Estimating Variable Returns to Scale Production Frontiers with Alternative Stochastic Assumptions

William E. Griffiths

University of Melbourne

Christopher J. O'Donnell

University of Queensland

First version: October 20, 2002

This version: January 15, 2003

Abstract

A stochastic production frontier model is formulated within the generalized production function framework popularized by Zellner and Revankar (1969) and Zellner and Ryu (1998). This framework is convenient for parsimonious modeling of a production function with variable returns to scale specified as a function of output. Two alternatives for introducing the stochastic inefficiency term and the stochastic error are considered, one where they are appended to the existing equation for the production relationship and one where the existing equation is solved for the log of output before the stochastic terms are added. The latter alternative is novel, but it is needed to preserve the usual definition of firm efficiency. The two alternative stochastic assumptions are considered in conjunction with two returns to scale functions, making a total of four models that are considered. A Bayesian framework for estimating all four models is described. The techniques are applied to USDA state-level data on agricultural output and four inputs. Posterior distributions for all parameters, firm efficiencies and the efficiency rankings of firms are obtained. The sensitivity of the results to the returns to scale specification and to the stochastic specification is examined.

1. **Introduction**

The estimation of stochastic production frontier models is a common procedure for assessing the efficiency of firms within an industry. Several versions of stochastic frontier models have been suggested in the literature, designed to accommodate the varying nature of data and specific characteristics of empirical applications. A typical model relevant for panel data involving observations on a number of firms over time can be written as

$$\log y_{it} = \log f(x_{it}, \beta) - z_i + u_{it} \tag{1}$$

In this equation y_{it} denotes output for the i-th firm in the t-th time period, x_{it} is a corresponding vector of inputs, β represents a vector of unknown parameters, and $f(x_{it}, \beta)$ is the deterministic part of the production frontier. It is assumed that the random errors u_{it} capture measurement and/or specification error and that they are independent normal random variables with mean zero and variance ω^{-1}. Each z_i is assumed to be a nonnegative random variable that describes the inefficiency of the i-th firm in terms of the distance of $\log y_{it}$ from the stochastic frontier $\log f(x_{it}, \beta) + u_{it}$. Alternative distributions that have been suggested in the literature for the z_i include the exponential, gamma, truncated normal and half-normal distributions. The inefficiency term is assumed to be constant over time, although this assumption can be relaxed. To measure efficiency (rather than inefficiency), and to make measurement of efficiency across firms comparable, it is conventional to use $\tau_i = \exp(-z_i)$ to denote the efficiency of the i-th firm. Since $0 \le \tau_i \le 1$, this measure allows us to say that the i-th firm is $100\tau_i\%$ efficient. Also, it has a natural interpretation as the ratio of mean output conditional on the inefficiency of the i-th firm (z_i), to mean output on the frontier (conditional on $z_i = 0$). This interpretation is expressed algebraically as

$$\frac{E(y_{it} \mid z_i)}{E(y_{it} \mid z_i = 0)} = \frac{f(x_{it}, \beta) \exp(-z_i) E[\exp(u_{it})]}{f(x_{it}, \beta) E[\exp(u_{it})]} = \exp(-z_i) = \tau_i \tag{2}$$

Reviews by Greene (1997) and Koop and Steel (2001) provide a convenient access to the extensive literature on stochastic production frontiers and its historical development.

Greene emphasizes the sampling theory approach while Koop and Steel focus on Bayesian inference.

In this paper we are concerned with Bayesian estimation of stochastic frontier models that exhibit variable returns to scale. One way to specify a model with variable returns to scale is to choose an appropriate form for the function $\log f(x_{it}, \beta)$. For example, a Cobb-Douglas specification is not satisfactory because it exhibits constant returns to scale. On the other hand, a translog specification for $\log f(x_{it}, \beta)$ yields a returns to scale function that is a linear function of the logs of the inputs; it can exhibit regions of increasing, constant and decreasing returns to scale. The translog model has some disadvantages, however. It does not automatically satisfy the regularity conditions of concavity and monotonicity, functions with several inputs require estimation of a large number of parameters, and the relationship between the substitutability of the inputs and the returns to scale may be a complicated undesirable one. An alternative approach without these problems, and the approach adopted in this paper, is the generalized production function specification pursued by Zellner and Revankar (1969), Revankar (1971), Zellner (1971, p.176), and Zellner and Ryu (1998). In this model the production function is assumed to be homothetic, implying it can be written as

$$y = g^*[f(x, \beta)] \tag{3}$$

where g^* is a monotonic transformation and $f(x, \beta)$ is a homogeneous function of degree μ. For the moment, we omit the i and t subscripts and the stochastic inefficiency and error terms. In the context of equation (3), the returns to scale (RTS) function is defined as

$$\text{RTS}(y) = \mu \frac{dy/y}{df/f} \tag{4}$$

A 'generalized production function' is obtained by specifying functions for $\text{RTS}(y)$ and $f(x, \beta)$, and solving the differential equation in (4) to yield an explicit representation of

the production relationship in equation (3). For example, Zellner and Revankar (1969) suggest the RTS function

$$RTS(y) = \frac{\mu}{1+\theta y} \tag{5}$$

that leads to the production relationship

$$\log y + \theta y = \log f(x,\beta) \tag{6}$$

If y is scaled such that $y \geq 1$, and the inequality restrictions $\theta > 0$, $\mu > 1+\theta$ hold, the RTS is greater than one for low outputs and decreases monotonically as output increases, leading to a U-shaped average cost curve that has a minimum when $RTS(y) = 1$. The RTS function in (5) achieves this desirable property with the introduction of only one additional parameter, θ. Also, the production relationship satisfies concavity and monotonicity regularity conditions as long as $f(x,\beta)$ is chosen to have these properties, and substitutability of the inputs is governed by the function $f(x,\beta)$.

If functions like (6) are considered desirable for modeling stochastic production frontiers with variable returns to scale, the next question that must be addressed is the way in which stochastic inefficiency and error terms are introduced. One possibility is that adopted by Kumbhakar (1988) and Kumbhakar et al (1991) who append the stochastic terms to equation (6). This strategy is a natural extension of the assumption employed in Zellner and Revankar (1969), Zellner (1971) and Zellner and Ryu (1998) where a single normally distributed error is attached to equation (6). In the panel data context described earlier Kumbhakar's strategy leads to the model

$$\log y_{it} + \theta y_{it} = \log f(x_{it},\beta) - z_i + u_{it} \tag{7}$$

One property of this function is that, in contrast to equation (1), the usual assumptions of homoskedastic z_i and u_{it} no longer imply that $\log y_{it}$ will be homoskedastic. It is not possible to find an analytical solution for $\log y_{it}$ from equation (7). Suppose, however, a numerical solution is available. It will be of the form

$$\log y_{it} = g[\theta, \, \log f(x_{it}, \beta) - z_i + u_{it}] \tag{8}$$

The variance of $\log y_{it}$ from (8) will, in general, depend on $f(x_{it}, \beta)$; and hence $\log y_{it}$ will be heteroskedastic. The existence of heteroskedasticity complicates the notion of the frontier. The efficiency measure in equation (2) will no longer be a simple function of z_i, but will depend on the input levels x_{it}. Specifically, if $u_{it} : N(0, \omega^{-1})$, then

$$f(y_{it} \mid z_i) = \left(\frac{\omega}{2\pi}\right)^{1/2} \left(\frac{1 + \theta y_{it}}{y_{it}}\right) \exp\left\{-\frac{\omega}{2}\left(\log y_{it} + \theta y_{it} - \log f(x_{it}, \beta) + z_i - u_{it}\right)^2\right\} \tag{9}$$

and the efficiency of the i-th firm can be written as

$$\frac{E(y_{it} \mid z_i)}{E(y_{it} \mid z_i = 0)} = \frac{\int (1 + \theta y_{it}) \exp\left\{-\frac{\omega}{2}\left(\log y_{it} + \theta y_{it} - \log f(x_{it}, \beta) + z_i - u_{it}\right)^2\right\} dy_{it}}{\int (1 + \theta y_{it}) \exp\left\{-\frac{\omega}{2}\left(\log y_{it} + \theta y_{it} - \log f(x_{it}, \beta) - u_{it}\right)^2\right\} dy_{it}} \tag{10}$$

The stochastic assumptions in (7) lead to a complex expression for efficiency involving integrals for which no closed form solution is apparent. However, despite the complexity of (10), it is nevertheless possible to use the cost function to find an alternative definition of inefficiency that is relatively simple. This interpretation was noted by Kumbhakar (1988); we consider it explicitly later in the paper.

A second way to introduce the stochastic inefficiency and error terms is to view them as being appended to the numerical solution for $\log y$ from equation (6). If the solution for $\log y$ is given by $\log y = g[\theta, \log f(x, \beta)]$, then this model can be written as

$$\log y_{it} = g[\theta, \, \log f(x_{it}, \beta)] - z_i + u_{it} \tag{11}$$

Including the stochastic terms in this way is more natural. Under the usual assumptions about z_i and u_{it}, $\log y_{it}$ is homoskedastic, and the inefficiency interpretation of z_i given in equation (2) still holds. Specifications like (11) have been overlooked in the literature,

probably because of the need to obtain the solution $\log y = g[\theta, \log f(x, \beta)]$. However, as we will see, estimation is still possible within the framework of Bayesian inference.

The objectives of this paper are to describe and illustrate Bayesian methodology for estimating stochastic frontiers like those described in equations (7) and (11). Two RTS functions are considered, that in equation (5) and one other. These two are a subset of five functions considered by Zellner and Ryu (1998) in their application to the transport equipment industry. We compare estimates of the parameters and the firm efficiencies under the two different RTS specifications and the two different stochastic specifications. As far as we are aware, Bayesian estimation of functions like (7) has not appeared in earlier literature, and estimation of functions like (11) has not been attempted using sampling theory or Bayesian inference. Our strategy for estimation, and the prior assumptions that we adopt, are modifications of those described in Koop and Steel (2001), adapted to accommodate the introduction of the RTS function and the different stochastic assumptions. One particular novelty in our work is the ability to estimate the parameters β and θ in equation (11), even although no analytical expression for g is available. This problem is overcome by solving for g numerically, within a Markov chain Monte Carlo algorithm. To ensure similar prior information is used for both functions we place a prior density function on the RTS when output is unity, and on the level of output for which average cost is a minimum $(RTS = 1)$. This prior is used to derive a prior on μ, the degree of homogeneity of the function f, and on the parameter in the RTS function.

One of the advantages of the Bayesian approach (also noted by Koop and Steel) is the ease with which useful inferences can be made about quantities that are quite complicated, or intractable functions from a sampling theory point of view. For example, we can provide posterior density functions for making finite-sample inferences about firm efficiencies, the efficiency ranking of each firm, the output at which average cost is a minimum and the probability that one firm is more efficient than another. We provide examples such as these from our application, thus illustrating the flexibility of the Bayesian approach.

The plan of the paper is as follows. The model and stochastic assumptions are presented in Section 2. Section 3 contains descriptions of the prior density functions. The

conditional posterior density functions and the Markov chain Monte Carlo (MCMC) algorithms are described in Section 4. The empirical application and results are presented in Section 5, with concluding remarks being made in Section 6.

2.   Models and Assumptions

The two models that we consider can be written as

$$\text{ZR}: \qquad \log y_{it} + \theta y_{it} = \log f(x_{it}, \beta) \qquad (12)$$

$$\text{NR}: \qquad \log y_{it} + \gamma(\log y_{it})^2 = \log f(x_{it}, \beta) \qquad (13)$$

The function ZR is attributable to Zellner and Revankar (1969); it was discussed in the introduction. The function NR is attributable to Nerlove (1963) and Ringstad (1967). It is derived from the RTS function

$$\text{RTS}(y) = \frac{\mu}{1 + 2\gamma \log y} \qquad (14)$$

For this function to decrease monotonically from a point above one to a point below one, producing a "U-shaped" average cost curve with a minimum at $\text{RTS} = 1$, we require the units of output to be such that $y \geq 1$; the required parameter restrictions are $\mu > 1$ and $\gamma > 0$.

The same function f is assumed for both models, namely a Cobb-Douglas of the form

$$\log f(x_{it}, \beta) = \beta_0 + \beta_1 \log x_{1it} + \beta_2 \log x_{2it} + \beta_3 \log x_{3it} + \beta_4 \log x_{4it} + \beta_5 \log(time)$$

$$= \beta_0 + \beta_1 \log\left(\frac{x_{1it}}{x_{4it}}\right) + \beta_2 \log\left(\frac{x_{2it}}{x_{4it}}\right) + \beta_3 \log\left(\frac{x_{3it}}{x_{4it}}\right) + \mu \log x_{4it} + \beta_5 \log(time)$$

$$(15)$$

where $\mu = \beta_1 + \beta_2 + \beta_3 + \beta_4$ is the degree of homogeneity of the function f with respect to the inputs x_1, x_2, x_3 and x_4. The second line in equation (15), where β_4 is replaced by μ, is the parameterisation used for estimation. The term $\log(time)$ is included to capture technical change that is assumed to be Hicks neutral. Including both technical change and

scale effects does raise questions about whether misspecifying one of these functions will lead to misleading inferences about the other. See, for example, Calem (1990). However, we believe that including some form of technical change is better than ignoring it.

At this point, it is convenient to give some information about the application, with further details being provided in Section 5. The data we use are annual USDA state-level data on agricultural output (y) and the four inputs, materials (x_1), capital (x_2), land (x_3), and labor (x_4). A "firm" is taken as a USDA farm production region, of which there are ten. Because there are more states than production regions, the subscript t is not a simple time index. It is an index for the values from all states in a given region, over the 26-year period 1971-1996. Since the number of states in each region varies, we write $t = 1,2,...,T_i$ and $i = 1,2,...,10$, where T_i is a multiple of 26. The variable *time* takes values 1 to 26 for each state.

The two models, each with two different stochastic assumptions, appear in equations (16) through (19). We use the abbreviation NR(het) to describe the Nerlove-Ringstad model with heteroskedastic $\log y$ and the abbreviation NR(hom) to describe the Nerlove-Ringstad model with homoskedastic $\log y$. ZR(het) and ZR(hom) are used for similar descriptions of the Zellner-Revankar models.

NR(het):
$$\log y_{it} + \gamma \left(\log y_{it} \right)^2 = \log f \left(x_{it}, \beta \right) - z_i + u_{it} \qquad (16)$$

NR(hom):
$$\log y_{it} = \frac{-1 + \sqrt{1 + 4\gamma \log f \left(x_{it}, \beta \right)}}{2\gamma} - z_i + u_{it} \qquad (17)$$

Equation (17) is obtained by solving equation (13) for $\log y_{it}$ and then appending the stochastic terms. Unlike the ZR(hom) model, a closed form solution for $\log y_{it}$ can be obtained. The second root of the quadratic has been omitted because it violates the restrictions on the RTS function that imply $\log y_{it} \geq 0$.

ZR(het):
$$\log y_{it} + \theta y_{it} = \log f(x_{it}, \beta) - z_i + u_{it} \qquad (18)$$

ZR(hom):
$$\log y_{it} = g[\theta, \log f(x_{it}, \beta)] - z_i + u_{it} \qquad (19)$$

In all cases the u_{it} are assumed to be independent $N\left(0, \omega^{-1}\right)$. Several alternatives have been suggested in the literature for the distribution of z_i, including the exponential, half normal and normal distributions. See Koop et al (1995) for a discussion of Bayesian estimation of the various alternatives. We follow Koop and Steel (2001) and assume the z_i are independent drawings from the exponential distribution

$$p\left(z_i\right) = \frac{1}{\lambda} \exp\left\{-\frac{z_i}{\lambda}\right\} \tag{20}$$

In addition, z_i and u_{it} are assumed to be independent.

For the 'homoskedastic' functions, $\tau_i = \exp(-z_i)$ is taken as the measure of inefficiency of the i-th firm. For the 'heteroskedastic' functions, we consider two measures of efficiency. The first is given in equation (10) for the ZR model; a similar expression can be derived for the NR model. We refer to these efficiency measures as those obtained from the output function. In the application the integrals in these equations were evaluated numerically for each x_{it} and for each parameter draw made using an MCMC algorithm. The second efficiency measure is that noted by Kumbhakar (1988), and is derived from the cost function. If a firm has knowledge of z_i and u_{it}, and, to achieve a given level of output y_{it}, it uses input levels that maximize profits, the cost function for the ZR(het) model can be written as

$$\log C_{it} = \psi_{it} + \frac{\log y_{it} + \theta y_{it}}{\mu} + \frac{z_i}{\mu} - \frac{u_{it}}{\mu} \tag{21}$$

where ψ_{it} depends on β and the input prices. We can interpret $\exp(z_i/\mu)$ as the proportional increase in costs necessary for the i-th firm to achieve a given level of output, relative to the cost for a firm that lies on the frontier. To give this measure an efficiency interpretation that can be compared with the other efficiency measures, in our application we report its inverse $\exp(-z_i/\mu)$. The same result also holds for the NR(het) model.

3. Prior Distributions

The parameters requiring prior distributions are $\beta_0, \beta_1, \beta_2, \beta_3, \mu, \beta_5, \lambda$ and ω for all functions, θ for the ZR models, and γ for the NR models. Caution needs to be exercised if improper priors are used for some of these parameters. For the case where $E[\log y_{it}]$ is a linear function of β, Fernández et al (1997) show that, in the absence of panel data, proper priors on both λ and ω are required to obtain proper posterior densities. The improper prior $p(\omega) \propto \omega^{-1}$ can be used when panel data are available, but a proper prior on λ still needs to be specified. Accordingly, we use $p(\omega) \propto \omega^{-1}$ as a prior for ω, and for λ^{-1} we use the exponential distribution

$$p\left(\lambda^{-1}\right) = c_\lambda \exp\left(-c_\lambda \lambda^{-1}\right) \tag{22}$$

Van den Broeck et al (1994) show that this prior density leads to a proper, but relatively non-informative, prior on the firm efficiencies, with a convenient setting for c_λ given by $c_\lambda = -\log(\tau^*)$ where τ^* is the prior median for the efficiency distribution.

A prior density function for the remaining parameters is obtained by:

1. Specifying proper priors for the RTS at the point $y = 1$ and the level of output for which $RTS = 1$.
2. Transforming these priors to a prior on μ and γ (or μ and θ).
3. Specifying uniform priors for $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_5)$.
4. Including the necessary inequality restrictions on all parameters.

Let

$r =$ the logarithm of RTS when $y = 1$, and

$q =$ the logarithm of output for which $RTS = 1$.

These parameters can be viewed as governing the magnitude of average cost when output is low, and how quickly average cost declines to reach its minimum. Ignoring for the moment inequality restrictions on these quantities, we take for their priors independent exponential distributions given by

$$p(q) = c_q \exp(-c_q q) \tag{23}$$

$$p(r) = c_r \exp(-c_r r) \tag{24}$$

Looking first at the implications of these priors for the parameters of the NR models, we note, from equation (14), that $RTS(1) = \mu$ and hence $r = \log \mu$. The resulting prior for μ is

NR: $$p(\mu) = p(r)\left|\frac{\partial r}{\partial \mu}\right| = \frac{c_r}{\mu^{c_r+1}} \tag{25}$$

To obtain an expression for q, the value of $\log y$ for which $RTS = 1$, we solve

$$\frac{\mu}{1 + 2\gamma q} = 1 \tag{26}$$

to obtain $q = (\mu - 1)/2\gamma$. Thus,

NR: $$p(\gamma \mid \mu) = p(q)\left|\frac{\partial q}{\partial \gamma}\right| = \frac{c_q(\mu-1)}{2\gamma^2}\exp\left\{-\frac{c_q(\mu-1)}{2\gamma}\right\} \tag{27}$$

Equations (25) and (27) define the joint prior $p(\gamma,\mu)$ for the NR models.

In the ZR models the expressions relating r and q to μ and θ are, from (5),

$$r = \log\left(\frac{\mu}{1+\theta}\right) \qquad\qquad q = \log\left(\frac{\mu-1}{\theta}\right) \tag{28}$$

leading to a joint prior for (μ,θ) given by

$$p(\mu,\theta) = p(r,q)\left\|\begin{matrix}\dfrac{\partial r}{\partial \mu} & \dfrac{\partial r}{\partial \theta} \\[2mm] \dfrac{\partial q}{\partial \mu} & \dfrac{\partial q}{\partial \theta}\end{matrix}\right\| = c_r c_q\left[\mu - (1+\theta)\right]\frac{(1+\theta)^{c_r-1}\theta^{c_q-1}}{\mu^{c_r+1}(\mu-1)^{c_q+1}} \tag{29}$$

The β_i are assigned uniform priors, but several inequality restrictions are imposed on the parameters. They are:

Both models: $\quad \beta_1 > 0, \ \beta_2 > 0, \ \beta_3 > 0, \ \beta_5 > 0, \ \beta_1 + \beta_2 + \beta_3 < \mu \tag{30}$

NR: $\quad \mu > 1, \ \gamma > 0 \tag{31}$

ZR: $\qquad \mu > 1 + \theta, \ \theta > 0$ $\qquad\qquad\qquad\qquad$ (32)

The restrictions in (30) are designed to make $f(x_{it}, \beta)$ an increasing function of all inputs and an increasing function of technical change. As mentioned earlier, the restrictions in (31) and (32) ensure U-shaped average cost curves with a minimum average cost.

To summarize the prior distributions on β, μ, γ and θ, and in preparation for a description of the MCMC algorithms used for estimation, it is convenient to introduce some new notation and to vary slightly some existing notation. Let $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_5)'$. That is, β contains all the parameters defined in the specification of $\log f(x_{it}, \beta)$ in equation (15), except for β_4 and μ. Also, let $x_{it}\beta$ equal the right-hand side of equation (15) after subtracting from it $\mu \log x_{4it}$. That is,

$$ x_{it}\beta = \log f(x_{it}, \beta) - \mu \log x_{4it} \qquad\qquad (33) $$

Let $\phi_{NR} = (\mu, \gamma)'$ and $\phi_{ZR} = (\mu, \theta)'$. Also, define $\delta_{NR} = (\beta', \phi_{NR})'$ and $\delta_{ZR} = (\beta', \phi_{ZR})'$. Defining δ_{NR} and δ_{ZR} is useful for describing the MCMC algorithm for the homoskedastic models; ϕ_{NR} and ϕ_{ZR} are useful for describing the algorithm for the heteroskedastic models. When δ or ϕ appears without a subscript, the equation where it appears will refer to both the ZR and NR models. Specifically, a generic version of the homoskedastic models in equations (17) and (19) will be written as

$$ \log y_{it} = g(x_{it}, \delta) - z_i + u_{it} \qquad\qquad (34) $$

A generic version of the heteroskedastic models in equations (16) and (18) is

$$ h(y_{it}, \phi) = x_{it}\beta - z_i + u_{it} \qquad\qquad (35) $$

with $h(y_{it}, \phi_{NR}) = \log y_{it} + \gamma(\log y_{it})^2 - \mu \log x_{4it}$ and $h(y_{it}, \phi_{ZR}) = \log y_{it} + \theta y_{it} - \mu \log x_{4it}$.

The prior densities for ϕ_{NR} and ϕ_{ZR} can be summarized as

$$ p(\phi_{NR}) \propto \frac{\mu - 1}{\gamma^2 \mu^{c_r + 1}} \exp\left\{ -\frac{c_q(\mu - 1)}{2\gamma} \right\} I(\phi_{NR}) \qquad\qquad (36) $$

$$p(\phi_{ZR}) \propto [\mu - (1+\theta)] \frac{(1+\theta)^{c_r-1} \theta^{c_q-1}}{\mu^{c_r+1}(\mu-1)^{c_q+1}} I(\phi_{ZR}) \tag{37}$$

where $I(\phi_{NR})$ and $I(\phi_{ZR})$ are indicator functions, equal to unity when the inequality restrictions in equations (31) and (32), respectively, hold, and zero otherwise. The density $p(\phi)$ will be used as notation for describing either (36) or (37). The prior densities for δ_{NR} and δ_{ZR} can be summarized as

$$p(\delta) \propto p(\phi) I(\beta \mid \phi) \tag{38}$$

where $I(\beta \mid \phi)$ is an indicator function equal to unity when the inequality restrictions in equation (30) hold. The indicator is expressed as conditional on ϕ because the restriction that β_4 be positive is expressed as $\beta_1 + \beta_2 + \beta_3 < \mu$, making it conditional on μ.

4. Estimation

Given the above models and prior densities, it is now possible to describe conditional posterior densities for the parameters and the inefficiency terms. Doing so enables us to set up Gibbs samplers for drawing observations from the joint posterior densities for all unobservables. These drawings can be used in turn to make inferences about unknown quantities of interest. The conditional posterior densities are slightly modified versions of those given in Koop and Steel (2001), adapted to allow for the nonlinear functions g and h, the priors $p(\delta)$ and $p(\phi)$, and the possibly different numbers of observations on each firm (region). In what follows $N = 10$ is the number of firms and $M = \sum_{i=1}^{N} T_i$ is the total number of observations. Also, following Koop and Steel, we use $f_G[w \mid b, c]$ to denote a gamma density with slope parameter b and scale parameter c. That is, $f_G(w \mid b, c) \propto w^{b-1} \exp(-cw)$. We use the notation y, x and z to denote all observations on output, the inputs and the inefficiencies, respectively.

The conditional posterior density for λ^{-1} is the same for both the heteroskedastic and homoskedastic models; it is given by the gamma distribution

$$p(\lambda^{-1} \mid y, x, z, \delta, \omega) = f_G\left[\lambda^{-1} \mid (N+1), \left(\sum_{i=1}^{N} z_i - \ln(\tau*)\right)\right] \tag{39}$$

For the remaining conditional posterior densities we first describe those for the homoskedastic models represented by equation (34), and then point out how they need to be modified for the heteroskedastic models represented by equation (35).

Model: $\log y_{it} = g(x_{it}, \delta) - z_i + u_{it}$

The conditional posterior density for the i-th inefficiency term z_i is a normal distribution truncated from below at zero with parameters μ_{zi} and σ_{zi}^2 given by

$$\mu_{zi} = \frac{1}{T_i}\sum_{t=1}^{T_i}(g(x_{it}, \delta) - \log y_{it}) - \frac{1}{T_i \omega \lambda} \qquad \qquad \sigma_{zi}^2 = \frac{1}{\omega T_i} \tag{40}$$

The precision parameter ω has the following gamma distribution as its conditional posterior density

$$p(\omega \mid y, x, z, \delta, \lambda) = f_G\left[\omega \left|\left(\frac{M}{2}\right), \left(\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(\log y_{it} - g(x_{it}, \delta) + z_i\right)^2\right)\right.\right] \tag{41}$$

Finally, the conditional posterior density function for δ is given by

$$p(\delta \mid y, x, z, \omega, \lambda) \propto \exp\left\{-\frac{\omega}{2}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(\log y_{it} - g(x_{it}, \delta) + z_i\right)^2\right\} p(\delta) \tag{42}$$

A Metropolis step was used to draw from this density, since it is not of a recognizable form. More details are given after we describe the conditional posterior densities for the heteroskedastic models.

Model: $h(y_{it}, \phi) = x_{it}\beta - z_i + u_{it}$

The conditional posterior density for the i-th inefficiency term z_i is again a normal distribution truncated from below at zero, with the same scale parameter $\sigma_{zi}^2 = 1/\omega T_i$, but the location parameter becomes

$$\mu_{zi} = \frac{1}{T_i} \sum_{t=1}^{T_i} \left(x_{it}\beta - h(y_{it},\phi) \right) - \frac{1}{T_i\omega\lambda} \tag{43}$$

The conditional posterior density for ω is the gamma distribution

$$p(\omega|\ y,\ x,\ z,\delta,\lambda) = f_G \left[\omega \left| \left(\frac{M}{2}\right),\ \left(\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(h(y_{it},\phi) - x_{it}\beta + z_i\right)^2\right) \right. \right] \tag{44}$$

For this model it is convenient to replace the conditional posterior density for δ with two conditional densities, one for β and one for ϕ. The conditional posterior density for β is a truncated multivariate normal distribution with truncations given by $I(\beta|\phi)$, and parameters

$$\overline{\beta} = (X'X)^{-1}X'(h+z) \qquad\qquad V_\beta = \omega^{-1}(X'X)^{-1} \tag{45}$$

where X is an $(M \times 5)$ matrix containing the observations x_{it}, and h and z are M – dimensional vectors containing $h(y_{it},\phi)$ and values z_i (suitably repeated), respectively. The conditional posterior density for ϕ is given by

$$p(\phi|\ y, x, z,\beta,\omega,\lambda) \propto |\ J\ |\exp\left\{ -\frac{\omega}{2}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(h(y_{it},\phi) - x_{it}\beta + z_i\right)^2 \right\} p(\phi) \tag{46}$$

where $|\ J\ |$ is the Jacobian of the transformation from u_{it} to y_{it}. For each of the models we have

$$|\ J_{NR}\ | = \prod_{i=1}^{N}\prod_{t=1}^{T_i}(1+2\gamma\log y_{it}) \qquad\qquad |\ J_{ZR}\ | = \prod_{i=1}^{N}\prod_{t=1}^{T_i}(1+\theta y_{it}) \tag{47}$$

Like equation (42), the density in equation (46) is not of a recognizable form; a Metropolis step was used to draw observations on ϕ from it.

Implementing the Gibbs Samplers

We now provide more details of how the Gibbs samplers were implemented and some of the difficulties that were encountered along the way. Each of the four models is considered in turn. For the NR(hom) model, starting values for δ, the z_i and ω were obtained by maximizing the normal likelihood function from equation (17) with the z_i

treated as fixed effects. To do so it was necessary to set β_0 a priori; different values of β_0 led to the same value for the maximized log-likelihood function. Using trial and error, we found an approximate setting for β_0 that was as small as possible, without any of the maximum likelihood estimates for the z_i being negative. Starting the MCMC algorithm at larger values of β_0 led to a period of 1000 to 2000 draws where the chain had clearly not converged. After approximately 2000 draws it settled down, yielding draws centered around a value similar to our ultimate choice for a starting value. Draws were made for $\lambda^{-1}, z_i, (i = 1, 2, ..., N)$ and ω from the density functions described in equations (39) to (41), with the values of the conditioning variables being their starting values or their most recent draws. To draw the z_i from their truncated normal distributions, we used the inverse distribution function method described, for example, by Albert and Chib (1996). For drawing δ from its conditional posterior density given in equation (42), we used an independence Metropolis step. At the j-th iteration a candidate value δ^* is drawn from a candidate generating density $q(\delta \mid y, x, z)$. The ratio

$$\alpha = \frac{p(\delta^* \mid y, x, z, \omega, \lambda) \, q(\delta_{(j-1)} \mid y, x, z)}{p(\delta_{(j-1)} \mid y, x, z, \omega, \lambda) \, q(\delta^* \mid y, x, z)} \tag{48}$$

is computed and δ^* is accepted as a draw $(\delta_{(j)} = \delta^*)$ if $\alpha \geq 1$; it is accepted as a draw with probability α if $\alpha < 1$. If δ^* is rejected, the draw from the previous iteration is taken as the current draw $(\delta_{(j)} = \delta_{(j-1)})$. For $q(\delta \mid y, x, z)$ we used a truncated multivariate normal distribution, truncated according to the inequality restrictions in $p(\delta)$, and with parameters $\overline{\delta}$ and $1.4\,\Sigma_\delta$, where $\overline{\delta}$ and Σ_δ were obtained from maximum likelihood estimation of the model

$$(\log y_{it} + z_i) = g(x_{it}, \delta) + u_{it} \tag{48}$$

The truncations in $p(\delta)$ turned out to be relatively mild and so a simple acceptance-rejection algorithm was satisfactory for drawing from $q(\delta \mid y, x, z)$. Note that additional normalizing constants caused by the truncations cancel in the ratio in (48). Also, there is

no need to set an a priori value for β_0 when we are conditioning on the z_i. Computing a value of the ratio α given in equation (48) requires evaluation of $p(\delta \mid y, x, z, \omega, \lambda)$ at δ^* and $\delta_{(j-1)}$. Increasing the spread of the candidate generating density to $1.8\Sigma_\delta$ did not lead to any marked difference in the results. For this and the other models, the proper priors were given parameter settings to make them relatively noninformative. Specifically, we set $c_r = 3.6$, $c_q = 0.5$ and $\tau^* = 0.8$. The Gibbs' sampler was used to generate 45000 observations, of which 5000 were discarded for a burn-in. Graphs of the generated observations did not suggest nonstationarity.

In the first attempt at estimating the ZR(hom) model, we followed a similar procedure to that just described for the NR(hom) model. However, because an analytical expression for g was not available, to specify likelihood functions for getting starting values, and for getting values $\overline{\delta}$ and Σ_δ for the candidate generating density, we used second-order Taylor series expansions around the mean of $\log y_{it}$ to approximate the equations $\log y_{it} = g(x_{it}, \delta)$ and $(\log y_{it} + z_i) = g(x_{it}, \delta)$. However, for drawing from the conditional densities in equations (40) to (42), and for evaluating the acceptance ratio in equation (48), $g(x_{it}, \delta)$ was obtained numerically as the solution for $\log y_{it}$ from the equation $\log y_{it} + \theta\, y_{it} = \log f(x_{it}, \beta)$. This solution must be evaluated for every observation in the data set, and for all parameter draws and proposed parameter draws. Such a task is time consuming, but nevertheless practical. Our first attempt at estimating the ZR(hom) model was judged to be unsuccessful because the Metropolis step got stuck on three occasions, leading to excessive repetitions of some values of δ. Believing the problem might be the large concentration of values for θ near the zero boundary (the posterior density for θ has a mode at zero), the model was reparameterized in terms of $\alpha = \theta^{0.3}$. The posterior density for α was approximately symmetric. This strategy did help, but there was still evidence of the algorithm 'sticking'. Changing the Metropolis step from an independence step to a random walk proposal for δ^* overcame the 'sticking' problem, but led to some (possibly excessive) drifting. We finally settled on a mixture of a random walk step and an independence step, weighted 0.75 and 0.25, respectively. The use of mixtures for generating proposals has been suggested by Geweke (1999). The

covariance matrices for the mixture components, and the mean for the independence component, were obtained from our earlier experiences and from experimentation. The reparameterization $\alpha = \theta^{0.3}$ was retained.

The algorithms used for the ZR(het) and NR(het) models were similar to those for the NR(hom) model. Starting values of the parameters were obtained from maximum likelihood estimation of equations (16) and (18), with the z_i treated as fixed effects and the value for β_0 set a priori. Drawing from the conditional posterior densities in equations (43) to (45) was relatively straightforward. An independence Metropolis step was used to generate draws on ϕ from the density in equation (46). The proposal density for this step was a normal distribution with mean $\bar{\phi}$ and covariance matrix $1.4\Sigma_\phi$ where $\bar{\phi}$ and Σ_ϕ were the maximum likelihood estimates and covariance matrix from estimating $h(y_{it}, \phi) = x_{it}\beta - z_i + u_{it}$, with β and z_i set equal to their conditioning values. To achieve approximate symmetry, the ZR(het) model was reparameterized in terms of $\alpha = \theta^{0.5}$.

5. The Application

The Economic Research Service (ERS) of the USDA compiles annual indexes of output, input use, and total factor productivity for the aggregate farm sector and for the individual states. A discussion of the methods and data used to construct the indexes, and some insights into farm production, can be found in Ball and Nehring (1998) and Ball et al (1999). We use state-level indexes for total output and for the inputs materials, capital, land and labor for the period 1971-1996. The 48 states (Alaska and Hawaii are excluded) are classified into ten farm production regions. In the context of our study, these ten regions were treated as firms whose efficiency is being assessed. With 48 states and 26 time periods, we have a total of 1248 observations. The allocation of these observations to the regions, and the states that make up each region are given in Table 1.

For results, we have chosen to report summary statistics from the posterior densities of all parameters and to illustrate the flexibility of the Bayesian approach by graphing posterior densities for a number of quantities of interest. Being able to draw

finite-sample inferences on various functions of parameters, such as firm efficiencies, relative efficiency, efficiency rankings, and the level of output for which average cost is a minimum, is an impelling advantage of Bayesian inference.

Table 2 contains the MCMC-estimated posterior means and standard deviations for all parameters as well as those for the RTS when $y = 1$, $\log y$ for which $RTS = 1$ (labelled as $\log y^*$), the standard deviation of u_{it} (denoted by $\omega^{-1/2}$), the standard deviation of inefficiency errors z_i (denoted by λ), and the proportion of the total variance of the error term that is attributable to the inefficiency errors, $\lambda^2/(\lambda^2 + \omega^{-1})$. Examining first the results for the β_k, we find that both models yield seemingly similar estimates, with the elasticity of $f(x,\beta)$ with respect to materials being much greater than those from the other inputs. Graphs of three of the complete posterior densities for β_1 and β_2 (see Figures 1 and 2) confirm that the models yield similar results for β_2, but show distinct differences for β_1. The densities for β_1 from the ZR models are more precise and located further to the left than their NR counterparts. In this and other figures three rather than four densities were graphed to avoid making the figures too congested; the omitted densities (usually ZR(het)) have no special characteristics. Overall, one can conclude that estimation of the β_k is sensitive to choice of model and stochastic assumptions, but the sensitivity is not enormous.

For estimation of μ, the homogeneity of the function $f(x,\beta)$, the ZR models suggests a smaller value, and more precise information about this value, relative to the results from the NR models (see Figure 3). Also, the heteroskedastic assumptions lead to larger less precise values than the corresponding homoskedastic assumptions. The posterior densities for the parameters in the RTS function, γ and θ, appear in Figures 4 and 5, respectively. These densities are very sensitive to choice of stochastic assumption. The posterior mean for γ from the NR(het) model is approximately double that for the NR(hom) model. The difference in posterior means for θ in the ZR models is over four times. The density for θ from the ZR(hom) model has a sharp mode at zero, while that

for the ZR(het) model shows some evidence of being bi-modal, with a mode at zero and an interior mode.

Figure 6 contains posterior densities for $\log y^*$, the logarithm of output at which the RTS function is equal to unity. The estimates for this quantity are quite sensitive to the stochastic assumption, but not sensitive to a choice between the ZR and NR models. Also, all the posterior densities have long tails to the right. The posterior probability of $\log y^*$ exceeding the maximum sample value of 6.7655 is 0.89 for both homoskedastic models, and 0.30 for both heteroskedastic models. The final observation that we make from Table 2 is the lack of sensitivity of estimation of the proportion of variance attributable to firm inefficiency; the posterior means and standard deviations are similar for both models and both stochastic assumptions.

Posterior information on the regional efficiencies appears in Table 3 (posterior medians and 90% probability intervals), in Tables 4 and 5 (posterior distributions on efficiency rankings), and in Figures 7, 8 and 9. For the homoskedastic models the efficiency measure is $\exp(-z_i)$. For the heteroskedastic models, posterior results for two measures are given; the cost function one is taken as $\exp(-z_i/\mu)$ and the output function one is that in equation (10), and its NR counterpart, averaged over all observations on a given region. Although the different measures do lead to different results, overall conclusions about the relative efficiencies of the different regions is remarkably insensitive to choice of model and choice of efficiency measure. The regions can be loosely placed in one of three groups, efficient regions (Northeast, Southeast, Pacific), inefficient regions (Delta States, Southern Plains) and moderately efficient regions (Lake States, Cornbelt, Northern Plains, Appalachia, Mountains). The robustness of the results with respect to model choice and efficiency measure is illustrated in Figures 7 and 8. These figures contain posterior densities of the efficiencies for one region from each group, namely, Appalachia (5), Southeast (6) and Delta States (7). For each region, all four densities have similar locations and shape.

From Table 3 we observe that the lowest median efficiency from all regions is 0.698 and the highest is 0.986. Values drawn from the densities ranged from 0.55 to 1.00.

The probability distributions for the efficiency rankings that appear in Tables 4 and 5 can be read in two ways. Each row represents a region and gives the posterior distribution for the rank of that region. For example, using the NR(hom) model in Table 3, the probability that region 2 is ranked 5-th is 0.181. Each column represents a rank and gives the posterior distribution for what region holds that rank. For example, again using the NR(hom) model, the probability that the most efficient region is region 1 is 0.010; the probability that the second most efficient region is region 6 is 0.791. The efficiency rankings suggest the four most efficient regions are 10, 6, 1 and 2 in that order, and the two least efficient regions are 7 and 8. The rankings of the remaining intermediate regions are more ambiguous; the probability distributions of these rankings have a greater spread and are more sensitive to choice of model and inefficiency measure. A similar conclusion is reached if the posterior medians in Table 3 are used to rank regions. The ease with which Bayesian inference can be used to provide posterior distributions for firm rankings, and probability statements about the relative efficiency of different firms, has also been noted by Atkinson and Dorfman (2001).

Suppose that the efficiency of one region relative to another is of interest. In this case we can get the posterior density of, for example, τ_3 / τ_4. This density is graphed in Figure 9 for three different efficiency measures. Most of the region covered by the densities is greater than one, suggesting that region 3 is likely to be more efficient than region 4. Estimating this probability from the MCMC observations, we obtain values ranging from to 0.92 to 0.98, depending on the chosen model and efficiency measure.

6. Concluding Remarks

We have demonstrated how Bayesian estimation can be used to make finite-sample inferences about parameters and firm efficiencies in a stochastic production frontier with a returns-to-scale function that depends on output. The stochastic errors in earlier studies of this kind were introduced in a way that facilitates estimation, but no longer retains the same inefficiency interpretation of the one-sided error. We show how the traditional inefficiency interpretation can be retained by appending stochastic errors *after* solving the deterministic production relationship for the logarithm of output. For the two models we considered, one analytical solution for $\log y$ was obtainable; the other solution had to be

obtained numerically. Solving the equation numerically does increase the computational burden, but it can be set up straightforwardly within a Markov chain Monte Carlo algorithm typically used for Bayesian estimation of stochastic frontiers. Bayesian inference is attractive because of the wide variety of quantities for which posterior information is readily attainable. In our application to state-level data on farm output and inputs, we obtain posterior densities on a variety of parameters, on regional efficiencies, on the output for which average cost is a minimum, on rankings of regional efficiencies, and on the relative efficiency of two regions. We find the efficiency rankings and production elasticities are generally not sensitive to the choice of model, but the homogeneity of the function and the output for which average cost is a minimum are heavily influenced by model choice. Our study is limited by the choice of only two returns to scale functions and because we did not consider functions other than the Cobb-Douglas for the homogeneous part of the function. Future work will consider and compare other combinations of functional forms.

References

Albert, J.H., Chib, S., 1996. Computation in Bayesian econometrics: An introduction to Markov chain Monte Carlo. In Hill, R.C. (Ed.), Advances in Econometrics Volume 11A: Computational Methods and Applications. Greenwich, JAI Press, pp.3-24.

Atkinson, S.E., Dorfman, J.H., 2001. Multiple comparisons with the best: Bayesian precision measures of efficiency rankings. Unpublished paper, University of Georgia, Athens.

Ball, V., Nehring, R., 1998. Patterns of state productivity growth in the U.S. farm sector. ERS Staff Paper, USDA, Washington.

Ball, V.E., Gollop, F., Hawke, A., Swinand, G., 1999. Patterns of state productivity growth in the U.S. farm sector: Linking state and aggregate models. American Journal of Agricultural Economics 81, 164-179.

Calem, P., 1990. On estimating technical progress and returns to scale. In: Sato, R. (Ed.), Conservation Laws and Symmetry. Boston, Kluwer, pp.

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

Geweke, J., 1999. Using simulation methods for Bayesian econometric models: inference, development and communication. Econometric Reviews 18, 1-74.

Greene, W.H., 1997. Frontier production functions. In: Pesaran, M.H., Schmidt, P. (Eds.), Handbook of Applied Econometrics, Volume II: Microeconometrics. Malden, Blackwell, pp.81-166.

Koop, G., Steel, M.F.J., 2001. Bayesian analysis of stochastic frontier models. In Baltagi, B.H. (Ed.), A Companion to Theoretical Econometrics. Oxford, Blackwell, pp.520-537.

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., 1988. On the estimation of technical and allocative inefficiency using stochastic frontier functions: The case of U.S. class 1 railroads. International Economic Review 29, 727-744.

Kumbhakar, S., Ghosh, S., McGuckin, J., 1991. A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms. Journal of Business and Economic Statistics 9, 279-286.

Nerlove, M., 1963. Returns to scale in electricity supply. In: Christ, C. et al (Eds.), Measurement in Econometrics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld. Stanford, Stanford University Press.

Revankar, N., 1971. A class of variable elasticity of substitution production functions. Econometrica 39, 61-71.

Ringstad, V., 1967. Econometric analyses based on a production function with neutrally variable scale elasticity. Swedish Journal of Economics, 115-133.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. New York, John Wiley.

Zellner, A., Revankar, N., 1969. Generalized production functions. Review of Economic Studies 36, 241-250.

Zellner, A., Ryu, H., 1998. Alternative functional forms for production, cost and returns to scale functions. Journal of Applied Econometrics 13, 101-127.

Table 1.   USDA ERS Production Regions

	Region	States	Observations
1	Northeast	Connecticut, Delaware, Massachusetts, Maryland, Maine, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont	286
2	Lake States	Michigan, Minnesota, Wisconsin	78
3	Cornbelt	Iowa, Illinois, Indiana, Missouri, Ohio	130
4	Northern Plains	Kansas, Nebraska, North Dakota, South Dakota	104
5	Appalachia	Kentucky, North Carolina, Tennessee, Virginia, West Virginia	130
6	Southeast	Alabama, Florida, Georgia, South Carolina	104
7	Delta States	Arkansas, Louisiana, Mississippi	78
8	Southern Plains	Oklahoma, Texas	52
9	Mountains	Arizona, Colorado, Idaho, Montana, New Mexico, Nevada, Utah, Wyoming	208
10	Pacific	California, Oregon, Washington	78

Table 2.   Posterior Means and Standard Deviations of Parameters

	$\log(y)$ homoskedastic		$\log(y)$ heteroskedastic	
	NR	ZR	NR	ZR
β_0	-0.2926	-0.2265	-0.3427	-0.2524
	(0.0592)	(0.0523)	(0.0677)	(0.0571)
β_1	0.7935	0.7551	0.8251	0.7633
	(0.0236)	(0.0175)	(0.0270)	(0.0187)
β_2	0.0714	0.0778	0.0592	0.0666
	(0.0278)	(0.0262)	(0.0274)	(0.0267)
β_3	0.0888	0.0858	0.0976	0.0938
	(0.0146)	(0.0140)	(0.0151)	(0.0146)
β_4	0.1430	0.1366	0.1489	0.1397
	(0.0249)	(0.0227)	(0.0249)	(0.0235)
β_5	0.1138	0.1087	0.1180	0.1107
	(0.0068)	(0.0062)	(0.0072)	(0.0064)
μ	1.0967	1.0553	1.1308	1.0635
	(0.0182)	(0.0071)	(0.0223)	(0.0089)
$\theta\ (\times 10^3)$		0.0272		0.1253
		(0.0293)		(0.0642)
γ	0.00585		0.01040	
	(0.00211)		(0.00277)	
RTS(1)	1.0967	1.0552	1.1308	1.0634
	(0.0182)	(0.0071)	(0.0223)	(0.0088)
$\log y*$[1]	8.886	8.568	6.4935	6.4747
	(2.075)	(1.963)	(0.9416)	(0.9824)
$\omega^{-1/2}$	0.1594	0.1594	0.1732	0.1620
	(.0031)	(.0031)	(.0052)	(.0035)
λ	0.2190	0.2260	0.2396	0.2274
	(0.0788)	(0.0827)	(0.0890)	(0.0831)
$\lambda^2/(\lambda^2+\omega^{-1})$	0.615	0.628	0.617	0.624
	(0.143)	(0.144)	(0.145)	(0.143)

[1] $y*$ is the output level where $RTS = 1$ and average cost is a minimum.

Table 3. Medians and 90% Probability Intervals for Regional Efficiencies

| | log(y) homoskedastic | | log(y) heteroskedastic | | | |
| | Output function | | Cost function | | Output function | |
region	NR	ZR	NR	ZR	NR	ZR
1	.924	.923	.929	.935	.925	.931
	(.869, .974)	(.862, .975)	(.865, .976)	(.868, .981)	(.858, .975)	(.862, .980)
2	.821	.809	.832	.825	.830	.820
	(.770, .867)	(.751, .858)	(.776, .877)	(.767, .868)	(.774, .876)	(.762, .865)
3	.804	.790	.817	.808	.816	.804
	(.753, .848)	(.732, .837)	(.762, .860)	(.751, .849)	(.760, .859)	(.746, .847)
4	.775	.766	.782	.778	.780	.773
	(.728, .815)	(.709, .808)	(.731, .822)	(.725, .816)	(.728, .820)	(.719, .811)
5	.799	.798	.805	.807	.800	.799
	(.754, .838)	(.742, .841)	(.754, .843)	(.753, .843)	(.749, .838)	(.745, .836)
6	.950	.950	.946	.950	.945	.948
	(.895, .990)	(.884, .992)	(.886, .987)	(.887, .989)	(.883, .986)	(.883, .989)
7	.732	.732	.736	.741	.730	.731
	(.688, .771)	(.679, .773)	(.686, .774)	(.689, .779)	(.680, .768)	(.679, .770)
8	.702	.691	.709	.703	.706	.699
	(.658, .742)	(.638, .733)	(.662, .748)	(.656, .741)	(.658, .746)	(.651, .738)
9	.789	.794	.788	.797	.780	.788
	(.745, .827)	(.736, .832)	(.738, .824)	(.744, .833)	(.730, .817)	(.734, .824)
10	.986	.983	.986	.986	.986	.986
	(.941, .999)	(.922, .999)	(.935, .999)	(.934, .999)	(.934, .999)	(.932, .999)

Table 4. Posterior Distributions for the Efficiency Ranking of Each Region: NR Model[1]

region	Rank									
	1	2	3	4	5	6	7	8	9	10
1 (hom) [2]	.010	.139	.851							
1 (het-c)	.013	.221	.766							
1 (het-o)	.012	.194	.794							
2 (hom)				.703	.181	.074	.034	.008		
2 (het-c)				.736	.197	.051	.013	.003		
2 (het-o)				.738	.205	.044	.010	.002		
3 (hom)				.160	.413	.252	.147	.028		
3 (het-c)				.203	.544	.197	.049	.008		
3 (het-o)				.215	.578	.168	.032	.007		
4 (hom)				.002	.014	.063	.197	.710	.013	
4 (het-c)				.001	.011	.086	.323	.574	.005	
4 (het-o)				.001	.013	.118	.396	.469	.004	
5 (hom)				.097	.282	.388	.176	.057		
5 (het-c)				.052	.216	.548	.142	.043		
5 (het-o)				.039	.183	.580	.155	.043		
6 (hom)	.066	.791	.143							
6 (het-c)	.040	.731	.228							
6 (het-o)	.039	.760	.201							
7 (hom)								.013	.894	.093
7 (het-c)								.006	.879	.114
7 (het-o)								.005	.839	.156
8 (hom)								.001	.092	.907
8 (het-c)									.114	.886
8 (het-o)									.155	.844
9 (hom)				.038	.110	.223	.445	.184	.001	
9 (het-c)				.008	.032	.118	.474	.366	.001	
9 (het-o)				.006	.022	.090	.407	.473	.002	
10 (hom)	.924	.070	.006							
10 (het-c)	.946	.048	.006							
10 (het-o)	.949	.046	.005							

[1] Blank table entries correspond to probabilities less than 0.0005.

[2] "hom" refers to the model with homoskedastic $\log y$; "het-c" refers to the cost function measure of efficiency in the model with heteroskedastic $\log y$; "het-o" refers to the output function measure of efficiency in the model with heteroskedastic $\log y$.
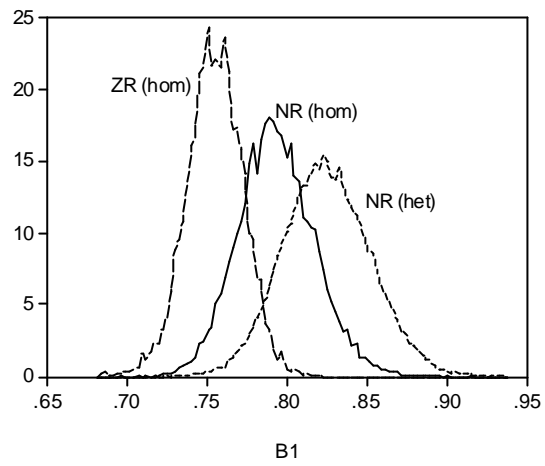
Table 5. Posterior Distributions for the Efficiency Ranking of Each Region: ZR Model[1]

region	Rank									
	1	2	3	4	5	6	7	8	9	10
1 (hom) [2]	.015	.124	.861							
1 (het-c)	.023	.248	.730							
1 (het-o)	.021	.239	.740							
2 (hom)				.597	.197	.129	.066	.011		
2 (het-c)				.688	.181	.085	.038	.008		
2 (het-o)				.699	.189	.074	.032	.007		
3 (hom)				.075	.207	.269	.381	.068	.001	
3 (het-c)				.136	.366	.270	.197	.031		
3 (het-o)				.159	.425	.239	.151	.026		
4 (hom)				.001	.008	.028	.112	.815	.035	
4 (het-c)				.001	.008	.039	.148	.783	.021	
4 (het-o)				.002	.011	.059	.199	.715	.015	
5 (hom)				.192	.362	.279	.141	.025		
5 (het-c)				.118	.314	.371	.161	.036		
5 (het-o)				.097	.272	.417	.167	.047		
6 (hom)	.092	.780	.128							
6 (het-c)	.058	.682	.260							
6 (het-o)	.056	.693	.250							
7 (hom)							.001	.034	.930	.035
7 (het-c)							.001	.021	.932	.046
7 (het-o)							.001	.014	.893	.092
8 (hom)									.035	.965
8 (het-c)									.046	.954
8 (het-o)									.092	.908
9 (hom)				.134	.226	.295	.299	.046		
9 (het-c)				.057	.130	.235	.456	.122		
9 (het-o)				.044	.103	.211	.450	.191	.001	
10 (hom)	.893	.096	.011							
10 (het-c)	.920	.070	.010							
10 (het-o)	.923	.068	.009							

[1] Blank table entries correspond to probabilities less than 0.0005.

[2] "hom" refers to the model with homoskedastic log y ; "het-c" refers to the cost function measure of efficiency in the model with heteroskedastic log y ; "het-o" refers to the output function measure of efficiency in the model with heteroskedastic log y .
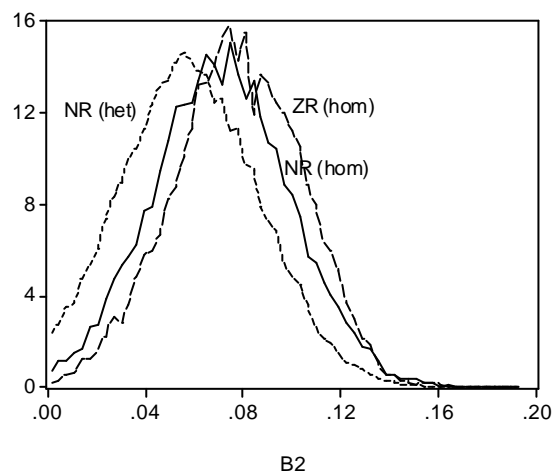
Figure 1.   Posterior densities for β_1
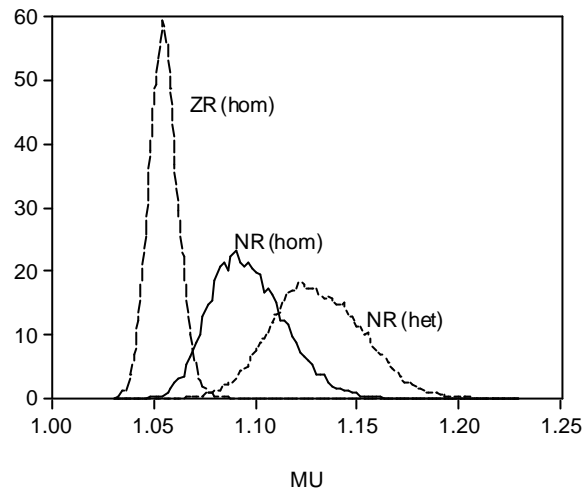

Figure 2.   Posterior densities for β_2

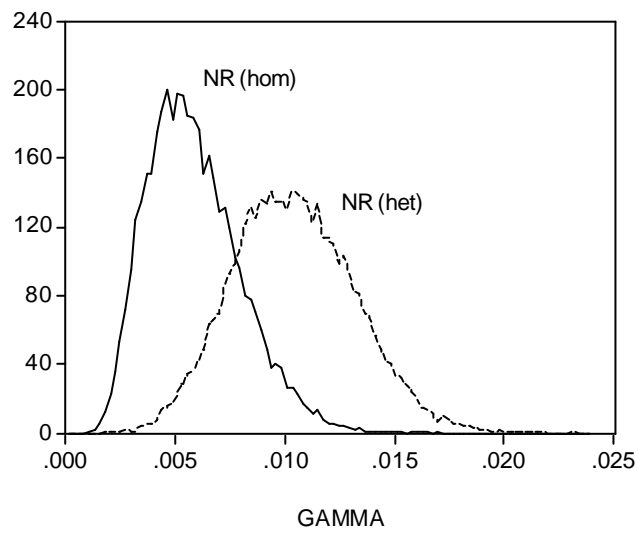Figure 3.   Posterior densities for μ

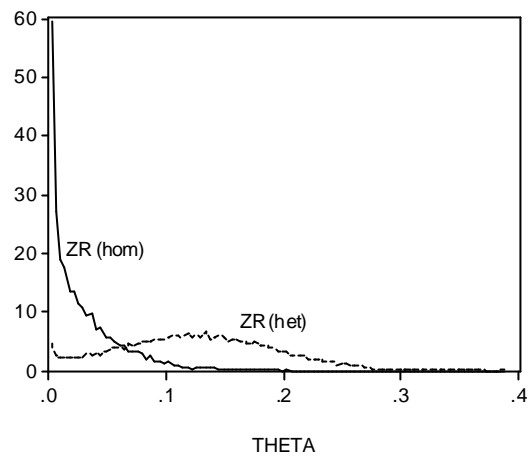
Figure 4.   Posterior densities for γ

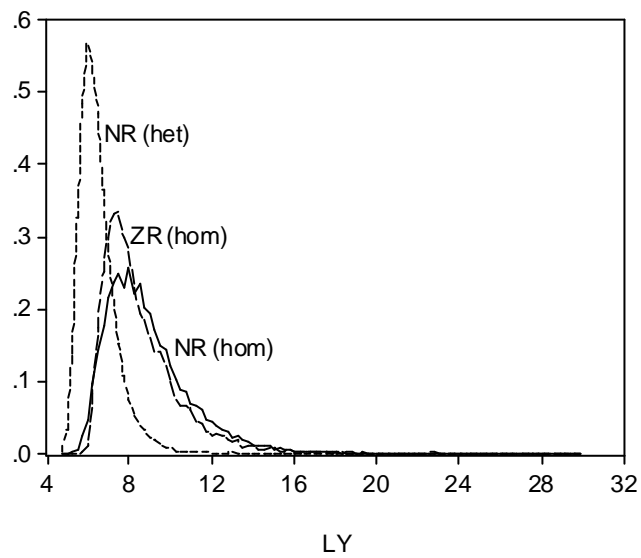Figure 5. Posterior densities for θ $(\times 10^3)$.

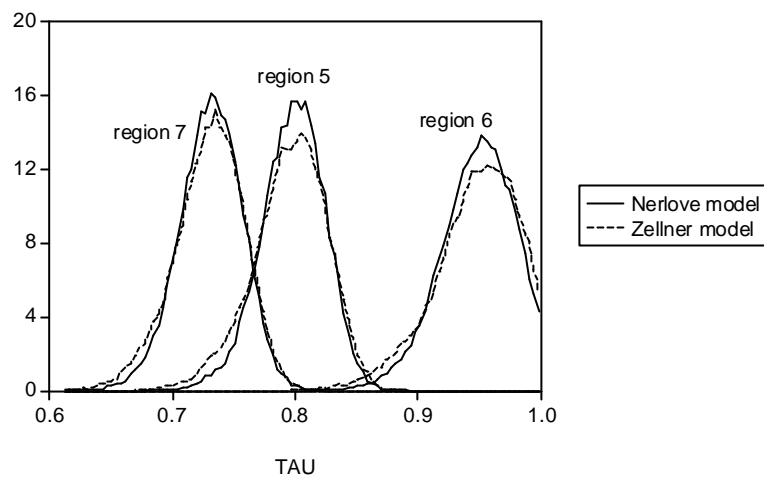

Figure 6. Posterior densities for $\log y^*$

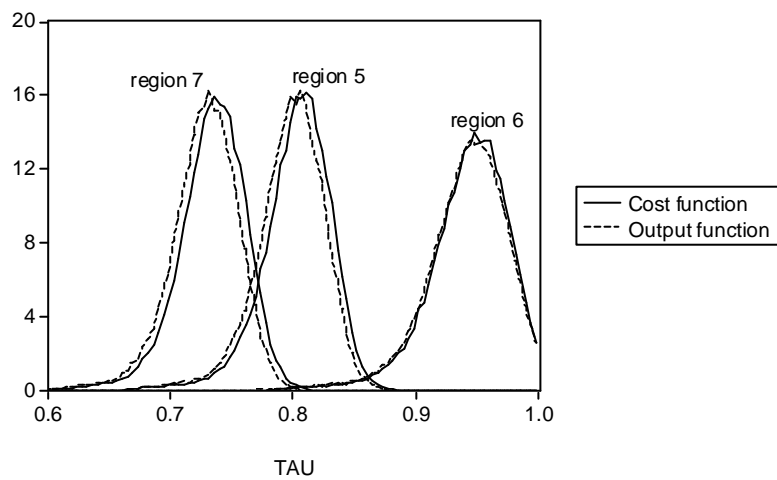Figure 7.   Posterior densities for selected efficiencies with log *y* homoskedastic

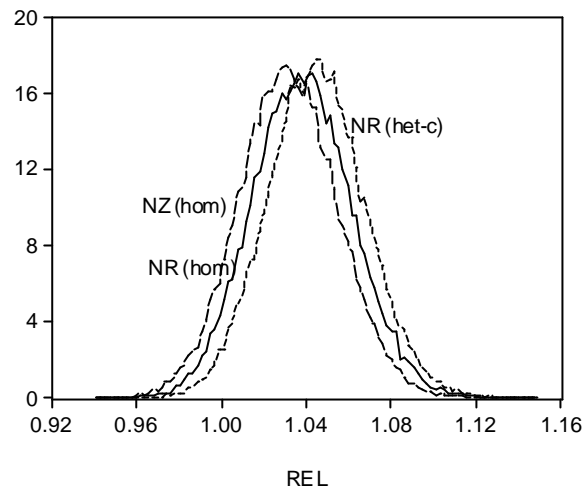Figure 8.   Posterior densities for selected efficiencies from model NR with log *y* heteroskedastic

Figure 9.   Posterior densities for the efficiency of region 3 relative to region 4.