Review Article

Some Recent Developments in Efficiency Measurement in Stochastic Frontier Models

Subal C. Kumbhakar[1] and Efthymios G. Tsionas[2]

[1] *Department of Economics, State University of New York, Binghamton, NY 13902, USA*
[2] *Department of Economics, Athens University of Economics and Business, 76 Patission Street,
   104 34 Athens, Greece*

Correspondence should be addressed to Subal C. Kumbhakar, kkar@binghamton.edu

This paper addresses some of the recent developments in efficiency measurement using stochastic frontier (SF) models in some selected areas. The following three issues are discussed in details. First, estimation of SF models with input-oriented technical efficiency. Second, estimation of latent class models to address technological heterogeneity as well as heterogeneity in economic behavior. Finally, estimation of SF models using local maximum likelihood method. Estimation of some of these models in the past was considered to be too difficult. We focus on the advances that have been made in recent years to estimate some of these so-called difficult models. We complement these with some developments in other areas as well.

1. Introduction

In this paper we focus on three issues. First, we discuss issues (mostly econometric) related to input-oriented (IO) and output-oriented (OO) measures of technical inefficiency and talk about the estimation of production functions with IO technical inefficiency. We discuss implications of the IO and OO measures from both the primal and dual perspectives. Second, the latent class (finite mixing) modeling approach is extended to accommodate behavioral heterogeneity. Specifically, we consider profit- (revenue-) maximizing and cost-minimizing behaviors with technical inefficiency. In our mixing/latent class model, first we consider a system approach in which some producers maximize profit while others simply minimize cost, and then we use a distance function approach, and mix the input and output distance functions (in which it is assumed, at least implicitly, that some producers maximize revenue while others minimize cost). In the distance function approach the behavioral assumptions are not explicitly taken into account. The prior probability in favor of profit (revenue) maximizing behavior is assumed to depend on some exogenous variables. Third, we consider stochastic frontier (SF) models that are estimated using local maximum likelihood (LML)

method to address the flexibility issue (functional form, heteroskedasticity, and determinants of technical inefficiency).

2. The IO and OO Debate

The technology (with or without inefficiency) can be looked at from either a primal or a dual perspective. In a primal setup two measures of technical efficiency are mostly used in the efficiency literature. These are (i) input-oriented (IO) technical inefficiency and (ii) output oriented (OO) technical inefficiency.[1] There are some basic differences between the IO and OO models so far as features of the technology are concerned. Although some of these differences and their implications are well-known except for Kumbhakar and Tsionas [1], no one has estimated a stochastic production frontier model econometrically with IO technical inefficiency using cross-sectional data.[2] Here we consider estimation of a translog production model with IO technical inefficiency.

2.1. The IO and OO Models

Consider a single output production technology where Y is a scalar output and X is a vector of inputs. Then the production technology with the IO measure of technical inefficiency can be expressed as

$$Y_i = f(X_i \cdot \Theta_i), \quad i = 1, \ldots, n, \tag{2.1}$$

where Y_i is a scalar output, $\Theta_i \leq 1$ is IO efficiency (a scalar), X_i is the $J \times 1$ vector of inputs, and i indexes firms. The IO technical inefficiency for firm i is defined as $\ln \Theta_i \leq 0$ and is interpreted as the rate at which all the inputs can be reduced without reducing output. On the other hand, the technology with the OO measure of technical inefficiency is specified as

$$Y_i = f(X_i) \cdot \Lambda_i, \tag{2.2}$$

where $\Lambda_i \leq 1$ represents OO efficiency (a scalar), and $\ln \Lambda_i \leq 0$ is defined as OO technical inefficiency. It shows the percent by which actual output could be increased without increasing inputs (for more details, see Figure 1).

It is clear from (2.1) and (2.2) that if $f(\cdot)$ is homogeneous of degree r then $\Theta_i^r = \Lambda_i$, that is, independent of X and Y. If homogeneity is not present their relationship will depend on the input quantities and the parametric form of $f(\cdot)$.

We now show the IO and OO measures of technical efficiency graphically. The observed production plan (Y, X) is indicated by the point A. The vertical length AB measures OO technical inefficiency, while the horizontal distance AC measures IO technical inefficiency. Since the former measures percentage loss of output while the latter measures percentage increase in input usage in moving to the production frontier starting from the inefficient production plan indicated by point A, these two measures are, in general, not directly comparable. If the production function is homogeneous, then one measure is a constant multiple of the other, and they are the same if the degree of homogeneity is one. In the more general case, they are related in the following manner: $f(X) \cdot \Lambda = f(X\Theta)$.

Although we consider technologies with a single output, the IO and OO inefficiency can be discussed in the context of multiple output technologies as well.
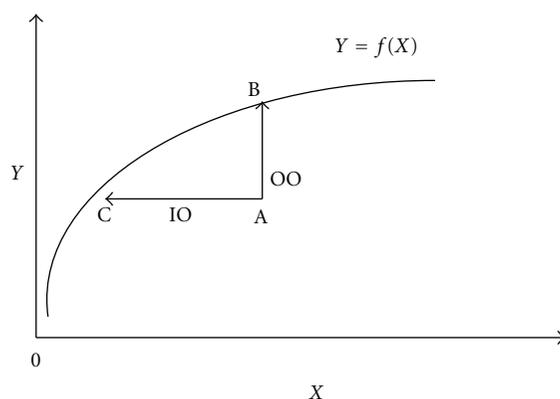
Figure 1: IO and OO technical inefficiency.

2.2. Economic Implications of the IO and OO Models

Here we ask two questions. First, does it matter whether one uses the IO or the OO representation so far as estimation of the technology is concerned? That means, whether features of the estimated technology such as elasticities, returns to scale, and so forth, are invariant to the choice of efficiency orientation. Second, are efficiency rankings of firms invariant to the choice of efficiency orientation? That is, does one get the same efficiency measures (converted in terms of either output loss or increase in costs) in both cases? It is not possible to provide general theoretical answers to these questions. These are clearly empirical issues so it is necessary to engage in applied research to get a feel for the similarities and differences of the two approaches.

Answers to these questions depend on the form of the production technology. If it is homogeneous, then there is no difference between these two models econometrically. This is because for a homogeneous function $r \ln \Theta_i = \ln \Lambda_i$, where r is the degree of homogeneity. Thus, rankings of firms with respect to $\ln \Lambda_i$ and $\ln \Theta_i$ will be exactly the same (one being a constant multiple of the other). Moreover, since $f(X) \cdot \Lambda = f(X)\Theta^r$, the input elasticities as well as returns to scale measures based on these two specifications of the technology will be the same.[3]

This is, however, not the case if the technology is nonhomogenous. In the OO model the elasticities and returns to scale will be independent of the technical inefficiency because technical efficiency (i.e., assumed to be independent of inputs) enters multiplicatively into the production function. This is not true for the IO model, where technical inefficiency enters multiplicatively with the inputs. This will be shown explicitly later for a nonhomogeneous translog production function.

2.3. Econometric Modeling and Efficiency Measurement

Using the lower case letters to indicate the log of a variable, and assuming that $f(\cdot)$ has a translog form the IO model can be expressed as

$$
\begin{aligned}
y_i = \beta_0 + (x_i - \theta_i 1_J)'\beta + \frac{1}{2}(x_i - \theta_i 1_J)'\Gamma(x_i - \theta_i 1_J) + \beta_T T_i + \frac{1}{2}\beta_{TT} T_i^2 \\
+ T_i(x_i - \theta_i 1_J)'\varphi + v_i, \quad i = 1, \ldots, n,
\end{aligned}
\tag{2.3}
$$

where y_i is the log of output, 1_J denotes the $J \times 1$ vector of ones, x_i is the $J \times 1$ vector of inputs in log terms, T_i is the trend/shift variable, β_0, β_T and β_{TT} are scalar parameters, β, φ are $J \times 1$ parameter vectors, Γ is a $J \times J$ symmetric matrix containing parameters, and v_i is the noise term. To make θ nonnegative we defined it as $-\ln \Theta = \theta$.

We rewrite the IO model above as

$$y_i = \left(\beta_0 + x_i'\beta + \frac{1}{2}x_i'\Gamma x_i + \beta_T T_i + \frac{1}{2}\beta_{TT}T_i^2 + x_i'\varphi T_i \right) - g(\theta_i, x_i) + v_i, \quad i = 1, \ldots, n, \qquad (2.4)$$

where $g(\theta_i, x_i) = -[(1/2)\theta_i^2 \Psi - \theta \Xi_i]$, $\Psi = 1_J'\Gamma 1_J$, and $\Xi_i = 1_J'(\beta + \Gamma x_i + \varphi T_i)$, $i = 1, \ldots, n$. Note that if the production function is homogeneous of degree r, then $\Gamma 1_J = 0, 1_J'\beta = r$, and $1_J'\varphi = 0$. In such a case the $g(\theta_i, x_i)$ function becomes a constant multiple of θ, (namely, $[(1/2)\theta_i^2 \Psi - \theta \Xi_i] = -r\theta_i$), and consequently, the IO model cannot be distinguished from the OO model. The $g(\theta_i, x_i)$ function shows the percent by which output is lost due to technical inefficiency. For a well-behaved production function $g(\theta_i, x_i) \geq 0$ for each i.

The OO model, on the other hand, takes a much simpler form, namely,

$$y_i = \left(\beta_0 + x_i'\beta + \frac{1}{2}x_i'\Gamma x_i + \beta_T T_i + \frac{1}{2}\beta_{TT}T_i^2 + x_i'\varphi T_i \right) - \lambda_i + v_i, \quad i = 1, \ldots, n, \qquad (2.5)$$

where we defined $-\ln \Lambda = \lambda$ to make it nonnegative.[4] The OO model in this form is the one introduced by Aigner et al. [2] and Meeusen and van den Broeck [3], and since then it has been used extensively in the efficiency literature. Here we follow the framework used in Kumbhakar and Tsionas [1] when θ is random.[5]

We write (2.4) more compactly as

$$y_i = z_i'\alpha + \frac{1}{2}\theta_i^2 \Psi - \theta \Xi_i + v_i, \quad i = 1, \ldots, n. \qquad (2.6)$$

Both Ψ and Ξ_i are functions of the original parameters, and Ξ_i also depends on the data (x_i and T_i).

Under the assumption that $v_i \sim \mathbb{N}(0, \sigma^2)$ and θ_i is distributed independently of v_i with the density function $f(\theta_i; \omega)$, where ω is a parameter, the probability density function of y_i can be expressed as

$$f(y_i; \mu) = \left(2\pi\sigma^2 \right)^{-1/2} \int_0^\infty \exp\left[-\frac{(y_i - z_i'\alpha - (1/2)\theta_i^2 \Psi + \theta_i \Xi_i)^2}{2\sigma^2} \right] f(\theta_i; \omega)d\theta_i, \quad i = 1, \ldots, n, \qquad (2.7)$$

where μ denotes the entire parameter vector.

We consider a half-normal and an exponential specification for the density $f(\theta_i; \omega)$, namely,

$$f(\theta_i; \omega) = \left(\frac{\pi\omega^2}{2} \right)^{-1/2} \exp\left(-\frac{\theta_i^2}{2\omega^2} \right), \quad \theta_i \geq 0,$$

$$f(\theta_i; \omega) = \omega \exp(-\omega\theta_i), \quad \theta_i \geq 0. \qquad (2.8)$$

The likelihood function of the model is then

$$l(\mu; y, X) = \prod_{i=1}^{n} f(y_i; \mu),$$ (2.9)

where $f(y_i; \mu)$ has been defined above. Since the integral defining $f(y_i; \mu)$ is not available in closed form we cannot find an analytical expression for the likelihood function. However, we can approximate the integrals using a simulation as follows. Suppose $\theta_{i,(s)}$, $s = 1, \ldots, S$ is a random sample from $f(\theta_i; \omega)$. Then it is clear that

$$f(y_i; \mu) \approx \widetilde{f}(y_i; \mu) \equiv S^{-1} \sum_{s=1}^{S} \exp\left[-\frac{\left(y_i - z_i'\alpha - (1/2)\theta_{i,(s)}^2 \Psi + \theta_{i,(s)} \Xi_i \right)^2}{2\sigma^2} \right],$$ (2.10)

and an approximation of the log-likelihood function is given by

$$\log l \approx \sum_{i=1}^{n} \log \widetilde{f}(y_i; \mu),$$ (2.11)

which can be maximized by numerical optimization procedures to obtain the ML estimator. For the distributions we adopted, random number generation is trivial, so implementing the SML estimator is straightforward.[6]

Inefficiency estimation is accomplished by considering the distribution of θ_i conditional on the data and estimated parameters

$$f(\theta_i \mid \widetilde{\mu}, D_i) \propto \exp\left[-\frac{\left(y_i - z_i'\widetilde{\alpha} - (1/2)\theta_i^2 \widetilde{\Psi} + \Xi_i \theta_i \right)^2}{2\widetilde{\sigma}^2} \right] f(\theta_i; \widetilde{\omega}), \quad i = 1, \ldots, n,$$ (2.12)

where a tilde denotes the ML estimate, and $D_i = [x_i, T_i]$ denotes the data. For example, when $f(\theta_i; \widetilde{\omega})$ is half-normal we get

$$f(\theta_i \mid \widetilde{\mu}, y, X) \propto \exp\left[-\frac{\left(y_i - z_i'\widetilde{\alpha} - (1/2)\theta_i^2 \widetilde{\Psi} + \theta_i \Xi_i \right)^2}{2\widetilde{\sigma}^2} - \frac{\theta_i^2}{2\widetilde{\omega}^2} \right], \quad \theta_i \geq 0, \ i = 1, \ldots, n.$$ (2.13)

This is not a known density, and even the normalizing constant cannot be obtained in closed form. However, the first two moments and the normalizing constant can be obtained by numerical integration, for example, using Simpson's rule.

To make inferences on *efficiency*, define efficiency as $r_i = \exp(-\theta_i)$ and obtain the distribution of r_i and its moments by changing the variable from θ_i to r_i. This yields

$$f_r(r_i \mid \widetilde{\mu}, D_i) = r_i^{-1} f(-\ln r_i \mid \widetilde{\mu}, y, X), \quad 0 < r_i \leq 1, \ i = 1, \ldots, n.$$ (2.14)

The likelihood function for the OO model is given in Aigner et al. [2] (hereafter ALS).[7] The Maximum likelihood method for estimating the parameters of the production function in the OO models are straightforward and have been used extensively in the literature starting from ALS.[8] Once the parameters are estimated, technical inefficiency (λ) is estimated from $E(\lambda \mid (v - \lambda))$—the Jondrow et al. [4] formula. Alternatively, one can estimate technical efficiency from $E(e^{-\lambda} \mid (v - \lambda))$ using the Battese and Coelli [5] formula. For an application of this approach see Kumbhakar and Tsionas [1].

2.4. Looking Through the Dual Cost Functions

2.4.1. The IO Approach

We now examine the IO and OO models when behavioral assumptions are explicitly introduced. First, we examine the models when producers minimize cost to produce the given level of output(s). The objective of a producer is to

$$
\begin{aligned}
&\text{Min.}\quad w'X \\
&\text{subject to}\quad Y = f(X \cdot \Theta)
\end{aligned}
\tag{2.15}
$$

from which conditional input demand functions can be derived. The corresponding cost function can then be expressed as

$$
w'X = C^a = \frac{C(w, Y)}{\Theta},
\tag{2.16}
$$

where $C(w, Y)$ is the minimum cost function (cost frontier) and C^a is the actual cost. Finally, one can use Shephard's lemma to obtain $X_j^a = X_j^*(w, Y)/\Theta \geq X_j^*(w, Y)$ for all j, where the superscripts a and $*$ indicate actual and cost-minimizing levels of input X_j.

Thus, the IO model implies (i) a neutral shift in the cost function which in turn implies that RTS and input elasticities are unchanged due to technical inefficiency, (ii) an equiproportional increase (at the rate given by θ) in the use of all inputs due to technical inefficiency, irrespective of the output level and input prices.

To summarize, result (i) is just the opposite of what we obtained in the primal case (see [6]). Result (ii) states that when inefficiency is reduced firms will move horizontally to the frontier (as expected by the IO model).

2.4.2. The OO Model

Here the objective function is written as

$$
\begin{aligned}
&\text{Min.}\quad w'X \\
&\text{subject to}\quad Y = f(X) \cdot \Lambda
\end{aligned}
\tag{2.17}
$$

from which conditional input demand functions can be derived. The corresponding cost function can then be expressed as

$$w'X = C^a = C\left(w, \frac{Y}{\Lambda}\right) \equiv C(w, Y) \cdot q(w, Y, \Lambda), \tag{2.18}$$

where as, before, $C(w, Y)$ is the minimum cost function (cost frontier) and C^a is the actual cost. Finally, $q(\cdot) = C(w, Y/\Lambda)/C(w, Y) \geq 1$. One can then use Shephard's lemma to obtain

$$X_j^a = X_j^*(w, Y)\left[q(\cdot) + \left\{\frac{C(w, Y)}{X_j^*}\right\}\left\{\frac{\partial q(\cdot)}{\partial w_j}\right\}\right] \geq X_j^*(w, Y) \quad \forall j, \tag{2.19}$$

where the last inequality will hold if the cost function is well behaved. Note that $X_j^a \neq X_j^*(w, Y)$ for all j unless $q(\cdot)$ is a constant.

Thus, the results from the OO model are just the opposite from those of the IO model. Here (i) inefficiency shifts the cost function nonneutrally (meaning that $q(\cdot)$ depends on output and input prices as well as Λ; (ii) increases in input use are not equiproportional (depends on output and input prices); (iii) the cost shares are not independent of technical inefficiency, (iv) the model is harder to estimate (similar to the IO model in the primal case).[9]

More importantly, the result in (i) is just the opposite of what we reported in the primal case. Result (ii) is not what the OO model predicts (increase in output) when inefficiency is eliminated. Since output is exogenously given in a cost-minimizing framework, input use has to be reduced when inefficiency is eliminated.

The results from the dual cost function models are just the opposite of what the primal models predict. Since the estimated technologies using cost functions are different in the IO and OO models, as in the primal case, we do not repeat the results based on the production/distance functions results here.

2.5. Looking Through the Dual Profit Functions

2.5.1. The IO Model

Here we assume that the objective of a producer is to

$$\text{Max.} \quad \pi = p \cdot Y - w'X \equiv p \cdot Y - \left(\frac{w}{\Theta}\right)'X \cdot \Theta, \tag{2.20}$$
$$\text{subject to} \quad Y = f(X \cdot \Theta),$$

from which unconditional input demand and supply functions can be derived. Since the above problem reduces to a standard neoclassical profit-maximizing problem when X is replaced by $X \cdot \Theta$, and w is replaced by w/Θ, the corresponding profit function can be expressed as

$$\pi^a = p \cdot Y - \left(\frac{w}{\Theta}\right)'X \cdot \Theta = \pi\left(\frac{w}{\Theta}, p\right) \equiv \pi(w, p) \cdot h(w, p, \Theta) \leq \pi(w, p), \tag{2.21}$$

where π^a is actual profit, $\pi(w, p)$ is the profit frontier (homogeneous of degree one in w and p) and $h(w, p, \Theta) = \pi(w/\Theta, p)/\pi(w, p) \leq 1$ is profit inefficiency. Note that the $h(w, p, \Theta)$

function depends on w, p, and Θ in general. Application of Hotelling's lemma yields the following expressions for the output supply and input demand functions:

$$Y^a = Y^*(w,p)\left[h(\cdot) + \left(\frac{\pi(w,p)}{Y^*}\right)\left(\frac{\partial h(\cdot)}{\partial p}\right)\right] \leq Y^*(w,p),$$

$$X_j^a = X_j^*(w,p)\left[h(\cdot) - \left(\frac{\pi(w,p)}{X_j^*}\right)\left(\frac{\partial h(\cdot)}{\partial w_j}\right)\right] \leq X_j^*(w,p) \quad \forall j, \tag{2.22}$$

where the superscripts a and $*$ indicate actual and optimum levels of output Y and inputs X_j. The last inequality in the above equations will hold if the underlying production technology is well behaved.

2.5.2. The OO Model

Here the objective function can be written as

$$\text{Max.} \quad \pi = p \cdot Y - w'X \equiv p \cdot \Lambda \cdot \frac{Y}{\Lambda} - w'X \cdot \Theta,$$

$$\text{subject to} \quad Y = f(X) \cdot \Lambda, \tag{2.23}$$

which can be viewed as a standard neoclassical profit-maximizing problem when Y is replaced by Y/Λ and p is replaced by $p \cdot \Lambda$, the corresponding profit function can be expressed as

$$\pi^a = \frac{p \cdot \Lambda \cdot Y}{\Lambda} - w'X = \pi(w, p \cdot \Lambda) \equiv \pi(w,p) \cdot g(w,p,\Lambda) \leq \pi(w,p), \tag{2.24}$$

where $g(w,p,\Lambda) = \pi(w,p\cdot\Lambda)/\pi(w,p) \leq 1$. Similar to the IO model using Hotelling's lemma, we get

$$Y^a = Y^*(w,p)\left[g(\cdot) + \left(\frac{\pi(w,p)}{Y^*}\right)\left(\frac{\partial g(\cdot)}{\partial p}\right)\right] \leq Y^*(w,p),$$

$$X_j^a = X_j^*(w,p)\left[g(\cdot) - \left(\frac{\pi(w,p)}{X_j^*}\right)\left(\frac{\partial g(\cdot)}{\partial w_j}\right)\right] \leq X_j^*(w,p) \quad \forall j. \tag{2.25}$$

The last inequality in the above equations will hold if the underlying production technology is well behaved.

   To summarize (i) a shift in the profit functions for both the IO and OO models is non-neutral. Therefore, estimated elasticities, RTS, and so on, are affected by the presence of technical inefficiency, no matter what form is used. (ii) Technical inefficiency leads to a decrease in the production of output and decreases in input use in both models, however, prediction of the reduction in input use and production of output are not the same under both models.

   Even under profit maximization that recognizes endogeneity of both inputs and outputs, it matters which model is used to represent the technology!! These results are

different from those obtained under the primal models and from the cost minimization framework. Thus, it matters (both theoretically and empirically) whether one uses an input- or output-oriented measure of technical inefficiency.

3. Latent Class Models

3.1. Modeling Technological Heterogeneity

In modeling production technology we almost always assume that all the producers use the same technology. In other words, we do not allow the possibility that there might be more than one technology being used by the producers in the sample. Furthermore, the analyst may not know who is using what technology. Recently, a few studies have combined the stochastic frontier approach with the latent class structure in order to estimate a mixture of several technologies (frontier functions). Greene [7, 8] proposes a maximum likelihood for a latent class stochastic frontier with more than two classes. Caudill [9] introduces an expectation-maximization (EM) algorithm to estimate a mixture of two stochastic cost frontiers with two classes.[10] Orea and Kumbhakar [10] estimated a four-class stochastic frontier cost function (translog) with time-varying technical inefficiency.

Following the notations of Greene [7, 8] we specify the technology for class j as

$$\ln y_i = \ln f\left(x_i, z_i, \beta_j\right)\big|_j + v_i\big|_j - u_i\big|_j, \tag{3.1}$$

where $u_i\big|_j$ is a nonnegative random term added to the production function to accommodate technical inefficiency.

We assume that the noise term for class j follows a normal distribution with mean zero and constant variance, σ_{vj}^2. The inefficiency term $u_t\big|_j$ is modeled as a half-normal random variable following standard practice in the frontier literature, namely,

$$u_i\big|_j \sim N_+\left(0, \omega_j^2\right) \quad u_i\big|_j \geq 0. \tag{3.2}$$

That is, a half-normal distribution with scale parameter ω_j for each class.

With these distributional assumptions, the likelihood for firm i, if it belongs to class j, can be written as [11]

$$l(i \mid j) = \frac{2}{\sigma_j} \phi\left\{\frac{\varepsilon(i \mid j)}{\sigma_j}\right\} \Phi\left\{-\frac{\lambda_j \varepsilon(i \mid j)}{\sigma_j}\right\}, \tag{3.3}$$

where $\sigma_j^2 = \omega_j^2 + \sigma_j^2$, $\lambda_j = \omega_j/\sigma_{vj}$ and $\varepsilon(i \mid j) = \ln y_i - \ln f(x_i, z_i, \beta_j)\big|_j$. Finally, $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of a standard normal variable.

The unconditional likelihood for firm i is obtained as the weighted sum of their j-class likelihood functions, where the weights are the prior probabilities of class membership. That is,

$$l(i) = \sum_{j=1}^{J} l(i \mid j) \cdot \pi_{ij}, \quad 0 \leq \pi_{ij} \leq 1, \quad \sum_{j} \pi_{ijt} = 1, \tag{3.4}$$

where the class probabilities can be parameterized by, for example, a logistic function. Finally, the log likelihood function is

$$\ln L = \sum_{i=1}^{n} \ln l(i) = \sum_{i=1}^{n} \ln \left\{ \sum_{j=1}^{J} l(i \mid j) \cdot \pi_{ij} \right\}. \tag{3.5}$$

The estimated parameters can be used to compute the conditional posterior class probabilities. Using Bayes' theorem (see Greene [7, 8] and Orea and Kumbhakar [10]) the posterior class probabilities can be obtained from

$$P(ji) = \frac{l(ij) \cdot \pi_{ij}}{\sum_{j=1}^{J} l(ij) \cdot \pi_{ij}}. \tag{3.6}$$

This expression shows that the posterior class probabilities depend not only on the estimated parameters in π_{ij}, but also on parameters of the production frontier and the data. This means that a latent class model classifies the sample into several groups even when the π_{ij} are fixed parameters (independent of i).

In the standard stochastic frontier approach where the frontier function is the same for every firm, we estimate inefficiency relative to the frontier for all observations, namely, inefficiency from $E(u_i \mid \varepsilon_i)$ and efficiency from $E[\exp(-u_i) \mid \varepsilon_i]$. In the present case, we estimate as many frontiers as the number of classes. So the question is how to measure the efficiency level of an individual firm when there is no unique technology against which inefficiency is to be computed. This is solved by using the following method,

$$\ln EF_i = \sum_{j=1}^{J} P(j \mid i) \cdot \ln EF_i(j), \tag{3.7}$$

where $P(j \mid i)$ is the posterior probability to be in the jth class for a given firm i (defined in (3.9)), and $EF_i(j)$ is its efficiency using the technology of class j as the reference technology. Note that here we do not have a single reference technology. It takes into account technologies from every class. The efficiency results obtained by using (3.10) would be different from those based on the most likely frontier and using it as the reference technology. The magnitude of the difference depends on the relative importance of the posterior probability of the most likely cost frontier, the higher the posterior probability the smaller the differences. For an application see Orea and Kumbhakar [10].

3.2. Modeling Directional Heterogeneity

In Section 2.3 we talked about estimating IO technical inefficiency. In practice most researchers use the OO model because it is easy to estimate. Now we address the question of choosing one over the other. Orea et al. [12] used a model selection test procedure to determine whether the data support the IO, OO, or the hyperbolic model. Based on such a test result, one may decide to use the direction that fits the data best. This implictly assumes that all producers in the sample behave in the same way. In reality, firms in a particular industry, although using the same technology, may choose different direction to move to the frontier. For example, some producers might find it costly to adjust input levels to attain

the production frontier, while for others it might be easier to do so. This means that some producers will choose to shrink their inputs while others will augment the output level. In such a case imposing one direction for all sample observations is not efficient. The other practical problem is that no one knows in advance, which producers are following what direction. Thus, we cannot estimate the IO model for one group and the OO model for another.

The advantage of the LCM is that it is not necessary to impose *a priori* criterion to identify which producers are in what class. Moreover, we can formally examine whether some exogenous factors are responsible for choosing the input or the output direction by making the probabilities function of exogenous variables. Furthermore, when panel data is available, we do not need to assume that producers follow one direction for all the time, so we can accommodate switching behaviour and determine when they go in the input (output) direction.

3.2.1. The Input-Oriented Model

Under the assumption that $v_i \sim \mathbb{N}(0, \sigma^2)$, and θ_i is distributed independently of v_i, according to a distribution with density $f_\theta(\theta_i; \omega)$, where ω is a parameter, the distribution of y_i has density

$$f_{IO}(y_i \mid z_i, \Delta) = \left(2\pi\sigma^2\right)^{-1/2} \int_0^\infty \exp\left[-\frac{\left(y_i - z_i'\alpha - (1/2)\theta_i^2\Psi + \theta_i\Xi_i\right)^2}{2\sigma^2}\right] f_\theta(\theta_i; \omega) d\theta_i,$$

$$i = 1, \dots, n, \tag{3.8}$$

where Δ denotes the entire parameter vector. We use a half-normal specification for θ, namely,

$$f_\theta(\theta_i; \omega) = \left(\frac{\pi\omega^2}{2}\right)^{-1/2} \exp\left(-\frac{\theta_i^2}{2\omega^2}\right), \quad \theta_i \geq 0. \tag{3.9}$$

The likelihood function of the IO model is

$$L_{IO}(\Delta; y, X) = \prod_{i=1}^n f_{IO}(y_i \mid z_i, \Delta), \tag{3.10}$$

where $f_{IO}(y_i \mid z_i, \Delta)$ has been defined in (3.8). Since the integral defining $f_{IO}(y_i \mid z_i, \mu)$ in (3.11) is not available in closed form, we cannot find an analytical expression for the likelihood function. However, we can approximate the integrals using Monte Carlo simulation as follows. Suppose $\theta_{i,(s)}$, $s = 1, \dots, S$ is a random sample from $f_\theta(\theta_i; \omega)$. Then it is clear that

$$f_{IO}(y_i \mid z_i, \mu) \approx \tilde{f}_{IO}(y_i \mid z_i, \mu)$$

$$\equiv \left(2\pi\sigma^2\right)^{-1/2} \left(\frac{\pi\omega^2}{2}\right)^{-1/2}$$

$$\times S^{-1} \sum_{s=1}^S \exp\left[-\frac{\left(y_i - z_i'\alpha - (1/2)\theta_{i,(s)}^2\Psi + \theta_{i,(s)}\Xi_i\right)^2}{2\sigma^2} - \frac{\theta_{i,(s)}^2}{2\omega^2}\right], \tag{3.11}$$

and an approximation of the log-likelihood function is given by

$$\log l_{\text{IO}} \approx \sum_{i=1}^{n} \log \widetilde{f}_{\text{IO}}(y_i \mid z_i, \mu), \tag{3.12}$$

which can be maximized by numerical optimization procedures to obtain the ML estimator. To perform SML estimation, we consider the integral in (3.11). We can transform the range of integration to $[0,1]$ by using the transformation $r_i = \exp(-\theta_i)$ which has a natural interpretation as IO technical efficiency. Then, (3.11) becomes

$$f_{\text{IO}}(y_i \mid z_i, \mu) = \left(2\pi\sigma^2\right)^{-1/2} \int_0^1 \exp\left[-\frac{\left(y_i - z_i'\alpha - (1/2)(\ln r_i)^2\Psi - \ln r_i\Xi_i\right)^2}{2\sigma^2}\right]$$
$$\times f_\theta(-\ln r_i; \omega) r_i^{-1} dr_i. \tag{3.13}$$

Suppose $r_{i,(s)}$ is a set of standard uniform random numbers, for $s = 1, \ldots, S$. Then the integral can be approximated using the Monte Carlo estimator

$$\widetilde{f}_{\text{IO}}(y_i \mid z_i, \mu) = \left(2\pi\sigma^2\right)^{-1/2} \left(\frac{\pi\omega^2}{2}\right)^{-1/2} G_i(\mu), \tag{3.14}$$

where $G_i(\mu) = S^{-1} \sum_{s=1}^{S} \exp[-(y_i - z_i'\alpha - (1/2)(\ln r_{i,(s)})^2\Psi - \ln r_{i,(s)}\Xi_i)^2/2\sigma^2 - (\ln r_{i,(s)})^2/2\omega^2 - \ln r_{i,(s)}]$.

The standard uniform random numbers and their log transformation can be saved in an $n \times S$ matrix before maximum likelihood estimation and reused to ensure that the likelihood function is a differentiable function of the parameters. An alternative is to maintain the same random number seed and redraw these numbers for each call to the likelihood function. This option increases computing time but implies considerable savings in terms of memory. An alternative to the use of pseudorandom numbers is to use the Halton sequence to produce quasi-random numbers that fill the interval $[0,1]$. The Halton sequence has been used in econometrics by Train [13] for the multinomial probit model, and Greene [14] to implement SML estimation of the normal-gamma stochastic frontier model.

3.2.2. The Output-Oriented Model

Estimation of the OO is easy since the likelihood function is available analytically. The model is

$$y_i = z_i'\alpha + v_i - \lambda_i, \quad i = 1, \ldots, n. \tag{3.15}$$

We make the standard assumptions that $v_i \sim N(0, \sigma_v^2)$, $\lambda_i \sim N_+(0, \sigma_\lambda^2)$, and both are mutually independent as well as independent of z_i. The density of y_i is [11, page 75]

$$f_{\text{OO}}(y_i \mid z_i, \mu) = \frac{2}{\rho} \varphi_N\left(\frac{e_i}{\rho}\right) \Phi_N\left(-\frac{e_i \tau}{\rho}\right), \tag{3.16}$$

where $e_i = y_i - z_i'\alpha$, $\rho^2 = \sigma_v^2 + \sigma_\lambda^2$, $\tau = \sigma_\lambda / \sigma_v$, and ϕ_N and Φ_N denote the standard normal pdf and cdf, respectively. The log likelihood function of the model is

$$\ln l_{OO}(\mu; y, Z) = n \ln\left(\frac{2}{\rho}\right) - \frac{n}{2}\ln\left(2\pi\rho^2\right) - \frac{1}{2\rho^2}\sum_{i=1}^{n} e_i^2 + \sum_{i=1}^{n} \ln \Phi_N\left(-\frac{e_i\tau}{\rho}\right). \qquad (3.17)$$

3.3. The Finite Mixture (Latent Class) Model

The IO and OO models can be embedded in a general model that allows model choice for each observation in the absence of sample separation information. Specifically, we assume that each observation y_i is associated with the OO class with probability p, and with the IO class with probability $1 - p$. To be more precise, we have the model

$$y_i = z_i'\alpha - \lambda_i + v_i, \quad i = 1, \ldots, n, \qquad (3.18)$$

with probability p, and the model

$$y_i = z_i'\alpha + \frac{1}{2}\theta_i^2\Psi - \theta_i\Xi_i + v_i, \quad i = 1, \ldots, n, \qquad (3.19)$$

with probability $1 - p$, where the stochastic elements obey the assumptions that we stated previously in connection with the OO and IO models. Notice that the technical parameters, α, are the *same* in the two classes. Denote the parameter vector by $\psi = [\alpha', \sigma^2, \omega^2, \sigma_v^2, \sigma_\lambda^2, p]'$. The density of y_i will be

$$f_{LCM}(y_i \mid z_i, \psi) = p \cdot f_{OO}(y_i \mid z_i, O) + (1 - p)f_{IO}(y_i \mid z_i, \Delta), \quad i = 1, \ldots, n, \qquad (3.20)$$

where $O = [\alpha', \sigma^2, \omega^2]'$, and $\Delta = [\alpha', \sigma_v^2, \sigma_\lambda^2]$ are subsets of ψ. The log likelihood function of the model is

$$\log l_{LCM}(\psi; y, Z) = \sum_{i=1}^{n} \ln f_{LCM}(y_i \mid z_i, \psi) = \sum_{i=1}^{n} \ln[p \cdot f_{OO}(y_i \mid z_i, O) + (1 - p)f_{IO}(y_i \mid z_i, \Delta)]. \qquad (3.21)$$

The log likelihood function depends on the IO density $f_{IO}(y_i \mid z_i, \Delta)$, which is not available in closed form but can be obtained with the aid of simulation using the principles presented previously to obtain

$$\log l_{LCM}(\psi; y, Z) = \sum_{i=1}^{n} \ln f_{LCM}(y_i \mid z_i, \psi) = \sum_{i=1}^{n} \ln\left[pf_{OO}(y_i \mid z_i, O) + (1 - p)\tilde{f}_{IO}(y_i \mid z_i, \Delta)\right], \qquad (3.22)$$

where $\tilde{f}_{IO}(y_i \mid z_i, \Delta)$ has been defined in (3.14) and $f_{OO}(y_i \mid z_i, O)$ in (3.16). This log likelihood function can be maximized using standard techniques to obtain the SML estimates of the LCM.

3.3.1. Technical Efficiency Estimation in the Latent Class Model

A natural output-based efficiency measure derived from the LCM is

$$\text{TE}_i^{\text{LCM}} = P_i \text{TE}_i^{\text{OO}} + (1 - P_i)\text{TE}_i^{\text{IO}}, \tag{3.23}$$

where

$$P_i = \frac{pf_{\text{OO}}\left(y_i \mid z_i, \widetilde{\text{O}}\right)}{pf_{\text{OO}}\left(y_i \mid z_i, \widetilde{\text{O}}\right) + (1 - p)f_{\text{IO}}\left(y_i \mid z_i, \widetilde{\Delta}\right)}, \quad i = 1, \ldots, \cdot n, \tag{3.24}$$

is the posterior probability that the *i*th observation came from the OO class. These posterior probabilities are of independent interest since they can be used to provide inferences on whether a firm came from the OO or IO universe, depending on whether, for example, $P_i \geq 1/2$ or $P_i < 1/2$. This information can be important in deciding which type of adjustment cost (input- or output-related) is more important for a particular firm.

From the IO component of the LCM we have the IO-related efficiency measure, say $\text{IOTE}_i^{\text{LCM}}$, and its standard deviation, say $\text{STE}_i^{\text{LCM}}$, that can be compared with IOTE_i and STE_i from the IO model. Similarly we can compare TE_i^{OO} with the output efficiency of the IO model $(\text{TE}_i^{\text{IO}})$ and/or the output efficiency of the OO component of the LCM.

3.4. Returns to Scale and Technical Change

Note that returns to scale (defined as $\text{RTS} = \sum_j \partial y / \partial x_j$) is not affected by the presence of technical inefficiency in the OO model. The same is true for input elasticities and elasticities of substitution (that are not explored here). This is because inefficiency in the OO model shifts the production function in a neutral fashion. On the contrary, the magnitude of technical inefficiency affects RTS in the IO models. Using the translog specification in (2.4), we get

$$\text{RTS}_i^{\text{IO}} = 1_J'\left(\beta + \Gamma x_i + \varphi T_i\right) - \theta_i 1_J'\Gamma, \tag{3.25}$$

whereas the formula for RTS in the OO model is

$$\text{RTS}_i^{\text{OO}} = 1_J'\left(\beta + \Gamma x_i + \varphi T_i\right). \tag{3.26}$$

We now focus on estimates of technical change from the IO and OO models. Again TC in the IO model can be measured conditional on θ $(\text{TC}^{\text{IO}}(\text{I}))$ and TC defined at the frontier $(\text{TC}^{\text{IO}}(\text{II}))$, namely,

$$\text{TC}^{\text{IO}}(\text{I}) = \beta_T + \beta_{TT}T_i + x_i'\varphi + 1_J'\varphi \; \theta_i,$$
$$\text{TC}^{\text{IO}}(\text{II}) = \beta_T + \beta_{TT}T_i + x_i'\varphi. \tag{3.27}$$

These two formulas will give different results if technical change is neutral and/or the production function is homogeneous (i.e., $1_J'\varphi \neq 0$). The formula for TC^{OO} is the same as

TC^{IO}(II) except for the fact that the estimated parameters in (3.3) are from the IO model, whereas the parameters to compute TC^{OO} are from the OO model.

It should be noted that in the LCM we enforce the restriction that the technical parameters, α, are the same in the IO and OO components of the mixture. This implies that RTS and TC will be the same in both components if we follow the first approach, but they will be different if we follow the second approach. In the second approach, a single measure of RTS and TC can be defined as the weighted average of both measures using the posterior probabilities, P_i, as weights. To be more precise, suppose $RTS_i^{IO,LCM}$(II) is the type II RTS measure derived from the IO component of the LCM, and $RTS_i^{OO,LCM}$ is the RTS measure derived from the OO component of the LCM. The overall LCM measure of RTS will be $RTS_i^{LCM} = P_i RTS_i^{OO,LCM} + (1 - P_i) RTS_i^{IO,LCM}$. Similar methodology is followed for the TC measure.

4. Relaxing Functional form Assumptions (SF Model with LML)

In this section we introduce the LML methodology [15] in estimating SF models in such a way that many of the limitations of the SF models originally proposed by Aigner et al. [2], Meeusen and van den Broeck [3], and their extensions in the last two and a half decades are relaxed. Removal of all these deficiencies generalizes the SF models and makes them comparable to the DEA models. Moreover, we can apply standard econometric tools to perform estimation and draw inferences.

To fix ideas, suppose we have a parametric model that specifies the density of an observed dependent variable y_i conditional on a vector of observable covariates $x_i \in X \subseteq R^k$, a vector of unknown parameters $\theta \in \Theta \subseteq R^m$, and let the density be $l(y_i; x_i, \theta)$. The parametric ML estimator is given by

$$\tilde{\theta} = \arg\max_{\theta \in \Theta}: \sum_{i=1}^{n} \ln l(y_i; x_i, \theta). \tag{4.1}$$

The problem with the parametric ML estimator is that it relies heavily on the parametric model that can be incorrect if there is uncertainty regarding the functional form of the model, the density, *and so forth.* A natural way to convert the parametric model to a nonparametric one is to make the parameter θ a function of the covariates x_i. Within LML this is accomplished as follows. For an arbitrary $x \in X$, the LML estimator solves the problem

$$\tilde{\theta}(x) = \arg\max_{\theta \in \Theta}: \sum_{i=1}^{n} \ln l(y_i; x_i, \theta) K_H(x_i - x), \tag{4.2}$$

where K_H is a kernel that depends on a matrix bandwidth H. The idea behind LML is to choose an anchoring parametric model and maximize a weighted log-likelihood function that places more weight to observations near x rather than weight each observation equally, as the parametric ML estimator would do.[11] By solving the LML problem for several points $x \in X$, we can construct the function $\tilde{\theta}(x)$ that is an estimator for $\theta(x)$, and effectively we have a fully general way to convert the parametric model to a nonparametric approximation to the unknown model.

Suppose we have the following stochastic frontier cost model:

$$y_i = x_i'\beta + v_i + u_i, \quad v_i \sim \mathbb{N}(0, \sigma^2), \quad u_i \sim \mathbb{N}(\mu, \omega^2), \quad u_i \geq 0, \quad \text{for } i = 1, \ldots, n, \ \beta \in R^k, \tag{4.3}$$

where y is log cost and x_i is a vector of input prices and outputs[12]; v_i and u_i are the noise and inefficiency components, respectively. Furthermore, v_i and u_i are assumed to be mutually independent as well as independent of x_i.

To make the frontier model more flexible (nonparametric), we adopt the following strategy. Consider the usual parametric ML estimator for the normal (v) and truncated normal (u) stochastic cost frontier model that solves the following problem [16]:

$$\tilde{\theta} = \arg \max_{\theta \in \Theta}: \sum_{i=1}^{n} \ln l(y_i; x_i, \theta), \tag{4.4}$$

where

$$l(y_i; x_i, \theta) = [\Phi(\psi)]^{-1} \Phi\left[\frac{\sigma^2 \psi + \omega(y_i - x'_i\beta)}{\sigma(\omega^2 + \sigma^2)^{1/2}}\right] [2\pi(\omega^2 + \sigma^2)]^{-1/2} \exp\left[-\frac{(y_i - x'_i\beta - \mu)^2}{2(\omega^2 + \sigma^2)}\right], \tag{4.5}$$

$\psi = \mu/\omega$, and Φ denotes the standard normal cumulative distribution function. The parameter vector is $\theta = [\beta, \sigma, \omega, \psi]$ and the parameter space is $\Theta = R^k \times R_+ \times R_+ \times R$. Local ML estimation of the corresponding nonparametric model involves the following steps. First, we choose a kernel function. A reasonable choice is

$$K_H(d) = (2\pi)^{-m/2} |H|^{-1/2} \exp\left(-\frac{1}{2} d' H^{-1} d\right), \quad d \in R^m, \tag{4.6}$$

where m is the dimensionality of θ, $H = h \cdot S$, $h > 0$ is a scalar bandwidth, and S is the sample covariance matrix of x_i. Second, we choose a particular point $x \in X$, and solve the following problem:

$$\tilde{\theta}(x) = \arg \max_{\theta \in \Theta}: \sum_{i=1}^{n} \left\{ -\ln \Phi(\psi) + \ln \Phi\left[\frac{\sigma^2 \psi + \omega(y_i - x'_i\beta)}{\sigma(\omega^2 + \sigma^2)^{1/2}}\right] \right.$$
$$\left. -\frac{1}{2} \ln(\omega^2 + \sigma^2) - \frac{1}{2} \frac{(y_i - x'_i\beta - \mu)^2}{(\omega^2 + \sigma^2)} \right\} K_H(x_i - x). \tag{4.7}$$

A solution to this problem provides the LML parameter estimates $\tilde{\beta}(x)$, $\tilde{\sigma}(x)$, $\tilde{\omega}(x)$, and $\tilde{\psi}(x)$. Also notice that the weights $K_H(x_i - x)$ do not involve unknown parameters (if h is known) so they can be computed in advance and, therefore, the estimator can be programmed in any standard econometric software.[13] For an application of this methodology to US commercial banks see Kumbhakar and Tsionas [17, 18] and Kumbhakar et al. [15].

5. Some Advances in Stochastic Frontier Analysis

5.1. General

There are many innovative empirical applications of stochastic frontier analysis in recent years. One of them is in the field of auctions, a particular field of game theory. Advances

and empirical applications in this field are likely to accumulate rapidly and contribute positively to the advancement of empirical Game theory and empirical IO. Kumbhakar et al. [19] propose Bayesian analysis of an auction model where systematic over-bidding and under-bidding is allowed. Extensive simulations are used to show that the new techniques perform well and ignoring measurement error or systematic over-bidding and under-bidding is important in the final results.

Kumbhakar and Parmeter [20] derive the closed-form likelihood and associated efficiency measures for a two-sided stochastic frontier model under the assumption of normal-exponential components. The model receives an important application in the labor market where employees and employers have asymmetric information, and each one tries to manipulate the situation to his own advantage, Employers would like to hire for less and employees to obtain more in the bargaining process. The precise measurement of these components is, apparently, important.

Kumbhakar et al. [21] acknowledge explicitly the fact that certain decision making units can be fully (i.e., 100%) efficient, and propose a new model which is a mixture of (i) a half-normal component for inefficient firms and (ii) a mass at zero for efficient firms. Of course, it is not known in advance which firms are fully efficient or not. The authors propose classical methods of inference organized around maximum likelihood and provide extensive simulations to explore the validity and relevance of the new techniques under various data generating processes.

Tsionas [22] explores the implications of the convolution $\varepsilon = v + u$ in stochastic frontier models. The fundamental point is that even when the distributions of the error components are nonstandard (e.g., Student-t and half-Student or normal and half-Student, gamma, symmetric stable, etc.) it is possible to estimate the model by ML estimation via the fast Fourier transform (FFT) when the characteristic functions are available in closed form. These methods can also be used in mixture models, input-oriented efficiency models, two-tiered stochastic frontiers, and so forth. The properties of ML and some GLS techniques are explored with an emphasis on the normal-truncated normal model for which the likelihood is available analytically and simulations are used to determine various quantities that must be set in order to apply ML by FFT.

Starting with Annaert et al. [23], stochastic frontier models have been applied very successfully in finance, especially the important issue of mutual funds performance. Schaefer and Maurer [24] apply these techniques to German funds to find that they "may be able to reduce its costs by 46 to 74% when compared with the best-practice complex in the sample." Of course, much remains to be done in this area and connect more closely stochastic frontier models with practical finance and better mutual fund performance evaluation.

5.2. Panel Data

Panel data have always been a source of inspiration and new models in stochastic frontier analysis. Roughly speaking, panel data are concerned with models of the form $y_{it} = \alpha_i + x'_{it}\beta + v_{it}$, where the α_i's are individual effects, random or fixed, x_{it} is a $k \times 1$ vector of covariates, β is a $k \times 1$ parameter vector and, typically, the error term $v_{it} \sim iidN(0, \sigma_v^2)$.

An important contribution in panel data models of efficiency is the incorporation of factors, as in Kneip et al. [25]. Factors arise from the necessity of incorporating more structure into frontier models, a point that is clear after Lee and Schmidt [26]. The authors use smoothing techniques to perform the econometric analysis of the model.

In recent years, the focus of the profession has shifted from the fixed effects model (e.g., Cornwell et al. [27]) to a so-called "true fixed effects model" (TFEM) first proposed by Greene [28]. Greene's model is $y_{it} = \alpha_i + x'_{it}\beta + v_{it} \pm u_{it}$, where $u_{it} \sim iidN_+(0, \sigma_v^2)$. In this model, the individual effects are separated from technical inefficiency. Similar models have been proposed previously by Kumbhakar [29] and Kumbhakar and Hjalmarsson [30], although in these models firm-effects were treated as persistent inefficiency. Greene shows that the TFEM can be estimated easily using special Gauss-Newton iterations without the need to explicitly introduce individual dummy variables, which is prohibitive if the number of firms (N) is large. As Greene [28] notes: "*the fixed and random effects estimators force any time invariant cross unit heterogeneity into the same term that is being used to capture the inefficiency. Inefficiency measures in these models may be picking up heterogeneity in addition to or even instead of inefficiency.*" For important points and applications see Greene [8, 31].

Greene's [8] findings are somewhat at odds with the perceived incidental parameters problem in this model, as he himself acknowledges. His findings motivated a body of research that tries to deal with the incidental parameters problem in stochastic frontier models and, of course, efficiency estimation. The incidental parameters problem in statistics began with the well-known contribution of Neyman and Scott [32] (see also [33]). In stochastic frontier models of the form: $y_{it} = \alpha_i + x'_{it}\beta + v_{it}$, for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The essence of the problem is that as N gets large, the number of unknown parameters (the individual effects α_i, $i = 1, \ldots, N$) increase at the same rate so consistency cannot be achieved. Another route to the incidental parameters problem is well-known in the efficiency estimation with cross-sectional data ($T = 1$), where JLMS estimates are not consistent.

To appreciate better the incidental parameters problem, the TFEM implies a density for the ith unit, say

$$p_i(\alpha_i, \delta; Y_i) \equiv p_i(\alpha_i, \beta, \sigma, \omega; Y_i), \quad \text{where } Y_i = [y'_i, x'_i]' \text{ is the data.} \tag{5.1}$$

The problem is that the ML estimator

$$\max_{\alpha_1, \ldots, \alpha_n, \delta} : \sum_{i=1}^{n} p_i(\alpha_i, \delta; Y_i) = \max_{\delta} : \sum_{i=1}^{n} p_i(\hat{\alpha}_i, \delta; Y_i) \tag{5.2}$$

is not consistent. The source of the problem is that the concentrated likelihood using $\hat{\alpha}_i$ (the ML estimator) will not deliver consistent estimators for all elements of δ.

In frontier models we know that ML estimators for β and $\sigma^2 + \omega^2$ seem to be alright but the estimator for ω or the ratio $\lambda = \omega/\sigma$ can be wrong. This is also validated in a recent paper by Chen et al. [34].

There are some approaches to correct such biases in the literature on nonlinear panel data models.

(i) Correct the bias to first order using a modified score (first derivatives of log likelihood).

(ii) Use a penalty function for the log likelihood. This can of course be related to [2] above.

(iii) Apply panel jackknife. Satchachai and Schmidt [35] has done that recently in a model with fixed effects but without one-sided component. He derives some interesting results regarding convergence depending on whether we have ties or

not. First differencing produces $O(T^{-1})$ but with ties we have $O(T^{-1/2})$ (for the estimator applied when you have a tie).

(iv) In line with (ii) one could use a modified likelihood of the form $p_i^*(\delta; Y_i) = \int p_i(\alpha_i, \delta; Y_i) w(\alpha_i) d\alpha_i$, where $w(\alpha_i)$ is some weighting function for which it is clear that there is a Bayesian interpretation.

Since Greene [8] derived a computationally efficient algorithm for the true fixed effects model, one would think that application of panel jackknife would reduce the first order bias of the estimator and for empirical purposes this might be enough. For further reductions in the bias there remains only the possibility of asymptotic expansions along the lines of related work in nonlinear panel data models. This point has not been explored in the literature but it seems that it can be used profitably.

Wang and Ho [36] show that "first-difference and within-transformation can be analytically performed on this model to remove the fixed individual effects, and thus the estimator is immune to the incidental parameters problem." The model is, naturally, less general than a standard stochastic frontier model in that the authors assume $u_{it} = f(z_{it}'\delta) u_i^+$, where u_i^+ is a positive half-normal random variable and $f()$ is a positive function. In this model, the dynamics of inefficiency are determined entirely by the function $f()$ and the covariates that enter into this function.

Recently, Chen et al. [34] proposed a new estimator for the model. If the model is $y_{it} = \alpha_i + x_{it}'\beta + v_{it} - u_{it}$, deviation from the mean gives

$$\tilde{y}_{it} = \tilde{x}_{it}'\beta + \tilde{v}_{it} - \tilde{u}_{it}. \tag{5.3}$$

Given $\beta = \hat{\beta}_{OLS}$, we have $\tilde{e}_{it} \equiv \tilde{y}_{it} - \tilde{x}_{it}'\beta = \tilde{v}_{it} - \tilde{u}_{it}$, when $\tilde{e}_{it} =$ "data." The distribution of $\tilde{e}_{it} = \tilde{v}_{it} - \tilde{u}_{it}$, belongs to the family of the multivariate closed skew-normal (CSN), so estimating λ and σ is easy. Of course, the multivariate CSN depends on evaluating a multivariate normal integral in R^{T+1}. With $T > 5$, this is not a trivial problem (see [37]).

There is reason to believe that "average likelihood" or a fully Bayesian approach can perform much better relative to sampling-theory treatments. Indeed, the true fixed effects model is nothing but another instance of the incidental parameters problem. Recent advances suggest that the best treatment can be found in "average" or "integrated" likelihood functions. For work in this direction, see Lancaster [33] and Arellano and Bonhomme [38], Arellano and Hahn [39, 40], Berger et al. [41], and Bester and Hansen [42, 43]. The performance of such methods in the context of TFE remains to be seen.

Tsionas and Kumbhakar [44] propose a full Bayesian solution to the problem. The approach is obvious in a sense, since the TFEM can be cast as a hierarchical model. The authors show that the obvious parameterization of the model does not perform well in simulated experiments and, therefore, they propose a new parameterization that is shown to effectively eliminate the incidental parameters problem. They also extend the TFEM to models with both individual and time effects. Of course, the TFEM is cast in terms of a random effects model so it is at first sight not directly related to [35].

6. Thoughts on Current State of Efficiency Estimation and Panel Data

6.1. Nature of Individual Effects

If we think about the model: $y_{it} = \alpha_i + x_{it}'\beta + v_{it} + u_{it}$, $u_{it} \geq 0$, as $N \rightarrow \infty$ one natural question is: Do we really expect ourselves to be so agnostic about the fixed effects as to allow α_{n+1} to

be completely different from what we already know about $\alpha_1, \ldots, \alpha_n$? This is rarely the case. But we do not adopt the true fixed effects model for that reason. There are other reasons. If we ignore this choice we can adopt a finite mixture of normal distribution for the effects.

In principle this can approximate well any distribution of the effects, so with enough latent classes we should be able to approximate the weight function $w(\alpha_i)$ quite well. That would pose some structure in the model, it would avoid the incidental parameters problem (if the number of classes grows slowly and at a lower rate than N) so for fixed T there should be no significant bias. For really small T a further bias correction device like asymptotic expansions or the jackknife could be used.

Since we do not adopt the true fixed effects model for that reason, why do we adopt it? Because the effects and the regressors are potentially correlated in a random effect framework so it is preferable to think of them as parameters. It could be that $\alpha_i = h(x_{i1}, \ldots, x_{iT})$, or perhaps $\alpha_i = H(\overline{x}_i)$, when $\overline{x}_i = T^{-1} \sum_{t=1}^{T} x_{it}$. Mundlak [45] first wrote about this model. In some cases it makes sense. Consider the alternative model:

$$
y_{it} = \alpha_{it} + x'_{it}\beta + v_{it} + u_{it},
$$
$$
\alpha_{it} = \overline{\alpha}(1 - \rho) + \rho\alpha_{it-1} + x'_{it}\gamma + \varepsilon_{it}.
\tag{6.1}
$$

Under stationarity we have the same implications with Mundlak's original model but in many cases it makes much more sense: mutual fund rating and evaluation is one of them. But even if we stay with Mundlak's original specification, many other possibilities are open. For small T, the most interesting case, approximation of $\alpha_i = h(x_{i1}, \ldots, x_{iT})$ by some flexible functional form should be enough for practical purposes. By "practical purposes" we mean bias reduction to order $O(T^{-1})$ or better. If the model becomes

$$
y_{it} = \alpha_i + x'_{it}\beta + v_{it} + u_{it} = h(x_{i1}, \ldots, x_{iT}) + x'_{it}\beta + v_{it} + u_{it}, \quad t = 1, \ldots, T,
\tag{6.2}
$$

adaptation of known nonparametric techniques should provide that rate of convergence. Reduction to

$$
y_{it} = \alpha_i + x'_{it}\beta + v_{it} + u_{it} = H(\overline{x}_i) + x'_{it}\beta + v_{it} + u_{it}, \quad t = 1, \ldots, T,
\tag{6.3}
$$

would facilitate the analysis considerably without sacrificing the rate. It is quite probable that h or H should be low order polynomials or basis functions with some nice properties (e.g., Bernstein polynomials are nice) that can overcome the incidental parameters problem.

6.2. Random Coefficient Models

Consider $y_{it} = \alpha_i + x'_{it}\beta_i + v_{it} + u_{it} \equiv z'_{it}\gamma_i + v_{it} + u_{it}$ [46]. Typically we assume that $\gamma_i \sim iidN_K(\overline{\gamma}, \Omega)$. For small to moderate panels (say $T = 5$ to 10) adaptation of the techniques in the paper by Chen et al. [34] would be quite difficult to implement in the context of fixed effects. The concern is again with evaluation of $(T + 1)$-dimensional normal integrals, when T is large. Here, again we are subject to the incidental parameters problem—we never really escape the "small sample" situation.

One way to proceed is the so-called CAR (conditionally autoregressive) prior (model) of the form:

$$\gamma_i \mid \gamma_j \sim N\left(\gamma_j, \varphi_{ij}^2\right) \quad \forall i \neq j, \quad \log \varphi_{ij}^2 = \delta_0 + \delta_1 \|w_i - w_j\|, \quad \text{if } \gamma_i \text{ is scalar (the TFEM).}$$

(6.4)

In the multivariate case we would need something like the BEKK factorization of a covariance matrix as in multivariate GARCH processes. The point is that the coefficients cannot be too dissimilar and their degree of dissimilarity depends on a parameter φ^2 that can be made a function of covariates, if any. Under different DGPs, it would be interesting to know how the Bayesian estimator of this model behaves in practice.

6.3. More on Individual Effects

Related to the above discussion, it is productive to think about sources, that is, where these α_is or γ_is come from. Of course we have Mundlak's [45] interpretation in place. In practice we have different technologies (represented by cost functions, say). The standard cost function $y_{it} = \alpha_0 + x'_{it}\beta + v_{it}$ with input-oriented technical inefficiency results in $y_{it} = \alpha_0 + x'_{it}\beta + v_{it} + u_{it}$, $u_{it} \geq 0$. Presence of allocative inefficiency results in a much more complicated model: $y_{it} = \alpha_0 + x'_{it}\beta + v_{it} + u_{it} + G_{it}(x_{it}, \beta, \xi_{it})$ where ξ_{it} is the vector of price distortions (see [47, 48]). So under some reasonable economic assumptions and common technology we end up with a nonlinear effects model through the $G(\cdot)$ function. Of course one can apply the TFEM here but that would not correspond to the true DGP. So the issues of consistency are at stake.

It is hard to imagine a situation where in a TFEM, $y_{it} = \alpha_i + x'_{it}\beta + v_{it} + u_{it}$ the α_is can be anything and they are subjected to no "similarity" constraints. We can, of course, accept that $\alpha_i \approx \alpha_0 + G_{it}(x_{it}, \beta, \xi_{it})$, so at least for the translog we should have a rough guide on what these effects represent under allocative inefficiency first-order approximations to the complicated G term are available when the ξ_{it}s are small. Of course then one has to think about the nature of the allocative distortions but at least that's an economic problem.

6.4. Why a Bayesian Approach?

Chen et al.'s transformation that used the multivariate CSN is one class of transformations, but there are many transformations that are possible because the TFEM does not have the property of information orthogonality [33]. The "best" transformation, the one that is "maximally bias reducing," cannot be taking deviations from the means because the information matrix is not block diagonal with respect to $(\alpha_i, \lambda = \omega/\sigma = \text{signal/noise})$. Other transformations would be more effective and it is not difficult to find them, in principle.

Recently Tsionas and Kumbhakar [44] considered a different model, namely, $y_{it} = \alpha_i + \delta_i^+ + x'_{it}\beta + v_{it} + u_{it}$, where $\delta_i^+ \geq 0$ is persistent inefficiency. They have used a Bayesian approach. Colombi et al. [49] used the same model but used classical (ML) approach to estimate the parameters as well as the inefficiency components. The finite sample properties of the Bayes estimators (posterior means and medians) in Tsionas and Kumbhakar [44] were found to be very good for small samples with λ values typically encountered in practice (of course one needs to keep λ away from zero in the DGP). The moral of the story is that in the random effects model, an integrated likelihood approach based on reasonable priors,

a nonparametric approach based on low-order polynomials or a finite mixture model might provide an acceptable approximation to parameters like λ.

Coupled with a panel jack knife device these approaches can be really effective in mitigating the incidental parameters problem. For one, in the context of TFE, we do not know how the Chen et al. [34] estimator would behave under strange DGPs-under strange processes for the incidental parameters that is. We have some evidence from Monte Carlo but we need to think about more general "mitigating strategies." The integrated likelihood approach is one, and is close to a Bayesian approach. Finite mixtures also hold great promise since they have good approximating properties. The panel jack knife device is certainly something to think about. Also analytical devices for bias reduction to order $O(T^{-1})$ or $O(T^{-2})$ are available from the likelihood function of the TFEM (score and information). Their implementation in software should be quite easy.

7. Conclusions

In this paper we presented some new techniques to estimate technical inefficiency using stochastic frontier technique. First, we presented a technique to estimate a nonhomogeneous technology using the IO technical inefficiency. We then discussed the IO and OO controversy in the light of distance functions, and the dual cost and profit functions. The second part of the paper addressed the latent class-modeling approach incorporating behavioral heterogeneity. The last part of the paper addressed LML method that can solve the functional form issue in parametric stochastic frontier. Finally, we added a section that deals with some very recent advances.

Endnotes

1. Another measure is hyperbolic technical inefficiency that combines both the IO and OO measures in a special way (see, e.g., [50], Cuesta and Zofio (1999), [12]). This measure is not as popular as the other two.

2. On the contrary, the OO model has been estimated by many authors using DEA (see, e.g., [51] and references cited in there).

3. Alvarez et al. [52] addressed these issues in a panel data framework with time invariant technical inefficiency (using a fixed effects models).

4. The above equation gives the IO model (when the production function is homogeneous) by labeling $\lambda_i = r\theta_i$.

5. Alvarez et al. [52] estimated an IO primal model in a panel data model where technical inefficiency is assumed to be fixed and parametric.

6. Greene [14] used SML for the OO normal-gamma model.

7. See also Kumbhakar and Lovell [11, pages 74–82] for the log-likelihood functions under both half-normal and exponential distributions for the OO technical inefficiency term.

8. It is not necessary to use the simulated ML method to estimate the parameters of the frontier models if the technical inefficiency component is distributed as half-normal, truncated normal, or exponential along with the normality assumption on the noise component. For other distributions, for example, gamma for technical inefficiency and normal for the noise component the standard ML method may not be ideal (see Greene

[14] who used the simulated ML method to estimate OO technical efficiency in the gamma-normal model).

9. Atkinson and Cornwell [53] estimated translog cost functions with both input- and output-oriented technical inefficiency using panel data. They assumed technical inefficiency to be fixed and time invariant. See also Orea et al. [12].

10. See, in addition, Beard et al. [54, 55] for applications using a non-frontier approach. For applications in social sciences, see, Hagenaars and McCutcheon [56]. Statistical aspects of the mixing models are dicussed in details in Mclachlan and Peel [57].

11. LML estimation has been proposed by Tibshirani [58] and has been applied by Gozalo and Linton [59] in the context of nonparametric estimation of discrete response models.

12. The cost function specification is discussed in details in Section 5.2.

13. An alternative, that could be relevant in some applications, is to localize based on a vector of exogenous variables z_i instead of the x_i's. In that case, the LML problem becomes

$$\widetilde{\theta}(z) = \arg\max_{\theta \in \Theta} : \sum_{i=1}^{n} \left\{ -\ln \Phi(\psi) + \ln \Phi \left[\sigma^2 \psi + \frac{\omega(y_i - x_i'\beta)}{\sigma(\omega^2 + \sigma^2)^{1/2}} \right] \right.$$
$$\left. -(1/2)\ln\left(\omega^2 + \sigma^2\right) - (1/2)\left(\frac{(y_i - x_i'\beta - \mu)^2}{(\omega^2 + \sigma^2)} \right) \right\} K_H(z_i - z),$$

where z are the given values for the vector of exogenous variables. The main feature of this formulation is that the β parameters as well as σ, ω, and ψ will now be functions of z instead of x.

References

[1] S. C. Kumbhakar and E. G. Tsionas, "Estimation of stochastic frontier production functions with input-oriented technical efficiency," *Journal of Econometrics*, vol. 133, no. 1, pp. 71–96, 2006.

[2] D. Aigner, C. A. K. Lovell, and P. Schmidt, "Formulation and estimation of stochastic frontier production function models," *Journal of Econometrics*, vol. 6, no. 1, pp. 21–37, 1977.

[3] W. Meeusen and J. van den Broeck, "Efficiency estimation from Cobb-Douglas production functions with composed error," *International Economic Review*, vol. 18, no. 2, pp. 435–444, 1977.

[4] J. Jondrow, C. A. Knox Lovell, I. S. Materov, and P. Schmidt, "On the estimation of technical inefficiency in the stochastic frontier production function model," *Journal of Econometrics*, vol. 19, no. 2-3, pp. 233–238, 1982.

[5] G. E. Battese and T. J. Coelli, "Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data," *Journal of Econometrics*, vol. 38, no. 3, pp. 387–399, 1988.

[6] C. Arias and A. Alvarez, "A note on dual estimation of technical efficiency," in *Proceedings of the 1st Oviedo Efficiency Workshop*, University of Oviedo, 1998.

[7] W. Greene, "Fixed and random effects in nonlinear models," Working Paper, Department of Economics, Stern School of Business, NYU, 2001.

[8] W. Greene, "Reconsidering heterogeneity in panel data estimators of the stochastic frontier model," *Journal of Econometrics*, vol. 126, no. 2, pp. 269–303, 2005.

[9] S. B. Caudill, "Estimating a mixture of stochastic frontier regression models via the em algorithm: a multiproduct cost function application," *Empirical Economics*, vol. 28, no. 3, pp. 581–598, 2003.

[10] L. Orea and S. C. Kumbhakar, "Efficiency measurement using a latent class stochastic frontier model," *Empirical Economics*, vol. 29, no. 1, pp. 169–183, 2004.

[11] S. C. Kumbhakar and C. A. K. Lovell, *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge, UK, 2000.

[12] L. Orea, D. Roibás, and A. Wall, "Choosing the technical efficiency orientation to analyze firms' technology: a model selection test approach," *Journal of Productivity Analysis*, vol. 22, no. 1-2, pp. 51–71, 2004.

[13] K. E. Train, *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, UK, 2nd edition, 2009.

[14] W. H. Greene, "Simulated likelihood estimation of the normal-gamma stochastic frontier function," *Journal of Productivity Analysis*, vol. 19, no. 2-3, pp. 179–190, 2003.

[15] S. C. Kumbhakar, B. U. Park, L. Simar, and E. G. Tsionas, "Nonparametric stochastic frontiers: a local maximum likelihood approach," *Journal of Econometrics*, vol. 137, no. 1, pp. 1–27, 2007.

[16] R. E. Stevenson, "Likelihood functions for generalized stochastic frontier estimation," *Journal of Econometrics*, vol. 13, no. 1, pp. 57–66, 1980.

[17] S. C. Kumbhakar and E. G. Tsionas, "Scale and efficiency measurement using a semiparametric stochastic frontier model: evidence from the U.S. commercial banks," *Empirical Economics*, vol. 34, no. 3, pp. 585–602, 2008.

[18] S. C. Kumbhakar and E. Tsionas, "Estimation of cost vs. profit systems with and without technical inefficiency," *Academia Economic Papers*, vol. 36, pp. 145–164, 2008.

[19] S. Kumbhakar, C. Parmeter, and E. G. Tsionas, "Bayesian estimation approaches to first-price auctions," to appear in *Journal of Econometrics*.

[20] S. C. Kumbhakar and C. F. Parmeter, "The effects of match uncertainty and bargaining on labor market outcomes: evidence from firm and worker specific estimates," *Journal of Productivity Analysis*, vol. 31, no. 1, pp. 1–14, 2009.

[21] S. Kumbhakar, C. Parmeter, and E. G. Tsionas, "A zero inflated stochastic frontier model," unpublished.

[22] E. G. Tsionas, "Maximum likelihood estimation of non-standard stochastic frontier models by the Fourier transform," to appear in *Journal of Econometrics*.

[23] J. Annaert, J. Van den Broeck, and R. V. Vennet, "Determinants of mutual fund underperformance: a Bayesian stochastic frontier approach," *European Journal of Operational Research*, vol. 151, no. 3, pp. 617–632, 2003.

[24] A. Schaefer, R. Maurer et al., "Cost efficiency of German mutual fund complexes," SSRN working paper, 2010.

[25] A. Kneip, W. H. Song, and R. Sickles, "A new panel data treatment for heterogeneity in time trends," to appear in *Econometric Theory*.

[26] Y. H. Lee and P. Schmidt, "A production frontier model with flexible temporal variation in technical inefficiency," in *The Measurement of Productive Efficiency: Techniques and Applications*, H. O. Fried, C. A. K. Lovell, and S. S. Schmidt, Eds., Oxford University Press, New York, NY, USA, 1993.

[27] C. Cornwell, P. Schmidt, and R. C. Sickles, "Production frontiers with cross-sectional and time-series variation in efficiency levels," *Journal of Econometrics*, vol. 46, no. 1-2, pp. 185–200, 1990.

[28] W. Greene, "Fixed and random effects in stochastic frontier models," *Journal of Productivity Analysis*, vol. 23, no. 1, pp. 7–32, 2005.

[29] S. C. Kumbhakar, "Estimation of technical inefficiency in panel data models with firm- and time-specific effects," *Economics Letters*, vol. 36, no. 1, pp. 43–48, 1991.

[30] S. C. Kumbhakar and L. Hjalmarsson, "Technical efficiency and technical progress in Swedish dairy farms," in *The Measurement of Productive Efficiency—Techniques and Applications*, H. O. Fried, C. A. K. Lovell, and S. S. Schmidt, Eds., pp. 256–270, Oxford University Press, Oxford, UK, 1993.

[31] W. Greene, "Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems," *Health Economics*, vol. 13, no. 10, pp. 959–980, 2004.

[32] J. Neyman and E. L. Scott, "Consistent estimates based on partially consistent observations," *Econometrica. Journal of the Econometric Society*, vol. 16, pp. 1–32, 1948.

[33] T. Lancaster, "The incidental parameter problem since 1948," *Journal of Econometrics*, vol. 95, no. 2, pp. 391–413, 2000.

[34] Y.-Y. Chen, P. Schmidt, and H.-J. Wang, "Consistent estimation of the fixed effects stochastic frontier model," in *Proceedings of the the European Workshop on Efficiency and Productivity Analysis (EWEPA '11)*, Verona, Italy, 2011.

[35] P. Satchachai and P. Schmidt, "Estimates of technical inefficiency in stochastic frontier models with panel data: generalized panel jackknife estimation," *Journal of Productivity Analysis*, vol. 34, no. 2, pp. 83–97, 2010.

[36] H. J. Wang and C. W. Ho, "Estimating fixed-effect panel stochastic frontier models by model transformation," *Journal of Econometrics*, vol. 157, no. 2, pp. 286–296, 2010.

[37] A. Genz, "An adaptive multidimensional quadrature procedure," *Computer Physics Communications*, vol. 4, no. 1, pp. 11–15, 1972.

[38] M. Arellano and S. Bonhomme, "Robust priors in nonlinear panel data models," Working paper, CEMFI, Madrid, Spain, 2006.

[39] M. Arellano and J. Hahn, "Understanding bias in nonlinear panel models: some recent developments," in *Advances in Economics and Econometrics, Ninth World Congress*, R. Blundell, W. Newey, and T. Persson, Eds., Cambridge University Press, 2006.

[40] M. Arellano and J. Hahn, "A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects," Working Papers, CEMFI, 2006.

[41] J. O. Berger, B. Liseo, and R. L. Wolpert, "Integrated likelihood methods for eliminating nuisance parameters," *Statistical Science*, vol. 14, no. 1, pp. 1–28, 1999.

[42] C. A. Bester and C. Hansen, "A penalty function approach to bias reduction in non-linear panel models with fixed effects," *Journal of Business and. Economic Statistics*, vol. 27, pp. 29–51, 2009.

[43] C. A. Bester and C. Hansen, "Bias reduction for Bayesian and frequentist estimators," unpublished.

[44] E. G. Tsionas, S. C. Kumbhakar et al., "Firm-heterogeneity, persistent and transient technical inefficiency," MPRA REPEC working paper, 2011.

[45] Y. Mundlak, "Empirical production function free of management bias," *Journal of Farm Economics*, vol. 43, no. 1, pp. 44–56, 1961.

[46] E. G. Tsionas, "Stochastic frontier models with random coefficients," *Journal of Applied Econometrics*, vol. 17, no. 2, pp. 127–147, 2002.

[47] S. C. Kumbhakar, "Modeling allocative inefficiency in a translog cost function and cost share equations: an exact relationship," *Journal of Econometrics*, vol. 76, no. 1-2, pp. 351–356, 1997.

[48] S. C. Kumbhakar and E. G. Tsionas, "Measuring technical and allocative inefficiency in the translog cost system: a Bayesian approach," *Journal of Econometrics*, vol. 126, no. 2, pp. 355–384, 2005.

[49] R. Colombi, G. Martini, and G. Vittadini, "A stochastic frontier model with short-run and long-run inefficiency random effects," Working Papers, Department of Economics and Technology Management, University of Bergamo, 2011.

[50] R. Färe, S. Grosskopf, and C. A. K. Lovell, *The Measurement of Efficiency of Production*, Kluwer Nijhoff Publishing, Boston, Mass, USA, 1985.

[51] S. C. Ray, *Data Envelopment Analysis*, Cambridge University Press, Cambridge, UK, 2004.

[52] A. Alvarez, C. Arias, and S. C. Kumbhakar, *Empirical Consequences of Direction Choice in Technical Efficiency Analysis*, SUNY, Binghamton, NY, USA, 2003.

[53] S. E. Atkinson and C. Cornwell, "Estimation of output and input technical efficiency using a flexible functional form and panel data," *International Economic Review*, vol. 35, pp. 245–256, 1994.

[54] T. Beard, S. Caudill, and D. Gropper, "Finite mixture estimation of multiproduct cost functions," *Review of Economics and Statistics*, vol. 73, pp. 654–664, 1991.

[55] T. R. Beard, S. B. Caudill, and D. M. Gropper, "The diffusion of production processes in the U.S. Banking industry: a finite mixture approach," *Journal of Banking and Finance*, vol. 21, no. 5, pp. 721–740, 1997.

[56] J. A. Hagenaars and A. L. McCutcheon, *Applied Latent Class Analysis*, Cambridge University Analysis, New York, NY, USA, 2002.

[57] G. McLachlan and D. Peel, *Finite Mixture Models*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, New York, NY, USA, 2000.

[58] R. Tibshirani, *Local likelihood estimation*, Ph.D. thesis, Stanford University, 1984.

[59] P. Gozalo and O. Linton, "Local nonlinear least squares: using parametric information in nonparametric regression," *Journal of Econometrics*, vol. 99, no. 1, pp. 63–106, 2000.

Advances in
Operations Research

Advances in
Decision Sciences

Journal of
Applied Mathematics

Algebra

Journal of
Probability and Statistics

The Scientific
World Journal

International Journal of
Differential Equations

International Journal of
Combinatorics

Advances in
Mathematical Physics
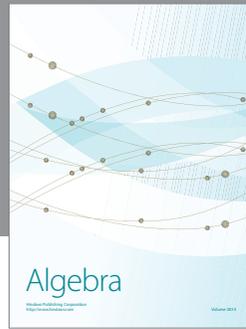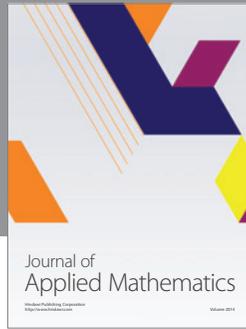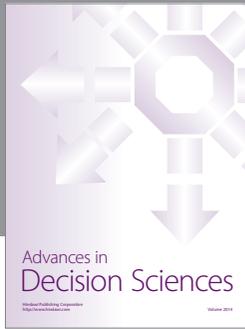
Hindawi

Submit your manuscripts at
http://www.hindawi.com

Journal of
Complex Analysis

Journal of
Mathematics

Mathematical Problems
in Engineering

Abstract and
Applied Analysis

Discrete Dynamics in
Nature and Society

International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces

International Journal of
Stochastic Analysis

Journal of
Optimization