MEASURING TECHNICAL AND ALLOCATIVE INEFFICIENCY
IN THE TRANSLOG COST SYSTEM: A BAYESIAN APPROACH*

Subal C. Kumbhakar
Department of Economics
State University of New York
Binghamton, NY 13902, USA.
Phone: (607) 777 4762, Fax: (607) 777 2681
E-mail: kkar@binghamton.edu

and

Efthymios G. Tsionas
Department of Economics
Athens University of Economics and Business
76 Patission Street, 104 34 Athens, Greece.
Phone: (301) 0820 3388, Fax: (301) 0820 3310
E-mail: tsionas@aueb.gr

Abstract

In this paper we propose simulation based Bayesian inference procedures in a cost system that includes the cost function and the cost share equations augmented to accommodate technical and allocative inefficiency. Markov Chain Monte Carlo techniques are proposed and implemented for Bayesian inferences on costs of technical and allocative inefficiency, input price distortions and over- (under-) use of inputs. We show how to estimate a well-specified translog system (in which the error terms in the cost and cost-share equations are internally consistent) in a random effects framework. The new methods are illustrated using panel data on U.S. commercial banks.

1. Introduction

Empirical estimation of efficiency in the stochastic frontier (SF) models (developed by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977)) involve estimation of a parametric production/cost/profit function with a composed error term consisting of a two-sided disturbance term that reflects exogenous shocks and a one-sided term that captures technical inefficiency.[1] Although the theory is well developed to estimate a system of equations either in the form of factor demand or cost function and cost share equations, the system approach is rarely applied in the efficiency literature.[2] The reason is that the error structure comprising noise, technical and allocative inefficiency complicates econometric estimation of the model. This is especially the case when one uses flexible functional forms to represent the underlying technology.[3] Joint estimation of technical and allocative inefficiency in a translog cost function presents a difficult problem (Greene (1980)).[4] The difficulty is that the cost function and the deviations of optimal shares from observed shares are complicated functions of allocative inefficiency. Although many attempts have been proposed, none have been entirely successful. Recently Kumbhakar (1997) proposed a solution for the Greene problem using a translog cost system, but empirical estimation of this model has been restricted to panel data models in which technical and allocative inefficiency are either assumed to be fixed parameters or functions of the data and unknown parameters (Atkinson and Cornwell (1994); Maietta (2000)). In this paper, we show that relatively simple econometric tools can be used to estimate technical and allocative inefficiency and perform exact inference in this model without assuming technical and allocative inefficiency as fixed parameters. Thus the main contribution of the paper is to show how to estimate a well-specified translog system (in which the error terms in the cost and cost-share equations are internally consistent) in a random effects framework.

More specifically, here we consider a Bayesian approach to address the Greene problem. Bayesian analysis of a stochastic frontier function was first proposed by van den Broeck, Koop, Osiewalski, and Steel (1994). The Gibbs sampler has been proposed as an effective numerical technique by Koop, Steel, and Osiewalski (1995) where it is shown that Gibbs sampling has an advantage over importance sampling. Koop, Osiewalski, and Steel (1997) proposed measuring technical inefficiency in panel data models where technical inefficiency is time-invariant. Fernandez, Koop, and Steel (2000)

---

[1] For a review of the efficiency literature see Bauer (1990), Greene (1993, 2001), Kumbhakar and Lovell (2000), and Koop and Steel (2001).
[2] On the contrary, estimation of a cost system is a common practice when measurement of input elasticities, returns to scale, productivity growth, etc., are sought (see, for example, Christensen and Greene (1976), Diewert and Wales (1987)).
[3] System approach with self-dual production function is used in Schmidt and Lovell (1979, 1980), Kumbhakar (1987), Kumbhakar et al. (1991), among many others.
[4] It is now labeled in the literature as the Greene problem (see Bauer (1990)). For a simpler functional form such as the Cobb-Douglas it is not a problem (see Schmidt and Lovell (1979)).

considered Bayesian estimation of a system of equations involving a multi-output production function without an explicit behavioral assumption (such as cost minimization or revenue/profit maximization). Here we consider a system approach that is derived from a translog cost function and the cost share equations. Thus, a cost minimization assumption is formally introduced in our model. We propose a Bayesian approach to estimate the translog cost system with only technical inefficiency, first. This model is different from the single equation cost function model of Koop et al. (1997). We then consider the cost system, in which both technical and allocative inefficiency are present. Although the former model is nested in the latter, estimation of the latter model is not a trivial extension of the former. Specialized numerical methods are needed to provide parameter inferences and measures of technical and allocative inefficiency.

We show that numerical analysis of the model from the Bayesian perspective can be facilitated using Markov Chain Monte Carlo (MCMC) procedures. Posterior analysis of the model resembles many features of the standard posterior analysis in the context of multivariate regression models. Exact finite sample posterior distributions are provided without resorting to asymptotic approximations. To account for the parametric restrictions across equations, we construct a semi-informative prior that allows for differing degrees of "correctness" of the restrictions. We also provide tools for efficiency measurement in both with and without allocative inefficiency models. Allocative inefficiency is modeled via price distortions from which inferences are drawn on input over- (under-) use. In other words, we draw (firm-specific) inferences on both price distortions and input over- (under-) use along with technical efficiency. The new methods are illustrated using panel data on U.S. commercial banks.

The remainder of the paper is organized as follows. The model with only technical inefficiency is developed in Section 2. This is followed by the model in which both technical and allocative inefficiency are modeled jointly. Section 4 deals with prior specification. The U.S. commercial banking data and empirical results are discussed in Section 5 while Section 6 concludes the paper.

2. A model with only technical inefficiency

We begin with a cost minimizing behavior where firms are allocatively efficient. Assuming that panel data is available and technical inefficiency is time-invariant, the cost system can be written as (Kumbhakar and Lovell (2000, p.155)

$$\ln C_{it}^a = \ln C_{it}^0 + v_{it} + u_i , \ i = 1,...,n , \ t = 1,...,T \tag{1}$$

$$S_{j,it}^a = S_{j,it}^0 + v_{j,it} , \quad j = 2,...,M \tag{2}$$

where C_{it}^a is the actual/observed cost of firm i in year t, $S_{j,it}^a$ is the observed cost share of input j ($j = 1,...,M$), C_{it}^0 is the cost frontier (cost without technical and allocative inefficiencies) and $S_{j,it}^0$ is the frontier cost share[5] of input j, v_{it} are the noise components, and $u_i \geq 0$ is time-invariant technical inefficiency, which can be interpreted as the percentage increase in cost due to technical inefficiency.

van den Broeck et al. (1994) and Koop *et al.* (1997) considered the cost function with time-invariant inefficiency that is modeled above in a Bayesian framework. However, they used a single equation approach and focused on estimating technical efficiency from the cost function alone. Another feature of the model (in (1) and (2)) is that it resembles a seemingly unrelated regression (SUR) model. A careful examination of the model reveals that it is also different from both the Fernandez, Koop, and Steel (2000) model and the SUR model. We extend the Koop *et al.* model to a system and the SUR model to accommodate technical inefficiency. Neither the technique proposed by Koop *et al.* (1997) and Fernandez et al. (2000) nor the standard Bayesian SUR technique (Griffiths (2001)) can be applied to estimate the model proposed above. Fernandez et al. (2000) present a system of equations associated with the distance function but the formulation is ad hoc. Also, the numerical techniques presented here are different from those in Fernandez et al. (2000).

We rewrite the above cost system in a generic form (which is a panel version of the SUR equation system extended to include time-invariant technical inefficiency):

$$
\begin{aligned}
y_1 &= X_1 \beta_1 + v_1 + u \otimes 1_T \\
y_2 &= X_2 \beta_2 + v_2 \\
&\vdots \\
y_M &= X_M \beta_M + v_M
\end{aligned}
\tag{3}
$$

where y_m is an $nT \times 1$ vector of observations[6] for the mth dependent variable ($m = 1,...,M$), X_m is an $nT \times k_m$ matrix of observations for the explanatory variables in the mth equation, β_m is a $k_m \times 1$ parameter error, v_m is an $nT \times 1$ random vector, u is a $n \times 1$ non-negative random vector representing time invariant technical inefficiency, and 1_T is a $T \times 1$ unit vector. Thus, n is the number of firms and each of these firms is observed for T time periods. The first equation in (3) is the translog cost function, and the remaining $M - 1$ equations are the associated cost share equations. We rewrite (3) as

---

[5] One cost share equation is dropped to avoid a singularity problem.
[6] It is straightforward to accommodate unbalanced panels. We assume technical inefficiency is time invariant.

$$y = X\beta + v + \begin{bmatrix} u \otimes 1_T \\ 0_{nT(M-1)} \end{bmatrix} \tag{4}$$

where $0_{nT(M-1)}$ is an $nT(M-1) \times 1$ vector of zeros, and the notations y and X are obvious. Regarding stochastic components we assume that

(i) $v \sim N_{nTM}(0_{nTM}, \Sigma \otimes I_{nT})$, where Σ is an $M \times M$ contemporaneous covariance matrix;

(ii) $u_i \sim IN(0, \sigma_u^2)$, $u_i \geq 0$ $(i=1,...,n)$, i.e., u follows a half-normal distribution,[7]

(iii) v and u are mutually independent, as well as independent of X.

With the above distributional assumptions the likelihood function of the model in (4) is given by

$$L(\beta, \Sigma^{-1}, \sigma_u; y, X) \propto \int_{\Re_+^n} |\Sigma^{-1}|^{n/2} \exp(-\tfrac{1}{2} tr\ A(\beta, u)\Sigma^{-1})\, p(u \mid \sigma_u)\, du \tag{5}$$

where

$$p(u \mid \sigma_u) = \left(\frac{\pi}{2}\sigma_u^2\right)^{-\frac{n}{2}} \exp(-\frac{1}{2\sigma_u^2} u'u), \quad u \in \Re_+^n \tag{6}$$

is the joint density function of u from (ii) above and

$$A(\beta, u) = \begin{bmatrix} (y_1 - X_1\beta_1 - u \otimes 1_T)'(y_1 - X_1\beta_1 - u \otimes 1_T) & \cdots & (y_1 - X_1\beta_1 - u \otimes 1_T)'(y_M - X_M\beta_M) \\ \vdots & & \vdots \\ (y_M - X_M\beta_M)'(y_1 - X_1\beta_1 - u \otimes 1_T) & \cdots & (y_M - X_M\beta_M)'(y_M - X_M\beta_M) \end{bmatrix} \tag{7}$$

For a Bayesian analysis we need to choose the prior density function of the parameters, viz., $p(\beta, \Sigma^{-1}, \sigma_u)$. Here we choose the following conditional structure:

$$p(\beta, \Sigma^{-1}, \sigma_u) \propto p(\beta \mid \Sigma^{-1}, \sigma_u)\, p(\Sigma^{-1})\, p(\sigma_u) \propto p(\beta)p(\Sigma^{-1})p(\sigma_u) \tag{8}$$

where

$$p(\sigma_u) \propto \sigma_u^{-(\underline{n}+1)} \exp\left(-\frac{\underline{q}}{2\sigma_u^2}\right), \quad \underline{n} \geq 0, \ \underline{q} \geq 0 \tag{9}$$

---

[7] Other distributions such as the exponential, truncated normal and gamma could be used. The relevance of these distributions in practical applications is an issue worth exploring in future research.

$$p(\Sigma^{-1}) \propto \left|\Sigma^{-1}\right|^{[\nu_\Sigma - (M+1)]/2} \exp\left(-\tfrac{1}{2} tr A_\Sigma \Sigma^{-1}\right) \qquad (10)$$

In (8) we assume, a priori, that β, Σ and σ_u are independent. The prior on σ_u^2 in (9) is inverted gamma. The prior for Σ^{-1} in (10) is Wishart with parameters ν_Σ and A_Σ. It reduces to the diffuse prior used by Zellner (1971, p. 242) when $\nu_\Sigma = 0$ and $A_\Sigma = 0_{M \times M}$. Regarding the prior on β, $p(\beta)$, we choose a form that can impose linear restrictions among the elements of β (that are derived from mathematical properties of the cost function). A suitable candidate for this is the semi-informative prior of Geweke (1993), i.e.,

$$G\beta \sim N_q\ (g,\ H) \qquad (11)$$

where G is a $q \times k$ matrix (where $k = \overset{M}{\underset{m=1}{\Sigma}} k_m$ is the total number of parameters and the rank of $G = q$), g is a $q \times 1$ vector, and H is a $q \times q$ matrix whose inverse exists. When $H \to 0_{q \times q}$, the prior in (11) allows exact imposition of the q linear restrictions. This is, indeed, the case here because the cost function we are estimating satisfies some mathematical properties that are exact (see, for example, Diewert (1982) for the properties of the cost function). As the elements of H diverge from $0_{q \times q}$, the prior becomes increasingly vague. As $\|H\| \to \infty$, the prior becomes improper.

2.1 Bayesian Inference

In sampling theory one starts computing the multiple integral in (5) and maximizes the likelihood function with respect to the parameters. Although the multiple integral in the likelihood function (5) can be computed analytically, it is likely that the log-likelihood function will be prone to numerical problems.

The same problems will be encountered in the Bayesian analysis of the posterior density function when the latent variables u are explicitly integrated out. This approach is called data augmentation. To get around such problems, we consider the posterior density function (the product of (5) and (8)) augmented by the latent inefficiency variables u:

$$p(\beta, \Sigma^{-1}, \sigma_u, u \mid y, X) \propto \left|\Sigma^{-1}\right|^{\frac{n - (M+1)}{2}} \exp(-\frac{1}{2} tr\, A(\beta, u)\Sigma^{-1}) \cdot \sigma_u^{-(n + \underline{n} + 1)} \exp\left(-\frac{q + u'u}{2\sigma_u^2}\right)$$
$$\times p(\beta) p(\Sigma^{-1}) p(\sigma_u) \qquad (12)$$

In (12) the latent variables u are treated as parameters in order to avoid the complicated likelihood function or posterior density function. This procedure facilitates considerably the use of MCMC techniques that will be used to perform inferences for this model, as we explain later. In the sampling-theory framework, this construction can be used to implement an EM algorithm.

Numerical Bayesian inference is performed using MCMC techniques[8], especially the Gibbs sampler. The philosophy of the method is simple. Given a posterior density function $p(\theta|Y)$ where $\theta = [\theta_1,...,\theta_p]'$ is the parameter vector, the objective is to simulate random draws $\{\theta^{(s)}, s = 1,...,S\}$ from the posterior. Once this is done the estimation problem is solved because (under quadratic loss) we can estimate θ from $\bar{\theta} = S^{-1}\sum_{s=1}^{S}\theta^{(s)}$ and we can compute second or other moments in a similar way, if they exist. We consider kernel densities of the individual elements of $\{\theta^{(s)}, s = 1,...,S\}$ to form approximations to marginal posterior density functions of parameters. The same is true for any function $f(\theta)$ of the parameter vector, since we have the draws $\{f(\theta^{(s)}), s = 1,...,S\}$. To generate draws from the posterior density function $p(\theta|Y)$ we consider the conditional density functions $p(\theta_j|\theta_{-j},Y)$ ($j = 1,...,p$) where θ_{-j} denotes all elements of θ except the j th element. The sequence $\{\theta^{(s)}, s = 1,...,S\}$ so generated is called Gibbs sampling sequence (Gelfand and Smith (1990), Tanner and Wong (1987)) and it converges in distribution to the posterior under fairly mild conditions (Roberts and Smith (1994)). This means that if the number of draws is large, then one can use the draws $\{\theta^{(s)}, s = 1,...,S\}$ as a sample from the posterior density function.

In the present model we generate random drawings from the following conditionals: (i) $\beta|\Sigma,\sigma_u,u,data$, (ii) $\Sigma|\beta,\sigma_u,u,data$, (iii) $\sigma_u|\beta,\Sigma,u,data$, and (iv) $u|\beta,\Sigma,\sigma_u,data$. We repeat this cycle S times to generate a sequence of length S for each one of these parameters. The draws so generated can be considered as a sample from the joint posterior density function of the parameters. The required conditional density functions to implement Gibbs sampling are as follows. For the regression parameters we have:

$$\beta|\Sigma^{-1},\sigma_u,u,y,X \sim N_k(\bar{\beta},\bar{V}) \qquad (13)$$

where the conditional posterior mean is

6

$$\bar{\beta} = [X'(\Sigma^{-1} \otimes I_{nT})X + G'H^{-1}G]^{-1}[X'(\Sigma^{-1} \otimes I_{nT})(y-U) + G'H^{-1}g],$$

$$U = \begin{bmatrix} u \otimes 1_T \\ 0_{nT(M-1)} \end{bmatrix} \text{ and the conditional posterior covariance matrix is}$$

$$\bar{V} = [X'(\Sigma^{-1} \otimes I_{nT})X + G'H^{-1}G]^{-1}.$$

The conditional posterior density function of Σ^{-1} is:

$$P(\Sigma^{-1} \mid \beta, \sigma_u, u, y, X) \propto |\Sigma^{-1}|^{\frac{n+v_\Sigma-(M+1)}{2}} \exp(-\tfrac{1}{2}tr[A_\Sigma + A(\beta,u)]\Sigma^{-1}) \tag{14}$$

which is a Wishart density function. The conditional posterior density function of σ_u is:

$$p(\sigma_u \mid \beta, \Sigma^{-1}, u, y, X) \propto \sigma_u^{-(n+\underline{n}+1)} \exp\left(-\frac{u'u+\underline{q}}{2\sigma_u^2}\right) \tag{15}$$

which implies

$$\frac{u'u+\underline{q}}{\sigma_u^2} \mid \beta, \Sigma^{-1}, u, y, X \sim \chi^2_{n+\underline{n}}. \tag{16}$$

It can be shown that the conditional posterior density function of latent inefficiencies is given by

$$u_i \mid \beta, \Sigma^{-1}, \sigma_u, y, X \sim N_1(\mu_i^*, \sigma_*^2), \quad u_i \geq 0, i = 1,...,n \tag{17}$$

where

$$\mu_i^* = T\sigma_*^2 \sum_{j=1}^m \bar{e}_{ji}\sigma^{1j}, i=1,...,n \; ; \sigma_*^2 = \frac{\sigma_u^2}{1+T\sigma^{11}\sigma_u^2} \; ; \; e_m = y_m - X_m\beta_m = [e_{m1},...,e_{mT}]', \quad \bar{e}_m = T^{-1}\sum_{t=1}^T e_{mt}$$

with $\bar{e}_m = [\bar{e}_{m1},...,\bar{e}_{mn}]'$ for $m=1,..,M$. The inverse contemporaneous covariance matrix is expressed as $\Sigma^{-1} = [\sigma^{ij}]$. Finally, the u_i's are independent in (17). To derive this result, notice that from (12) we can decompose the posterior as $p(u \mid \beta, \Sigma^{-1}, \sigma_u, y, X) \cdot p(\beta, \Sigma^{-1}, \sigma_u \mid y, X)$, in which the second part is irrelevant. We follow the technique presented in Tsionas (1999) to draw from (17). This technique utilizes acceptance sampling based on an exponential blanketing density whose parameter is chosen to maximize the acceptance rate.

---

[8] For a review of MCMC methods in econometrics, see Geweke (1999).

To set up the Gibbs sampler we draw random numbers from conditional posterior density functions (13), (14), (15), (16) and (17). This task is straightforward because these density functions are from well-known families like the normal, truncated normal and Wishart. Therefore, the Gibbs sampler provides a straightforward numerical approach to Bayesian analysis of a translog cost system involving technical inefficiency.

2.2 Efficiency measurement

In this section, we describe efficiency measurement based on the concept of posterior predictive efficiency developed by van den Broeck et al (1994). Consider a yet unobserved firm for which the data on the dependent variables are in y^0 (a $MT \times 1$ vector) and the data on the explanatory variables are in X^0 (a $MT \times k$ matrix), i.e.,

$$
y^0 = \begin{bmatrix} y_1^0 \\ \vdots \\ y_M^0 \end{bmatrix}, \qquad X^0 = \begin{bmatrix} X_1^0 \\ \vdots \\ X_M^0 \end{bmatrix}
$$

and y_m^0 is $T \times 1$ vector and X_m^0 is $MT \times k_m$ matrix ($m = 1, \dots, M$). Define $\mu_0^* = T\sigma_*^2 \sum_{m=1}^{M} \bar{e}_m^0 \sigma^{1m}$,

$\bar{e}_m^0 = T^{-1} \sum_{t=1}^{T} e_{mt}^0$, $m = 1, \dots, M$ and $e^0 = y^0 - X^0\beta = [e_1^0, \dots, e_M^0]'$.

From (17) the conditional posterior density function of the latent inefficiency for the yet unobserved firm is

$$
u_0 \big| \beta, \Sigma^{-1}, \sigma_u, y, X, X^0 \sim N(\mu_0^*, \sigma_*^2), \ u_0 \geq 0 \tag{18}
$$

Let $r_0 = \exp(-u_0)$ be the efficiency of the firm, $r_0 \in (0,1)$. Then

$$
p(r_0 \big| \beta, \Sigma^{-1}, \sigma_u, y, X, X^0) = \left(\frac{\pi}{2} \sigma_*^2 \right)^{-\frac{1}{2}} \Phi\left(\frac{\mu_0^*}{\sigma^*} \right) r_0^{-1} \exp\left[-\frac{(\ln r_0 + \mu_0^*)^2}{2\sigma_*^2} \right], r_0 \in (0,1) \tag{19}
$$

It is necessary to integrate out the model parameters (viz., β and Σ^{-1}) to obtain the marginal density function of r_0. For this we write (19) as

$$
p(r_o \big| y, X, X^o) = \int p(r_o \big| \beta, \Sigma^{-1}, \sigma_u, y, X, X^o) p(\beta, \Sigma^{-1}, \sigma_u \big| y, X) d\beta \cdot d\sigma_u \cdot d\Sigma^{-1} \tag{20}
$$

An approximation of (20) can be computed using the standard estimator

$$p(r_o \mid y, X, X^o) \propto S^{-1} \sum_{s=1}^{S} p(r_o \mid \beta_{(s)}, \Sigma_{(s)}^{-1}, \sigma_{u,(s)}, y, X, X^o) \tag{21}$$

where $\{\beta_{(s)}, \Sigma_{(s)}^{-1}, \sigma_{u,(s)}; s = 1,...,S\}$ is the set of posterior draws. The posterior predictive density function in (21) can be presented graphically to draw inferences about the efficiency level of a yet unobserved firm, after normalization to make it a proper density function.

In practice it is important to report efficiency measures for the observed firms as well. The density function of $u_i \mid \beta, \Sigma^{-1}, \sigma_u, y, X$ is given in (17). Thus, if

$$r_i = \exp(-u_i) \tag{22}$$

then a straightforward modification of (19) can be used to obtain the firm-specific efficiency density function.

Moments of r_i can be computed easily, and the density $p(r_i \mid y, X)$ can be approximated using the standard estimator based on the set of posterior draws. The mean and/or median of r can be used to predict efficiency. We do this as follows: Given the draws $\{u_i^{(s)}, s = 1,..,S\}$ for the s th iteration of the Gibbs sampler, we compute $r_i^{(s)} = \exp(-u_i^{(s)})$. Since $u_i^{(s)}$ is a draw from the conditional density function $r_i \mid \beta, \Sigma^{-1}, \sigma_u, y, X$ it follows that

$$\bar{r}_i = S^{-1} \sum_{s=1}^{S} r_i^{(s)} \tag{23}$$

represents average firm-specific technical efficiency.

2.3 Prior elicitation

Given the functional forms of the prior density function, in practice, we have to choose the hyperparameters to match whatever prior knowledge we may have. Although it is difficult to have prior notions about parameters like β or Σ (other than restrictions imposed by economic theory) it is sometimes possible to utilize prior information about inefficiency. Given the prior in (9) for parameter σ_u the objective in this section is to choose the hyperparameters $\underline{n}$ and $\underline{q}$ in some satisfactory way.

Since $r_i = \exp(-u_i)$, $r_i \in (0,1)$, we have

$$p(r_i \mid \sigma_u) = \left(\frac{\pi}{2} \sigma_u^2 \right)^{-\frac{1}{2}} r_i^{-1} \exp\left(-\frac{(\ln r_i)^2}{2\sigma_u^2} \right), \quad i = 1, \ldots, n \tag{24}$$

We can either treat u_i as a model parameter and use the prior in (6) or consider it to be a part of the model. Both the interpretations give the same posterior results. The prior of σ_u in (9) depends on the hyperparameters $\underline{n}$ and $\underline{q}$. These parameters may be elicited as follows. To facilitate prior elicitation, we used numerical quadrature to compute the mean $\bar{r}$ and variance $\bar{s}^2$ of the marginal prior for values in the range $\underline{n} \in [1, 100]$ and $\underline{q} \in [0.001, 5]$. Then we computed the following regressions (with 5,000 observations) to approximate prior elicitation.[9]

$$\ln(\underline{q}) = 6.163 + 3.032 \cdot \ln \bar{r} + 1.090 \cdot \ln \bar{s}^2, \quad R^2 = 0.800 \tag{25}$$

$$\ln(\underline{n}/\underline{q}) = -1.157 + 2.022 \cdot \ln \bar{r} - 1.024 \cdot \ln \bar{s}^2, \quad R^2 = 0.999 \tag{26}$$

For any desired prior mean efficiency and prior variance, these regressions can be used to obtain approximately the right values of the hyperparameters $\underline{n}$ and $\underline{q}$. More precise prior elicitation can be accomplished using exact quadrature methods with the implied prior probability density function for efficiency, $p(r)$. Alternatively, it can be approximated using simulation techniques. Given a sample of values $\{\sigma_u^{(s)}; s = 1, \ldots, S\}$ from the prior, one could draw from $u^{(s)} \mid \sigma_u^{(s)} \sim \mid N(0, \sigma_u^{2(s)}) \mid$, compute $r^{(s)} = \exp(-u^{(s)})$, and approximate the marginal prior $p(r)$ using a histogram of the $r^{(s)}$.

It can be shown that the posterior density function is finitely integrable and that parameters and efficiency measures have finite first and second moments. The most important result is that with cross-sectional data and a flat prior on Σ^{-1} the posterior does not exist, and we need a prior that keeps Σ probabilistically "away from zero". Posterior moments exist under standard conditions. A Technical Appendix detailing these statements is available upon request.

3. A model with both technical and allocative inefficiency

In this section, we consider a model that allows for both technical and allocative inefficiency. Perhaps the simplest way to deal with allocative inefficiency is to argue that the share equation residuals

represent deviations from first-order conditions and, therefore, they represent allocative distortions. This modeling approach, however, fails to take into account the link between allocative inefficiency and its impact on cost. Here we follow Kumbhakar (1997) who, following the definition of allocative inefficiency from Schmidt and Lovell (1979), derived the exact relationship between allocative inefficiency and cost therefrom in the context of the translog cost function. It solves the Greene problem theoretically. No estimation technique is, however, offered. And we are not aware of any application where the Greene problem is solved using a flexible cost function and treating allocative inefficiency as random. In a sampling theory framework empirical application of the Kumbhakar (1997) model is difficult because of the computational complexity of the model, especially when allocative inefficiency is represented by random variables à la Schmidt and Lovell (1979).

Assume ξ_j represents (time-invariant) allocative inefficiency for the input pair $(j,1)$ so that the relevant input price vector (often labeled as shadow price vector) to the firm is $(w^* \equiv (w_1, w_2^*,..., w_M^*)) = (w_1, w_2 \exp(\xi_2),..., w_M \exp(\xi_M))$, where $\xi_2,...,\xi_M$ are random variables. Kumbhakar (1997) showed that the translog system (with a single output) can be written as follows.[10]

$$\ln C_{it}^a = \ln C_{it}^* + \ln G_{it} + v_{it} + u_i, \ i = 1,...,n, \ t = 1,...,T \tag{27}$$
$$S_{j,it}^a = S_{j,it}^0 + \eta_{j,it} \tag{28}$$

where C_{it}^a, $S_{j,it}^a$, $S_{j,it}^0$, v_{it} and u_i are the same and defined in Section 2. The arguments of C_{it}^* are w_{it}^* and y_{it} while those in C_{it}^0 (defined in Section 2) are w_{it} and y_{it}. The $\eta_{j,it}$ and $\ln G_{it}$ are functions of allocative inefficiency, $\xi_2,...,\xi_M$ (defined below). We rewrite (27) as $\ln C_{it}^a = \ln C_{it}^0 + \ln C_{it}^{AL} + v_{it} + u_i$ where $\ln C_{it}^{AL} (= \ln C_{it}^* - \ln C_{it}^0 + \ln G_{it})$ can be interpreted as the percentage increase in cost due to allocative inefficiency.[11] For a translog functional form C_{it}^0, $S_{j,it}^0$, $\ln C_{it}^{AL}, \ln G_{it}$ and $\eta_{j,it}$ are:

$$\ln C_{it}^0 = \alpha_0 + \sum_j \alpha_j \ln w_{j,it} + \gamma_y \ln y_{it} + \tfrac{1}{2}\gamma_{yy}(\ln y_{it})^2 + \tfrac{1}{2}\sum_j \sum_k \beta_{jk} \ln w_{j,it} \ln w_{k,it}$$
$$+ \sum_j \gamma_{jy} \ln w_{j,it} \ln y_{it} + \alpha_t t + \tfrac{1}{2}\alpha_{tt} t^2 + \beta_{yt} \ln y_{it} t + \sum_j \beta_{jt} \ln w_{j,it} t, \tag{29}$$

$$S_{j,it}^0 = \alpha_j + \sum_k \beta_{jk} \ln w_{k,it} + \gamma_{jy} \ln y_{it} + \beta_{jt} t, \tag{30}$$

---

[9] If one needs highly precise priors then it is necessary to conduct the simulation that will give the exact priors.
[10] The multiple output generalization of this result is straightforward.
[11] This is non-negative given strict concavity of the cost function.

$$\ln C_{it}^{AL} = \ln G_{it} + \sum_j \alpha_j \xi_{j,i} + \sum_j \sum_k \beta_{jk} \xi_{j,i} \ln w_{k,it} + \tfrac{1}{2} \sum_j \sum_k \beta_{jk} \xi_{j,i} \xi_{k,i} + \sum_j \gamma_{jy} \xi_{j,i} \ln y_{it} + \sum_j \beta_{jt} \xi_{j,i} t, \quad (31)$$

$$G_{it} = \sum_j S_{j,it}^* \exp(-\xi_{j,i}), \tag{32}$$

where

$$S_{j,it}^* = \alpha_j + \sum_k \beta_{jk} \ln w_{k,it}^* + \gamma_{jy} \ln y_{it} + \beta_{jt} t \equiv S_{j,it}^0 + \sum_k \beta_{jk} \xi_k. \tag{33}$$

Finally,

$$\eta_{j,it} = \frac{S_{j,it}^0 \left\{ 1 - G_{it} \exp(\xi_{j,i}) \right\} + \sum_k \beta_{jk} \xi_k}{G_{it} \exp(\xi_{j,i})}. \tag{34}$$

Thus, $\eta_{j,it}$ are the deviations of the actual cost shares from their optimum values, and are non-linear functions of allocative inefficiency, $\xi_2,...,\xi_M$, and data.

If we denote the vectors of all observations on log cost and $M-1$ cost shares by $y_1, y_2,..., y_M$, the matrix of cost function regressors (observations on log prices, log output, their squares and interactions) by X_1, the matrix of cost share equation regressors (observations on log prices and log output by X_2, then we can write the translog cost system in (27) and (28) as

$$y_1 = X_1(\xi)\beta_1 + \ln G(\xi, \beta) + v_1 + u \otimes 1_T$$

$$\tag{35}$$

$$y_j = X_2 \beta_j + \eta_{j-1}(\xi, \beta) + v_j, \quad j = 2,..., M$$

where we have appended error terms v_j ($j \geq 2$) in the share equations to capture, for example, measurement errors in cost share equations. The matrix $X_1(\xi)$ denotes X_1 when $w_{j,it}$ are replaced by $w_{j,it}^* \equiv w_{j,it} \exp(\xi_{j,i})$ so that $X_1(0_{M-1}) = X_1$. Finally, β denotes the entire parameter vector. The system in (35) is a nonlinear seemingly unrelated regression model with nonlinear random effects.

We continue to assume, as before, that $v \sim N_{MnT}(0, \Sigma \otimes I_T)$. Furthermore, we assume that $\xi_i = [\xi_{2,i}, ..., \xi_{M,i}]' \sim N_{M-1}(0, \Omega)^{12}$, $i = 1, ..., n$. Then the above model represents a system of nonlinear regression equations with random effects. We write the system compactly as

$$y = X(\xi)\beta + \phi(\xi, \beta) + v + \begin{bmatrix} u \otimes 1_T \\ 0_{(M-1)nT} \end{bmatrix} \tag{36}$$

where $\beta = [\beta_1', ..., \beta_M']'$, $\phi(\xi, \beta) = \begin{bmatrix} \ln G(\xi, \beta) \\ \eta(\xi, \beta) \end{bmatrix}$, $X(\xi) = \begin{bmatrix} X_1(\xi) \\ X_2 \otimes 1_{M-1} \end{bmatrix}$

We assume that $v \sim N_{nTM}(0, \Sigma \otimes I_{nT})$, $\xi \sim N_{n(M-1)}(0, \Omega \otimes I_n)$, and both are independent of each other as well as independent of X. Regarding the priors we have

$$p(\beta, \Sigma^{-1}, \sigma_u, \Omega^{-1}) \propto p(\beta)p(\Sigma^{-1})p(\sigma_u)p(\Omega^{-1}).$$

The priors on σ_u, Σ^{-1}, and β are the same as in (9), (10) and (11), and we choose a Wishart prior for Ω^{-1}, viz.,

$$p(\Omega^{-1}) \propto |\Omega^{-1}|^{(v_\Omega - M)/2} \exp\left(-\tfrac{1}{2}A_\Omega \Omega^{-1}\right)$$

where v_Ω and A_Ω are parameters of the prior density function. The augmented posterior density function of the model is

$$p(\beta, \Sigma^{-1}, \sigma_u, u, \xi, \Omega^{-1} \mid y, X) \propto |\Sigma^{-1}|^{\frac{n-(M+1)}{2}} \exp(-\tfrac{1}{2}tr\, A(\beta, \xi, u)\Sigma^{-1}) \cdot \sigma_u^{-(n+\underline{n}+1)}$$

$$\times \exp\left(-\frac{q + u'u}{2\sigma_u^2}\right) |\Omega^{-1}|^{(n-m)/2} \exp\left(-\tfrac{1}{2}tr\, Q(\xi)\Omega^{-1}\right) p(\beta \mid \Sigma^{-1}, \sigma_u, \Omega) \tag{37}$$

where $Q(\xi) = \sum_{i=1}^{n} \xi_i \xi_i'$ and $A(\beta, \xi, u)$ is similar to (7) except that we have the ξ terms in it, viz.,

---

[12] It is easy to assume a non-zero mean for the ξ's. Allowing for a non-zero mean, say μ, required somewhat tight priors at least in our application. Although μ is clearly identified from the share equation constant terms, it does not appear that we have "proper empirical identification" to use a term suggested to us by an anonymous referee. For that reason, we opt for setting μ = 0 in this application, reflecting our prior notion that average allocative inefficiency is likely to be small.

$$A(\beta,\xi,u) = \begin{bmatrix} \left(e_1(\xi,\beta,u)-u\otimes 1_T\right)'\left(e_1(\xi,\beta,u)-u\otimes 1_T\right) & \cdots & e_1(\xi,\beta,u)'e_M(\xi,\beta,u) \\ \vdots & & \vdots \\ e_M(\xi,\beta,u)'\left(e_1(\xi,\beta,u)-u\otimes 1_T\right) & \cdots & e_M(\xi,\beta,u)'e_M(\xi,\beta,u) \end{bmatrix}$$

where $e_m(\xi,\beta,u) = y_m - X_m(\xi)\beta_m - \phi_m(\xi,\beta)$, $m = 1,...,M$. The prior density functions for parameters other than Ω are the same as before.

The kernel posterior density function of parameters is

$$p(\beta,\Sigma^{-1},\sigma_u,\Omega^{-1} \mid y,X) = \int_{\mathfrak{R}^n}\int_{\mathfrak{R}^n_+} p(\beta,\Sigma^{-1},\sigma_u,u,\xi,\Omega^{-1} \mid y,X)\,du\,d\xi \; ,$$

which doesn't have a closed form analytical solution.[13] For this reason, inference in this model is a challenge.[14]

3.1 Bayesian inference

To perform Bayesian analysis of this model we utilize MCMC methods associated with the posterior distribution. In other words, we construct a Markov chain defined by conditional density functions of parameters. In this Markov chain, random draws are made from each posterior conditional distribution. The conditional posterior distributions required for implementing MCMC techniques are as follows.

---

[13] It is possible to integrate explicitly the augmented posterior with respect to Σ and Ω but the resulting expressions are highly complicated, depend on the latent variables u and ξ, and thus it is not clear how this could be useful in estimation.

[14] Alternative computational approaches to estimating nonlinear random effect models are available in the statistics literature, although systems of nonlinear equations are hard to find. Pinheiro and Bates (1995) discuss the theory and computational techniques for nonlinear random effect models, and conclude that adaptive Gaussian quadrature is one of the best methods to obtain approximations to the integrated likelihood function (the integration is with respect to ξ). However, when ξ_i is defined in more than one dimension, Gaussian quadrature is subject to curse of dimensionality. Laplace approximation is another alternative, and has been investigated by many authors, including Vonesh and Chinchilli (1997), Wolfinger (1993), and Wolfinger and Lin (1997). In the related literature on generalized nonlinear models with random effects, numerical quadrature techniques have been considered and analyzed by Longford (1994), McCulloch (1994), Liu and Pierce (1994), and Diggle, Liang, and Zeger (1994). These techniques have, however, unknown degrees of accuracy with respect to approximating the likelihood function. It is clear that an error ε in the computation of integrals with respect to a given ξ_j for a particular observation is magnified to $NJ\varepsilon$ for the log-likelihood if we have N observations and J ξ_j's, where J = M-1. Thus the error can be substantial. In addition, if J is large the curse of dimensionality will interfere with our inability to use these techniques in practically relevant applications.

3.2 Conditional posterior of β

The conditional posterior density function for β is

$$p(\beta \mid \Sigma^{-1}, \sigma_u, u, \xi, \Omega^{-1}, y, X) \propto \exp(-\tfrac{1}{2} tr\, A(\beta, \xi, u)\Sigma^{-1}) \cdot p(\beta \mid \Sigma^{-1}, \sigma_u, \Omega) \qquad (38)$$

This density function is not from any known family so random number generation is difficult. One can, however, use the Metropolis-Hastings algorithm (Tierney (1994)). Our objective is to generate random draws from a distribution with density $f(x)$ but a direct sampling is not possible. However, there is a density $g(x)$ from which random number generation is easy. So we proceed as follows. Given an initial condition $x^{(0)}$, we generate a candidate $y \sim g(x)$. The next draw will be either $x^{(0)}$ or the candidate y. More specifically, we set $x^{(1)} = y$ with probability $\alpha(x^{(0)}, y) = \min\left\{1, \dfrac{f(y)/g(y)}{f(x^{(0)})/g(x^{(0)})}\right\}$, else we set $x^{(1)} = x^{(0)}$. We continue this process until we get S draws. Apparently, for this procedure to work, $g(x)$ must be a good approximation to $f(x)$, otherwise we will be rejecting a lot of the candidates, meaning that effectively we will be unable to explore $f(x)$, as desired. The method is applied in the current setting in the following manner. Let $f(\beta \mid \xi, u, \sigma_u, \Sigma, y, X)$ be the exact posterior conditional, and $g(\beta \mid \xi, u, \sigma_u, \Sigma, y, X)$ be the posterior corresponding to the multivariate Student-t proposal density[15] for β when $\phi(\xi, \beta) = 0$. Suppose the current draw is $\beta^{(i)}$.

Define $\alpha\left(\beta^{(i)}, \widetilde{\beta}\right) = \min\left\{1, \dfrac{f(\widetilde{\beta} \mid \xi, u, \sigma_u, \Sigma, y, X)/g(\widetilde{\beta} \mid \xi, u, \sigma_u, \Sigma, y, X)}{f(\beta^{(i)} \mid \xi, u, \sigma_u, \Sigma, y, X)/g(\beta^{(i)} \mid \xi, u, \sigma_u, \Sigma, y, X)}\right\}$. Then, with probability $\alpha\left(\beta^{(i)}, \widetilde{\beta}\right)$ we accept the proposal $\widetilde{\beta}$, else we maintain the current draw $\beta^{(i)}$. If the acceptance rate of this proposal is not satisfactory, we can always modify the proposal to some extent. For example, we can multiply its covariance matrix by a certain constant that can be tuned to maintain a satisfactory acceptance rate.

This proposal is attractive because it automatically allows for imposition of all theoretical restrictions via the prior $p(\beta)$. In our application, the acceptance rate of this proposal is near 70%. We

---

[15] It is necessary to maintain the acceptance probability bounded so the degrees of freedom of the Student-t should be less than the degrees of freedom of the usual approximate Student-t posterior for β. In our empirical work we sample from Student-t with 40 degrees of freedom but have found that normal proposals produce only trivial changes.

also impose the monotonicity and concavity restrictions at each sample point using the rejection method, exactly as in the model with only technical inefficiency.

3.3 Conditional posterior of Σ

The posterior conditional density function of Σ is given by

$$p(\Sigma^{-1} \mid \beta, \sigma_u, u, \xi, \Omega^{-1}, y, X) \propto \left|\Sigma^{-1}\right|^{\frac{n + v_\Sigma - (M+1)}{2}} \exp(-\tfrac{1}{2} tr[A_\Sigma + A(\beta, \xi, u)]\Sigma^{-1}) \qquad (39)$$

which is the density function of a Wishart distribution. Given β, ξ, u the matrix $A(\beta, \xi, u)$ is a known constant. So generating random draws from the above density function is straightforward.

3.4 Conditional posterior of Ω

The posterior conditional density function of Ω is

$$p(\Omega^{-1} \mid \beta, \Sigma^{-1}, \sigma_u, u, \xi, y, X) \propto \left|\Omega^{-1}\right|^{(n + v_\Omega - M)/2} \exp\left(-\tfrac{1}{2} tr[A_\Omega + Q(\xi)]\Omega^{-1}\right) \qquad (40)$$

which is also the density function of a Wishart distribution since $A_\Omega + Q(\xi)$ is a given matrix of constants.

3.5 Conditional posterior of σ_u

The posterior conditional density function of σ_u satisfies

$$\frac{\underline{q} + u'u}{\sigma_u^2} \mid \beta, \Sigma^{-1}, \Omega^{-1}, u, \xi, y, X \sim \chi_{n+\underline{n}}^2 \qquad (41)$$

from which random number generation is simple.

3.6 Conditional posterior of u

For latent technical inefficiency, it can easily be shown, using the techniques developed in the previous section, that

$$u_i \big| \beta, \Sigma^{-1}, \Omega^{-1}, \sigma_u, \xi, y, X \sim N_1(\mu_i^*, \sigma_*^2), \quad u_i \geq 0, \ i = 1...n \tag{42}$$

where

$$\mu_i^* = T\sigma_*^2 \sum_{j=1}^m \bar{e}_{ji}(\xi, \beta, u)\sigma^{1j}, \ i = 1,...,n \ ; \ \sigma_*^2 = \frac{\sigma_u^2}{1 + T\sigma^{11}\sigma_u^2} \ ;$$

$$e_m(\xi, \beta, u) = y_m - X_m(\xi)\beta_m - \phi_m(\xi, \beta) = [e_{m1}(\xi, \beta, u),...,e_{mT}(\xi, \beta, u)]', \ m = 1,...,M$$

$$\bar{e}_m(\xi, \beta, u) = T^{-1} \sum_{t=1}^T e_{mt}(\xi, \beta, u), \ m = 1,...,M \ \text{ is a } n \times 1 \text{ vector with}$$

$$\bar{e}_m(\xi, \beta, u) = [\bar{e}_{m1}(\xi, \beta, u),...,\bar{e}_{mn}(\xi, \beta, u)]', \ m = 1,...,M$$

and the inverse contemporaneous covariance matrix is expressed as $\Sigma^{-1} = [\sigma^{ij}]$. This distribution of u_i is truncated normal, and since the u_i's are independent in their joint posterior conditional distribution, random draws can be generated sequentially for each $i = 1,...,n$ as in Tsionas (1999).

3.7 Conditional posterior of ξ

Finally, we consider the posterior conditional density function of ξ given by

$$p(\xi \mid \beta, \Sigma^{-1}, \sigma_u, u, \Omega^{-1}, y, X) \propto \exp(-\tfrac{1}{2} tr \, A(\beta, \xi, u)\Sigma^{-1}) \cdot \exp\left(-\tfrac{1}{2} tr Q(\xi)\Omega^{-1}\right) \tag{43}$$

Generating random draws from this joint density function is not straightforward because the density function is not from any known form. One promising possibility to obtain a reasonably good proposal density function is to linearize the cost share equations (i.e., to make them linear in the ξ's). The resulting approximate posterior of each ξ_i will be normal so in practice we can use a Student-t proposal density function[16] to maintain the acceptance probability bound. We can easily obtain a random draw from this posterior, and use a Metropolis rule to maintain the correct posterior. The normal approximation is, in fact, very simple to use. The task is to linearize the cost function and share equations with respect to ξ and use the approximation to obtain a multivariate normal or Student-*t* density function for the ξ's that can be used as proposal density function for the Metropolis-Hastings step. In Appendix A we show that partial derivatives of the cost function with respect to ξ's are exactly zero, so we can use only the share equations to derive the linear approximation so the normal approximation has a particularly simple

---

[16] We use a Student-t with 10 degrees of freedom but we have found the results remain the same when we use a normal proposal.

structure for the translog cost system. The acceptance rate of the Metropolis chain is over 85% for the data we analyzed, which is a satisfactory approximation.[17] Let ξ_i be the current draw from the approximate density function. Let $f(\xi_i \mid y, X, \beta, \Sigma, \Omega)$ be the exact conditional density function of ξ_i given the data and the parameters, and $g(\xi_i \mid y, X, \beta, \Omega)$ be the normal approximation, *i.e.*, the pdf of the multivariate normal resulting from the normal approximation. Clearly, both density functions are available in closed form. The candidate $\xi_i \sim f(\xi_i \mid y, X, \beta, \Sigma, \Omega)$ will either be accepted or rejected in favor of the previous draw, say $\xi_i^{(0)}$, according to the following rule. Let

$$a\left(\xi_i^{(0)}, \xi_i\right) = \min\left\{1, \ \frac{f(\xi_i \mid y, X, \beta, \Sigma, \Omega)/g(\xi_i \mid y, X, \beta, \Omega)}{f(\xi_i^{(0)} \mid y, X, \beta, \Sigma, \Omega)/g(\xi_i^{(0)} \mid y, X, \beta, \Omega)}\right\}$$

Either we accept this draw with probability $a\left(\xi_i^{(0)}, \xi_i\right)$, or reject it and take $\xi_i^{(0)}$ as the draw. The overall acceptance rate of this procedure is over 85% in the data set we have analyzed. It is necessary to ensure that the approximation to the exact conditional posterior of ξ_i is satisfactory. We cannot claim that this will always be the case but we suspect that when this is not so, the approximation can be improved adaptively by linearizing around a point $\overline{\xi}_i$ (different from zero) that could be the current posterior mean of ξ_i.[18]

3.8 Joint measurement of technical and allocative inefficiency

The previous model is capable of providing measures of technical and allocative inefficiency for each firm, and for a yet unobserved firm (in which case we make posterior predictive inferences, *i.e.*, predictive inferences conditional on the observed data). Our problem here is as follows. Suppose $f(\theta, D)$

---

[17] Following the sampling theory literature we could have used Gaussian quadrature to integrate out ξ's from the likelihood function or the posterior distribution, and then use the Metropolis-Hastings algorithm (MHA) to provide inferences for β, Σ, Ω. The required posterior would not, of course, be available in closed form. We did not opt for this technique because the MHA does not take account of the special features of the problem, namely linearity of the system conditional on ξ's. Moreover, for complicated posteriors, the MHA can result in high autocorrelation of the draws, making reliable exploration of the posterior a troublesome task.

[18] We have tried this approximation in our application and found that the acceptance rate increased slightly. However, we decided not to use it in reporting the final results since the original proposal performed rather well. Another advantage of the algorithm is that it can be vectorized easily to generate all ξ_i's at once by exploiting properties of the normal distribution. Consequently, fitting the nonlinear translog random effect model is not considerably more time consuming compared to the translog system with only technical inefficiency.

represents any function of the parameters θ (inclusive, perhaps, of any latent variables like u and ξ) and the data D. The objective is to estimate the posterior expectation $Ef(\theta,D) = \int f(\theta,D)p(\theta|D)d\theta$.

Given draws $\theta^{(s)}$, $s = 1,...,S$ from the posterior density function $p(\theta|D)$, this expectation can be approximated by $E[f(\theta,D)|D] \approx S^{-1}\sum_{s=1}^{S} f(\theta^{(s)},D)$.

Next, we describe the procedure to obtain technical and allocative inefficiency measures. Measurement of technical inefficiency in the present model is exactly the same as in the model without allocative inefficiency presented in section 2. Define $r_i = \exp(-u_i)$ to be the efficiency index of firm $i = 1,...,n$. The firm-specific efficiency measure is provided by the mean of the posterior density function of r_i, viz., $E(r_i|y,X)$. This measure can be obtained by averaging $r_i^{(s)}$, where $r_i^{(s)}$ denotes the s th draw for r_i. Once a draw $u_i^{(s)}$ from the conditional posterior of u_i becomes available, this can be computed easily. The posterior predictive density function can also be obtained easily. Since $u_i|\sigma_u$ is half normal, the density function of $r_i|\sigma_u$ can be obtained easily. The marginal density function of r_i is then given by $p(r_i) = \int_0^{\infty} p(r_i|\sigma_u)p(\sigma_u|data)d\sigma_u \approx S^{-1}\sum_{s=1}^{S} p(r_i|\sigma_u^{(s)})$ where $\sigma_u^{(s)}$, $s = 1,...,S$ denotes the posterior draws for σ_u.

Regarding allocative inefficiency we follow a similar procedure. First, the departure of observed prices from shadow prices, namely the ξ_j's, are of interest. Given the posterior draws for ξ, say $\xi_{j,i}^{(s)}$, a firm-specific measure is provided by $\overline{\xi}_{j,i} = S^{-1}\sum_{s=1}^{S} \xi_{j,i}^{(s)}$ where S indicates the number of draws. This gives the percentage deviation of observed prices from shadow prices for input pair $(j,1)$ for firm i. The deviations themselves can be estimated from $\overline{\lambda}_{j,i} = S^{-1}\sum_{s=1}^{S} \lambda_{j,i}^{(s)}$, where $\lambda_{j,i}^{(s)} = \exp(\xi_{j,i}^{(s)})$, $s = 1,...,S$.

Another measure of interest is the percentage increase in costs due to allocative inefficiency for each firm. This is given by $\ln C_{it}^{AL}$ and depends on the data, the ξ s and the parameters in β. Clearly, this measure can be computed for each draw of β and ξ. It can be averaged with respect to the draws, and provide temporal and firm-specific measures for $\ln C_{it}^{AL}$ or C_{it}^{AL}. A posterior predictive density function for $\ln C^{AL}$, referring to an as of yet unobserved (or a typical) firm is computed as follows. Given a vector of prices and output for that firm (for example, the sample averages of these variables) $\ln C^{AL}$ is

19

computed for each draw, and then averaged with respect to the draws. This provides information regarding the distribution of allocative inefficiency for a typical firm conditional on the data, and can be used for predictive purposes. Parameter uncertainty is fully taken into account in these computations in standard Bayesian fashion since these measures are averaged against the posterior distribution of parameters. The implication, of course, is that we do not have to resort to asymptotic approximations of the "plug-in" variety. Similar principles are followed to compute the firm-specific measures for $\ln C^{AL}$.

Further information can be obtained from Bayesian inferences regarding whether certain inputs are under- or over- utilized. This information is not provided by the ξ's alone. Since $\partial C^* / \partial w_j = x_j^{TE}$ and $\partial C^0 / \partial w_j = x_j^{OP}$ where x_j^{TE} and x_j^{OP} denote technically efficient and optimal (both technically and allocatively efficient) quantities of input x_j (with firm and time subscripts omitted) and $C^* = C^a \big|_{u=0}$, non-optimal use of input x_j relative to input x_1 (for example), can be obtained from the formula[19]

$$\kappa_j = (x_j / x_1)^{TE} / (x_j / x_1)^{OP} = \exp(-\xi_j)\left(1 + \eta_j / S_j^0\right) / \left(1 + \eta_1 / S_1^0\right) \text{ for } j = 2,...,m.$$

Consequently, if $\kappa_j > (<)$ 1 then input x_j is over- (under-) used relative to input x_1. Note that these measures are firm-specific and time-varying. Using the MCMC algorithm we have one draw for each of the κ_j. This draw depends on all parameters of the system. The final measure is obtained by averaging across all draws using the standard estimator. This operation is equivalent to integrating parameter uncertainty out.

It is clear that the model developed in this section provides all the information that we need to evaluate technical and allocative inefficiency for each firm in the sample using a cost system (consisting of the cost function and cost share equations). There are four basic advantages of the model we proposed in this section. First, technical and allocative inefficiency are modeled in a way that is consistent with the cost minimization problem of standard microeconomic theory. Second, all inferences are for the given data, so no asymptotic approximations are used. Third, using a systems approach ensures that we obtain more precise estimates of technical efficiency relative to a single equation approach. Finally, it solves the Greene problem from an empirical point of view.[20]

---

[19] Alternatively, one can define non-optimal use of input x_j from $\kappa_j = x_j^{TE} / x_j^{OP}$

$= \exp(-\xi_j)\left(1 + \eta_j / S_j^0\right) \sum_k (S_k^0 + \eta_k) \exp(-\eta_k)$ for $j = 1,...,m$. Thus, if $\kappa_j > (<)$ 1 then input x_j is over-used (under-used).

[20] An anonymous referee mentioned that a fixed effects model would also have these advantages. While this is true, we argue that there are situations (for example, cross-sectional models) where one cannot use a fixed effects model. Furthermore, in a fixed effects model technical inefficiency is defined relative to the best firm in the sample. This is, however, not the case in the present model.

4. Prior specifications

In our empirical application we use two priors (A and B). Both priors have $\underline{n} = 1$. Prior A has $\underline{q} = 0.1$ and prior B has $\underline{q} = 1$. Prior median efficiency is 71% and 41.8% respectively. We choose $\underline{n} = 1$ because $\underline{n}$ represents the number of observations in a fictitious experiment that provides a random sample $u_1, \dots u_{\underline{n}}$ with variance $\underline{q} / \underline{n}$. Therefore, prior information exists but is not particularly precise.

For the semi-informative prior to accommodate the restrictions implied by economic theory we assume $G\beta \sim N_q(0, \varepsilon^2 I_q)$ where $\varepsilon = 10^{-5}$. This prior is extremely tight and practically implies exact imposition of the restrictions. The exact form of the restrictions for this application is presented in Appendix B. Since these are mathematical restrictions to be satisfied by any cost function, we decided to impose these constraints exactly. Nonetheless, we argue that the proposed method allows one to use different degrees of correctness. We use informative priors for both Σ and Ω which are both inverted Wishart with $\nu_\Sigma = \nu_\Omega = 1$ degree of freedom and scale matrix $A_\Sigma = A_\Omega = 10^{-5} I$. These priors are proper but very diffuse.

5. Data and empirical results
5.1 Data

The data for this study is taken from the commercial bank and bank holding company database managed by the Federal Reserve Bank of Chicago. It is based on the Report of Condition and Income (Call Report) for all U.S. commercial banks that report to the Federal Reserve banks and the FDIC. In this paper we used the data for the years 1996-2000 and selected a random sample of 500 commercial banks.

The commercial banking industry is one of the largest and most important sectors of the US economy. The structure of the banking industry has undergone rapid changes in the last two decades, mostly due to extensive consolidation. Justification of mergers and acquisitions is often provided in terms of economies of scale and efficiency. Here we focus on the efficiency arguments by estimating a flexible cost system. Previous banking efficiency studies (see the survey by Berger and Humphrey (1997)) based on cost function estimation mostly focused on technical inefficiency. The reason for this is that estimation of the translog system with both technical and allocative inefficiency was not feasible before. From this perspective, this is the first banking study in which a translog cost function system with technical and

allocative inefficiency are jointly estimated without making them deterministic functions of data and unknown parameters.

In the banking literature there is controversy regarding the choice of inputs and outputs. Here we follow the intermediation approach (Kaparakis et al. (1994) in which banks are viewed as financial firms transforming various financial and physical resources into loans and investments. The output variables are: installment loans (to individuals for personal/household expenses) (y_1), real estate loans (y_2), business loans (y_3), federal funds sold and securities purchased under agreements to resell (y_4), other assets (assets that cannot be properly included in any other asset items in the balance sheet) (y_5). The input variables are: labor (x_1), capital (x_2), purchased funds (x_3), interest-bearing deposits in total transaction accounts (x_4) and interest-bearing deposits in total nontransaction accounts (x_5). For each input the price is obtained by dividing total expenses on it by the corresponding input quantity. Thus, for example, the price of labor (w_1) is obtained from expenses on salaries and benefits divided by the number of full time employees (x_1). The same approach is used to obtain w_2 through w_5. Total cost is then defined as the sum of the expenses on these five inputs. To impose the linear homogeneity restrictions, we normalize total cost and all the prices with respect to w_5.

5.2 Empirical results

5.2.1 Technical efficiency

The model for technical efficiency only with a single output is outlined in equation (1). Here we write it out fully for five outputs and five inputs using the translog functional form

$$\ln C_{it}^a = \alpha_0 + \sum_j \alpha_j \ln w_{j,it} + \sum_m \gamma_m \ln y_{m,it} + \tfrac{1}{2}\sum_m \sum_q \gamma_{mq} \ln y_{m,it} \ln y_{q,it} + \tfrac{1}{2}\sum_j \sum_k \beta_{jk} \ln w_{j,it} \ln w_{k,it}$$

$$+ \sum_j \sum_m \gamma_{jm} \ln w_{j,it} \ln y_{m,it} + \alpha_t t + \tfrac{1}{2}\alpha_{tt} t^2 + \sum_m \beta_{mt} \ln y_{m,it} t + \sum_j \beta_{jt} \ln w_{j,it} t + v_{1,it} + u_i$$

and

$$S_{j,it}^a = \alpha_j + \sum_k \beta_{jk} \ln w_{k,it} + \sum_m \gamma_{jm} \ln y_{m,it} + \beta_{jt} t + v_{j+1,it}, \quad m,q = 1,...,5; j,k = 1,...,4.$$

The assumptions on the noise components ($v_1,…,v_5$), and technical inefficiency component, u are as before and are not repeated here.

In Figure 1 we provide the posterior predictive technical efficiency as well as kernel density estimates for firm-specific technical efficiency measures for both priors A and B. We considered two models: (i) the translog cost system with only technical inefficiency, and (ii) the translog cost system with

both technical and allocative inefficiency. In Figures 1a and 1b we provide the posterior predictive technical efficiency distributions, which give the density function of technical efficiency for a typical or yet unobserved firm. The results from the two systems (with and without allocative inefficiency) are quite similar. This doesn't mean that allocative inefficiency can be ignored.[21] Since the overall cost efficiency is the product of technical and allocative efficiency, i.e., $OE = TE \cdot AE$ (Farrell (1957)), and both TE and AE are less than unity – the estimated TE in the technical inefficiency only model (where OE = TE by construction) is likely to be biased upward. In Figures 1c and 1d we report kernel densities of firm-specific efficiency measures. For each bank in the sample, its technical efficiency measure is the mean of the distribution of technical efficiency of that bank, conditional on its data, and unconditional on the parameters. In other words, parameter uncertainty is accounted for in estimating technical efficiency. Figures 1c and 1d present the kernel densities of these bank-specific efficiency means. Results from both models show that efficiency values below 70% are highly improbable.[22]

In Figure 2 we report efficiency rankings (technical) of banks from models with only technical inefficiency and with both technical and allocative inefficiency. Generally, the correlations between these rankings are fairly high but for some specific banks large differences are observed. Thus, if the focus is individual bank efficiency the high correlation of efficiency ranking between models may not be useful in choosing between models.

5.2.2. Allocative efficiency

We now discuss results obtained from the system with both technical and allocative inefficiency. The density functions of allocative inefficiency (price distortion) (ξ_j) are reported in Figure 3a and 3b, and some summary measures are reported in Table 1. Each graph[23] provides the kernel density estimate of bank-specific average of ξ_j. It may be useful to describe again how these density functions arise. For the i th bank in the sample the Gibbs sampler provides a random draw $\xi_{(i)j}^{(s)}$ during iteration $s = 1,...,S$ for input pair $(j,1)$. Recall that this is a draw from the distribution of $\xi_{(i)j}$ conditional on the data and all

---

[21] Note that ignoring allocative inefficiency, if any, makes the model misspecified that results wrong parameter estimates since the allocative inefficiency component in a translog cost function depends on outputs and input prices in a non-linear fashion.

[22] We examine convergence of our MCMC sampling schemes using Geweke's convergence diagnostics, as well as by running multiple chains starting from over-dispersed initial conditions and checking whether final marginal posteriors are close. Models with or without allocative inefficiency seemed very robust to initial conditions and passed convergence tests. Additionally, we have examined autocorrelation functions of posterior draws. To reduce the autocorrelation, we use the batching technique that is standard in the simulation literature. Batch means display practically zero autocorrelations. Results in graphical form are available upon request.

[23] We do not report posterior predictive distributions for allocative inefficiency measures since we are interested mostly in inferences regarding the banks in our sample.

other parameters of the model. What we need is the density function of $\xi_{(i)j}$ conditional on the data only.

To average out parameter uncertainty we compute $\overline{\xi}_{(i)j} = S^{-1}\sum_{s=1}^{S}\xi_{(i)j}^{(s)}$ which provides a price distortion

measure for each input and each bank. Figure 4 presents kernel densities of $\overline{\xi}_{(i)j}$ across banks for any

given input $j = 1,...,4$. These density functions do not vary widely with the prior of the technical

efficiency parameter σ_u. They are concentrated around zero (so we can claim that, banks on average do

not seem to have significant relative price distortions) but they differ in terms of spread and overall shape.

For labor (input 1), relative price distortions can be as large as 8% in absolute value. For other inputs, the

spread is much lower and distortions range from minus 4% to plus 4%. The difference in spreads reflects

the fact that for labor (relative to input 5) banks seem to misperceive prices to a greater extent compared

to other inputs. Finally, the fact that these density functions are not particularly tight means that banks are

quite heterogeneous in terms of allocative inefficiency.

The density functions of κ_j (relative over- (under-) use of each input) are reported in Figures 3c

and 3d, and their summary measures are reported in Table 1. These density functions are kernel density

estimates of bank-specific measures derived following the formula given in the previous paragraph. These

density functions are mostly centered around unity meaning that, on average, banks do not make many

allocative mistakes in using their inputs. The considerable spread suggests that banks are highly

heterogeneous in their relative input mis-allocation. More specifically, κ_1 ranges from 0.836 to 1.182

suggesting that banks may under-utilize labor (relative to input 5) by as much as 17.4% or over-utilize it

by as much as 18.2%. The remaining κ_j ranges roughly from 0.92 to 1.08 suggesting under-utilization

by 8% and over-utilization by 8%. Thus we observe the presence of considerable allocative inefficiency

in the sample banks.

When banks fail to allocate their inputs properly, costs will increase. We label this as the cost of

allocative inefficiency. In Figure 4 we provide kernel densities of bank-specific measures of percentage

increases in cost due to allocative inefficiency, $\ln C^{AL}$. The density functions of $\ln C^{AL}$ are highly skewed

to the right. On average, allocative inefficiency increased cost by 10% (meaning that on average, cost

allocative efficiency is 90%), although there are fewer banks for which this is much higher. From the

practical point of view optimal use of inputs is driven by the motive of attaining high cost efficiency.

Since some inputs are costlier than others a relatively cheap (expensive) input can be over- (under-) used

more than some other inputs that are relatively costly. Here we show how to obtain information on

allocative inefficiency (defined in terms of price distortion), non-optimal input use and finally the cost of

non-optimal input use, along with cost of technical inefficiency.

We have tried several prior tightness parameters for the regression parameters, β, viz., ε ranging from 10^{-12} to 0.1 without noticing large differences in final results related to technical and allocative inefficiency. We have also tried larger values for the prior tightness parameter. In these cases the results started to differ notably. This type of outcome is quite reasonable. With small values of the prior tightness we can interpret the model as a cost-share system, at least in an approximate sense. We necessarily lose this interpretation as the value of prior tightness increases. In this case the interpretation of technical efficiency becomes ambiguous so it should be expected that the results become more sensitive on the prior.

6. Conclusions

In this paper we developed Bayesian tools for making inferences on firm-specific technical and allocative inefficiency using a system approach. The system considered here is based on the cost minimization behavior of producers. The main contribution of the paper is the estimation of a well-specified translog system (in which the error terms in the cost and cost-share equations are internally consistent) in a random effects framework. This solves the Greene problem by using a model that is theoretically consistent and estimating it without treating the inefficiencies as parametric functions of the data and unknown parameters. First, we analyzed the model with only technical inefficiency and then we introduced allocative inefficiency. The model with only technical inefficiency is a standard seemingly unrelated regressions system conditional on the latent inefficiency variable. The model with both technical and allocative inefficiency is a nonlinear seemingly unrelated regression with nonlinear random effects. We showed that simulation-based numerical Bayesian analysis can be used to provide inferences on parameters and more importantly on functions of interest, viz., technical efficiency, allocative inefficiency, price distortions for each input, non-optimal input use for each input, etc., for each firm. The new techniques are applied to a panel of U.S. banks. We compared the results obtained from two system approaches, namely with and without allocative inefficiency. Results show some important differences in efficiency estimates across models.

References

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

Atkinson, S. E., and C. Cornwell, 1994, Parametric estimation of technical and allocative inefficiency with panel data, International Economic Review 35, 231-244.

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

Berger, A. N. and D. B. Humphrey, 1997, Efficiency of financial institutions: International survey and directions for future research, European Journal of Operational Research 98, 175-212.

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

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

Chen, L., Z. Qin, and J.S. Liu, 2000, Exploring hybrid Monte Carlo in Bayesian computation, in: E.I. George, ed., Bayesian Methods with Applications to Science, Policy, and Official Statistics, Selected Papers from ISBA 2000: The Sixth World Meeting of the International Society for Bayesian Analysis, 71-80.

Christensen, L. R., and W.H. Greene, 1976, Economies of scale in U. S. electric power generation, Journal of Political Economy 84, 655-76.

Diewert, W.E., and T.J. Wales, 1987, Flexible functional forms and global curvature conditions, Econometrica 55, 43-68.

Diewert, W.E., 1982, Duality approaches to microeconomic theory, in: K.J. Arrow and M.D. Intriligator, eds., Handbook of Mathematical Economics, Vol II (North-Holland, Amsterdam).

Diggle, P.J., K.Y. Liang, and S.L. Zeger, 1994, Analysis of longitudinal data (Clarendon Press, Oxford).

Farrell, M. J., 1957, The measurement of productive efficiency, Journal of the Royal Statistical Society, Series A, 120, 253-81.

Fernandez, C., G. Koop, and M.F.J. Steel, 2000, A Bayesian analysis of multiple output stochastic frontiers, Journal of Econometrics 98, 47-79.

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

Gelfand, A..E. and A.F.M. Smith, 1990, Sampling based approaches to calculating marginal densities, Journal of the American Statistical Association 85, 398-409.

Geweke, J., 1993, Bayesian treatment of the independent Student-t linear model, Journal of Applied Econometrics 8, S19-S40.

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

Greene, W.H., 1980, On the estimation of a flexible frontier production model, Journal of Econometrics 13:1, 101-15.

Greene, W.H., 1993, The econometric approach to efficiency analysis, in: H.O. Fried, C.A.K. Lovell and S.S. Schmidt, eds., The measurement of productive efficiency: Techniques and applications (Oxford University Press, Oxford) 68-119.

Greene, W.H., 2001, New developments in the estimation of stochastic frontier models with panel data, (Department of Economics, Stern School of Business, New York University, NY).

Griffiths, W.E., 2001, Bayesian inference in the seemingly unrelated regressions models, in D. Giles, ed., Computer-Aided Econometrics, Marcel Dekker (forthcoming).

Kalirajan, K. P., 1990, On Measuring Economic Efficiency, Journal of Applied Econometrics 5, 75-85.

Kaparakis, E.I., S.M. Miller, and A. Noulas, 1994, Short-run cost-inefficiency of commercial banks: A flexible stochastic frontier approach, Journal of Money, Credit and Banking 26, 875-893.

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

Koop, G. and M.F.J. Steel, 2001, Bayesian analysis of stochastic frontier models, in: Baltagi B., ed., A Companion to Theoretical Econometrics (Blackwell, Oxford) 520-573.

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

Kumbhakar, S.C., 1987, The specification of technical and allocative inefficiency in stochastic production and profit frontiers, Journal of Econometrics 34, 335-48.

Kumbhakar, S.C., 1997, Modeling allocative inefficiency in a translog cost function and cost share equations: An exact relationship, Journal of Econometrics 76, 351-356.

Kumbhakar, S.C. and C.A.K Lovell, 2000, Stochastic frontier analysis (Cambridge University Press, New York).

Liu, Q. and D.A. Pierce, 1994, A note on Gauss-Hermite quadrature, Biometrika 81, 624-629.

Longford, N.T., 1994, Logistic regression with random coefficients, Computational Statistics and Data Analysis 17, 1-15.

Maietta, O.W., 2002, The decomposition of cost efficiency into technical and allocative efficiency with panel data of Italian dairy farms, European Review of Agricultural Economics, 27, 473-495.

McCulloch, C.E., 1994, Maximum likelihood variance components estimation for binary data, Journal of the American Statistical Association 89, 330-335.

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

Pinheiro, J.C. and D.M. Bates, 1995, Approximations to the log-likelihood function in the nonlinear mixed-effects model, Journal of Computational and Graphical Statistics 4, 12-35.

Roberts, G.O. and A.F.M. Smith, 1994, Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms, Stochastic Processes and their Applications 49, 207-216.

Schmidt, P., and C.A.K. Lovell, 1979, Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers, Journal of Econometrics 9, 343-66.

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

Tierney, L., 1994, Markov chains for exploring posterior distributions (with discussion), Annals of Statistics 22, 1701-1762.

Tsionas, E.G., 1999, Full likelihood inference in normal-gamma stochastic frontier models, Journal of Productivity Analysis 13, 179-201.

Vonesh, E.F. and V.M. Chinchilli, 1997, Linear and nonlinear models for the analysis of repeated measurements (Marcel Dekker, New York).

Wolfinger, R.D., 1993, Laplace's approximation for nonlinear mixed models, Biometrika 80, 791-795.

Wolfinger, R.D. and X. Lin, 1997, Two Taylor-series approximation methods for nonlinear mixed models, Computational Statistics and Data Analysis 25, 465-490.

Zellner, A., 1971, Introduction to Bayesian inference in econometrics (Wiley, New York).

APPENDIX A. Random draws from conditional posterior density function of ξ

In this Appendix, we show how to derive the normal approximation to the conditional posterior of ξ, and then use a Metropolis update to maintain the correct posterior density function. We assume that u is known and we set it to zero without loss of generality (meaning, it is subtracted from y_1). In general, any nonlinear system with random effects can be written as

$$y_{1,it} = f_{1,it}(\xi_i; \beta) + v_{1,it}$$
$$\vdots$$
$$y_{M,it} = f_{M,it}(\xi_i; \beta) + v_{M,it}, \quad i = 1,..,n; \; t = 1,..,T$$

where $\xi_i \sim N_{M-1}(0, \Omega)$ and $v_{it} \equiv [v_{1,it}, ..., v_{M,,it}]' \sim N_M(0, \Sigma)$ independently, and $f_{m,it}$ denotes a given nonlinear function for the mth equation, firm i and year t. We can take a first-order Taylor expansion with respect to ξ, consider all observations for the i th firm, and write the system in obvious notation as

$$y_{1,i} = f_{1,i}(0;\beta) + Z_{1,i}\xi_i + v_{1,i}$$

$$\vdots$$

$$y_{M,i} = f_{M,i}(0;\beta) + Z_{M,i}\xi_i + v_{M,i}$$

where $\left[v'_{1,i},...,v'_{M,i}\right]' \sim N_M\left(0, \Sigma \otimes I_T\right)$, and $Z_{1,i},...,Z_{M,i}$ are matrices representing the first derivatives with respect to ξ evaluated at the point of approximation. Therefore, $Z_{m,i} = \dfrac{\partial f_{m,i}}{\partial \xi_i}\bigg|_{\xi=0}$ whose dimension is $T \times (M-1)$. Apparently, the $Z_{m,i}$'s are functions of the data as well as β. If we combine this linear system with the distributional assumption about ξ_i we have

$$0_{M-1} = I_{M-1}\xi_i + e_i, \ e_i \sim N_{M-1}(0,\Omega)$$

Therefore, we can use standard results from mixed estimation to obtain

$$\xi_i \sim N_{M-1}\left(\tilde{\xi}_i, \left[Z_i'\left(\Sigma^{-1} \otimes I_T\right)Z_i + \Omega^{-1}\right]^{-1}\right)$$

where

$$\tilde{\xi}_i = \left[Z_i'\left(\Sigma^{-1} \otimes I_T\right)Z_i + \Omega^{-1}\right]^{-1} Z_i'\left(\Sigma^{-1} \otimes I_T\right)\zeta_i, \ i=1,..,n; \zeta_i = \left[\zeta_{1,i},...,\zeta_{M,i}\right]';$$

$$\zeta_{m,i} = y_{m,i} - f_{m,i}(0;\beta), \ m=1,..,M \ \text{and} \ Z_i = \left[Z_{1,i}'...Z_{M,i}'\right]'.$$

This represents the normal approximation to the posterior conditional density function of any nonlinear system with time-invariant random effects. In practice instead of a normal density function we use a Student-t with 10 degrees of freedom.

These general results must now be specialized to the translog cost-share system we analyze. The task is to find the first derivatives of the cost function and share equations with respect to the ξ's, and evaluate them at $\xi = 0_{M-1}$. Omitting i,t subscripts and error terms for simplicity, the cost function is

$$\ln C^a = \ln C^0 + \ln C^{AL}$$

$$\ln C^{AL} = \ln G + \sum_j \alpha_j \xi_j + \sum_j \gamma_{jy}\xi_j \ln y + \sum_j\sum_k \beta_{jk}\xi_j \ln w_k + \tfrac{1}{2}\sum_j\sum_k \beta_{jk}\xi_j\xi_k$$

and $\ln C^0$ is the usual translog cost function. Clearly, assuming all restrictions implied by the theory in place, we get

$$\frac{\partial \ln C^{AL}}{\partial \xi_j} = \frac{\partial \ln G}{\partial \xi_j} + \alpha_j + \gamma_{jy}\ln y + \sum_k \beta_{jk}\ln w_k + \sum_k \beta_{jk}\xi_k$$

Since

$$G = \sum_l S_l^* \exp(-\xi_l), \text{ where } S_l^* = \alpha_l + \sum_k \beta_{lk}(\ln w_k + \xi_k) + \gamma_{ly} \ln y$$

we obtain

$$\frac{\partial \ln G}{\partial \xi_j} = G^{-1}\left\{\sum_l \beta_{lj} \exp(-\xi_l) - \exp(-\xi_j)S_j^*\right\}$$

Therefore, $\left.\dfrac{\partial \ln C^{AL}}{\partial \xi_j}\right|_{\xi=0} = 0$. Since $\left.\ln C^{AL}\right|_{\xi=0} = 0$, the cost function contributes nothing to the conditional posterior of ξ up to a first order of approximation. This is particularly important because the cost function is the most complicated function of the system, and omitting the cost function from further consideration results in computational gain. Next, we consider the share equations. These are given by

$$S_m^a = S_m^0 + \eta_m$$

$$\eta_m = \frac{S_m^0\{1 - G\exp(\xi_m)\} + \sum_k \beta_{mk}\xi_k}{G\exp(\xi_m)} \quad \text{for } m = 1,..,M-1.$$

Clearly, $\left.\eta_l\right|_{\xi=0} = 0$, and $\left.G\right|_{\xi=0} = \sum_k \left.S_k^*\right|_{\xi=0} = 1$. Moreover, it is easy to show that

$$\left.\frac{\partial[G\exp(\xi_m)]}{\partial \xi_j}\right|_{\xi=0} = \left(S_j^0\right)^2 + \beta_{jj} \text{ if } m = j, \text{ and } S_m^0\left(1 + S_j^0\right) + \beta_{mj}, \text{ if } m \neq j.$$

After some algebra, the derivatives of the allocative inefficiency term with respect to ξ's are

$$\left.\frac{\partial \eta_m}{\partial \xi_j}\right|_{\xi=0} = -S_j^0\left(1 - S_j^0\right) + \beta_{jj} \text{ if } m = j, \text{ and } S_m^0 S_j^0 + \beta_{mj}, \text{ if } m \neq j.$$

These partial derivatives are simple functions of the data and β, and can be computed very easily at no cost conditional on the $\beta's$. Therefore, we can set up the matrices Z_i, and these matrices can, in turn, be used to obtain a draw from the approximate multivariate Student-t posterior conditional density function of ξ_i's.

Appendix B. Prior restrictions imposed by economic theory

Consider the translog cost function:

$$\ln C = \alpha_0 + \alpha' \ln w + \beta' \ln q + \tfrac{1}{2}\ln w' A \ln w + \tfrac{1}{2}\ln q' B \ln q + \ln w' D \ln q + \gamma_t t + \tfrac{1}{2}\gamma_{tt}t^2$$
$$+ \delta' \ln q \cdot t + \theta' \ln w \cdot t \tag{B.1}$$

where $\ln w$ is the $m \times 1$ vector of log prices, $\ln q$ is the $s \times 1$ vector of log outputs, α_0 is a constant, α, θ and β, δ are $m \times 1$ and $s \times 1$ vectors, and A, B, D are matrices of dimensions $m \times n$, $s \times s$ and $m \times s$ respectively. The share equations will be in the form

$$S = m + F \ln w + G \ln q + \lambda t \tag{B.2}$$

where S is the $m \times 1$ vector of input shares, F, G are matrices of dimensions $m \times m$ and $m \times s$ respectively, and λ is an $m \times 1$ vector.

We impose exactly the restrictions that A and B are symmetric. Homogeneity implies

$$1'_m A = 0, \ 1'_S D = 0, \ 1'_m \alpha = 1$$

where 1_m denotes the $m \times 1$ unit vector. Moreover, we have the cross-equation restrictions

$$m = \alpha, \ F = a, \ G = D, \ \theta = \lambda$$

In the system (B.1)-(B.2) we have a total of 110 unrestricted parameters; provided we impose exactly the symmetry restrictions on A and B as well as homogeneity we have to account for the 44 cross-equation restrictions in (D.6)-(D.9). Given our conventions, if γ denotes the 110×1 unrestricted parameter vector the restrictions are as follows:

$\gamma_{67} - \gamma_2 = 0$	$\gamma_{89} - \gamma_4 = 0$
$\gamma_{68} - \gamma_7 = 0$	$\gamma_{90} - \gamma_9 = 0$
$\gamma_{69} - \gamma_8 = 0$	$\gamma_{91} - \gamma_{13} = 0$
$\gamma_{70} - \gamma_9 = 0$	$\gamma_{92} - \gamma_{16} = 0$
$\gamma_{71} - \gamma_{10} = 0$	$\gamma_{93} - \gamma_{17} = 0$
$\gamma_{72} - \gamma_{11} = 0$	$\gamma_{94} - \gamma_{18} = 0$
$\gamma_{73} - \gamma_{42} = 0$	$\gamma_{95} - \gamma_{52} = 0$
$\gamma_{74} - \gamma_{43} = 0$	$\gamma_{96} - \gamma_{53} = 0$
$\gamma_{75} - \gamma_{44} = 0$	$\gamma_{97} - \gamma_{54} = 0$
$\gamma_{76} - \gamma_{45} = 0$	$\gamma_{98} - \gamma_{55} = 0$
$\gamma_{77} - \gamma_{46} = 0$	$\gamma_{99} - \gamma_{56} = 0$
$\gamma_{78} - \gamma_3 = 0$	$\gamma_{100} - \gamma_5 = 0$
$\gamma_{79} - \gamma_8 = 0$	$\gamma_{101} - \gamma_{10} = 0$
$\gamma_{80} - \gamma_{12} = 0$	$\gamma_{102} - \gamma_{14} = 0$
$\gamma_{81} - \gamma_{13} = 0$	$\gamma_{103} - \gamma_{17} = 0$
$\gamma_{82} - \gamma_{14} = 0$	$\gamma_{104} - \gamma_{19} = 0$
$\gamma_{83} - \gamma_{15} = 0$	$\gamma_{105} - \gamma_{20} = 0$
$\gamma_{84} - \gamma_{47} = 0$	$\gamma_{106} - \gamma_{57} = 0$
$\gamma_{85} - \gamma_{48} = 0$	$\gamma_{107} - \gamma_{58} = 0$
$\gamma_{86} - \gamma_{49} = 0$	$\gamma_{108} - \gamma_{59} = 0$
$\gamma_{87} - \gamma_{50} = 0$	$\gamma_{109} - \gamma_{60} = 0$
$\gamma_{88} - \gamma_{51} = 0$	$\gamma_{110} - \gamma_{61} = 0$

Since all restrictions are linear, they can be put in the form $G\gamma = g$ where G is 44×110, and g is 44×1. Based on this formulation a semi-informative prior can be specified in the form $G\gamma \sim N_{44}(g, H)$ where H is 44×44.

Table 1. Posterior results for functions of interest*

	Prior A	*Prior B*
Posterior predictive technical efficiency	0.963 (0.032)	0.963 (0.035)
Firm-specific technical efficiency	0.868 (0.247)	0.859 (0.257)
$\ln C^{AL}$	0.099 (0.088)	0.098 (0.088)
ξ_1	-0.0016 (0.033)	-0.0016 (0.033)
ξ_2	-0.0056 (0.011)	-0.0055 (0.011)
ξ_3	0.0029 (0.024)	0.0028 (0.024)
ξ_4	0.0011 (0.008)	0.0009 (0.008)
κ_1	1.0038 (0.059)	1.0038 (0.058)
κ_2	1.017 (0.026)	1.017 (0.027)
κ_3	0.991 (0.035)	0.991 (0.034)
κ_4	0.996 (0.025)	0.996 (0.025)

* The entries are the posterior means. Posterior standard deviations appear in parentheses.

Figure 1a. Posterior predictive technical efficiency distribution (prior A)
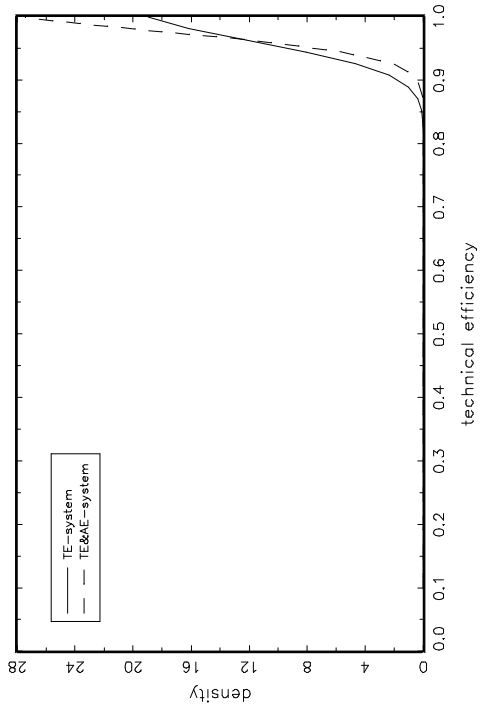
Figure 1b. Posterior predictive technical efficiency distribution (prior B)
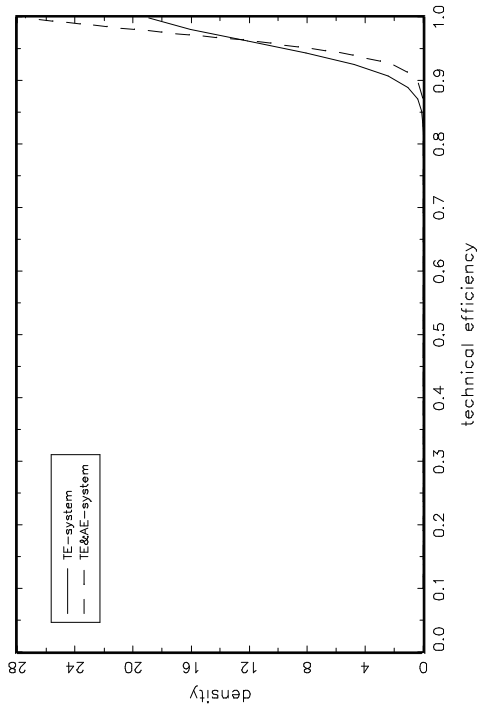
Figure 1c. Technical efficiency distribution (prior A)
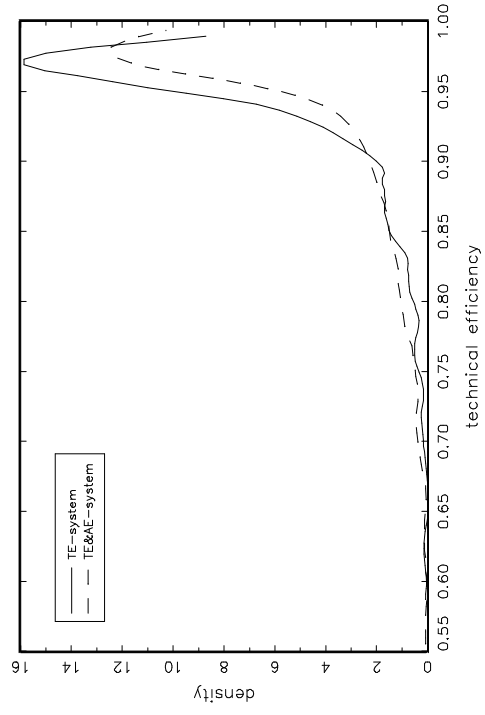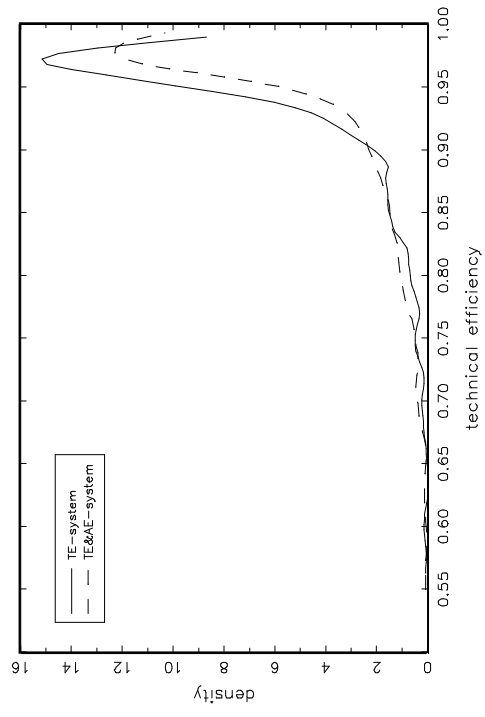
Figure 1d. Technical efficiency distribution (prior B)

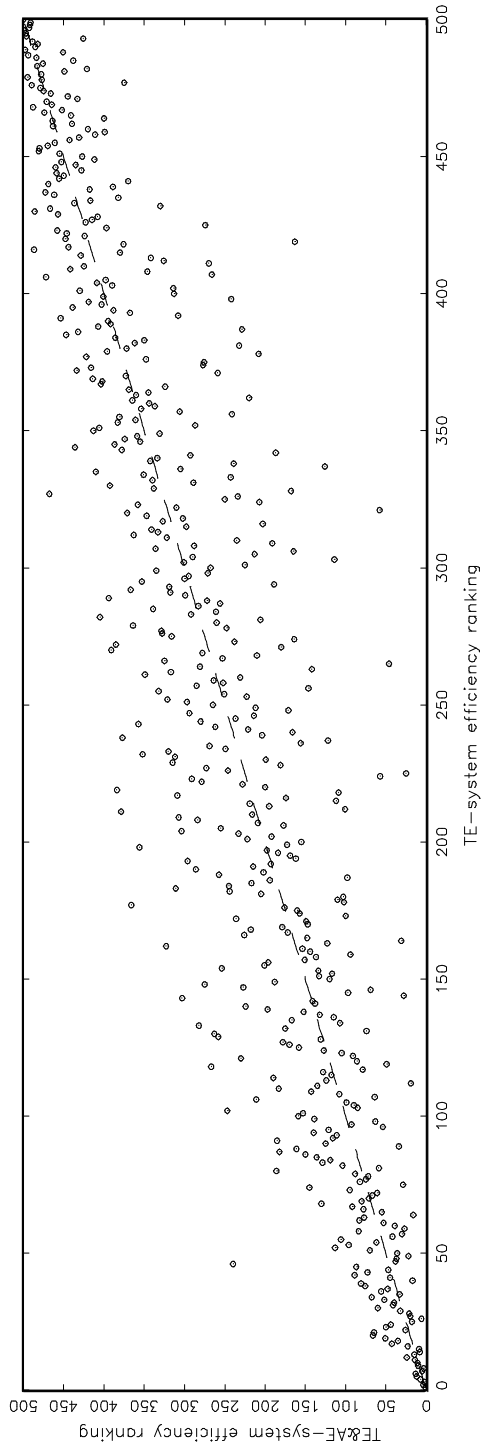Figure 2a. Comparison of efficiency rankings, prior A

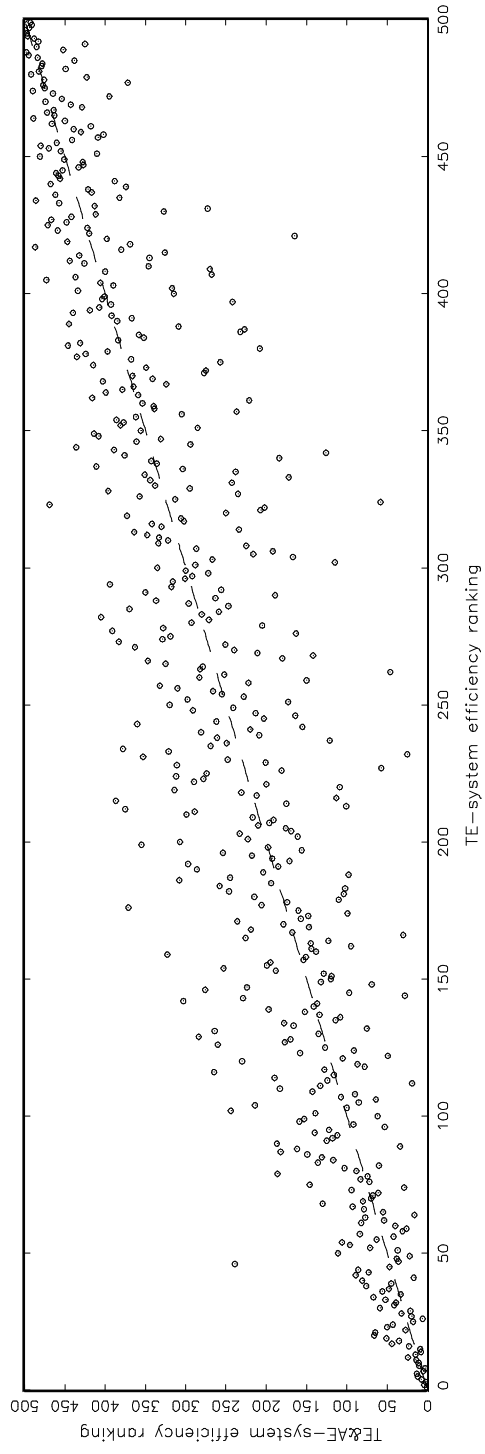Figure 2b. Comparison of efficiency rankings, prior B

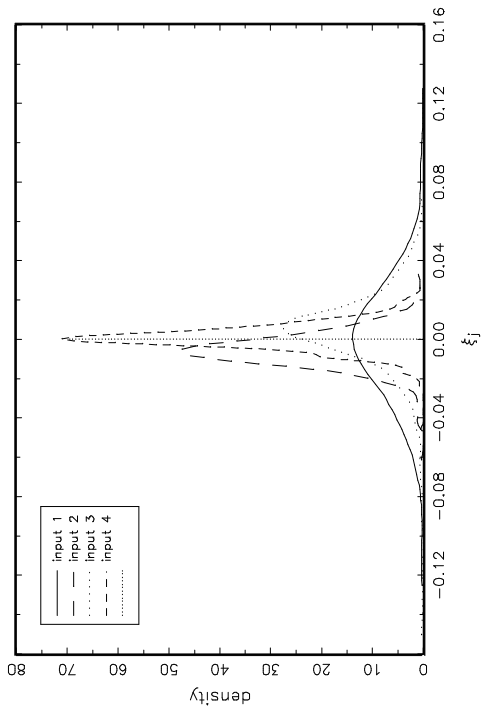Figure 3a. Distribution of price distortions, prior A
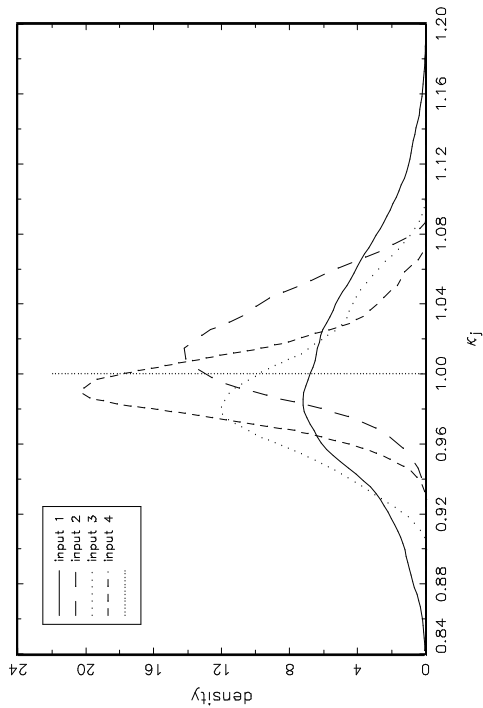
Figure 3b. Distribution of price distortions, prior B

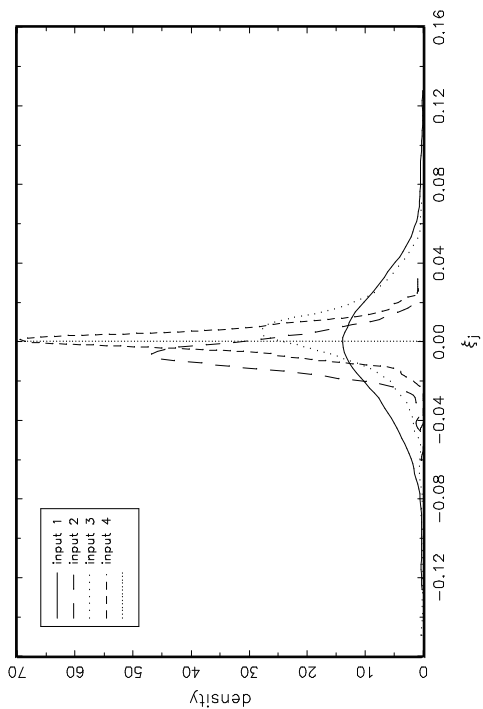Figure 3c. Distribution of allocative distortions, prior A

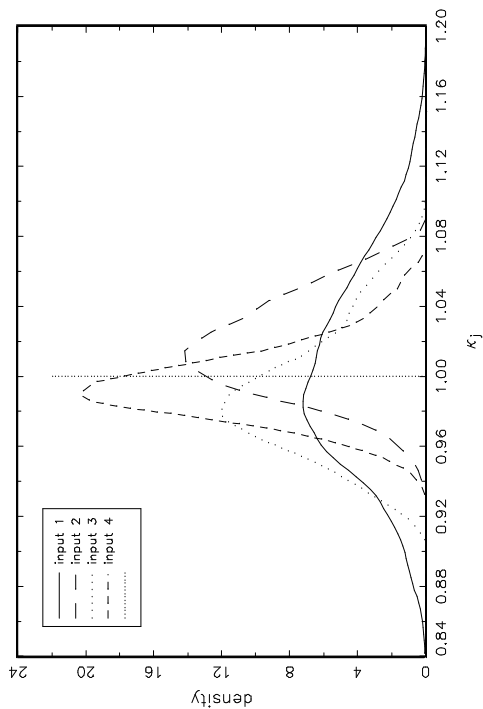Figure 3d. Distribution of allocative distortions, prior B

Figure 4. Distribution of cost of allocative inefficiency