STOCHASTIC FRONTIER MODELS WITH RANDOM COEFFICIENTS

Efthymios G. Tsionas

Department of Economics, Athens University of Economics and Business,
Athens, Greece.

Abstract

The paper proposes a stochastic frontier model with random coefficients to separate technical inefficiency from technological differences across firms, and free the frontier model from the restrictive assumption that all firms must share exactly the same technological possibilities. Inference procedures for the new model are developed based on Bayesian techniques, and computations are performed using Gibbs sampling with data augmentation to allow finite-sample inference for underlying parameters and latent efficiencies. An empirical example illustrates the procedure.

JEL Classification Codes: C11, C23, D24.

Keywords: Stochastic frontier models; technical efficiency; random coefficient models; Bayesian analysis; Markov Chain Monte Carlo; Gibbs sampling.

Acknowledgments: Gary Koop kindly provided the data. The author had useful discussions with Bill Greene, Gary Koop, Mark Steel, P. A. V. B. Swamy, and George Tavlas. All remaining errors are the author's responsibility.

Address: 76 Patission Street, 104 34 Athens, Greece.
Tel. (++301) 820 3392, Fax: (++301) 823 8249 & 820 3301, Internet: tsionas@aueb.gr

1. Introduction

Beginning with Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977), efficiency measurement using stochastic frontier models has become a standard tool in the arsenal of applied economist. Stochastic frontier models start with a production or cost function (known as the frontier) and decompose the error term into two components. The first component is measurement error. The second component is a one-sided disturbance, which reflects the fact that some decision-making units can be below the frontier. Considerable effort has been devoted in modeling the one-sided disturbance by using exponential (Meeusen and van den Broeck, 1977), truncated normal (Aigner *et al*., 1977) or *gamma* distributions (Greene, 1990 and Tsionas, 2000). Both sampling theory and Bayesian methods of inference have been explored in connection with estimation and efficiency measurement using stochastic frontiers (see Stevenson, 1990, van den Broeck *et al*., 1994 and Koop, Osiewaski and Steel, 1997). For an overview of stochastic frontier models see Bauer (1990). Koop and Steel (2001) present a recent overview of the Bayesian approach to stochastic frontiers. See also Kalirajan and Shand (1999).

According to the frontier approach, all firms share exactly the same production possibilities and differ only with respect to their degree of inefficiency, which can be due to a host of causes including, for example, differences in the managerial input. Therefore, traditional methods assume all firms face a common frontier, but their actual output might lie below this frontier and the vertical distance is a measure of inefficiency. If the researcher deals with a group of firms, which are known to share exactly the same production possibilities, the traditional stochastic frontier model is adequate. However, in practice, firms have different technologies for a variety of reasons. For example, adoption of a new technology is costly, and firms adopt new technologies only with considerable lags, see for example Reinganum (1989, p.383). If costs related to installation and personnel training differ across firms, at any given point in time there will be some variability in the types of technology used by firms. Therefore, in practice, production possibilities are expected to differ in a cross-section of firms, and a set of different technologies may simultaneously coexist at any given time. If that is the case, efficiency measurement cannot proceed under the assumption of common technology.

Assuming that firms share the same technology when it is inappropriate to do so, will result in misleading efficiency measurement and confusion between technological differences and technology-specific inefficiency. More specifically, inability to produce efficiently given a set of inputs is mixed up with using a different (possibly inferior) technology. Obviously, however, these two concepts have little in common. If, for example, there are two firms, firm A which uses the latest technology and firm B which uses an older version of this technology, it is natural that firm A produces more output compared to B, in at least a subset of the factor space. Suppose both firms are fully efficient relative to their technology. If one incorrectly fits a common frontier, then one will find that firm B is inefficient. If, however, it is not optimal for firm B to adopt the latest technology because of high firm-specific costs or other reasons, it follows one is misguided in concluding that firm B is inefficient. The relative difference in output reflects technological differences, not inferior practice. In the real world, of course, there are both technology differentials and technical inefficiency (relative to firm-specific technologies) so a need arises for models, which can account for both features.

Needless to say efficiency analysis in the context of random coefficient stochastic frontiers is asking a lot from the data, and strong identifying assumptions are needed in order to obtain meaningful estimates. One particular assumption is the distributional structure of random coefficients across firms. Here, we adopt a multivariate normal distribution (which is the standard in the traditional random coefficient literature) but this assumption cannot always be correct. A reasonable alternative would be a finite mixture of normals if it is suspected that a given number of possible technologies exist. The essence of the approach remains the same but computations become more involved.

Another issue is whether it makes sense to define inefficiency simply based on the intercepts of the frontier model, and ignore variation in the remaining parameters: One could certainly claim that firm B in the previous example must be inefficient since it uses an inferior technology. The validity of this claim depends on the reasons why an inferior technology is used. If the firm is not familiar with the existence of a new technology or neglects to adopt it because of inferior management practices it must certainly be the case that it is inefficient. If, however, it is not *profitable* for the firm to adopt the new technology because of high adjustment costs, this is no longer true. If managerial practices are sound and usage of an old technology is optimal for the firm (at a given point in time) should output loss from using the old technology be included in inefficiency? I adopt the view that this would not accord well with the concept of inefficiency as traditionally used in economics: What this concept tries to capture is output loss from inferior practices that can be improved without significant cost to the firm. However, output loss from inferior technology usage is not a choice that can be altered at no cost.

Besides, there are practical problems with an inefficiency concept that takes account of variation in all parameters: Some firms would be inefficient in certain subsets of the factor space, and efficient in others. Then it would not be apparent how the inefficiency contributions of different subsets can be aggregated into a single measure. Indeed, it is not apparent that a single measure even exists. Another point is that if we allow inefficiency to include the entire output difference at a given factor level (resulting from inferior technologies or not) then it would not be sensible to measure inefficiency without an explicitly dynamic model. The reason is that adoption of a new technology is an inherently dynamic process. This would complicate the analysis beyond the purpose of the present paper but the idea could be pursued in future research.

Kalirajan and Obwona (1994) have specified a random coefficient average production function model and measured inefficiency using the residuals from a frontier derived by using the maximum response coefficients. Their motivation for using a random coefficient model was not that firm technology may be heterogeneous but rather, they wanted to avoid the assumption that the frontier is a neutral shift of the conventional production function. Although their contribution is important, their inefficiency measure includes both technological differences and firm-specific inefficiency. In addition, they were able to derive only relative, not absolute inefficiency measures. On related issues, see Akhavein, Swamy, Taubman, and Singamsetti (1997).

The present paper frees the stochastic frontier model from the assumption that the frontier is common to all firms and proposes a random coefficient stochastic frontier model where (absolute) firm-specific efficiency can be separated from technological differentials across firms. On random coefficient models, see Swamy and Tavlas (1995, 2001). Exact finite sample results for parameters as well as latent efficiencies are derived using a Bayesian analysis of the model. The computations are organized around Markov Chain Monte Carlo methods, and especially the Gibbs sampler with data augmentation. Computations are not more difficult than computations for the usual stochastic frontier model or the random coefficient model without one-sided disturbances.

The remaining of the paper is organized as follows. The next section presents the model. Section 3 develops Bayesian inference procedures for the stochastic frontier model with random coefficients. An empirical application is analyzed in section 4. The final section concludes the paper.

2. The model

Consider the model

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta}_i + v_{it} - u_{it}, \quad i = 1,..,N, \; t = 1,..,T \tag{1}$$

where $\mathbf{x}_{it}$ is a $(K\text{-}1)\times 1$ vector of observations for the explanatory variables (typically logs of inputs) , y_{it} is the i*th* observation of year t for the dependent variable (typically the log of production), v_{it} is measurement error, distributed as *i.i.d* $\mathrm{N}(0,\sigma^2)$), u_{it} is a non-negative disturbance, $\boldsymbol{\beta}_i$ is a $(K-1)\times 1$ vector of random coefficients, and α is an intercept which is assumed to be non-random (because of the presence of the random error term v_{it} in equation (1)). The case $T = 1$ corresponds to cross-section data.

The error term u_{it} reflects an inefficiency component that forces production to be below the frontier. If $u_i = 0$ then the firm is fully efficient. Although (1) is assumed to represent a production function, in the case of cost function we have

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta}_i + v_{it} + u_{it}, \quad i = 1,..,N, \; t = 1,..,T \tag{2}$$

so if the signs of y_{it} and $\mathbf{x}_{it}$ are reversed all results of the production frontier apply to the cost frontier as well. To complete the specification of the model, it is assumed that u_{it} is exponentially distributed with parameter θ, *i.e.* u_{it}'s are *i.i.d* with distribution whose density is given by

$$f(u_{it}) = \theta \exp(-\theta u_{it}) \tag{3}$$

Alternative specifications for u_{it} will be considered as well. Parameters $\boldsymbol{\beta}_i$ are distributed according to a ($K-1$)-variate normal distribution as follows.

$$\boldsymbol{\beta}_i \sim \mathrm{N}(\overline{\boldsymbol{\beta}}, \boldsymbol{\Omega}),\ i = 1,..,N \tag{4}$$

where $\overline{\boldsymbol{\beta}}$ is a $(K-1)\times 1$ vector of parameter means, and $\boldsymbol{\Omega}$ is a $(K-1)\times(K-1)$ positive definite covariance matrix. Clearly, usual stochastic frontier models (van den Broeck *et al.,* 1994) can be obtained as a special case by fixing $\boldsymbol{\Omega} = \mathbf{0}_{K-1,K-1}$.

The assumptions of the model are that $\boldsymbol{\beta}_i \mid \overline{\boldsymbol{\beta}}, \boldsymbol{\Omega}$ are independent, and e_{it} as well as u_{it} are independent of $\mathbf{x}_{it}$. The independence assumption is standard practice in the frontier literature. Apparently, this model is a hierarchical model with two levels of latent variables, namely u_{it} and $\boldsymbol{\beta}_i$. Each firm has its own production function with parameters $\boldsymbol{\beta}_i$ which reflects heterogeneity of firms in their technology. That is, at the first stage nature gives each firm a specific set of technological coefficients $\boldsymbol{\beta}_i$ from distribution (4). At the second stage, each firm experiences a shock, which determines its inefficiency level u_{it} from an exponential distribution with parameter θ.

This random coefficient formulation is due to Hildreth and Houck (1968) and is a special case of Swamy's specification (Swamy, 1970). At this point it must be explained why in (1) only the slope coefficients are random whereas the intercept is fixed. If we write (1) in the equivalent form

$$y_{it} = (\alpha + v_{it}) + \mathbf{x}'_{it}\boldsymbol{\beta}_i - u_{it} \tag{5}$$

it is clear that all coefficients (including the intercept) are random but there is no measurement error in the model. Clearly it is not possible to have both measurement error and a random intercept. This is not limited to the stochastic frontier model but is endemic to all random coefficient models. For this reason I make the choice to have measurement error appear explicitly in the model and let the intercept be constant. It must be noted, however, that with panel data the following formulation is possible:

$$y_{it} = (\alpha + \xi_i) + \mathbf{x}'_{it}\boldsymbol{\beta}_i + v_{it} - u_{it} \tag{6}$$

where ξ_i is an error term so that $\alpha + \xi_i$ is the random intercept, and v_{it} is the usual noise. Therefore, with panel data it is possible to have a (cross-sectional) random intercept and noise at the same time. When panel data is not available, there is limited confidence in the estimates of intercept and inefficiency parameter u_{it}. Finally, in the case of panel data an assumption like $u_{it} = u_i$ (for all $t = 1,..,T$) is common.

Substituting (4) in (1), one obtains

$$y_{it} = \alpha + \mathbf{x}'_{it}\overline{\boldsymbol{\beta}} + e_{it} - u_{it} \tag{7}$$

where e_{it} is distributed as independent $\mathrm{N}(0, \sigma^2 + \mathbf{x}'_{it}\boldsymbol{\Omega}\mathbf{x}_{it})$. If $\mathbf{z}'_{it} = [1 \ \mathbf{x}'_{it}]$ and $\boldsymbol{\delta} = [\alpha \ \overline{\boldsymbol{\beta}}']'$, it follows that (7) can be written as

$$y_{it} = \mathbf{z}'_{it}\boldsymbol{\delta} + e_{it} - u_{it} \tag{8}$$

Therefore, the assumption of random coefficients in (1) implies a stochastic frontier with normal, heteroscedastic measurement error. Not accounting for this heteroscedasticity has important consequences for estimation of stochastic frontiers by maximum likelihood, as the resulting estimates will be inconsistent. This is known in the frontier literature since Caudill and Ford (1993) and Caudill, Ford and Gropper (1995). The difference with this literature is that heteroscedasticity in the present paper is quadratic in the regressors with inequality restrictions on the parameters owing to the symmetry and positive definiteness of the covariance matrix $\mathbf{\Omega}$. Caudill, Ford and Gropper (1995) on the other hand explicitly model the logarithm of variance of the one-sided error as a linear function of exogenous variables.

It should be mentioned that there is a sense in which parameters $\boldsymbol{\beta}_i$ cannot simply be dropped from the analysis, as we have done in going from (1) and (4) to (7). If there are regularity restrictions in the model (for example, $\boldsymbol{\beta}_i$'s be positive or returns to scale be less than an upper bound), then (4) should be restricted to the regularity region. In (7), the error term e_{it} should account for this. In particular, e_{it} would be distributed independently as $N(0, \sigma^2 + \mathbf{x}_{it}'\mathbf{\Omega}\mathbf{x}_{it})$ truncated to an appropriate regularity region that would have to depend on $\bar{\boldsymbol{\beta}}$ as well as $\mathbf{x}_{it}$'s. Since this complicates the analysis considerably, regularity conditions will not be imposed upon the $\boldsymbol{\beta}_i$'s. The same approach has been followed by Kalirajan and Obwona (1994) and all random coefficient models to the author's knowledge.

It is interesting to note that (8) makes clear why one can compute technical inefficiency although frontiers are firm specific. In model (1) with no restriction on the firm-specific coefficients it would not be possible to identify both a firm-specific frontier and the firm's technical efficiency relative to that frontier. Assumption (4) places enough structure in the model to convert it to a common frontier with heteroscedastic disturbances, as in (8). This best practice technology can be used to identify technical inefficiency at the cost of introducing a special form of heteroscedasticity. This best practice technology is still an average best practice frontier but the assumption of homoscedasticity does not hold any more.

Another issue is whether or not the present model runs the risk of falsely identifying inefficiency as heteroscedasticity. This would be the equivalent of falsely identifying misspecification as inefficiency in the standard frontier approach. If the standard frontier applies, and a stochastic-coefficient frontier is adopted, then all the elements of matrix $\mathbf{\Omega}$ in (4) should be close to zero, in which case the disturbance e_{it} in (7) or (8) is homoscedastic and the model reduces to the standard frontier. Therefore, in the absence of other misspecifications (which would invalidate the traditional frontier as well) the stochastic-coefficient frontier cannot falsely identify inefficiency as heteroscedasticity or random variation in parameters.

The present model is also a reasonable alternative to standard frontiers that employ a large number of individual dummies to identify inefficient firms in frontiers with slope coefficients that are common across firms. In these models, dummy variables are used to get different intercept term for each firm, and then inefficiency is estimated as the deviation of each intercept from its maximum value. Although these fixed-effect models have certain advantages, they require the estimation of a large number of parameters. This is the reason why Battese and Coelli (1988) have proposed a frontier model with random intercept, and slope coefficients that are the same across firms. The present model can be thought of as a generalization of this model to the case of random slope parameters, when the assumption of common technology across firms cannot be maintained.

The log-likelihood function of model (7) is

$$L(\alpha,\bar{\boldsymbol{\beta}},\sigma,\theta;\mathbf{y},\mathbf{X}) = NT\ln\theta + (\theta^2/2)\sum_{t=1}^{T}\sum_{i=1}^{N} w_{it} + \sum_{t=1}^{T}\sum_{i=1}^{N}[\ln\Phi(\frac{-\varepsilon_{it}-\theta w_{it}}{w_{it}^{1/2}}) + \theta\varepsilon_{it}] \quad (9)$$

where $\varepsilon_{it} \equiv y_{it} - \alpha - \mathbf{x}_{it}'\bar{\boldsymbol{\beta}}$, $w_{it} \equiv \sigma^2 + \mathbf{x}_{it}'\boldsymbol{\Omega}\mathbf{x}_{it}$ $(i = 1,..,N,\ t = 1,..,T)$, $\Phi(\cdot)$ denotes the standard normal cumulative distribution function, and $\mathbf{y}$ and $\mathbf{X}$ are $NT\times 1$ and $NT\times(K-1)$ matrices in obvious notation. Maximum likelihood estimates can be computed by maximizing the likelihood function in (9) using standard gradient methods or the related E-M algorithm. Bayesian analysis of the model can be based on importance sampling (van den Broeck *et al*., 1994) by using the log-likelihood function in (8). In this paper, the Gibbs sampler is used to explore the posterior distribution of the model. The results developed in this paper, especially the conditional distributions implied by the joint posterior distribution can, however, be useful for researchers interested in maximizing the log-likelihood function using the E-M algorithm. Contrary to the E-M algorithm, however, the Gibbs sampler produces the entire set of posterior distributions of parameters and inefficiency measures, not just the mode of the likelihood function.

To proceed with Bayesian analysis of the model, (1) and (7) can be viewed as a hierarchical linear model *i.e.* $\boldsymbol{\beta}_i$'s are viewed as parameters with prior given by (4). Priors are also placed on the hyperparameters $\overline{\boldsymbol{\beta}}$ and $\boldsymbol{\Omega}$. Empirical researchers may often find it desirable to assume that $\boldsymbol{\Omega} = diag[\omega_1\, \omega_2 \ldots \omega_{K-1}]$ *i.e.* variance parameters are a priori independent. Although this assumption is not adopted in the present paper, inferences are developed for this special case as well.

A prior for the model in (1) and (7) can be specified as follows:

$$p(\sigma^2, \theta, \boldsymbol{\Omega}) \propto \sigma^{-(\underline{N}_1+1)} \exp(-\underline{q}_1 / 2\sigma^2) \theta^{\underline{N}_2 - 1} \exp(-\underline{q}_2 \theta) \,|\, \boldsymbol{\Omega} \,|^{-(K+\underline{\nu}+1)/2} \exp(-1/2 tr \boldsymbol{\Omega}^{-1} \underline{\mathbf{S}}) \quad (10)$$

The prior for α and $\overline{\boldsymbol{\beta}}$ is flat, i.e. $p(\alpha, \overline{\boldsymbol{\beta}}) \propto const.$ reflecting that the applied researcher is likely to have no prior information about parameter means or the researcher may not want to impose prior information. The prior for θ is *gamma*, the prior for σ is of the inverted *gamma* type, and the prior for $\boldsymbol{\Omega}$ is inverted Wishart (Zellner, 1971). If $\underline{q}_1 = 0$ and $\underline{N}_1 = 0$ the prior for σ would be a standard non-informative Jeffreys' prior, *i.e.* $p(\sigma) \propto \sigma^{-1}$. The presence of non-zero $\underline{q}_1$ is required to make the posterior distribution well defined (see Fernandez, Osiewalski and Steel, 1997). One may set $\underline{q}_1$ to a very small number (for example 10^{-6}).

When $\underline{N}_2 = 1$, the prior for θ is exponential with parameter $\underline{q}_2 \equiv -\ln r^*$. This prior has been introduced by van den Broeck *et al.* (1994) where it has been shown that r^* is prior median efficiency, by utilizing the definition $r_{it} \equiv \exp(-u_{it})$. The assumption $r^* = 1$ imposes the prior belief that all decision-making units are fully efficient. Such a prior belief is likely to be rather restrictive in practice. In addition Fernandez, Osiewalski and Steel, (1997, Proposition 1) show that with a non-informative prior for θ the posterior is not well defined. Finally, the restrictions $\underline{\nu} = 0$ and $\underline{\mathbf{S}} = \mathbf{0}$ provide a Jeffreys prior for the elements of covariance matrix $\boldsymbol{\Omega}$. In practice, one may set $\underline{\nu} = 1$ and $\underline{\mathbf{S}} = 10^{-6} \mathbf{I}_{K-1}$.

It must be mentioned that in the prior (10) $(\alpha, \overline{\boldsymbol{\beta}})$ are jointly independent of θ. This is a common assumption (see van den Broeck *et al*., 1994) and its motivation is in the fact that it appears impossible to see how inefficiency priors (via priors on θ) could depend on specific values of $y_i = \alpha + \mathbf{x}_i'\boldsymbol{\beta}_i + v_i - u_i$. Apparently, inefficiency measures (and thus θ) depend on the estimated frontier, and thus parameters $(\alpha, \overline{\boldsymbol{\beta}})$ as well, which would suggest that the priors of $(\alpha, \overline{\boldsymbol{\beta}})$ and θ could not be independent. This argument, however, presupposes that the data is available which cannot be the case when the prior is formulated. It is possible to set aside a portion of the sample (*i.e.* a number of firms for a number of periods) and use this sub-sample to deliver joint priors of $(\alpha, \overline{\boldsymbol{\beta}})$ and θ. The details of this approach would be useful to be explored in future research. It must be mentioned, however, that computations under this assumption would be complicated and the elegance of Gibbs sampling would be lost.

The posterior distribution corresponding to (4) and (7) is

$$p(\alpha, \overline{\boldsymbol{\beta}}, \{\boldsymbol{\beta}_i\}, \sigma, \boldsymbol{\Omega}, \theta, \mathbf{u} \mid \mathbf{y}, \mathbf{X}) \propto \theta^{NT+\underline{N}_2-1} \sigma^{-(NT+\underline{N}_1+1)} \exp(\ln \mathrm{r}^* \theta) \mid \boldsymbol{\Omega} \mid^{-(N+K+\underline{\nu})/2}$$
$$\prod_{i=1}^{N} \exp[-1/2\sigma^2 (y_{it} + u_{it} - \alpha - \mathbf{x}_{it}'\boldsymbol{\beta}_i)^2 - \theta u_{it}] \exp[-1/2(\boldsymbol{\beta}_i - \overline{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta}_i - \overline{\boldsymbol{\beta}}) - 1/2 tr \underline{\mathbf{S}}\boldsymbol{\Omega}^{-1}] \tag{11}$$

It is easy to prove that the integral of (11) with respect to latent inefficiencies $\{u_i\}$ as well as latent individual response coefficients $\{\boldsymbol{\beta}_i\}$ gives a posterior distribution whose kernel is the prior multiplied by the likelihood function, the logarithm of which is given by equation (9). The practice of treating latent variables as parameters is known as data augmentation, and its rationale will be explained in the next section.

3. Posterior analysis by Gibbs sampling

The Gibbs sampler (Gelfand and Smith, 1990, Smith and Roberts, 1993, and Casella and George, 1992, for a review) is an iterative scheme which utilizes the conditional distributions implied by the kernel of a posterior distribution $p(\mathbf{a} \mid \mathbf{y})$ where $\mathbf{a} \in \boldsymbol{\Theta} \subseteq \mathbf{R}^m$

denotes the parameter vector, $\mathbf{\Theta}$ is the parameter space, $\mathbf{y}$ denotes the data, and m stands for the number of parameters. To produce a (non-random) sample $\{\mathbf{a}^{(i)}, i = 1,.., M\}$ which converges in distribution to $p(\mathbf{a} \mid \mathbf{y})$ one starts from initial conditions $a^{(0)}$. Define $\mathbf{a}_{-j} = [a_1 \ldots a_{i-1}\ a_{i+1} \ldots a_{\mathrm{m}}]'$, $j = 1,..., m$, as the parameter whose j^{th} element is omitted. Then one uses the following iteration for $i = 1,..., M$:

$$\text{Draw } a_1^{(i)} \text{ from } p(a_1 \mid \mathbf{a}_{-1}^{(i-1)}, \mathbf{y})$$

$$\text{Draw } a_2^{(i)} \text{ from } p(a_2 \mid \mathbf{a}_{-2}^{(i-1)}, \mathbf{y})$$

$$\ldots.$$

$$\text{Draw } a_m^{(i)} \text{ from } p(a_m \mid \mathbf{a}_{-m}^{(i-1)}, \mathbf{y}),$$

The posterior expectation of the vector function of the parameters $\mathbf{f}(\mathbf{a})$ which is

$$E[\mathbf{f}(\mathbf{a}) \mid \mathbf{Y}] = \frac{\int_{\mathbf{\Theta}} \mathbf{f}(\mathbf{a}) p(\mathbf{a} \mid \mathbf{Y}) d\mathbf{a}}{\int_{\mathbf{\Theta}} p(\mathbf{a} \mid \mathbf{Y}) d\mathbf{a}} \tag{12}$$

can be estimated using the Monte Carlo average $M^{-1} \sum_{m=1}^{M} \mathbf{f}(\mathbf{a}^{(m)})$. This includes as special cases, estimation of posterior moments of any order, as well as estimation of marginal posterior densities of parameters or functions of the parameters. Data augmentation (Tanner and Wong, 1987) is an extension of the Gibbs sampler to the case where the posterior involves integrals with respect to a set of latent variables that cannot be computed analytically. Let $\boldsymbol{\xi}$ denote the latent variables. In this case, the posterior is $p(\mathbf{a}, \boldsymbol{\xi} \mid \mathbf{y})$. The latent variables can be thought of as parameters with appropriate priors, and Gibbs sampling can be applied to the simplified posterior $p(\mathbf{a}, \boldsymbol{\xi} \mid \mathbf{y})$. The marginal posterior $p(\mathbf{a} \mid \mathbf{y})$ (and its marginals) can be calculated in the usual way.

In order to use the Gibbs sampler, knowledge of the conditional distributions is required. The conditional distributions for the stochastic frontier model with random coefficients are presented below. Existence of the augmented posterior

$p(\alpha,\bar{\boldsymbol{\beta}},\{\boldsymbol{\beta}_i\},\sigma,\boldsymbol{\Omega},\theta,\mathbf{u}\,|\,\mathbf{y},\mathbf{X})$, its moments and existence of efficiency distributions and their moments is proved in the theorems of Appendix A. Functional forms of distributions are reviewed in Appendix B.

Conditional distribution of $\boldsymbol{\beta}_i$

The conditional distribution of $\boldsymbol{\beta}_i$ can be shown to be

$$\boldsymbol{\beta}_i\,|\,\alpha,\bar{\boldsymbol{\beta}},\sigma,\boldsymbol{\Omega},\theta,\mathbf{u},\mathbf{y},\mathbf{X}\sim \mathsf{N}_{K-1}(\hat{\boldsymbol{\beta}}_i,\hat{\mathbf{V}}_i),\ \text{i=1,..,N} \tag{13}$$

where $\mathsf{N}_p(\mathbf{m},\mathbf{V})$ denotes the $p-$variate normal distribution with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{V}$, and

$$\hat{\boldsymbol{\beta}}_i=\left(\frac{\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{x}'_{it}}{\sigma^2}+\boldsymbol{\Omega}^{-1}\right)^{-1}\left(\frac{\sum_{t=1}^{T}\mathbf{x}_{it}(y_{it}+u_{it}-\alpha)}{\sigma^2}+\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{\beta}}\right) \tag{14}$$

$$\hat{\mathbf{V}}_i=\left(\frac{\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{x}'_{it}}{\sigma^2}+\boldsymbol{\Omega}^{-1}\right)^{-1} \tag{15}$$

Although there are $NT(K-1)$ such parameters to draw for each Gibbs iteration, random number generation from their conditional distribution is straightforward.

Conditional distribution of σ^2

The conditional distribution of σ^2 is given by

$$\frac{\underline{q}_1+\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it}+u_{it}-\alpha-\mathbf{x}'_{it}\boldsymbol{\beta}_i)^2}{\sigma^2}\,|\,\alpha,\bar{\boldsymbol{\beta}},\{\boldsymbol{\beta}_i\},\boldsymbol{\Omega},\theta,\mathbf{u},\mathbf{y},\mathbf{X}\sim\chi^2(NT+\underline{N}_1) \tag{16}$$

The efficient way to generate from this distribution, is to notice that it is *gamma*$(1/2, NT/2)$ and then use standard simulation algorithms for the *gamma* distribution.

Conditional distribution of $\bar{\boldsymbol{\beta}}$ and α

(7) implies that the posterior conditional distribution of $\bar{\boldsymbol{\beta}}$ is

$$p(\bar{\boldsymbol{\beta}} \mid \alpha, \{\boldsymbol{\beta}_i\}, \sigma, \boldsymbol{\Omega}, \theta, \mathbf{u}, \mathbf{y}, \mathbf{X}) \propto \exp[-1/2 \sum_{i=1}^{N} (\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}})' \boldsymbol{\Omega}^{-1} (\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}})] \qquad (17)$$

implying that the conditional distribution of $\bar{\boldsymbol{\beta}}$ is

$$\bar{\boldsymbol{\beta}} \mid \alpha, \{\boldsymbol{\beta}_i\}, \sigma, \boldsymbol{\Omega}, \theta, \mathbf{u}, \mathbf{y}, \mathbf{X} \sim \mathsf{N}_{K-1}(N^{-1} \sum_{i=1}^{N} \boldsymbol{\beta}_i, N^{-1} \boldsymbol{\Omega}) \qquad (18)$$

The conditional distribution of α is

$$\alpha \mid \bar{\boldsymbol{\beta}}, \{\boldsymbol{\beta}_i\}, \sigma, \boldsymbol{\Omega}, \theta, \mathbf{u}, \mathbf{y}, \mathbf{X} \sim \mathsf{N}((NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} (y_{it} + u_{it} - \mathbf{x}_{it}' \bar{\boldsymbol{\beta}}), (NT)^{-1} \sigma^2) \qquad (19)$$

If one is willing to assume that the parameter covariance matrix $\boldsymbol{\Omega}$ is diagonal, and $\boldsymbol{\Omega} = diag[\omega_1, ..., \omega_{K-1}]$ then for each $j = 1,.., K-1$,

$$\bar{\beta}_j \mid \alpha, \{\boldsymbol{\beta}_i\}, \sigma, \boldsymbol{\Omega}, \theta, \mathbf{u}, \mathbf{y}, \mathbf{X} \sim \mathsf{N}(N^{-1} \sum_{i=1}^{N} \beta_{ij}, N^{-1} \omega_j) \qquad (20)$$

where $\bar{\beta}_j$ is the j^{th} element of $\bar{\boldsymbol{\beta}}$, and β_{ij} is the j^{th} element of $\boldsymbol{\beta}_i$. The $\bar{\boldsymbol{\beta}}_j$'s are conditionally independent because of the diagonal parameter covariance matrix. However, the $\bar{\beta}_j$'s are not independent in the marginal posterior distribution.

Conditional distribution of $\boldsymbol{\Omega}$

With the Jeffreys' prior for $\boldsymbol{\Omega}$, its posterior conditional distribution is

$$p(\boldsymbol{\Omega} \mid \alpha, \bar{\boldsymbol{\beta}}, \{\boldsymbol{\beta}_i\}, \sigma, \theta, \mathbf{u}, \mathbf{y}, \mathbf{X}) \propto |\boldsymbol{\Omega}|^{-(N+K+\underline{\nu})/2} \exp[-1/2tr\{\boldsymbol{\Omega}^{-1}(\mathbf{S}+\underline{\mathbf{S}})\}] \tag{21}$$

where $\mathbf{S} = \sum_{i=1}^{N} (\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}})(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}})'$, and "*tr*" denotes the matrix trace operator. The above is an inverted Wishart distribution (Zellner, 1971, page 395). Random number generation from this distribution is facilitated by the property that if $\mathbf{z}_1, \mathbf{z}_2, .., \mathbf{z}_N$ are (K-1)-dimensional random vectors, each distributed as *i.i.d* $\mathsf{N}_{K-1}(0, \mathbf{S}^{-1})$ then $(\sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i')^{-1}$ is distributed according to the $IW(\mathbf{S}, N, K-1)$ distribution, see Zellner (1971, p.389). Geweke (1996, section 3.1) proposes a similar but more efficient procedure for generating random numbers from the Wishart distribution.

When $\boldsymbol{\Omega}$ is assumed diagonal, the posterior conditional distribution of each diagonal element is given by

$$\frac{\underline{\mathbf{S}}_i + \sum_{i=1}^{N} [\beta_{ij} - \bar{\beta}_j]^2}{\omega_j} \mid \alpha, \bar{\boldsymbol{\beta}}, \{\boldsymbol{\beta}_i\}, \sigma, \theta, \mathbf{u}, \mathbf{y}, \mathbf{X} \sim \chi^2(N + \underline{\nu} + K - 2), \text{ j=1,.., } K\text{-1} \tag{22}$$

In the case of two regressors (K=2) in which case there is a single ω_j, the degrees of freedom in the above conditional are equal to N, as is the case in the usual normal, K-variate linear model with a single (disturbance) variance parameter.

Conditional distribution of u_{it}

These parameters are *a posteriori* independent, and the conditional distribution of the i*th* component is given by a truncated normal:

$$u_{it} \mid \alpha, \overline{\boldsymbol{\beta}}, \{\boldsymbol{\beta}_i\}, \sigma, \boldsymbol{\Omega}, \theta, \mathbf{y}, \mathbf{X} \sim \mathrm{N}(\alpha + \mathbf{x}'_{it}\overline{\boldsymbol{\beta}} - y_{it} - \theta w_{it}, w_{it}) \cdot \mathbf{1}(u_{it} \geq 0) \tag{23}$$

where $\mathbf{1}(A)$ is equal to one if event A is true, and zero otherwise. For the w_{it}'s see the discussion following equation (9). This conditional posterior distribution generalizes the truncated normal conditional distribution of these parameters in the homoscedastic normal-exponential stochastic frontier model (see Koop, Steel and Osiewalski, 1995). Generating random numbers from these distributions can be accomplished by rejection sampling. To draw from $N(\mu,\sigma^2)$ truncated below at zero, one may draw from the non-truncated distribution and retain the draw only if it is non-negative. This is efficient provided the mean μ is not "too negative" (say $\mu > -\sigma$). If $\mu < -\sigma$ this procedure becomes inefficient, in which case one may use acceptance sampling using an exponential source density, $p(x) = \lambda \exp(-\lambda x)\, 1(x \geq 0)$. Then, it can be shown that the optimal parameter of the exponential distribution is $\lambda^* = (x^* - \mu)/\sigma^2$ where x^* is the unique positive root of the quadratic $x^2 - \mu x - \sigma^2 = 0$. If $g(x) \equiv \lambda^* x - \frac{(x-\mu)^2}{2\sigma^2}$ the exponential draw should be accepted with probability $\exp\{g(x) - g(x^*)\}$.

Conditional distribution of θ

The conditional distribution of θ is *gamma*,

$$\theta \mid \alpha, \overline{\boldsymbol{\beta}}, \{\boldsymbol{\beta}_i\}, \sigma, \boldsymbol{\Omega}, \mathbf{u}, \mathbf{y}, \mathbf{X} \sim G(NT + \underline{N}_2, \sum_{t=1}^{T} \sum_{i=1}^{N} u_{it} - \ln r^*) \tag{24}$$

Random number generation from this distribution is straightforward.

Based upon these conditional distributions, the Gibbs sampler consists of iterating the following scheme.

1. Draw $\{\boldsymbol{\beta}_i\}$ using its conditional distribution in (13) for $i = 1,..,N$.

2. Draw σ using its conditional distribution in (16).

3. Draw $\overline{\boldsymbol{\beta}}$ and α using their conditional distributions in (18) and (19).

4. Draw $\boldsymbol{\Omega}$ using its conditional distribution in (21).

5. Draw $\{u_{it}\}$ using its conditional distribution in (23) for $i = 1,..,N$, $t = 1,..,T$.

6. Draw θ using its conditional distribution in (24).

The ordering of the parameters in the above scheme is arbitrary.

Alternative inefficiency distributions

Several alternative distributions for u_{it} have been proposed in connection with the traditional stochastic frontier model, the most prominent of which are the half-normal and *gamma* distributions. The methods proposed so far in this paper can be extended in a straightforward manner to allow for these alternative inefficiency distributions. Suppose $u_{it} \sim \mathrm{N}(0,\sigma_u^2)$ $(u_{it} \geq 0)$ *i.e.* u_{it} follows a half-normal distribution with probability density function

$$p(u_{it} \mid \sigma_u) = (\frac{\pi}{2}\sigma_u^2)^{-1/2} \exp(-\frac{u_{it}^2}{2\sigma_u^2}) \cdot \mathbf{1}(u_{it} \geq 0) \tag{25}$$

For parameter σ_u we may adopt the (conditionally conjugate) prior

$$p(\sigma_u) \propto \sigma_u^{-(\underline{N}_u+1)} \exp(-\frac{\underline{q}_u}{2\sigma_u^2}), \ \underline{N}_u, \sigma_u \geq 0 \tag{26}$$

The kernel posterior distribution in (11) would have to be modified as follows:

$$\begin{aligned} & p(\alpha, \overline{\boldsymbol{\beta}}, \{\boldsymbol{\beta}_i\}, \sigma, \boldsymbol{\Omega}, \sigma_u, \mathbf{u} \mid \mathbf{y}, \mathbf{X}) \propto \sigma_u^{-(\underline{N}_u+NT+1)} \sigma^{-(NT+\underline{N}_1+1)} \mid \boldsymbol{\Omega} \mid^{-(N+K+\underline{v})/2} \cdot \\ & \exp[-1/2\sigma^2 \sum_{t=1}^{T}\sum_{i=1}^{N} (y_{it} + u_{it} - \alpha - \mathbf{x}'_{it}\boldsymbol{\beta}_i)^2 - 1/2\sigma_u^2 \{\sum_{t=1}^{T}\sum_{i=1}^{N} u_{it}^2 + \underline{q}_u\}] \cdot \\ & \exp[-1/2(\boldsymbol{\beta}_i - \overline{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta}_i - \overline{\boldsymbol{\beta}}) - 1/2 tr \underline{\mathbf{S}}\boldsymbol{\Omega}^{-1}] \end{aligned} \tag{27}$$

It is clear that conditional posterior distributions of parameters $\alpha, \bar{\boldsymbol{\beta}}, \{\boldsymbol{\beta}_i\}, \sigma, \boldsymbol{\Omega}$ remain as before, and we have the following new results. For parameter σ_u the posterior conditional distribution is

$$\frac{\sum_{t=1}^{T}\sum_{i=1}^{N} u_{it}^2 + \underline{q}_u}{\sigma_u^2} \mid \alpha, \bar{\boldsymbol{\beta}}, \{\boldsymbol{\beta}_i\}, \sigma, \boldsymbol{\Omega}, \mathbf{u}, \mathbf{y}, \mathbf{X} \sim \chi^2(NT + \underline{N}_u) \tag{28}$$

For the inefficiency parameter we have

$$u_{it} \mid \alpha, \bar{\boldsymbol{\beta}}, \{\boldsymbol{\beta}_i\}, \sigma, \boldsymbol{\Omega}, \sigma_u, \mathbf{y}, \mathbf{X} \sim N(-\frac{\sigma_u^2 e_{it}}{\sigma^2 + \sigma_u^2}, \frac{\sigma^2 \sigma_u^2}{\sigma^2 + \sigma_u^2}),\ u_{it} \geq 0,\ i = 1,..,N,\ \ t = 1,..,T \tag{29}$$

where $e_{it} = y_{it} - \alpha - \mathbf{x}'_{it}\boldsymbol{\beta}_i$. Drawing random numbers from this distribution is straightforward given the analysis already provided. It is also possible to adopt a gamma distribution for inefficiency. The details of Bayesian efficiency analysis with gamma inefficiency in the traditional stochastic frontier model are provided in Tsionas (2000). Here, we may adopt a simplified version of the gamma distribution, known as the Erlang form that restricts the shape parameter to be integer. The probability density function of the inefficiency parameter is

$$p(u_{it} \mid \theta) = \frac{\theta^J}{\Gamma(J)} u_{it}^{J-1} \exp(-\theta u_{it}) \cdot \mathbf{1}(u_{it} \geq 0) \tag{30}$$

Here, J is a positive shape parameter that can be treated as constant. When $J = 1$ we obtain the exponential model. In practice, values of J =1,2,3 would be sufficient to allow generalization of the exponential model without running the risk of bringing u_{it} too close to normality so as to make it nearly indistinguishable from the normal distribution of v_{it}. The kernel posterior distribution in (7) has now to be modified as

$$p(\alpha,\overline{\boldsymbol{\beta}},\{\boldsymbol{\beta}_i\},\sigma,\boldsymbol{\Omega},\theta,\mathbf{u}\mid\mathbf{y},\mathbf{X})\propto\theta^{NT+\underline{N}_2-1}\sigma^{-(NT+\underline{N}_1+1)}\exp(\ln r^*\theta)\,|\,\boldsymbol{\Omega}\,|^{-(N+K+\underline{v})/2}\cdot$$
$$\exp[-1/2\sigma^2\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it}+u_{it}-\alpha-\mathbf{x}'_{it}\boldsymbol{\beta}_i)^2-\theta\sum_{t=1}^{T}\sum_{i=1}^{N}u_{it}]\cdot\{\prod_{t=1}^{T}\prod_{i=1}^{N}u_{it}\}^{J-1} \quad (31)$$
$$\exp[-1/2(\boldsymbol{\beta}_i-\overline{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta}_i-\overline{\boldsymbol{\beta}})-1/2tr\underline{\mathbf{S}}\boldsymbol{\Omega}^{-1}]$$

For parameter θ, the conditional posterior distribution is the same with the exponential model, and the conditional kernel posterior distribution of u_{it} becomes

$$p(u_{it}\mid\alpha,\overline{\boldsymbol{\beta}},\{\boldsymbol{\beta}_i\},\sigma,\boldsymbol{\Omega},\theta,\mathbf{y},\mathbf{X})\propto u_{it}^{J-1}f_N(u_{it}\mid\alpha+\mathbf{x}'_{it}\boldsymbol{\beta}_i-y_{it}-\theta w_{it},\ w_{it})\cdot\mathbf{1}(u_{it}\geq 0) \quad (32)$$

where all variables and parameters have been defined in connection with the exponential model. Drawing random numbers from the posterior conditional distribution of u_{it} above can be accomplished using efficient numerical procedures presented in Tsionas (2000).

Efficiency measurement

Following van den Broeck *et al.* (1994) *firm-specific efficiency measurement* is based on the conditional distribution $p(u_{it}\mid\alpha,\overline{\boldsymbol{\beta}},\{\boldsymbol{\beta}_i\},\sigma,\boldsymbol{\Omega},\theta,\mathbf{y},\mathbf{X})=p(u_{it}\mid\alpha,\overline{\boldsymbol{\beta}},\sigma,\boldsymbol{\Omega},\theta,\mathbf{y},\mathbf{X})$. Efficiency of the i*th* firm is defined as $r_{it}=\exp(-u_{it})$. Using a change of variables, the distribution of efficiency conditionally on parameters and the data is given by

$$p(r_{it}\mid\alpha,\overline{\boldsymbol{\beta}},\sigma,\boldsymbol{\Omega},\theta,\mathbf{y},\mathbf{X})=p(u_{it}\mid\alpha,\overline{\boldsymbol{\beta}},\sigma,\boldsymbol{\Omega},\theta,\mathbf{y},\mathbf{X})r_{it}^{-1} \quad (33)$$

evaluated at $u_{it}=-\ln(r_{it})$. The distribution $p(u_{it}\mid\alpha,\overline{\boldsymbol{\beta}},\sigma,\boldsymbol{\Omega},\theta,\mathbf{y},\mathbf{X})$ is the posterior conditional distribution of u_{it} given the parameters, and was given in (23). Accounting for parameter uncertainty results in integrating out the parameters to obtain

$$p(r_{it}\mid\mathbf{y},\mathbf{X})=\int p(r_{it}\mid\alpha,\overline{\boldsymbol{\beta}},\sigma,\boldsymbol{\Omega},\theta,\mathbf{y},\mathbf{X})d\alpha d\overline{\boldsymbol{\beta}}d\sigma d\boldsymbol{\Omega}d\theta \quad (34)$$

This integral is not available analytically but can be approximated using the Monte Carlo estimate

$$\hat{p}(r_{it} \mid \mathbf{y}, \mathbf{X}) = M^{-1} \sum_{m=1}^{M} p(r_{it} \mid \alpha^{(m)}, \overline{\boldsymbol{\beta}}^{(m)}, \sigma^{(m)}, \boldsymbol{\Omega}^{(m)}, \theta^{(m)}, \mathbf{y}, \mathbf{X}) \tag{35}$$

where superscript m denotes the value of the parameter during the m*th* iteration of the Gibbs sampler. An estimate of firm-specific efficiency can be obtained using the mean of this distribution, $\hat{r}_{it} = \int_0^1 \hat{p}(r_{it} \mid \mathbf{y}, \mathbf{X})$ which can be computed using a Simpson's rule once values of $p(r_{it} \mid \mathbf{y}, \mathbf{X})$ have been computed pointwise.

4. Application

We consider the data collected by Christensen and Greene (1976) for N =123 electric utility companies in the United States in 1970. The data are listed in the Appendix to Greene (1990) and have been used by van den Broeck *et al.* (1994). There are three production factors (labor, capital and fuel with respective prices p_L, p_K, p_F) and the specification of the cost function is

$$y_i = -\beta_{0i} - \beta_{1i} \ln Q_i - \beta_{2i} \ln^2 Q_i - \beta_{3i} \ln(p_{Li} / p_{Fi}) - \beta_{4i} \ln(p_{Ki} / p_{Fi}) + v_i - u_i,\ i = 1,.., N \tag{36}$$

where $y_i = -\ln(C_i / p_{Fi})$, Q_i denotes output and C_i is the cost of the *ith* firm. This is a Cobb-Douglas cost function but permits returns to scale to vary with output. van den Broeck *et al.* (1994) claim that r^* =0.875 is a reasonable value for prior median efficiency. In stochastic frontier models (more so in cross-sectional frontier studies) the results show some sensitivity to the choice of this parameter. If panel data are available and one is willing to make the assumption that efficiency is constant over time the importance of r^* for posterior inferences is likely to be more limited. What the researcher judges to be a reasonable value for prior median efficiency should guide the choice of r^*. Such information may come from previous studies or simply the investigator's beliefs. Regarding values of the remaining prior hyperparameters, I have specified $\underline{N}_1 = \underline{N}_2 = 1$, $\underline{q}_1 = \underline{q}_2 = 10^{-6}$, $\underline{\nu} = 1$, and $\underline{\mathbf{S}} = 10^{-6}\mathbf{I}_{K-1}$.

Gibbs sampling with data augmentation has been implemented for this data set using 10,000 iterations, 5,000 of which were discarded to mitigate start-up effects. As it seems to be the rule with stochastic frontier models, convergence was not found to be a problem (Koop *et al*., 1995, Koop and Steel, 2001). Gibbs samplers were started from different initial conditions (corresponding to different values of θ, σ and Ω) and their behavior was compared after a burn-in period of 5,000 iterations. The results were reasonably robust. Formal convergence diagnostics based on multiple Markov chains (Gelman and Rubin, 1992 or Brooks and Gelman, 1996) could not reject that Gibbs sampling output has converged to the posterior distribution within 5,000 iterations. The convergence results are available on request.

The empirical results (posterior means and posterior standard deviations) are reported in Table 1. The fixed coefficients stochastic frontier model has been estimated by van den Broeck *et al*. (1994) using importance sampling techniques. Their results are also reported for comparison purposes. The diagonal elements of $\mathbf{\Omega}$, the parameter correlation matrix (derived from the parameter covariance matrix $\mathbf{\Omega}$) along with posterior standard deviations, are also reported. Parameter variances indicate considerable heterogeneity of β_3 and β_4 (the coefficients of capital and labor) which range from 0.03 to 0.206, and from 0.005 to 0.121 respectively.

Posterior means of cost function parameters are reasonably close for both models. The largest differences pertain to posterior means of σ^2, standard errors of cost function parameters and θ. Differences in σ^2 and standard errors can be attributed to the heteroscedastic nature of the stochastic frontier model. Differences in θ are more important and carry implications for efficiency measurement. According to van den Broeck *et al*. (1994) the posterior mean of θ is 11.27 (standard deviation 3.31) while according to the stochastic coefficient frontier model it is 75.12 (standard deviation 19.11). If we define $r_i = \exp(-u_i)$ to be individual efficiency for the i^{th} firm, the posterior predictive distribution of this quantity is given by

$$f(r \mid \mathbf{y}, \mathbf{X}) = \int_0^{\infty} f(u_i \mid \theta) p(\theta \mid \mathbf{y}, \mathbf{X}) d\theta = E_{\theta}(\theta r^{\theta - 1} \mid \mathbf{y}, \mathbf{X}) \tag{37}$$

where E_{θ} signifies posterior expectation with respect to θ, see van den Broeck *et al.* (1994). The above quantity is posterior to the data on all observed firms but prior to the yet unobserved output of some firm whose efficiency is sought. This is the Bayesian counterpart of the classical characteristic of "average" (as opposed to individual) efficiency.

The random coefficient model implies that near perfect efficiency is about seven times more likely compared to the fixed coefficients stochastic frontier model. Indeed, plots of the predictive density imply that values greater than 0.99 are by far more likely in the random coefficient, rather than the fixed coefficient stochastic frontier. Since θ can be as high as about 110, such plots are not very helpful. What is important is that such large values of θ imply nearly perfect efficiency.

Model comparison involving the fixed coefficient and the random coefficient frontier model can be accomplished using Bayes factors. If $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ denote the parameters of two models M_1 and M_2 whose prior distributions are $p(\boldsymbol{\theta}_1)$ and $p(\boldsymbol{\theta}_2)$, and their likelihood functions are $L(\boldsymbol{\theta}_1; \mathbf{y})$ and $L(\boldsymbol{\theta}_2; \mathbf{y})$, the Bayes factor in favor of M_1 and against M_2 is given by

$$B_{12} = \frac{\int L(\boldsymbol{\theta}_1; \mathbf{y}) p(\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1}{\int L(\boldsymbol{\theta}_2; \mathbf{y}) p(\boldsymbol{\theta}_2) d\boldsymbol{\theta}_2} \tag{38}$$

Computing Bayes factors is nontrivial, except in the simplest of models. Following the approach of Chib (1995) a simulation approximation to the Bayes factor and the marginal likelihood can be computed provided the conditional distributions have known integration constants and drawing random numbers from the conditional distributions is feasible. This approach requires repeated Gibbs sampling from each conditional distribution. In the context of random coefficient stochastic frontier models, this is feasible. The Bayes factor in favor of the fixed coefficient frontier model and against the stochastic coefficient frontier model was computed to be 3.102×10^{-6} which suggests a clear superiority of the stochastic coefficient frontier model. It should be noted that a non-informative prior can be specified for α and $\bar{\boldsymbol{\beta}}$ since these parameters appear in both the random-coefficient and the fixed-coefficient specifications.

What does this result mean in economic terms? The reason why derivation of efficiency measures and identification of inefficient firms is an important task in stochastic frontier analysis is that such firms will go bankrupt in the long run. On the other hand, many applied researchers would feel that the finding of perfect efficiency is reasonable in the long run but questionable in the short run. The proper identification of efficient and inefficient firms is an empirical issue, and depends crucially on the form of production function and associated assumptions about similar parameters in the sample. When parameters are not the same across firms, usual stochastic frontiers based on a common technology will provide misleading efficiency measures, for essentially the same reasons that heteroscedastic frontiers will give biased results if heteroscedasticity is ignored, see Caudill, Ford and Gropper, 1995. The stochastic-coefficient frontier is one way of dealing with the problem of technological differentials across firms, which could be responsible for inefficiency findings in environments, which are essentially efficient but differentiated in terms of production capabilities. In such environments it is quite likely that the forces of competition work strong enough to make sure firms provide every effort to maximize efficiency. Of course, this statement is not universal, and leaves open the possibility of finding inefficiency even after technological differences have been accounted for. The extent to which this happens is, again, an empirical issue.

Regarding firm-specific efficiency measures $\hat{r}_i$, these were computed for each firm. The minimum value was close to 0.99 suggesting near perfect efficiency once heterogeneous technology has been taken into account using the random coefficient approach. These results are different with what has been obtained in previous research using this data set (van den Broeck *et al.,* 1994) as expected. Van den Broeck *et al.* (1994, Table 6) find mean efficiencies which range from 0.830 to 0.910 depending on the model specification (when $r^* = 0.875$) and even lower when $r^* = 0.50$ (Table 9). The data suggest, however, that part of this estimated inefficiency may be technological differences and firms are almost fully efficient.

Conclusions

The paper developed a random coefficient stochastic frontier model to account for firm heterogeneity in efficiency measurement. Accounting for such heterogeneity is important because, as a rule, different firms face different production possibilities. Ignoring this reality may be seriously misleading as far as efficiency is concerned. The paper developed a Gibbs sampling approach to posterior analysis of stochastic frontier models with random coefficients. The new methods were applied to the electric utility data of Christensen and Greene (1976), previously analyzed by Greene (1990) and van den Broeck *et al*. (1994) using conventional, fixed-coefficient stochastic frontier models.

References

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

Akhavein JD, Swamy PAVB, Taubman SB, and Singamsetti RN. 1997. A general method of deriving the inefficiencies from a profit function. *Journal of Productivity Analysis* **8**: 71-93

Battese G.E., Coelli TJ. 1988. Prediction of firm-level technical efficiencies with a generalised frontier production function and panel data. *Journal of Econometrics* **38**: 387-399.

Bauer PW. 1990. Recent developments in the econometric estimation of frontiers. *Journal of Econometrics* **46**: 39-56.

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

Brooks S, Gelman A. 1996. General methods for monitoring convergence of iterative simulations. Manuscript.

Casella G, George E. 1992. Explaining the Gibbs sampler. *The American Statistician* **46**: 167-174.

Caudill, SB, Ford JM. 1993. Biases in frontier estimation due to heteroscedasticity. *Economics Letters* **41**: 17-20.

Caudill SB, Ford JM, Gropper DM. 1995. Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business and Economic Statistics* **13**: 105-111.

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

Christensen, LR, Greene WH. 1976. Economies of scale in U.S. electric power generation. *Journal of Political Economy* **84**: 655-676.

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

Gelfand AE, Smith AFM. 1990. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**: 398-409.

Gelman A, Rubin D. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* **7**: 457-511

Geweke J. 1996. Simulation-based Bayesian inference for economic time series, Research Department, Federal Reserve Bank of Minneapolis, Working Paper 570.

Greene WH. 1990. A gamma-distributed stochastic frontier model. *Journal of Econometrics* **46**: 141-163.

Hildreth C, Houck JP. 1968. Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association* **63**: 584-595.

Kalirajan KP, Obwona MB. 1994. Frontier production function: The stochastic coefficients approach. *Oxford Bulletin of Economics and Statistics* **56**: 87-96.

Kalirajan KP, Shand RT. 1999. Frontier production functions and technical efficiency measures. *Journal of Economic Surveys* **13**: 149-172.

Koop G, Osiewalski J, Steel MFJ. 1997. Bayesian efficiency analysis through individuals effects: Hospital cost frontiers. Journal of Econometrics **76**: 77-105.

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

Koop G, Steel MFJ, Osiewalski J. 1995. Posterior analysis of stochastic frontier models using Gibbs sampling. *Computational Statistics* **10**: 353-373.

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

Reinganum JF. 1989. The timing of innovation: Research, Development and Diffusion. In *Handbook of Industrial Organization*. Schmalensee, R, Willig RD (eds). Amsterdam: Elsevier Science Publishers.

Roberts GO, Smith AFM 1994. Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications* **49**: 207-216.

Smith AFM, Roberts CO. 1993. Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society* **B55**: 3-23.

Stevenson RE. 1990. Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics* **13**: 57-66.

Swamy PAVB. 1970. Efficient inference in random coefficient regression models. Econometrica 38, 311-323.

Swamy P A V B, Tavlas GS. 1995. Random coefficient models: Theory and applications. *Journal of Economic Surveys* **9**: 165-196.

Swamy PAVB, Tavlas GS. 2001. Random coefficient models. In *A Companion to Theoretical Econometrics*. Baltagi B (ed). Blackwell, 410-428.

Tanner MA, Wong W-H. 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**: 528-550.

Tsionas EG. 2000. Full likelihood inference in normal-gamma stochastic frontier models. *Journal of Productivity Analysis* **13**: 179-201.

Zellner A. 1971. *Introduction to Bayesian inference in econometrics*. Wiley: New York.

Table 1. Empirical results for random coefficient stochastic frontier

Parameter	*Random Coefficients*[(a)]	*Fixed Coefficients*[(b)]
$\bar{\beta}_0$	-7.416 (.381)	-7.479 (.345)
$\bar{\beta}_1$	.445 (.034)	.428 (.042)
$\bar{\beta}_2$	.023 (.002)	.029 (.003)
$\bar{\beta}_3$	.247 (.068)	.249 (.065)
$\bar{\beta}_4$	.043 (.049)	.045 (.062)
σ^2	.0061 (.0033)	.013 (.004)
θ	75.12 (19.11)	11.27 (3.31)

Notes: Posterior standard errors appear in parentheses. Parameter means are denoted by $\bar{\beta}_i$, σ^2 stands for the equation variance and θ for the parameter of the inefficiency exponential distribution.

APPENDIX A. Gibbs sampling with data augmentation using 10,000 iterations and discarding 5,000 to mitigate start-up effects.

(b) Taken from van den Broeck *et al.* (1994, p.294, Table 4).

Posterior estimates of parameter variances

β_1	β_2	β_3	β_4
.00949 (.0052)	$3.37\ 10^{-5}$ ($1.58\ 10^{-5}$)	.107 (.056)	.071 (.037)

Notes: These are the diagonal elements of the parameter covariance matrix $\mathbf{\Omega}$. Posterior standard errors in parentheses.

Posterior estimates of parameter correlation matrix

	β_1	β_2	β_3	β_4
β_1	1.0	-.898 (.089)	.612 (.211)	-.933 (.178)
β_2		1.0	-.895 (.177)	.733 (.200)
β_3			1.0	-.389 (.288)
β_4				1.0

Notes: The parameter correlation matrix is derived from the parameter covariance matrix $\mathbf{\Omega}$. Posterior standard deviations are reported in parentheses.

APPENDIX A. Existence of posterior distribution and posterior moments

In what follows it is assumed that parameter $\alpha = 0$ without any loss of generality in order to ease notation. FOS denotes Fernandez, Osiewalski and Steel (1997). Furthermore, let $K' = K - 1$.

Theorem 1. Assume the following prior

$$p(\sigma^2, \theta, \mathbf{\Omega}) \propto \sigma^{-(\underline{N}_1+1)} \exp(-\underline{q}_1 / 2\sigma^2)\theta^{\underline{N}_2-1} \exp(-\underline{q}_2\theta) \,|\, \mathbf{\Omega} \,|^{-(\underline{\nu}+1)/2} \exp(-1/2 tr \mathbf{\Omega}^{-1} \underline{\mathbf{S}}) \quad \text{(A.1)}$$

where $\underline{N}_1, \underline{N}_2, \underline{\nu} \geq 0$, $\underline{q}_1, \underline{q}_2 > 0$ and $\underline{\mathbf{S}}$ is a positive definite matrix. Then the augmented posterior exists.

Proof. We have to show that the augmented posterior has a finite integral. The integral of the augmented posterior is given by

$$\int p(\overline{\boldsymbol{\beta}}, \theta, \sigma, \mathbf{\Omega}, \{\boldsymbol{\beta}_i\}, \mathbf{u} \,|\, \mathbf{y}, \mathbf{X}) d\overline{\boldsymbol{\beta}} d\theta d\sigma d\mathbf{\Omega} d\{\boldsymbol{\beta}_i\} d\mathbf{u} \propto$$
$$\int \theta^{N+\underline{N}_2-1} \sigma^{-(N+\underline{N}_1+1)} \exp\{-[\underline{q}_1 + \sum_{i=1}^{N} (y_i + u_i - \mathbf{x}_i' \boldsymbol{\beta}_i)^2] / 2\sigma^2 - [\underline{q}_2 + \sum_{i=1}^{N} u_i]\theta\}$$
$$|\, \mathbf{\Omega} \,|^{-(N+\underline{\nu}+1)/2} \exp\{-1/2 tr[\underline{\mathbf{S}} + \sum_{i=1}^{N} (\boldsymbol{\beta}_i - \overline{\boldsymbol{\beta}})(\boldsymbol{\beta}_i - \overline{\boldsymbol{\beta}})'] \mathbf{\Omega}^{-1}\} d\overline{\boldsymbol{\beta}} d\theta d\sigma d\mathbf{\Omega} d\{\boldsymbol{\beta}_i\} d\mathbf{u} \quad \text{(A.2)}$$

After integrating with respect to σ, θ and $\mathbf{\Omega}$ using properties of the inverted *gamma*, gamma and inverted Wishart distributions the above is proportional to

$$\int \{\underline{q}_1 + \sum_{i=1}^{N} (y_i + u_i - \mathbf{x}_i' \boldsymbol{\beta}_i)^2\}^{-(N+\underline{N}_1)/2} \{\underline{q}_2 + \sum_{i=1}^{N} u_i\}^{-(N+\underline{N}_2)} \,|\, \underline{\mathbf{S}} + \sum_{i=1}^{N} (\boldsymbol{\beta}_i - \overline{\boldsymbol{\beta}})(\boldsymbol{\beta}_i - \overline{\boldsymbol{\beta}})' \,|^{-(N+\underline{\nu})/2} d\overline{\boldsymbol{\beta}} d\{\boldsymbol{\beta}_i\} d\mathbf{u}$$
$$\text{(A.3)}$$

Each term of the integrand is bounded below if $\underline{q}_1, \underline{q}_2 > 0$ and $\underline{\mathbf{S}}$ is a positive definite matrix, so the integral is finite. ###

Theorem 2. The posterior does not exist with a non-informative prior on σ, θ or Ω.

Proof. We will consider the posterior $p(\overline{\boldsymbol{\beta}}, \sigma, \theta, \mathbf{\Omega}, \mathbf{u} \,|\, \mathbf{y}, \mathbf{X})$. We have

$$\int p(\overline{\boldsymbol{\beta}}, \sigma, \theta, \mathbf{\Omega}, \mathbf{u} \,|\, \mathbf{y}, \mathbf{X}) d\overline{\boldsymbol{\beta}} d\sigma d\theta d\mathbf{\Omega} d\mathbf{u} \propto$$
$$\int \theta^N \prod_{i=1}^{N} (\sigma^2 + \mathbf{x}_i' \mathbf{\Omega} \mathbf{x}_i)^{-1/2} \exp\{-\sum_{i=1}^{N} \frac{(y_i + u_i - \mathbf{x}_i' \overline{\boldsymbol{\beta}})^2}{2(\sigma^2 + \mathbf{x}_i' \mathbf{\Omega} \mathbf{x}_i)} - \theta \sum_{i=1}^{N} u_i\} p(\sigma, \theta, \mathbf{\Omega}) d\overline{\boldsymbol{\beta}} d\sigma d\theta d\mathbf{\Omega} d\mathbf{u}$$
$$\text{(A.4)}$$

where $p(\sigma,\theta,\mathbf{\Omega}) \propto \sigma^{-1}\theta^{-1} |\mathbf{\Omega}|^{-(K'+1)/2}$ and $K' \equiv K-1$. Let $\mathbf{\Omega S} = \mathbf{S\Lambda}$ be the spectral decomposition of $\mathbf{\Omega}$ where $\mathbf{S}$ is the orthogonal matrix of its eigenvectors and $\mathbf{\Lambda}$ the diagonal matrix that contains its eigenvalues on the diagonal. Define $\mathbf{z}_i = \mathbf{S}'\mathbf{x}_i$. Since $\mathbf{x}_i'\Omega\mathbf{x}_i \geq 0$ and $\mathbf{x}_i'\mathbf{\Omega}\mathbf{x}_i \leq \lambda_{\max}\mathbf{z}_i'\mathbf{z}_i \leq D$ where D is a given positive bound, a lower bound to the above integral is

$$\int \theta^N (\sigma^2 + D)^{-N/2} \exp\{-(\mathbf{y}+\mathbf{u}-\mathbf{X}\overline{\boldsymbol{\beta}})'(\mathbf{y}+\mathbf{u}-\mathbf{X}\overline{\boldsymbol{\beta}})/2\sigma^2 - \theta\sum_{i=1}^{N} u_i\} p(\sigma,\theta,\mathbf{\Omega}) d\overline{\boldsymbol{\beta}} d\sigma d\theta d\mathbf{\Omega} d\mathbf{u} \tag{A.5}$$

Let

$$Q(\overline{\boldsymbol{\beta}},\mathbf{u}) = \int (\sigma^2 + D)^{-N/2} \exp\{-(\mathbf{y}+\mathbf{u}-\mathbf{X}\overline{\boldsymbol{\beta}})'(\mathbf{y}+\mathbf{u}-\mathbf{X}\overline{\boldsymbol{\beta}})/2\sigma^2\} p(\sigma) d\sigma \leq \int \sigma^{-(N+1)} \exp\{-(\mathbf{y}+\mathbf{u}-\mathbf{X}\overline{\boldsymbol{\beta}})'(\mathbf{y}+\mathbf{u}-\mathbf{X}\overline{\boldsymbol{\beta}})/2\sigma^2\} d\sigma < \infty \tag{A.6}$$

The integral in (A.5) after integrating with respect to θ can be written as

$$\int Q(\overline{\boldsymbol{\beta}},\mathbf{u}) \{\sum_{i=1}^{N} u_i\}^{-N} p(\mathbf{\Omega}) d\overline{\boldsymbol{\beta}} du d\mathbf{\Omega} \tag{A.7}$$

which diverges since $\int \{\sum_{i=1}^{N} u_i\}^{-N} du$ does not exist. ###

Theorem 3. If $\mathbf{u} \in \mathbf{U}$ and $\int_{\overline{\mathbf{U}}} p(\mathbf{u}) d\mathbf{u} = \infty$ for some bounded $\overline{\mathbf{U}} \subseteq \mathbf{U}$ the posterior does not exist and the predictive distribution $p(\mathbf{y}) = \infty$ for all $\mathbf{y} \in \mathbf{R}^N$.

Proof. The proof closely follows FOS, proof of Theorem 2. With a prior $p(\mathbf{u})$ for the latent inefficiency terms, a slight modification of arguments leading to (A.5) gives that $\int p(\overline{\boldsymbol{\beta}},\sigma,\mathbf{\Omega},\mathbf{u} | \mathbf{y},\mathbf{X}) d\overline{\boldsymbol{\beta}} d\sigma d\mathbf{\Omega} d\mathbf{u}$ is proportional to a function with lower bound given by

$$\int p(\mathbf{u}) \int (\sigma^2 + D)^{-N/2} \exp\{-(\mathbf{y}+\mathbf{u}-\mathbf{X}\overline{\boldsymbol{\beta}})'(\mathbf{y}+\mathbf{u}-\mathbf{X}\overline{\boldsymbol{\beta}})/2\sigma^2\} p(\sigma,\mathbf{\Omega}) d\overline{\boldsymbol{\beta}} d\sigma d\mathbf{\Omega} d\mathbf{u} \propto \int p(\mathbf{u}) Q(\overline{\boldsymbol{\beta}},\mathbf{u}) p(\mathbf{\Omega}) d\overline{\boldsymbol{\beta}} d\mathbf{\Omega} d\mathbf{u} \tag{A.8}$$

Since $\overline{\mathbf{U}}$ is bounded and $Q(\overline{\boldsymbol{\beta}},\mathbf{u})$ is bounded in $\overline{\boldsymbol{\beta}}$ we are left with $\int_{\overline{\mathbf{U}}} p(\mathbf{u}) d\mathbf{u} = \infty$, by assumption. ###

Theorem 4. For $0 \leq m < N - K'$, the posterior expectation $E(\overline{\beta}_l^m | \mathbf{y},\mathbf{X})$ exists with informative priors on σ,θ and $\mathbf{\Omega}$.

Proof. We consider

$$\int |\bar{\beta}_l^m| \, p(\bar{\boldsymbol{\beta}},\sigma,\theta,\boldsymbol{\Omega},\mathbf{u} \mid \mathbf{y},\mathbf{X}) d\bar{\boldsymbol{\beta}} d\sigma d\theta d\boldsymbol{\Omega} d\mathbf{u} \propto$$

$$\int |\bar{\beta}_l^m| \theta^N \prod_{i=1}^{N} (\sigma^2 + \mathbf{x}_i'\boldsymbol{\Omega}\mathbf{x}_i)^{-1/2} \exp\{-\sum_{i=1}^{N} \frac{(y_i + u_i - \mathbf{x}_i'\bar{\boldsymbol{\beta}})^2}{2(\sigma^2 + \mathbf{x}_i'\boldsymbol{\Omega}\mathbf{x}_i)} - \theta \sum_{i=1}^{N} u_i\} p(\sigma,\theta,\boldsymbol{\Omega}) d\bar{\boldsymbol{\beta}} d\sigma d\theta d\boldsymbol{\Omega} d\mathbf{u} \tag{A.9}$$

Let $V_i \equiv V_i(\sigma,\boldsymbol{\Omega}) = \sigma^2 + \mathbf{x}_i'\boldsymbol{\Omega}\mathbf{x}_i$, $y_i^* = y_i V_i^{-1/2}$, $\mathbf{x}_i^* = \mathbf{x}_i V_i^{-1/2}$, $u_i^* = u_i V_i^{-1/2}$ and $V(\sigma,\boldsymbol{\Omega}) = diag[V_i(\sigma,\boldsymbol{\Omega}), i = 1,..., N]$. Define also $\bar{\boldsymbol{\beta}}^* = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}(\mathbf{y}^* + \mathbf{u}^*)$ and $(N-K')s_*^2 = (\mathbf{y}^* + \mathbf{u}^* - \mathbf{X}^*\bar{\boldsymbol{\beta}}^*)'(\mathbf{y}^* + \mathbf{u}^* - \mathbf{X}^*\bar{\boldsymbol{\beta}}^*)$.

It follows that

$$\int |\bar{\beta}_l^m| \, p(\bar{\boldsymbol{\beta}},\sigma,\theta,\boldsymbol{\Omega},\mathbf{u} \mid \mathbf{y},\mathbf{X}) d\bar{\boldsymbol{\beta}} d\sigma d\theta d\boldsymbol{\Omega} d\mathbf{u} \propto$$

$$\int |\bar{\beta}_l^m| \theta^N \prod_{i=1}^{N} V_i(\sigma,\boldsymbol{\Omega})^{-1/2} \exp\{-1/2(\mathbf{y} + \mathbf{u} - \mathbf{X}\bar{\boldsymbol{\beta}})' V(\sigma,\boldsymbol{\Omega})^{-1} (\mathbf{y} + \mathbf{u} - \mathbf{X}\bar{\boldsymbol{\beta}}) - \theta \sum_{i=1}^{N} u_i\} p(\sigma,\theta,\boldsymbol{\Omega}) d\bar{\boldsymbol{\beta}} d\sigma d\theta d\boldsymbol{\Omega} d\mathbf{u} =$$

$$\int |\bar{\beta}_l^m| \theta^N \prod_{i=1}^{N} V_i(\sigma,\boldsymbol{\Omega})^{-1/2} \exp\{-1/2[(N-K')s_*^2 + (\bar{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}^*)'\mathbf{X}^{*\prime}\mathbf{X}^*(\bar{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}^*)] - \theta \sum_{i=1}^{N} u_i\} p(\sigma,\theta,\boldsymbol{\Omega}) d\bar{\boldsymbol{\beta}} d\sigma d\theta d\boldsymbol{\Omega} d\mathbf{u} \leq$$

$$\int \sigma^{-N} \exp\{-1/2(N-K')s_*^2\} |\bar{\beta}_l^m| f_N(\bar{\boldsymbol{\beta}}; \bar{\boldsymbol{\beta}}^*, s_*^2(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}) \exp(-\theta \sum_{i=1}^{N} u_i) p(\sigma,\theta,\boldsymbol{\Omega}) d\bar{\boldsymbol{\beta}} d\sigma d\theta d\boldsymbol{\Omega} d\mathbf{u} \tag{A.10}$$

With a proper prior $p(\sigma,\theta,\boldsymbol{\Omega})$, it follows that $E(\bar{\beta}_l^m \mid \mathbf{y},\mathbf{X})$ exists because it is taken against a normal distribution, so it is well defined by standard existence results for the heteroscedastic linear model. ###

Theorem 5. If $m > -(N-K')$ then the posterior expectation $E(\sigma^m \mid \mathbf{y},\mathbf{X})$ exists.

Proof. The proof, based on proof of Theorem 1, is easy.

Theorem 6. Under the conditions of Theorem 1, all marginal moments of positive order of efficiency measures are finite, *i.e.*

$$E\{\prod_{i=1}^{N} \exp(-m_i u_i \mid \mathbf{y},\mathbf{X})\} < \infty .$$

Proof. Since $0 < \exp(-m_i u_i) \leq 1$ the proof is immediate. ###

Based on Theorem 1, with a proper prior on θ and $\mathbf{\Omega}$, existence of product moments of θ and $\mathbf{\Omega}$ is also immediate.

APPENDIX B. Distributions

If the scalar random variable X follows the normal distribution with mean μ and variance σ^2, this is denoted by $X \sim \mathsf{N}(\mu,\sigma^2)$ and its probability density function is given by

$$f_N(x \mid \mu,\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp[-(x-\mu)/2\sigma^2] \qquad \text{(B.1)}$$

If X is a vector random variable in R^k, normally distributed with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{V}$, this is denoted by $X \sim \mathsf{N}_k(\mathbf{m},\mathbf{V})$ and its probability density function is given by

$$f_{\mathsf{N},k}(x \mid \mathbf{m},\mathbf{V}) = (2\pi)^{-1/2} \mid \mathbf{V} \mid^{-1/2} \exp[-1/2(x-\mathbf{m})'\mathbf{V}^{-1}(x-\mathbf{m})] \qquad \text{(B.2)}$$

If the scalar, non-negative random variable X follows the *gamma* distribution with shape parameter P and scale parameter θ, this is denoted by $X \sim G(P,\theta)$ and its probability density function is given by

$$f_G(x \mid P,\theta) = \theta^{P-1}\Gamma(P)^{-1} x^{P-1} \exp(-\theta x),\ x \geq 0 \qquad \text{(B.3)}$$

The mean of the distribution is $E(X) = P/\theta$. The variance of the distribution is $Var(X) = P/\theta^2$. The *gamma* function is defined by $\Gamma(z) = \int_0^\infty x^{z-1}\exp(-x)dx$. The special case P=1, produces the exponential distribution. If ν is a positive integer, and $X \sim G(\nu/2,1/2)$ then $X \sim \chi^2(\nu)$, *i.e.* the chi-square distribution with ν degrees of freedom.

If the scalar, non-negative random variable X follows the inverted *gamma* distribution with shape parameter P and scale parameter θ, this is denoted by $X \sim IG(P,\theta)$ and its probability density function is given by

$$f_{IG}(x \mid P,\theta) = \theta^{P-1}\Gamma(P)^{-1} x^{-(P+1)} \exp(-\theta x^{-1}),\ x \geq 0 \qquad \text{(B.4)}$$

The $m(m+1)/2$ distinct elements of an $m \times m$ symmetric positive definite matrix $\mathbf{A}$ are distributed according to the Wishart $W(\mathbf{\Sigma},\nu,m)$ distribution if the probability density function is given by

$$p(\mathbf{A} \mid \mathbf{\Sigma},\nu,m) = k \mid \mathbf{\Sigma} \mid^{-\nu/2} \mid \mathbf{A} \mid^{(\nu-m-1)/2} \exp(-1/2 tr\mathbf{\Sigma}^{-1}\mathbf{A}),\ \mid \mathbf{A} \mid > 0 \qquad \text{(B.5)}$$

where $k^{-1} = 2^{\nu m/2} \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma(\nu + 1 - i)/2)$, $m \leq \nu$ and $\mathbf{\Sigma}$ is a symmetric positive definite matrix.

The $m(m+1)/2$ distinct elements of an $m \times m$ symmetric positive definite matrix $\mathbf{G}$ are distributed according to the inverted Wishart $W(\mathbf{H}, \nu, m)$ distribution if the probability density function is given by

$$p(\mathbf{G} \mid \mathbf{H}, \nu, m) = k \mid \mathbf{H} \mid^{\nu/2} \mid \mathbf{G} \mid^{-(\nu+m+1)/2} \exp(-1/2 tr \mathbf{G}^{-1}\mathbf{H}), \ \mid \mathbf{G} \mid > 0 \qquad \text{(B.6)}$$

where k was defined above, $m \leq \nu$ and $\mathbf{H}$ is a symmetric positive definite matrix. It must be noted that the joint probability density function of the $m(m+1)/2$ distinct elements of $\mathbf{G}^{-1}$ is a $W(\mathbf{H}^{-1}, \nu, m)$.