The Frontier Approach to the Measurement of Productivity and Technical Efficiency

Vania Sena[1]

LUBS

University of Leeds

July 9, 2003

Abstract

In 1957, Farrell proposed to measure technical (in)efficiency as the realised deviation from a frontier isoquant. Since then, the research has developed several methods to derive the production frontier and it has also extended its scope in applying frontier techniques to the measurement of total factor productivity. In this paper, I present the core techniques for the measurement of technical efficiency and productivity based on the notion of frontier and introduce the more recent methodological advances in the field.

1 Introduction

Recent discussions on the sources of productivity and efficiency differentials have pushed forward the debate on efficiency and total factor productivity (TFP) measurement on the research agenda. Indeed, once it is agreed that high levels of efficiency and productivity are desirable goals for an economy, then it is important to define and measure them in ways that respect economic theory and, at the same time, provide useful information to both managers and policy-makers. Among the different methods, the ones based on the notion of *best practice frontier* have become very popular. The reasons for their success can be ascribed to two factors: first, they have deep roots in the economic theory. Indeed, in this approach, efficiency is measured as the distance from a best practice frontier (or the boundary of the production possibility set), computed in accordance with the axioms of the production theory; equally, productivity change is measured as the variation over time of the firm's distance from the frontier and is decomposed into changes in technical efficiency, technical change and change in scale, where the first is measured by how far the firm is from the frontier, the second by shifts in the frontier itself and the last by the movements of the firm along the production function curvature. Second, the concept of a distance from a standard allows us to operationalise the concept of inefficiency and TFP, providing ready-to use information for decision-makers.

The idea of measuring a firm's performance with respect to a best practice frontier goes back to the 1950s. Koopmans (1951) defined technical efficiency as the capability of a firm to maximize output for given inputs. However this notion does not offer any guidance concerning the degree of in-

efficiency. This issue was addressed by Farrell (1957). He extended the work initiated by Koopmans and suggested measuring inefficiency as the observed deviation from a frontier isoquant. However, the production possibility set that economic theory associates with any productive activity is unknown. Therefore, the subsequent research has focused on the best way to identify the frontier of the production possibilities set. Two methodologies are now available: a) parametric methods, based on the econometric estimation of the frontier and b) non-parametric methods (called this way as they do not require a functional form for the technology) based on linear programming techniques such as Data Envelopment Analysis (DEA, henceforth). Later, during the 1980s, frontier techniques have been usefully extended to measure productivity growth. Not surprisingly, it is possible to distinguish between a parametric approach to the measurement of productivity change, and a non-parametric one, linked mainly to the development of the Malmquist index (Fare *et al.*, 1992).

Where is the current research in this field going? Three main research areas seem to be of particular interest for the applied economist. The first deals with the use of semiparametric econometrics to estimate stochastic frontiers. Indeed, parametric methods have often been criticised on the ground that they require both a functional form for the production technology and a distributional assumption for the inefficiency component. An attempt to accommodate this criticism has been made by employing semiparametric econometrics to relax at least one of the two assumptions. This has produced an array of semi-parametric estimators of potential interest to the applied economist. The second research area is concerned with the statistical proper-

ties of the DEA estimators. Non-parametric methods typically do not allow us to carry out any statistical inference on the computed efficiency scores and therefore recent research has been devoted to the identification of the statistical properties of the estimators, to the design of possible hypotheses tests and to use the bootstrap for constructing confidence intervals. Finally, the last research area deals with the treatment of undesirable (from a social standpoint) outputs when measuring productivity. Farrell based measures of productivity usually consider as an improvement in productivity the simultaneous expansion in outputs (whether bad or good), for given amount of inputs. In this type of framework, reductions in one output (the bad output) would be treated as a decrease of productivity. Therefore, new measures of productivity have been derived so that a firm can be credited for reducing undesirable output and penalised vice-versa.

The purpose of this paper is to provide an introduction to different methods for the measurement of technical efficiency and TFP based on the notion of the*best practice frontier*, to assess their strenghts and shortcomings when used in empirical work and to present some methodological advances in this field. However, it is not meant to be an exhaustive survey on the topic.[2]. Papers whose techniques are of potential interest to the applied economist are included.

After introducing Farrell's measure of technical efficiency (Section 2), I will present the parametric methods to the frontier estimation (Section 3): after briefly presenting the cross-sectional models (Section 3.1), I will focus on panel data models (Section 3.2) and then illustrate how semiparametric econometrics can be usefully applied to frontier models (Section 3.3). Af-

terwards, the DEA models are presented (Section 4), along with the recent results on inference (Section 4.1) and bootstrapping (Section 4.2). The frontier approach to the measurement of productivity is then presented (Section 5), followed by an illustration of the parametric approach (Section 5.1) first and then of the non-parametric one (Malmquist index) (Section 5.2). Next I show how the Malmquist index can be extended to take into account the production of undesirable outputs by using the directional distance functions (Section 6). Finally Section 7 offers some concluding remarks.

2 Farrell's measure of technical efficiency

Farrell (1957) proposed to measure technical inefficiency as one minus the equiproportionate reduction in all inputs for given outputs. A score of unity indicates technical efficiency because no equiproportionate input reduction is feasible, and a score less than unity measures the severity of technical inefficiency. This measure is an input-oriented one; it is, however, straightforward to convert it to an output-oriented measure. This is defined as one minus the equiproportionate expansion in output for given inputs, with a score of one indicating technical efficiency and vice-versa.

Let me define these measures formally, starting from the input-oriented one. Let producers use inputs $x = (x_1, ..., x_n) \in R^n_+$ to produce outputs $y = (y_1, ..., y_m) \in R^m_+$. Production technology can be represented by an input set:

$$L(y) = \{x : (y, x) \qquad \text{is} \qquad \text{feasible}\} \tag{1}$$

which for every $y \in R_+^m$ has an isoquant:

$$IsoqL(y) = \{x : x \in L(y), \lambda x \notin L(y), \forall \lambda \in [0,1)\} \tag{2}$$

and an efficient subset:

$$EffL(y) = \{x : x \in L(y), x' \notin L(y), \forall x' \leq x\} \tag{3}$$

Shephard (1953) introduced the input distance function to provide a functional representation of a multiple output technology. The input distance function can be defined as (with I indicating that it is an input-oriented measure):

$$D_I(y,x) = \{max\lambda : (x/\lambda) \in L(y)\} \tag{4}$$

where $D_I(y,x) \geq 1$ and it follows from (2) that:

$$IsoqL(y) = \{x : D_I(y,x) = 1\} \tag{5}$$

The Farrell input-oriented measure of technical efficiency (DF_I) can now be given a formal interpretation as:

$$DF_I(y,x) = min\{\lambda : \lambda x \in L(y)\} \leq 1 \tag{6}$$

It follows from (6) that:

$$DF_I(x,y) = \frac{1}{D_I(y,x)} \tag{7}$$

and:

$$IsoqL(y) = \{x : DF_I(y, x) = 1\} \tag{8}$$

that is, the Farrell index is the inverse of the Shepard distance function, which measures the maximum amount by which an input vector can be shrunk along a ray while holding the output level constant. The value of the distance function equals unity if and only if observed inputs equal minimum potential inputs.

Consider, now, the output-oriented measure of technical efficiency. Production technology can be represented by an output set $P(x)$ defined as:

$$P(x) = \{y : (x, y) \qquad \text{is feasible}\} \tag{9}$$

For every $x \in R^n_+$, the output set has an isoquant defined as:

$$IsoqP(x) = \{y : y \in P(x), \theta y \notin P(x), \forall \theta \in (1, +\infty)\} \tag{10}$$

and an efficient subset:

$$EffP(x) = \{y : y \in P(x), y' \notin P(x), \forall y' \geq y\} \tag{11}$$

An alternative representation of the technology is the Shepard's output distance function (where o indicates that the measure is output-oriented):

$$D_o(x, y) = min\{\theta : (y/\theta) \in P(x)\} \tag{12}$$

where $D_o(x, y) \leq 1$. The output-oriented measure of technical efficiency is defined as:

$$DF_o(x,y) = max\{\theta : \theta y \in P(x)\} \geq 1 \tag{13}$$

and it follows from (13) that:

$$DF_o(x,y) = \frac{1}{D_o(x,y)} \tag{14}$$

and consequently:

$$IsoqP(x) = \{y : DF_o(x,y) = 1\} \tag{15}$$

This completes the formal derivation of the Farrell measures. The logic behind them should be clear: they measure as inefficiency the extent to which a firm's actual inputs usage (or output production) can be radially contracted (expanded) towards the boundary of the production possibility set (or frontier) and still allowing the firm to produce the same amount of output (or using the same amount of inputs). To implement these measures, it is necessary to identify the frontier a firm faces and then measure how far the latter is from it. As mentioned in the Introduction, two main methodologies are available in the literature: the parametric one, based on the econometric estimation of the frontier, with the residual being identified as the measure of inefficiency and the non-parametric one, based on linear programming techniques. We now turn to the analysis of these two methodologies starting with the parametric approach.

3 The econometric approach to efficiency measurement

3.1 The cross-sectional model

Consider a cross-section of n producers using a vector of inputs $\mathbf{x}$ to produce y. In this case, using a Cobb-Douglas functional form, the production technology can be represented by:

$$lny_i = \alpha + \beta ln\mathbf{x}_i + u_i + v_i \qquad i = 1, ..., n \tag{16}$$

where β is a vector of technology parameters to be estimated; for the estimation, an error term is attached to (16). The error term is usually assumed to have two different components: a stochastic one, v_i, picking up all the random factors which can affect (positively and negatively) production, having the usual Gaussian properties; and a systematic component, $u_i \leq 0$, measuring (in)efficiency and so all the factors which systematically affect production adversely, i.e. which do not allow the firm to produce as much output as its inputs usage would allow. For estimation purposes, it is assumed that it is distributed independently of v_i and that it is truncated. Parametric methods use econometrics to estimate the best practice frontier (i.e. the production technology of the best performer in the industry under analysis) and at the same time, after deriving a measure of the estimated residual, they measure technical efficiency as:

$$TE_i = exp(-u_i)|v_i \tag{17}$$

Different estimators for (16) have been suggested; historically, the first ones go under the label of deterministic methods and are linked to Afriat (1972) and Richmond (1974). They assume that the estimated residual $(u_i + v_i)$ measures the technical inefficiency of the decision making unit, sweeping away any other source of stochastic variation in the dependent variable. They estimate (16) by OLS, which allows us to estimate the 'average' practice within the industry; the best practice frontier is then derived by shifting up the constant of the estimated average production function by either the maximum positive residual (COLS) or the residuals' mean (MOLS, DOLS). Obviously, the main problem with deterministic methods is that they do not allow us to disentangle the stochastic shock from the inefficiency in the residual; therefore, they have been abandoned in favour of stochastic methods, due to Aigner *et al.* (1977), Meeusen and van den Broeck (1977) and Battese and Corra (1977). (16) is now estimated by Maximum Likelihood (ML) and the estimated residual is then decomposed into inefficiency and stochastic noise, by using the formula proposed by Jondrow, Lovell, Materov and Schmidt (JLMS) (1982). However, both the ML estimation and the JLMS decomposition require distributional assumptions on the two components of the error. Therefore, different formulations of the log-likelihood to maximise and of the JLMS decomposition have been derived under different distributions of the inefficiency.[3] Obviously, the key question is to what extent the efficiency scores and their ranking are sensitive to the different distributional assumptions. The answer to this question is not well documented in the literature. However, empirical studies where different distributional assumptions have been used for comparison show that both the rankings and

the efficiency scores are very similar across different distributions (Greene, 1993).

3.2 Panel Data Models

In the cross-sectional model, technical efficiency indices cannot be separated from firm-specific effects which, although not related to efficiency, can still enter in the inefficiency component of the residual. While this problem has been recognized long ago, it has been solved only since panel data have been available. Two methodologies have been developed to measure technical efficiency using panel data-sets: the former is based on the traditional panel data estimators (fixed and random effects) while the latter employs ML estimators. The traditional panel data estimator was initially proposed by Schmidt and Sickles (1984). They specified the following production frontier model:

$$lny_{it} = \alpha_0 + \beta lnx_{it} + u_i + v_{it} \qquad t = 1, ..., T \qquad i = 1, ..., N \qquad (18)$$

Outputs and inputs can now vary across time and producers. Statistical noise (v_{it}) varies over producers and time, but technical inefficiency (u_i) varies only over producers. The firm-specific inefficiency term can be merged with the constant to obtain a conventional panel data model:

$$lny_{it} = \alpha_i + \beta lnx_{it} + v_{it} \qquad (19)$$

(19) can be estimated either by the fixed effects (FE) (by using dummy variables to account for individual effects or, alternatively, by applying OLS

on the deviations of the time means) or by the random effects (RE) estimator (a GLS estimator), where the inefficiency component is allowed to be random. Efficiency scores are then computed by comparing the estimated α_i of each producer to its maximum estimated value. The main advantages of the FE estimator are that a) it is distribution free, i.e. no assumption on the two error components is required, and b) the fixed effects and the regressors can be correlated. However, the FE estimator suffers from some major shortcomings, which limit its use in the empirical analysis. First, efficiency is measured by using the fixed effects that by definition can pick up all firms' sources of heterogeneity, not necessarily related to inefficiency; so in a sense, the main benefit of using a panel data is lost by using the FE estimator. Second, it is not possible to infer statistically to what extent the estimated effects are significantly different from each other. Third, time-invariant variables cannot be introduced among the regressors. Finally, for large panels, it is practically intractable, as the number of parameters to estimate tends to be high. RE estimator shares with the FE estimator the advantage of not requiring a specific distribution on the efficiency component. However, it requires the effects to be uncorrelated with the variables in the model. This can be a particularly unreasonable assumption when modelling production relationships where inefficiency may be related to the usage of inputs and quality.

ML frontier estimation methods were introduced by Pitt and Lee (1981) and Battese and Coelli (1988), among others. Their starting point is the production model specified in (19); now, however, the effects (α_i) are assumed to follow a one-sided distribution (either half-normal or truncated) and to

be independent of the remaining variables. The production frontier is estimated by using Maximum Likelihood and the resulting error is decomposed into the stochastic shock and the inefficiency component, the latter being also used to compute the efficiency scores. The main advantage is that ML estimation allows to gain (statistical) efficiency in the estimation (as long as the assumption of independence is not a problem) and to have a tighter parameterisation allowing direct individual specific estimates of the inefficiency term in the model.

All the models presented so far, however, share a common shortcoming: they assume that inefficiency is time-invariant and clearly this is a problem if the time series is long. Several approaches have been suggested to solve this problem. Cornwell, Schmidt and Sickles (CSS) (1990) proposed a time-varying technical inefficiency panel data model where firm-specific temporal changes in technical inefficiency are allowed. More specifically, they replace α_i by α_{it} in (19) where:

$$\alpha_{it} = \theta_{i1} + \theta_{i2}t + \theta_{i3}t^2 \tag{20}$$

where the θ's are the parameters to be estimated. This way, efficiency can vary over time and in a different manner for each producer. Lee and Schmidt (1993) proposed a different generalisation where technical inefficiency effects are defined by the product of individual firms' effect and time effects:

$$\alpha_{it} = \theta_t \alpha_i \tag{21}$$

where $\theta_t = \Sigma_t \delta_t$ with δ_t being a dummy variable for each period t. In this model, the stochastic component is time-invariant while the time-varying

component is an *ad hoc* structure assumed to be common across firms. It seems appropriate only for short panels as the number of parameters to be estimated can increase easily otherwise. This model is usually estimated by using instrumental variable estimators. Two objections can be made to the CSS and Lee and Schmidt models. The first is of a more general nature: both models approximate the evolution of efficiency over time by using a set of time dummies or a time trend. Obviously, they do not allow us to control for the possibility of technical change (usually done by introducing a time trend), or movements of the frontier over time. So, by using these models, it is implicitly assumed that no technical change has occurred over time, which can be unreasonable if the panel is quite long. The second objection is of more practical nature: both specifications require a lot of parameters to be estimated and if the panel is particularly long, they can be beyond the capabilities of an average PC. Therefore, their applicability seems to be limited to very short panels and not surprisingly they are not used very often in the empirical literature.

A parallel development has occurred in the field of ML estimation of stochastic frontiers. Battese and Coelli (1992) proposed a stochastic production model for (un)balanced panel data where the temporal variation of technical inefficiency is modelled through an error component model. More specifically, technical efficiency is computed as $u_{it} = exp(-\eta(t-T))$ where η is a parameter to be estimated; u_{it} are the non-negative random variables which are assumed to account for technical inefficiency in production and are assumed to be i.i.d. as truncations at zero of the $N(\mu, \sigma_u^2)$. Battese and Coelli (1995) proposed a panel data frontier model where the non-negative

technical inefficiency term is assumed to follow a truncated distribution with different means for each DMU. So the distribution of the technical efficiency can now be modelled as a function of observed variables. In this setting, the inefficiency effects can be defined as:

$$u_{it} = z_{it}\delta + w_{it} \tag{22}$$

where z_{it} is a vector of observable explanatory variables, δ is a vector of unknown parameters and w_{it} is a random variable defined by the truncation of a Normal distribution with $-z_{it}\delta$ as the truncation point. By allowing the inefficiency to have different means and to model the impact of exogenous variables on the inefficiency, this technique is quite useful to explain the determinants of inefficiency and it is a valid alternative to the old two-stage procedure (where efficiency scores were first computed by the parametric approach and then regressed on exogenous factors).

Finally, just a few words on how to choose among the different models suggested by the literature. From the above discussion, it is clear that they impose different restrictions on the data and have different properties. In a short panel, where it is not possible to assume independence between the effects and the regressors, RE and ML estimators cannot be used and only the FE can. However, in this case, a lot of effort must be put to make sure that inefficiency is the only source of heterogeneity as picked up by the fixed effects. If the regressors and the effects can be maintained to be independent, then ML is more efficient than the traditional panel estimators because it exploits distributional information the other estimators do not.[4] In a long panel, it is advisable to derive time-variant measures of technical efficiency.

In this case, the distribution free approaches (based on the traditional panel estimators) are not easy estimators to implement as they impose an *ad hoc* structure of time variation (which prevents from controlling for technical change) and at the same time they are computationally cumbersome. So, if the distributional assumption on the inefficiency is not a problem, MLE appears to be the best model to use for its computational simplicity. In addition, it also allows us to model the impact that external factors have on the distribution of the inefficiencies.

3.3 Semiparametric methods

As mentioned, parametric methods based on ML estimation require both a distributional assumption on the error components and a functional form for the production technology. Both requirements can be a source of misspecification and attempts have been made to relax any of the two assumptions, by using semi-parametric econometrics. However, at the moment, in spite of the fact that there exists a significant number of semiparametric estimators, these cannot be considered to be a unified corpus of alternative estimators to the parametric ones; on the contrary, they tackle specific problems arising from the implementation of parametric estimators. Anyway, within their limited scope, they produce results of potential interest for the applied economist.

The various semi-parametric estimators proposed in the literature are now considered. A first set of estimators has been developed to estimate a panel data frontier model like in (19), where no distribution on the effects is assumed and different types of correlation between the effects and the regressors are allowed. We know from the previous section, that traditional panel data

estimators do not require a distributional assumption on the effects; however, they differ in that the FE estimator allows for correlation between the effects and *all* the regressors, while the RE estimator requires independence between the two sets of variables. Suppose now there is correlation between a *sub-set* of regressors and the effects in the model to estimate. In this case, the traditional panel estimators cannot be used. A first alternative was proposed by Hausmann and Taylor (1981) who proposed a hybrid between the within estimator and the GLS estimator. However, if the distribution of the effects is not known, this estimator is not very efficient. So semi-parametric estimators seem to be a more efficient (statistically speaking) alternative to the parametric estimators.

Park, Sickles and Simar (1998) consider the semiparametric estimation of a stochastic frontier model with panel data with a known linear functional form and an unknown density for the individual effects, that is:

$$y_{it} = \alpha_i + \beta \mathbf{x}_{it} + v_{it} \qquad i = 1, ..., N \qquad t = 1, ..., T \tag{23}$$

where y_{it} and $\mathbf{x}_{it}$ have the usual interpretation, v_{it} are i.i.d. random variables from $N(0, \sigma^2)$ and α_i are i.i.d. from an unknown density h whose support is bounded above. It is assumed that v_{it} and α_i are independent. In this context, the frontier is given by $\beta \mathbf{x}_{it} + B(h)$ where $B(h)$ is the upper boundary of the support of the density h while technical efficiency of the i-th observation is $\alpha_i - B(h)$.

They derive the efficient semiparametric estimator for (23) where different types of dependency between the regressors and the individual effects are allowed (More details on the semi-parametric estimation of a frontier

model are available on request). More specifically, they consider three types of models: a) Model I where no orthogonality restriction between effects and regressors is imposed; in this case, they show that the semiparametric efficient estimator is the Within estimator; b) Model II, where it is assumed a dependency between a subset of regressors and the effects. In this case, the semiparametric efficient estimator is:

$$\hat{\beta} = \hat{\beta}^{IV} + N^{-1}\Sigma_{i=1}^{N}\hat{\tilde{l}} \tag{24}$$

where $\hat{\beta}^{IV}$ is the consistent preliminary estimator of β (proved to be the Hausman-Taylor Instrumental Variable estimator) and $\hat{\tilde{l}}$ is an estimator of the so-called efficient influence function, which is estimated by a non-parametric kernel estimator, with $(1 + Tq)$ dimensions, where T is the number of time periods and q is the number of regressors correlated with the effects. c) Model III, which allows for correlation between the long-run movements of the regressors and the effects. This type of dependency may appear less obvious than the one in Model II, but it is relevant in industries where long-run changes in the employed factors are accompanied by changes in efficiency over time (due to possible improvement in the quality of inputs, for instance). In this case, the efficient semiparametric estimator is the same as in Model II, where the main difference is in the dimensions of the kernel estimator, now $(1 + q)$.

There are not many empirical applications of these estimators. Adams *et al.* (1999) have applied these estimators to the US banking industry, while Park *et al.* (1998) have used them to estimate the production frontier for a set of North-American and European airlines. From these, however, it is

possible to identify advantages and disadvantages of these estimators. First, from a statistical standpoint, they are definitely more efficient than the traditional panel estimators. Indeed, Park *et al.* (1998) show that there are 17-33 per cent efficiency (in statistical terms) gains in using the semi-parametric efficient estimator. However, most of these results are asymptotic and therefore require sizeable panel data. Clearly, the next step in the research would be to define the properties of these estimators in small samples. At the same time, too long panels may involve the use of multidimensional kernels that can be difficult to estimate (this is particularly true for Model II). So, at the moment, it appears as if the applicability of these estimators is limited to short panels and to models where there are not complicated patters of correlation between the effects and the variables. In spite of this, these estimators deserve attention as they allow us to model an important source of misspecification (i.e. correlation between regressors and effects) in empirical work.

A second set of semiparametric estimators has been suggested by Fan, Li and Weersink (1996) and Huang and Fu (1999) to estimate a cross-sectional frontier model where the functional form is unspecified but the composite error follows a known distribution. Their starting point is the typical stochastic production frontier model:

$$y_i = g(x_i) + e_i + u_i \qquad i = 1, ..., n \tag{25}$$

where g is a function unknown to the researcher; e_i and u_i are then the two error components, where the usual assumptions apply. The logic behind these estimators is quite simple. We know previously that parametric

estimation of a cross-sectional frontier model is based on ML estimation; this requires the derivation of a log-likelihood which is then maximised with respect to two parameters, σ^2 and λ. If $g(x_i)$ is known, then σ^2 can be computed independently and therefore the log-likelihood can maximised only with respect to λ, making the estimation of the parameters much easier. If the functional form is unknown, then these can be estimated by some non-parametric estimator. Fan, Li and Weersink (1996) suggest to write:

$$g(x) = E(y_i|x_i) + \mu \tag{26}$$

with $\mu = (2^{0.5}\sigma_u)/\pi^{0.5}$ and then estimate $E(y_i|x_i)$ by the following kernel estimator, $\hat{E}(y_i|x_i) = \Sigma_{j=1}^{n} y_j K((x_i - x_j)/h)/[\Sigma_{j=1}^{n} K((x_i - x_j)/h)]$. This way, $\hat{\sigma}^2$ can be calculated separately and a new log-likelihood function (called pseudo-likelihood) can be derived which allows us to derive a pseudo-likelihood estimate of λ. These estimates are then used to decompose the error in the two components, by using the JLMS decomposition formula. Huang and Fu (1999) build upon the method of Fan *et al.* by suggesting the use of the average derivative estimator to estimate non-parametrically $E(y_i|x_i)$ and then to derive the variance of the residual by maximising the pseudo-likelihood function as defined by Fan *et al.* The two error components are then calculated by the equivalent of the JLMS formula. Simulations show that both estimators perform like parametric estimators when having to estimate a correctly specified g and that they perform adequately in finite samples. These semiparametric estimators are quite appealing in that they allow us to estimate the (in)efficiency by avoiding a major source of misspecification, that is the functional form. Computationally speaking, they can be applied very

easily as they are straightforward applications of non-parametric estimators. Therefore, they are a viable alternative when estimating frontier models in settings where the use of restrictive functional forms may create problems.

4 Data Envelopment Analysis

The linear programming approach to the construction of the production frontier is known as Data Envelopment Analysis (DEA). It is non-parametric as it does not require an explicit functional form and constructs the frontier from the observed input-output ratios by linear programming techniques. The first DEA model was formulated by Charnes, Cooper and Rhodes (CCR) (1978) and was expressed mathematically as:

$$max_{\theta,\lambda}\theta \tag{27}$$

subject to:

$$\mathbf{X}\lambda \leq x_0 \tag{28}$$

$$\theta y_0 \leq \mathbf{Y}\lambda \tag{29}$$

$$\lambda \geq 0 \tag{30}$$

where $\mathbf{X}$ is an n by I input matrix with columns x_i, $\mathbf{Y}$ is an m by I output matrix with columns y_i and λ is an I by 1 intensity vector. In the DEA problem the performance of a producer is evaluated in terms of his ability to expand its output vector subject to the constraints imposed by best observed practice. If radial expansion is possible for a producer, its

optimal $\theta > 1$, while if radial expansion is not possible, its optimal $\theta = 1$.

The CCR problem imposes constant returns to scale on the technology (expressed by the fact that λs are free to vary between zero and infinity). A second version of the DEA problem was suggested by Banker, Charnes and Cooper (1984) (BCC) which relaxes the assumption of constant returns to scale. The mathematical format of the BBC problem is the same as in (27-30), where the only difference is that (30) is now substituted by the sum of λs constrained to be equal to one.

DEA does not allow us to model stochastic shocks to production, i.e. it is deterministic. Therefore the computed efficiency scores may be biased by factors which are external to the production process. Not surprisingly, some attempts have been made to incorporate stochastic components into the linear programming problem. Varian (1985) first introduced a two-sided deviation to include the random noise and to calculate the efficiency measure free of such random noise. Land *et al.* (1993) provided the so-called chance-constrained efficiency analysis which allowed the deterministic frontier to capture the effect of random noise without being stochastic. The data requirements of the chance-constrained efficiency measurement, however, are too many. Indeed, it is necessary to have information on the expected values of all variables, along with their variance and covariance matrices and the probability levels at which feasibility constraints are to be satisfied. Therefore, this approach is too informationally demanding to be implemented easily.

4.1 Statistical properties of DEA estimators

A common critique to the DEA (frequently voiced within the "parametric circles") is that no statistical inference can be carried out on the computed efficiency scores. Indeed, in its earlier versions, the DEA-based efficiency scores were not accompanied by the equivalent of standard errors. However, nowadays, this critique is unfair as there exists a considerable amount of research that has characterised the statistical properties of DEA estimators, has designed hypotheses tests on DEA scores and employed bootstrapping to construct confidence intervals.

Let us go in order. Before being able to define the statistical properties of the DEA estimators and to carry out any type of inference, it is necessary to establish the statistical model of the Data Generating Process (DGP), here indicated as P, which generates the data upon which DEA is then applied. This task has been taken on by Simar and Wilson, (2000), who defined the statistical model as follows:

A1. Inputs and outputs in P are freely disposable.

A2. The sample observations, (x_i, y_i) are i.i.d. random variables with probability density function $f(x, y)$ with support over P.

A3. The probability of observing units on P must approach unity as the sample size increases.

A4. For all (x, y) in the interior of Φ, $\theta(x, y)$ (the true Farrell measure of technical efficiency) is differentiable in both its arguments.

The statistical properties of the DEA estimators can now be derived. Banker (1993) showed that the DEA estimator is an asymptotic weakly consistent estimator of an arbitrary monotone and concave production function

with one input and one (or more) output(s), where deviations from the production frontier are considered stochastic variations in the technical efficiency of individual observations. Next, Korostelev *et al.* (1995) have analysed the convergence rate of this DEA estimator (in the same Banker's framework) and show that as the number of outputs increases, the number of observations must increase at an exponential rate. The same result is obtained in a multivariate framework by Kneip *et al.* (1998). As for the sampling distribution of these estimators (necessary to be able to make inference), Gijbels *et al.* (1999) prove that in the univariate case (one input, one output) the estimator's sampling distribution approximates an F-distribution.

Once the statistical properties of the DEA estimators have been defined, it is possible to construct hypothesis tests on the DEA efficiency scores. Banker (1993, 1996) has suggested a battery of hypotheses tests allowing us to test the DEA model specification, the type of returns to scale the technology exhibits and whether the inefficiency scores computed from two samples are significantly different. We consider here in detail the different tests for this last hypothesis (The other two types of tests are available on request). Suppose we apply DEA to two different samples of firms (having size N_1 and N_2 respectively) and we want to test whether the computed efficiency scores, θ, are different across them. If the efficiency scores follow an exponential distribution in both samples, then the mean efficiency score for the first sample is $1+\mu_1$, while for the second sample it is $1+\mu_2$: in this case the null hypothesis can be specified as $H_0 : \mu_1 = \mu_2 = \mu$ (indicating that both types of firms have the same inefficiency distribution) and the alternative hypothesis as $H_1 : \mu_1 > \mu_2$ (implying that on average the firms of

the first group are less efficient than the second type). So the following test statistics can be used:

$$T_{EX} = \frac{\Sigma_{j \in N_1}(\hat{\theta}_j^B - 1)/N_1}{\Sigma_{j \in N_2}(\hat{\theta}_j^B - 1)/N_2} \tag{31}$$

which is distributed as an $F(2N_1, 2N_2)$-statistic. If the efficiency scores follow a Normal distribution, $N(0, \mu_i)$ with $i = 1, 2$, under the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu$ and the alternative hypothesis $H_1 : \mu_1 > \mu_2$, (31) changes to:

$$T_{HN} = \frac{\Sigma_{j \in N_1}(\hat{\theta}_j^B - 1)^2/N_1}{\Sigma_{j \in N_2}(\hat{\theta}_j^B - 1)^2/N_2} \tag{32}$$

which is distributed as an $F(N_1, N_2)$-statistic. Finally, if the efficiency scores do not follow any distribution, a non-parametric Smirnov test can be used.

The possibility of carrying out hypotheses tests on the DEA efficiency score is definitely appealing; however, so far, the scope for applying these tests is quite limited as their properties are mostly asymptotic; indeed in small samples the tests statistics do not follow an F distribution. This makes them difficult to be used in the bulk of applied work which is usually based on small samples. Therefore subsequent research has explored the feasibility within the DEA context of bootstrapping, which allows us to approximate the sampling distribution of $\hat{\theta}$ and so to build confidence intervals.

4.2 Bootstrapping

Bootstrap is a data-based simulation method for statistical inference which can be used to build confidence intervals for parameters in situations where these cannot be derived analytically. The basic idea involves sampling with replacement to produce random samples of size n from the original data. Each of these samples is known as a bootstrap sample and provides an estimate of the parameters of interest. Repeating the sampling a large number of times provides information on the variability of the estimator. Extending bootstrap techniques to the DEA environment is quite straightforward and Simar and Wilson (1998) propose the following procedure. Assume the data χ_n are generated from the DGP $P = P(\phi, f(x, y))$, where ϕ is the unknown production set and $f(x, y)$ has been defined above. By using the data χ_n, it is possible to get a consistent estimator of P, of the unknown production set, $\hat{\phi}$, and of the efficiency scores, $\hat{\theta}(x_o, y_o)$. The efficiency scores estimates can be considered as a new population from which it is possible to draw a new dataset (or pseudo-sample), $\chi^* = (x_i^*, y_i^*)$. By applying DEA to the pseudo-sample, new estimators of the production set, $\hat{\phi}^*$ and of the efficiency scores, $\hat{\theta}^*$, can be computed. This operation can be repeated B times so to get B pseudo-samples and pseudo-estimates of the efficiency scores. The empirical distribution of these pseudo-estimates gives an approximation of the unknown sampling distribution of the efficiency scores[5].

If the distribution of $(\hat{\theta}(x_0, y_0) - \theta(x_0, y_0))$ was known, then it would be possible to find values of a_α and b_α such that:

$$Pr(-b_\alpha \leq \hat{\theta}(x_0, y_0) - \theta(x_0, y_0) \leq -a_\alpha) = (1 - \alpha) \tag{33}$$

Using the bootstrap values of the pseudo-estimates $\hat{\theta}^*$, we can find values of $\hat{b}_\alpha$ and $\hat{a}_\alpha$[6] such that:

$$Pr(-\hat{b}_\alpha \leq \hat{\theta}^*(x_0, y_0) - \hat{\theta}(x_0, y_0) \leq -\hat{a}_\alpha | \hat{P}(\chi_n)) = (1 - \alpha) \tag{34}$$

So the bootstrap approximation of (33) is:

$$Pr(-\hat{b}_\alpha \leq \hat{\theta}(x_0, y_0) - \theta(x_0, y_0) \leq -\hat{a}_\alpha) \approx (1 - \alpha) \tag{35}$$

The estimated $(1 - \alpha)$-percent confidence interval is:

$$\hat{\theta}(x_0, y_0) + \hat{a}_\alpha \leq \theta(x_0, y_0) \leq \hat{\theta}(x_0, y_0) + \hat{b}_\alpha \tag{36}$$

An important issue when using the bootstrap in a frontier context is to ensure that in generating the pseudo-samples and the pseudo-estimators, these are consistent, i.e. the following property holds:

$$(\hat{\theta}^* - \hat{\theta}) | \hat{P}(\chi_n) \approx (\hat{\theta} - \theta) | P \tag{37}$$

Now a common version of the boostrap (called naïve bootstrap) produces inconsistent pseudo-estimators (Efron and Tibshirani, 1993), as they do not consider the fact that in the frontier context the distribution functions have a bounded, unknown support on P. To solve the problem, Simar and Wilson suggest to draw pseudo-data sets from a smooth, consistent, kernel estimate $\hat{f}(\theta)$ of the marginal density of the original estimates $\hat{\theta}$[7].

Finally, the procedure suggested by Simar and Wilson is quite easy to implement as it requires the use of basic bootstrap techniques, just modified to take into account of the frontier context. So at the moment, it is the best

way to carry out some inference in the DEA framework.

5 TFP growth and its components

As mentioned in the Introduction, frontier techniques have been extended so to be able to measure productivity growth. Traditional methods to the measurement of productivity have usually assumed that firms are located on the frontier, i.e. they are efficient. This implied that productivity movements were assumed to be due to shifts of the frontier (or technical change). However, if a firm is inefficient, then the contribution of efficiency change to productivity change has to be acknowledged. We know that each time period a firm faces a best practice frontier, it defines the maximum output a firm can produce for a given inputs. The frontier is determined by the state of technical progress at that time. A firm is not always located on the best practice frontier, but can be located anywhere in the production possibility set (and therefore regarded as technically inefficient). Any movement of productivity over time can be decomposed into two parts: movements of the frontier due to changes in the technological capabilities of the firm (technical change) and movements of the firm towards (or far from) the frontier as it is more (or less) successful at reducing internal inefficiency. In addition, if we allow the production technology to have decreasing returns to scale, then there is an additional component of productivity change, the scale component, as the firm from one period to the other moves along the production technology and exploits its curvature. The distinctive feature of the frontier approach is in the fact that the change in efficiency is now a component of TFP. Obviously, the legitimate question, now, is to what extent this is a significant compo-

nent, from an empirical standpoint. Generally speaking, efficiency change is quite relevant and is an important source of productivity variation. However, more striking is the relative importance of the efficiency change component over the other components. Indeed, in studies carried out on private sector organisations, efficiency change (while still being a source of productivity variation) is less important than technical change, for instance (Fare *et al.*, 2001). By contrast in public sector organisations reductions in inefficiency can still be a sizeable source of productivity gains (Linna, 1998). This can be easily explained by the fact that in the private sector organisational slack is smaller than in the public sector.

There are two main methodologies for the measurement of productivity change using frontier techniques: the parametric approach, based on the stochastic frontier analysis, and the non-parametric approach, linked mainly to the developments of the Malmquist index. We will now focus indeed on the two different approaches, starting with the parametric one.

5.1 The Parametric Approach

The first to apply parametric methods to the measurement of productivity change were Nishimizu and Page (1982) who, by using a parametric, deterministic approach, measured productivity growth as the sum of the (deterministic) frontier technical change and change in efficiency. The key idea was to generalise the Solow model in which technical change and productivity change were identical and to allow for inefficiency. Afterwards, the Nishimizu and Page approach has been extended to make use of the possibilities of the stochastic frontier approach. Consider the following production function for

producer i at time t:

$$y_{it} = f(\mathbf{x}_{it}, \beta, t)exp(\epsilon_{it}) \tag{38}$$

where y_{it} is the output at time t for production unit i and $\mathbf{x}_{it}$ is a n-vector of inputs; $\epsilon_{it} = u_{it} + v_{it}$ is the residual. The productivity change index is defined as the difference between the rate of change of output and the rate of change of inputs:

$$T\dot{F}P = \dot{y} - \dot{\mathbf{x}} \tag{39}$$

If we totally differentiate (38) and insert into (39), we get:

$$T\dot{F}P = \Delta T + (\epsilon - 1)\Sigma_n(\frac{\epsilon_n}{\epsilon})\dot{\mathbf{x}} + \Delta TE \tag{40}$$

where ΔT is a measure of the rate of technical change, ΔTE is a measure of the rate of technical efficiency, ϵ_n are the elasticities of output with respect to each input and ϵ is the sum of all the elasticities, providing a measure of the returns to scale characterising the production frontier. This way, productivity change is decomposed into a technical change component, a technical efficiency change component and a scale component.

Assume a translog functional form is used for f in (38):

$$lny_{it} = \beta_0 + \beta_L lnL_{it} + \beta_t t + \beta_{Lt} tlnL_{it} + 0.5\beta_{LK} lnL_{it} lnK_{it} + 0.5\beta_{tt} t^2 + \beta_K lnK_{it} + \beta_{Kt} tlnK_{it} + v_{it} + u_{it} \tag{41}$$

then, (40) becomes:

$$\Delta T = \hat{\beta}_t + \hat{\beta}_{tt}t + \Sigma_{n=1}^{2}\hat{\beta}_{nt}lnx_{nit} \tag{42}$$

$$\Delta TE = \hat{\dot{u}_{it}} \tag{43}$$

$$\hat{\epsilon}_L = \hat{\beta}_L + \hat{\beta}_K lnK_{it} + \hat{\beta}_L t \tag{44}$$

$$\hat{\epsilon} = \Sigma_{n=1}^{2}(\hat{\beta}_n + \hat{\beta}_n lnx_{it} + \hat{\beta}_{nt}t) \tag{45}$$

where $1 = L$, $2 = K$ and x is a generic input. (41) is structurally similar to the panel data models considered above; therefore in principle all the panel data estimators which allow us to compute time-variant efficiency scores could be used. However, estimators based on traditional panel data techniques use time dummies to compute efficiency change and so they do not allow us to distinguish between technical efficiency change and technical change in the estimation of (41). Therefore, ML estimation is again the best option as long as the assumptions on the error components and of independence do not create problems. A nice feature of the parametric approach (and therefore its main strength) is that it allows us to test the type of returns to scale the technology exhibits and the statistical significance of each component of TFP, allowing us to know to what extent efficiency change (among the others) is significant in explaining productivity movements. However, its main weakness is that it does not provide disaggregated (i.e. producer-specific) information on technical change and scale change (unlike the Malmquist index) and this explains why it has been overshadowed by the development of Malmquist to the point that the frontier approach to productivity change seems to identify with the Malmquist index.

5.2 The Malmquist index

The Malmquist index, allowing us to measure a firm's productivity growth, was introduced by Caves *et al.* (CCD) (1982). The Malmquist index measures productivity change by comparing the position of a firm in two adjacent time periods with respect to a best practice frontier, measured by a distance function. The index was named after Malmquist who in 1953 proposed a quantity index (based on ratios of distance functions) for use in consumption analysis. CCD define an output-based Malmquist productivity index relative to a single technology at time t as:

$$M^t = \frac{D_o^t(x^{t+1}, y^{t+1})}{D_o^t(x^t, y^t)} \tag{46}$$

where the output-input pairs used at time t and at time $t+1$ are compared to the technology available at time t, measured by the distance function D_o^t. The equivalent index for time $t+1$ is:

$$M^{t+1} = \frac{D_o^{t+1}(x^{t+1}, y^{t+1})}{D_o^{t+1}(x^t, y^t)} \tag{47}$$

Later on, Fare *et al.* (1989) employed the geometric mean of the two output-based Malmquist indices, (46) and (47) to define the following index of productivity growth:

$$M = \left[\frac{D_o^t(x^{t+1}, y^{t+1})}{D_o^t(x^t, y^t)} \frac{D_o^{t+1}(x^{t+1}, y^{t+1})}{D_o^{t+1}(x^t, y^t)} \right]^{\frac{1}{2}} \tag{48}$$

where the first ratio measures the change in technical efficiency from time t to time $t+1$ and the second ratio measures the technology change. $D_o^t(x^t, y^t)$ and $D_o^{t+1}(x^{t+1}, y^{t+1})$ are the output distance functions measured at time t and

$t+1$, respectively. $D_o^t(x^{t+1}, y^{t+1})$ is the mixed-period distance output function measuring the maximum proportional change in output required to make the observed input-output pair at period $t+1$ feasible relative to the technology of period t, while $D_o^{t+1}(x^t, y^t)$ measures the maximum proportional change in output required to make the input-output pair observed at time t, feasible relative to the technology of time $t+1$.

(48) can be reformulated to highlight the roles of technical progress and of technical efficiency change:

$$M = \frac{D_o^{t+1}(x^{t+1}, y^{t+1})}{D_o^t(x^t, y^t)} \left[\frac{D_o^t(x^{t+1}, y^{t+1})}{D_o^{t+1}(x^{t+1}, y^{t+1})} \frac{D_o^t(x^t, y^t)}{D_o^{t+1}(x^t, y^t)} \right]^{\frac{1}{2}} \tag{49}$$

with technical efficiency change being the first ratio and technical change the product of the two ratios in the brackets. A Malmquist index greater (smaller) than one indicates an increase (decrease) in productivity; the same is true for its components.

It is noticeable that no mention has been made of the scale component in the above decomposition of the Malmquist index. Indeed, as such, the index can provide a correct measure of productivity growth only when the underlying technology has constant returns to scale, i.e. there is no scale component in the productivity growth. This implies that each distance function in (48) must be computed under the assumption of constant returns to scale. In the presence of non-constant returns to scale, the Malmquist index provides an inaccurate measure of productivity change even if the distance functions are computed under variable returns to scale (Grifell-Tatje and Lovell, 1995). Therefore, a raft of new productivity indexes (Ray and Desli, 1997; Grifell-Tatje and Lovell, 1999; Balk, 2001) have been proposed, having in common

the idea of scaling the Malmquist index by a term which accounts for the scale component. There is a heated debate on what is the best decomposition and so at the moment this issue seems pretty unsettled.

The Malmquist index is very popular to measure and decompose productivity growth. It is used very often as an alternative to both the Tornquist and Fisher indices as, unlike them, it does not require information on prices, but only on quantities. So, it is very useful when measuring productivity in public sector organisations, where information on input and output prices is rarely available. Another reason for its success is that it allows us to easily identify the sources of productivity growth and provides a lot of producer-specific information on technical change and the scale component for time period, which can be very useful for decision-making. Computationally speaking, both parametric and non-parametric methods can be used in principle to estimate the distance functions making up the index. In the empirical literature, it is quite common to use DEA. In earlier days, it was not possible to carry out statistical inference on the computed distance functions; nowadays it is possible to use bootstrap to compute confidence intervals on each component of the Malmquist index (Simar and Wilson, 1999). However, the main weakness of DEA (namely that it is a deterministic method) is still there and so the computed distance functions may include the effect of factors not related to technical efficiency and technical change. Of course stochastic frontier methods can be used (Grifell-Tatje *et al.*, 2001). They can be appealing as they allow us to disentangle the stochastic noise from the efficiency measures. However, the econometric estimation of distance functions is not trivial; indeed, in the case of a distance function, output is

endogenous and some IV estimator is necessary in this context. Therefore, the use of a parametric method for the estimation of the Malmquist index does not seem a feasible route. The best option left to the researcher is to try to specify the DEA model (underlying the Malmquist index) in the best possible way so to minimise the impact of external factors on the computed distance functions.

6 TFP and undesirable outputs

When evaluating the performance of producers it makes sense to credit them for the provision of desirable outputs and penalise them for the production of undesirable outputs; that is, good and bad outputs should be treated asymmetrically. However, for a long time, the problem of undesirable outputs has been ignored in the performance measurement literature. Indeed, the traditional output-oriented Farrell measure cannot treat good and bad outputs asymmetrically because it seeks to increase both simultaneously. In 1989, Fare *et al.* (1989) modified the Farrell measure to permit an asymmetric treatment of outputs, allowing the maximum radial expansion of all desirable outputs and contraction of all undesirable outputs, holding inputs constant. More recently, directional distance functions have proved extremely useful in measuring performance to account for such outputs (Chung emphet al., 1997; Ball *et al.*, 1998). Their distinctive feature is that they measure the amount by which an output vector can be translated radially from itself to the technology frontier in a pre-assigned direction. In addition, they are a generalisation of Shepard's output distance functions (Chambers emphet al., 1998) and this relationship has been used to compute a Malmquist-type

productivity index (so-called Malmquist-Luenberger (ML) index) that allows us to measure productivity in the presence of undesirable outputs (Chung *et al.*, 1997).

To define a directional distance function, it is important to make some assumptions on the production process. Consider a productive process with the following output set:

$$P(x) = \{(y, b) : x\text{can produce}(y, b)\}. \tag{50}$$

where $y \in R_+^M$ denotes the desirable output, $b \in R_+^I$ is the undesirable output and $x \in R_+^N$ is an input. Desirable and undesirable outputs are produced simultaneously, implying that the desirable output is "nulljoint" with the undesirable output: the only way not to produce the undesirable output is by not producing the desirable output, that is if $(y, b) \in P(x)$ and $b = 0$ then $y = 0$. This simultaneous production of desirable and undesirable outputs implies that the reduction of the undesirable output is costly (i.e. there is weak disposability of undesirable outputs) as it requires the firm to reduce the desirable output. Therefore $(y, b) \in P(x)$ and $0 \leq \theta \leq 1$ imply $(\theta y, \theta b) \in P(x)$. Finally, it is assumed that the desirable output is freely disposable, i.e. $(y, b) \in P(x)$ and $y^{'} \leq y$ imply $(y^{'}, b) \in P(x)$.

Under these assumptions, the technology can be completely characterised by a directional output distance function. Formally, a directional output distance function is defined as:

$$\overrightarrow{D}_0(x, y, b; g) = \{sup(\beta : y, b) + \beta g \in P(x))\} \tag{51}$$

where g is the vector of 'directions' by which outputs are scaled. Direc-

tional distance functions are a generalisation of the Shepard output distance function (Chambers *et al.*, 1998). More specifically, let $g = (y, -b)$ (implying that the positive output is expanded and the negative output is contracted), then:

$$\overrightarrow{D}_0(x, y, b; g) = \frac{1}{D_0(x, y, b; g)} - 1 \tag{52}$$

(52) can be used to define the output-oriented ML index as:

$$ML_o = [\frac{1 + \overrightarrow{D}_o^t(x^t, y^t, b^t; g)}{1 + \overrightarrow{D}_o^t(x^{t+1}, y^{t+1}, b^{t+1}; g)} \frac{1 + \overrightarrow{D}_o^{t+1}(x^t, y^t, b^t; g)}{1 + \overrightarrow{D}_o^{t+1}(x^{t+1}, y^{t+1}, b^{t+1}; g)}]^{0.5} \tag{53}$$

The ML index (as the Malmquist index) can also be decomposed into two components of technical efficiency (MLEFFCH) and technological change (MLTECH):

$$MLEFFCH_o = \frac{1 + \overrightarrow{D}_o^t(x^t, y^t, b^t; g)}{1 + \overrightarrow{D}_o^{t+1}(x^{t+1}, y^{t+1}, b^{t+1}; g)} \tag{54}$$

$$MLTECH_o = [\frac{1 + \overrightarrow{D}_o^{t+1}(x^t, y^t, b^t; g)}{1 + \overrightarrow{D}_o^t(x^t, y^t, b^t; g)} \frac{1 + \overrightarrow{D}_o^{t+1}(x^{t+1}, y^{t+1}, b^{t+1}; g)}{1 + \overrightarrow{D}_o^t(x^{t+1}, y^{t+1}, b^{t+1}; g)}]^{0.5} \tag{55}$$

so that their product is equal to ML_0. The efficiency change term is equivalent to the ratio of the directional distance functions in two adjacent time periods. The technical change term is the geometric mean of the shift in technology evaluated at x^{t+1} and the shift in technology evaluated at x^t. The ML index indicates an increase in productivity if its value is greater than one

and productivity decline if its value is less than one. The directional distance functions composing the ML index can be computed using a modification of the DEA problem where the bad output is now weakly disposable.

A classical example where the ML index is relevant is the evaluation of the performance of a firm having to reduce pollutants (bad output) to meet environmental regulations. The hospital sector is another example where directional distance functions can provide useful information (Dismuke and Sena, 2001). Traditionally, when the performance of an hospital is evaluated, they are assumed to 'produce' discharges and typically no distinction is made between their discharge status (i.e. dead or alive). Therefore, dead and alive discharges are treated in an equivalent manner and when the traditional Malmquist index is used, any effort of the hospital management to reduce the mortality rate may be recorded as a productivity decrease. However, from the social standpoint, what really matters is the ability of the hospital to reduce dead discharges and increase the alive ones. In this case, a ML index is more suited to measure productivity growth than the traditional Malmquist index as the former allows to derive a measure of productivity growth which, by rewarding the hospital's efforts to reduce mortality, gives a picture of the hospital's performance more responding to society's expectations.

7 Conclusions

An overview of the frontier approach to the measurement of technical efficiency and productivity has been provided. Both parametric and non-parametric approaches have been presented, where the former is based on the econometric estimation of the production frontier and the latter uses

linear programming techniques to construct the frontier.

Stochastic methods have the advantage of controlling for random shocks to the production and measurement errors in the inputs and outputs. However, they can be subject to specification errors, even if, by using semiparametric estimators, this problem can be addressed. Another problem is the number of parameters to be estimated, which can simply be too many, especially when using panel data. The main weakness of the non-parametric approach is the fact that it is a deterministic method. In spite of attempts to solve this problem (such as with chance constrained DEA), it is still impossible to control for random shocks to production. On the other hand, what was regarded to be the main problem with DEA, namely the lack of statistical inference, has now been solved and there is a substantial literature on hypothesis tests and how to construct confidence intervals by bootstrapping the DEA efficiency scores. The main advantage of DEA is that it provides a lot of producer-specific information, like dual solution values, slacks and so on. It is really impossible to suggest one approach over the other, as they both have positive and negative features; in a sense, they could be used jointly as they provide complementary information. At any rate, it is clear that the frontier approach offers an interesting set of tools to measure efficiency and TFP and so contribute to decision-making within both private and public organisations.

Endnotes

[1]The author wishes to thank Toke Aidt, Huw Dixon, Peter Simmons, Peter Smith, Gabriel Talmain and Emmanuel Thanassoulis for useful comments on previous drafts of the work. The usual disclaimer applies. Address for correspondence: LUBS, University of Leeds, Leeds, LS2 9JT, United Kingdom; telephone: 0044-113-3434514, e-mail:vs@lubs.leeds.ac.uk.

[2]Interested readers can refer to Kumbhakar and Lovell (2000) and Cooper *et al.* (2000) for thorough introductions to parametric and non-parametric methods, respectively.

[3]The most common distributions in the empirical work are the Half-Normal, the Truncated and the Exponential. Other distributions have been suggested such as the Gamma; however, given its complexity, it is rarely used in empirical work. See also Lovell and Kumbhakar (2000).

[4]Hallam and Machado (1996) and Bravo-Ureta and Ahmad (1996) have compared the different panel data estimators. They find that MLE provides better measures of the efficiency scores; however, they also seem to conclude that the two approaches (MLE and traditional panel estimators) may generate similar rankings.

[5]The quality of the approximation depends on the value of B. The bigger the number of draws, the smaller the approximation error.

[6]$\hat{b}_\alpha$ and $\hat{a}_\alpha$ are unknown and can be computed by sorting the values $\hat{\theta}(x_0, y_0) - \theta(x_0, y_0)$ in increasing order and then delete $\alpha/2$ percent of the elements at either end of the sorted list; then $-\hat{b}_\alpha$ and $-\hat{a}_\alpha$ are set equal to

the endpoints of the truncated sorted array.

[7]This is equivalent to assuming that the distribution of inefficiencies is homogenous. This assumption can be relaxed (as in the so-called heterogeneous approach) but at a cost of increased complexity and computational burden.

References

Adams R M, Berger A N and Sickles R C (1999) 'Semiparametric Approaches to Stochastic Panel Frontiers with Application in the Banking Industry', *Journal of Business and Economics Statistics*, 17, 3, 349-358.

Afriat S (1972) 'Efficiency Estimation of Production Functions', *International Economic Review*, 13:3, October, 568-598.

Aigner D J , Lovell C A K and Schmidt P.J. (1977) 'Formulation and Estimation of Stochastic Frontier Production Function Models', *Journal of Econometrics*, 6, 1, 21-37.

Ahmad M and Bravo-Ureta B E (1996) 'Technical Efficiency Measures for Dairy Farms Using Panel Data: A Comparison of Alternative Model Specifications' *Journal of Productivity Analysis*, 7(4), 399-415.

Ball E, Fare R, Grosskopf S and Nehring R. (1998) 'Productivity of the US Agricultural Sector: the case of Undesirable Outputs', *mimeo.*

Balk B M (2001) 'Scale Efficiency and Productivity Change', *Journal of Productivity Analysis*, 15, 159-183.

Banker R D, Charnes A, Cooper W W (1984) 'Some Models for Estimating Technical and Scale Efficiency in Data Envelopment Analysis', *Management Science*, 30, 9, 1078-1092.

Banker R D (1993) 'Maximum Likelihood, Consistency and Data Envelopment Analysis', *Management Science*, 39:10, 1265-1273.

Banker R D (1996) 'Hypothesis Tests Using Data Envelopment Analysis', *Journal of Productivity Analysis*, 7, 139-159.

Battese G E and Coelli T J (1988) 'Prediction of Firm-Level Technical Efficiencies with a Generalised Frontier Production Function and Panel Data', *Journal of Econometrics*, 38, 387-399.

Battese G E and Coelli T J (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 E and Coelli T J (1995) 'A Model for Technical Efficiency Effects in a Stochastic Frontier Production Function for Panel Data', *Empirical Economics*, 20, 325-332.

Battese, G E and Corra G S (1977) 'Estimation of a Production Frontier Model: With Application to the Pastoral Zone of Eastern Australia', *Australian Journal of Agricultural Economics*, 21(3), 169-79.

Caves D, Christensen L and Diewert W E (1982) 'The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity', *Econometrica*, 50(6), 1393-1414.

Chambers R G, Chung Y and Fare R (1998) 'Profit, Directional Distance Functions and Nerlovian Efficiency', *Journal of Optimization Theory and*

Applications, 70, 407-419.

Charnes A, Cooper W W and Rhodes E (1978) 'Measuring the Efficiency of Decision-making Units', *European Journal of Operation Research*, 2(6), 429-444.

Chung Y, Fare R and Grosskopf S (1997) 'Productivity and undesirable outputs: a directional distance function approach', *Journal of Environmental Management*, 51, 229-240.

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

Cooper W W, Seiford L M and Tone K (2000) *'Data envelopment analysis: A comprehensive text with models, applications, references and DEA Solver software'* London: Kluwer.

Dismuke C E and Sena V (2001) 'Is There a Trade-Off Between Quality and Productivity? The Case of Diagnostic Technologies in Portugal', *Annals of Operations Research*, 107, 101-116.

Efron B and Tibshirani R J (1993) *An Introduction to the Bootstrap*, London: Chapman and Hall.

Fan Y, Li Q and Weersink A (1996) 'Semiparametric Estimation of Stochastic Production Frontier', *Journal of Business and Economics Statistics*, 14, 460-468.

Fare R, Grosskopf S, Lindgren B and Roos P (1992)'Productivity in Swedish Pharmacies: A Malmquist Input Index Approach', *Journal of Productivity Analysis*, 3(1/2), 270-285.

Fare R, Grosskopf S, Lovell C A K and Pasurka C (1989) 'Multilateral Productivity Comparisons when Some Outputs are Undesirable: A Nonparametric Approach', *The Review of Economics and Statistics*, 71, 1, 90-98.

Fare R, Grosskopf, S and Lee W F (2001) 'Productivity and Technical Change: The Case of Taiwan', *Applied Economics*, 33(15), 1911-25.

Farrell M J (1957) 'The Measurement of Productive Efficiency',*Journal of the Royal Statistical Society*, Series A, General, 120, 253-281.

Fuentes H J, Grifell-Tatje E and Perelman S (2001) 'A Parametric Distance Function Approach for Malmquist Productivity Index Estimation' ,*Journal-of-Productivity-Analysis*, 15(2), 79-94.

Gijbels I, Mammen E, Park B U and Simar L (1999) 'On Estimation of Monotone and Concave Frontier Functions', *Journal of the American Statistical Association*, 94, 220-228.

Greene W (1993) 'The Econometric Approach to Efficiency Analysis', in Fried H, Lovell C A K and Schmidt P (eds) *The Measurement of Productive Efficiency*, Oxford: Oxford U P.

Grifell-Tatje E and Lovell C A K (1995) 'A Note on the Malmquist Productivity Index', *Economics Letters*, 47, 169-175.

Grifell-Tatje E and Lovell C A K (1999) 'A Generalised Malmquist Productivity Index', *TOP*, 7(1), 81-101.

Grosskopf S (1993) 'Efficiency and Productivity', in Fried H, Lovell C A K and Schmidt P (eds) *The Measurement of Productive Efficiency*, Oxford: Oxford U P.

Hallam D and Machado F (1996) 'Efficiency Analysis with Panel data: a Study of Portuguese Diary Farms', *European Review of Agricultural Economics*, 23(1), 79-83.

Hausmann J A and Taylor E E (1981) 'Panel Data and Unobservable Individual Effects', *Econometrica*, 49, 1377-1398.

Huang C J and Fu T T (1999) 'An Average Derivative Estimation of Stochastic Frontiers', *Journal of Productivity Analysis*, 12, 45-53.

Jondrow J, Lovell C A K, Materov I S and Schmidt P (1982) 'On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model', *Journal of Econometrics*, 19, 2/3, 233-238.

Kneip A, Park B U and Simar L (1998) 'A Note on the Convergence of Nonparametric DEA Estimators for Production Efficiency Scores', *Econometric Theory*, 14, 783-793.

Koopmans T C (1951) 'An Analysis of Production as an Efficient Combination of Activities', in TC Koopmans (ed) *Activity Analysis of Production and Allocation*, Cowles Commission for Research in Economics, Monograph 13, New York: Wiley.

Korostolev A P, Simar L and Tsybakov, A B (1995) 'On Estimation of Monotone and Convex Boundaries', *Publications de l'Institut de Statistique de l'Universite' de Paris*, 39:1, 3-18.

Kumbhakar S and Lovell C A K (2000) *Stochastic Frontier Analysis*, Cambridge: Cambridge U P.

Land K C, Lovell C A K and Thore S (1993) 'Chance-constrained Data Envelopment Analysis', *Managerial and Decision Economics*, 14:6, 541-554.

Lee Y H and Schmidt P (1993) 'A Production Frontier Model with Flexible Temporal Variation in Technical Efficiency', in Fried H, Lovell C A K and Schmidt P (eds) *The Measurement of Productive Efficiency*, Oxford: Oxford U P.

Linna M (1998) 'Measuring Hospital Cost Efficiency with Panel Data Models', *Health Economics*, 7(5), 415-27.

Meeusen W and van Den Broeck J (1977) 'Efficiency Estimation for Cobb-Douglas production Functions with Composed Error', *International Economic Review*, 18(2), 435-444 .

Nishimizu M and Page J M (1982) 'Total Factor Productivity Growth, Technological Progress and Technical Efficiency Change: Dimensions of Productivity Change in Yugoslavia 1965-78', *Economic Journal*, 92, 920-936 .

Park B U, Sickles R C and Simar L (1998) 'Stochastic Panel Frontiers: a Semiparametric approach', *Journal of Econometrics*, 84, 273-301.

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

Ray S and Desli E (1997) 'Productivity Growth, Technical Progress and Efficiency Change in Industrialised Countries: a Déjà vu', *American Economic Review*, 87, 1033-1039.

Richmond J (1974) 'Estimating the Efficiency of Production', *International Economic Review*, 15(2), 515-521.

Schmidt P and Sickles R C (1984) 'Production Frontiers and Panel Data', *Journal of Business and Economic Statistics*, 2, 367-374.

Simar L (1992) '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(1/2), 171-191.

Simar L and Wilson P (1998) 'Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models', *Management Science*, 44(11), 49-61.

Simar L and Wilson P (1999) 'Estimating and Bootstrapping Malmquist Indices', *European Journal of Operations Research*, 115, 459-471.

Simar L and Wilson P (2000) 'Statistical Inference in Nonparametric Frontier Models: the State of the Art', *Journal of Productivity Analysis*, 13, 49-78.

Varian H (1985) 'Nonparametric Analysis of Optimising Behaviour with Measurement Error', *Journal of Econometrics*, 30:1/2, 445-458.