Multiple-Output Production With Undesirable Outputs:
An Application to Nitrogen Surplus in Agriculture

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.

Many production processes yield both good outputs and undesirable ones (*e.g.* pollutants). In this paper, we develop a generalization of a stochastic frontier model which is appropriate for such technologies. We discuss efficiency analysis and, in particular, define technical and environmental efficiency in the context of our model. Methods for carrying out Bayesian inference are developed and applied to a panel data set of Dutch dairy farms.

*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

1 Introduction

There are many applications in economics where an individual unit (*e.g.* a farm, firm, country or individual) produces undesirable outputs such as pollution, in addition to desirable ones. As a leading example, note that the environmental problems caused by modern agriculture present an increasingly worrying policy problem. In the application considered in this paper, Dutch dairy farms produce not only good outputs (which we will informally call "goods"), such as milk, but also undesirable outputs (or "bads"), such as excessive nitrogen due to the creation of manure and application of chemical fertilizers. It is thus important to understand the nature of the best-practice technology available to farmers for turning inputs into good and bad outputs. Furthermore, it is important to see how individual farms measure up to this technology. In other words, evaluation of farm efficiency, both in producing as many good outputs and as few undesirable outputs as possible, is crucial. In this paper, we describe how extensions of stochastic frontier models can be used to shed light on these issues.

Stochastic frontier models are commonly used in the empirical study of efficiency and productivity. 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). An introductory survey of the use of Bayesian methods with stochastic frontier models is given in Koop and Steel (2000). All of these papers assume that a single good output is produced and ignore the possible presence of undesirable by-products. In the single output case, a sensible definition of efficiency can be easily found (*i.e.* the ratio of actual output produced to the maximum that could have possibly been produced with the inputs used). The maximum possible output is estimated from a (single-equation) production frontier. The extension to the case of multiple good outputs is more complicated since multivariate distributions must be used and various ways of defining efficiency exist. Fernández, Koop and Steel (2000) provide a solution to these complications. The present paper builds on our previous work to allow for some of the outputs to be bad. This extension involves not only a careful discussion of how to define the production technology for turning inputs into outputs, but also how to measure efficiency relative to this technology. We distinguish between technical and environmental efficiency. The former is the standard efficiency concept

which compares actual to maximum possible output, extended to a multi-output setting in the way suggested by Fernández *et al.* (2000). Environmental efficiency is a new concept we develop in an attempt to address the question "How much pollution reduction could be achieved, without sacrificing good outputs, by adopting best-practice technology?".

To fix some basic ideas, y, b and x will, throughout the paper, denote vectors of good outputs, bad outputs and inputs, respectively. The best-practice technology for turning inputs into outputs is given by a relationship between the inputs and best-practice vectors of good and bad outputs:

$$f(y_{bp}, b_{bp}, x) = 0. \tag{1.1}$$

In the simplest case of a single good output and no bads, the relationship above is typically assumed to allow for expressing the scalar y_{bp} as a function of x:

$$y_{bp} = h(x),$$

where $h(x)$ is known as the production frontier and y_{bp} corresponds to the maximum output level that can be obtained with input vector x. The technical efficiency of a firm producing y with these inputs is then defined as:

$$\tau \equiv \frac{y}{y_{bp}} = \frac{y}{h(x)},$$

where $0 \leq \tau \leq 1$. Statistical estimation of technical efficiency can be done by adding measurement error to the model and making appropriate distributional and functional form assumptions (see, *e.g.*, Koop and Steel, 2000).

In the multiple good outputs case, still without bads, Fernández *et al.* (2000) also assume that the relationship in (1.1) is separable, in the sense that there exist non-negative functions $\theta(\cdot)$ and $h(\cdot)$ such that:

$$\theta(y_{bp}) = h(x),$$

where y_{bp} is now a vector. $\theta(y) = constant$ maps out the output combinations that are technologically equivalent, and is thus defined as the production equivalence surface. By analogy with the single output case, $h(x)$ defines the maximum output [as measured by $\theta(y)$] that can be produced with inputs x and is referred to as the production frontier. For a firm producing y, technical efficiency is defined as:

$$\tau \equiv \frac{\theta(y)}{\theta(y_{bp})} = \frac{\theta(y)}{h(x)}.$$

2

Fernández *et al.* (2000) provide a detailed justification for and illustration of this efficiency definition. In essence, it corresponds to the usual definition of efficiency in the single output case using the "aggregate output" $\theta(y)$ in place of the single output.

In the general multiple output case with both goods and bads, the present paper argues for a relationship in (1.1) that can be written as:

$$\theta(y_{bp}) = h_1(x) \text{ and } \kappa(b_{bp}) = h_2(y_{bp}),$$

for non-negative functions $\theta(\cdot)$, $h_1(\cdot)$, $\kappa(\cdot)$ and $h_2(\cdot)$. In other words, the general relationship can be broken down into two equations involving the "aggregate goods" $\theta(y)$, the "goods' production frontier" $h_1(x)$, the "aggregate bads" $\kappa(b)$, and the "bads' production frontier" $h_2(y)$. The assumption that the amount of good outputs produced depends only on the inputs, while production of bad outputs depends on the amount of good outputs produced is likely to be reasonable in many cases. If not, modifications of the present model can be implemented with only slight alterations to the framework. For a firm producing (y, b), we can define technical efficiency in the usual way:

$$\tau_1 \equiv \frac{\theta(y)}{\theta(y_{bp})} = \frac{\theta(y)}{h_1(x)},$$

and environmental efficiency as:

$$\tau_2 \equiv \frac{h_2(y)}{\kappa(b)},$$

i.e. the minimum possible bad aggregate output divided by the actual one. Both τ_1 and τ_2 lie in the interval [0,1]. There is some earlier work in this area, which treats pollutants as though they were inputs in the production process (see, *e.g.*, Koop, 1998 or Reinhard, Lovell and Thijssen, 1999). We argue that it is more sensible to treat them as outputs as we do in our model, thus viewing them as by-products of the production process. Nevertheless, below we discuss the relationship between our model and the alternative of simply treating pollutants as an input. In essence, we find that our joint distribution for goods and bads implies a conditional distribution for good outputs given bads which is quite similar to the specification of, *e.g.*, Koop (1998). Thus, in a sense, the previous work in this area focusses only on certain aspects of our model (*i.e.* a certain conditional distribution), and our approach encompasses others in the literature. A key issue in this respect, which

will be discussed in the paper, is how economic regularity conditions are imposed on the production process.

The introduction has outlined the basic theoretical model we will use. The following sections of the paper develop a statistical model, which is then used with both artificial data and the empirical application involving Dutch dairy farms. In particular, in the next section, we formally describe the sampling model used in this paper. In the third section we look at various implications of our sampling model in terms of the underlying economic theory. The fourth and fifth sections describe the prior and the algorithm used for posterior inference. The sixth section discusses the performance of our methods using artificial data, while the seventh presents our empirical application involving Dutch dairy farms. The final section concludes. Details of the Markov chain Monte Carlo (MCMC) algorithm used for computation are given in the appendix.

2   The Sampling Model

We begin by defining notation. The cross-sectional unit of analysis will generally be referred to as a firm, which could be a farm, individual or country, etc. We have data from a panel of $i = 1, ..., N$ firms, where the ith firm has been observed for $t = 1, ..., T_i$ time periods. For ease of presentation, the average period of observation is denoted by $T = \sum_{i=1}^{N} T_i / N$, so that the total number of observations is NT. The ith firm in the tth period produces p good outputs and m bad outputs, respectively denoted as $y_{(i,t)} = (y_{(i,t,1)}, \ldots, y_{(i,t,p)})' \in \Re_+^p$ and $b_{(i,t)} = (b_{(i,t,1)}, \ldots, b_{(i,t,m)})' \in \Re_+^m$, using k inputs $x_{(i,t)} = (x_{(i,t,1)}, ..., x_{(i,t,k)})'$.

As already indicated in the Introduction, we require two major extensions with respect to the usual single-output stochastic frontier model. One is that the model has to account for more than one output, and here we follow the same strategy as in Fernández *et al.* (2000) where we define production equivalence surfaces through a parametric aggregator function. The second difficulty is that we need to distinguish between good and bad outputs, and this is what constitutes the main value added of the present paper. Of course, there are many ways of dealing with this issue and we could consider just treating the two types of output differently in the same aggregator, thus effectively reducing the problem to a unidimensional frontier. This would follow the line of reasoning suggested by the

deterministic Data Envelopment Analysis literature (such as in Pittman, 1983, and Färe, Grosskopf, Lovell and Pasurka, 1989) and more recently by Koop (1998) and Reinhard *et al.* (1999), who use a statistical approach but include bads as inputs in a stochastic frontier for a single good. However, this does not allow for a natural separation of the efficiency of technical aspects of production on the one hand and environmental aspects on the other. Therefore, we propose separate modelling of goods and bads, with links provided through the frontier for the bads and through correlations of both measurement and inefficiency error terms. In addition, prior links will be introduced in Section 4.

First, let us model the production technology of the good outputs $y_{(i,t)}$. As proposed in Fernández *et al.* (2000), we extend the single output case to multiple outputs through defining the goods aggregator:

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

with $\alpha_j \in (0,1)$ for all $j = 1, \ldots, p$, $\sum_{j=1}^{p} \alpha_j = 1$ and with $q > 1$ to ensure a negative elasticity of transformation between any two outputs. The function in (2.1) is closely related to the "constant elasticity of transformation" specification of Powell and Gruen (1968), and implies an elasticity of transformation equal to $1/(1-q)$. The role of the p unknown parameters in α and q is further discussed in Fernández *et al.* (2000).

Interpreting $\theta_{(i,t)}$ as aggregate good output, we define a production equivalence surface as the $(p-1)$-dimensional surface with a constant value for $\theta_{(i,t)}$. Now that we have reduced the problem to a unidimensional aggregate, we define the NT-dimensional vector

$$\log \theta = (\log \theta_{(1,1)}, \log \theta_{(1,2)}, \ldots, \log \theta_{(1,T_1)}, \ldots, \log \theta_{(N,T_N)})', \tag{2.2}$$

and we model $\log \theta$ through the usual stochastic frontier

$$\log \theta = V\beta - Dz + \varepsilon_g. \tag{2.3}$$

In (2.3), the matrix $V = (v(x_{(1,1)}), \ldots, v(x_{(N,T_N)}))'$ groups a number of exogenous regressors, where each $v(x_{(i,t)})$ is a function of the inputs $x_{(i,t)}$. The particular choice of $v(\cdot)$ defines the specification of the production frontier (*e.g.* Cobb-Douglas, translog, etc.). Typically, we will want to impose regularity conditions on the frontier coefficients β, which assign an

economic meaning to the surface $V\beta$ in (2.3) as a frontier (*e.g.* monotonicity conditions ensuring that production does not decrease as inputs increase). In the empirical section of this paper we use a Cobb-Douglas frontier (*i.e.* it contains an intercept and is linear in the logs of the inputs) and, accordingly, all the elements of β except for the intercept are restricted to be non-negative.

Another important element of (2.3) is the vector of *technical* inefficiencies $Dz \in \Re_+^{NT}$, where the matrix D is fixed. Choosing D gives some structure to the inefficiencies, as explained in Fernández, Osiewalski and Steel (1997). In this paper we shall follow the usual practice in assuming that inefficiencies for each firm are constant over time (*i.e.* a standard individual effects setup). For a balanced panel (where $T_i = T$ for all i), this corresponds to $D = I_N \otimes \iota_T$, where $\otimes$ denotes the Kronecker product and ι_T a T-dimensional vector of ones. For an unbalanced panel (where T_i varies with i, as is the case in our Dutch dairy farm application), we take the obvious generalization of this. In both cases, $z \in \Re_+^N$ becomes a vector of firm-specific inefficiencies. Since the dependent variable in (2.3) has been logged, it follows that technical efficiency for the ith firm is defined as $\tau_{1i} = \exp(-z_i)$. In the present paper we will assume that D has this individual effects form, but extensions to more general structures can straightforwardly be implemented.

Now we turn to the analysis of the bads, $b_{(i,t)}$, for which we specify a very similar model. First we aggregate the m components of $b_{(i,t)}$ into

$$\kappa_{(i,t)} = \left(\sum_{j=1}^{m} \gamma_j^r \, b_{(i,t,j)}^r \right)^{1/r}, \qquad (2.4)$$

with $\gamma_j \in (0,1)$ for all $j = 1, \ldots, m$ and such that $\sum_{j=1}^{m} \gamma_j = 1$ and now with $0 < r < 1$. In the case of bads a positive elasticity of transformation is more appropriate on the basis of economic theory considerations. We define $\log \kappa$ analogously to $\log \theta$ [see (2.2)] and model this through the following stochastic frontier:

$$\log \kappa = U\delta + Mv + \varepsilon_b, \qquad (2.5)$$

where $U = (u(y_{(1,1)}), \ldots, u(y_{(N,T_N)}))'$, *i.e.* the matrix U is a function of the good outputs. This reflects the idea that the frontier for the bads should be measured relative to the amount of goods produced. Note that we define U as a function of the p-variate $y_{(i,t)}$ rather than of the aggregated scalar counterpart $\theta_{(i,t)}$ alone, since the aggregation in (2.1)

relates only to the production of the good outputs, and it may well be that the influence of the various components of $y_{(i,t)}$ on the production of unwanted outputs is very different from how they appear in (2.1). It is quite likely that different technologically equivalent $y_{(i,t)}$ vectors, *i.e.* situated on the same production equivalence surface corresponding to a particular value of $\theta_{(i,t)}$, can have very different consequences for the minimal amount of aggregated undesirable outputs we can achieve. As with β in the goods' frontier, we shall impose regularity conditions on δ, so that a larger amount of goods cannot be commensurate with a smaller amount of bads. In the empirical section of this paper we use a Cobb-Douglas specification and, accordingly, all the elements of δ except for the intercept are restricted to be non-negative.

$U\delta$ will define the smallest feasible (frontier) production of the aggregate undesirable outputs for a given amount of desirable outputs. If there is any systematic (positive) deviation, this is labelled *environmental* inefficiency, which is grouped in the vector $Mv \in \Re_+^{NT}$. Generally, we can impose different structures through choosing the fixed matrix M. In this paper, we assume that firms have a constant environmental inefficiency over time (*i.e.* individual effects), so that $M = D$ is as described above, and the vector $v \in \Re_+^N$ groups the environmental inefficiencies. The environmental efficiency of firm i is, thus, $\tau_{2i} = \exp(-v_i)$. Again, extensions to more general M, not necessarily equal to D, are straightforward.

We still need to introduce stochastics into the sampling model. We shall consider the distribution on the technical and environmental inefficiencies, (z, v), as part of the prior and discuss it in the next section. Although non-Bayesian analyses that treat the inefficiencies as part of the sampling model (by assigning a probability distribution to them) are not equivalent to those that merely treat them as fixed unknown parameters, this distinction is formally irrelevant in a Bayesian context, where any unknown quantity is made stochastic. Fernández *et al.* (1997) argue that interpreting distributional assumptions on (z, v) as a hierarchical prior is the most revealing way of thinking about the model.

The terms ε_g in (2.3) and ε_b in (2.5) capture the usual measurement error and model imperfections, and as such will be assigned a symmetric distribution. We allow for the two error terms to be correlated for the same firm and time period, and assume a bivariate Normal distribution. That is, if we let $f_N^R(\varepsilon|a, A)$ denote the R-variate Normal p.d.f. with

mean a and covariance matrix A, evaluated at ε, we can write:

$$p(\varepsilon_g, \varepsilon_b | \Sigma) = f_N^{2NT} \left(\begin{array}{c} \varepsilon_g \\ \varepsilon_b \end{array} \middle| 0, \Sigma \otimes I_{NT} \right) \tag{2.6}$$

where Σ is a 2×2 P.D.S. matrix.

We note that (2.6) is not sufficient to define a probability distribution on all $p + m$ components of $(y_{(i,t)}, b_{(i,t)})$, as it merely specifies a joint distribution for the aggregated goods and bads, $(\theta_{(i,t)}, \kappa_{(i,t)})$. Since these aggregators, respectively defined in (2.1) and (2.4), depend on unknown parameters, they can not constitute sufficient statistics and we can not simply stop here. That is, a valid sampling model must be based on a $(p + m)$-variate distribution for $(y_{(i,t)}, b_{(i,t)})$, and basing the analysis on (2.6) alone is only enough when $p = m = 1$, *i.e.* when we have a single good and a single bad. Thus, we will complete (2.6) with stochastic assumptions in the $p + m - 2$ remaining dimensions, along the lines suggested by Fernández *et al.* (2000) for the multiple goods case with no bads ($m = 0$). First, for the goods, when $p > 1$ we define the weighted output shares:

$$\eta_{(i,t,j)} = \frac{\alpha_j^q y_{(i,t,j)}^q}{\sum_{l=1}^p \alpha_l^q y_{(i,t,l)}^q}, \; j = 1, \ldots, p, \tag{2.7}$$

group them into $\eta_{(i,t)} = (\eta_{(i,t,1)}, \ldots, \eta_{(i,t,p)})'$, and assume independent sampling from

$$p(\eta_{(i,t)} | s) = f_D^{p-1}(\eta_{(i,t)} | s), \tag{2.8}$$

where $s = (s_1, \ldots, s_p)' \in \Re_+^p$ and $f_D^{p-1}(\cdot | s)$ is the p.d.f. of a Dirichlet distribution with parameter s. Similarly, if $m > 1$ we define a weighted vector of shares for the bads:

$$\zeta_{(i,t,j)} = \frac{\gamma_j^r b_{(i,t,j)}^r}{\sum_{l=1}^m \gamma_l^r b_{(i,t,l)}^r}, \; j = 1, \ldots, m, \tag{2.9}$$

stack them to form $\zeta_{(i,t)} = (\zeta_{(i,t,1)}, \ldots, \zeta_{(i,t,m)})'$, and assume independent sampling from

$$p(\zeta_{(i,t)} | h) = f_D^{m-1}(\zeta_{(i,t)} | h), \tag{2.10}$$

where $h = (h_1, \ldots, h_m)' \in \Re_+^m$. Intuition for equations (2.7)-(2.10) can be obtained by noting that the earlier part of this section only specified a model for aggregate good and bad outputs. For instance, (2.3) models only the production equivalence surface, but (2.7) and (2.8) model output components within a production equivalence surface.

8

Now, $(2.1) - (2.10)$ lead to a sampling distribution for Y, B which are matrices of dimensions $NT \times p$ and $NT \times m$, respectively, with elements ordered in the same manner as $\log \theta$ in (2.2). Taking into account the Jacobian of the transformation, we obtain the following sampling density:

$$p(Y, B | \beta, z, \delta, v, \Sigma, \alpha, \gamma, q, r, s, h) = f_N^{2NT} \left(\begin{array}{c} \log \theta \\ \log \kappa \end{array} \Bigg| \begin{array}{c} V\beta - Dz \\ U\delta + Mv \end{array}, \Sigma \otimes I_{NT} \right) \quad (2.11)$$

$$\prod_{i,t} \left[f_D^{p-1}(\eta_{(i,t)} | s) \left(\prod_{j=1}^{p} q^{1-\frac{1}{p}} \frac{\eta_{(i,t,j)}}{y_{(i,t,j)}} \right) \right] \prod_{i,t} \left[f_D^{m-1}(\zeta_{(i,t)} | h) \left(\prod_{j=1}^{m} r^{1-\frac{1}{m}} \frac{\zeta_{(i,t,j)}}{b_{(i,t,j)}} \right) \right].$$

3 Implications of the Sampling Model

It is important to relate the statistical model specified in the previous section to the underlying economic theory. This we have partially done in the previous sections, the purpose of the present section is to provide additional insight by looking at some implications of the sampling model. In particular, we will examine the marginal distributions for the goods and the bads, as well as the conditionals, derived from the sampling distribution in (2.11). Examining these distributions is relevant because *e.g.* the marginal distribution of the goods will relate to the common case where only the goods are modelled and the bads are ignored, whereas the conditional distribution of the goods given the bads relates to the models where the bads are treated as additional inputs in the production process.

Since the distribution in (2.11) assumes independent sampling across firms and time periods, we will consider a single observation and suppress the (i, t) subscripts. In addition, our individual effects setting implies that the inefficiencies are merely z and v. All of the expressions in this section should be understood as conditioned on the model parameters [as in (2.11)] but we will not explicitly indicate this.

Much of the intuitive motivation related to frontiers and inefficiencies can be obtained by looking at the special case where there is only one good and one bad. Note that in this case, $\log \theta = \log y$, $\log \kappa = \log b$ and the shares η and ζ do not enter the analysis. Furthermore, economic intuition relates largely to the stochastic frontiers of equations (2.3) and (2.5), which only involve $\log \theta$ and $\log \kappa$ instead of the entire vectors y and b. These considerations suggest focussing on the marginal and conditional distributional properties

9

of $\log\theta$ and $\log\kappa$ derived from the joint distribution of $(\log\theta, \eta, \log\kappa, \zeta)$. Using (2.11) with a Cobb-Douglas structure for the bads' frontier $[i.e.\ U = (1, \log y_1, \ldots, \log y_p)]$ and making a change of variables from (y, b) to $(\log\theta, \eta, \log\kappa, \zeta)$, we obtain:

$$p(\log\theta, \eta, \log\kappa, \zeta) = f_N^2 \begin{pmatrix} \log\theta \\ \log\kappa \end{pmatrix} \Big| \begin{array}{c} V\beta - z \\ \overline{\delta}(V\beta - z) + h(\eta) + v \end{array}, W \Big) f_D^{p-1}(\eta|s) f_D^{m-1}(\zeta|h), \quad (3.1)$$

where $\overline{\delta} \equiv \sum_{j=2}^{p+1} \delta_j$, $h(\eta) \equiv \delta_1 + \sum_{j=1}^{p} \delta_{j+1} \log(\alpha_j^{-1} \eta_j^{1/q})$ and, denoting by σ_{ij} the $(i, j)^{th}$ element of Σ, the elements of W are:

$$w_{11} = \sigma_{11}, \ w_{12} = \overline{\delta}\sigma_{11} + \sigma_{12}, \ w_{22} = [det(\Sigma) + w_{12}^2]/\sigma_{11}.$$

3.1   Marginal distributions of $(\log\theta, \eta)$ and $(\log\kappa, \zeta)$

Note that this is equivalent to examining the marginal distributions of y and b, which respectively correspond to modelling the goods ignoring the bads and modelling the bads ignoring the goods.

From (3.1) it is immediate that, in the marginal distribution, $\log\theta$ and η are independent. The distribution of $\log\theta$ is Normal with,

$$E[\log\theta] = V\beta - z, \ V[\log\theta] = w_{11}, \quad (3.2)$$

whereas the distribution of η is given by the p.d.f. $f_D^{p-1}(\eta|s)$. It is interesting to note that this corresponds exactly to the specification in Fernández $et\ al.$ (2000), who consider the problem with multiple goods and no bads.

Marginally, $\log\kappa$ and ζ are also independent. The distribution of $\log\kappa$ is a bit complicated, due to the presence of the term $h(\eta)$ in its mean. The conditional distribution of $\log\kappa$ given η is Normal with

$$E[\log\kappa|\eta] = \overline{\delta}(V\beta - z) + h(\eta) + v, \ V[\log\kappa|\eta] = w_{22}, \quad (3.3)$$

so that the marginal distribution of $\log\kappa$ is a (continuous) location mixture of Normals, each of them defined as in (3.3), with mixing distribution $f_D^{p-1}(\eta|s)$. The mean and variance of the resulting marginal distribution may not be very revealing since the latter need not be unimodal. Nevertheless, it can be seen that

$$E[\log\kappa] = \overline{\delta}E[\log\theta] + E[h(\eta)] + v = \delta_1 + \sum_{j=1}^{p} \delta_{j+1} E[\log y_j] + v,$$

provided that the mean exists. The distribution of ζ is given through the p.d.f. $f_D^{m-1}(\zeta|h)$.

What economic insight can be obtained from the marginal distributions of $\log\theta$ and $\log\kappa$? Firstly, remember that in the specification in the previous section, we imposed economic regularity conditions to ensure that the goods' production frontier was non-decreasing in inputs and the bads' production frontier was non-decreasing in good outputs. We now see that these conditions are identical to those required to impose regularity on the marginal distributions.

Secondly, note that $E[\log\kappa|\eta]$ implies that different output shares (*i.e.* different η's) can lead to different amounts of bads. This seems quite reasonable as producing one of the good outputs may well be more polluting than producing another one. In other words, whereas the position of a firm on the goods' production equivalence surface should, by definition, not affect its technical efficiency, it can well affect its frontier for the bads. If we had just allowed the amount of bads to depend on aggregate good output [*i.e.* $U = (1, \log\theta)$], we would not have obtained this degree of flexibility. As a hypothetical example, consider cattle farms which produce meat and leather. Since production of leather is very polluting, farms on the same good production equivalence surface which choose to produce more leather relative to meat will cause more pollution.

3.2 Conditional distributions of $\log\theta$ and $\log\kappa$

Here we focus on the conditional distributions of $\log\theta$ given (η, b) and of $\log\kappa$ given (ζ, y). From (3.1), it is immediate that both these distributions are Normal with means given by:

$$
\begin{aligned}
E[\log\theta|\eta, b] &= V\beta\left(1 - \overline{\delta}\frac{w_{12}}{w_{22}}\right) + \frac{w_{12}}{w_{22}}[\log\kappa - h(\eta)] - \left[\left(1 - \overline{\delta}\frac{w_{12}}{w_{22}}\right)z + \frac{w_{12}}{w_{22}}v\right], (3.4) \\
E[\log\kappa|\zeta, y] &= \left(\overline{\delta} - \frac{w_{12}}{w_{11}}\right)E[\log\theta] + \frac{w_{12}}{w_{11}}\log\theta + h(\eta) + v.
\end{aligned}
$$

These conditional properties illuminate some important features of our model:

Firstly, note that the conditional mean of $\log\theta$ given (η, b) depends on $\log\kappa$. Such an equation forms the starting point of analyses where bads are treated as inputs in the production process (*e.g.* Koop, 1998 or Reinhard *et al.*, 1999). A consideration of the one good-one bad case is especially revealing in this regard. In this case, $h(\eta)$ would not be present and $\log\theta$ and $\log\kappa$ would simply be $\log y$ and $\log b$. If V contained the logs of inputs, then we would obtain a Cobb-Douglas production frontier where b was simply

treated as an input. In other words, the treatment of pollutants as inputs arises naturally in our framework if we focus only on the conditional distribution of y given b. In the general setting with multiple goods and bads, using the variable $\log \kappa$ as an input for the goods' frontier requires correcting it to $\log \kappa - h(\eta)$, to account for the fact that the aggregate amount of bads depends on the position of the firm on the goods' production equivalence surface.

Similarly, although perhaps less interesting, the second equation in (3.4) shows the implications of our model when only the bads are modelled stochastically and y is treated as a fixed input.

In order to interpret the means in (3.4) as frontiers minus (or plus) an inefficieny term, we need to impose economic regularity conditions, so that aggregate goods increase with inputs and $\log \kappa - h(\eta)$, and aggregate bads increase with $E[\log \theta]$ and goods. In the common case where V corresponds to a Cobb-Douglas specification, these conditions simply amount to requiring non-negative coefficients for all the variables, which is equivalent to

$$- \min\left(\overline{\delta}\sigma_{11}, \frac{\sigma_{22}}{\overline{\delta}}\right) \leq \sigma_{12} \leq 0. \tag{3.5}$$

In our empirical work, this regularity condition is imposed in addition to the ones discussed in Section 2.

It is interesting to note that, if we impose the regularity condition in (3.5), the single inefficiency term which appears when the bads are treated as an additional input [*i.e.* as in the first equation in (3.4)] is a weighted average of the technical inefficiency z obtained from the equation $\log \theta = V\beta - z$ (ignoring the symmetric error) and the "inefficiency" $v/\overline{\delta}$ obtained naively from the equation $\log \kappa = U\delta + v$ (from which $\log \theta = [\log \kappa - h(\eta)]/\overline{\delta} - v/\overline{\delta}$).

Finally, in this paper we treat the position of firms in the goods' production equivalence surface, *i.e.* the value of η, as stochastic, with distribution given by $f_D^{p-1}(\eta|s)$, and we will assume a weakly informative prior for s. However, if we had data on the prices of the different goods we could take a more informative approach, since it is logical to assume that firms will aim to be at the point on the goods' production equivalence surface where profit is maximised (this point will be given by $\eta_j \propto (p_j/\alpha_j)^{q/(q-1)}$, $j = 1, \ldots, p$, if profit is measured by $\sum_j p_j y_j$, where $p_1, \ldots, p_p$ denote the prices of the p different goods). Thus, additional information on prices can be easily incorporated in our analysis. However, we

can see from (3.3) that different values of η can lead to different amounts of pollution and, generally, the profit maximising point will not be optimal from an environmental point of view. It can be seen from the expression of $h(\eta)$ that a way to achieve pollution reduction is to decrease production of goods y_j with a large value of δ_{j+1} in favour of goods with a smaller value of the corresponding δ coefficient. These considerations mean that inference on δ is of great potential relevance for policy development (*e.g.* for making decisions on taxing or subsidizing outputs in order to ensure that the profit maximising point is adequate from an environmental point of view). To provide a hypothetical example of this, consider a farm which produces meat and leather, and assume leather production yields more pollution than meat production. In this case, by concentrating mainly on meat the farm will reduce the production of pollution. This, however, may not maximize profit. A government interested in minimizing the pollution from leather tanning might consider taxing leather and/or subsidizing meat in order to encourage firms to shift towards less-polluting meat production.

In summary, an examination of certain conditional and marginal distributions provides additional motivation for and insight into our model. Furthermore, this reveals an additional economic regularity condition [*i.e.* (3.5)] that will be imposed on the production technology.

4   The Prior

Given the relative complexity of the model, we shall not investigate the existence of the posterior distribution under classes of improper priors. See Fernández *et al.* (1997) for such an analysis in the simpler standard stochastic frontier model (also covering a range of other models with unobserved heterogeneity in the location). Thus, we shall assume a proper prior from the start. We shall conduct a careful prior elicitation in some crucial dimensions while being relatively flat in other, less crucial, dimensions. In Section 6, we will ascertain the influence of the prior through simulation studies.

In particular, we shall use the following proper prior structure:

$$p(\beta, \delta, \Sigma, z, v, \alpha, \gamma, q, r, s, h) = p(\beta, \delta, \Sigma)p(z, v)p(\alpha)p(\gamma)p(q)p(r)p(s)p(h), \qquad (4.1)$$

where independence is assumed between most parameters, but not necessarily between the

parameters of both frontiers (β and δ) and between the vectors of inefficiency terms (z and v). In addition, restriction (3.5) links Σ and δ. Building upon our earlier work with stochastic frontiers (see *e.g.* Koop, Osiewalski and Steel, 1997) and using the intuition that only T_i observations are available for the inefficiency terms of firm i, it is the prior on (z, v) that is most critical to the analysis. Thus, we shall spend considerable effort on the elicitation of a reasonable prior for (z, v).

Prior for (z, v):

For each $i = 1, \ldots, N$, we take a truncated Normal inefficiency distribution:

$$p(z_i, v_i | \mu_i, \Omega) = f_N^2((z_i, v_i)' | \mu_i, \Omega) f^{-1}(\mu_i, \Omega) I_{\Re_+^2}(z_i, v_i) \tag{4.2}$$

where $f(\mu_i, \Omega)$ is the integrating constant of the truncated Normal, $I_{\Re_+^2}(\cdot)$ is the indicator function for $\Re_+^2$ and we assume independence between firms. The reason for taking a truncated Normal is mainly the ease with which correlation can be handled. Applying a Farlie-Gumbel-Morgenstern transformation to the Exponential distributions often used in our previous, single-equation work, only allows for quite moderate correlations, namely in the range from [-1/4,1/4]. Adopting more specific bivariate Exponential distributions leads to serious restrictions, as *e.g.* in Gumbel's (1960) analytic family or the empiric method by Lawrance and Lewis (1983), where the sign of the correlation has to be pre-specified.

It is often desirable to allow for firms' efficiencies to depend on their characteristics (*e.g.* size, ownership, etc.). In order to incorporate this, we allow for the inefficiencies to depend on d variables in $g_i = (1, g_{i2}, \ldots, g_{id})'$ through the mean of the underlying Normal in (4.2):

$$\mu_i = (\phi, \psi)' g_i, \tag{4.3}$$

and assume the following prior on the parameters ϕ and ψ:

$$p(\phi, \psi | \Omega) = f_N^{2d} \left(\begin{array}{c} \phi \\ \psi \end{array} \middle| 0, \Omega \otimes A(d) \right), \tag{4.4}$$

with

$$A(d) = c \left(\begin{array}{cc} a - \frac{d-1}{3} & 0' \\ 0 & I_{d-1} \end{array} \right) \tag{4.5}$$

where we need to choose $a > (d-1)/3$ and $c > 0$. We could easily generalize this to let both elements of μ_i also depend on different explanatory variables. In that case, it might

be preferable to use priors that assume independence between elements of ϕ and ψ that correspond to different variables. The prior inefficiency distribution is now completed by an Inverted Wishart prior on Ω:

$$p(\Omega) = f^2_{IW}\left(\Omega | \Omega_0, \nu_0\right), \tag{4.6}$$

which has expectation equal to the matrix $\Omega_0/(\nu_0 - 3)$ if the number of degrees of freedom $\nu_0 > 3$.

Some intuition and explanation of this prior is in order at this stage in order to help in eliciting the prior hyperparameters a, c, Ω_0 and ν_0. Firstly, the specification in (4.5) depends on a standardization of the variable g_i, such that $0 \le g_{ij} \le 1$ for $j = 2, \ldots, d$. In particular, from (4.3)-(4.5) we can obtain the prior for μ_i:

$$p(\mu_i | \Omega) = f^2_N\left(\mu_i | 0, c\left[a - \frac{d-1}{3} + \sum_{j=2}^{d} g^2_{ij}\right]\Omega\right). \tag{4.7}$$

Note that, because we have imposed that $g_{ij} \in [0, 1]$, $\sum_{j=2}^{d} g^2_{ij}$ is in the range $[0, d-1]$. If g_{ij} were Uniformly distributed, the expected value of the latter sum would be equal to $(d-1)/3$. This suggests the following approximation:

$$p(\mu_i | \Omega) \approx f^2_N(\mu_i | 0, ac\Omega), \tag{4.8}$$

which should be reasonably accurate if a is large w.r.t. $|\sum_{j=2}^{d} g^2_{ij} - (d-1)/3|$ and is exact when $d = 1$. Given that we want our prior elicitation to be applicable for any choice of g_i in the appropriate range, we want to choose a large enough. For (empirically relevant) situations with small d (say, $d \le 3$) we find that choosing $a = 3$ gives reasonable results. This implies that we have a larger variance for the first element of ϕ and ψ than for the others, which reflects a moderate prior belief that firm characteristics influence the efficiency distributions less than the common intercept. In addition, the prior on the first element of ϕ and ψ is then not much affected by the choice of (small) d. To accommodate the case where $d > 1$, we also need to choose a value for c. This can be done on the basis of comparing prior variances of the last $d - 1$ components of ϕ and ψ with those of the underlying Normal on inefficiencies in (4.2). We feel that $c = 2$ is a reasonable value as it allows for sufficient prior uncertainty on ϕ and ψ, given a prior efficiency distribution

(obtained by simulation) that is roughly in line with our prior beliefs. Smaller values of c led to a prior on the coefficients (ϕ,ψ) that we found too tight. In addition, we choose a diagonal Ω_0 and we set $\nu_0 = 6$ which is the smallest (and hence least informative) integer value for which the prior covariance of Ω exists. Thus, we are only left with the choice of $\omega_{ii}^0 (i = 1, 2)$.

It is always desirable to elicit priors in terms of easily interpretable quantities. In the present case, technical efficiency $\tau_{1i} = \exp(-z_i)$ and environmental efficiency $\tau_{2i} = \exp(-v_i)$ have a clear interpretation. Subsection 6.1 reports the results of a calibration exercise where we investigate the implications of values for ω_{ii}^0 on prior efficiencies.

Prior for other parameters:

All of the other parameters of the model (*i.e.* $\beta, \delta, \Sigma, \alpha, \gamma, q, r, s, h$) are assigned relatively flat, yet proper priors. In selecting these priors, we base ourselves on the experience gained with the multiple good output analysis, with no bads, of Fernández *et al.* (2000). In particular, we adopt the following structure for the rest of the prior:

$$p(\beta, \delta, \Sigma) \propto f_N \left(\begin{array}{c} \beta \\ \delta \end{array} \middle| b_0, H_0^{-1} \right) f_{IW}^2(\Sigma|\Sigma_0, \lambda_0) I_{\mathcal{RR}}(\beta, \delta, \Sigma), \tag{4.9}$$

$$p(\alpha) = f_D^{p-1}(\alpha|a_0), \tag{4.10}$$

$$p(\gamma) = f_D^{m-1}(\gamma|g_0), \tag{4.11}$$

$$p(q) \propto f_G(q|1, q_0) I_{(1,\infty)}(q), \tag{4.12}$$

where $f_G(.|a, b)$ denotes a Gamma density function with shape parameter a and mean a/b (if $a = 1$, we have an Exponential),

$$p(r) \propto f_G(r|1, r_0) I_{(0,1)}(r), \tag{4.13}$$

$$p(s) = \prod_{j=1}^{p} f_G(s_j|1, k_j), \tag{4.14}$$

and, finally,

$$p(h) = \prod_{j=1}^{m} f_G(h_j|1, n_j). \tag{4.15}$$

We make the relatively noninformative choices of $b_0 = 0$, H_0 equal to 0.0001 times the identity matrix of the appropriate dimension, $\Sigma_0 = I_2$ and $\lambda_0 = 2$. $I_{\mathcal{RR}}(\beta, \delta, \Sigma)$ denotes the indicator function for the region where economic regularity conditions are met. In

16

the previous sections we have shown the form of this region when V and U are chosen to imply Cobb-Douglas frontiers [note that, in this case, (3.5) means that the regularity conditions also involve Σ]. For the Dirichlet priors in (4.10) and (4.11) we set a_0 and g_0 to appropriately dimensioned vectors of ones. For the truncated exponential priors in (4.12) and (4.13) we set $r_0 = q_0 = 10^{-6}$. In (4.14) and (4.15) we set $k_j = n_j = 1$, thus centering $\eta_{(i,t)}$ and $\zeta_{(i,t)}$ at the equal output share values.

5   The Posterior MCMC Algorithm

The resulting posterior from combining the sampling model in (2.11) and the prior in Section 4 does not lend itself to immediate analytical analysis. Instead, we shall use a Markov chain Monte Carlo (MCMC) algorithm on the space of the original parameters in (2.11) and the inefficiencies (z, v), augmented with ϕ, ψ, Ω, [*i.e.* the parameters of the hierarchical prior of (z, v) in (4.2)-(4.3)]. The Markov chain will be constructed from Gibbs steps for $(z, v), (\beta, \delta), \Sigma$, where we can draw immediately from the conditionals, and Normal random walk Metropolis samplers for $\Omega, (\phi, \psi), \alpha, \gamma, q, r, s, h$, since the conditionals for the latter do not have a well-known form. We fine-tune results from preliminary runs in order to select the variance for the increments in the random walk Metropolis samplers. The relevant conditional posterior distributions are described in more detail in the appendix.

6   Simulation Experiments

In this section, we carry out simulation experiments in an attempt both to understand the prior better and examine the performance of our posterior MCMC algorithm in controlled circumstances.

6.1   Prior on Technical and Environmental Efficiencies

As discussed above, the prior on the efficiencies is most crucial in our analysis. Remember that our recommended prior elicitation procedure uses parameters such as ϕ, ψ and Ω which are difficult to interpret in an intuitive fashion. Accordingly, it is useful to examine the implications of our prior given the prior hyperparameters selected in Section 4. We concentrate here on the base case where $a = 3$ and $c = 2$ in (4.5) and will examine the choice

of ω_{ii}^0 for the specifications with $d = 1$ (no explanatory variables) and $d = 2$. Note that the prior for the inefficiencies (and, hence, the efficiencies) depends on g_i [see (4.3)]. Since our standardization requires $0 \leq g_{ij} \leq 1$ for $j = 2, \ldots, d$, we will take $g_{i2} = 0.5$ when $d = 2$. The priors for other choices of g_{i2} are very similar to those reported here. Table 1 lists some relevant prior marginal properties of the efficiency τ_i for various choices of $\omega_{ii}^0 (i = 1, 2)$. Prior 2.5 and 97.5 percentiles are denoted by "2.5%" and "97.5%". In addition, the column "Corr." indicates prior correlation between technical and environmental efficiencies. There is always a small positive correlation and marginal prior p.d.f.'s for the efficiencies all have a mode at one (full efficiency). For larger values of ω_{ii}^0 a secondary mode at zero appears. Note that the priors for the efficiencies are very similar for both values of d. On the basis of these simulation results, we will select the value $\omega_{ii}^0 = 0.65$ for both technical ($i = 1$) and environmental ($i = 2$) efficiencies.

The properties of the prior for Ω, ϕ and ψ, defined through (4.4)-(4.6), are analytically known. For example, the marginal prior distributions of elements of ϕ and ψ are Student-t with $\nu_0 - 1$ degrees of freedom, location 0 and if $\nu_0 > 3$ we obtain for the variances

$$V[\phi_j] = \frac{\omega_{11}^0}{\nu_0 - 3} A_{jj}(d), \ \ V[\psi_j] = \frac{\omega_{22}^0}{\nu_0 - 3} A_{jj}(d), \tag{6.1}$$

where $\omega_{ii}^0 (i = 1, 2)$ and $A_{jj}(d)(j = 1, ..., d)$ are the diagonal elements of Ω_0 and $A(d)$, respectively. Nevertheless, it is useful to present some prior properties of ϕ and ψ for the chosen prior in Table 2. Since the prior we adopt treats technical and environmental efficiencies in the same manner, Table 2 presents results only for $\phi = (\phi_1, ..., \phi_d)'$. Results for ψ are identical. It can be seen that the prior for ϕ is fairly non-informative. The prior mean of the diagonal elements of Ω is 0.2167. The half-Normal inefficiency distribution (*i.e.* $\phi = 0$) is an extremely common one in the literature. By keeping the prior centered over $\phi = 0$, we allow some flexibility relative to this well-known benchmark. As remarked in Section 4, the relative magnitudes of a and c imply that we are more flexible on ϕ_1 than on ϕ_2.

The prior p.d.f. of the chosen efficiency distribution is plotted in Figures 1 through 4 below for the case $d = 2$. This prior on the efficiencies combines our desired prior median with a reasonable spread.

6.2 Artificial Data

Working with artificially generated data allows us to empirically investigate three crucial issues:

- Whether 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 degree to which the prior influences sample information.

- Whether the numerical properties of the MCMC algorithm used are satisfactory and convergence is achieved with the number of draws used.

In this subsection, we present results for two different artificial data sets. One of these is generated using parameter values that are in a region that our prior favours, the other is generated with parameter values in a region that receives less prior support.

We generate both artificial data sets with $T = 5$, $N = 250$, 2 good outputs, 2 bad outputs, 1 input and 1 explanatory variable in the mean of the efficiency distribution. Hence, $p = 2$, $m = 2$, $k = 1$ and $d = 2$. Both frontiers are assumed to be Cobb-Douglas (*i.e.* intercept plus log-linear), which implies that β is of dimension $k + 1 = 2$ and δ is of dimension $m + 1 = 3$.

The input is generated in a simple manner. First we generate a draw from a N(0,0.04) distribution for each of the $N = 250$ individuals and take the exponential. We treat the resulting numbers as the initial levels of the 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 explanatory variable in the efficiency distribution, g_{i2}, is drawn from the Uniform[0,1].

To complete the specification of our first data generating process we set:

$\beta = 0.5\iota_2; \delta_1 = 0.2, \delta_2 = \delta_3 = 0.5; \sigma_{11} = \sigma_{22} = 1, \sigma_{12} = -0.25; \Omega = 0.25I_2;$

$\phi = \psi = 0_2; q = 1.5; r = 0.75; s = h = \iota_2; \alpha = \gamma = 0.5\iota_2.$

For the second data generating process, we set $\phi = \psi = \iota_2$ but otherwise leave all the other parameters unchanged. The two data sets are based on different seeds for the random number generators.

Table 3 contains posterior results for the two artificial data sets. The columns labelled "2.5%" and "97.5%" are the 2.5 and 97.5 percentiles of the posterior (*i.e.* the lower and upper bounds of a 95% posterior probability interval). Of course, results based on only two artificial data sets should be treated as suggestive rather than definitive. Nevertheless, the general message of Table 3 is encouraging. The fact that all except one of the 95% posterior probability intervals contain the true values used to generate the artificial data indicates that our Bayesian approach is reliable. An examination of posterior results for different starting values and different length runs indicate our MCMC algorithm works well if the number of burn-in and included replications is sufficiently high. The results presented are based on 55,000 drawings (40,000 for data set 1) with the initial 5,000 discarded.

However, Table 3 does reveal some findings which are worthwhile discussing in more detail. Firstly, many of the 95% posterior probability intervals are quite wide. In particular, comparing prior percentiles for ϕ and ψ with the posterior ones, it is clear the data have been only somewhat informative. We have noted this in our previous work with stochastic frontier models. In essence, stochastic frontier models ask the data to estimate a frontier and then decompose each firm-specific deviation from the frontier into two parts. This is quite difficult to do unless there is a great deal of data. Here we have set $T = 5$ which is quite small (although most applications will likely involve values of this magnitude). Furthermore, we are working with a much more complicated multiple-output model which requires the estimation of production equivalence surfaces in addition to frontiers and decomposition of errors. In view of all these issues, we would argue that the 95% posterior probability intervals are not unreasonably wide. Secondly, an examination of results for the second data set reveals that the prior does not have a strong effect on posterior results, even in cases where the likelihood and prior are located in rather different regions.

In order to further investigate how we learn about efficiencies, Figures 1 through 4 contrast the prior efficiency distribution with an out-of-sample predictive efficiency distribution, obtained by integrating out (4.2) with the posterior distribution on (ϕ, ψ, Ω) and transforming to efficiencies. This can be interpreted as reflecting our posterior knowledge about the efficiency of an unobserved firm, say f, and will be denoted by $\tau_{jf}(j = 1, 2)$ in the sequel. Formally, denote the inefficiency of a typical, unobserved firm by (z_f, v_f). We then have $\tau_{1f} = \exp(-z_f)$ and $\tau_{2f} = \exp(-v_f)$. The true efficiency distibution, obtained

by fixing (ϕ, ψ, Ω) at the values used to generate the data, is also plotted on these graphs. In all of these figures, we have set $g_{i2} = 0.5$.

It can be seen that the data does allow us to learn about the efficiency distribution of an unobserved firm. That is, the densities labelled "Prior" and "Predictive" in the figures are different from one another and the latter tends to be reasonably close to the density labelled "True".

The artificial data results suggest that, overall, our model and methods are reasonably good at extracting the information from data sets of realistic size. The somewhat wide posterior density intervals are an accurate reflection of the difficulty inherent in trying to get precise estimates with small data sets in stochastic frontier models. It is worth stressing that Bayesian stochastic frontier methods allow for the calculation of finite sample measures of uncertainty (*e.g.* 95% posterior density intervals) for all parameters and the efficiencies themselves. This is something that is very difficult to do reliably using non-Bayesian statistical methods (see, *e.g.*, Horrace and Schmidt, 1996).

7   Application to Nitrogen Surplus In Dairy Farms

The data set used in this paper was compiled by the Agricultural Economics Research Institute in the Hague using data on highly specialized dairy farms that were in the Dutch Farm Accountancy Data Network, a stratified random sample. Reinhard *et al.* (1999) and Reinhard (1999) describe the data in detail. The panel is unbalanced and we have 1,545 observations on $N = 613$ dairy farms in the Netherlands for some or all of 1991-94. There are no explanatory variables to be included in the efficiency distribution and we assume Cobb-Douglas forms for both frontiers. The dairy farms produce three outputs, $p = 2$ of these are good and $m = 1$ is bad, using $k = 3$ inputs:

- Good outputs: Milk (millions of kg) and Non-milk (millions of 1991 Guilders).

- Bad output: Nitrogen surplus (thousands of kg).

- Inputs: Family labor (thousands of hours), Capital (millions of 1991 Guilders) and Variable input (thousands of 1991 Guilders).

Variable input includes *inter alia* hired labor, concentrates, roughage and fertilizer. Non-milk output contains meat, livestock and roughage sold. The definition of capital includes land, buildings, equipment and livestock.

Table 4 provides empirical results for this data set based on 55,000 drawings with a burn-in of 5,000. Note that "RTS" means returns to scale, while τ_{1f} and τ_{2f} are the technical and environmental efficiencies for an unobserved (typical) farm (see Section 6).

The posterior properties of these efficiency measures and some of the most important parameters are given in Table 4. Most of the parameters are reasonably precisely estimated. The production frontier for goods shows increasing returns to scale, while the production frontier for bads shows slightly decreasing returns to scale. That is, it seems that increasing good output production implies a less than proportionate increase in Nitrogen surplus. It is interesting to note the relative magnitudes of the coefficients in the production frontier for Nitrogen surplus. These coefficients can be interpreted as elasticities, and the elasticity of bad output production with respect to milk production is much greater than the elasticity with respect to non-milk production. This finding indicates that it is the milk production side of dairy farming that is most associated with the production of Nitrogen.

The shape of the production equivalence surface for the good outputs depends on the parameters α and q. The values $\alpha_1 = \alpha_2 = \frac{1}{2}$ and $q = 1$, imply a linear production equivalence surface with a one-for-one tradeoff (in the units of measurement mentioned above) between the two outputs (see Fernández *et al.*, 2000, for details). Table 4 indicates that we are not too far from this case.

With stochastic frontier models, interest typically centers on the efficiencies. Space precludes the presentation of results on environmental and technical efficiency for each individual farm. The information in Table 4 about the posterior technical and environmental efficiencies of a typical (unobserved) firm is supplemented by Figures 5 and 6 which plot these posterior efficiency distributions along with those which correspond to the minimum, median and maximum efficient farms. These latter firms are selected based on the posterior mean efficiencies in a previous run.

The posteriors for τ_{1f} and τ_{2f} exhibit a high variance, indicating a great degree of spread in efficiencies across farms. Some farms are much more efficient than others in terms of both their production of good and bad outputs. Furthermore, these predictives are very different

from the prior efficiency distribution (see Figures 1-4) so we have learned something from the data.

It is also apparent that environmental efficiency tends to be lower than technical efficiency. In fact, the posterior median of the environmental efficiency for a typical farm is only 0.345, indicating that the typical farm is only 35% as efficient as the (hypothetical) best-practice farm in terms of pollution control. Thus, the typical farm is producing three times as much Nitrogen surplus than is consistent with best practice! This result seems to be driven by a minority of farms who are leading the way in minimizing Nitrogen surplus with a majority of farms falling far behind this lead. This finding seems less strange when we consider the incentives farmers face. There is an incentive to eliminate technical inefficiencies (*i.e.* profit maximization), and this is reflected in higher levels of technical efficiency. However, there were, at the time this data was collected, no financial incentives for farms to improve their environmental efficiency. Note that Reinhard *et al.* (1999) also find lower environmental efficiencies than technical ones (although they use a different definition of efficiencies in a classical framework with bads treated as inputs).

The posteriors for technical and environmental efficiency for the individual farms are less dispersed, but still fairly wide. A key advantage of Bayesian methods is that they lead to finite sample measures of uncertainty (*e.g.* 95% posterior density intervals) for firm-specific efficiencies. In empirical work, it often turns out that such 95% posterior density intervals are quite wide and the present application is no exception. Ranking farms in terms of their efficiencies could be of interest for policy purposes. A naive use of such a ranking could be highly misleading, in view of the posterior uncertainty in farm efficiencies. Perhaps a more realistic use of efficiency results is to see whether different groups of farms can be statistically distinguished from one another in terms of their efficiencies (*e.g.* can the best and worst farms be reliably differentiated in terms of their 95% posterior density intervals?). Figures 5 and 6 indicate that this is possible. That is, the efficiency distributions of the minimum, median and maximum firms do not overlap much. Tables 5 and 6 present additional evidence on this issue. Based on the posterior means in an earlier run, we rank all farms in terms of their efficiency performance and select the five farms which are at the quartiles and label these the minimum, 25th percentile, median, 75th percentile and maximum efficient farms. We then calculate the probability that, *e.g.,* the farm we label the

minimum efficient really is less efficient than the farm that we label the median efficient. We do this for every combination of farms. The tables indicate that we are able to distinguish well between the various quartiles, particularly for the minimum and maximum farms.

Finally, the correlation between τ_{1f} and τ_{2f} is 0.249, indicating that there is a slight tendency for farms which are technically inefficient to also be be environmentally inefficient. This result is reinforced by the fact that the correlation between the vectors of posterior means of farm-specific technical and environmental efficiencies is 0.365. A somewhat similar finding was reported by Reinhard *et al.* (1999).

8   Conclusion

In this paper we have introduced a framework for measuring environmental and technical efficiency in the context of a model for multiple good and bad outputs. A Markov chain Monte Carlo algorithm is developed to conduct Bayesian statistical inference. Results using artificial and real data show that this algorithm is computationally practical and yields reasonable results.

An application to Dutch dairy farms indicates that:

- Farms tend to be more efficient technically than environmentally.

- The (small) positive correlation between efficiencies indicates that farms which tend to be less efficient technically also tend to be less efficient environmentally.

- However, there is a large spread of efficiencies. This manifests itself in large differences between the 2.5 and 97.5 percentiles of the posteriors for the technical and environmental efficiencies of a typical farm.

- Increasing returns to scale seem to exist for good output production, while slightly decreasing returns exists for bad output production.

We hesitate to draw policy conclusions based solely on this one set of empirical results for one model specification. However, to illustrate the types of issues that our model can be used to address, we offer the following comments. The relatively large degree of environmental inefficiency indicates that pollution can be reduced in many farms at little cost

24

in terms of foregone good output. That is, if inefficient farms were to adopt best-practice technology and move towards their environmental production frontiers, production of pollutants could be reduced at no cost to milk or non-milk production. The positive correlation between the two types of efficiencies indicates that improving environmental efficiency could be associated with improvements in technical efficiency. Hence, policies aimed at improving efficiency (*e.g.* by educating farmers in best-practice technology) could have large payoffs. Furthermore, the pattern of returns to scale results indicate that larger farms have advantages. Hence, policies which promote rationalization of farms (*e.g.* encouraging larger farms to purchase smaller farms) could result both in more production of milk and non-milk outputs (due to increasing returns in the good production frontier) and less pollution (due to decreasing returns in the environmental production frontier).

Appendix: Posterior Conditional Densities

From combining (2.11) and the prior described in Section 4, we obtain the following conditional posterior density functions:

Conditional for (z, v)**:**

$$p(z, v | Y, B, \phi, \psi, \Omega, \beta, \delta, \Sigma, \alpha, \gamma, q, r, s, h) \propto f_N^{2N} \left(\begin{array}{c} z \\ v \end{array} | d, R \right) I_{\Re_+^{2N}}(z, v), \qquad (A.1)$$

where, since $D = M$ (otherwise a slight generalization is required),

$$R^{-1} = \left(\begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right) \Sigma^{-1} \left(\begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right) \otimes D'D + \Omega^{-1} \otimes I_N, \qquad (A.2)$$

and, defining $G = (g_1, ..., g_N)'$,

$$d = R \left[\left(\begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right) (\Sigma^{-1} \otimes D') \left(\begin{array}{c} V\beta - \log \theta \\ U\delta - \log \kappa \end{array} \right) + (\Omega^{-1} \otimes I_N) \left(\begin{array}{c} G\phi \\ G\psi \end{array} \right) \right]. \qquad (A.3)$$

If $D'D$ is diagonal (as is the case for the "individual effects" option used here, which corresponds to $D'D = diag(T_1, \ldots, T_N)$), we have conditional independence across observational units and can draw (z_i, v_i) separately from a univariate truncated Normal for each $i = 1, \ldots, N$, which greatly simplifies the numerical effort.

Conditional for (β, δ)**:**

$$p(\beta, \delta | Y, B, z, v, \phi, \psi, \Omega, \Sigma, \alpha, \gamma, q, r, s, h) \propto f_N \left(\begin{array}{c} \beta \\ \delta \end{array} | b_*, H_*^{-1} \right) I_{\mathcal{RR}}(\beta, \delta, \Sigma), \qquad (A.4)$$

with

$$H_* = H_0 + \begin{pmatrix} V' & 0 \\ 0 & U' \end{pmatrix} (\Sigma^{-1} \otimes I_{NT}) \begin{pmatrix} V & 0 \\ 0 & U \end{pmatrix} \qquad (A.5)$$

$$b_* = H_*^{-1} \left[H_0 b_0 + \begin{pmatrix} V' & 0 \\ 0 & U' \end{pmatrix} (\Sigma^{-1} \otimes I_{NT}) \begin{pmatrix} \log\theta + Dz \\ \log\kappa - Mv \end{pmatrix} \right]. \qquad (A.6)$$

Conditional for Σ:

$$p(\Sigma|Y, B, z, v, \phi, \psi, \Omega, \beta, \delta, \alpha, \gamma, q, r, s, h) \propto f_{IW}^2(\Sigma|\Sigma_0 + A_*' A_*, \lambda_0 + NT) I_{\mathcal{RR}}(\beta, \delta, \Sigma), \quad (A.7)$$

where we have defined the $NT \times 2$ matrix

$$A_* = (\log\theta - V\beta + Dz, \log\kappa - U\delta - Mv). \qquad (A.8)$$

Conditional for Ω:

$$p(\Omega|Y, B, z, v, \phi, \psi, \beta, \delta, \Sigma, \alpha, \gamma, q, r, s, h) \qquad (A.9)$$

$$\propto \prod_{i=1}^{N} f^{-1}(\mu_i, \Omega) f_{IW}^2\left(\Omega\Big|\frac{\nu_0}{1+ac}\Omega_0 + (z - G\phi, v - G\psi)'(z - G\phi, v - G\psi), \nu_0 + N\right).$$

Conditional for ϕ, ψ:

$$p(\phi, \psi|Y, B, z, v, \Omega, \beta, \delta, \Sigma, \alpha, \gamma, q, r, s, h) \propto \prod_{i=1}^{N} f^{-1}[(\phi, \psi)' g_i, \Omega] \qquad (A.10)$$

$$f_N^{2d}\left(\begin{pmatrix} \phi \\ \psi \end{pmatrix} \Big| (\{I_2 \otimes [G'G + A^{-1}(d)]^{-1} G'\} \begin{pmatrix} z \\ v \end{pmatrix}, \Omega \otimes [G'G + A^{-1}(d)]^{-1}\right)$$

Conditional for α:

$$p(\alpha|Y, B, z, v, \phi, \psi, \Omega, \beta, \delta, \Sigma, \gamma, q, r, s, h) \qquad (A.11)$$

$$\propto \prod_{j=1}^{p} \alpha_j^{a_{0j} + qNTs_j - 1} \prod_{i,t} \left(\sum_{l=1}^{p} \alpha_l^q y_{(i,t,l)}^q\right)^{-\sum_l s_l} \exp -\frac{1}{2} \mathrm{tr}\Sigma^{-1} A_*' A_*,$$

defining $a_0 = (a_{01}, \ldots, a_{0p})'$.

Conditional for γ:

$$p(\gamma|Y, B, z, v, \phi, \psi, \Omega, \beta, \delta, \Sigma, \alpha, q, r, s, h) \qquad (A.12)$$

$$\propto \prod_{j=1}^{m} \gamma_j^{g_{0j} + rNTh_j - 1} \prod_{i,t} \left(\sum_{l=1}^{m} \gamma_l^r b_{(i,t,l)}^r\right)^{-\sum_l h_l} \exp -\frac{1}{2} \mathrm{tr}\Sigma^{-1} A_*' A_*,$$

defining $g_0 = (g_{01}, \ldots, g_{0m})'$.

Conditional for q:

$$p(q|Y, B, z, v, \phi, \psi, \Omega, \beta, \delta, \Sigma, \alpha, \gamma, r, s, h) \tag{A.13}$$

$$\propto \quad q^{NT(p-1)} I_{(1,\infty)}(q) \exp - \left[q q_0 + \sum_{j=1}^{p} \left\{ s_j \sum_{i,t} (\ln \sum_{l=1}^{p} \alpha_l^q y_{(i,t,l)}^q - q \ln \alpha_j y_{(i,t,j)}) \right\} + \frac{1}{2} \mathrm{tr} \Sigma^{-1} A_*' A_* \right].$$

Conditional for r:

$$p(r|Y, B, z, v, \phi, \psi, \Omega, \beta, \delta, \Sigma, \alpha, \gamma, q, s, h) \tag{A.14}$$

$$\propto \quad r^{NT(m-1)} I_{(0,1)}(r) \exp - \left[r r_0 + \sum_{j=1}^{m} \left\{ h_j \sum_{i,t} (\ln \sum_{l=1}^{m} \gamma_l^r b_{(i,t,l)}^r - r \ln \gamma_j b_{(i,t,j)}) \right\} + \frac{1}{2} \mathrm{tr} \Sigma^{-1} A_*' A_* \right].$$

Conditional for the elements of s:

$$p(s_j|Y, B, z, v, \phi, \psi, \Omega, \beta, \delta, \Sigma, \alpha, \gamma, q, r, s_{(-j)}, h) \tag{A.15}$$

$$\propto \quad \left[\frac{\Gamma(\sum_{l=1}^{p} s_l)}{\Gamma(s_j)} \right]^{NT} \exp - s_j \left\{ k_j + \sum_{i,t} (\ln \sum_{l=1}^{p} \alpha_l^q y_{(i,t,l)}^q - q \ln \alpha_j y_{(i,t,j)}) \right\},$$

where $s_{(-j)}$ denotes the vector s without the jth element.

Conditional for the elements of h:

$$p(h_j|Y, B, z, v, \phi, \psi, \Omega, \beta, \delta, \Sigma, \alpha, \gamma, q, r, s, h_{(-j)}) \tag{A.16}$$

$$\propto \quad \left[\frac{\Gamma(\sum_{l=1}^{m} h_l)}{\Gamma(h_j)} \right]^{NT} \exp - h_j \left\{ n_j + \sum_{i,t} (\ln \sum_{l=1}^{m} \gamma_l^r b_{(i,t,l)}^r - r \ln \gamma_j b_{(i,t,j)}) \right\},$$

where $h_{(-j)}$ denotes the vector h without the jth element.

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.

Bauer, P. (1990), "Recent Developments in the Econometric Estimation of Frontiers," *Journal of Econometrics*, 46, 39-56.

Färe, R., Grosskopf, S., Lovell, C.A.K., and Pasurka, C. (1989), "Multilateral Productivity Comparisons When Some Outputs Are Undesirable: A Nonparametric Approach," *Review of Economics and Statistics*, 71, 90-98.

Fernández, C., Koop, G., and Steel, M.F.J. (2000), "A Bayesian Analysis of Multiple-output Production Frontiers, *Journal of Econometrics*, forthcoming.

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.

Gumbel, E.J. (1960), "Bivariate Exponential Distributions," *Journal of the American Statistical Association*, 55, 698-707.

Horrace, W., and Schmidt, P. (1996), "Confidence Statements for Efficiency Estimates From Stochastic Frontier Models, *Journal of Productivity Analysis*, 7, 257-282.

Koop, G. (1998), "Carbon Dioxide Emissions and Economic Growth: A Structural Approach," *Journal of Applied Statistics*, 25, 489-515.

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., and Steel, M.F.J. (2000), "Bayesian Analysis of Stochastic Frontier Models," in *Companion to Econometrics*, ed. B. Baltagi, Oxford: Basil Blackwell, forthcoming.

Lawrance, A.J. and Lewis, P.A.W. (1983), "Simple Dependent Pairs of Exponential and Uniform Random Variables," *Operations Research*, 31, 1179-1197.

Meeusen, W., and van den Broeck, J. (1977), "Efficiency Estimation From Cobb-Douglas Production Functions With Composed Errors. *International Economic Review*, 8, 435-444.

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.

Reinhard, S. (1999), "Econometric Analysis of Economic and Environmental Efficiency of Dutch Dairy Farms, PhD dissertation, Wageningen University.

Reinhard, S., Lovell, C.A.K., and Thijssen, G. (1999), "Econometric Application of Technical and Environmental Efficiency: An Application to Dutch Dairy Farms," *American Journal of Agricultural Economics*, 81, 44-60.

Table 1: Prior Elicitation (efficiency properties)

	Mean	St.Dev.	Median	2.5%	97.5%	Corr.
$d = 1$						
$\omega_{ii}^0 = 0.10$	0.837	0.164	0.896	0.411	0.997	0.107
$\omega_{ii}^0 = 0.30$	0.752	0.226	0.827	0.211	0.995	0.083
$\omega_{ii}^0 = 0.65$	0.681	0.267	0.758	0.101	0.993	0.090
$\omega_{ii}^0 = 1.00$	0.635	0.287	0.704	0.058	0.991	0.069
$\omega_{ii}^0 = 2.00$	0.561	0.312	0.610	0.019	0.987	0.068
$d = 2$						
$\omega_{ii}^0 = 0.10$	0.837	0.166	0.897	0.408	0.997	0.098
$\omega_{ii}^0 = 0.30$	0.753	0.226	0.828	0.210	0.995	0.100
$\omega_{ii}^0 = 0.65$	0.679	0.267	0.754	0.103	0.993	0.082
$\omega_{ii}^0 = 1.00$	0.640	0.286	0.711	0.060	0.992	0.096
$\omega_{ii}^0 = 2.00$	0.564	0.312	0.617	0.018	0.988	0.070

Table 2: Properties of Coefficients for the Chosen Prior

	Mean	St.Dev.	Median	2.5%	97.5%
$d = 1$					
ϕ_1	0.000	1.140	0.000	-2.270	2.270
$d = 2$					
ϕ_1	0.000	1.075	0.000	-2.141	2.141
ϕ_2	0.000	0.658	0.000	-1.311	1.311

Table 3: Posterior Results for Artificial Data Sets

Parameter	Actual	Median	Set 1 2.5%	97.5%	Actual	Median	Set 2 2.5%	97.5%
β_1	0.5	0.496	0.187	0.812	0.5	0.042	-0.365	0.509
β_2	0.5	0.544	0.280	0.812	0.5	0.267	0.022	0.596
δ_1	0.2	0.150	-0.240	0.493	0.2	-0.595	-1.660	0.983
δ_2	0.5	0.481	0.366	0.603	0.5	0.461	0.370	0.559
δ_3	0.5	0.502	0.382	0.628	0.5	0.416	0.314	0.531
σ_{11}	1.0	0.988	0.907	1.077	1.0	0.973	0.893	1.064
σ_{12}	-0.25	-0.251	-0.491	-0.003	-0.25	-0.141	-0.339	-0.008
σ_{22}	1.0	1.030	0.907	1.211	1.0	0.967	0.878	1.088
ω_{11}	0.25	0.413	0.142	0.898	0.25	0.341	0.241	0.622
ω_{12}	0.0	-0.115	-0.456	0.105	0.0	-0.074	-0.231	0.032
ω_{22}	0.25	0.217	0.077	0.715	0.25	0.235	0.166	0.509
ϕ_1	0.0	-0.170	-1.102	1.036	1.0	0.140	-0.949	0.771
ϕ_2	0.0	-0.411	-1.425	0.194	1.0	1.140	0.699	2.088
ψ_1	0.0	0.353	-0.717	0.911	1.0	1.855	-0.287	2.951
ψ_2	0.0	0.258	-0.199	1.186	1.0	0.847	0.444	1.507
q	1.5	1.288	1.027	1.685	1.5	1.819	1.367	2.446
r	0.75	0.676	0.563	0.862	0.75	0.751	0.617	0.926
α_1	0.5	0.478	0.436	0.521	0.5	0.488	0.455	0.526
γ_1	0.5	0.504	0.459	0.552	0.5	0.504	0.460	0.544
s_1	1.0	1.299	0.843	1.813	1.0	0.759	0.523	1.138
s_2	1.0	1.385	0.906	1.925	1.0	0.849	0.572	1.295
h_1	1.0	1.156	0.800	1.485	1.0	1.013	0.768	1.391
h_2	1.0	1.184	0.829	1.555	1.0	0.987	0.743	1.295

Table 4: Posterior Results for Dutch Dairy Farm Data Set

	Median	2.5%	97.5%
β_1(Intercept)	-3.533	-3.694	-3.226
β_2(Labour)	0.120	0.090	0.150
β_3(Capital)	0.537	0.504	0.572
β_4(Variable)	0.487	0.463	0.509
RTS Goods	1.145	1.115	1.173
δ_1(Intercept)	2.578	2.262	2.890
δ_2(Milk)	0.889	0.858	0.921
δ_3(Non-milk)	0.081	0.065	0.098
RTS Bads	0.971	0.940	1.001
τ_{1f}	0.620	0.415	0.880
τ_{2f}	0.345	0.198	0.599
q	1.004	1.000	1.019
α_1	0.534	0.510	0.565

Table 5: Probability that Farm in Column is Less Technically Efficient than Farm in Row

Farm	Min	25th	Median	75th
25th	1.000			
Median	1.000	0.930		
75th	1.000	0.996	0.861	
Max	1.000	1.000	1.000	1.000

Table 6: Probability that Farm in Column is Less Environmentally Efficient than Farm in Row

Farm	Min	25th	Median	75th
25th	1.000			
Median	1.000	0.821		
75th	1.000	0.920	0.776	
Max	1.000	1.000	1.000	1.000

Figure 2: Environmental Efficiency Densities

artificial data set 1

True
Predictive
Prior

Figure 1: Technical Efficiency Densities

artificial data set 1

True
Predictive
Prior
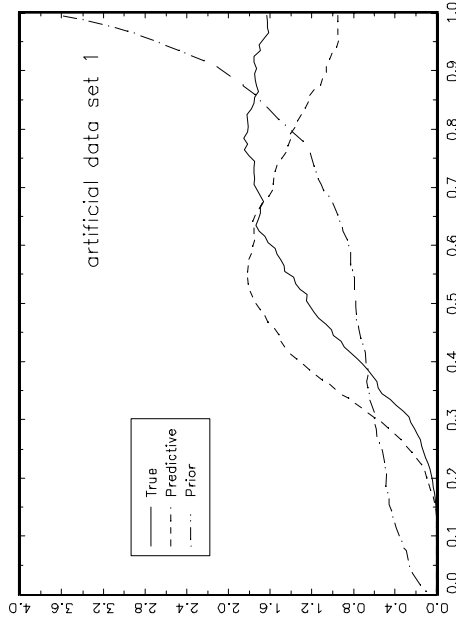
Figure 4: Environmental Efficiency Densities

artificial data set 2

True
Predictive
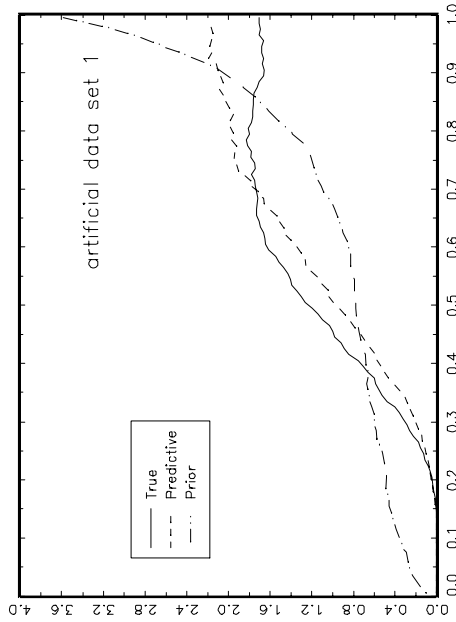Prior

Figure 3: Technical Efficiency Densities

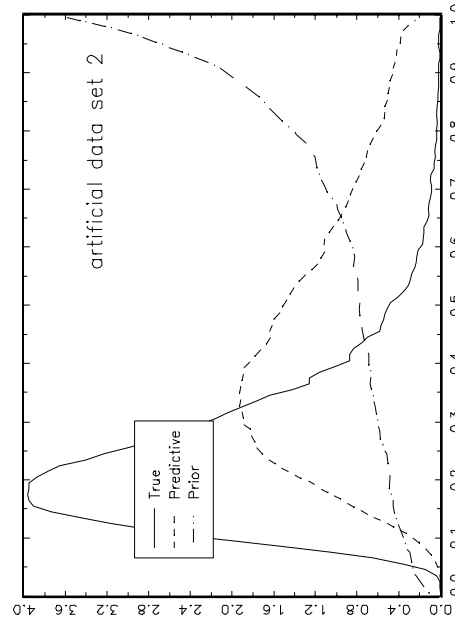artificial data set 2

True
Predictive
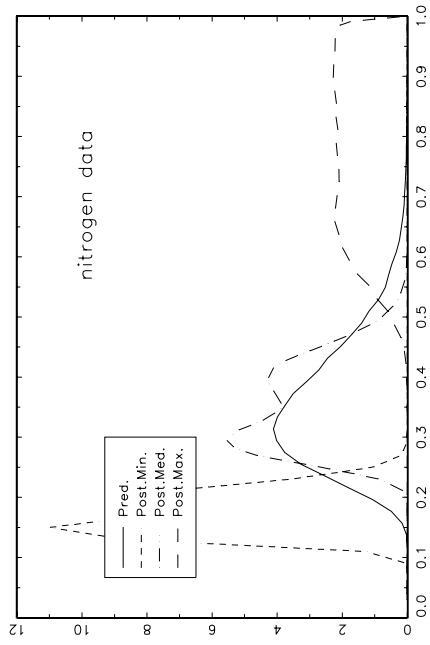Prior

Figure 6: Posterior Density of Environmental Efficiency

Figure 5: Posterior Density of Technical Efficiency