, Lovell, Materov and Schmidt (1982) measure of technical efficiency. The FSEs were separated into the two groups using the posterior mean of α (1.723) as the threshold value. Since the posterior standard deviation of α is quite small (0.002) the effect of uncertainty about this parameter is quite small.

switching variables based on the age of capital stock, human capital (representing the input quality), and a time trend to capture structural switching or structural transition.

We did not consider in this paper, the case where the threshold variables and/or the inputs are endogenous. In practice, these cases may arise for various reasons; for example, the firm may choose (or switch to) a different production technology due to some self-selection reasons, of which the determinants are the variables used in the regime switching rule. Consequently, ε_{it} will be correlated with v_{it} and u_{it} leading to the endogeneity of the threshold variables. For this case, the presence of endogeneity of threshold variables does not pose any fundamental estimation problem under Bayesian framework as long as the afore mentioned correlation is modelled explicitly (see for example Lai (2013)), since it is only a matter of estimation of a few more correlation parameters. Lai (2013) considers a “within” transformation approach and least squares method to handle the endogeneity of the threshold variables in the stochastic frontier framework. Interested readers are referred to this paper for more details.

Finally, given our analysis discussed in this paper, it would be interesting to extend our models to the smooth transition threshold models where the indicator function in (1) is replaced by a smooth distribution function. We will leave this extension for future research.

References

Anatolyev, S. (2004), Inference when a nuisance parameter is weakly identified under the null hypothesis, *Economics Letters*, Volume 84, Issue 2, 245–254.

Andrews, D.W.K and Ploberger, W. (1994), Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica*, 62, 1383–1414.

Aigner, D., K. Lovell, and P. Schmidt, 1977, Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* 6, 21-37.

Bai, J., 1997, Estimating multiple breaks one at a time, *Econometric Theory* 13, 315-352.

Bawuens L., M. Lubrano and J.-F. Richard, 1999, Bayesian inference in dynamic econometric models, Oxford University Press.

van den Broeck, J., G. Koop, J. Osiewalski, and M.F.J. Steel, 1994, Stochastic frontier models: A Bayesian perspective, *Journal of Econometrics* 61, 273-303.

Chen, M.-H., Q.-M. Shao, and J.G. Ibrahim, 2000, Monte Carlo Methods in Bayesian Computation, Springer, New York.

Chib, S., 1995, Marginal likelihood from the Gibbs sampler, *Journal of the American Statistical Association* 90, 1313-1321.

Davies, R.B. (1987), Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, 74, 33–43.

DiCiccio, T.J., R.E. Kass, A. Raftery, and L. Wasserman, 1997, Computing Bayes factors by combining simulation and asymptotic approximation, *Journal of the American Statistical Association* 92, 903-915.

Fernandez, C., J. Osiewalski, and M.F.J. Steel, 1997, On the use of panel data in stochastic frontier models, *Journal of Econometrics* 79, 169-193.

Gelfand, A.E., and D.K. Dey, 1994, Bayesian model choice: Asymptotics and exact calculations, *Journal of the Royal Statistical Society* B 56, 501-514.

Gelmann, A. (2006), Prior distributions for variance parameters in hierarchical models, *Bayesian Analysis* 1, Number 3, 515–533

Geweke, J., 1989, Bayesian inference in econometric models using Monte Carlo integration, *Econometrica* 57, 1317-1340.

Geweke J., 1992, Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in J.M. Bernardo, A.F.M. Smith, A.P. Dawid, and J.O. Berger (eds) *Bayesian Statistics 4*, 169-193, New York: Oxford University Press.

Geweke, J., 1999, Using simulation methods for Bayesian econometric models: Inference, development and communication (with discussion and rejoinder), *Econometric Reviews* 18: 1-126.

Geweke J., 2006, Interpretation and inference in mixture models: Simple MCMC works, discussion paper, Department of Economics, University of Iowa.

Geweke, J., W. McCausland, and J. J. Stevens, 2003, Using simulation methods for Bayesian econometric models, in E. A. Giles (ed), Computer-aided Econometrics, New York, Marcel Dekker, 209-261.

Greene, W.H., 1993, The econometric approach to efficiency analysis, in H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, *The measurement of productive efficiency: Techniques and applications*, Oxford University Press, New York, 68-119.

Greene, W.H., 2001, New developments in the estimation of stochastic frontier models with panel data, Remarks prepared for the 7th European Workshop on Efficiency and Productivity Analysis, University of Oviedo, Spain.

Greene, W.H., 2004, Distinguishing Between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organization's Panel Data on National Health Care Systems, *Journal of Health Economics* (forthcoming).

Hansen, B.E., 1996, Inference When Nuisance Parameters is not Identified under the Null Hypothesis, *Econometrica*, 64, pp. 413-430.

Hansen, B.E. 2000, Sample Splitting and Threshold Estimation, *Econometrica*, 68, pp. 575-603.

Jondrow, J., C. A. K. Lovell, I. S. Materov and P. Schmidt (1982), On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model, *Journal of Econometrics* 19:2/3 (August), 233-38.

Kass, R., and A. Raftery, 1995, Bayes factors, *Journal of the American Statistical Association* 90, 773-795.

Kim, C.-J., and C. R. Nelson 1999, State-Space Models with Regime-Switching: Classical and Gibbs-Sampling Approaches with Applications, MIT Press.

Koop, G., J. Osiewalski and M.F.J. Steel, 2000a, A stochastic frontier analysis of output level and growth in Poland and Western economies, *Economics of Planning* 33, 185-202.

Koop, G., J. Osiewalski and M.F.J. Steel, 2000b, Measuring the sources of output growth in a panel of countries, *Journal of Business and Economic Statistics* 18, 284-299.

Koop, G., and M.F.J. Steel, 2001, Bayesian analysis of stochastic frontier models. In *A Companion to Theoretical Econometrics*, Baltagi B. (ed). Blackwell, 520-573.

Kumbhakar, S.C. and C.A.K. Lovell, 2000, *Stochastic Frontier Analysis*, Cambridge University Press, New York, NY.

Lai, H-P, 2013, Estimation of threshold stochastic frontier model in the presence of an endogenous sample split variable, *Journal of Productivity Analysis, 40*, 227-237.

Lewis, S.M. and Raftery, A.E., 1997, Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator, *Journal of the American Statistical Association*, *92*, 648-655.

Limam, Y.R. and S.M. Miller, 2004, Explaining Economic Growth: Factor Accumulation, Total Factor Productivity Growth, and Production Efficiency Improvement, *Quarterly Review of Economics and Finance*, forthcoming.

Liu, J.S., 1994, The collapsed Gibbs sampler in Bayesian computations with application to a gene regulation problem, *Journal of the American Statistical Association* 89, pp. 958-966.

Mbaga, M., R. Romain, B. Larue, and L. Lebel, 2003, Assessing technical inefficiency of Quebec dairy farms, *Canadian Journal of Agricultural Economics* 51, pp. 121-137.

Meeusen W., and J. van den Broeck, 1977, Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review* 8, 435-444.

Meng, X.-L., and W.H. Wong, 1996, Simulating ratios of normalizing constants via a simple identity: A theoretical explanation, Statitica Sinica 6, 831-860.

Orea, L. and S.C. Kumbhakar, 2004, Efficiency Measurement Using a Stochastic Frontier Latent Class Model, *Empirical Economics*, 29, 69-83.

Staiger, D. and J.H Stock (1997), Instrumental variables regression with weak instruments, *Econometrica*, 65, 557–586.

J.H Stock, J.H Wright (2000), GMM with weak identification, *Econometric*a, 68, 1055–1096.

Tierney L., and J.B. Kadane, 1986, Accurate approximations for posterior moments and marginal densities, *Journal of the American. Statistical Association*, 81, 82-86.

Tierney, L., R.R. Kass and J.B. Kadane, 1989, Fully exponential Laplace approximations to expectations and variances of nonpositive functions, *Journal of the American. Statistical Association*, 84, 710-716.

Tsionas, E.G., 2002, Stochastic Frontier Models with Random Coefficients, *Journal of Applied Econometrics*, 17, pp. 127-147.

Tsionas, E.G., S.C. Kumbhakar, 2004, Markov Switching Stochastic Frontier Model, *The Econometrics Journal*, 7, pp.1-28.

Zellner, A., 1986, On assessing prior distributions and Bayesian regression analysis with g-prior distributions, in P.K. Goel and A. Zellner (eds.), *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti*. Amsterdam, North Holland.

APPENDIX A: Marginal Likelihood Considerations

An important issue is whether the posterior distribution is sufficiently close to normality to justify approximation of the marginal likelihood using the Laplace approximation.[10] We use quantile-quantile expressions of the quantity $F = (\theta - \bar{\theta})' \bar{\Sigma}^{-1} (\theta - \bar{\theta}) \xrightarrow{D} \chi^2_k$ (under normality of the joint posterior), where $\bar{\theta}$ is the posterior mean, $\bar{\Sigma}$ is the posterior covariance matrix and k is the dimensionality of the parameter vector (which varies from model to model). Available upon request, are typical quantile-quantile plots of some of the models estimated in this paper. The empirical cdf of F is computed using the MCMC draws taken every other tenth to mitigate the impact of autocorrelation in the estimation of posterior covariance matrix. Although the posterior distribution is non-normal the deviations do not seem significant enough and justify normality as a reasonable approximation. This means that Laplace's method should be a reasonable approximation to the marginal likelihood of the models analyzed in this paper.

Next we take up the more general issue of how well the Laplace approximation behaves in estimating the marginal likelihood of stochastic frontier models. To this end we consider a stochastic frontier model of the form $y_{it} = \beta_1 + \beta_2 X_{it1} + \beta_3 X_{it2} + v_{it} - u_{it}$, with $i = 1,...,n$, $t = 1,...,T$, $v_{it} \sim i.i.d.\ N(0, \sigma_v^2)$, $u_{it} \sim \left| i.i.d.\ N(0, \sigma_u^2) \right|$. The parameter choices are $\beta_1 = -1$, $\beta_2 = \beta_3 = 0.5$, $\sigma_v = \sigma_u = 0.1$, the regressors are generated as $i.i.d.\ N(0,1)$ and they are not fixed in repeated samples. The prior is $\frac{Q_v}{\sigma_v^2} \sim \chi^2_{N_v}$, where $Q_v = 0.01$ and $N_v = 1$, $\frac{Q_u}{\sigma_u^2} \sim \chi^2_{N_u}$, and we consider alternative choices of the hyperparameters N_u and Q_u. Clearly, these technical inefficiency densities are quite different in terms of what they imply about prior efficiency. It should also be mentioned that choosing $\sigma_v = \sigma_u$ in the parameterization is not only empirically plausible but also a relatively hard case for estimation and inference since the "signal to noise ratio" is equal to one. For the regression parameters we assume $\beta \sim N_3(0,\ g \cdot (X'X)^{-1})$ where $g = 100$. This is Zellner's g-prior distribution (Zellner, 1986). We have also experimented with the prior $\beta \sim N_3(0,\ g \cdot I_k)$ but qualitatively the results in terms of marginal likelihood were not significantly different.

We will consider a Monte Carlo experiment with 100 data sets. For each data set the model is analyzed using the Gibbs sampler with 5500 iterations, the first 500 of which are discarded to mitigate start up effects. Standard convergence diagnostics (Geweke, 1992) indicate that convergence is obtained quite early when we start the Gibbs sampler from least squares quantities (with $\sigma_v = \sigma_u = s$ where s is the residual standard deviation). From the 5000 draws that are left, we take every other tenth to approximate the marginal likelihood. Regarding the sample size we consider both cross-sectional and panel data and our choices are dictated by what is reasonable in terms of data sets actually used in practice.

Several methods are used to approximate the marginal likelihood and three things seem to be worth mentioning. If we take Chib's approximation as the closest to the right answer, then (a) log

[10] For a general discussion of Bayes factors see Kass and Raftery (1995).

marginal likelihood resulting from Laplace approximation using the Bartlett adjustment is by far the closest to the Chib approximation, (b) the approximation is much better in panel data rather than in cross-sectional data. Finally, (c) if the objective is, as usual, to compute Bayes factors then it does not really matter which method is used since for most methods differences of log marginal likelihoods relative to the Chib approximation are more or less constant across different configurations of the sample size.

Given the relative ease of computing Laplace approximations it seems that this method coupled with a Bartlett adjustment provides a close approximation to the value obtained by the more accurate method due to Chib. Chib's method is relatively cumbersome in implementation since it involves repeated Gibbs sampling fixing in sequentially every element of the parameter vector to its value taken at the point of approximation (typically the posterior mean). All other methods can be implemented more or less in an automatic way since they only require coding the likelihood function and the prior distribution. Of course the results reported here cannot be taken as comprehensive but they can be taken as indicative of how different approximations to the marginal likelihood behave in a set up that is empirically plausible and relevant. Moreover, the results reported here are relevant in the sense that the stochastic frontier model is highly non-normal by construction. To our knowledge this is the only Monte Carlo evaluation of alternative marginal likelihood estimators in stochastic frontier models.

APPENDIX B: Multiple Threshold Extension

We can extend the model in Section 6 as follows:

$$y_{it} = \sum_{r=1}^{R+1} \beta_r' I(\alpha_{r-1} < q_{it} \leq \alpha_r) x_{it} + v_{it} - u_{it}, \tag{B.1}$$

where $\alpha_0 = -\infty, \alpha_{R+1} = \infty$ and R is the number of regimes. We assume $\alpha_1 < \alpha_2 < ... < \alpha_R$. We assume $q_{it} = z_{it}'\gamma$ where z_{it} is an $m \times 1$ vector of thresholding variables. Conditionally on $\alpha = \ \alpha_1, ..., \alpha_R\ '$ the model is a standard Bayesian stochastic frontier and we can follow the techniques laid out in Section 4. This is because we can write (B.1) in the form:

$$y_{it} = \beta' w_{it}\ \ \alpha\ + v_{it} - u_{it}, \tag{B.2}$$

where $\beta = \left[\beta_1', ..., \beta_R'\right]'$ and

$$w_{it}\ \ \alpha\ = \left[I(q_{it} \leq \alpha_1) x_{it}', I(\alpha_1 < q_{it} \leq \alpha_2) x_{it}', ..., I(\alpha_R < q_{it} \leq \alpha_{R+1}) x_{it}'\right]'.$$

The conditional posterior distributions of the elements of vector α are as follows.
For α_1 we have:

$$p\ \ \alpha_1 \mid \alpha_{(-1)}, \beta, y, X, Z, u\ \propto \exp\ -\tfrac{1}{2\sigma_v^2} \sum_{i,t: q_{it} \leq \alpha_1} \left[y_{it} + u_{it} - \beta_1' x_{it}\right]^2\ ,$$

$$p\ \ \alpha_2 \mid \alpha_{(-2)}, \beta, y, X, Z, u\ \propto \exp\ -\tfrac{1}{2\sigma_v^2} \sum_{i,t: \alpha_1 < q_{it} \leq \alpha_2} \left[y_{it} + u_{it} - \beta_2' x_{it}\right]^2\ , ... \tag{B.3}$$

$$p\ \ \alpha_R \mid \alpha_{(-R)}, \beta, y, X, Z, u\ \propto \exp\ -\tfrac{1}{2\sigma_v^2} \sum_{i,t: q_{it} > \alpha_R} \left[y_{it} + u_{it} - \beta_{R+1}' x_{it}\right]^2\ ,$$

where the notation $\alpha_{(-r)}$ denotes all elements of α with the exception of the rth element, subject to the restrictions

$$\alpha_1 < \alpha_2 < ... < \alpha_R. \tag{B.4}$$

Draws from these conditional posterior distributions can be obtained using a Metropolis-Hastings algorithm as in Section 6.
It might be best to draw all elements of α simultaneously as the restrictions in (B.4) can be incorporated in a straightforward way.
Conditionally on the αs, the conditional posterior distribution of γ is:

$$p\left(\gamma \mid \alpha,\beta,y,X,Z,u\right) \propto \exp\left\{-\tfrac{1}{2\sigma_v^2}\sum\nolimits_{i,t}\left[y_{it}+u_{it}-\sum\nolimits_{r=1}^{R+1}\beta_r' x_{it} I\left(\alpha_{r-1} < z_{it}'\gamma \le \alpha_r\right)\right]^2\right\}. \tag{B.5}$$

Although the distribution does not belong to a standard family, we can use again the Metropolis-Hastings algorithm to provide random draws. For example, in the random walk Metropolis-Hastings algorithm, a candidate is drawn: $\gamma_* \sim N\left(\gamma^{(s-1)}, V\right)$ where $\gamma^{(s-1)}$ is the previous draw and $V = hI$ for some constant $h > 0$ which is determined so that approximately ¼ of all candidates are accepted. The acceptance rule is

$$\gamma^{(s)} = \gamma_*, \text{ with probability } \min\left\{1,\ \frac{p\left(\gamma_* \mid \alpha,\beta,y,X,Z,u\right)}{p\left(\gamma^{(s-1)} \mid \alpha,\beta,y,X,Z,u\right)}\right\}, \tag{B.6}$$

else $\gamma^{(s)} = \gamma^{(s-1)}$.

Table 1. Geweke's Convergence Diagnostic (Single Threshold)

	Model 1	Model 2	Model 3	Model 4
CD, Parameters	0.471 – 3.541	0.212 – 1.344	0.415 – 1.212	0.200 – 1.071
CD, Latent var.	0.303 – 2.919	0.313 – 1.510	0.210 – 0.917	0.810 – 1.444
RNE	0.114 (0.035)	0.310 (0.431)	0.265 (0.414)	0.317 (0.215)
NSE	0.015 (0.021)	0.0011 (0.0016)	0.0022 (0.0039)	0.0015 (0.0022)

Notes: CD is Geweke's (1992) convergence diagnostic (absolute value of the *t*-statistic for testing the difference of means in the first 50% and last 20% of the draws. RNE is relative numerical efficiency and NSE is the numerical standard error). For RNE and NSE reported are statistics for the structural parameters. In parentheses reported are statistics for the latent variables. All reported statistics are medians across structural parameters and latent variables. All statistics are also medians across the threshold variables age and education.

Table 2. Posterior statistics for Model 4

	Threshold variable		
	age of capital	education	trend
constant	-0.069 (0.015) 2.535 (0.134)	-0.102 (0.015) 2.438 (0.071)	-0.068 (0.015) 2.392 (0.071)
labor	0.274 (0.007) 0.253 (0.013)	0.200 (0.008) 0.346 (0.009)	0.256 (0.009) 0.274 (0.007)
capital	0.712 (0.006) 0.707 (0.011)	0.731 (0.006) 0.675 (0.007)	0.716 (0.007) 0.714 (0.006)
trend	0.012 (0.0006) -0.001 (0.0007)	0.006 (0.0008) 0.001 (0.0005)	0.012 (0.003) -0.002 (0.0007)
α	2.375 (0.022)	1.663 (0.028)	8.895 (0.667)
σ_v	0.169 (0.003)	0.162 (0.003)	0.169 (0.003)
σ_u	0.057 (0.015)	0.054 (0.014)	0.060 (0.017)
FSE	0.956 (0.007)	0.958 (0.006)	0.954 (0.008)
LML	-567.70	-531.80	-591.30

Notes: Numbers in parentheses are posterior standard deviations. For each cell, the first row gives the Bayesian results for the first regime and the second row corresponds to the second regime. FSE is firm specific efficiency (with s.d. in parentheses) and LML is the log marginal likelihood. The coefficients of regional dummies were restricted to be common in the two regimes. The coefficients of regional dummies are not reported. The LML value of a simple half-normal stochastic production frontier was 512.46. Detailed estimates for this model are not reported.

Table 3. Posterior statistics for Model 3

	Threshold variable		
	age of capital	education	trend
constant	0.015 (0.019) 3.034 (0.095)	0.024 (0.031) -0.045 (0.017)	-0.113 (0.031) -0.051 (0.018)
labor	0.246 (0.009) 0.377 (0.010)	0.222 (0.010) 0.294 (0.009)	0.253 (0.012) 0.280 (0.008)
capital	0.736 (0.008) 0.636 (0.008)	0.707 (0.009) 0.732 (0.009)	0.730 (0.011) 0.703 (0.007)
trend	-0.002 (0.0008) 0.002 (0.0004)	0.0061 (0.001) 0.001 (0.0004)	0.011 (0.0025) -0.002 (0.0007)
α	2.177 (0.007)	1.723 (0.002)	9.790 (0.705)
σ_v	0.165 (0.006) 0.085 (0.006)	0.213 (0.006) 0.086 (0.004)	0.169 (0.006) 0.163 (0.004)
σ_u	0.146 (0.017) 0.217 (0.01)	0.082 (0.021) 0.112 (0.009)	0.082 (0.019) 0.078 (0.016)
FSE	0.875 (0.072)	0.925 (0.033)	0.939 (0.013)
LML	-431.33	-412.57	-422.70

Notes: Numbers in parentheses are posterior standard deviations. In each cell, the first row gives results for the first regime and the second row corresponds to the second regime. FSE is firm specific efficiency (with s.d. in parentheses) and LML is the log marginal likelihood. The coefficients of regional dummies are not reported.

Table 4. Posterior statistics for Models 1 and 2

	Model 2	*Model 1*
constant	2.180 (0.008) 3.071 (0.006)	1.457 (0.060) 2.753 (0.028)
labor	0.318 (0.001) 0.252 (0.001)	0.176 (0.006) 0.346 (0.003)
capital	0.710 (0.0009) 0.689 (0.006)	0.806 (0.005) 0.669 (0.001)
trend	-0.001 ($4x10^{-5}$) 0.012 (0.0001)	-0.0006 ($3x10^{-4}$) 0.0016 ($3x10^{-4}$)
σ_v	0.003 ($7x10^{-5}$) 0.003 ($6x10^{-5}$)	0.002 ($3.8x10^{-5}$) 0.005 (0.0002)
σ_u	0.312 (0.005) --	0.385 (0.006) 0.396 (0.008)
Regime switching determinants		
constant	-1.447 (0.389)	1.258 (0.678)
education	0.289 (0.05)	-0.161 (0.095)
age capital	0.247 (0.081)	0.088 (0.132)
trend	0.015 (0.003)	-0.109 (0.008)
FSE	0.797 (0.136)	0.758 (0.154)
LML	-551.66	-543.10

Notes: Numbers in parentheses are posterior standard deviations. In each cell, the first row gives results for the first regime and the second row corresponds to the second regime. FSE is firm specific efficiency (with s.d. in parentheses) and LML is the log marginal likelihood. The coefficients of regional dummies are not reported.

Figure 1: Posterior Distributions Labor and Capital Elasticities, Technical Change and Threshold Parameter from Model 3 - Education

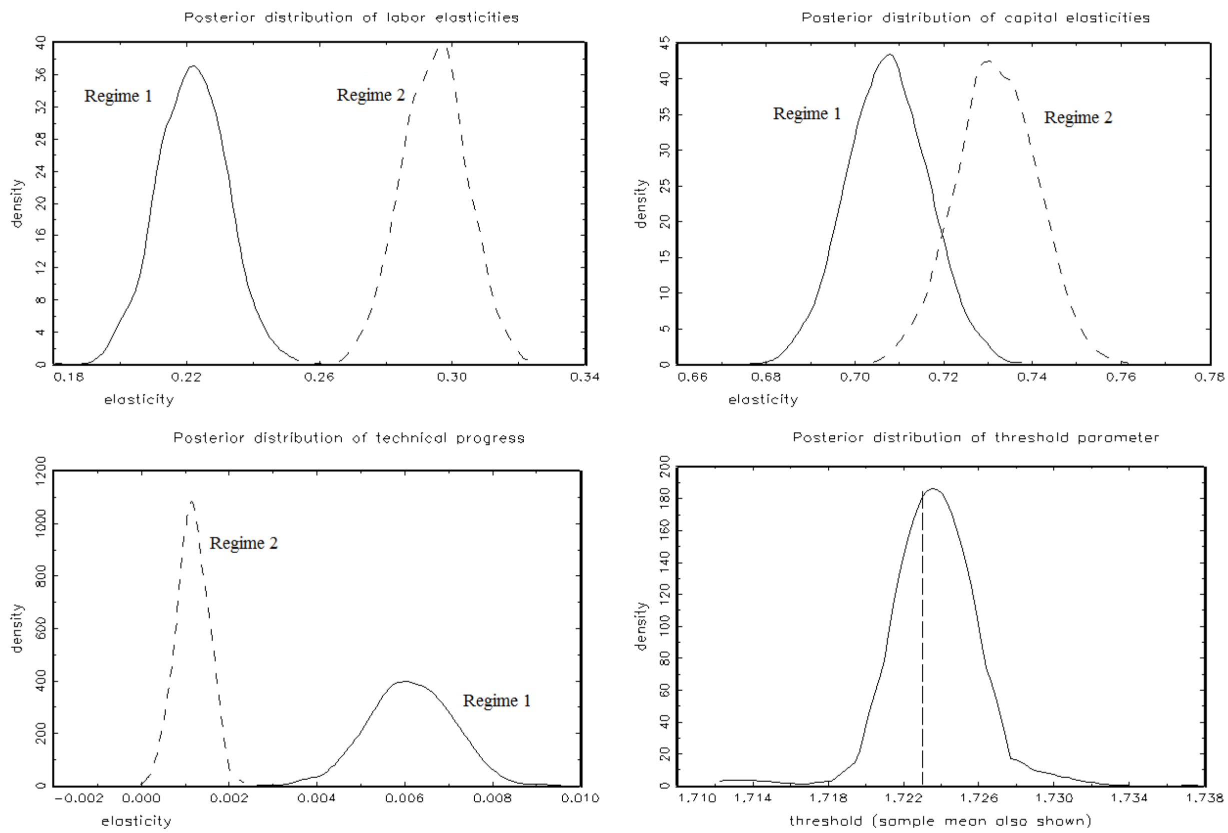

Figure 2: Distribution of FSE from Model 3 - Education