A Survey of Stochastic Frontier Models and Likely Future Developments

Christine Amsler, Young Hoon Lee, and Peter Schmidt*

This paper summarizes the literature on stochastic frontier production function models. It covers the definition of technical efficiency, the basic cross-sectional stochastic frontier model, and the stochastic frontier model with panel data and time-invariant as well as time-varying technical inefficiency. It also discusses models in which technical inefficiency depends on explanatory variables. Finally, it discusses the problem of inference on the inefficiencies and makes some predictions about likely future developments in the field.

Keywords: Stochastic frontiers, Production functions, Technical efficiency, Panel data, Efficiency measurement

JEL Classification: C10, C20, D24

I. Introduction

This paper is a survey of stochastic frontier models. Stochastic frontier models were introduced by Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977). Since then a very large

* Associate Professor, Department of Economics, Michigan State University, USA, (E-mail) amsler@msu.edu; Professor, Department of Economics, Sogang University, Sinsu-dong #1, Mapo-gu, Seoul 121-742, Korea, (Tel) +82-2-705-8772, (Fax) +82-2-704-8599, (E-mail) yhnlee@sogang.ac.kr; Professor, Department of Economics, Michigan State University, USA, School of Economics, Yonsei University, Korea, (E-mail) schmidtp@msu.edu, respectively. The second author acknowledges that this research is financially supported by the 2008 Sogang University Research Fund (10045). Paper presented at the 16th *Seoul Journal of Economics* International Symposium held at Seoul National University, Seoul, 27 November 2008.

literature has developed on this topic, and a comprehensive survey would be at least book-length (*e.g.*, Kumbhakar and Lovell 2000). Of necessity our survey will be selective. Not surprisingly, we will pay the most attention to those aspects of the literature to which we have contributed. The omission of other topics does not mean that we consider them unimportant.

The plan of the paper is as follows. Section 2 defines technical efficiency, the concept whose measurement is the point of these models. Section 3 considers the basic cross-sectional stochastic frontier model, and Section 4 discusses models in which technical inefficiency depends on explanatory variables. Section 5 covers the stochastic frontier model with panel data and time-invariant technical inefficiency. Section 6 discusses panel data models in which technical inefficiency changes over time. Section 7 considers the problem of inference on the inefficiencies. Finally, Section 8 gives our concluding remarks and some predictions about likely future developments in the field.

II. Definition of Technical Efficiency and Inefficiency

Technical inefficiency can be defined as the failure to produce maximal possible output, given input levels. Comparing actual output to maximal possible output gives rise to an "output based" inefficiency measure. Alternatively, technical inefficiency can be thought of as the failure to use the minimal possible inputs to produce a given output level. Comparing the actual inputs to the minimal possible inputs gives rise to an "input based" inefficiency measure.

Figure 1 illustrates the input-based definition of technical efficiency proposed in the classic paper by Farrell (1957). Suppose that we have one output and two inputs, so that the production function is $y=f(X_1, X_2)$ where y is output and X_1 and X_2 are inputs. Suppose that a firm produces output y_0 using input quantities (X_1^1, X_2^1). This is represented as point B on the graph. Point B is above the isoquant for output level y_0, $Isoq(y_0)$. It could produce output level y_0 at point A, which has the same input proportions as B but is on $Isoq(y_0)$. The input-based measure of the technical efficiency of this firm is defined as OA/OB (where OA and OB are the distances of points A and B from the origin), and its input-based technical inefficiency is 1−OA/OB. More formally, the Farrell input-based efficiency measure is defined as

$$TE_I = Min\{\lambda \ni (y,\ \lambda X)\ is\ feasible\}.$$

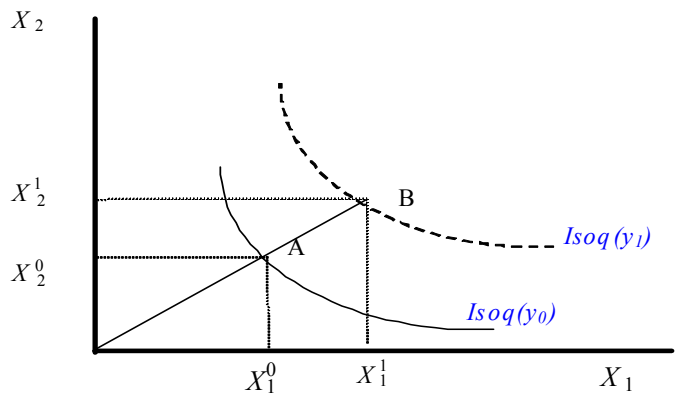

FIGURE 1
THE INPUT-BASED MEASURE OF TECHNICAL EFFICIENCY

Alternatively, the firm using inputs (X_1^1, X_2^1) could increase its output to y_1, the output level corresponding to the isoquant on which point B is located. The output based measure of the technical efficiency of this firm would be y_0/y_1, the ratio of actual output to potential output, given the input levels; its output-based technical inefficiency would be $1-y_0/y_1$. More formally, the output-based efficiency measure is defined as

$$TE_O = Min\{\theta \ni (y/\theta,\ X)\ \ is\ feasible\}.$$

In this paper, we will consider output-based efficiency measures. Also, we will consider only the case of a single output. In this case it is natural and convenient to think in terms of production functions (rather than the corresponding isoquants). The production frontier is the production function that gives maximal possible output, given inputs, and technical efficiency is measured simply as the ratio of actual output to the frontier output, given the input quantities used.

III. Cross Section Stochastic Frontier Models

The first production frontier models were deterministic. Let Y be output in levels and y be output in logs. The frontier for y is $f(x)$, and $y \leq f(x)$: actual output is always less than or equal to the frontier. We

express this inequality with a one-sided (non-positive) additive error term: $y=f(x)-u$, with $u\geq 0$. Exponentiating, we have $Y=e^{y}=e^{f(x)}e^{-u}$. Therefore, $e^{-u}=Y/e^{f(x)}=$ actual output divided by possible output = technical efficiency (TE) and technical inefficiency $=1-e^{-u}$. However, $1-e^{-u}$ is approximately equal to u (the approximation is quite good for small values of u) and often we will simply refer to u as technical inefficiency.

Empirically, we will generally want to use a linear function (which includes Cobb-Douglas or translog technologies), and the linear deterministic production frontier model is

$$y_i=\alpha_0+x_i'\beta-u_i,\quad u_i\geq 0,\quad i=1,\ 2,\ \cdots,\ N \tag{1}$$

where y_i is log output, x_i is a $K\times 1$ vector of inputs (generally in logs), β is the vector of regression coefficients and u_i is technical inefficiency. The objective is not only to estimate β but also to estimate u_i.

Aigner and Chu (1968) estimate the frontier using linear and quadratic programming:

$$LP:\ Min\sum_{i=1}^{N}|y_i-\alpha_0-x_i'\beta| \quad \text{subject to } y_i\leq\alpha_0+x_i'\beta \text{ for all } i$$

$$QP:\ Min\sum_{i=1}^{N}[y_i-\alpha_0-x_i'\beta]^2 \quad \text{subject to } y_i\leq\alpha_0+x_i'\beta \text{ for all } i,$$

where the minimization is with respect to α_0 and β. Technical inefficiency of firm i is calculated as the difference between actual output and the estimated frontier.

Stochastic production frontier models, proposed by Aigner, Lovell, and Schmidt (1977) (hereafter ALS1977) and Meeusen and van den Broeck (1977), make the production frontier stochastic. The model is of the form:

$$y_i=\alpha_0+x_i'\beta+\varepsilon_i,\quad \varepsilon_i=v_i-u_i,\quad i=1,\ 2,\ \cdots,\ N \tag{2}$$

The "composed error" $\varepsilon_i=v_i-u_i$ is made up of both a statistical noise term v_i and the technical inefficiency $u_i\geq 0$. The frontier is $\alpha_0+x_i'\beta+v_i$, which is stochastic because it includes v_i. Identification of this model requires strong assumptions. Specific distributional assumptions need to be made for v and for u. For example it is often assumed

that v is normal and that u is half-normal. Also v, u and x are assumed to be independent. This is a strong assumption since it rules out the possibility that a firm's input choices are influenced by its level of technical inefficiency.

The estimates of the parameters of the model are usually obtained by maximum likelihood estimation (MLE), that is, by maximization of the likelihood function:

$$\ln L = \sum_{i=1}^{N} \ln k(y_i - \alpha_0 - x_i'\beta) \tag{3}$$

where $k(\varepsilon) = \int_0^\infty h(u,\ \varepsilon+u)\,du$, $h(u,\ v) = f(v)\,g(u)$ and $f(v)$ and $g(u)$ are the probability density functions of u and v, respectively.

Different models can be generated by different assumptions about the distribution of u. For example, ALS1977 considered the case that u was exponential as well as the case that it was half-normal. Stevenson (1980) assumed a general truncated normal distribution and Greene (1980, 1990) assumed a gamma distribution. Empirically, the choice of distributional assumptions matters; different assumptions yield different results. Kumbhakar and Lovell (2000) discuss this issue at some length. Only very recently (Wang, Amsler, and Schmidt 2008) have goodness of fit tests been developed to allow one to test these distributional assumptions.

The main focus is on the estimation of technical inefficiency. We cannot simply calculate technical inefficiency by subtracting y_i from the frontier, since the frontier contains the statistical noise v_i term which is not observable. We can estimate ε_i as $\hat{\varepsilon}_i = y_i - \hat{\alpha}_0 - x_i'\hat{\beta}$ but this is an estimate of $\varepsilon_i = v_i - u_i$, and we need somehow to separate u_i from v_i. The standard estimate, suggested by Jondrow, Lovell, Materov, and Schmidt (1982), is the conditional expectation of u_i given $\varepsilon_i = v_i - u_i$, evaluated at the fitted values of ε_i (*i.e.*, $\hat{\varepsilon}_i$) and the estimated values of the parameters. With a half normal assumption for u, the estimate is

$$\hat{u}_i = E(u_i \mid \varepsilon_i) = \mu_i^* + \sigma_* \left[\frac{\phi(-\mu_i^*/\sigma_*)}{1 - \Phi(-\mu_i^*/\sigma_*)}\right] \tag{4}$$

where $\mu_i^* = -\varepsilon_i\,\sigma_u^2/\sigma^2$, $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$, $\sigma^2 = \sigma_u^2 + \sigma_v^2$ and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and cumulative distribution functions, respectively.

It is obvious that $\hat{u}_i$ is not a consistent estimate of u_i since we need to estimate N "parameters" based on N observations. In fact $\hat{u}_i$ does not converge in probability to any limit, since the variability of v_i remains no matter how large N is. To put this another way, $\text{var}(u_i|\varepsilon_i) > 0$ independently of N. The expected value of $\hat{u}_i$ equals $E(u_i)$ since $E(\hat{u}_i) = E[E(u_i|\varepsilon_i)] = E(u_i)$ by the law of iterated expectations. However, $\hat{u}_i$ is not unbiased in the conditional sense: $E(\hat{u}_i|u_i) \neq u_i$. Rather, as shown by Wang and Schmidt (2008), $\hat{u}_i$ is a shrinkage toward the mean of u.

In fact, Jondrow, Lovell, Materov, and Schmidt showed that u_i conditional on ε_i is distributed as $N^+(\mu_i^*, \sigma_*^2)$. Horrace and Schmidt (1996) showed how to construct confidence intervals for technical inefficiencies using this distribution.

IV. Models with Inefficiency That Depends on Explanatory Variables

In this Section, we consider stochastic frontier models in which observable characteristics of the firms affect their levels of technical inefficiency. As before, let y be log output, let x be a vector of functions (usually logs) of inputs, and $u \geq 0$ be the one-sided error reflecting technical inefficiency. Now we also specify a set of variables z that affect u. Generally the variables in z are either functions of inputs or measures of the environment in which the firm operates. Thus it is possible that x and z overlap. We can write u as $u(z, \delta)$ to reflect its dependence on z and some parameters δ. Different models correspond to different specifications of $u(z, \delta)$.

We will say that the model has the scaling property if

$$u(z, \delta) = h(z, \delta) \cdot u^* , \tag{5}$$

where $h(z, \delta) \geq 0$, and where $u^* \geq 0$ has a distribution that does not depend on z. We will call $h(z, \delta)$ the scaling function and u^* the basic random variable. In models with the scaling property, changes in z change the scale but not the shape of $u(z, \delta)$. The scaling property is discussed in more detail in Álvarez, Amsler, Orea, and Schmidt (2006).

A prominent example of a model that has the scaling property is the scaled half-normal model, or RSCFG model, of Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill, Ford, and Gropper (1995). In this model it is assumed that u is distributed as

$N^+(0,\ \sigma(z,\ \theta)^2)$. This is equivalent to assuming that u is distributed as $\sigma(z,\ \theta)$ times a variable distributed as $N^+(0,\ 1)$. Thus $\sigma(z,\ \theta)$ corresponds to the scaling function $h(z,\ \delta)$ above. The various papers make different suggestions for the function $\sigma(z,\ \theta)$. For example, Caudill, Ford, and Gropper specify $\sigma(z,\ \gamma)=\exp(z'\gamma)$.

A well known and popular model that does not have the scaling property is the KGMHLBC model of Kumbhakar, Ghosh, and McGuckin (1991), Huang and Liu (1994), and Battese and Coelli (1995). This is a truncated normal model in which the mean of the pre-truncation normal depends on z and some parameters θ. That is, u is distributed as $N^+(\mu(z,\ \theta),\ \sigma^2)$. Since the degree of truncation varies with μ, the shape of the distribution of u changes when z changes. All three of the papers listed above suggest a linear specification of μ: $\mu=\alpha+z'\delta$.

In the RSCFG model, the expectation of u is monotonic in z so long as the specification for σ is monotonic in z. Similarly, in the KGMHLBC model, the expectation of u is monotonic in z (though the relationship is complicated) so long as the specification of μ is monotonic in z. Wang (2002) proposes a model in which the relationship of the expectation of u to z could be non-monotonic. He does this by assuming that the distribution of u is $N^+(\mu,\ \sigma^2)$, where both μ and σ depend on z and some parameters. Specifically, he assumes that $\mu=z'\delta$ and $\sigma^2=\exp(z'\gamma)$.

In Wang's model the z each have two different coefficients, one for the mean and one for the variance. In the RSCFG model and the KGMHLBC model, the z each have only one coefficient. If one wishes to restrict attention to models in which each of the z has only one coefficient, scaling models may be attractive, primarily because the coefficients in the scaling function are easy to interpet. In particular, a reasonable competitor to the RSCFG and KGMHLBC models would be the scaled Stevenson model, which is simply the scaled version of the truncated normal model of Stevenson (1980).

Once the error distribution is specified, the model is estimated by maximum likelihood. Wang and Schmidt (2002) refer to this as a one step procedure. This is different from a two step procedure in which the steps are: (i) Estimate a model ignoring the effect of z on u. (ii) Fit another model using z to explain the estimated inefficiencies $\hat{u}$. Two step procedures are not recommended because, as Wang and Schmidt show, there are serious biases at each step.

V. Panel Data Stochastic Frontier Models with Time-Invariant Inefficiency

Cross-sectional stochastic frontier models rely on two kinds of strong assumptions. Specific distributional assumptions need to be made for noise and for technical inefficiency; and the errors must be independent of the inputs. Even with these strong assumptions, the estimates of technical inefficiency are not consistent. Panel data allow us to relax some or all of these assumptions, and they allow consistent estimation of technical inefficiency. However, these advantages come at a price, because they depend on the additional assumption that technical inefficiency is time invariant, or that it varies in a restricted way over time. In this Section we consider the case that technical inefficiency is time invariant.

Pitt and Lee (1981) and Schmidt and Sickles (1984) were the first to consider stochastic frontier models with panel data. They considered the model with time invariant inefficiencies:

$$y_{it}=\alpha_0+x'_{it}\beta-u_i+v_{it}, \qquad i=1,\ 2,\ \cdots,\ N, \quad t=1,\ 2,\ \cdots,\ T \tag{6}$$

This equation can be converted to a standard panel data model:

$$y_{it}=x'_{it}\beta+\alpha_i+v_{it}, \qquad i=1,\ 2,\ \cdots,\ N, \quad t=1,\ 2,\ \cdots,\ T \tag{7}$$

where $\alpha_i=\alpha_0-u_i$. Note that $\alpha_i\leq\alpha_0$ and $\alpha_i=\alpha_0$ only when $u_i=0$. Therefore, a smaller individual-specific intercept implies a lower level of technical efficiency.

It is clear that $TE_i=\exp(-u_i)=\exp(\alpha_i-\alpha)$ is an absolute efficiency measure, in the sense that it compares the firm's efficiency to the absolute standard of $TE=1$. We can also consider relative efficiency measures that compare the firm's efficiency to that of the most efficient of the N firms in the sample. To define such measures, we write the intercepts in ranked order

$$\alpha_{(1)}\leq\alpha_{(2)}\leq\cdots\leq\alpha_{(N)}\leq\alpha_0 \tag{8}$$

so that (N) is the index of the best firm in the sample and its intercept is $\alpha_{(N)}$. We then write the technical inefficiency terms (the u_i) in reverse ranked order, so that

$$0 \leq u_{(N)} \leq u_{(N-1)} \leq \cdots \leq u_{(1)} \tag{9}$$

With these definitions it is the case that $\alpha_{(i)} = \alpha_0 - u_{(i)}$. Now we can define the relative efficiency measures $u_i^* = u_i - u_{(N)} = \alpha_{(N)} - \alpha_i \geq 0$ and $TE_i^* = \exp(-u_i^*) \leq 1$. Note that $u_i^* \leq u_i$ and $TE_i^* \geq TE_i$; efficiency levels are higher when measured relative to the best of the N firms than when they are measured relative to the absolute standard of $TE = 1$.

A. Estimation with Distributional Assumptions

Pitt and Lee (1981) considered the model (7) under essentially the same assumptions as in the cross-sectional stochastic frontier model. This treatment of the model requires distributional assumptions for the two error terms: $v_i \sim iid\ N(0, \sigma_v^2)$, $u_i \sim iid\ N^+(0, \sigma_u^2)$ (or some other one-sided distribution), and u, v, and x are independent of each other. They derived the joint density function of $\varepsilon_{it} = v_{it} - u_i$ for all t from the assumed densities of u_i, v_{i1}, $\cdots$, v_{iT}, and then estimated the model by MLE.

To estimate technical efficiency for a firm, Battese and Coelli (1988) suggested the following. The estimate of u_i is $\hat{u}_i = E(u_i | \varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT}) = E(u_i | \varepsilon_i) = E(u_i | \bar{\varepsilon}_i)$, where $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT})'$ and $\bar{\varepsilon}_i = 1/T \sum \varepsilon_{it}$. These are evaluated at the estimated values of the ε_{it} and the estimated values of the other parameters. Similarly the estimate of TE_i is $\hat{TE}_i = E[\exp(-u_i) | \varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT})$. The formula for $\hat{u}_i$ is the same as in equation (4), except that ε_i and σ_v^2 are replaced by $\bar{\varepsilon}_i$ and σ_v^2/T, respectively. Note that this estimate measures absolute efficiencies since we are measuring the distance of $\hat{u}_i$ from zero, not from $u_{(N)}$.

B. Fixed Effects Estimation

This estimation method considers equation (7) as the regression model. We treat the α_i as fixed, so we do not need to impose any distributional assumptions. Also we allow correlation between technical inefficiency and the inputs. But we assume the strict exogeneity of the noise, in the sense that $E[v_{it} | x_{i1}, x_{i2}, \cdots, x_{iT}) = 0$.

This model can be estimated using the conventional "fixed effects" or "within" estimator. This can be defined in three different but equivalent ways. The first is ordinary least squares (OLS) on equation (7), treating the parameters as β, α_1, $\cdots$, α_N. The second is OLS with dummies for the N firms:

$$y=X\beta+D\alpha+v, \quad D=I_N\otimes 1_T,\ \alpha=(\alpha_1,\ \cdots,\ \alpha_N)' \tag{10}$$

where 1_T is a $T\times 1$ vector of ones. The third is OLS after the within transformation:

$$(y_{it}-\bar{y}_i)=(x_{it}-\bar{x}_i)'\beta+(v_{it}-\bar{v}_i), \quad i=1,\ 2,\ \cdots,\ N, \quad t=1,\ 2,\ \cdots,\ T \tag{11}$$

where $\bar{y}_i=1/T\sum_t y_{it}$ and $\bar{x}_i$, $\bar{v}_i$ and are defined similarly.

The individual α_i are estimated as the coefficients of the dummies in equation (10). Equivalently, $\hat{\alpha}_i=\bar{y}_i-\bar{x}_i'\hat{\beta}$ where $\hat{\beta}$ is the within estimator.

Note that the coefficients of time invariant regressors are not identified in this approach. They are linearly dependent with the individual dummies in equation (10), or equivalently they become zero after the within transformation. For example, the input "land" might be constant in panel data for farms, and then it cannot be included in the model.

The estimator of the production function parameters ($\hat{\beta}$) is consistent and asymptotically normal as $NT\to\infty$ (either $N\to\infty$ or $T\to\infty$). The estimator of the firm specific intercepts ($\hat{\alpha}_i$) is consistent as $T\to\infty$. This condition is necessary for $p\lim \bar{v}_i=0$ in the representation $\hat{\alpha}_i=\alpha_i-\bar{x}_i'(\hat{\beta}-\beta)+\bar{v}_i$. This is somewhat unfortunate since the assumption that technical efficiency is time-invariant is less plausible when T is large.

Schmidt and Sickles (1984) suggested the following estimates of technical inefficiency, based on the within estimates:

$$\hat{\alpha}_0=\max_j \hat{\alpha}_j,\ \hat{u}_i=\hat{\alpha}_0-\hat{\alpha}_i \text{ and } \hat{TE}_i=\exp(-\hat{u}_i) \tag{12}$$

If we think of N as fixed, these estimates are clearly estimates of relative technical inefficiency. That is, as $T\to\infty$ with N fixed, $\hat{\alpha}_0$ is a consistent estimator of $\alpha_{(N)}$, $\hat{u}_i$ is a consistent estimator of u_i^*, $\hat{TE}_i$ and is a consistent estimator of TE_i^*. However, as $N\to\infty$ relative and absolute efficiencies should become the same. That is, as $N\to\infty$, $u_{(N)}\to_p 0$ so that $\alpha_{(N)}\to_p \alpha_0$, $u_i^*\to_p u_i$ and $TE_i^*\to_p TE_i$. Thus we expect that, as both $N\to\infty$ and $T\to\infty$, the estimates in equation (12) should be consistent estimates of absolute efficiency. However, Park and Simar (1994) showed that consistent estimation of absolute efficiency requires $N\to\infty$ and $T\to\infty$, but also the additional condition that $1/\sqrt{\mathrm{T}}\ln N\to 0$. Thus it is required that N grows slowly relative to T.

It is important to realize that $\hat{\alpha}_0 = \max_j \hat{\alpha}_j$ is biased upward as an estimate of $\alpha_{(N)} = \max_j \alpha_j$, for finite T. This is true because $\hat{\alpha}_0 \geq \hat{\alpha}_{(N)}$ and $E[\hat{\alpha}_{(N)}] = \alpha_{(N)}$, and basically reflects the fact that the largest $\hat{\alpha}_i$ is more likely to contain positive estimation error than negative. This bias is larger when T is smaller, when N is larger, and when the variance of statistical noise is larger relative to the variance of technical inefficiency. It implies that in finite samples $\hat{u}_i^*$ is biased upward as an estimate of u_i^* and $\hat{TE}_i^*$ is biased downward as an estimate of TE_i^*. Empirically, the fixed effects approach typically yields lower levels of estimated technical efficiency than the MLE approach.

VI. Panel Data Stochastic Frontier Models with Time-Varying Efficiency

The stochastic frontier production model with time-varying efficiency is defined by

$$y_{it} = \alpha_t + x_{it}\beta + v_{it} - u_{it} = x_{it}\beta + \alpha_{it} + v_{it}, \quad i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T \quad (13)$$

where $\alpha_{it} = \alpha_t - u_{it}$ is the intercept for firm i in period t. Note that we allow a time-varying common intercept, α_t. Clearly we cannot expect to estimate all of the u_{it} (or α_{it}) without some assumptions about their temporal pattern or correlation structure. Therefore, different models have emerged as different choices for the form of α_{it} (or, equivalently, u_{it}).

Cornwell, Schmidt, and Sickles (1990, CSS) proposed the model in which $\alpha_{it} = W_t'\delta_i$, where W_t is a vector of observed functions of time. They considered the specific case that α_{it} was quadratic in t, so that $W_t = [1, t, t^2]$ and $\alpha_{it} = \delta_{i0} + \delta_{i1}t + \delta_{i2}t^2$. Thus, the intercept for each firm is quadratic in time, but the form of the quadratic varies over firms.

Kumbhakar (1990) and Battese and Coelli (1992, BC) suggested the model that $u_{it} = \theta_t(\eta)u_i$. Here $\theta_t(\eta)$ depends on t and on some parameters η. It determines the temporal pattern of technical inefficiency. Specifically, Kumbhakar set

$$\theta_t(b, c) = [1 + \exp(bt + ct^2)]^{-1} \text{ and BC set } \theta_t(\eta) = \exp[\eta(T - t)].$$

Lee and Schmidt (1993, LS) and Ahn, Lee, and Schmidt (2001) considered a model that is similar to the models of Kumbhakar and

BC, but more flexible. They set $\alpha_{it}=\theta_t\alpha_i$, where the θ_t are unrestricted parameters to be estimated. Thus the temporal pattern of technical inefficiency is completely unrestricted. This model nests the models of Kumbhakar (1990) and BC in which inefficiencies vary over time in specific exponential forms. Of course, there are more parameters to estimate, since η contains the $T-1$ parameters θ_t for $t=2, \cdots, T$, with a normalization that $\theta_1=1$.

The models of the previous two paragraphs imply that the temporal pattern of inefficiency is the same for each firm, though the magnitude varies with u_i or α_i. (This statement assumes that the α_i are all of the same sign.) The CSS model does not have that property. Another model that does not have that property was proposed by Cuesta (2000), who assumed $\alpha_{it}=\theta_{it}\alpha_i$ where $\theta_{it}=\exp[\eta_i(T-t)]$. Now η_i depends on i, whereas in the BC model it did not. Another model that does not have the property that the temporal pattern of technical inefficiency is the same for all firms is the group-specific model of Lee (2006, 2009). The firms are put into groups, such that all of the firms in a given group have the same temporal pattern of inefficiency, but this pattern differs across groups. Specifically, $\alpha_{it}=\theta_{gt}\alpha_i$ where $i\in group\ g$. θ_{gt} can be treated as a parameter or alternatively a functional form such as $\theta_{gt}=\exp[\eta_g(T-t)]$ can be imposed on θ_{gt}.

Ahn, Lee, and Schmidt (2007, ALS) applied a multi-factor model to the stochastic frontier model. This model was suggested as an extension to the single factor model of LS and Ahn, Lee, and Schmidt (2001). The multi-factor model specifies

$$\alpha_{it}=\theta_{1t}\delta_{1i}+\theta_{2t}\delta_{2i}+\cdots+\theta_{pt}\delta_{pi}=\sum_{j=1}^{p}\theta_{jt}\delta_{ji} \tag{14}$$

Therefore, this model reduces to LS if the number of factors is one ($p=1$). The model also nests CSS as the special case that $p=3$ and $\theta_{1t}=1$, $\theta_{2t}=t$ and $\theta_{3t}=t^2$. Therefore this model nests all of the specifications of BC, Kumbhakar (1990), CSS, and LS.

We now turn to the estimation of the models. Kumbhakar (1990) and BC suggested random effects estimation in which a distributional assumption was made for u_i. The same approach can be applied to all of the models in which there is a single u_i (or α_i) per firm, that is, to all of the models listed above except the CSS model and the multifactor model. The estimates of the parameters of the model are consistent as $N\rightarrow\infty$ with T fixed. Intuitively, these models are similar in spirit to cross-sectional models and a large number of firms is

required to consistently estimate the parameters of the distribution of u.

All of the models listed above can also be estimated by fixed effects. For those models where the number of parameters does not depend on T (*i.e.*, for all of the above models except the factor models), the fixed effects estimates of the parameters of the model are clearly consistent as $T \to \infty$ with N fixed. Comparing this to the discussion of the previous paragraph, it is reasonable to argue that random effects models based on a distributional assumption are natural when N is large and T is small, whereas fixed effects estimates are natural when N is small and T is large. However, fixed effects estimates can also be used when N is large, where the motivation would be to avoid making a distributional assumption for inefficiency. In that case, there is a potential "incidental parameters problem" because the number of parameters increases with sample size (N). However, CSS show that there is no incidental parameters problem in their model. Han, Orea, and Schmidt (2005) provide a valid fixed effects treatment of models like the Kumbhakar (1990) and BC models. For factor models, the relevant asymptotic theory is provided in Ahn, Lee, and Schmidt (2001), Ahn, Lee, and Schmidt (2007), Bai and Ng (2002), and Bai (2003).

Once we have consistent estimates of the α_{it}, estimated technical inefficiency is obtained in a manner similar to the case of fixed effects and time-invariant technical inefficiency. We define

$$\hat{\alpha}_t = \max_j \hat{\alpha}_{jt}, \ \hat{u}_{it} = \hat{\alpha}_t - \hat{\alpha}_{it} \text{ and } \hat{TE}_{it} = \exp(-\hat{u}_{it}). \tag{15}$$

We can now make statements similar to those we made in the time-invariant case. Our estimates of relative technical inefficiency should be consistent as $T \to \infty$. Furthermore, as $N \to \infty$ relative and absolute technical inefficiency should become the same. Therefore, as both $N \to \infty$ and $T \to \infty$, we hope to obtain a consistent estimate of absolute technical inefficiency. However, there is no rigorous proof of this result (similar in spirit to Park and Simar 1994) currently available, and it is not known whether the additional condition needed in the time-invariant case ($1/\sqrt{T} \ln N \to 0$) also applies here.

VII. Inference on Inefficiencies

So far in this paper we have discussed the estimation of technical inefficiency. That discussion is in terms of point estimates. Now we will discuss how to perform inference on inefficiency levels. Specifically we

will consider the construction of confidence intervals for u. We will discuss the cross-sectional case and the case of panel data with time-invariant technical inefficiency. The extension of this analysis to cases in which inefficiency depends on explanatory variables, or varies over time, is tedious but not conceptually difficult.

A. Inference with a Distributional Assumption

The simplest case to consider is the original cross-sectional stochastic frontier model in which the error is $\varepsilon = v - u$ where v is normal and u is half normal. In this case the point estimate of u is $\hat{u} = E(u \mid \varepsilon)$, evaluated at $\varepsilon = \hat{\varepsilon}$, as proposed by Jondrow, Lovell, Materov, and Schmidt (1982). However, Horrace and Schmidt (1996) observed that Jondrow, Lovell, Materov, and Schmidt had additionally shown that the distribution of u conditional on ε is $N^{+}(\mu_{*}, \sigma_{*}^{2})$ where and $\mu_{*} = \varepsilon\sigma_{u}^{2}/(\sigma_{u}^{2} + \sigma_{v}^{2})$ and $\sigma_{*}^{2} = \sigma_{u}^{2}\sigma_{v}^{2}/(\sigma_{u}^{2} + \sigma_{v}^{2})$. Therefore this distribution, evaluated at $\varepsilon = \hat{\varepsilon}$, can be used to create confidence intervals for u. These should be accurate since the only approximation involved is the fact that we must evaluate the conditional distribution at estimated values $(\hat{\varepsilon},\ \hat{\sigma}_{u}^{2},\ \hat{\sigma}_{v}^{2})$.

This procedure also extends to the case of panel data with time invariant inefficiency and a distributional assumption. One uses the distribution of u conditional on $(\varepsilon_1, \cdots, \varepsilon_T)$, which is also a truncated normal distribution, given by Battese and Coelli (1988).

B. Bayesian Inference

The Jondrow, Lovell, Materov, and Schmidt result has a Bayesian flavor to it. It treats the parameters of the model as known (*i.e.*, it treats the estimated parameters as if they were the true parameters) and conditions on ε, which would be equivalent to conditioning on the data (y and x) if the parameters were known. A true Bayesian procedure would put a prior distribution on the parameters and on u (*i.e.*, on each of the u_i) and would condition on the data. Bayesian analyses of the stochastic frontier model have been proposed and described in a series of papers, notably Koop, Steele, and Osiewalski (1995) and Koop, Osiewalski, and Steele (1997).

Kim and Schmidt (2000) have compared Bayesian and classical analyses and found little difference in results, if the assumptions on u match up. For example, MLE applied to a model in which u is assumed to be exponential is not very different from a Bayesian

analysis with an exponential prior for u. As another example, Koop, Osiewalski, and Steele (1997) define a "Bayesian fixed effects model" in the setting of panel data, and this gives results that are similar to those from the fixed effects analysis discussed in Section 5.2 above.

There are some computational advantages to being a Bayesian, especially the availability of Markov Chain Monte Carlo sampling methods. There is no need for the numerical maximization of a likelihood function, as there is with classical MLE. From a classical point of view, specifying a prior for the parameters is troublesome, but for large samples the data should dominate the prior, and one can argue that these "asymptotics" (that the posterior depends little on the choice of prior) have the advantage of being visible.

C. Multiple Comparisons with the Best

Multiple comparisons with the best (MCB) is a statistical technique that yields confidence intervals for differences in parameter values between all populations and the best population. In the context of fixed effects estimation with panel data, Horrace and Schmidt (1996, 2000) have suggested its use to construct confidence intervals for the relative technical inefficiencies $u_i^* = u_i - u_{(N)} = \alpha_{(N)} - \alpha_i$, which are indeed differences from the best.

As above, let firms be indexed by $i = 1, 2, \cdots, N$ and let (N) be the index of the best firm. MCB constructs a set S of possibly best populations, and a set of intervals (L_i, U_i) such that

$$P[(N) \in S \text{ and } L_i \leq \alpha_{(N)} - \alpha_i \leq U_i \text{ for all i}] \geq 1 - c \tag{16}$$

where $1-c$ is a chosen confidence level (*e.g.*, 0.95). Thus with a given confidence level we have a set of populations that includes the best, and joint confidence intervals for all differences from the best. MCB was developed by Hsu (1981, 1984) and Edwards and Hsu (1983). A general exposition can be found in Hochberg and Tamhane (1987), Hsu (1996) and Horrace and Schmidt (2000).

To perform MCB, we need an estimate of the vector $(\alpha_1, \cdots, \alpha_N)'$ that is normally distributed, with a variance matrix that is known up to a constant (scale). In typical MCB applications to the efficiency measurement problem, the fixed effects estimates $\hat{\alpha}_i$ will be used. The normality of these estimates requires either that the errors v_{it} are normal, or that T is big enough that a central limit theorem applies. However, because

this is a fixed-effects treatment, no assumption about the distribution of the u_i is needed.

MCB produces confidence intervals that are quite conservative. That is, they are valid, in the sense that their coverage rate is indeed at least $1-c$, but they are often very wide.

D. Bootstrapping

We can use bootstrapping to construct confidence intervals for functions of the fixed effects estimates. The inefficiency measures $\hat{u}_i^*$ are functions of the fixed effects estimates and so bootstrapping can be used for inference on these measures.

We begin with a very brief discussion of bootstrapping in the general setting in which we have a parameter θ, and there is an estimate $\hat{\theta}$ based on a sample $z_1, \cdots, z_n$ of i.i.d. random variables. The estimator $\hat{\theta}$ is assumed to be regular enough so that $\sqrt{n}\,(\hat{\theta}-\theta)$ is asymptotically normal. The following bootstrap procedure will be repeated many times, say for $b=1, \cdots, B$ where B is large. For iteration b, construct pseudo data $z_1^{(b)}, \cdots, z_n^{(b)}$ by sampling randomly with replacement from the original data $z_1, \cdots, z_n$. From the pseudo data, construct the estimate $\hat{\theta}^{(b)}$. The basic result of the bootstrap is that, under fairly general circumstances, the asymptotic (large n) distribution of $(\sqrt{n}\,(\hat{\theta}^{(b)}-\hat{\theta})$ conditional on the sample is the same as the (unconditional) asymptotic distribution of $\sqrt{n}\,(\hat{\theta}-\theta)$. Thus for large n the distribution of $\hat{\theta}$ around the unknown θ is the same as the bootstrap distribution of $\hat{\theta}^{(b)}$ around $\hat{\theta}$, which is revealed by a large number (B) of draws.

We now consider the application of the bootstrap to the specific case of the fixed effects estimates. Our discussion follows Simar (1992). Let the fixed effects estimates be $\hat{\beta}$ and $\hat{\alpha}_i$, from which we calculate $\hat{u}_i^*(i=1, \cdots, N)$. Let the residuals be $\hat{v}_{it}=y_{it}-\hat{\alpha}_i-x_{it}'\hat{\beta}(i=1, \cdots, N, t=1, \cdots, T)$. The bootstrap samples will be drawn by resampling these residuals, because the v_{it} are the quantities analogous to the z's in the previous paragraph, in the sense that they are assumed to be i.i.d., and the $\hat{v}_{it}$ are the observable versions of the v_{it}. (The sample size n above corresponds to NT.) So, for bootstrap iteration $b\,(=1, \cdots, B)$ we calculate the bootstrap sample $\hat{v}_{it}^{(b)}$ and the pseudo data $y_{it}^{(b)}=\hat{\alpha}_i+x_{it}'\hat{\beta}+\hat{v}_{it}^{(b)}$. From these data we get the bootstrap estimates $\hat{\beta}^{(b)}$, $\hat{\alpha}_i^{(b)}$ and $\hat{u}_i^{*(b)}$, and the bootstrap distribution of these estimates is used to make inferences about the parameters.

We note that the estimates $\hat{u}_i$ depend on the quantity $\max_j \hat{\alpha}_j$. Since

"max" is not a smooth function, it is not immediately apparent that this quantity is asymptotically normal, and if it were not the validity of the bootstrap would be in doubt. A rigorous proof of the validity of the bootstrap for this problem is given by Hall, Härdle, and Simar (1995). They prove the equivalence of the following three statements: (i) $\max_j \hat{\alpha}_j$ is asymptotically normal. (ii) The bootstrap is valid as $T \to \infty$ with N fixed. (iii) There are no ties for $\max_j \alpha_j$, that is, there is a unique index i such that $\alpha_i = \max_j \alpha_j$. There are two important implications of this result. First, the bootstrap will not be reliable unless T is large. Second, this is especially true if there are near ties for $\max_j \alpha_j$, in other words, when there is substantial uncertainty about which firm is best.

We wish to use the bootstrap to construct a confidence interval for u_i^*. That is, for a given confidence level c, we seek lower and upper bounds L_i, U_i, such that $P[L_i \leq u_i^* \leq U_i] = 1 - c$. The simplest version of the bootstrap for the construction of confidence intervals is the percentile bootstrap. Here we simply take L_i and U_i to be the upper and lower $c/2$ fractiles of the bootstrap distribution of the $\hat{u}_i^{*(b)}$.

The percentile bootstrap intervals are accurate for large T but may be inaccurate for small to moderate T. This is a general statement, but in the present context there is a specific reason to be worried, which is the finite sample upward bias in $\max_j \hat{\alpha}_j$ as an estimate of $\max_j \alpha_j$. This will be reflected in incorrect centering of the interval and poor coverage. Simar and Wilson (1998) develop a bias corrected percentile bootstrap, as follows. As above, let $\hat{\theta}$ be the original estimate and $\hat{\theta}^{(b)}$ be the b^{th} bootstrap estimate. Define *estimated bias* $= \bar{\hat{\theta}}^{boot} - \hat{\theta}$ where $\bar{\hat{\theta}}^{boot}$ is the average of the B bootstrap estimates. Now define the bias corrected bootstrap values $\tilde{\theta}^{(b)} = \hat{\theta}^{(b)} - 2(\textit{estimated bias})$ and apply the percentile bootstrap using the bias corrected bootstrap values $\tilde{\theta}^{(b)}$. Note that estimated bias is subtracted twice, once to get the bootstrap values to center on the original estimates, and a second time to get them to center on the true θ.

Simulation evidence in Kim, Kim, and Schmidt (2007) indicates that the bias corrected percentile bootstrap is the best currently available method for constructing confidence intervals for inefficiency levels without making a distributional assumption.

VIII. Concluding Remarks and Comments on Likely Future Developments

The original stochastic frontier model of 1977 was a fully parametric model. It assumed a specific functional form for the deterministic portion of the frontier, and it assumed specific distributions for noise and for technical inefficiency. This model has been extended in a large number of directions: alternative distributional assumptions, other types of frontiers (cost functions, distance functions, ···), systems of equations, panel data, allowance for exogenous determinants of inefficiency, *etc.* No doubt such extensions and elaborations of the model will continue. However, it is probably fair to say that, as long as the model is fully parametric, the issues of how to estimate technical inefficiency and how to perform inference about it have basically been solved. Now the more interesting developments are likely to involve attempts to weaken the assumptions that need to be made.

One of the main arguments in favor of data envelopment analysis (DEA) and free disposal hull (FDH) methods in efficiency analysis is that they do not require a parametric specification of the frontier. Recent work on the stochastic frontier model similarly has aimed to not require a parametric specification of the deterministic part of the frontier (the regression function). Of course we can always estimate a regression consistently by purely nonparametric methods like kernels or nearest neighbors, but there ought to be advantages of imposing the restrictions that economic theory dictates. There has been a little work by econometricians on nonparametric methods with shape restrictions (*e.g.*, Tripathi 2000; Tripathi and Kim 2003). More recently there has been work that has more aggressively linked stochastic frontier models to DEA and FDH, notably Kuosmanen (2006, 2008). He estimates stochastic frontier models subject only to constraints like free disposability and convexity, and shows that the results have piecewise linear forms analogous to DEA. This is interesting and valuable work. We predict that in the foreseeable future the methodology will exist for routine application of the stochastic frontier model without a parametric specification of the frontier.

Avoiding distributional assumptions for noise and inefficiency is a more challenging task. The fixed effects panel data model does this successfully, but at some costs, such as the need for a large number of time series observations per firm, and the assumption that inefficiency is time invariant (or changes in a restricted way over time). Even then,

the problem of inference on the inefficiencies has not been solved very successfully. More sophisticated statistical analysis (improved bootstraps, the jackknife, *etc.*) will likely improve the situation, but the fixed effects model is probably not the long term future of the field.

If we take a random effects perspective, then there is a fundamental identification problem in that the most we can "observe" is $\varepsilon = v - u$, whereas fundamentally we are interested in u. This is the so-called "deconvolution problem" and it can never be solved without some fairly strong assumptions. As a trivial example, if v and u are both normal, they are not separately identified. Of course normal u are ruled out in the present context, but nothing prevents u from being almost normal (*e.g.*, $N^{+}(3,1)$). The assumption that v is normal does not seem to bother people, so that is a reasonable starting point, and if that assumption is made it is interesting to ask what kinds of regularity have to be assumed on u for its distribution to be identified and, more importantly, for individual values of u to be estimable and inference about them to be possible. This strikes us as the most difficult and yet most promising task for future work.

An alternative strategy is to continue to use parametric models but to find good ways to test their assumptions. Two of the authors of this paper are working on goodness of fit tests, for example (Wang, Amsler, and Schmidt 2008), something that seems long overdue.

(Received 17 November 2008; Revised 28 January 2009)

References

Ahn, S. C., Lee, Y. H., and Schmidt, P. "GMM Estimation of Linear Panel Data Models with Time-Varying Individual Effects." *Journal of Econometrics* 101 (No. 2 2001): 219-55.

_____. "Stochastic Frontier Models with Multiple Time-Varying Individual Effects." *Journal of Productivity Analysis* 27 (No. 1 2007): 1-12.

Aigner, D. J., and Chu, S. F. "On Estimating the Industry Production Function." *American Economic Review* 58 (1968): 226-39.

Aigner, D. J., Lovell, C. A. K., and Schmidt, P. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6 (No. 1 1977): 21-37.

Álvarez, A., Amsler, C., Orea, L., and Schmidt, P. "Interpreting and Testing the Scaling Property in Models Where Inefficiency

Depends on Firm Characteristics." *Journal of Productivity Analysis* 25 (No. 3 2006): 201-12.

Bai, J. "Inferential Theory for Factor Models of Large Dimensions." *Econometrica* 71 (No. 1 2003): 135-71.

Bai, J., and Ng, S. "Determining the Number of Factors in Approximate Factor Models." *Econometrica* 70 (No. 1 2002): 191-221.

Battese, G. E., and Coelli, T. J. "Prediction of Firm-Level Technical Efficiency with a Generalized Frontier Production Function and Panel Data." *Journal of Econometrics* 38 (No. 3 1988): 387-99.

_____. "Frontier Production Functions, Technical Efficiency and Panel Data with Application to Paddy Farmers in India." *Journal of Productivity Analysis* 3 (No. 2 1992): 153-69.

_____. "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data." *Empirical Economics* 20 (No. 2 1995): 325-32.

Caudill, S. B., and Ford, J. M. "Biases in Frontier Estimation Due to Heteroskedasticity." *Economics Letters* 41 (No. 1 1993): 17-20.

Caudill, S. B., Ford, J. M., and Gropper, D. M. "Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroskedasticity." *Journal of Business and Economic Statistics* 13 (No. 1 1995): 105-11.

Cornwell, C., Schmidt, P., and Sickles, R. "Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels." *Journal of Econometrics* 46 (Nos. 1-2 1990): 185-200.

Cuesta, R. A. "A Production Model with Firm-Specific Temporal Variation in Technical Inefficiency: with Application to Spanish Dairy Farms." *Journal of Productivity Analysis* 13 (No. 2 2000): 139-49.

Edwards, D. G., and Hsu, J. C. "Multiple Comparisons with the Best Treatment." *Journal of the American Statistical Association* 78 (1983): 965-71.

Farrell, M. J. "The Measurement of Productive Efficiency." *Journal of the Royal Statistal Society,* Series A. 120 (No. 3 1957): 253-90.

Greene, W. H. "Maximum Likelihood Estimation of Econometric Frontier Functions." *Journal of Econometrics* 13 (No. 1 1980): 27-56.

_____. "A Gamma-Distributed Stochastic Frontier Model." *Journal of Econometrics* 46 (Nos. 1-2 1990): 141-63.

Hall, P., Härdle, W., and Simar, L. "Iterated Bootstrap with Applications to Frontier Models." *Journal of Productivity Analysis* 6 (No. 1

1995): 63-76.

Han, C., Orea, L., and Schmidt, P. "Estimation of a Panel Data Model with Parametric Temporal Variation in Individual Effects." *Journal of Econometrics* 126 (No. 2 2005): 241-67.

Hochberg, Y., and Tamhane, A. C. *Multiple Comparison Procedures*. New York: John Wiley and Sons, 1987.

Horrace, W. C., and Schmidt, P. "Confidence Statements for Efficiency Estimates from Stochastic Frontier Models." *Journal of Productivity Analysis* 7 (No. 3 1996): 257-82.

_____. "Multiple Comparisons with the Best with Economic Applications." *Journal of Applied Econometrics* 15 (No. 1 2000): 1-26.

Hsu, J. "Simultaneous Confidence Intervals for All Distances from the Best." *Annals of Statistics* 9 (1981): 1026-34.

_____. "Constrained Simultaneous Confidence Intervals for Multiple Comparisons with the Best." *Annals of Statistics* 12 (1984): 1145-50.

_____. *Multiple Comparisons: Theory and Methods*. London: Chapman and Hall, 1996.

Huang, C. J., and Liu, J. T. "Estimation of a Non-Neutral Stochastic Frontier Production Function." *Journal of Productivity Analysis* 5 (No. 2 1994): 171-80.

Jondrow, J., Lovell, C. A. K., Materov, I. S., and Schmidt, P. "On the Estimation of Technical Efficiency in the Stochastic Frontier Production Model." *Journal of Econometrics* 19 (Nos. 2-3 1982): 233-38.

Kim, M., Kim, Y., and Schmidt, P. "On the Accuracy of Bootstrap Confidence Intervals for Efficiency Levels in Stochastic Frontier Models with Panel Data." *Journal of Productivity Analysis* 28 (No. 3 2007): 165-81.

Kim, Y., and Schmidt, P. "A Review and Empirical Comparison of Bayesian and Classical Approaches to Inference on Efficiency Levels in Stochastic Frontier Models with Panel Data." *Journal of Productivity Analysis* 14 (No. 2 2000): 91-118.

Koop, G., Osiewalski, J., and Steele, M. F. "Bayesian Efficiency Analysis through Individual Effects: Hospital Cost Frontiers." *Journal of Econometrics* 76 (Nos. 1-2 1997): 77-105.

Koop, G., Steele, M. F., and Osiewalski, J. "Posterior Analysis of Stochastic Frontier Models Using Gibbs Sampling." *Computational Statistics* 10 (1995): 353-73.

Kumbhakar, S. C. "Production Frontiers, Panel Data, and Time-Varying Technical Inefficiency." *Journal of Econometrics* 46 (Nos. 1-2 1990): 201-11.

Kumbhakar, S. C., Ghosh, S., and McGuckin, J. T. "A Generalized Production Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms." *Journal of Business and Economic Statistics* 9 (No. 3 1991): 279-86.

Kumbhakar, S. C., and Lovell, C. A. K. *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press, 2000.

Kuosmanen, T. Stochastic Nonparametric Envelopment of Data: Combining Virtues of SFA and DEA in a Unified Framework. Discussion Paper, MTT Agrifood Research Finland, 2006.

_____. "Representation Theorem for Convex Nonparametric Least Squares." *The Econometrics Journal* 11 (No. 2 2008): 308-25.

Lee, Y. H. "A Stochastic Production Frontier Model with Group-Specific Temporal Variation in Technical Efficiency." *European Journal of Operational Researc*h 174 (No. 3 2006): 1616-30.

_____. "The Group-Specific Stochastic Frontier Models with Parametric Specifications." Forthcoming in *European Journal of Operational Research*, 2009.

Lee, Y. H., and Schmidt, P. "A Production Frontier Model with Flexible Temporal Variation in Technical Inefficiency." In H. Fried, C. A. K. Lovell, and S. Schmidt (eds.), *The Measurement of Productive Efficiency: Techniques and Applications*. Oxford: Oxford University Press, 1993.

Meeusen, W., and van den Broeck, J. "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error." *International Economic Review* 18 (No. 2 1977): 435-44.

Park, B. U. and Simar, L. "Efficient Semiparametric Estimation in a Stochastic Frontier Model." *Journal of American Statistics Association* 89 (1994): 929-36.

Pitt, M., and Lee, L. F. "The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry." *Journal of Development Economics* 9 (No. 1 1981): 43-64.

Reifschneider, D., and Stevenson, R. "Systematic Departures from the Frontier: A Framework for the Analysis of Firm Inefficiency." *International Economic Review* 32 (No. 3 1991): 715-23.

Schmidt, P., and Sickles, R. C. "Production Frontiers and Panel Data." *Journal of Business and Economic Statistics* 2 (No. 4 1984): 367-74.

Simar, L. "Estimating Efficiencies from Frontier Models with Panel Data: a Comparison of Parametric, Non-Parametric and Semi-Parametric Methods with Bootstrapping." *Journal of Productivity Analysis* 3 (No. 2 1992): 171-203.

Simar, L., and Wilson, P. "Sensitivity of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models." *Management Science* 44 (No. 1 1998): 49-61.

Stevenson, R. E. "Likelihood Functions for Generalized Stochastic Frontier Estimation." *Journal of Econometrics* 13 (No. 1 1980): 57-66.

Tripathi, G. "Local Semiparametric Efficiency Bounds under Shape Restrictions." *Econometric Theory* 16 (No. 5 2000): 729-39.

Tripathi, G. and Kim, Woocheol. "Nonparametric Estimation of Homogeneous Functions." *Econometric Theory* 19 (No. 4 2003): 640-63.

Wang, H. J. "Heteroscedasticity and Non-Monotonic Efficiency Effects in a Stochastic Frontier Model." *Journal of Productivity Analysis* 18 (No. 3 2002): 241-53.

Wang, H. J. and Schmidt, P. "One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels." *Journal of Productivity Analysis* 18 (No. 2 2002): 129-44.

Wang, W. S., Amsler, C., and Schmidt, P. Goodness of Fit Tests in Stochastic Frontier Models. Unpublished Manuscript, Michigan State University, 2008.

Wang, W. S., and Schmidt, P. "On the Distribution of Estimated Technical Efficiency in Stochastic Frontier Models." *Journal of Econometrics* 148 (No. 1 2009): 36-45.