



Edinburgh School of Economics
Discussion Paper Series
Number 21

*A Bayesian analysis of multiple-output production
frontier*

Carmen Fernandez (University of Bristol)
Gary Koop (University of Edinburgh)
Mark F J Steel (University of Edinburgh)

Date
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

A Bayesian Analysis of Multiple-Output Production Frontiers

Carmen Fernández

*Department of Mathematics, University of Bristol
Bristol, BS8 1TW, U.K.*

Gary Koop

*Department of Economics, University of Edinburgh,
Edinburgh, EH8 9JY, U.K.*

Mark Steel*

*Department of Economics, University of Edinburgh,
Edinburgh, EH8 9JY, U.K.*

In this paper we develop Bayesian tools for estimating multi-output production frontiers in applications where only input and output data are available. Firm-specific inefficiency is measured relative to such frontiers. Our work has important differences from the existing literature, which either assumes a classical econometric perspective with restrictive functional form assumptions, or a non-stochastic approach which directly estimates the output distance function. Bayesian inference is implemented using a Markov chain Monte Carlo algorithm. A banking application shows the ease and practicality of our approach.

Keywords: Banking data, efficiency, productivity, Markov Chain Monte Carlo
JEL classification: C11, D24

*Corresponding author; Department of Economics, University of Edinburgh, 50 George Square, Edinburgh, EH8 9JY, United Kingdom; Phone: +44-131-650 8352, Fax: +44-131-650 4514; E-mail: Mark.Steel@ed.ac.uk

We acknowledge funding by the ESRC under grant number R34515. This research originated when the first author held a Research Training Grant (ERBFMBICT 961021) under the Training and Mobility of Researchers Programme, financed by the European Commission, and the first and third author were affiliated with CentER, Tilburg University, The Netherlands. Stimulating comments from two anonymous Referees and discussions with Subal Kumbhakar and Jacek Osiewalski are gratefully acknowledged. We would also like to thank Robert Adams, Allen Berger and Robin Sickles for providing the banking data used in this paper.

1 Introduction

In many areas of economics, interest centres on the estimation of production technology and the related issue of how to measure a firm's efficiency relative to this technology. This paper discusses these issues for the case where a firm uses inputs to produce several outputs jointly. A Bayesian approach is used, building on our previous work with single output technology (*e.g.* Fernández, Osiewalski and Steel, 1997 and Koop, Osiewalski and Steel, 1997), and an empirical implementation using Markov chain Monte Carlo (MCMC) methods is presented.

To motivate the present paper, it is worthwhile to briefly discuss some of the relevant literature. The theoretical starting point in most analyses of multiple-output technology is a transformation function, which describes the production technology through a relationship of the form:

$$f(y, x) = 0,$$

where y is a vector of p outputs and x is a vector of inputs. Lau (1972) provides an early theoretical discussion of such technologies. If the transformation function is separable then we can write it as:

$$g(y) = h(x).$$

More discussion of separability and other assumptions made in the course of this paper is provided in the Conclusion. Econometric estimation of such a model, even if separability is assumed, is difficult for two reasons:

i) A single equation, such as that provided by the transformation function, is not enough to provide a valid statistical model for $p > 1$ endogenous variables (*i.e.* y). So merely adding an error onto this function is not an acceptable way to describe the process generating y . In general, we need a p -equation system and merely writing out a seemingly unrelated regression model for the p outputs will ignore the restrictions implied by jointness in production.

ii) The left-hand side of the equation, $g(y)$, typically depends on unknown parameters, which means that $g(y)$ can not constitute a sufficient statistic for estimating $h(\cdot)$.

Perhaps as a result of these problems, there has been very little empirical work which directly estimates the multi-output transformation function. Most researchers working in the area estimate cost functions and/or demand systems when faced with multiple outputs, and avoid the problems caused by i) and ii) (see, *e.g.* Kumbhakar, 1996, for a discussion of these issues). However, such a solution requires that the researcher have available data on costs and/or prices. In the common case where only y and x are observed, the researcher is forced to deal directly with the transformation function.

There is a large literature (*e.g.* Färe and Primont, 1990) which implicitly estimates this function in order to evaluate firm-specific productivity using non-econometric approaches. This approach assumes a deterministic transformation function (*i.e.* no measurement error in the data) and uses linear programming techniques. For many cases, such an approach is undoubtedly reasonable. However, in noisy data sets it might be preferable to formally model measurement error and adopt an econometric approach (see Koop *et al.* 1997, 2000 for a more detailed discussion of these issues). The only econometric approaches we have been able to find in the literature are those of Adams, Berger and Sickles (1996, 1999) and Löthgren (1997), all of which assume separability in the transformation function. Adams *et al.* (1996, 1999) consider a linear $g(y)$ and normalize the coefficient of one of the outputs to 1, while putting the remaining outputs on the right-hand side. Problems which arise due to the correlation of the remaining outputs with the error term are resolved through semiparametric efficient methods, but, of course, results are not invariant to the (arbitrary) choice of the dependent variable. Löthgren (1997) makes $g(y)$ linear in the polar coordinates of y and models the norm of the output vector as a function of the polar angles and the inputs, thus ignoring the multivariate character of the output vector. Although promising, both are really single-output methods (conditioning on $p - 1$ output dimensions), and the linearity restriction on $g(y)$ is perhaps objectionable in some contexts.

In the present paper, we derive econometric methods which allow for $g(y)$ to have general forms. We assume a constant elasticity of transformation form, but the basic ideas extend to any form for $g(y)$. In addition, we shall take account of the multivariate character of the data by specifying a p -dimensional sampling model. To establish some terminology, note that $g(y) = \text{constant}$ maps out the output combinations that are technologically equivalent. Hence, it is referred to as the production equivalence surface, which is $(p - 1)$ -dimensional. By analogy with the single output case, $h(x)$ defines the maximum output [as measured by $g(y)$] that can be produced with inputs x and is referred to as the production frontier. If measurement error did not exist, all firms would lie on or within the frontier and deviations from the frontier would be interpreted as firm-specific inefficiency.

Our methods are Bayesian and, as discussed in our previous work (Koop *et al.* 1997, 2000), allow us to calculate exact finite sample properties of all features of interest (including firm-specific efficiency) and surmount some difficult statistical problems involved with classical estimation of stochastic frontier models.¹ We would like to stress that the purpose of the present paper is not to argue that the Bayesian approach is invariably superior to the classical econometric or linear programming approaches. Each has its advantages and disadvantages. Rather we would argue that the Bayesian method should be one important tool, among others, in the kit of techniques for the researcher working with multiple-output production frontier models. The purpose of the present paper is to develop this tool in the context of a novel way of modelling multi-output frontiers.

The remainder of the paper is organized as follows: The second section presents and motivates the basic model used in the paper. The third section discusses prior elicitation, while the fourth derives methods for Bayesian inference using MCMC techniques. The fifth

¹For instance, in a single output cross-sectional context Jondrow *et al.* (1982) show that the efficiency of a particular firm can be estimated using classical methods, but not consistently. Depending on the assumptions made about the efficiency distribution, this problem can carry over to panel data contexts. Furthermore, it is difficult to obtain standard errors for measures of efficiency. Horrace and Schmidt (1996) provides a recent discussion of various methods to construct confidence intervals for efficiencies. Note that these involve strong assumptions and treat estimates of parameters as though they are equal to true values.

section contains applications and the sixth section concludes and comments on possible extensions.

2 The Model

We consider a set of NT observations corresponding to outputs of N different firms (or agents) over T time periods. The output of firm i ($i = 1, \dots, N$) at time t ($t = 1, \dots, T$) is p -dimensional and is given by the vector $y_{(i,t)} = (y_{(i,t,1)}, \dots, y_{(i,t,p)})' \in \mathfrak{R}_+^p$.

In this section we extend the stochastic frontier model with composed error, so far developed for the case of a single output, to multiple outputs. This is done through the following transformation of the p -dimensional output vector:

$$\theta_{(i,t)} = \left(\sum_{j=1}^p \alpha_j^q y_{(i,t,j)}^q \right)^{1/q}, \quad (2.1)$$

with $\alpha_j \in (0, 1)$ for all $j = 1, \dots, p$ and such that $\sum_{j=1}^p \alpha_j = 1$ and with $q > 1$. For fixed values of $\alpha = (\alpha_1, \dots, \alpha_p)'$, q and $\theta_{(i,t)}$, (2.1) defines a $(p - 1)$ -dimensional surface in \mathfrak{R}_+^p corresponding to all the p -dimensional vectors of outputs $y_{(i,t)}$ that are technologically equivalent. In other words, (2.1) plots the production equivalence surface. By way of illustration, consider $p = 2$ (*i.e.* two outputs); Figure 1 depicts the combinations of outputs that are technologically equivalent (*i.e.* with the same value of $\theta_{(i,t)}$, which is unity in the figure) for different values of α_1 and q .

In the literature, the transformation in (2.1) is sometimes referred to as “constant elasticity of transformation” (see Powell and Gruen, 1968 and Kumbhakar, 1987), where the elasticity of transformation of any two outputs is given by $1/(1 - q)$. Restricting $q > 1$ ensures negativity of the latter. Our parameterization in (2.1) is different from the one in Kumbhakar (1987) and assigns clearly defined roles to α and q . As will become clear, α ensures invariance to rescaling of the individual outputs, which is crucial since a natural common measurement unit for all p components is often lacking. On the other hand, q

serves to describe the elasticity of transformation, regardless of the value of α . This makes the adoption of prior independence between α and q (see Section 3) more reasonable.

Equation (2.1) thus defines a transformation from the multivariate output vector $y_{(i,t)}$ to the univariate quantity $\theta_{(i,t)}$. Given this transformation (the parameters of which we estimate from the data) the basic problem of finding firm-specific efficiencies is essentially the same as in the single-output case. If we interpret the value $\theta_{(i,t)}$ as a kind of “aggregate output”, then it is sensible to define $\delta_{(i,t)} = \log(\theta_{(i,t)})$, group these transformed outputs in an NT -dimensional vector

$$\delta = (\delta_{(1,1)}, \delta_{(1,2)}, \dots, \delta_{(1,T)}, \dots, \delta_{(N,T)})', \quad (2.2)$$

and model δ through the following stochastic frontier model:

$$\delta = V\beta - Dz + \sigma\varepsilon. \quad (2.3)$$

In the latter equation, $V = (v(x_{(1,1)}), \dots, v(x_{(N,T)}))'$ denotes an $NT \times k$ matrix of exogenous regressors, where $v(x_{(i,t)})$ is a k -dimensional function of the inputs $x_{(i,t)}$ corresponding to firm i ($i = 1, \dots, N$) at time t ($t = 1, \dots, T$). The particular choice of $v(\cdot)$ defines the specification of the production frontier: *e.g.* $v(x_{(i,t)})$ is the vector of all logged inputs for a Cobb-Douglas technology, whereas a translog frontier also involves squares and cross products of these logs. The corresponding vector of regression coefficients is denoted by $\beta \in \mathcal{B} \subseteq \mathfrak{R}^k$. Often, theoretical considerations will lead to regularity conditions on β , which will restrict the parameter space \mathcal{B} to a subset of \mathfrak{R}^k , still k -dimensional and possibly depending on x . For instance, we typically want to ensure that the marginal products of inputs are positive.

Technological inefficiency is captured by the fact that firms may lie below the frontier, thus leading to a vector of inefficiencies $\gamma \equiv Dz \in \mathfrak{R}_+^{NT}$, where D is an exogenous $NT \times M$ ($M \leq NT$) matrix and $z \in \mathcal{Z}$ with $\mathcal{Z} = \{z = (z_1, \dots, z_M)' \in \mathfrak{R}^M : Dz \in \mathfrak{R}_+^{NT}\}$. Through different choices of D , we can accommodate various amounts of structure on the vector γ of

inefficiencies. For instance, taking $D = I_{NT}$, the NT -dimensional identity matrix, leads to an inefficiency term which is specific to each different firm and time period. $D = I_N \otimes \iota_T$, where ι_T is a T -dimensional vector of ones and \otimes denotes the Kronecker product, implies inefficiency terms which are specific to each firm, but constant over time (*i.e.* “individual effects”). In our empirical illustrations in Section 5 we make the latter choice for D . Fernández *et al.* (1997) provides a detailed description of other possible choices for D . Since we are working in terms of δ , the log of the aggregate output, the efficiency corresponding to firm i and period t will be defined as $\exp(-\gamma_{(i,t)})$ where $\gamma_{(i,t)}$ is the appropriate element of γ .

Figure 2 provides an illustration of how efficiency is defined for the case with two outputs and $\alpha_1 = \frac{1}{2}$. For a given level of inputs x_0 , Figure 2a depicts the production equivalence curves (drawn lines) corresponding to the frontier for two different values of q ; the inside curve applies when $q = 1.5$ whereas the exterior one is for $q = 5$. Consider a firm producing outputs as in point A while using the inputs x_0 . If we take $q = 1.5$, we can use the corresponding production equivalence curve (dashed line) that passes through A and the efficiency will be given by the value of θ [as defined in (2.1), dropping the subscript (i, t)] in that curve divided by its value in the curve corresponding to the frontier. Note that efficiency can thus be measured anywhere along the production equivalence curves. Similarly, if $q = 5$ the other production equivalence curves apply. Since the transformation θ in (2.1) is homogeneous of degree one in the output vector, our efficiency measure is equivalent to the output distance function measure of Shephard (1970) used in *e.g.* Färe and Primont (1990) and Pittman (1983). Graphically, we can represent the output distance function as the radial measure of technical efficiency given by the ratio OA/OF (which is equal to OB/OD) in Figure 2 for $q = 1.5$, and by OA/OE (=OC/OD) for $q = 5$. For the simple case where we only have one input, Figure 2b puts these efficiencies in the context of the production frontier $h(x) = \exp\{v(x)'\beta\}$, where the distance OD is the value corresponding to input level x_0 . The production of the firm that produces A is now

transformed to either point B (for $q = 1.5$) or point C (for $q = 5$) and the corresponding efficiency is measured as above.

It should be clear from Figure 2a that, given α and q , (2.1) induces an ordering of points in p -dimensional output space in terms of the values of θ . Once this ordering is established, we can use the univariate stochastic frontier in (2.3).

Finally, the model in (2.3) also captures the fact that the frontier is not known exactly, but needs to be estimated from the data. This leads to the two-sided error term ε . Following usual practice, we shall assume in the sequel that ε follows an NT -dimensional Normal distribution with zero mean and identity covariance matrix. Thus, ε corresponds to NT independent replications from the standard univariate Normal distribution. In addition, we allow for a scale factor $\sigma \in \mathfrak{R}_+$. The resulting Normal probability density function (p.d.f.) of δ given (β, z, σ) shall be denoted as follows:

$$p(\delta|\beta, z, \sigma) = f_N^{NT}(\delta|V\beta - Dz, \sigma^2 I_{NT}). \quad (2.4)$$

The essence of the multi-output stochastic frontier model is described through equation (2.3), which extends both the concept of technological frontier as well as inefficiency to the situation of multiple outputs. However, due to the multivariate nature of the problem, equation (2.4) is not enough to define a sampling density for the observables $y_{(i,t)}$ and conduct posterior inference, since it corresponds to NT independent replications from a univariate (instead of a p -variate) distribution. In other words, (2.4) only provides a likelihood function for δ , but not for the individual $y_{(i,t,j)}$'s.

In order to complete our sampling model, we need to introduce stochastics in the $p - 1$ remaining dimensions through considering the distribution of the outputs within each of the production equivalence surfaces. Defining

$$\eta_{(i,t,j)} = \frac{\alpha_j^q y_{(i,t,j)}^q}{\sum_{l=1}^p \alpha_l^q y_{(i,t,l)}^q}, \quad j = 1, \dots, p, \quad \text{and} \quad \eta_{(i,t)} = (\eta_{(i,t,1)}, \dots, \eta_{(i,t,p)})', \quad (2.5)$$

we assume independent sampling ($i = 1, \dots, N; t = 1, \dots, T$) from

$$p(\eta_{(i,t)}|s) = f_D^{p-1}(\eta_{(i,t)}|s), \quad (2.6)$$

where $s = (s_1, \dots, s_p)' \in \mathfrak{R}_+^p$ and $f_D^{p-1}(\eta|s)$ denotes the p.d.f. of a $(p-1)$ -dimensional Dirichlet distribution with parameter s [see Appendix 1]. Note that $\eta_{(i,t)}$ can loosely be interpreted as a vector of output shares, $\sum_{j=1}^p \eta_{(i,t,j)} = 1$ and $0 \leq \eta_{(i,t,j)} \leq 1$. The Dirichlet is a reasonably flexible distribution which satisfies the preceding properties and is commonly used to model shares which sum to one. Other specifications are possible (*e.g.* the additive logistic Normal is used in Osiewalski, 1997), and these can be accommodated in the general approach outlined here by making appropriate changes to the MCMC algorithm.

We can now show that equations (2.4) and (2.6) lead to a p.d.f. for the $NT \times p$ matrix of output observations, Y , where all the elements of Y are positive. Equations (2.4) – (2.6) lead to the following sampling density for Y :

$$p(Y|\beta, z, \sigma, \alpha, q, s) = f_N^{NT}(\delta|V\beta - Dz, \sigma^2 I_{NT}) \prod_{i,t} f_D^{p-1}(\eta_{(i,t)}|s) \prod_{i,t,j} q^{1-(1/p)} \frac{\eta_{(i,t,j)}}{y_{(i,t,j)}}, \quad (2.7)$$

where $i = 1, \dots, N; t = 1, \dots, T; j = 1, \dots, p$. Note that the first two factors in (2.7) come directly from (2.4) and (2.6), whereas the last factor is the Jacobian of the transformation.

3 The Prior

In order to conduct Bayesian inference, we need to complement the likelihood function in (2.7) with a prior distribution on the parameters $(\beta, z, \sigma, \alpha, q, s)$. We shall choose a proper prior distribution with the following product structure

$$p(\beta, z, \sigma, \alpha, q, s) = p(\beta)p(z)p(\sigma)p(\alpha)p(q)p(s). \quad (3.1)$$

It is worthwhile briefly noting that use of improper priors in stochastic frontier models can cause problems (*i.e.* inability to calculate meaningful Bayes factors or even the lack

of existence of the posterior itself, see Fernández *et al.*, 1997, for details). In this paper, we choose values for hyperparameters which imply relatively noninformative but proper priors. In serious applications, the researcher might, alternatively, want to carefully elicit more informative priors. In Section 5 we will investigate the sensitivity of posterior inference with respect to our choices of prior hyperparameters.

Prior for z :

The distribution of z determines the distribution of the inefficiency vector $\gamma = Dz$. In many situations, it may be reasonable to assume that inefficiencies depend on some exogenous firm characteristics or on the time period, described through variables in some matrix $W = (w_{lj})$. Here $j = 1, \dots, r$ indexes the number of explanatory variables and $l = 1, \dots, M$ refers to the dimension of z as described in Section 2.

In order to implement this, we introduce a hierarchical structure which adds an r -dimensional extra parameter vector $\phi = (\phi_1, \dots, \phi_r)' \in \mathfrak{R}_+^r$. Given ϕ , z has the following p.d.f. with support on \mathcal{Z} :

$$p(z|\phi) \propto \prod_{l=1}^M f_G(z_l|1, \lambda_l(\phi)) I_{\mathcal{Z}}(z), \quad (3.2)$$

where $f_G(z|a, b)$ denotes the p.d.f. of a Gamma distribution with mean a/b and variance a/b^2 . Note that, when \mathcal{Z} has a Cartesian product structure, the M components of z are independent given ϕ . In general, we take $\lambda_l(\phi)$ to depend on ϕ in the following way

$$\lambda_l(\phi) = \prod_{j=1}^r \phi_j^{w_{lj}}, \quad (3.3)$$

where w_{lj} are (zero-one) dummy variables with $w_{l1} = 1$. The zero-one character of the variables w_{lj} is useful in that it leads to a very tractable conditional distribution, as we shall see in Section 4.

The r components of ϕ are taken to be a priori independent with p.d.f.

$$p(\phi) = \prod_{j=1}^r f_G(\phi_j|e_j, g_j), \quad (3.4)$$

with positive prior hyperparameters e_j and g_j . In all our empirical illustrations we choose $e_1 = 1$ and $g_1 = -\log(0.80)$. When $r = 1$ and D is a selection matrix (*i.e.* each row of D has one single nonzero entry, which is unity), as is commonly the case, these values imply a relatively flat prior on the individual efficiencies with prior median efficiency set at 0.80, as discussed in Koop *et al.* (1997). In our empirical illustration using banking data we take $r = 3$, and we additionally set $e_j = g_j = 1$ ($j = 2, 3$). These are relatively noninformative values which centre the prior for ϕ_j over 1. Since the ϕ_j 's enter in product form [see equation (3.3)], this prior is centred over the point where w_{lj} ($j = 2, 3$) has no effect on efficiency.

It is worth mentioning that this specification for the inefficiency is most reasonably interpreted as a prior, but classical econometricians would probably interpret it as part of the likelihood function. With hierarchical models, there is some arbitrariness as to what gets labelled “prior” information and what gets labelled as the “likelihood”. Bayesian inference, however, is entirely unaffected by the particular labelling one favours. See Fernández *et al.* (1997) for a discussion about the possible dangers of considering the marginal likelihood obtained by integrating out z .

Finally, we remark that the prior in (3.2) assumes that $\mathcal{Z} \subseteq \mathfrak{R}_+^M$, which is the case for virtually any reasonable empirical application. However, if some components of z can take negative values, we need to consider prior distributions for z that reflect this (*e.g.* a Normal prior truncated to \mathcal{Z}). Our results can easily be modified for this situation.

Prior for β :

The prior assumed on the frontier parameters has p.d.f.

$$p(\beta) \propto f_N^k(\beta|b_0, H_0^{-1})I_{\mathcal{B}}(\beta), \quad (3.5)$$

i.e. a k -dimensional Normal distribution with mean b_0 and covariance matrix H_0^{-1} , truncated to the regularity region \mathcal{B} . In our application, we assume a Cobb-Douglas form for the frontier. Hence, $I_{\mathcal{B}}(\beta)$ corresponds to simply restricting the elements of β (except the inter-

cept) to be non-negative. In our empirical illustrations we set $b_0 = 0_k$ and $H_0 = 10^{-4} \times I_k$. Since the elements of β can be interpreted as factor elasticities they are likely less than 1. Due to the logarithmic transformation, the intercept will also certainly be much less than one prior standard deviation from the mean. Hence, our prior is quite noninformative.

Prior for σ :

We define the prior distribution on the scale σ , through a Gamma distribution on the precision $h = \sigma^{-2}$:

$$p(h) = f_G(h|n_0/2, a_0/2). \quad (3.6)$$

In our empirical illustrations, we set $n_0/2 = 1$ (which leads to an Exponential prior for h) and $a_0/2 = 10^{-6}$. These values imply large prior uncertainty.

Prior for α :

Since the components of α are all in the interval $(0, 1)$ and sum up to one, an obvious choice of a prior distribution is a Dirichlet with p.d.f.

$$p(\alpha) = f_D^{p-1}(\alpha|a), \quad (3.7)$$

where the hyperparameter $a = (a_1, \dots, a_p)' \in \mathfrak{R}_+^p$. We use the diffuse choice of $a = \iota_p$, which makes the prior Uniform over the $(p - 1)$ -dimensional unit simplex.

Prior for q :

For q we take an Exponential prior truncated to the interval $(1, \infty)$ in accordance with economic theory,

$$p(q) \propto f_G(q|1, d)I_{(1, \infty)}(q). \quad (3.8)$$

A diffuse choice for d , used in our empirical illustrations, is $d = 10^{-6}$.

Prior for s :

We assume p independent Gamma distributions for the components of s :

$$p(s) = \prod_{j=1}^p p(s_j) = \prod_{j=1}^p f_G(s_j|b_j, c_j). \quad (3.9)$$

In our empirical illustrations, we make the diffuse choice of $b_j = 1$ and $c_j = 10^{-6}$ for all j .

4 Bayesian Inference and MCMC Algorithm

The likelihood function in (2.7) together with the prior distribution defined through (3.1) – (3.9) defines a Bayesian model, *i.e.* a joint distribution of the observables and parameters. Posterior inference is based on the conditional distribution of the parameters given the observables. Since this posterior distribution is not analytically tractable, we shall employ an MCMC sampler to generate drawings from it. We shall implement this by partitioning the parameter components into several blocks and sequentially drawing according to conditional distribution of each of these blocks given the data and the remaining parameter blocks². In this section we describe all these conditional distributions and indicate which methods were used for random variate generation. The parameter ϕ introduced in Section 3 [see (3.2) – (3.4)] will also be included in the sampler as it simplifies the drawing from z , and is of interest in itself when $r > 1$.

Conditional posterior distribution of z :

The conditional posterior distribution of z has p.d.f. proportional to the product of (2.7) and (3.2). In the practically relevant situation where D is of full column-rank, the latter product leads to

$$p(z|Y, \phi, \beta, \sigma, \alpha, q, s) \propto f_N^M(z|m, R)I_Z(z), \quad (4.1)$$

i.e. a truncated M -variate Normal distribution, with mean

$$m = (D'D)^{-1}\{D'(V\beta - \delta) - \sigma^2\lambda(\phi)\}, \text{ where } \lambda(\phi) = (\lambda_1(\phi), \dots, \lambda_M(\phi))', \quad (4.2)$$

and covariance matrix

$$R = \sigma^2(D'D)^{-1}. \quad (4.3)$$

²For some blocks these conditionals are easy to draw from and then we shall draw from them directly (Gibbs steps), while for others the conditionals are non-standard distributions. In the latter case, we shall use random walk Metropolis samplers to generate the corresponding drawings. It can easily be shown that this leads to a Markov chain with the posterior distribution as its limiting distribution. See *e.g.* Tierney (1994).

Drawing from (4.1) is much simplified if $D'D$ is diagonal (as is the case in our applications), since then all elements of z are conditionally independent and can be drawn separately.

Conditional posterior distribution of ϕ :

We shall draw separately each of the r components, ϕ_j , conditioning upon the observables and the remaining parameters (also including the remaining components of ϕ). The p.d.f. of this distribution is proportional to the product of (3.2) and (3.4). The fact that the variables w_{lj} only take the value zero or one, leads to the following simple expression for the p.d.f. of this conditional distribution:

$$p(\phi_j|Y, \{\phi_h, h \neq j\}, z, \beta, \sigma, \alpha, q, s) = f_G \left(\phi_j | e_j + \sum_{l=1}^M w_{lj}, g_j + \sum_{l=1}^M (w_{lj} z_l \prod_{h \neq j} \phi_h^{w_{lh}}) \right). \quad (4.4)$$

Conditional posterior distribution of β :

Multiplying (2.7) and (3.5) leads to the following truncated Normal p.d.f. for β :

$$p(\beta|Y, z, \phi, \sigma, \alpha, q, s) \propto f_N^k(\beta|b_*, H_*^{-1}) I_{\mathcal{B}}(\beta), \quad (4.5)$$

where

$$H_* = H_0 + \sigma^{-2} V'V, \quad (4.6)$$

and

$$b_* = H_*^{-1} \{H_0 b_0 + \sigma^{-2} V'(\delta + Dz)\}. \quad (4.7)$$

Conditional posterior distribution of σ :

We shall compute the conditional posterior distribution of $h = \sigma^{-2}$. The latter distribution has p.d.f. proportional to the product of (2.7) and (3.6), which results in

$$p(h|Y, z, \phi, \beta, \alpha, q, s) = f_G \left(h | \frac{n_0 + NT}{2}, \frac{a_0 + (\delta - V\beta + Dz)'(\delta - V\beta + Dz)}{2} \right). \quad (4.8)$$

Conditional posterior distribution of α :

We remind the reader that this is a $(p - 1)$ -dimensional distribution, where each of the p components of α are in between zero and one and sum to one [see (2.1)]. From (2.7) and (3.7) we obtain the following p.d.f. for this conditional distribution:

$$p(\alpha|Y, z, \phi, \beta, \sigma, q, s) \propto \prod_j \alpha_j^{a_j + s_j q NT - 1} \prod_{i,t} \left(\sum_j \alpha_j^q y_{(i,t,j)}^q \right)^{-\sum_j s_j} \exp \left\{ -\frac{1}{2\sigma^2} (\delta - V\beta + Dz)' (\delta - V\beta + Dz) \right\}, \quad (4.9)$$

where, as before, $i = 1, \dots, N$; $t = 1, \dots, T$; $j = 1, \dots, p$, and with the domain of α indicated at the beginning of this paragraph. Note that δ , defined through (2.1) – (2.2) depends on α and, therefore, leads to a non-trivial factor for this conditional distribution.

In contrast to all the conditional distributions presented so far, drawing from the conditional distribution in (4.9) is not straightforward. There are many ways of drawing from non-standard distributions and we have experimented with some of them (in dimensions up to $p = 4$). In our experience, a simple random walk Metropolis-Hastings algorithm³ works very well and we chose the latter for our applications. In particular, we use a $(p - 1)$ -variate Normal candidate-generating density (proposals outside the support of α are never accepted). The mean of the Normal is given by the previous draw and the variance is calibrated so that the acceptance probability is reasonable.⁴ The acceptance probability is merely the ratio of (4.9) evaluated at the new candidate draw to (4.9) evaluated at the last accepted draw.

³In a Metropolis-Hastings algorithm we draw candidate values from an arbitrary distribution and we switch to the candidate value with a certain probability. Otherwise, the chain stays at the current value. Because we use a symmetric candidate generator, we really use the simpler Metropolis implementation. See *e.g.* Chib and Greenberg (1995) for an intuitive explanation.

⁴Chib and Greenberg (1995) discuss how the optimal acceptance probability is roughly 0.45 in one dimensional problems and this goes to 0.23 as the number of dimensions goes to infinity. These numbers are relevant to the case where the target and candidate generating densities are Normal. These conditions do not exactly fit the present case, but nevertheless we use the above numbers as rough guidelines.

Conditional posterior distribution of q :

The product of (2.7) and (3.8) leads to the following p.d.f. for $q > 1$:

$$p(q|Y, z, \phi, \beta, \sigma, \alpha, s) \propto q^{NT(p-1)} \exp(-dq) \exp\{-A(q)\} I_{(1, \infty)}(q), \quad (4.10)$$

where

$$A(q) = \frac{1}{2\sigma^2} (\delta - V\beta + Dz)' (\delta - V\beta + Dz) + \sum_{i,t,j} s_j \log \left(\frac{\sum_{l=1}^p \alpha_l^q y_{(i,t,l)}^q}{\alpha_j^q y_{(i,t,j)}^q} \right), \quad (4.11)$$

where, as usual, $i = 1, \dots, N$; $t = 1, \dots, T$; $j = 1, \dots, p$. Note that, from (2.1), δ depends on q . In this case, too, we use a random walk Metropolis-Hastings algorithm with a Normal candidate-generating density calibrated as above.

Conditional posterior distribution of s :

We shall draw separately each of the p components of s . The conditional posterior distribution of s_j , $j = 1, \dots, p$ (also given the remaining components of s) has p.d.f. on $(0, \infty)$:

$$p(s_j|Y, z, \phi, \beta, \sigma, \alpha, q, \{s_h : h \neq j\}) \propto \frac{\Gamma(\sum_l s_l)^{NT} s_j^{b_j-1}}{\Gamma(s_j)^{NT}} \exp \left[-s_j \left\{ c_j + \sum_{i,t} \log \left(\frac{\sum_l \alpha_l^q y_{(i,t,l)}^q}{\alpha_j^q y_{(i,t,j)}^q} \right) \right\} \right], \quad (4.12)$$

where $i = 1, \dots, N$; $t = 1, \dots, T$; $l = 1, \dots, p$. In order to draw from the latter distribution, a random walk Metropolis-Hasting algorithm is used in the same manner as for q or α .

In our empirical illustrations, we take a final run of 15,000 replications from our MCMC algorithm and discard the initial 5,000 to mitigate start-up effects. Preliminary runs are used to calibrate the variances/covariances for our Normal candidate generating densities used in the Metropolis-Hastings algorithms. Careful study of the numerical properties of MCMC algorithms in stochastic frontier models in previous work (*e.g.* Koop *et al.*, 1997, 2000), indicated that posteriors are well-behaved and convergence is easy to achieve. Results for the present model indicate similarly good computational properties for the multi-output case. In Appendix 2 we shall comment on some formal convergence criteria for our applications.

5 Applications

5.1 Simulated Data

To evaluate the possibilities for conducting inference, it is useful to first consider the performance of our Bayesian methodology with artificial data. For the sake of brevity, we present results here for only one simulated data set. Results for other data sets reinforce the general message of this section: that Bayesian methods are relatively easy to apply and yield sensible and accurate results.

Analyzing simulated data allows us to empirically verify three crucial issues:

- * all parameters are identified and reasonably precise inference can be conducted on the basis of a sample that is of an order of magnitude likely to occur in practice;
- * the influence of the prior is not overly strong, and does not dominate the sample information;
- * the numerical properties of the MCMC algorithm used are satisfactory and convergence is achieved with the number of draws used.

On all these counts, the performance is quite good, as evidenced by the results.

We generate an artificial data set with $T = 10$ time periods, $N = 400$ individuals, $p = 3$ outputs and $k = 2$ regressors with coefficients given by $\beta = 0.5 \times \iota_2$. The explanatory variables in the 4000×2 matrix V in (2.3) are drawn in a simple way: The first explanatory variable (*i.e.* the first column of V) is a vector of ones corresponding to the intercept. The second column we draw as follows: First we generate a drawing from a $N(0,1)$ distribution for each of the $N = 400$ individuals and take the exponential. We treat the resulting numbers as the initial levels of the second explanatory variable for each of the N individuals. Then we assume, for each individual, that inputs evolve over time according to a Normal random walk with standard deviation equal to 0.01. If negative values for inputs are obtained, they are discarded and redrawn. In this way, we complete the matrix V . The scale σ in (2.3) is taken to be equal to 0.5, whereas the matrix $D = I_N \otimes \iota_T$, *i.e.* we assume

individual-specific efficiencies, and we take $r = 1$ in (3.3) with $\phi = 5$. In addition, $q = 2$, $\alpha = (1/3)\iota_3$ and $s = \iota_3$.

Table 1 presents the median, the interquartile range and the 0.025 and 0.975 quantiles of the posterior for all parameters. We note that posterior medians are all near the true values, relative to interquartile ranges. The true values are always contained in the central interval covering 95% of the posterior mass.

The Metropolis-Hastings algorithm used to draw values for α , q and s performs very well indeed: the acceptance probabilities range from 0.180 to 0.545. Evidence on convergence of the Markov chain is presented in Appendix 2.

Finally, Figure 3 presents the marginal posterior densities of the parameters that appear in the production equivalence surfaces [see (2.1)], namely q and α . The corresponding prior densities are also indicated (with appropriate scaling factors to facilitate visual comparison). Clearly, the data evidence dominates the prior and posterior inference is usefully concentrated around the values used to generate the data.

The influence of the prior can be assessed more fully by conducting a prior sensitivity analysis. We tried various other values of the prior hyperparameters in (3.4)-(3.9) and concluded that posterior results are virtually unaffected by substantial changes in the prior. For example, changing prior median efficiency to 0.5 or to 0.95, taking $H_0 = I_k$ and multiplying a_0 , d and c_j by a factor 1000 leaves the results unchanged.

5.2 Banking Data

We now apply our Bayesian methods to a data set that has been used in Berger (1993) and Adams *et al.* (1996, 1999). This banking data set contains observations on $N = 798$ limited branching banks in the United States for $T = 10$ years (1980-89). The data set contains $p = 3$ outputs (real estate loans; commercial and industrial loans; installment loans), 5 inputs (average number of employees; physical capital; purchased funds; demand deposits; retail, time and savings deposits) and one explanatory variable for bank-specific efficiency:

w_{l2} = bank size, $l = 1, \dots, N$ [see (3.2)-(3.3)]. Adams *et al.* (1996) treat w_{l2} as an input, but we feel it more reasonably might be something which affects efficiency. In order to investigate whether there is any influence of the input mix on the efficiency distribution (which could occur if inefficient firms tend to choose suboptimal input combinations, as suggested by a referee), we have also included dummies⁵ for input ratios. We found that only one of those, namely the ratio between physical capital and retail, time and savings deposits, has a substantial effect on the inefficiency distribution. Posterior odds in favour of excluding each of the other regressors are of the order 10 to 1.⁶ Thus, in our final specification, both bank size (w_{l2}) and the latter input ratio (w_{l3}) are included as variables affecting $\lambda_l(\phi)$ in (3.3).

We shall follow Adams *et al.* (1996, 1999) in assuming a Cobb-Douglas technology for the frontier. Adding an intercept, this leads to $k = 6$ regressors in (2.3), where $v(\cdot)$ is a 6-dimensional function with the first component equal to unity and the remaining ones corresponding to the logarithms of each of the 5 inputs mentioned above. Furthermore, we take $D = I_N \otimes \iota_T$, *i.e.* bank-specific efficiencies which are constant over time.

The main purpose of the present empirical illustration is to show that our methods are computationally feasible in a serious application of the sort that is currently being done in the literature. We do not attempt to contribute to the banking efficiency literature, and hence our discussion of data issues is superficial. The interested reader is referred to Berger (1993) for details and to Berger and Humphrey (1997) for an extensive survey of the literature.

⁵As discussed in the previous section, it is convenient if the explanatory variables for efficiency are 0-1 dummies. Hence, w_{l2} is a dummy variable which equals 1 if a bank's time-averaged size is above the all-bank average, and the input ratio dummies are defined in a similar way. In constructing the latter, we have arbitrarily chosen the retail, time and savings deposits as the denominator, *i.e.* we limit ourselves to four ratios of the 10 theoretical possibilities, in order to keep r small.

⁶In this case, we can use the Savage-Dickey density ratio, which gives the Bayes factor in favour of the point null hypothesis and can easily be computed as the ratio of the marginal posterior versus the prior at $\phi_j = 1$. We assumed unitary prior odds for all models considered. See Verdinelli and Wasserman (1995) for a discussion.

One advantage of the Bayesian approach is that the entire posterior p.d.f. (or its properties) of any firm-specific efficiency [*i.e.* $\exp(-z_i)$] can be calculated. In contrast, classical approaches (such as Adams *et al.* 1996, 1999) construct only point estimates for firm specific efficiency (and the consistency of these estimates depends on T , which can be small in many applications; here $T = 10$)⁷. Furthermore, in Adams *et al.* (1996, 1999), efficiency distributions are calculated which are kernel smoothed histograms of these point estimates of firm-specific efficiency. In previous work (*e.g.* Koop *et al.*, 1997), we have argued that a better measure of an overall efficiency distribution is based on a predictive notion. Given our individual effects choice for D , equation (3.2) implies that the inefficiency z_f of a hypothetical, unobserved firm is given by

$$p(z_f|\phi) = f_G(z_f|1, \lambda_f(\phi)), \quad (5.1)$$

where $\lambda_f(\phi)$ is defined in (3.3) and depends on $r - 1$ firm characteristics (w_{f2}, \dots, w_{fr}) . If we now integrate out the parameter ϕ in (5.1) using its posterior distribution, we are left with a predictive distribution with p.d.f.

$$p(z_f|Y) = \int p(z_f|\phi)p(\phi|Y)d\phi$$

for the inefficiency of a hypothetical out-of-sample firm. Using the relationship between inefficiency and efficiency [efficiency is equal to $\exp(-z_f)$], we can transform this into a measure of predictive efficiency. We refer to this out-of-sample firm efficiency as “predictive efficiency”. Note that it better captures the uncertainty inherent in any measure of the dispersion of efficiencies across firms than the alternative of merely plotting point estimates (without any standard errors) for each firm. In other words, it directly answers the question: “Given all the information contained in the data about the efficiencies of every firm, what can we say about the efficiency of a new, as yet unobserved, firm in that industry?”.

⁷However, see Kim and Schmidt (1999) for an empirical comparison between inference on efficiency using some more sophisticated classical methods and Bayesian methods.

Table 2 presents posterior quantiles for the parameters, returns to scale and the elasticity of transformation. In addition, we present results for predictive efficiency distributions,⁸ the efficiency distributions of three representative observed firms⁹ as well as quantities relating to the deviation with respect to the frontier. To motivate the latter, note that the total deviation of firm i at time t from the frontier is given by $u_{(i,t)} = \sigma\varepsilon_{(i,t)} - z_i$. We define squared deviation from the frontier as: $SDF = E(u_{(i,t)}^2 | \sigma, \phi) = \sigma^2 + 2\lambda_i(\phi)^{-2}$. In the case where we have no explanatory variables in the efficiency distribution (*i.e.* $r = 1$), $\lambda_i(\phi) = \phi$ does not vary over firms and SDF is a single measure. Since firm characteristics enter the efficiency distribution in our application, $\lambda_i(\phi)$ is different for “large” and “small” firms and we compute two measures of SDF. We can decompose SDF into the part due to the symmetric error (σ^2) and that due to the one-sided error. The last rows of Table 2 present the fraction of SDF due to the symmetric error labelled as “Fraction from ε ”. Note that these measures are random variables, and we plot their posterior quantiles. Figure 4 plots the posterior p.d.f.’s of some key parameters as well as some efficiency measures. The Metropolis-Hastings algorithms used accept in between 21% and 60% of all proposed values.

Adams *et al.* (1996, 1999) use identical data, but a different methodology. Some of our parameters are incomparable to their’s. For the frontier parameters we obtain posterior medians which are roughly in accordance with the classical point estimates in Adams *et al.* (1999), where bank size is not included in the analysis, but quite different from those in Adams *et al.* (1996), who treat bank size as an input. The main differences with the latter

⁸Since $r > 1$, we have to make a choice for which type of firm we calculate the predictive efficiency distribution. We chose to present results for what turn out to be the the most extreme of the four possible firm types: a big firm with high capital/retail, time and savings deposits ratio ($w_{12} = w_{13} = 1$) and a small firm with a low ratio between physical capital and retail, time and savings deposits ($w_{12} = w_{13} = 0$). For ease of reference, we shall refer to these firm types in the text as “large” and “small”.

⁹These representative firms are chosen based on a crude measure of size of output. That is, we simply sum all three outputs together to get a measure of aggregate output. We then choose the three firms which have the minimum, median and maximum aggregate output levels, respectively.

paper are for physical capital, where we find a much smaller elasticity, and for retail, time and savings deposits, where we find a larger elasticity than they do. Another result relating to parameter estimates worth noting is that we obtain only slightly increasing returns to scale (*i.e.* the posterior median of returns to scale is 1.043 with an interquartile range of 0.011), whereas Adams *et al.* (1996) find substantially increasing returns to scale.

The w_{ij} variables seem to have some effect on the efficiency distribution. Both ϕ_2 and ϕ_3 are quite different from 1. This can be seen clearly in Figure 4d and from the fact that the 0.025 quantile of the posterior of ϕ_2 is close to 1 and that of ϕ_3 exceeds 1, implying a predominantly positive effect on mean efficiency if the dummies are unity (*i.e.* for “large” firms).

The shape of the production equivalence surface depends on q and α . The median of q is 1.025 and its spread is fairly small. A glance at Figure 1 indicates the virtually linear shape for the production equivalence surface implied by values in this range. The implied elasticity of transformation between any two outputs tends to take rather large negative values. Furthermore, α_3 is larger than α_1 and α_2 , the effect of which can be visualised by comparing Figure 1a to 1b. Loosely speaking, the third output (installment loans) is receiving more weight in the output aggregator given by (2.1). Alternatively put, we need to trade in more than one dollar of real estate or commercial loans¹⁰ to produce one extra dollar of installment loans with any given set of inputs. Plots of the posteriors of these parameters are given in Figures 4b and c. In contrast, Adams *et al.* (1996, 1999) put most weight on real estate loans.

With regards to SDF, note that regardless of whether we use the efficiency distribution for a large or small firm, the symmetric error always has a very small role to play. In particular, it virtually never accounts for more than 7.5 per cent of the total expected squared deviation.

The focus of most stochastic frontier applications is on efficiency and, as we have argued

¹⁰At median values, this amounts to almost \$2 of real estate loans or commercial loans

above, it is here where the Bayesian approach makes some important contributions. Figure 4a plots the posteriors for three representative observed firms (see footnote 9) and it can be seen that these p.d.f.'s are relatively spread out, especially for the bank with maximum aggregate output. Merely presenting point estimates of firm specific efficiencies misses this important information. In order to compare the efficiencies of different firms, we could simply define as a parameter of interest any function of their respective efficiencies (such as the difference) and conduct inference on its quantiles or plot the entire p.d.f. of this quantity. Given the sensitive policy issues which often result when one ranks individual firms based on point estimates of efficiency, the ability to plot the entire p.d.f. of a firm's efficiency is highly desirable.. Figure 4a clearly indicates that the largest bank (in the sense of footnote 9) is more efficient than the median bank, which in turn is more efficient than the smallest bank.

Figure 4a also plots predictive efficiency distributions for, as yet, unobserved firms¹¹. We consider both a large and a small bank (as defined in footnote 8). These distributions are obviously much more dispersed than their counterparts for observed firms, as we can only learn about the inefficiency z_f of an unobserved firm through the posterior distribution of ϕ . Although, as described above, these predictive efficiency distributions have a different interpretation from the "Distributions of Relative Efficiencies" given in Adams *et al.* (1996, 1999), our plots are roughly similar to theirs in location, yet even more dispersed¹². In other words, a wide variety of efficiency behaviour seems to be present in this data set. Tail quantiles of the posterior distributions clearly indicate that almost no useful inference on efficiency can be conducted without having observed outputs and inputs of the particular bank. Some information can be gleaned from knowing the type of bank as measured by W

¹¹In order to enhance the informational content of these plots, we have multiplied the scale by a factor 10.

¹²The fact that our distributions are more dispersed than those in Adams *et al.* (1996, 1999) might be due to the fact that we are incorporating parameter uncertainty in our firm-specific efficiencies, but we should also bear in mind the difference in interpretation.

(note the difference in both distributions of z_f), but a lot of uncertainty remains if we can not identify the bank.

In order to have an estimable model, we needed to make a host of assumptions outlined in Sections 2 and 3. Most of these are easily testable by simple extensions of the model (see the comments in the Conclusion), but relaxing some would require a much larger effort. However, we can produce a diagnostic check of the overall adequacy of our Bayesian model using a predictive device. In the same fashion as we derive predictive efficiency distributions based on (5.1), we can generate a full predictive distribution for a set of p -dimensional firm outputs, say y_{f1}, \dots, y_{fn} , by just integrating the sampling model with the posterior. This is easily implemented through our MCMC method, and we can randomly select a prediction sample from the full sample of 798 firms and compare the predictive distribution for each individual component of y_{f1}, \dots, y_{fn} at each time period with the actually observed number. Recording at which predictive percentile the observations occur allows us to contrast the empirical (observed) distribution with the predictive distribution that our model generates (*e.g.* the fraction of the observations that are less than the first percentile of their corresponding predictive distribution should ideally be close to 1%). Thus, all model assumptions (both the sampling model and the prior) are then subjected to a very challenging test, which is the confrontation with observed reality. We have chosen $n = 10$ firms at random, which leaves us with $n \times T = 100$ observations to be predicted for each individual output component. The plots in Figure 5 are calibration plots, which compare the predictive percentiles¹³ with the empirical ones generated by these 100 observations, for each output component separately. Even though components 2 and 3 are not as accurately predicted as component 1, the performance is still quite acceptable, in our opinion. This illustrates the adequacy of our model in capturing the main features of these data and provides considerable (diagnostic) evidence for an overall validation of all our modelling assumptions. Such predictive assessment could also be conducted on the

¹³We examine the 1st, 2nd, 5th, 10th, 25th, 50th, 75th, 90th, 95th, 98th, and 99th percentiles.

basis of tail areas for a particular measure of discrepancy, leading to posterior predictive p -values as in Meng (1994) and Gelman, Meng and Stern (1996). Alternatively, properties of realized error terms in the sense of Zellner (1975) could be analyzed.

The present analysis was conducted without any strong subjective prior input, except for the fact that we elicited a prior median efficiency of 0.8 and used that in the weakly informative prior in (3.4). Other prior information that might be at hand concerns the frontier parameters; as we have little substantive prior information to bring to bear on this (besides, of course, the regularity conditions), we have based the presented results upon the very flat prior in (3.5) with $b_0 = 0$ and $H_0 = 10^{-4} \times I_k$. However, we have also verified that changing this to the much more informative prior with prior mean elasticities equal to 0.2 and $H_0 = I_k$, does not affect the results. If, in addition, we also multiply the values of a_0 , d and c_j by a factor 1000 and base (3.5) on a prior median efficiency of 0.95 or 0.5, we find that the results are, again, virtually unaffected. This testifies to the fact that the data are really driving the results presented here.

We have found that researchers in the area are sometimes worried about the computational demands of the MCMC algorithms for Bayesian analyses of stochastic frontier models. In this section, we have worked with a large data set of empirical relevance. On a 500 Mhz. Pentium-III PC, 15,000 replications takes roughly 5 hours using non-optimized code in Gauss-VM (including a predictive calibration plot). This is a substantial, but not prohibitive, computational cost (*i.e.* a few short runs of less than an hour each to calibrate the Metropolis-Hastings algorithm followed by one final overnight run suffice to estimate the model). Furthermore, calibration of the Metropolis-Hastings steps proved to be very easy and convergence occurred quite quickly (see Appendix 2). Smaller data sets (*e.g.* as used in Koop *et al.*, 2000) have commensurately lower computational costs. But the bottom line worth stressing is that, for data sets of the size used in serious empirical applications, our methods are computationally practical.

6 Conclusion

In this paper, we consider econometric estimation of multiple-output production frontiers when only data on inputs and outputs are available. Little econometric work has been done in this area and two problems hinder simple regression-based inference. Firstly, the production frontier only defines one equation, but we have multiple endogenous variables. Secondly, the desire to have flexible functional forms leads to models where unknown parameters appear on both sides of the equation defining the production technology. We show how Bayesian methods can be used to surmount these problems and can be implemented using Markov Chain Monte Carlo methods. Empirical illustrations involving simulated and real data indicate the practicality and sensibility of our approach.

The model adopted in this paper embodies a number of assumptions, most of which can fairly easily be relaxed or changed. The only assumption that does not fall in this category is that of separability, mentioned in the Introduction. Relaxing this assumption would change the statistical structure of the model, and constitutes a major future challenge. We agree there are applications where this assumption might be unappealing (say, if different outputs require a different input mix), so we think this is an important research area. Extensions to other production equivalence surfaces [outside those defined in (2.1)] and production frontiers (such as flexible functional forms, see Koop, Osiewalski and Steel, 1994) are easily implemented in our framework. The assumption of homoskedasticity in (2.3) can easily be relaxed with panel data, *e.g.* by making σ time-specific. Such models were tried in Koop, Osiewalski and Steel (2000) and generally found to be too unstructured in a growth of countries application, but other applications (with large N) may render this an interesting extension. Normality in (2.4) could also easily be relaxed, in particular if we consider the class of scale mixtures of Normals (see Geweke, 1993, and Fernández and Steel, 2000, for an obvious extension to the MCMC algorithm in this case). Furthermore, the distribution on the inefficiency term can also be chosen differently, as shown in Koop, Steel

and Osiewalski (1995). Finally, we could try different distributions of the output shares in (2.6), as explained in the text of Section 2. All these model assumptions can be assessed formally through posterior odds and inference on quantities that are not model-specific (such as efficiencies) can be conducted through averaging the inference given each model with the corresponding posterior model probabilities as weights (Bayesian model averaging). We hope the model and methodology explained here will prove a useful framework for investigating the merits of these and alternative assumptions in various empirical contexts.

Appendix 1: Dirichlet Distribution

A $(p - 1)$ -dimensional Dirichlet distribution with parameter $s = (s_1, \dots, s_p)'$ ($s_j > 0$ for all j) has support on

$$\left\{ y = (y_1, \dots, y_{p-1})' \in \mathfrak{R}^{p-1} : 0 \leq y_j \leq 1 \text{ and } \sum_{j=1}^{p-1} y_j \leq 1 \right\}$$

and is given through the p.d.f.

$$p(y) = f_D^{p-1}(y|s) = \frac{\Gamma(\sum_j s_j)}{\prod_j \Gamma(s_j)} \prod_j y_j^{s_j-1}, \quad (A.1)$$

where the index $j \in \{1, \dots, p\}$ and we have defined $y_p = 1 - \sum_{l=1}^{p-1} y_l$.

Appendix 2: Convergence Results

In previous work (e.g. Koop *et al.*, 1995) we have found stochastic frontier models to be well-behaved and our MCMC algorithms to converge quickly. Informal evidence (e.g. from doing many different runs from diverse starting points) indicates that the desirable properties we have found for single-output also hold for multiple-output stochastic frontier models. Nevertheless, it is worthwhile to provide some more formal diagnostics.

Many convergence diagnostics have been suggested and Cowles and Carlin (1996) provide a survey. Popular ones are due to Gelman and Rubin (1992) and Geweke (1992). Gelman and Rubin (1992) suggest taking m parallel MCMC chains from over-dispersed starting values. They then reason that, if the algorithm has converged, the difference in posterior means of any function of interest across chains should be “small”. “Small” is measured relative to the posterior variance of the function of interest. They create a statistic based on B^* = the variance between means from the m parallel chains and W^* = the mean of the m within-chain variances. Here we cannot directly use this approach since we have no proof that means and variances exist. However, we can derive measures that are in the spirit of Gelman and Rubin by replacing “mean” with “median” and “variance” with “interquartile range” in the definitions of W^* and B^* above. These are presented below for all parameters with $m=5$ and 5,000 start-up (or “burn-in”) replications followed by 15,000 recorded replications. The over-dispersed starting values are randomly drawn as follows:

1. $h = \sigma^{-2}, q, \phi_j$ for $j = 1, \dots, r$ and s_i for $i = 1, \dots, p$ are each drawn independently from the Gamma distribution with mean 10 and variance 100. Values of $q < 1$ are discarded and redrawn.
2. β_i for $i = 1, \dots, k$ are drawn independently from the Normal with mean 0 and variance 1. Negative values for any coefficient (other than the intercept) are discarded and redrawn.
3. α is drawn from the Uniform over the unit simplex.

The convergence diagnostic recommended in Geweke (1992) is also based on posterior means and central limit theorems and, hence, cannot be directly used here. Nevertheless, we can present some information based on medians and interquartile ranges which is similar in spirit. In particular, we present the posterior median of each parameter based on the first 10% and last 50% of the replications (after discarding the burn-in replications). If the difference between these is small relative to the interquartile range, this is evidence of convergence.

Table A1: Convergence Information for Simulated Data

	Gelman/Rubin		Geweke		
	W^* =Median IQR	B^* =IQR Medians	Med. first 10%	Med. last 50%	IQR (all reps)
β_1	0.0321	0.0002	0.5079	0.5103	0.0328
β_2	0.0065	0.0001	0.4963	0.4962	0.0066
σ	0.0078	0.0001	0.4968	0.4967	0.0078
α_1	0.0061	0.0001	0.3342	0.3345	0.0062
α_2	0.0063	0.0002	0.3341	0.3344	0.0061
α_3	0.0063	0.0002	0.3317	0.3313	0.0063
q	0.1067	0.0063	2.0056	1.9817	0.1177
ϕ	0.7217	0.0063	5.9493	5.9982	0.7212
s_1	0.0814	0.0035	0.9640	0.9837	0.0906
s_2	0.0816	0.0074	0.9696	0.9924	0.0894
s_3	0.0823	0.0036	0.9666	0.9814	0.0916

Table A1 indicates that convergence has been achieved for the artificial data. The difference in posterior medians across chains with widely varying starting values is very small relative to the posterior interquartile ranges. In particular, B^* is at least 10 times as small as W^* for every parameter and is usually more than 100 times as small. The last three columns, which present evidence relating to Geweke's (1992) convergence diagnostic, also indicate convergence. That is, the differences between posterior medians at the beginning and the end of the chain is very small relative to the posterior interquartile range for every parameter.

Table A.2 presents the same information for the banking data. Results using the Gelman/Rubin approach are not quite as strong as for the simulation data, but nevertheless convergence is indicated. As might be expected, the parameters which relate to the production equivalence surface exhibit the poorest convergence properties. However, the fact that posterior results are so similar despite widely differing starting values and the evidence from the Geweke approach indicate that we need not worry about the eventual convergence of the algorithm.

Table A2: Convergence Information for Banking Data

	Gelman/Rubin		Geweke		
	W^* =Median IQR	B^* =IQR Medians	Med. first 10%	Med. last 50%	IQR (all reps)
β_1	0.1088	0.0051	0.4404	0.4395	0.1078
β_2	0.0167	0.0003	0.1777	0.1788	0.0161
β_3	0.0069	0.0002	0.0293	0.0295	0.0072
β_4	0.0041	0.0001	0.0608	0.0608	0.0041
β_5	0.0099	0.0004	0.2980	0.2974	0.0103
β_6	0.0100	0.0001	0.4755	0.4757	0.0101
σ	0.0016	0.0001	0.1418	0.1418	0.0016
α_1	0.0052	0.0012	0.2568	0.2581	0.0058
α_2	0.0059	0.0010	0.2798	0.2786	0.0057
α_3	0.0063	0.0004	0.4636	0.4635	0.0065
q	0.0336	0.0129	1.0252	1.0253	0.0333
ϕ_1	0.1135	0.0035	1.5582	1.5721	0.1131
ϕ_2	0.1271	0.0044	1.1642	1.1554	0.1253
ϕ_3	0.1391	0.0027	1.3413	1.3381	0.1400
s_1	0.2425	0.0834	4.1940	4.1708	0.2427
s_2	0.1186	0.0478	2.2966	2.3313	0.1224
s_3	0.1925	0.0668	3.6875	3.6965	0.1929

References

- Adams, R., Berger, A. and Sickles, R., 1996. Computation and inference in semiparametric efficient estimation, manuscript.
- Adams, R., Berger, A. and Sickles, R., 1999. Semiparametric approaches to stochastic panel frontiers with applications in the banking industry, *Journal of Business and Economic Statistics*, 17, 349-358.
- Berger, A., 1993. "Distribution-free" estimates of efficiency in the U.S. banking industry and tests of the standard distributional assumptions, *Journal of Productivity Analysis*, 4, 261-292.
- Berger, A. and Humphrey, D., 1997. Efficiency of financial institutions: International

- survey and directions for future research, *European Journal of Operational Research*, 98, 175-212.
- Chib, S. and Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm, *The American Statistician* 49, 327-335.
- Cowles, M. and Carlin, B., 1996. Markov chain Monte Carlo convergence diagnostics: A comparative review, *Journal of the American Statistical Association*, 91, 883-904.
- Färe, R. and Primont, D., 1990. A distance function approach to multioutput technologies, *Southern Economic Journal*, 879-891.
- 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.
- Fernández, C. and Steel, M.F.J., 2000. Bayesian regression analysis with scale mixtures of Normals, *Econometric Theory*, 16, forthcoming.
- Gelman, A., Meng, X.L., and Stern, H., 1997. Posterior predictive assessment of model fitness via realized discrepancies, (with discussion) *Statistica Sinica*, 6, 733-807.
- Gelman, A. and Rubin, D., 1992. Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, 457-511.
- Geweke, J., 1992, Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in *Bayesian Statistics 4*, eds. J. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith, Oxford: Oxford University Press, 169-193.
- Geweke, J., 1993, Bayesian treatment of the independent Student- t linear model, *Journal of Applied Econometrics*, 8, S19-S40.
- Horrace, W. and Schmidt, P., 1996, Confidence statements for efficiency estimates from stochastic frontier models, *Journal of Productivity Analysis*, 7, 257-282.
- Jondrow, J., Lovell, C.A.K., Materov, I. and Schmidt, P., 1982. On the estimation of technical inefficiency in the stochastic frontier production model, *Journal of Econometrics*, 19, 233-238.

- Kim, Y. and Schmidt, P., 1999. An empirical comparison of Bayesian and classical approaches to inference on efficiency levels in stochastic frontier models with panel data, manuscript.
- Koop, G., Osiewalski, J., Steel, M.F.J., 1994. Bayesian efficiency analysis with a flexible cost function, *Journal of Business and Economic Statistics*, 12, 93-106.
- Koop, G., Osiewalski, J., Steel, M.F.J., 1997. Bayesian efficiency analysis through individual effects: Hospital cost frontiers, *Journal of Econometrics*, 76, 77-105.
- Koop, G., Osiewalski, J., Steel, M.F.J., 2000. The components of output growth: A stochastic frontier analysis, *Oxford Bulletin of Economics and Statistics*, 62, forthcoming.
- Koop, G., Steel, M.F.J., Osiewalski, J., 1995. Posterior analysis of stochastic frontier models using Gibbs sampling, *Computational Statistics*, 10, 353-373.
- Kumbhakar, S., 1987. The specification of technical and allocative inefficiency of multi-product firms in stochastic production and profit frontiers, *Journal of Quantitative Economics*, 3, 213-223.
- Kumbhakar, S., 1996. Efficiency measurement with multiple outputs and multiple inputs, *Journal of Productivity Analysis*, 7, 225-255.
- Lau, L., 1972. Profit functions of technologies with multiple inputs and outputs, *Review of Economics and Statistics*, 54, 281-289.
- Löthgren, M., 1997. Generalized Stochastic Frontier Production Models, *Economics Letters*, 57, 255-259.
- Meng, X.L., 1994. Posterior predictive p -values, *Annals of Statistics*, 22, 1142-1160.
- Osiewalski, J., 1997. On the use of the consumption efficiency index in empirical demand studies, manuscript.
- Pittman, R., 1983. Multilateral productivity comparisons with undesirable outputs, *Economic Journal*, 93, 883-891.

- Powell, A. and Gruen, F., 1968. The constant elasticity of transformation production frontier and linear supply system, *International Economic Review*, 9, 315-328.
- Shephard, R.W., 1970. *Theory of Cost and Production Functions* (Princeton University Press, Princeton).
- Tierney, L., 1994. Markov chains for exploring posterior distributions, (with discussion) *Annals of Statistics*, 22, 1701-1762.
- 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.
- Zellner, A., 1975. Bayesian analysis of regression error terms, *Journal of the American Statistical Association*, 70, 138-144.

Table 1: Posterior Quantiles using Simulated Data

	True Value	Median	IQR	0.025 quant.	0.975 quant.
β_1	0.500	0.512	0.033	0.466	0.561
β_2	0.500	0.496	0.007	0.487	0.506
σ	0.500	0.497	0.008	0.485	0.508
α_1	0.333	0.334	0.006	0.325	0.344
α_2	0.333	0.335	0.006	0.325	0.343
α_3	0.333	0.331	0.006	0.322	0.340
q	2.000	1.978	0.118	1.828	2.136
ϕ	5.000	5.994	0.723	4.980	7.142
s_1	1.000	0.985	0.091	0.871	1.117
s_2	1.000	0.996	0.089	0.884	1.126
s_3	1.000	0.983	0.092	0.875	1.117

Table 2: Posterior Quantiles with Banking Data

	Median	IQR	0.025 quant.	0.975 quant.
β_1	0.441	0.108	0.287	0.594
β_2	0.179	0.016	0.154	0.203
β_3	0.029	0.007	0.019	0.039
β_4	0.061	0.004	0.055	0.067
β_5	0.298	0.010	0.284	0.312
β_6	0.476	0.010	0.461	0.490
σ	0.142	0.002	0.139	0.144
ϕ_1	1.563	0.113	1.402	1.737
ϕ_2	1.155	0.125	0.986	1.352
ϕ_3	1.336	0.140	1.148	1.551
α_1	0.257	0.006	0.249	0.264
α_2	0.279	0.006	0.271	0.288
α_3	0.464	0.007	0.454	0.475
q	1.025	0.033	1.001	1.067
s_1	4.187	0.243	3.837	4.407
s_2	2.319	0.122	2.140	2.432
s_3	3.738	0.193	3.431	3.932
RTS	1.043	0.011	1.027	1.059
Elast. of transf.	-53.63	97.51	-1466	-14.96
Pred. eff., small	0.644	0.419	0.095	0.983
Pred. eff., large	0.750	0.320	0.217	0.989
Eff. of min. bank	0.265	0.017	0.241	0.291
Eff. of med. bank	0.507	0.032	0.461	0.556
Eff. of max. bank	0.766	0.069	0.669	0.873
SDF, small	0.838	0.119	0.683	1.037
SDF, large	0.365	0.079	0.269	0.494
Fraction from ε , small	0.024	0.003	0.019	0.030
Fraction from ε , large	0.055	0.012	0.041	0.075

Fig. 1. Production equivalence curves for bivariate output. As the dashing of the lines becomes finer, they represent $q = 1$, $q = 2$, $q = 4$ and $q = \infty$, respectively. All curves correspond to $\theta = 1$.

Fig. 1a. $\alpha_1 = 1/2$.

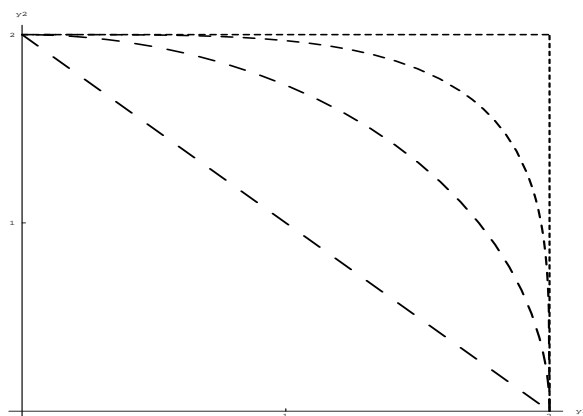


Fig. 1b. $\alpha_1 = 1/3$.

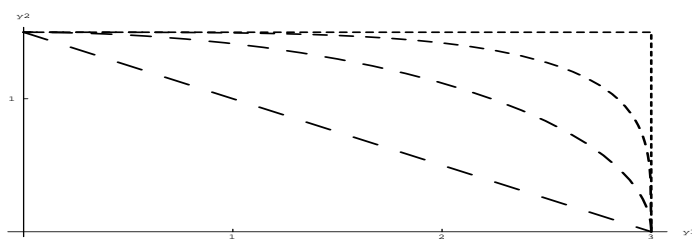


Fig. 2. Efficiency of production A indicated through production equivalence curves for $p = 2$ and $\alpha_1 = 1/2$ as well as through the production frontier with $k = 1$.

Fig. 2a. Production equivalence curves for $q = 1.5$ (flatter solid line) and $q = 5$ (more curved solid line) and given input level x_0 . Efficiency of point A is OB/OD and OC/OD , respectively.

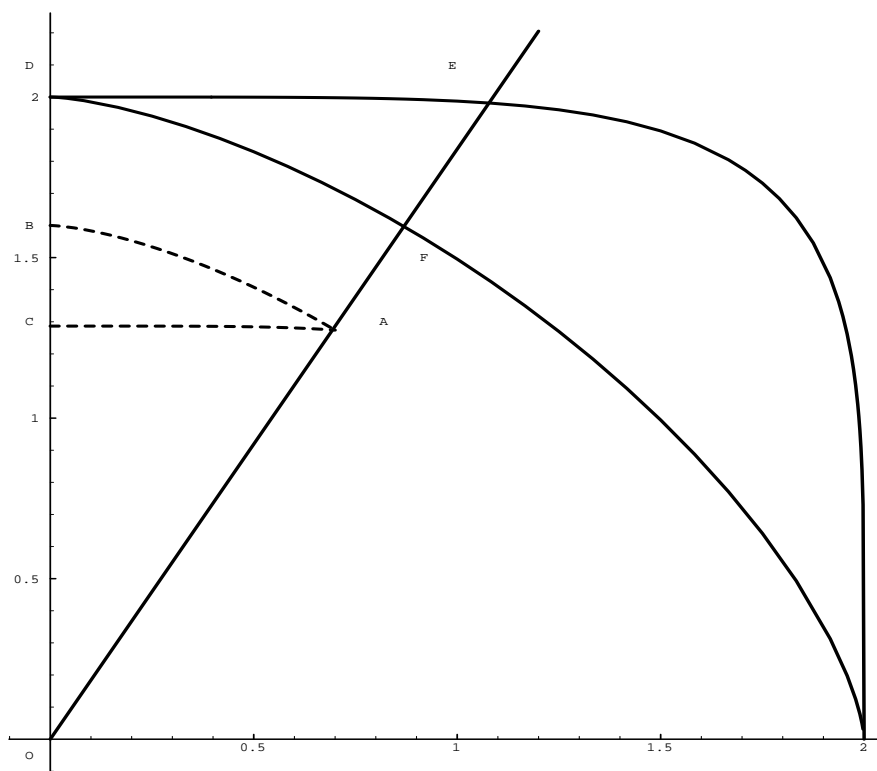


Fig. 2b. Production frontier. At the input level x_0 the relevant efficiencies are, again, indicated.

