
Edinburgh School of Economics

**Discussion Paper Series**

Number 19

# *Bayesian Analysis of Stochastic Frontier Models*

**Gary Koop** (University of Edinburgh)
**Mark F J Steel** (University of Edinburgh)

Date
May 1999

***Published by***
School of Economics
University of Edinburgh
30 -31 Buccleuch Place
Edinburgh EH8 9JT
+44 (0)131 650 8361
http://www.ed.ac.uk/schools-departments/economics

THE UNIVERSITY *of* EDINBURGH

# Bayesian Analysis of Stochastic Frontier Models

GARY KOOP and MARK F.J. STEEL
*Department of Economics, University of Edinburgh,*
*Edinburgh, EH8 9JY, U.K.*
*E-mail: G.Koop or Mark.Steel @ed.ac.uk*

October 9, 1998
revised: May 21, 1999

# 1 Introduction

Stochastic frontier models are commonly used in the empirical study of firm[1] efficiency and productivity. The seminal papers in the field are Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977), while a recent survey is provided in Bauer (1990). The ideas underlying this class of models can be demonstrated using a simple production model[2] where output of firm $i$, $Y_i$, is produced using a vector of inputs, $X_i$, $(i = 1 \ldots N)$. The best-practice technology for turning inputs into output depends on a vector of unknown parameters, $\beta$, and is given by:

$$Y_i = f(X_i; \beta). \tag{1}$$

This so-called production frontier captures the maximum amount of output that can be obtained from a given level of inputs. In practice, actual output of a firm may fall below the maximum possible. The deviation of actual from maximum output is a measure of inefficiency and is the focus of interest in many applications. Formally, equation (1) can be extended to:

$$Y_i = f(X_i; \beta)\tau_i, \tag{2}$$

where $0 \leq \tau_i \leq 1$ is a measure of firm-specific efficiency and $\tau_i = 1$ indicates firm $i$ is fully efficient.

In this chapter we will discuss Bayesian inference in such models. We will draw on our previous work in the area: van den Broeck, Koop, Osiewalski and Steel (1994), Koop, Osiewalski and Steel (1994, 1997, 1998a,b) and Koop, Steel and Osiewalski (1995), whereas

[1]Throughout this chapter, we will use the term "firm" to refer to the cross-sectional unit of analysis. In practice, it could also be the individual or country, *etc.*

[2]In this chapter we focus on production frontiers. However, by suitably redefining $Y$ and $X$, the methods can be applied to cost frontiers.

theoretical foundations can be found in Fernández, Osiewalski and Steel (1997). It is worthwhile to digress briefly as to why we think these models are worthy of serious study. Efficiency measurement is very important in many areas of economics[3] and, hence, worthy of study in and of itself. However, stochastic frontier models are also close to other classes of models and can be used to illustrate ideas relating to the linear and nonlinear regression models; models for panel data, variance components, random coefficients and, generally, models with unobserved heterogeneity. Thus, stochastic frontier models can be used to illustrate Bayesian methods in many areas of econometrics. To justify our adoption of the Bayesian paradigm, the reader is referred to our work in the area. Suffice it to note here that the competitors to the Bayesian approach advocated here are the classical econometric stochastic frontier approach (see Bauer, 1990 for a survey) and the deterministic or non-parametric Data Envelopment Analysis (DEA) approach (see, *e.g.*, Färe, Grosskopf and Lovell, 1994). Each of the three approaches has strengths and weaknesses, some of which will be noted in this chapter.

This chapter is intended to be reasonably self-contained. However, we do assume that the reader has a basic knowledge of Bayesian methods as applied to the linear regression model (*e.g.* Judge, Griffiths, Hill, Lütkepohl and Lee, 1985, Chap. 4 or Poirier, 1995, pp. 288-309 and 524-550). Furthermore, we assume some knowledge of simulation methods. Koop (1994, pp. 12-26) provides a simple survey of some of these methods. Osiewalski and Steel (1998) focuses on simulation methods in the context of stochastic frontier models. Casella and George (1992) and Chib and Greenberg (1995) are good expository sources for Gibbs sampling and Metropolis-Hastings algorithms, respectively. Geweke (1999) is a complete survey of both Bayesian methods and computation.

The remainder of the chapter is organized as follows. The second section considers the stochastic frontier model with cross-sectional data beginning with a simple log-linear

[3]In addition to standard microeconomic studies of firm efficiencies, stochastic frontier models have been applied to *e.g.* environmental issues and macroeconomic growth studies.

model then considering a nonlinear extension and one where explanatory variables enter the efficiency distribution. The third section discusses the issues raised by the availability of panel data.

# 2 The Stochastic Frontier Model with Cross-Sectional Data

## 2.1 Introduction and Notation

The model given is equation (2) implicitly assumes that all deviations from the frontier are due to inefficiency. This assumption is also typically made in the DEA approach. However, following standard econometric practice, we add a random error to the model, $\zeta_i$, to capture measurement (or specification) error[4], resulting in:

$$Y_i = f(X_i; \beta)\tau_i\zeta_i. \tag{3}$$

The addition of measurement error makes the frontier stochastic, hence the term "stochastic frontier models". We assume that data for $i = 1 \dots N$ firms is available and that the production frontier, $f(\cdot)$, is log-linear (*e.g.* Cobb-Douglas or translog). We define $X_i$ as a $1 \times (k+1)$ vector (*e.g.* $X_i = (1\ L_i\ K_i)$ in the case of a Cobb-Douglas frontier with two inputs, $L$ and $K$) and, hence, (3) can be written as:

$$y_i = x_i\beta + v_i - z_i, \tag{4}$$

[4]This error reflects the stochastic nature of the frontier and we shall conveniently denote it by "measurement error". The treatment of measurement error is a crucial distinction between econometric and DEA methods. Most economic data sets are quite noisy and, hence, we feel including measurement error is important. DEA methods can be quite sensitive to outliers since they ignore measurement error. Furthermore, since the statistical framework for DEA methods is nonparametric, confidence intervals for parameter and efficiency estimates are very difficult to derive. However, econometric methods require the researcher to make more assumptions (*e.g.* about the error distribution) than do DEA methods. Recently, there has been some promising work on using the bootstrap with DEA methods which should lessen some of the criticisms of DEA (see Simar and Wilson, 1998a,b, and the references contained therein).

where $\beta = (\beta_0 \dots \beta_k)'$, $y_i = ln(Y_i)$, $v_i = ln(\zeta_i)$, $z_i = -ln(\tau_i)$ and $x_i$ is the counterpart of $X_i$ with the inputs transformed to logarithms. $z_i$ is referred to as inefficiency and, since $0 \leq \tau_i \leq 1$, it is a non-negative random variable. We assume that the model contains an intercept with coefficient $\beta_0$. Equation (4) looks like the standard linear regression model, except that the "error" is composed of two parts. This gives rise to another name for these models, *viz.* "composed error models".

For future reference, we define $y = (y_1 \dots y_N)'$, $v = (v_1 \dots v_N)'$, $z = (z_1 \dots z_N)'$ and the $N \times (k+1)$ matrix $x = (x_1' \dots x_N')'$. Also, let $f_G(a|b,c)$ denote the density function of a Gamma distribution with shape parameter $b$ and scale $c$ so that $a$ has mean $b/c$ and variance $b/c^2$. $p(d) = f_N^r(d|g,F)$ indicates that $d$ is $r$-variate Normal with mean $g$ and covariance matrix $F$. We will use $I(.)$ to denote the indicator function; *i.e.* $I(G) = 1$ if event $G$ occurs and is otherwise 0. Furthermore, $I_N$ will indicate the $N \times N$ identity matrix and $\iota_N$ and $N \times 1$ vector of ones. Sample means will be indicated with a bar, *e.g.* $\overline{y} = \frac{1}{N}\iota_N' y$.

## 2.2 Bayesian Inference

In order to define the sampling model[5], we make the following assumptions about $v_i$ and $z_i$ for $i = 1 \dots N$:

1. $p(v_i|h^{-1}) = f_N^1(v_i|0, h^{-1})$ and the $v_i$s are independent.

2. $v_i$ and $z_l$ are independent of one another for all $i$ and $l$.

3. $p(z_i|\lambda^{-1}) = f_G(z_i|1, \lambda^{-1})$ and the $z_i$s are independent.

The first assumption is commonly made in cross-sectional analysis, but the last two require some justification. Assumption 2. says that measurement error and inefficiency

[5]We use the terminology "sampling model" to denote the joint distribution of $(y, z)$ given the parameters and shall base the likelihood function on the marginal distribution of $y$ given the parameters.

are independent of one another. Assumption 3. is a common choice for the non-negative random variable, $z_i$, although others (*e.g.* the half-Normal) are possible. Ritter and Simar (1997) show that the use of very flexible one-sided distributions for $z_i$ such as the unrestricted Gamma may result in a problem of weak identification. Intuitively, if $z_i$ is left too flexible, then the intercept minus $z_i$ can come to look too much like $v_i$ and it may become virtually impossible to distinguish between these two components with small data sets. The Gamma with shape parameter 1 is the Exponential distribution, which is sufficiently different from the Normal to avoid this weak identification problem.[6] In addition, van den Broeck *et al.* (1994) found the Exponential model the least sensitive to changes in prior assumptions in a study of the most commonly used models. Note that $\lambda$ is the mean of the inefficiency distribution and let $\theta = (\beta', h, \lambda)'$ denote the parameters of the model.

The likelihood function is defined as:

$$L(y;\theta) = \prod_{i=1}^{N} p(y_i|x_i,\theta),$$

which requires the derivation of $p(y_i|x_i,\theta) = \int p(y_i|x_i,z_i,\theta)p(z_i|\theta)dz_i$. This is done in Jondrow, Lovell, Materov and Schmidt (1982) for the Exponential model and in van den Broeck *et al.* (1994) for a wider class of inefficiency distributions. However, we do not repeat the derivation here, since we do not need to know the explicit form of the likelihood function. To understand why isolating the likelihood function is not required, it is necessary to explain the computational methods that we recommend for Bayesian inference in stochastic frontier models.

Bayesian inference can be carried out using a posterior simulator which generates draws from the posterior, $p(\theta|y,x)$. In this case, Gibbs sampling with data augmentation is a natural choice for a posterior simulator. This algorithm relies on the fact that sequential draws,

[6] In van den Broeck *et al.* (1994) and Koop, Steel and Osiewalski (1995), the Erlang distribution (*i.e.* the Gamma distribution with fixed shape parameter, here chosen to be 1, 2 or 3) was used for inefficiency. The computational techniques necessary to work with Erlang distributions are simple extensions of those given in this section.

$\theta^{(s)}$ and $z^{(s)}$, from the conditional posteriors $p(\theta|y, x, z^{(s-1)})$ and $p(z|y, x, \theta^{(s)})$, respectively, will converge to draws from $p(\theta, z|y, x)$ from which inference on the marginal posteriors of $\theta$ or of functions of $z$ (such as efficiencies) can immediately be derived. In other words, we do not need to have an analytical formula for $p(\theta|y, x)$ (and, hence, the likelihood function), but rather we can suffice with working out the full conditional distributions $p(\theta|y, x, z)$ and $p(z|y, x, \theta)$. Intuitively, the former is very easy to work with since, conditional on $z$, the stochastic frontier model reduces to the standard linear regression model[7]. If $p(\theta|y, x, z)$ as a whole is not analytically tractable, we can split up $\theta$ into, say, $\beta$ and $(h, \lambda)$ and draw sequentially from the full conditionals $p(\beta|h, \lambda, y, x, z)$, $p(h, \lambda|\beta, y, x, z)$ and $p(z|y, x, \beta, h, \lambda)$. However, before we can derive the Gibbs sampler, we must complete the Bayesian model by specifying a prior for the parameters.

The researcher can, of course, use any prior in an attempt to reflect his/her prior beliefs. However, a proper prior for $h$ and $\lambda^{-1}$ is advisable: Fernández, Osiewalski and Steel (1997) show that Bayesian inference is not feasible (in the sense that the posterior distribution is not well-defined) under the usual improper priors for $h$ and $\lambda^{-1}$. Here, we will assume a prior of the product form: $p(\theta) = p(\beta)p(h)p(\lambda^{-1})$. In stochastic frontier models, prior information exists in the form of economic regularity conditions. It is extremely important to ensure that the production frontier satisfies these, since it is highly questionable to interpret deviations from a non-regular frontier as representing inefficiency. In an extreme case, if the researcher is using a highly flexible (or nonparametric) functional form for $f(\cdot)$ it might be possible for the frontier to fit the data nearly perfectly. It is only the imposition of economic regularity conditions that present this overfitting. The exact form of the economic regularity conditions depend on the specification of the frontier. For instance, in the Cobb-Douglas case, $\beta_i \geq 0$, $i = 1 \ldots k$ ensures global regularity of the production frontier. For

[7] As shown in Fernández, Osiewalski and Steel (1997), the use of the full model with data augmentation also allows for the derivation of crucial theoretical results on the existence of the posterior distribution and moments.

the translog specification things are more complicated and we may wish only to impose local regularity. This requires checking certain conditions at each data point (see Koop, Osiewalski and Steel, 1998b). In either case, we can choose a prior for $\beta$ which imposes economic regularity. As emphasized by Fernández, Osiewalski and Steel (1997), a proper or bounded prior is sufficient for $\beta$. Thus, it is acceptable to use a Uniform (flat) prior:

$$p(\beta) \propto I(E), \tag{5}$$

where $I(E)$ is the indicator function for the economic regularity conditions. Alternatively, a Normal prior for $\beta$ is proper and computationally convenient. In this chapter, we will use $p(\beta)$ as a general notation, but assume it is either truncated Uniform or truncated Normal. Both choices will easily combine with a Normal distribution to produce a truncated Normal posterior distribution.

For the other parameters, we assume Gamma priors:

$$p(h) = f_G(h|a_h, b_h) \tag{6}$$

and

$$p(\lambda^{-1}) = f_G(\lambda^{-1}|a_\lambda, b_\lambda). \tag{7}$$

Note that, by setting $a_h = 0$ and $b_h = 0$ we obtain $p(h) \propto h^{-1}$, the usual noninformative prior for the error precision in the Normal linear regression model. Here, the use of this improper prior is precluded (see Theorem 1 (ii) of Fernández, Osiewalski and Steel, 1997), but small values of these hyperparameters will allow for Bayesian inference (see Proposition 2 of Fernández, Osiewalski and Steel, 1997) while the prior is still dominated by the likelihood function. The hyperparameters $a_\lambda$ and $b_\lambda$ can often be elicited through consideration of the efficiency distribution. That is, researchers may often have prior information about the shape or location of the efficiency distribution. As discussed in van den Broeck *et al.*

(1994), setting $a_\lambda = 1$ and $b_\lambda = -ln(\tau^*)$ yields a relatively noninformative prior which implies the prior median of the efficiency distribution is $\tau^*$. These are the values for $a_\lambda$ and $b_\lambda$ used in the following discussion.

The Gibbs sampler can be developed in a straightforward manner by noting that, if $z$ were known, then we could write the model as $y + z = x\beta + v$ and standard results for the Normal linear regression model can be used. In particular, we can obtain

$$p(\beta|y, x, z, h, \lambda^{-1}) = f_N^{k+1}(\beta|\widehat{\beta}, h^{-1}(x'x)^{-1})p(\beta), \tag{8}$$

where

$$\widehat{\beta} = (x'x)^{-1}x'(y + z).$$

Furthermore,

$$p(h|y, x, z, \beta, \lambda^{-1}) = f_G(h|a_h + \frac{N}{2}, b_h + \frac{(y - x\beta + z)'(y - x\beta + z)}{2}). \tag{9}$$

Also, given $z$, the full conditional posterior for $\lambda^{-1}$ can easily be derived:

$$p(\lambda^{-1}|y, x, z, \beta, h) = f_G(\lambda^{-1}|N + 1, z'\iota_N - \ln(\tau^*)). \tag{10}$$

Equations (8), (9) and (10) are the full conditional posteriors necessary for setting up the Gibbs sampler *conditional on z*. To complete the posterior simulator, it is necessary to derive the posterior distribution of $z$ conditional on $\theta$. Noting that we can write $z = x\beta - y + v$,where $v$ has p.d.f. $f_N^N(v|0, h^{-1}I_N)$ and $z_i$ is *a priori* assumed to be i.i.d. $f_G(z_i|1, \lambda^{-1})$,[8] we obtain:

[8]The assumption that the inefficiencies are drawn from the Exponential distribution with unknown common mean $\lambda$ can be interpreted as a hierarchical prior for $z_i$. Alternatively, a classical econometrician would interpret this distributional assumption as part of the sampling model. This difference in interpretation highlights the fact that the division into prior and sampling model is to some extent arbitrary. See Fernández, Osiewalski and Steel (1997) for more discussion of this issue.

$$p(z|y,x,\beta,h,\lambda^{-1}) \propto f_N^N(z|x\beta - y - h^{-1}\lambda^{-1}\iota_N, h^{-1}I_N)\prod_{i=1}^{N} I(z_i \geq 0). \qquad (11)$$

A Gibbs sampler with data augmentation on $(\beta, h, \lambda^{-1}, z)$ can be set up by sequentially drawing from (8), (9), (10) and (11), where $(\beta, h)$ and $\lambda^{-1}$are independent given $z$, so that (10) can be combined with either (8) or (9) and there are only three steps in the Gibbs. Note that all that is required is random number generation from well-known distributions, where drawing from the high-dimensional vector $z$ is greatly simplified as (11) can be written as the product of $N$ univariate truncated Normals.

Given posterior simulator output, posterior properties of any of the parameters or of the individual $\tau_i$s can be obtained.[9] The latter can be calculated using simulated draws from (11) and transforming according to $\tau_i = \exp(-z_i)$. It is worth stressing that the Bayesian approach provides a finite sample distribution of the efficiency of each firm. This allows us to obtain both point and interval estimates, or even *e.g.* $P(\tau_i > \tau_j|y, x)$. The latter is potentially crucial since important policy consequences often hinge on one firm being labelled as more efficient in a statistically significant sense. Both DEA and classical econometric approaches typically only report point estimates. The DEA approach is nonparametric and, hence, confidence intervals for the efficiency measures obtained are very hard to derive.[10] Distributional theory for the classical econometric approach is discussed in Jondrow *et al.* (1982) and Horrace and Schmidt (1996). These papers point out that, although point estimates and confidence intervals for $\tau_i$ can be calculated, the theoretical justification is not that strong. For example, the maximum likelihood estimator for $\tau_i$ is inconsistent and

[9] Note that we have not formally proven that the posterior mean and variance of $\theta$ exist (although numerical evidence suggests that they do). Hence, we recommend using the posterior median and interquartile range of $\theta$ to summarize properties of the posterior. Since $0 \leq \tau_i \leq 1$, we know that all posterior moments exist for the firm specific efficiencies.

[10] Recent work on bootstrapping DEA frontiers is promising to surmount this problem and this procedure seems to be gaining some acceptance.

the methods for constructing confidence intervals assume unknown parameters are equal to their point estimates. For this reason, it is common in classical econometric work to present some characteristics of the efficiency distribution as a whole (*e.g.* estimates of $\lambda$) rather than discuss firm specific efficiency. However, firm specific efficiencies are often of fundamental policy importance and, hence, we would argue that an important advantage of the Bayesian approach is its development of finite sample distributions for the $\tau_i$s.

## 2.3 Extensions

There are many ways of extending the previous model. For instance, we could allow for different distributions for $z_i$ (see Koop, Steel and Osiewalski, 1995) or for many outputs to exist (see Fernández, Koop and Steel, 1998). Here we focus on two other extensions which are interesting in and of themselves, but also allow us to discuss some useful Bayesian techniques.

### 2.3.1 Explanatory variables in the Efficiency Distribution

Consider, for instance, a case where data is available for many firms, but some are private companies and others are state-owned. Interest centers on investigating whether private companies tend to be more efficient than state owned ones. This type of question can be formally handled by stochastic frontier models if we extend them to allow for explanatory variables in the efficiency distribution. Let us suppose that data exists on $m$ variables which may affect the efficiency of firms (*i.e.* $w_{ij}$, for $i = 1 \ldots N$ and $j = 1 \ldots m$). We assume $w_{i1} = 1$ is an intercept and $w_{ij}$ are 0-1 dummy variables for $j = 2 \ldots m$. The latter assumption could be relaxed at the cost of increasing the complexity of the computational methods. Since $\lambda$, the mean of the inefficiency distribution, is a positive random variable, a logical extension of the previous model is to allow it to vary over firms in the following manner:

$$\lambda_i^{-1} = \prod_{j=1}^{m} \phi_j^{w_{ij}}, \tag{12}$$

where the $\phi_j > 0$ are unknown parameters. Note that if $\phi_j = 1$ for $j = 2 \ldots m$ then this model reduces to the previous one. To aid in interpretation, observe how this specification allows, for instance, for private and state-owned firms to have different inefficiency distributions. If $w_{i2} = 1$ indicates that firm $i$ is private, then $\phi_2 > 1$ implies that the mean of the inefficiency distribution is lower for private firms and, hence, that private firms tend to be more efficient than state-owned ones. We stress that such a finding would not imply that every private firm is more efficient than every state-owned one, but rather that the former are drawing their efficiencies from a distribution with a higher mean. Such a specification seems very suitable for many sorts of policy issues and immediately allows for out-of-sample predictions.

For the new parameters, $\phi = (\phi_1 \ldots \phi_m)'$, we assume independent Gamma priors: $p(\phi) = p(\phi_1) \ldots p(\phi_m)$ with $p(\phi_j) = f_G(\phi_j | a_j, b_j)$ for $j = 1 \ldots m$. If the explanatory variables have no role to play (*i.e.* $\phi_2 = \ldots = \phi_m = 1$ ), then $\phi_1$ is equivalent to $\lambda^{-1}$ in the previous model. This suggests one may want to follow the prior elicitation rule discussed above and set $a_1 = 1$ and $b_1 = -ln(\tau^*)$. The other prior hyperparameters, $a_j$ and $b_j$ for $j = 2 \ldots m$, can be selected in the context of particular applications with moderate values for these parameters yielding a relatively noninformative prior. See Koop, Osiewalski and Steel (1997) for details.

A posterior simulator using Gibbs sampling with data augmentation can be set up as a straightforward extension of the one considered above. In fact, the posterior conditionals for $\beta$ and $h$ (*i.e.* equations (8) and (9)) are completely unaffected and the conditional for $z$ in (11) is only affected in that $\lambda^{-1}\iota_N$ must be replaced by the vector $\eta = (\lambda_1^{-1} \ldots \lambda_N^{-1})'$, where $\lambda_i^{-1}$ is given in equation (12). It can also be verified that for $j = 1 \ldots m$:[11]

[11]This is where the assumption that the $w_{ij}$s are 0-1 dummies is crucial.

$$p(\phi_j|y,x,z,\beta,h,w,\phi^{(-j)}) = f_G(\phi_j|a_j + \sum_{i=1}^{N} w_{ij}, b_j + \sum_{i=1}^{N} w_{ij} z_i \prod_{s \neq qj} \phi_s^{w_{is}}), \tag{13}$$

where $\phi^{(-j)} = (\phi_1 \ldots \phi_{j-1}, \phi_{j+1} \ldots \phi_m)'$. Hence, Bayesian inference in this model can again be conducted through sequential drawing from tractable distributions.

So far, we have focussed on posterior inference. This stochastic frontier model with varying efficiency distribution can be used to illustrate Bayesian model comparison. Suppose $m = 2$ and we are interested in calculating the Bayes factor comparing model $M_1$ where $\phi_2 = 1$ (*e.g.* there is no tendency for state-owned and private firms to differ in their efficiency distributions) against model $M_2$ with $\phi_2 \neq 1$. The prior for $M_2$ is given above. Define $\psi = (\beta, h, \phi^{(-2)})'$ as the parameters in the model $M_1$ and let $p_l(.)$ indicate a density under $M_l$ for $l = 1, 2$. If we make the reasonable assumption that $p_2(\psi|\phi_2 = 1) = p_1(\psi)$, then the Bayes factor in favor of $M_1$ can be written as the Savage-Dickey density ratio (see Verdinelli and Wasserman, 1995):

$$B_{12} = \frac{p_2(\phi_2 = 1|y,x,w)}{p_2(\phi_2 = 1)}, \tag{14}$$

the ratio of posterior to prior density values at the point being tested. Note that the denominator of (14) is trivial to calculate since it is merely the Gamma prior for $\phi_2$ evaluated at a point. The numerator is also easy to calculate using (13). As Verdinelli and Wasserman (1995) stress, a good estimator of $p(\phi_2 = 1|y,x,w)$ on the basis of $R$ Gibbs replications is:

$$\frac{1}{R}\sum_{r=1}^{R} p(\phi_2 = 1|y,x,z^{(r)},\beta^{(r)},h^{(r)},w,\phi^{(-2)(r)}), \tag{15}$$

where superscript $^{(r)}$ denotes the $r^{th}$ draw in the Gibbs sampling algorithm. That is, we can just evaluate (13) at $\phi_2 = 1$ for each draw and average. Bayes factors for hypotheses such as this can be easily calculated without recourse to evaluating the likelihood function or adding steps to the simulation algorithm (as in the more general methods of Gelfand and Dey, 1994 and Chib, 1995, respectively).

### 2.3.2 Nonlinear Production Frontiers

The previous models both assumed that the production frontier was log-linear. However, many common production functions are inherently nonlinear in the parameters (*e.g.* the constant elasticity of substitution or CES or the asymptotically ideal model or AIM, see Koop, Osiewalski and Steel, 1994). However, the techniques outlined above can be extended to allow for an arbitrary production function. Here we assume a model identical to the stochastic frontier model with common efficiency distribution (i.e. m=1) except that the production frontier is of the form:[12]

$$y_i = f(x_i; \beta) + v_i - z_i. \tag{16}$$

The posterior simulator for everything except $\beta$ is almost identical to the one given above. Equation (10) is completely unaffected, and (9) and (11) are slightly altered by replacing $x\beta$ by $f(x;\beta) = (f(x_1;\beta)\ldots f(x_N;\beta))'$.

However, the conditional posterior for $\beta$ is more complicated, having the form:

$$p(\beta|y,x,z,h,\lambda^{-1}) \propto \exp\left(-\frac{h}{2}\sum_{i=1}^{N}(y_i - f(x_i;\beta) + z_i)^2\right) p(\beta). \tag{17}$$

Equation (17) does not take the form of any well-known density and the computational algorithm selected will depend on the exact form of $f(x;\beta)$. For the sake of brevity, here we will only point the reader in the direction of possible algorithms that may be used for drawing from (17). Two major cases are worth mentioning. First, in many cases, it might be possible to find a convenient density which approximates (17) well. For instance, in the case of the AIM model a multivariate-$t$ density worked well (see Koop, Osiewalski and Steel, 1994). In this case, importance sampling (Geweke, 1989) or an independence chain Metropolis-Hastings algorithm (Chib and Greenberg, 1995) should work well. On

[12] The extension to a varying efficiency distribution as in (12) is trivial and proceeds along the lines of the previous model.

the other hand, if no convenient approximating density can be found, a random walk chain Metropolis-Hastings algorithm might prove a good choice (see Chib and Greenberg, 1995). The precise choice of algorithm will be case-specific and, hence, we do not discuss this issue in any more detail here.

# 3 The Stochastic Frontier Model with Panel Data

## 3.1 Time-invariant Efficiency

It is increasingly common to use panel data[13] in the classical econometric analysis of the stochastic frontier model. Some of the statistical problems (*e.g.* inconsistency of point estimates of firm specific efficiency) of classical analysis are alleviated with panel data and the assumption of a particular distributional form for the inefficiency distribution can be dispensed with at the cost of assuming time-invariant efficiencies (*i.e.* treating them as "individual effects"). Schmidt and Sickles (1984) is an early influential paper which develops a relative efficiency measure based on a fixed effects specification and an absolute efficiency measure based on a random effects specification. In this paper, we describe a Bayesian alternative to this classical analysis and relate the random/fixed effects distinction to different prior structures for the efficiency distribution.

Accordingly, assume that data is available for $i = 1 \ldots N$ firms for $t = 1 \ldots T$ time periods. We will extend the notation of the previous section so that $y_i$ and $v_i$ are now $T \times 1$ vectors and $x_i$ a $T \times k$ matrix containing the $T$ observations for firm $i$. Note, however, that the assumption of constant efficiency over time implies that $z_i$ is still a scalar and $z$ an $N \times 1$ vector. For future reference, we now define $y = (y_1' \ldots y_N')'$ and $v = (v_1' \ldots v_N')'$ as $NT \times 1$ vectors and $x = (x_1' \ldots x_N')'$ as an $NT \times k$ matrix. In contrast to previous

[13] Of course, many of the issues which arise in the stochastic frontier model with panel data also arise in traditional panel data models. It is beyond the scope of the present chapter to attempt to summarize the huge literature on panel data. The reader is referred to Matyas and Sevestre (1996) or Baltagi (1995) for an introduction to the broader panel data literature.

notation, $x_i$ does not contain an intercept. We assume that the stochastic frontier model can be written as:

$$y_i = \beta_0 \iota_T + x_i \delta + v_i - z_i \iota_T, \tag{18}$$

where $\beta_0$ is the intercept coefficient and $v_i$ is i.i.d. with p.d.f. $f_N^T(v_i|0, h^{-1} I_T)$. As discussed in Fernández, Osiewalski and Steel (1997), it is acceptable to use an improper noninformative prior for $h$ when $T > 1$ and, hence, we assume $p(h) \propto h^{-1}$. We discuss different choices of priors for $\beta_0$ and $z_i$ in the following material.

### 3.1.1 Bayesian Fixed Effects Model

Equation (18) looks like a standard panel data model (see, *e.g.*, Judge *et al.*, 1985, Chap. 13). The individual effect in the model can be written as:

$$\alpha_i = \beta_0 - z_i,$$

and the model rewritten as:

$$y_i = \alpha_i \iota_T + x_i \delta + v_i. \tag{19}$$

Classical fixed effects estimation of (19) proceeds by making no distributional assumption for $\alpha_i$, but rather using firm-specific dummy variables. The Bayesian analogue to this is to use flat, noninformative priors for the $\alpha_i$s.[14] Formally, defining $\alpha = (\alpha_1 \dots \alpha_N)'$, we then adopt the prior:

$$p(\alpha, \delta, h) \propto h^{-1} p(\delta). \tag{20}$$

[14]Note that this implies we now deviate from Assumption 3 in subsection 2.2.

The trouble with this specification is that we cannot make direct inference about $z_i$ (since $\beta_0$ is not separately identified) and, hence, the absolute efficiency of firm $i$: $\tau_i = \exp(-z_i)$. However, following Schmidt and Sickles (1984), we define relative inefficiency as:

$$z_i^{rel} = z_i - \min_j(z_j) = \max_j(\alpha_j) - \alpha_i. \tag{21}$$

In other words, we are measuring inefficiency relative to the most efficient firm (*i.e.* the firm with the highest $\alpha_i$).[15] Relative efficiency is defined as $r_i^{rel} = \exp(-z_i^{rel})$ and we assume that the most efficient firm has $r_i^{rel} = 1$.

It is worth noting that this prior seems like an innocuous noninformative prior, but this initial impression is false since it implies a rather unusual prior for $r_i^{rel}$. In particular, as shown in Koop, Osiewalski and Steel (1997), $p(r_i^{rel})$ has a point mass of $N^{-1}$ at full efficiency and is $p(r_i^{rel}) \propto 1/r_i^{rel}$ for $r_i^{rel} \in (0,1)$. The latter is an L-shaped improper prior density which, for an arbitrary small $a \in (0,1)$, puts an infinite mass in $(0,a)$ but only a finite mass in $(a,1)$. In other words, this "noninformative" prior strongly favors low efficiency.

Bayesian inference in the fixed effects model can be carried out in a straightforward manner, by noting that for Uniform $p(\delta)$, (19)-(20) is precisely a Normal linear regression model with Jeffreys' prior. The vector of regression coefficients $(\alpha' \; \delta')'$ in such a model has a $(N+k)$-variate Student-$t$ posterior with $N(T-1)-k$ degrees of freedom (where we have assumed that $N(T-1) > k$, which implies $T > 1$). For typical values of $N$, $T$ and $k$ the degrees of freedom are enormous and the Student-$t$ will be virtually identical to the Normal distribution. Hence, throughout this subsection we present results in term of this Normal approximation.

Using standard Bayesian results for the Normal linear regression model (*e.g.* Judge *et*

[15] It is worth noting that the classical econometric analysis assigns the status of most efficient firm to one particular firm and measures efficiency relative to this. The present Bayesian analysis also measures efficiency relative to the most efficient firm, but allows for uncertainty as to which that firm is.

*al.*, 1985, Chap. 4), it follows that the marginal posterior for $\delta$ is given by (for general $p(\delta)$):

$$p(\delta|y,x) = f_N^k(\delta|\widehat{\delta}, \widehat{h}^{-1}S^{-1})p(\delta), \tag{22}$$

where

$$\widehat{\delta} = S^{-1}\sum_{i=1}^{N}(x_i - \iota_T\overline{x_i})'(y_i - \overline{y_i}\iota_T), \tag{23}$$

$$S = \sum_{i=1}^{N} S_i,\ \overline{x_i} = \frac{1}{T}\iota_T' x_i$$

and

$$S_i = (x_i - \iota_T\overline{x_i})'(x_i - \iota_T\overline{x_i}).$$

Note that (23) is the standard "within estimator" from the panel data literature. Finally,

$$\widehat{h}^{-1} = \frac{1}{N(T-1)-k}\sum_{i=1}^{N}(y_i - \widehat{\alpha_i}\iota_T - x_i\widehat{\delta})'(y_i - \widehat{\alpha_i}\iota_T - x_i\widehat{\delta}),$$

where $\widehat{\alpha_i}$ is the posterior mean of $\alpha_i$ defined below.

The marginal posterior of $\alpha$ is the $N$-variate Normal with means

$$\widehat{\alpha_i} = \overline{y_i} - \overline{x_i}\widehat{\delta},$$

and covariances

$$cov(\alpha_i, \alpha_j) = \widehat{h}^{-1}\left(\frac{\Delta(i,j)}{T} + \overline{x_i}S^{-1}\overline{x_j}'\right),$$

where $\Delta(i,j) = 1$ if $i = j$ and 0 otherwise. Thus, analytical formulae for posterior means and standard deviations are available and, if interest centers on these, posterior simulation methods are not required. However, typically interest centers on the relative efficiencies which are a complicated nonlinear function of $\alpha$, *viz.*,

$$r_i^{rel} = \exp(\alpha_i - \max_j(\alpha_j)), \tag{24}$$

and, hence, posterior simulation methods are required. However, direct Monte Carlo integration is possible since the posterior for $\alpha$ is multivariate Normal and can easily be simulated. These simulated draws of $\alpha$ can be transformed using (24) to yield posterior draws of $r_i^{rel}$. However, this procedure is complicated by the fact that we do not know which firm is most efficient (*i.e.* which firm has largest $\alpha_j$) and, hence, is worth describing in detail.

We begin by calculating the probability that a given firm, $i$, is the most efficient:

$$P(r_i^{rel} = 1|y,x) = P(\alpha_i = \max_j(\alpha_j)|y,x), \tag{25}$$

which can be easily calculated using Monte Carlo integration. That is, (25) can simply be estimated by the proportion of the draws of $\alpha$ which have $\alpha_i$ being the largest.

Now consider the posterior for $r_i^{rel}$ over the interval $(0,1)$ (*i.e.* assuming it is *not* the most efficient):

$$p(r_i^{rel}|y,x) = \sum_{j=1,j\neq i}^{N} p(r_i^{rel}|y,x,r_j^{rel} = 1)P(r_j^{rel} = 1|y,x). \tag{26}$$

Here $P(r_j^{rel} = 1|y,x)$ can be calculated as discussed in the previous paragraph. In addition, $p(r_i^{rel}|y,x,r_j^{rel} = 1)$ can be calculated using the same posterior simulator output. That is, assuming firm $j$ is most efficient, then $r_i^{rel} = \exp(\alpha_i - \alpha_j)$ which can be evaluated from those draws of $\alpha$ that correspond to $\alpha_j = \max_l(\alpha_l)$. Hence, posterior analysis of the relative efficiencies in a Bayesian fixed effects framework can be calculated in a straightforward manner.[16]

[16]This procedure can be computationally demanding since $P(r_j^{rel} = 1|y,x)$ and $p(r_i^{rel}|y,x,r_j^{rel})$ must be calculated for every possible $i$ and $j$. However, typically, $P(r_j^{rel} = 1|y,x)$ is appreciable (*e.g.* $> 0.001$) for only a few firms and the rest can be ignored (see Koop, Osiewalski and Steel, 1997, p. 82).

### 3.1.2 Bayesian Random Effects Model

The Bayesian fixed effects model described above might initially appeal to researchers who do not want to make distributional assumptions about the inefficiency distribution. However, as we have shown above, this model is implicitly making strong and possibly unreasonable prior assumptions. Furthermore, we can only calculate relative, as opposed to absolute, efficiencies. For these reasons, it is desirable to develop a model which makes an explicit distributional assumption for the inefficiencies. With such a model, absolute efficiencies can be calculated in the spirit of the cross-sectional stochastic frontier model of section 2, since the distribution assumed for the $z_i$s allows us to separately identify $z_i$ and $\beta_0$. In addition, the resulting prior efficiency distributions will typically be more in line with our prior beliefs. Another important issue is the sensitivity of the posterior results on efficiency to the prior specification chosen. Since $T$ is usually quite small, it makes sense to "borrow strength" from the observations for the other firms by linking the inefficiencies. Due to Assumption 3 in subsection 2.2, this is not done through the sampling model. Thus, Koop, Osiewalski and Steel (1997) define the difference between Bayesian fixed and random effects models through the prior for $z_i$. In particular, what matters are the prior links that are assumed between the $z_i$s. Fixed effects models assume, *a priori*, that the $z_i$s are fully separated. Random effects models introduce links between the $z_i$s, typically by assuming they are all drawn from distributions that share some common unknown parameter(s). In Bayesian language, the random effects model then implies a hierarchical prior for the individual effects.

Formally, we define a Bayesian random effects model by combining (18) with the prior:

$$p(\beta_0, \delta, h, z, \lambda^{-1}) \propto h^{-1} p(\delta) f_G(\lambda^{-1}|1, -ln(\tau^*)) \prod_{i=1}^{N} f_G(z_i|1, \lambda^{-1}). \tag{27}$$

That is, we assume noninformative priors for $h$ and $\beta_0$, whereas the inefficiencies are again assumed to be drawn from the Exponential distribution with mean $\lambda$. Note that the $z_i$s are now linked through this common parameter $\lambda$, for which we choose the same prior as in section 2.

Bayesian analysis of this model proceeds along similar lines to the cross-sectional stochastic frontier model presented in section 2. In particular, a Gibbs sampler with data augmentation can be set up. Defining $\beta = (\beta_0 \ \delta')'$ and $X = (\iota_{NT} : x)$ the posterior conditional for the measurement error precision can be written as:

$$p(h|y, x, z, \beta, \lambda^{-1}) = f_G\left(h|\frac{NT}{2}, \frac{1}{2}\left[y - X\beta + (I_N \otimes \iota_T)z\right]'\left[y - X\beta + (I_N \otimes \iota_T)z\right]\right). \quad (28)$$

Next we obtain:

$$p(\beta|y, x, z, h, \lambda^{-1}) = f_N^{k+1}\left(\beta|\overline{\beta}, h^{-1}(X'X)^{-1}\right)p(\delta), \quad (29)$$

where

$$\overline{\beta} = (X'X)^{-1}\left[y + (I_N \otimes \iota_T)z\right].$$

The posterior conditionals for the inefficiencies takes the form:

$$p(z|y, x, \beta, h, \lambda^{-1}) \propto f_N^N(z|\,(\iota_N : \widetilde{x})\,\beta - \widetilde{y} - (Th\lambda)^{-1}\iota_N, (Th)^{-1}I_N)\prod_{i=1}^{N} I(z_i \geq 0), \quad (30)$$

where $\widetilde{y} = (\overline{y}_1 \ldots \overline{y}_N)'$ and $\widetilde{x} = (\overline{x}_1' \ldots \overline{x}_N')'$.

Furthermore, the posterior conditional for $\lambda^{-1}$, $p(\lambda^{-1}|y, x, z, \beta, h)$, is the same as for the cross-sectional case (*i.e.* equation 10).

Using these results, Bayesian inference can be carried out using a Gibbs sampling algorithm based on (10), (28), (29) and (30). Although the formulas look somewhat complicated, it is worth stressing that all conditionals are either Gamma or truncated Normal.

## 3.2 Extensions

Extending the random effects stochastic frontier model to allow for a nonlinear production function or explanatory variables in the efficiency distribution can easily be done in a similar fashion as for the cross-sectional model (see subsection 2.3 and Koop, Osiewalski and Steel, 1997). Furthermore, different efficiency distributions can be allowed for in a straightforward manner and multiple outputs can be handled as in Fernández, Koop and Steel (1998). Here we concentrate on extending the model in a different direction. In particular, we free up the assumption that each firm's efficiency, $\tau_i$, is constant over time. Let us use the definitions of $X$ and $\beta$ introduced in the previous subsection and write the stochastic frontier model with panel data as:

$$y = X\beta - \gamma + v, \tag{31}$$

where $\gamma$ is a $TN \times 1$ vector containing inefficiencies for each individual observation and $y$ and $v$ are defined as in subsection 3.1. In practice, we may want to put some structure on $\gamma$ and, thus, Fernández, Osiewalski and Steel (1997) propose to rewrite it in terms of an $M$-dimensional vector $u$ as:

$$\gamma = Du, \tag{32}$$

where $M \leq TN$ and $D$ is a known $TN \times M$ matrix. Above, we implicitly assumed $D = I_N \otimes \iota_T$ which implies $M = N$ and $u_i = \gamma_{it} = z_i$. That is, firm-specific inefficiency was constant over time. However, a myriad of other possibilities exist. For instance, $D$ can correspond to cases where clusters of firms or time periods share common efficiencies, or parametric time dependence exists in firm-specific efficiency. Also note the case $D = I_{TN}$, which allows each firm in each period to have a different inefficiency (*i.e.* $\gamma = u$). Thus, we are then effectively back in the cross-section framework without exploiting the panel structure of the data. This case is considered in Koop, Osiewalski and Steel (1998a,b), where interest centered on the change in efficiency over time.[17] With all such specifications, it is possible to conduct Bayesian inference by slightly altering the posterior conditionals presented above in an obvious manner.

However, as discussed in Fernández, Osiewalski and Steel (1997), it is very important to be careful when using improper priors on any of the parameters. In some cases, improper priors imply that the posterior does not exist and, hence, valid Bayesian inference cannot be carried out. Intuitively, the inefficiencies can be interpreted as unknown parameters. If there are too many of these, prior information becomes necessary. As an example of the types of results proved in Fernández, Osiewalski and Steel (1997), we state one of their main theorems:

*Theorem (Fernández, Osiewalski and Steel, 1997, Theorem 1)*

*Consider the general model given in (31) and (32) and assume the standard noninformative for* $h: \ p(h) \propto h^{-1}$. *If rank*$(X:D) < TN$ *then the posterior distribution exists for any bounded or proper* $p(\beta)$ *and any proper* $p(u)$. *However, if rank*$(X:D) = TN$, *then the posterior does not exist.*

The Bayesian random effects model discussed above has rank$(X:D) < TN$, so the posterior does exist even though we have used an improper prior for $h$. However, for the case where efficiency varies over time and across firms (*i.e.* $D = I_{TN}$), more informative priors are required in order to carry out valid Bayesian inference. Fernández, Osiewalski and Steel (1997, Prop. 2) show that a weakly informative (not necessarily proper) prior on $h$ that penalizes large values of the precision is sufficient.

[17]Koop, Osiewalski and Steel (1998a,b) also allow the frontier to shift over time and interpret such shifts as technical change. In such a framework, it is possible to decompose changes in output growth into components reflecting input change, technical change and efficiency change. The ability of stochastic frontier models with panel data to calculate such decompositions is quite important in many practical applications. Also of interest are Baltagi and Griffin (1988) and Baltagi, Griffin and Rich (1995), which develop a more general framework relating changes in the production function with technical change in a non-stochastic frontier panel data model.

# 4 Summary

In this chapter, we have described a Bayesian approach to efficiency analysis using stochastic frontier models. With cross-sectional data and a log-linear frontier, a simple Gibbs sampler can be used to carry out Bayesian inference. In the case of a nonlinear frontier, more complicated posterior simulation methods are necessary. Bayesian efficiency measurement with panel data is then discussed. We show how a Bayesian analogue of the classical fixed effects panel data model can be used to calculate the efficiency of each firm relative to the most efficient firm. However, absolute efficiency calculations are precluded in this model and inference on efficiencies can be quite sensitive to prior assumptions. Accordingly, we describe a Bayesian analogue of the classical random effects panel data model which can be used for robust inference on absolute efficiencies. Throughout we emphasize the computational methods necessary to carry out Bayesian inference. We show how random number generation from well-known distributions is sufficient to develop posterior simulators for a wide variety of models.

# 5 References


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

Baltagi, B. 1995: *Econometric Analysis of Panel Data*, New York: John Wiley and Sons.

Baltagi, B. and Griffin, J. 1988: A general index of technical change, *Journal of Political Economy*, 90, 20-41.

Baltagi, B.; Griffin, J and Rich, D. 1995: The measurement of firm-specific indexes of technical change, *Review of Economics and Statistics*, 77, 654-663.

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

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

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

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

Chib, S. and Greenberg, E. 1995: Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49, 327-335.

Färe, R.; Grosskopf, S. and Lovell, C.A.K. 1994: *Production Frontiers*. Cambridge, U.K.: Cambridge University Press.

Fernández, C.; Koop, G. and Steel, M.F.J. 1998: A Bayesian analysis of multiple output production frontiers. manuscript.

Fernández, C.; Osiewalski, J. and Steel, M.F.J. 1997: On the use of panel data in stochastic frontier models with improper priors. *Journal of Econometrics*, 79, 169-193.

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

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

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

Horrace, W. and Schmidt, P. 1996: Confidence statements for efficiency estimates from stochastic frontiers. *Journal of Productivity Analysis*, 7, 257-282.

Jondrow, J.; Lovell, C.A.K.; Materov, I.S. and Schmidt, P. 1982: On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19, 233-238.

Judge, G.; Griffiths, W.; Hill, R.C.; Lütkepohl, H. and Lee, T.-C. 1985: *The Theory* *and Practice of Econometrics*, Second edition. New York: John Wiley and Sons.

Koop, G. 1994: Recent progress in applied Bayesian econometrics. *Journal of Economic Surveys*, 8, 1-34.

Koop, G.; Osiewalski, J. and Steel, M.F.J. 1994: Bayesian efficiency analysis with a flexible form: The AIM cost function. *Journal of Business and Economic Statistics*, 12, 93-106.

Koop, G.; Osiewalski, J. and Steel, M.F.J. 1997: Bayesian efficiency analysis through individual effects: Hospital cost frontiers. *Journal of Econometrics*, 76, 77-105.

Koop, G.; Osiewalski, J. and Steel, M.F.J. 1998a: Modeling the sources of output growth in a panel of countries. *Journal of Business and Economic Statistics*, forthcoming.

Koop, G.; Osiewalski, J. and Steel, M.F.J. 1998b: The components of output growth: A stochastic frontier analysis. *Oxford Bulletin of Economics and Statistics*, forthcoming.

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

Matyas, L. and Sevestre, P. (editors) 1996: *The Econometrics of Panel Data*, Dordrecht: Kluwer Academic Publishers.

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

Osiewalski, J. and Steel, M.F.J. 1998: Numerical tools for the Bayesian analysis of stochastic frontier models. *Journal of Productivity Analysis,* 10, 103-117.

Poirier, D. 1995: *Intermediate Statistics and Econometrics: A Comparative Approach.* Cambridge, Mass.: The MIT Press.

Ritter, C. and Simar, L. 1997: Pitfalls of Normal-Gamma stochastic frontier models. *Journal of Productivity Analysis*, 8, 167-182.

Schmidt, P. and Sickles, R. 1984: Production frontiers and panel data. *Journal of Business and Economic Statistics*, 2, 367-374.

Simar, L. and Wilson, P.W. 1998a: A general methodology for bootstrapping in nonparametric frontier models. manuscript.

Simar, L. and Wilson, P.W. 1998b: Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44, 49-61.

Verdinelli, I. and Wasserman, L. 1995: Computing Bayes factors using a generalization of the Savage-Dickey Density Ratio. *Journal of the American Statistical Association*, 90, 614-618.