Universidad de SanAndrés

"The Econometrics Approach to the Measurement of Efficiency: A Survey"

Por Martín Rossi (Universidad de San Andrés).

D.T.: Nº 117 **Febrero 2015**

Vito Dumas 284, B1644BID, Victoria, San Fernando, Buenos Aires, Argentina
Teléfono 4725-7020, Fax 4725-7010
Email: economia@udesa.edu.ar

The Econometrics Approach to the Measurement of Efficiency: A Survey

Martín A. Rossi*
Universidad de San Andrés

Abstract

I present a survey on the econometric approach to the measurement of efficiency, focusing on the models used in empirical applications. I describe both models for cross sectional data and models for panel data. Finally, I survey the recent literature on models with time varying technical efficiency.

JEL Classification: O3.
Keywords: stochastic frontiers, productivity, technical change.

February 2015

* Universidad de San Andres, Vito Dumas 284, B1644BID, Victoria, Buenos Aires, Argentina. TE: 5411-4725-6948. Email: mrossi@udesa.edu.ar.

1. Introduction

The generally accepted meaning of the term efficiency is that it refers to the effectiveness with which the production process in the firm converts valuable inputs into valuable outputs. In practice, however, efficiency is often conceived as a physical rather than a value concept. It is useful to label such physical measures of efficiency as measures of "technical efficiency". Koopmans (1951) provided a definition of technical efficiency: a producer is technically efficient if, and only if, it is impossible to produce more of any output without producing less of some other output or using more of some input.

What we usually call inefficiency is a residual concept and the interpretation of this residual is influenced by one's view of the world. Given a strong enough faith in optimising behaviour, measured inefficiency must represent the effects of unmeasured inputs (Stigler, 1976). Advocates of optimising behaviour and no systematic differences among firms would be likely to interpret the residual, aside form statistical noise, as measurement error in either inputs or outputs and raise questions about whether all inputs and outputs have been captured.

On the other hand, some people argue that inefficiency is a pervasive fact of life (Schmidt, 1985). People make mistakes, and people manage firms. Their mistakes have consequences and those can be measured.

Therefore, the debate can be focused upon the existence or not of systematic differences in performance among firms, and upon the interpretation given to such differences. In this survey I assume that differences might exist, that differences might reflect inefficiencies, and that these inefficiencies can be measured.

2. The frontier

An efficiency measure is the distance of firm's observed practice to the frontier.[1] Of course, the task of measuring efficiency would be greatly simplified if this frontier were known. Unfortunately, usually the frontier is unknown and thus it has to be estimated. There are a number of choices to be made in order to be able to estimate a frontier, and these options will potentially give rise to different performance evaluations.

The first decision is how to construct the frontier. There are two alternatives: (i) a theoretically defined function based on engineering knowledge of the process of the industry, or (ii) an empirical function constructed from estimates based on observed data. Farrell (1957), in his path-breaking paper, argues in favour of the latter:

> "In a first place, it is very difficult to specify a theoretical efficient function [...]. Thus, the more complex the process, the less accurate is the theoretical function likely to be. Also, partly because of this, and partly because the more complex the process, the more scope it allows to human frailty, the theoretical function is likely to be wildly optimistic. If the measures are to be used as some sort of yardstick for judging the success of individuals plants, firms, or industries, this is likely to have unfortunate psychological effects; it is far better to compare performances with the best actually achieved than with some unattainable level" (Farrell, 1957, p. 255).

Following Farrell's suggestion, in this work I analyse individual performance in relation to best-observed practice. If efficiency is measured against best observed practice the result would be a measure of *relative* efficiency, where the firm is being compared with the other firms in the

[1] Other performance measures are partial productivity indices or output/input ratios. Most technologies, however, employ several inputs or produce many outputs, or both; therefore, improvement in a partial productivity index might be explained by input-output substitution, and not necessarily by an efficiency improvement in input usage or output production.

sample. Therefore, being found fully efficient does not imply that a firm cannot enhance its performance; it just means that no other firm in the sample is performing as well as it is.

The following decision is about the kind of technological relationship that is going to be estimated. Technology can be described in terms of a production frontier (output distance, input distance, production function, input requirement), or in terms of an economic frontier (cost function, profit function, revenue function). Technical efficiency is defined in terms of distance to a production frontier, and economic efficiency in terms of distance to an economic frontier. Whereas technical efficiency is a purely physical notion that can be measured without recourse to price information and without having to impose a behavioural objective on producers, economic frontiers require both price information and the imposition of an appropriate behavioural objective on producers. In this paper I focus on production frontiers because of absence of price data.

The measurement of productive efficiency is often oriented in some way, in the sense that either inputs or outputs are considered exogenous–outside the control of the firm. An input oriented efficiency measure indicates the ability of a firm to reduce the input consumption for a given level of output. Conversely, output orientation measures the ability of a firm to increase output for a given level of inputs. Which is adopted depends on the characteristics of the industry under analysis.

A remaining choice involves the estimation technique. The two competing paradigms are parametric and non-parametric techniques. Parametric methods estimate the frontier by means of econometric tools, whereas non-parametric methods use linear programming techniques.

The most widely used non-parametric approach is the so-called Data Envelopment Analysis (DEA), introduced by Charnes, Cooper, and Rhodes (1978). In this methodology, firms are considered efficient if there are no other firms, or linear combination of firms, which produce more of at least one output (given the inputs) or use less of at least one input (given the outputs).

The principal advantage of non-parametric approaches is that no functional form of the frontier is imposed *a priori* on the data. Parametric approaches, in turn, require the specification of a particular functional form for the technology, which implies that they might label inefficiency something that actually reflects a misspecification of the model. The empirical evidence, however, suggests that this problem is not very relevant when flexible functional forms—like the translog—are used in the analysis (Coelli and Perelman, 2001).

Non-parametric methods generally estimate the efficient frontier without making any assumption about the distribution of the error term, which makes hypothesis testing difficult. Parametric methods permit the testing of hypotheses such as those relating to the significance of included environmental variables (variables beyond the firms' control), technical change, and so on.

3. Parametric approach

For convenience, in what follows I use an input requirement function formulation. An input requirement function is the correct specification when firms use one dominant input to produce an exogenous output bundle. Any result for the input requirement model has a counterpart for the production function model or the input distance function model by one or more appropriate changes of sign. Density and log likelihood functions are only shown for models of cross-sectional data; panel data models are extensions of cross-sectional ones. The latter can be found in the original references or in Kumbhakar and Lovell (2000).

The simple deterministic input requirement frontier can be defined as

$$L_i = Y_i\beta + \varepsilon_i, \tag{1.1}$$

where L_i denotes the appropriate function (e.g., logarithm) of the input for the *i*-th sample firm $(i = 1,...,N)$, Y_i is a $(1\times k)$ vector of appropriate functions of the outputs associated with the *i*-th sample firm (the first element would be generally be one for all firms), and β is a $(k\times 1)$

unknown parameter vector to be estimated (which includes β_0, the intercept). The error term $\varepsilon_i = u_i \geq 0$ represents technical inefficiency of the *i*-th firm, in the sense that it measures the excess of input from its minimal possible value given by the frontier.

Under the assumption that the error term is uncorrelated with the regressors, ordinary least squares (OLS) estimates of the parameters in (1.1) are consistent (with the exception of the constant term), though not efficient.[2] The constant term can be consistently estimated by shifting the estimated frontier downwards until all residuals except one are positive. This procedure is called corrected ordinary least squares (COLS) and gives corrected residuals

$$\hat{u}_i = \hat{e}_i - \min_i \hat{e}_i ,$$

where $\hat{e}_i$ are the OLS residuals. Proof of the consistency of the COLS estimator appears in Gabrielsen (1975) and Greene (1980).

In general, technical efficiency can be calculated as the ratio of estimated (minimum feasible) to observed input. In most applications the log of the input variable appears on the left-hand side of the estimating equation. In those cases technical efficiency measures are

$$TE_i = \exp(-u_i). \tag{1.2}$$

When the frontier is deterministic every discrepancy between the individual firm performance and the frontier is considered due to inefficiency, ignoring the possibility of a firm performance being affected not only by inefficiencies in the management of its resources but also by factors beyond its control and not considered as regressors. This is often considered a serious drawback of deterministic approaches. Besides, deterministic approaches are very sensitive to the presence of outliers since a single outlier observation can have strong effects on the estimated efficiency measures.

[2] If the distribution of ε_i were known, the parameters in equation (1.1) could be estimated more efficiently by maximum likelihood.

Deterministic techniques are, in a sense, polar opposites of OLS estimates: OLS attributes all variation in input not associated to variations in outputs to random shocks, whereas deterministic approaches attribute all variation in input not associated to variations in outputs to technical inefficiency. An alternative to these polar cases would be a model that attributes variation in input not associated to variations in outputs to some combination of random shocks and technical inefficiency.

Following this idea, Aigner, Lovell, and Schmidt (1977) and Meeusen and van de Broeck (1977) proposed the so-called stochastic frontiers, which are based on the idea that deviations from the frontier could be partially out of the control of the firm. In such specifications, the input of each firm is bounded below by a frontier that is stochastic in the sense that its placement is allowed to vary randomly across firms. From an economic standpoint this technique permits firms to be technically inefficient relative to their own frontier. Inter-firm variation of the frontier presumably captures the effects of exogenous shocks beyond the control of the firm. Intra-firm variation captures randomness under the control of the firm, i.e. inefficiency (Schmidt and Lovell, 1979).

The stochastic production frontier specification is similar to the one given by equation (1.1), with the difference that now $\varepsilon_i = v_i + u_i$ is a composite error term, where v_i is an unrestricted variable representing random noise and u_i is the (non-negative) inefficiency term. All $v's$ and $u's$ are assumed to be independent of each other, and of the regressors. The u_i component cannot be directly observed; therefore it has to be inferred from the composite error term. In order to perform this decomposition and establish which part of the composite term corresponds to random noise and which part to inefficiency, it is necessary to assume some distribution for both components. The noise component is usually assumed to be independently and identically distributed $N(0, \sigma_v^2)$. For the inefficiency term, several functional forms have been proposed:

half-normal (Aigner, Lovell, and Schmidt, 1977), exponential (Meeusen and van den Broeck, 1977), truncated normal (Stevenson, 1980), and gamma (Greene, 1990).

The choice of assumption implies a trade-off between flexibility and simplicity. The (two - parameter) truncated normal and gamma distributions are more flexible than the (one-parameter) half-normal and exponential distributions. This greater flexibility, however, comes at a cost. As pointed out by Ritter and Simar (1997), there are difficulties associated with the estimation of the two parameters of these distributions when the sample size is not big enough.

In the applied literature, the vast majority of the papers use the half-normal distribution,[3] though recently there is an increasing tendency to use the more flexible truncated normal distribution.

The density function of the half-normal distribution, u_i distributed as the absolute value of a $N(0,\sigma_u^2)$ variable,[4] is

$$f(u^{HN}) = \frac{2}{\sqrt{2\pi}\sigma_u}\exp\left(-\frac{u^2}{2\sigma_u^2}\right),\ u \geq 0\,.$$

The density of the truncated normal distribution, u_i distributed as a $N(\mu,\sigma_u^2)$ variable truncated below at zero,[5] is

$$f(u^{TN}) = \frac{1}{\left[1-\Phi\left(-\frac{\mu}{\sigma_u}\right)\right]\sqrt{2\pi}\sigma_u}\exp\left(-\frac{(u-\mu)^2}{2\sigma_u^2}\right),\ u \geq 0\,,$$

where $\Phi(.)$ represents the standard normal cumulative density function.

The half-normal distribution assumes that the mode of inefficiency is equal to zero, and that the likelihood of inefficient behaviour decreases monotonically with inefficiency. In order to relax these assumptions, a truncated normal distribution can be used. Fortunately, it is not

[3] Bauer (1990a), Bravo-Ureta and Pinheiro (1993), and Coelli (1995) have also made this observation.
[4] For notational simplicity, I drop the observation subscript "*i*" in this and the following density functions.

necessary to chose *a priori* between (one of) them: given that the half-normal distribution is a particular case of the truncated normal ($\mu = 0$) it is possible to test one against the other.

The noise component, v_i, is assumed to be distributed as $N(0, \sigma_v^2)$:

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left(-\frac{v^2}{2\sigma_v^2} \right).$$

Given that v_i and u_i are assumed independent, the joint probability density function of v_i and u_i is the product of their individual densities:

$$f(u^{HN}, v) = \frac{1}{\pi\sigma_u\sigma_v} \exp\left(-\frac{u^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2} \right), u \geq 0$$

and

$$f(u^{TN}, v) = \frac{1}{\left[1 - \Phi\left(-\frac{\mu}{\sigma_u} \right) \right] 2\pi\sigma_u\sigma_v} \exp\left(-\frac{(u-\mu)^2}{2\sigma_u^2} - \frac{(v)^2}{2\sigma_v^2} \right), \; u \geq 0$$

for the half-normal and truncated normal distributions respectively.

Making the transformation $v_i = \varepsilon_i - u_i$, the joint distribution of u_i and ε_i is

$$f(u^{HN}, \varepsilon) = \frac{1}{\pi\sigma_u\sigma_v} \exp\left(-\frac{u^2}{2\sigma_u^2} - \frac{(\varepsilon - u)^2}{2\sigma_v^2} \right), \; u \geq 0$$

in the half-normal case, and

$$f(u^{TN}, \varepsilon) = \frac{1}{\left[1 - \Phi\left(-\frac{\mu}{\sigma_u} \right) \right] 2\pi\sigma_u\sigma_v} \exp\left(-\frac{(u-\mu)^2}{2\sigma_u^2} - \frac{(\varepsilon - u)^2}{2\sigma_v^2} \right), \; u \geq 0$$

in the truncated normal case.

The density functions of ε_i, obtained by integrating $f(u_i, \varepsilon_i)$ over the range of u_i, are

[5] Or $u_i \sim N^+(\mu, \sigma_u^2)$.

$$f(\varepsilon^{N/HN}) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\left[1-\Phi\left(\frac{-\varepsilon\lambda}{\sigma}\right)\right]$$

and

$$f(\varepsilon^{N/TN}) = \frac{1}{\sigma}\phi\left(\frac{\varepsilon-\mu}{\sigma}\right)\left[1-\Phi\left(-\frac{\mu}{\sigma\lambda}-\frac{\varepsilon\lambda}{\sigma}\right)\right]\left[1-\Phi\left(-\frac{\mu}{\sigma_u}\right)\right]^{-1}$$

for the normal/half-normal case and normal/truncated normal, respectively, where $\phi(.)$ represents the standard normal density, $\sigma^2 = \sigma_u^2 + \sigma_v^2$ and $\lambda = \sigma_u / \sigma_v$. If the assumptions made about v_i and u_i are correct the shape of the estimated error must be positively skewed in both cases.

Direct estimates of the stochastic frontier with cross-sectional data may be obtained by either modified ordinary least squares (MOLS) or maximum likelihood (ML) methods.

MOLS estimation requires two steps. The first step is independent of the assumption made for the inefficiency term and it consists basically in an OLS estimation of the model, which gives consistent and unbiased estimates for the slope parameters, but biased estimates for the constant term. This can be seen more clearly if the stochastic frontier is rewritten as

$$L_i = \left[\beta_0 + E(u_i)\right] + Y_i\tilde{\beta} + \left[v_i + u_i - E(u_i)\right] = \beta_0^* + Y_i\tilde{\beta} + \varepsilon_i^*, \qquad (1.3)$$

where the vector $\tilde{\beta}$ does not include the intercept. The error term ε_i^* has zero mean, and therefore OLS estimates of (1.3) gives consistent estimates of $\tilde{\beta}$.

The second step involves the estimation of the parameters β_0, σ_u^2, and σ_v^2. At this stage it is necessary to make an assumption about the distribution of u_i. MOLS can be applied by using each of the four distributions mentioned above (and eventually other non-negative distributions). Greene (1993, 1997) discusses this method for the normal/exponential and normal/gamma specifications, and Harris (1992) does so for the normal/truncated normal model. Here I apply the

method of moments approach to the normal/half-normal model analysed by Olson, Schmidt, and Waldman (1980).

Assuming that u_i follows a half-normal distribution, then the mean of ε_i is

$$E(\varepsilon_i^{N/HN}) = E(u_i) = \left(\frac{2}{\pi}\right)^{1/2} \sigma_u ,$$

and the second and third moments are

$$E\left(\varepsilon_i^2\right) = \frac{\pi - 2}{\pi}\sigma_u^2 + \sigma_v^2$$

and

$$E\left(\varepsilon_i^3\right) = -\left(\frac{2}{\pi}\right)^{1/2}\left(1 - \frac{4}{\pi}\right)\sigma_u^3 .$$

Since $E(u_i)$ is a constant, ε_i^* has the same second and third moments as ε_i. Therefore, the moments of the OLS residuals can be used to obtain consistent estimates of σ_u^2 and σ_v^2. Then, the biased OLS intercept is corrected using the estimated expected value of technical inefficiency:

$$\hat{\beta}_0 = \hat{\beta}_0^* - \hat{E}(u_i) .$$ [6]

The frontier estimated with this procedure is simply the average response function shifted downwards, implying that the technological parameters of the frontier are the same as those of the average function.

As first shown in Olson, Schmidt, and Waldman (1980), there are two potential difficulties (or indications of misspecifications) when using MOLS in a half-normal setting. The first one emerges when the skewness of the OLS residuals has the incorrect sign, which causes a negative estimate of σ_u. In such cases it is common to set $\sigma_u^2 = 0$ and consider a model with pure noise where all the firms are considered efficient. The second difficulty appears when the variance of

[6] Expressions for the standard errors of the MOLS estimators of the intercept and the variance parameters are presented in Coelli (1995).

the OLS residual is smaller than the variance of u_i,[7] thus rendering a negative estimate of σ_v. In these cases it is usual to set $\sigma_v^2 = 0$, yielding a model where all the distance to the frontier is due to inefficiency.

The second approach to the estimation of stochastic frontiers is ML. Introducing specific distributions for both components of the error term, assuming that v_i and u_i are independent, and that the regressors are exogenous, the asymptotic properties of the ML estimates can be proved in the usual way (Førsund, Lovell, and Schmidt, 1980). ML estimates are obtained by maximising the following log likelihood functions:

$$L(\beta,\lambda,\sigma) = \frac{N}{2}\ln 2\pi - N\ln\sigma + \sum_{i=1}^{N}\ln\Phi\left(\varepsilon_i\lambda\sigma^{-1}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\varepsilon_i^2$$

and

$$L(\beta,\lambda,\sigma,\mu) = \frac{N}{2}\ln 2\pi - N\ln\sigma + \sum_{i=1}^{N}\ln\left[1-\Phi\left\{\sigma^{-1}\left(-\frac{\mu}{\lambda}-\varepsilon_i\lambda\right)\right\}\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(\varepsilon_i-\mu)^2 - N\ln\left[1-\Phi\left(-\frac{\mu}{\sigma}\sqrt{\lambda^{-2}+1}\right)\right]$$

for the half-normal and truncated normal distributions. In the empirical applications I use the parameterization proposed by Battese and Corra (1977): $\gamma = \sigma_u^2/\left(\sigma_v^2+\sigma_u^2\right)$. The log-likelihoods are obtained replacing λ by $\sqrt{\frac{\gamma}{1-\gamma}}$.

Comparing the two approaches, ML estimates are asymptotically more efficient than MOLS estimates, provided the distributional assumptions used to obtain the likelihood function are correct. The impact of an error in the specification of the inefficiency distribution, however, is more important in ML estimates since the MOLS estimator introduces the assumption only in the

[7] The variance of u_i is $\frac{\pi-2}{\pi}\sigma_u^2$, not σ_u^2.

second step and hence the estimation of the slope parameters is not affected by a misspecification of this distribution. ML estimates incorporate this information from the beginning, and therefore any misspecification of the distribution of u_i would affect the estimation of the slope parameters.[8]

Once the frontier has been estimated, regardless of whether it is computed by MOLS or ML, the following step is to obtain the individual efficiency measures. The problem here is how to extract the information about the unobservable term u_i contained in the observable term ε_i. Jondrow et al. (1982) present a procedure for the half-normal model. The conditional distribution of u given ε is

$$f(u/\varepsilon) = \frac{f(u,\varepsilon)}{f(\varepsilon)} = \frac{1}{\sqrt{2\pi}\sigma_*} \exp\left[-\frac{(u-\mu_*)^2}{2\sigma_*^2} \right] \Bigg/ \left[1 - \Phi\left(\frac{-\mu_*}{\sigma_*} \right) \right].$$

where $\sigma_*^2 = \sigma_v^2 \frac{\sigma_u^2}{\sigma^2}$ and $\mu_* = \varepsilon \frac{\sigma_u^2}{\sigma^2}$. Since $f(u/\varepsilon)$ is distributed as $N^+(\mu_*, \sigma_*^2)$, either the mean or the mode of this distribution can serve as a point estimator for u_i. The mean and the mode are given by

$$E[u_i/\varepsilon_i] = \mu_{*i} + \sigma_* \left[\frac{\phi(-\mu_{*i}/\sigma_*)}{1-\Phi(-\mu_{*i}/\sigma_*)} \right] = \sigma_* \left[\frac{\phi(\varepsilon_i \lambda/\sigma)}{1-\Phi(-\varepsilon_i \lambda/\sigma)} + \frac{\varepsilon_i \lambda}{\sigma} \right]$$

and

$$M[u_i/\varepsilon_i] = \begin{cases} \varepsilon_i \left(\sigma_u^2 / \sigma^2 \right) & \text{if } \varepsilon_i \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

When output is in logs, individual technical efficiency scores can be obtained by substituting either $E(u_i/\varepsilon_i)$ or $M(u_i/\varepsilon_i)$ into equation (1.2).

The intuition behind the above results is as follows. On one extreme case, when σ_u^2 is equal to zero there is no contribution from inefficiency and hence all firms are said to be fully efficient.

[8] In fact, MOLS estimates of inefficiency would be robust to misspecification of the distribution of v_i.

Exactly the other extreme case arises when σ_v^2 is equal to zero. In this case there is no contribution from the noise term and, therefore, all differences between the frontier and the observed behaviour is considered inefficiency. Finally, there are in between cases in which some part of the observed residual is accounted as noise and the remainder as inefficiency.

Battese and Coelli (1988) proposed an alternative point estimator for technical efficiency:

$$TE_i = E\left(\exp\left[-u_i / \varepsilon_i\right]\right) = \left[\frac{1-\Phi\left(\sigma_* - \frac{\mu_{*_i}}{\sigma_*}\right)}{1-\Phi\left(-\frac{\mu_{*_i}}{\sigma_*}\right)}\right] \exp\left(-\mu_{*_i} + \frac{1}{2}\sigma_*^2\right).$$

The Battesi and Coelli estimator is preferred to the Jondrow et al. estimator when u_i is not close to zero (Murillo-Zamorano 2004). In any case, whichever point estimator is used, all of them share an important defect: they are not consistent estimates of technical efficiency. It is possible, however, to get confidence intervals for any of the alternative technical efficiency point estimators commented above. Thus, Hjalmarsson, Kumbhakar, and Heshmati (1996) propose confidence intervals for the Jondrow et al. (1982) efficiency estimator, and Bera and Sharma (1996) for the Battese and Coelli (1988) one.

Models for panel data

In general, stochastic frontier models with cross-sectional data are subject to three serious drawbacks (Schmidt and Sickles, 1984). Firstly, the inefficiency term estimations are not consistent—in fact, there is only one observation available for each firm. Secondly, both model estimation and separation between inefficiency and noise call for specific assumptions to be made about the distribution of either term. Finally, it might be incorrect to assume that inefficiency is independent of the regressors.

The preceding problems are potentially solvable using panel data. The first drawback can be handled if the number of observations on each firm is large enough. As pointed out by

Kumbhakar and Lovell (2000), however, this benefit can be overstated since in practice many panels are relatively short. Second, having access to panel data allows the researcher to avoid any assumption about the distribution of the inefficiency (though at the cost of assuming that inefficiency is time invariant). Finally, not all panel data estimation techniques require the assumption of independence of the technical inefficiency term from the regressors.

The simplest panel data model, which assumes that inefficiency is time invariant, is

$$L_{it} = \alpha + Y_{it}\beta + \nu_{it} + u_i, \tag{1.4}$$

where the sub-index i represents firms ($i = 1, ..., N$), t the period of time ($t = 1, ..., T$), L_{it} denotes the appropriate function of the input, Y_{it} is a vector of appropriate functions of the outputs, α and β are parameters of the model, ν_{it} is an independent and identically distributed random error with zero mean and constant variance $\left(\sigma_v^2\right)$ representing random events outside the firms' control, and u_i is a non-negative random disturbance with constant variance $\left(\sigma_u^2\right)$ representing technical inefficiency. The random variables ν_{it} and u_i are assumed to be independent from each other.

The above model can be re-expressed as

$$L_{it} = \alpha_i + Y_{it}\beta + \nu_{it}, \tag{1.4'}$$

where $\alpha_i = \alpha + u_i$.

The model in equation (1.4') can be estimated using two different standard techniques. The first approach is called "fixed-effects" and considers the individual effect α_i as a specific parameter of the firm. The second approach is called "random-effects" and considers the individual effect as a component of the random term of the model.

The model (1.4') can be estimated using OLS by treating the α_i's as parameters of dummy variables which identify each individual—usually called Least Square Dummy Variable (LSDV) estimator. That is,

$$L_{it} = \sum_{i}^{N} D_i \alpha_i + Y_{it}\beta + \nu_{it},$$

where D_i is a dummy variable that takes the value of one if the observation corresponds to the *i*-th individual and zero otherwise. Consistency of LSDV requires either (i) $corr(Y_{it}, v_{it}) = 0$ and $T \to \infty$; or (ii) $corr(Y_{it}, v_{is}) = 0 \ \forall s,t.$ Condition (ii) (strict exogeneity) rules out any feedback from past productivity shocks to current levels of outputs.

In this setting, the estimation of the relative inefficiency measures uses the fact that the most efficient firm in the sample must have a value of u_i equal to zero, which implies $\min_i(\alpha_i) = \alpha$. In other words, the smallest estimated firm intercept is used to define the intercept parameter so that all firm effects are estimated to be zero or positive. Then, u_i can be estimated as

$$\hat{u}_i = \hat{\alpha}_i - \min_i(\hat{\alpha}_i). \tag{1.5}$$

The similarity of the fixed-effects model to the COLS model based on cross-sectional data should be apparent. If the input is expressed in logs, the individual efficiency measures are obtained by substituting the expression in equation (1.5) into equation (1.2).

The properties of the estimators can be better understood by realising that the OLS estimator of β is equivalent to the OLS estimates of the model in deviations from individual-specific mean values (also called the "within model"):

$$(L_{it} - \bar{L}_i) = (Y_{it} - \bar{Y}_i)\beta + \nu_{it} - \bar{v}_i, \tag{1.6}$$

where $\bar{L}_i$, $\bar{Y}_i$, and $\bar{v}_i$ are individual means $\left(\text{i.e., } \bar{L}_i = \frac{1}{T}\sum_{s=1}^{T} L_{is}\right)$. In this case, the individual effects can be estimated as

$$\hat{\alpha}_i = \bar{L}_i - \bar{Y}_i\hat{\beta}.$$

The estimator of α_i is only consistent when T tends to infinity. That is, the consistent estimation of the firms' efficiency measures requires a large number of time periods—i.e. a large number of observations on each of the firms.

The fixed-effects model precludes the inclusion of time invariant regressors. In the presence of time invariant attributes that affect firms' performance, the random-effects model can be used instead if one is willing to assume independence between the individual effects and the regressors.

The estimation of equation (1.4) with OLS gives consistent estimates of the parameters of the model when the individual effects are not correlated with the explanatory variables. The covariance matrix, however, is not scalar; therefore, it is possible to get more efficient estimates using generalised least squares (GLS). In particular, for each individual

$$Var(v_{it} + u_i) = \sigma_v^2 + \sigma_u^2$$

and

$$Cov(v_{it} + u_i, v_{it+1} + u_i) = \sigma_u^2,$$

which are the result of the assumptions of independence between u_i and v_{it}, the absence of serial correlation of v_{it}, and the time invariance of u_i. These assumptions imply the following ($T \times T$) covariance matrix for each individual:

$$\Omega_i = \begin{bmatrix} \sigma_v^2 + \sigma_u^2 & \sigma_u^2 & \dots & \sigma_u^2 \\ \sigma_u^2 & \sigma_v^2 + \sigma_u^2 & \dots & \sigma_u^2 \\ & & \dots & \\ \sigma_u^2 & \sigma_u^2 & \dots & \sigma_v^2 + \sigma_u^2 \end{bmatrix}.$$

The individual effects are assumed uncorrelated across the individuals, that is, $Cov(v_{it} - u_i, v_{jt+k} - u_j) = 0, \forall k$. Therefore, the ($NT \times NT$) covariance matrix for all the individuals can be represented as

$$\Omega = \begin{bmatrix} \Omega_1 & 0 & \dots & 0 \\ 0 & \Omega_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \Omega_N \end{bmatrix}.$$

The GLS estimator can be written as

$$\hat{\beta}^{GLS} = (Y'\Omega^{-1}Y)^{-1}Y'\Omega^{-1}L,$$

where Y is the $(NT \times K)$ matrix of regressors (K is the number of right-hand side variables) and L is the $(NT \times 1)$ dependent variable.

In empirical applications, however, σ_v^2 and σ_u^2 are unknown and GLS estimation is infeasible. A feasible estimator can be obtained by estimating σ_v^2 as the variance of the within estimation (1.6), and realising that $\sigma_v^2 + \frac{\sigma_u^2}{T}$ is just the variance of the "between model"

$$\bar{L}_i = \alpha + \bar{Y}_i\beta + \bar{v}_i + u_i.$$

An estimate of σ_u^2 can be recovered from both estimates. The inefficiency term can be estimated as

$$\hat{u}_i = \frac{\sum_t^T \hat{\varepsilon}_{it}}{T},$$

where

$$\hat{\varepsilon}_{it} = L_{it} - \hat{\alpha}^{GLS} - Y_{it}\hat{\beta}^{GLS}.$$

If the input is in logs, the individual efficiency measures can be calculated as in (1.2).

In the unrealistic situation where σ_v^2 and σ_u^2 are known, the estimates of the parameters of the model are consistent whenever N or T tends to infinity.[9] When σ_v^2 and σ_u^2 are unknown, the

[9] If T is fixed, this again requires "strict exogeneity" or $corr(Y_{it}, v_{is}) = 0 \ \forall s,t.$

consistent estimation of σ_u^2 requires $N \to \infty$. On the other hand, a consistent estimate of the individual efficiency measures again requires $T \to \infty$.

The consistency of the GLS estimates depends on the individual effects being uncorrelated with all the explanatory variables. This assumption can be tested using a Hausman test, which in this case is based on the significance of the differences between feasible GLS and within estimates. Under the null hypothesis of no correlation between the individual effects and the regressors both estimators within ($\hat{\beta}^W$) and GLS ($\hat{\beta}^{GLS}$) are consistent. Therefore, under the null hypothesis of no correlation these estimators would not differ significantly:

$$H_0 : \beta^W - \beta^{GLS} = 0 .$$

Under the null hypothesis the within estimator is inefficient, and under the alternative hypothesis of correlation between the individual effects and the explanatory variables the GLS estimator is inconsistent, which implies that in that case one should expect significant differences between both estimators ($H_1 : \beta^W - \beta^{GLS} \neq 0$). The test of this hypothesis can be done using a Wald test:

$$W = (\hat{\beta}^W - \hat{\beta}^{GLS})' \Sigma^{-1} (\hat{\beta}^W - \hat{\beta}^{GLS}),$$

where

$$\Sigma = Var(\hat{\beta}^W - \hat{\beta}^{GLS}) = Var(\hat{\beta}^W) - Var(\hat{\beta}^{GLS}) .$$

This simple expression for the variance of the difference is one of the main results found in Hausman (1978): the variance of the difference between an inefficient and an efficient (under the null hypothesis) estimator is the difference between the variances. The Wald statistic has a chi-squared distribution with degrees of freedom equal to the number of parameters that are being compared.

Summing up, the fixed-effects model does not require the assumption that the inefficiency term and the regressors are orthogonal to each other, but at the cost of not allowing the inclusion of constant regressors. In the presence of time invariant attributes of the firms that are omitted from the model, these would be captured in the fixed effects, mixing with the inefficiency term when they should be classified otherwise. The random-effects model, in turn, allows the inclusion of time invariant regressors in the model, although at the cost of having to assume that the inefficiency term is independent from all the regressors.[10] This can be a strong assumption since if a firm knows its efficiency level this could affect its output choices.

Both fixed-effects and random-effects models do not make any assumption about the distribution of the inefficiency term. If the researcher is willing to assume some distribution of the inefficiency term, and to assume independence between the efficiency effects and the regressors, a stochastic maximum likelihood estimate is feasible. This approach is widely used in empirical analysis and therefore is described in detail in what follows.

The stochastic frontier input requirement function model with panel data is written as[11]

$$L_{it} = Y_{it}\beta + \varepsilon_{it},$$

where L_{it}, Y_{it}, and β are defined as before. The error term is specified as $\varepsilon_{it} = v_{it} + u_i$. The v_{it} are assumed to be independently and identically distributed $N(0, \sigma_v^2)$, independent of the u_i. The u_i are assumed to be independently and identically distributed defined as the truncation (at zero) of the $N(\mu, \sigma_u^2)$ distribution. In addition, it is assumed that v_{it} and u_i are independently distributed of the regressors.

[10] There is a third alternative, not so much used in this literature, where only a subset of the regressors are correlated with the inefficiency disturbance and then a Hausman-Taylor (1981) estimator can be employed.
[11] This is an adaptation of Battese and Coelli (1988) production function model.

The parameters of the model can be consistently and efficiently estimated using ML, provided the assumptions regarding the distribution of inefficiency and the independence between technical inefficiency and the explanatory variables are correct.

In order to obtain individual efficiency measures, u_i has to be inferred from the observable, ε_{it}. Battese and Coelli (1988) provide the formula for this decomposition and show that the condition for consistency is again $T \to \infty$. The assumption that the firm effects are time-invariant, however, is unlikely to be realistic for a sufficiently long period of time; firms may discover, after a period of time, the extent of their inefficiency and adjust their practices to become more efficient (assuming they have the incentives to do so).

In this context, the Hausman test can be used to contrast the distributional assumptions of the error components. Under the null hypothesis of correct distributional assumptions both estimators GLS and ML are consistent, though GLS is inefficient.[12] Under the alternative, the Battese and Coelli (1988) stochastic frontier model is inconsistent (GLS is still consistent) and therefore there should be significant differences between the two estimators.

Models with time varying technical efficiency

Until now I have assumed that inefficiency is constant over time. This assumption, however, may not be as innocuous as it appears. In economics terms, the assumption implicitly implies that the agents do not realise the degree of their inefficiency or, if they do realise, they do not try to reduce it or they are not able to do it. In statistical terms, imposing this restriction without formally testing its appropriateness may result in inconsistency of estimators for the parameters of the model as well as for technical inefficiency (Kumbhakar, 1990).

[12] The inefficiency of GLS is due to the non-normality of the composite error term.

The assumption of time-invariant inefficiency can be relaxed by imposing some structure on how the inefficiency evolves over time.[13] One possibility is the Cornwell, Schmidt, and Sickles (1990) specification, which allows the individual effects to evolve over time as a quadratic function

$$u_{it} = \zeta_{i1} + \zeta_{i2} t + \zeta_{i3} t^2 .$$

In this model the inefficiency term is a quadratic function of time, and the form may not be the same across firms. Cornwell, Schmidt, and Sickles propose three different estimators for this model, all of them being some variant of least squared: within, GLS, or Hausman-Taylor. Which one is actually selected depends on the number of assumptions the researcher is willing to make about the independence of the firm effects.

The Cornwell, Schmidt, and Sickles specification is very flexible, but has two important disadvantages: it is very demanding in terms of data (the additional parameters to be estimated are 3N) and it does not allow the separation of technical inefficiency changes (catching-up) from technical change (shifts in the frontier).

Trying to solve these problems, Kumbhakar (1990) proposes a stochastic frontier model that allows inefficiency to change over time. Unlike Cornwell, Smith, and Sickles (1990), in Kumbhakar's model the temporal variation of technical inefficiency is not modelled through the intercept of the frontier, but through an error component. The model specifies the inefficiency of firm i at time t as the product of a one-sided random variable, u_i, and a given exponential function, $\zeta(t)$:

$$u_{it} = \zeta(t) u_i$$
$$\zeta(t) = \left[1 + \exp(bt + ct^2) \right]^{-1}$$

[13] An alternative is to model inefficiency as being statistically independent over time, treating the panel as cross-sectional data.

where *b* and *c* are unknown parameters to be estimated, and u_i are assumed to be independent and identically distributed as truncations at zero of the $N(\mu,\sigma^2)$ distribution. The function $\zeta(t)$ can be monotonically increasing or decreasing, concave, or convex, depending on the sign and magnitude of *b* and *c*.

Battese and Coelli (1992) extended Kumbhakar's model in order to allow for unbalanced panel data.[14] They specify the inefficiency as an exponential function

$$u_{it} = u_i \exp\left[-\eta(t-T)\right],$$

where u_i is independent and identically distributed $\left|N\left(0,\sigma_u^2\right)\right|$, and η is the only additional parameter to be estimated. In this specification, since the exponential function, $\exp\left[-\eta(t-T)\right]$, has a value of one when *t*=*T*, the random variable u_i can be considered as the technical inefficiency effect for the *i*-th firm in the last period of the panel (but can be estimated even if firm i is not observed in period T). For earlier periods, the technical efficiency effects are the products of the technical inefficiency effect for the *i*-th firm in the last period of observation, and the value of the exponential function, which depends on the parameter η and the number of periods before the last observation. If η is positive then the model shows decreasing inefficiency effects, while if η is negative the inefficiency effects are increasing.

The advantages of the Kumbhakar (1990) and Battese and Coelli (1992) specifications are that they are less data demanding than the Cornwell, Schmidt, and Sickles (1990) model, and that technical inefficiency changes over time can be distinguished from technical change.[15] A disadvantage is that the ordering of the firms according to the magnitude of the technical inefficiency effects is the same at all time periods.

[14] An unbalanced panel is one in which the group size differs across groups.
[15] Technical change can be incorporated by means of a time trend in the frontier, or by the inclusion of time dummy variables.

Trying to avoid this disadvantage Cuesta (2000) extended the Battese and Coelli (1992) representation by modelling the inefficiency term as

$$u_{it} = u_i \exp\left[-\eta_i (t - T)\right],$$

where η_i are firm specific parameters to be estimated. The Battese and Coelli (1992) model is nested in this model and therefore it is possible to impose the restrictions in order to test the hypothesis of a common pattern of inefficiency change across firms, $H_0 : \eta_1 = \eta_2 = ... = \eta_i = \eta$. The ML estimator of Cuesta's model, however, can be inconsistent since the number of parameters increases with the sample—the "incidental parameter" problem.

Models incorporating exogenous influences on efficiency

The analysis of technical efficiency has two components. The first is the estimation of a frontier that serves as a benchmark against which to estimate producers' efficiency. The second component is equally important, although much less frequently explored. It concerns the incorporation of exogenous variables that can exert an influence on firms' performance.

The main difference between environmental variables and variables influencing efficiency is that the former are assumed to influence the structure of the technology by which inputs are converted into output, whereas the latter influence the efficiency with which inputs are converted to outputs. Unfortunately, there is no universal rule to distinguish both types of variables; thus the final decision depends on the researcher's view of the world.

Early empirical papers in which the issue of the explanation of inefficiency effects was raised include Pitt and Lee (1981) and Kalirajan (1981). These papers adopt a two-stage approach, first estimating the frontier and the firms' efficiency levels, and afterwards regressing these efficiency levels on a vector of variables, say Z. The two-step procedure, however, gives biased results because the model estimated at the first step suffers from an omitted variable problem. More importantly, Wang and Schmidt (2002) show that the bias of the two-step

procedure is substantial and affects the estimates of both technological parameters and efficiency measures.

This issue was first addressed by Kumbhakar, Gosh, and McGukin (1991) and Reifschneider and Stevenson (1991), who propose one-step stochastic frontier models in which the relationship between Z and technical efficiency is imposed in estimating the technology and the firms' efficiency level.

In this paper I use the Battese and Coelli (1995) specification of this one-step procedure, which extended the Kumbhakar, Gosh, and McGukin (1991) model in order to allow for panel data. In the Battese and Coelli model, the noise term is assumed to be independent and identically distributed $N\left(0,\sigma_v^2\right)$. The u_{it} are non-negative random variables assumed to be independently distributed such that u_{it} is the truncation (at zero) of the normal distribution with mean μ_{it} and variance σ_u^2. The μ_{it} is defined by

$$\mu_{it} = Z_{it}\xi,$$

where Z_{it} is a $(1\times m)$ vector of explanatory variables associated with technical inefficiency and ξ is a $(m\times 1)$ vector of unknown coefficients.[16]

Technical efficiency of production for the *i*-th firm at the *t*-th time period is $TE_{it} = \exp(-u_{it})$, and the prediction of the technical efficiencies is based on its conditional expectation, given the model assumptions.

It is important to notice that estimates of the coefficients on the *Z* variables in the stochastic frontier model are not directly comparable to the coefficients obtained from an OLS regression with the same *Z* variables on the right hand side. The coefficient on Z_k in the inefficiency effects model of the stochastic frontier measures the amount that the mean of the normal distribution

[16] Related to the previous section, one component of Z_{it} could be time (t).

prior to truncation shifts with Z_k. This amount is not necessarily equal to the change in expected input use when changing Z_k, which is more comparable to OLS estimates.

As shown in Huang and Liu (1994) and Battese and Broca (1997),

$$\frac{\partial E\left(ln\ L_{it}\right)}{\partial Z_{k,it}} = C_{it}\left(\frac{\partial \mu_{it}}{\partial Z_{k,it}}\right),$$

where

$$C_{it} = 1 - \frac{1}{\sigma}\left\{\frac{\phi\left(\frac{\mu_{it}}{\sigma} - \sigma\right)}{\Phi\left(\frac{\mu_{it}}{\sigma} - \sigma\right)} - \frac{\phi\left(\frac{\mu_{it}}{\sigma}\right)}{\Phi\left(\frac{\mu_{it}}{\sigma}\right)}\right\}$$

and ϕ and Φ represent the density and distribution functions of the standard normal random variable.

References

Aigner, D., C.A. Knox Lovell, and P. Schmidt (1977). "Formulation and Estimation of Stochastic Frontier Production Function Models." Journal of Econometrics 6, 21--27.

Battese, G. and T. Coelli (1988). "Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data." Journal of Econometrics 38, 387-399.

Battese, G. and T. Coelli (1992). "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India." Journal of Productivity Analysis 3, 153-169.

Battese, G. and T. Coelli (1995). "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data." Empirical Economics 20, 325–332.

Battese, G. and G. Corra (1977). "Estimation of a Production Frontier Model: With Application to the Pastoral Zone of Eastern Australia." Australian Journal of Agricultural Economics 21, 39–56.

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

Bravo-Ureta, B. and A. Pinheiro (1993). "Efficiency Analysis of Developing Country Agriculture: A Review of the Frontier Function Literature." Agricultural and Resource Economics Review 22, 88-101.

Charnes, A., W. Cooper, and E. Rhodes (1978). "Measuring the Efficiency of Decision Making Units." European Journal of Operational Research, 2 (6), 429-444.

Coelli, T. (1995). "Estimators and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis." Journal of Productivity Analysis 6, 247-268.

Coelli, T. and S. Perelman (2001). "Medición de la Eficiencia Técnica en Contextos Multiproducto." In A. Alvarez, *La Medición de la Eficiencia y la Productividad*. Ediciones Pirámide.

Coelli, T., D. Prasada Rao and G. Battese (1998). *An Introduction to Efficiency and Productivity Analysis*. Kluwer Academic Publishers.

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

Cuesta, R. (2000). "A Production Model with Firm-Specific Temporal Variation in Technical Inefficiency: with Application to Spanish Dairy Farms." Journal of Productivity Analysis 13, 139-158.

Farrell, M. (1957). "The Measurement of Productive Efficiency." Journal of the Royal Statistical Society, Series A, Part III, Vol. 120, 253:281.

Førsund, F., C.A. Knox Lovell and P. Schmidt (1980). "A Survey of Frontier Production Functions and of their Relationship to Efficiency Measurement." Journal of Econometrics 13, 5-25.

Gabrielsen, A. (1975). "On Estimating Efficient Production Functions." Working Paper A-85, Department of Humanities and Social Sciences, Bergen, Norway.

Greene, W. (1993). "The Econometric Approach to Efficiency Analysis." In H. Fried, C. A. Knox Lovell and P. Schmidt (eds.), *The Measurement of Productive Efficiency*. New York: Oxford University Press, 68-119.

Greene, W. (1997). "Frontier Production Functions." In M. Pesaran and P. Schmidt (eds.), *Handbook of Applied Econometrics*, Volume II: Microeconomics. Oxford: Blackwell Publishers.

Grenne, W. (1980). "Maximum Likelihood Estimation of Econometric Frontier Functions." Journal of Econometrics 13, 27-56.

Grenne, W. (1990). “A Gamma-Distributed Stochastic Frontier Model.” Journal of Econometrics 46, 141--163.

Harris, C. (1992). “Technical Efficiency in Australia: Phase I.” In R. Caves (ed.), *Industrial Efficiency in Six Nations*. MIT Press, Cambridge.

Hausman, J. (1978). “Specification Tests in Econometrics.” Econometrica 46, 1251-1272.

Hausman, J. and W. Taylor (1981). “Panel Data and Unobservable Individual Effects.” Econometrica 49, 1377-1398.

Jondrow, J., C.A. Knox Lovell, I. Materov, and P. Schmidt (1982). “On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model.” Journal of Econometrics 19, 233-238.

Kalijaran, K. (1981). “An Econometric Analysis of Yield Variability in Paddy Production.” Canadian Journal of Agricultural Economics 29, 283-294.

Koopmans, T. (1951). “An Analysis of Production as an Efficient Combination of Activities”. In T. Koopmans (ed.), *Activity Analysis of Production and Allocation*. Cowles Commission for Research in Economics, Monograph No.13, New York, Wiley.

Kumbhakar, S. (1990). “Production Frontiers, Panel Data, and Time-Varying Technical Inefficiency.” Journal of Econometrics 46, 201-211.

Kumbhakar, S., S. Gosh and J. McGukin (1991). “A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in US Dairy Farms.” Journal of Business and Economic Statistics 9, 279-286.

Kumbhakar, S. and C.A. Knox Lovell (2000). *Stochastic Frontier Analysis*. Cambridge University Press.

Meeusen, W. and J. van den Broeck (1977). “Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error.” International Economic Review 18, 435--444.

Olson, J., P. Schmidt and D. Waldman (1980). “A Monte Carlo Study of Estimators of the Stochastic Frontier Production Function.” Journal of Econometrics 13, 67-82.

Pitt, M. and L. Lee (1981). “Measurement and Sources of Technical Inefficiency in the Indonesian Weaving industry. “ Journal of Development Economics 9, 43-64.

Reifschneider, D. and R. Stevenson (1991). “Systematic Departures from the frontier: A Framework from the Analysis of Firm Inefficiency.” International Economic Review 32, 715-723.

Ritter, C., and L. Simar (1997). “Pitfalls of Normal-Gamma Stochastic Frontier Models.” Journal of Productivity Analysis 8, 167--182.

Sampedro, C. (2001). “La Estimación de Eficiencia en Modelos con Datos de Panel.” In A. Alvarez, *La Medición de la Eficiencia y la Productividad*. Ediciones Pirámide.

Schmidt, P. (1985). “Frontier Production Functions”. Econometric Reviews 4, 289--328.

Schmidt, P. and C.A. Knox Lovell (1979). “Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontiers.” Journal of Econometrics 9, 343-366.

Schmidt, P. and R. Sickles (1984). “Production Frontiers and Panel Data.” Journal of Business & Economic Statistics 2, 367-374.

Stevenson, R. (1980). “Likelihood Functions for Generalized Stochastic Frontier Estimations.” Journal of Econometrics 13, 57--66.

Stigler, G. (1976). “The Existence of X-Efficiency.” The American Economic Review 66, 213-216.

Wang, H. and P. Schmidt. (2002). “One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels.” Journal of Productivity Analysis 18 (2), 129-144.