ECONOMIC DISCUSSION PAPERS

Efficiency Series Paper 02/2004

Scale and the Efficiency Production Function

Joaquín Millán

Departamento de Economía

Universidad de Oviedo

Available online at: www.uniovi.es/economia/edp.htm

UNIVERSIDAD DE OVIEDO

DEPARTAMENTO DE ECONOMÍA

PERMANENT SEMINAR ON EFFICIENCY AND PRODUCTIVITY

SCALE AND THE EFFICIENCY PRODUCTION FUNCTION[♣]

Joaquín Millán[♠]

Efficiency Series Paper 02/2004

Abstract: This paper introduces the 'efficiency production function' and addresses the question of what kind of efficiency effects can be independently isolated and identified from scale effects in parametric production models. Sato introduced the concept of holotheticity in order to analyze the family of production functions under which the effect of technical progress is completely transformed to apparent scale effects. In this paper, the concept of holotheticity is extended to cover the analysis of technical efficiency. Some by-results concerning the orientation of efficiency are presented.

Keywords: Efficiency production functions, homotheticity, holotheticity, returns to scale, technical efficiency.

[♣] This paper is a revised version of 'Identification of Efficiency and Scale Effects in Parametric Frontier Models' presented at the II Hellenic Workshop on Efficiency and Productivity, in Patras 2003. This research has been funded by grant ALG 2001-3631 of the Spanish DGICYT.

[♠] Dept. of Agricultural Economics, ETSIA, Polytechnic University of Madrid, Spain.

E-mail: jmillan@eco.etsia.upm.es.

1. Introduction

Parametric efficiency analysis requires the specification of a functional form and orientation towards the frontier for the efficiency index. One alternative for this is to model technical efficiency as a parameter and the other is based on assumptions on the statistical structure of the inefficiency elements. Although both approaches are related, this paper mainly discusses the first one because its focus is on the relationship between identifiable efficiency parameters and functional structure. As is argued below, the differences between the meanings of ‘efficiency’ and ‘productivity’ are, to a great extent, conventional. For simplicity, together with its importance in parametric production analysis as in Good, Nadiri and Sickles (1997), only the one-output multiple input-production function is analyzed. Moreover, the approach here is deterministic, except for a few comments on applications.

The analysis of efficiency as an unknown variable has previously been presented in a different way. Griliches (1957) argued that since firms may have different quantities of unobserved ‘managerial ability’, the exclusion of this variable from the analysis can produce biased estimates of the parameters of the model if any of the included explanatory variables is correlated with managerial ability. This ‘managerial ability’, which can be read ‘efficiency parameter’, is considered an omitted variable. Recently, Alvarez and Arias (2003) have used the idea in connection with more modern tools in efficiency analysis. They use an example to illustrate how increasing output with a fixed level of managerial ability can lead to an increase in observed economic inefficiency, and explore the role of managerial ability and output in an empirical translog cost model, focusing on the elasticity of size in milk production.

The present paper differs in approach because efficiency is analyzed in a similar way to that which characterizes technical change. There is more than mere similarity in the formulations of technical change and efficiency production functions. Solow (1957:312) notes that he is "... using the phrase 'technical change' as a shorthand expression for *any kind of shift* in the production function. Thus slowdowns, speedups, improvements in education of the labor force, and all sorts of things appear as 'technical change'." However, the ‘shift’ of the production frontier is the basis of efficiency analysis! Thus, the parametric modeling of efficiency is closely related to the widely used approach to technical change, with technology as a parameter of the production function, formulating the named technical change production function. The technology parameters enter the production function in a specific way. As an example, Solow assumes that technical progress is Hicks-neutral and that the production function is linear homogeneous. The same approach is followed in the efficiency production function, with efficiency as a parameter. Output-oriented and input-oriented efficiency are two ways of modeling neutral (homothetic) efficiency changes.

Sato (1980, 1981) has studied in depth the problems of identifying technical change production functions and scale economies using Lie´s theory of transformation groups. It is argued here that the analysis can also be applied to the efficiency production functions. The problem is non-trivial because technical change (efficiency) and scale economies are not directly observable in applied production economics. The problem of non-identification of technical change and scale parameters occurs as a result of a general structure of transformation groups that Sato defines as "holothetic" production functions.

The economies of scale are related to the functional structure. Constant returns to scale mean the production function is linear homogeneous. The variable economies of scale emerge in many ways. They may generate a homogeneous but non-linear production function (constantscale elasticity). The may generate a homothetic production function with variable scale elasticity. They may even generate a non-homothetic production function. There is a question related to variable returns to scale in addition to the problem of identification of efficiency and scale economies: Are the structure characteristics of the estimated technologies (mainly substitution rates) invariant to the ways technical efficiency is oriented? Alvarez, Arias and Kumbhakar (2003) and Kumbhakar and Tsionas (2003) have addressed recently this question.

Some specific problems in parametric efficiency analysis concerning scale are addressed using the concept of ‘efficiency production function’ and the tools of group transformation theory. This paper focuses on the relationship between the functional structure and invariance, or at least equivalence results, concerning the magnitude and orientation of the efficiency parameter and scale. The presentation uses the results in Sato (1980, 1981) extensively while keeping the mathematics at the minimum. The only requirement is the analysis of the fulfillment by the ‘efficiency production function’ of the properties needed for the application of the group transformation theorems in Sato, concerning ‘technical change production functions’.

The paper is organized as follows. Section 2 presents the efficiency production function and summarizes some results concerning scale and the orientation of efficiency indexes under homotheticity. In section 3, the concept of the ‘holothetic’ production function in Sato (1980) is presented with the main properties concerning technical change and transformation groups then those directly useful for efficiency analysis are explained. Section 4 is devoted to discussing some theoretical and empirical problems in the recent efficiency literature using the results previously presented. Finally, in section 5, some conclusions and suggestions close the paper.

2. The homogeneous efficiency production function

Assume that there are several inputs **x** and one output y that is subject to the following neoclassical production function, smooth (differentiable as pleased) and with the other usual regularity conditions, without imposing homotheticity:

$$y = f(\mathbf{x}, t) \tag{1}$$

where t is an index of the state of the technology.

The expression (1) is the ‘technical change production function’. It is essential, in this form, to specify the way in which technology enters the production function. The usual procedure has been to formulate certain hypotheses concerning the way in which technical progress has affected certain important variables that are derived from the production function, for example in Hicks-neutrality.

This paper presents the ‘efficiency production function’, which can be formulated in an analogous way as:

$$y= f(\mathbf{x}, E) \tag{2}$$

where E denotes an efficiency or ‘management ability’ parameter.

The application of the ‘efficiency production function’ begins by presenting the output-orientation of efficiency. The output-oriented parameter of technical efficiency represents the ratio of observed output to potential output. It was first explicitly employed by Timmer (1971), although a log version of the inverse of the output-oriented parameter of technical efficiency, which can be estimated using linear or quadratic programs, appears in Aigner and Chu (1968). The output-oriented technical efficiency index can be defined as the proportion inwhich a firm can increase output from a given quantity of inputs. This index can be represented as:

$$E^O = \min\left\{\Theta \middle| \left(\mathbf{x}, \frac{y}{\Theta}\right) \in T\right\} \qquad (3)$$

where T represents the technology set and $E^O \leq 1$. In the simple (one output) efficiency production function the parametric formulation is easy. The production function with output-

oriented technical efficiency index is:

$$y = E^O \cdot f(\mathbf{x}) \qquad (4)$$

Note that the idea of the efficiency production function is implicit in the specification of inefficient technologies. Moreover, observe that an upward shift ('super-efficiency' or 'technical progress') can be considered, simply taking a value of E^O greater than one. The effect of output-oriented inefficiency is a downward shift of the production function at the efficient level, keeping the isoquant map invariant.

The consequence of the output transformation is the conservation of the isoquant map, except in the numerical value of the isoquants. This property is called invariance, and it is the main characteristic to be observed in the analysis that follows. A family of curves $f(\mathbf{x}) = K$ is said to be invariant under a group if every transformation of the group transforms each curve into a curve in the family.

Only the output value for the efficiency frontier and the efficiency parameter are needed to formalize the structure of the output transformation. The output oriented transformation from $y=f(\mathbf{x})$ to y' is defined by:

$$TO_E: y' = E^O \cdot y \qquad [\, y=f(\mathbf{x})] \qquad (5)$$

In order to apply the theory of transformation groups, the efficiency transformation TO_E must satisfy the three Lie Group properties presented in Eisenhart (1933) and Sato (1980, 1981):

(A) **Composition:** The result of the successive performance of two transformations T_E and T_F is the same as that of the single transformation $T_{E \cdot F}$.

(B) **Inverse:** The value E of the transformation T_E determines the transformation inverse to that obtained by using the value E^{-1}

(C) **Identity** The value $E_0 = 1$ gives the identical transformation.

When the transformation considered satisfies the Assumptions A, B, and C, then the type of efficiency transformation T_E is said to possess the Lie group properties. It is clear that TO_E satisfies properties A, B and C. Moreover, TO_E is a specific one-variable application of a well-known class of transformations named uniform magnification, perspective, dilatation or homothetic transformations (Eisenhart, 1933:42), resulting in a mere re-scaling of the output levels. Thus, we can present the following invariance Lemma:

Lemma 1. The structure of the technology is invariant to the output oriented efficiency transformation.

A consequence of the previous lemma is the fundamental role of the input-related transformations identifying changes in the isoquant maps. Sato (1981: 21) asks what types of exogenous change will leave a given isoquant map (or factor price frontier) unaltered or invariant. He does not consider the possibility of the output transformation, although it is an even more direct re-labeling of the isoquant map. It is very interesting because he addresses the possibility of the explanation of higher output than expected due to technical change ('output oriented', in efficiency parlance) or to factor accumulation and scale economies('input transformations'). It is remarkable that he explicitly formulates the problem to be analyzed in the following form: "Assume that when exogenous technical progress is introduced, it will not change the form of the production function f, but it will change the output level by affecting the way in which the factor inputs are combined." (Sato, 1981: 22). In fact, Sato develops an input approach using only the functions $x_i=x_i$ ($\mathbf{x}$,t).

It emerges that Sato identifies the output orientation and the input orientation. He is omitting the intermediate step presented here. However, from the property 'if a set of transformations form a group, the transforms of these transformations form a group' [Eisenhart (1933: 18)], the lemma of invariance under output oriented transformations says that the output transformation does belong to the same group of the resulting input transformations. Hence, the conclusions by Sato presented in the next section are right, because they follow directly resulting in the invariance property of the output transformation, and the possibility of an input transformation unidentifiable from an output transformation. When the consequence of a particular input transformation results in the same family (invariance) is the input transformation equivalent to the output transformation.

In this paper, only the homothetic input transformation underlying input-oriented efficiency (and also graph hyperbolic efficiency) measurement is considered. An efficiency production function that assumes that a firm producing a given level of output may be using more inputs than the minimum necessary leads to the input-oriented measure of technical efficiency. The input-oriented measure of technical efficiency can be defined as the maximum equi-proportional reduction in all inputs that still permits production of a given quantity of output. That is:

$$E^I = \min\{\Theta \mid (\Theta\mathbf{x}, y) \in T\} \tag{6}$$

where E^I is the input-oriented index of technical efficiency ($E^I \leq 1$). The input-based measure of efficiency is introduced in Farrell (1957) under constant returns to scale. Färe and Lovell (1978) first distinguished between input-oriented and output-oriented indexes of technical efficiency and showed that they are equivalent under constant returns to scale. Kopp (1981) extends the conceptual interpretation of the input-oriented measure by Farrell to more general returns to scale functional forms.

Other possibilities arise for the index direction choices. An example is the hyperbolic measure introduced in Färe, Grosskopf and Lovell (1985) that simultaneously assumes the proportionate contraction of inputs and expansion of outputs. The hyperbolic index is commonly defined as:

$$E^H = \min\left\{\Theta \middle| \left(\Theta\mathbf{x}, \frac{y}{\Theta}\right) \in T\right\} \tag{7}$$

where E^H is the hyperbolic index of technical efficiency.

The efficiency production function with input-oriented technical efficiency parameter is formulated as:

$$y = f(E^I \cdot x) \tag{8}$$

while the production function with hyperbolic-oriented technical efficiency index is:

$$y = E^H \cdot f(E^H \cdot x) \tag{9}$$

Next, assuming that the production function f(**x**) is homogeneous of degree k and imagine a proportional increase λ in input quantities resulting in a new output y'. The efficiencyparameters (E^O , E^I, E^H) and the degree of returns to scale k are unknown. With output-oriented technical efficiency index, the transformation results in:

$$y' = E^O \cdot f(\lambda \cdot \mathbf{x}) = E^O \cdot \lambda^k \cdot f(\mathbf{x}) = A \cdot f(\mathbf{x}) \tag{10}$$

With input-oriented technical efficiency index, the same scale transformation is:

$$y' = f(E^I \cdot \lambda \cdot \mathbf{x}) = (E^I \cdot \lambda)^k \cdot f(\mathbf{x}) = (E^I)^k \cdot \lambda^k \cdot f(x) = A \cdot f(\mathbf{x}) \tag{11}$$

Finally, with hyperbolic orientation, the transformation is:

$$y' = E^H \cdot f(E^I \cdot \lambda \cdot \mathbf{x}) = E^H \cdot (E^H \cdot \lambda)^k \cdot f(\mathbf{x}) = (E^H)^{1+k} \cdot \lambda^k \cdot f(\mathbf{x}) = A \cdot f(\mathbf{x}) \tag{12}$$

Then, the equivalence

$$E^O = (E^I)^k = (E^H)^{1+k} \tag{13}$$

Atkinson and Cornwell (1994: 247), Greene (1997), and Kumbhakar and Tsionas (2003), using the dual cost function, remark on the equality $E^O = (E^I)^k$. However, they do not consider that, in fact, the parameters E^O, E^I and k are not identified. To quote Greene (1997:113) "…the production function is homogeneous, the effect of the economies of scale can be removed by rescaling the estimated disturbance." If the production function is homogeneous, then the efficiency measure and the scale elasticity cannot be identified[1]. The output-oriented measure is not identified because the only parameter identified is A ($=E^{O}\cdot\lambda^{k}$). Constant returns to scale or linear homogeneity mean k = 1. Using the above equations:

$$E^{O} = E^{I} = (E^{H})^{2} \tag{14}$$

The first equality is the well-known result of equality of the input-oriented and the output-oriented efficiency measures under constant returns to scale (Färe and Lovell, 1978). The second equality suggests that it seems better to give the same value to the hyperbolic measure when input-oriented and output-oriented measures do coincide. Thus, a 'new' definition of graph hyperbolic measures emerges as:

$$E^{H^{*}} = \min\left\{\Theta^{2} \,\middle|\, \left(\Theta\mathbf{x}, \frac{y}{\Theta}\right) \in T\right\} \tag{15}$$

However, the efficiency path is the same for E^{H} in (7) and E^{H*} in (15), because it is a mere transformation of the parameter. The issue of measurement is tackled more directly in Millán (2003).

Turning back to variable returns to scale, under constant elasticity of scale k the parameters (E^{O}, E^{I}, E^{H} or E^{H*}) and k are not identified in (10)-(12). They are not 'essential parameters' (Eisenhart, Sato). Only parameter A can be identified from a particular series of observations of inputs and output if the production function is homogeneous. The above results can be extended to homothetic production functions, those that can be represented by an increasing transformation of a linear homogeneous production function (the transformation results in the same parameter A). Homothetic functions are more easily studied using the dual.

One area of application of the duality theory initiated in Shephard (1953) is homotheticity, because it results in a multiplicatively separable structure of the cost

[1] After giving an interpretetation of input-oriented efficiency valid in terms of cost reduction for non-homothetic production functions and presenting other efficiency indices, Kopp (1981) illustrates by an example the different measures proposed. He uses a Cobb-Douglas with increasing returns to scale. Rather obviously in his text, this function was not estimated. The problem that emerges from the analysis here is how the degree of returns to scale can be estimated. In fact, any arbitrary degree of returns to scale could be imposed.

function. Shephard (1953) defines the homothetic neoclassical production function to be of the form H(f(**x**)) where f(**x**) has the properties of a production function homogeneous of degree one, and H(·) is any positive, strictly increasing (continuous) function with H(0)=0. C(y,**r**) being the neoclassical cost function, the homothetic neoclassical cost function takes the form:

$$C(y,\mathbf{r}) = H^{-1}(y)\cdot u(\mathbf{r}) \tag{16}$$

where u(**r**) is the homogeneous function of degree one in the input price vector **r** indexing the level of prices for the unitary cost function.

The efficiency cost function C(y,**r**,E) is related to the efficiency production function y=f(**x**,E). There are alternative explanations of cost efficiency, as mentioned in Greene (1997:113). The output-oriented efficiency cost function associated with the specification of the efficiency production function in (4) can be expressed as:

$$C(y,\mathbf{r},E^O) = \min_x \{\mathbf{r'}\cdot\mathbf{x} \mid y = E^O\cdot f(\mathbf{x})\} = C(y/E^O,r) \tag{17}$$

In an analogous way, the input-oriented efficiency cost function can be expressed as, using the primal in (8):

$$C(y,\mathbf{r},E^I) = \min_x \{\mathbf{r'}\cdot\mathbf{x} \mid y = f(E^I\cdot\mathbf{x})\} = C(y,\mathbf{r})/E^I \tag{18}$$

The production function y = f(**x**) being homothetic, the output-oriented efficiency homothetic cost function is:

$$C(y,\mathbf{r},E^O) = H^{-1}\cdot(y/E^O)\cdot u(\mathbf{r}) = H^{-1}(y)\cdot u(\mathbf{r})/H(E^O) \tag{19}$$

Analogously, the input-oriented efficiency homothetic cost function is:

$$C(y,\mathbf{r},E^I) = H^{-1}(y)\cdot u(\mathbf{r})/E^I \tag{20}$$

Note that H(·) or $H^{-1}(\cdot)$ are not generally known, and thus cannot be identified from E^O or E^I in (19) and (20). Moreover, when prices are given as is often assumed in cross-sectionanalysis, C=C(y,E), and the situation is the same that under homotheticity[2]. Under constant returns to scale H(·) = 1, and the equality of E^O and E^I is obtained.

[2] It is the usual case when the average cost frontier is estimated, as in Alvarez and Arias (2003).

The more obvious ‘new’ conclusion based on the above results is that neutral efficiency can be identified from scale effects only when the production is non-homothetic. Until now, the result has been presented as a correspondence between input-oriented or output-oriented measures under homotheticity. Some questions about the generalization of the previous results arise naturally. These are whether there are alternative ways of describing efficiency under homogeneity of the production function, or if there is any functional structure concerning returns to scale that can be separately identified from any alternative specification of the efficiency parameter.

3. Holothetic production functions

Sato (1980, 1981) analyzes in depth the relationships between scale and technical change forms using Sophus Lie’s theory of transformation groups. Until now, the presentation has highlighted the analogy between efficiency production function and technical change production function. In fact, if the efficiency transformation satisfies the required properties for a transformation group, then the proof for the technical progress production function in Sato’s work is valid for the efficiency production functions. This type of efficiency may be referred to as a Lie type of efficiency or holothetic efficiency. Holothetic means complete-transforming, and hence no new name is needed.

In the analysis of the technical change production function by Sato, the operation is addition (+), the inverse being (–t) and identity $t_0 = 0$. A similar procedure is possible with the efficiency production function, taking into account that the most common functional forms (Cobb-Douglas and translog) are in logarithmic form. Taking logarithms in E in the transformation T_E gives a new transformation with (+, -E, 0) instead of ($\cdot$, E^{-1}, 1). However, the presentation in levels seems more natural. The input oriented transformation is defined by:

$$TI_E: x_i' = E^I \cdot x_i \qquad [\ y=f(\mathbf{x}')] \qquad (21)$$

It is clear that TI_E satisfies the Lie Group properties A-C of composition, inverse and identity. The transformation in (21) belongs to the class of homothetic transformations or dilatations. Thus, the concepts and results from Sato concerning technical change production functions can be applied to efficiency production functions with input-orientation.

Perhaps the assumptions A, B and C are too restrictive for any kind of efficiency in practice. It is reasonable to think of inefficiencies, such as congestion, which do not satisfy these restrictions. However, there are two important issues here. The first one is that when inefficiencies arise in the non-regular region of a production frontier, as in congestion, the design of a production function without optimal properties is the particular issue to be explained. The second one is that each regular efficiency production function satisfying the Lie Group properties possesses the holothetic property compatible with at least one type of efficiency having these restrictions. The most important result is holotheticity lemma:

Fundamental Lemma of Holotheticity: 'A family of input-oriented efficiency production functions is holothetic under a given type of efficiency (technical progress) if and only if it is invariant under a group.'

Thus, one can always derive at least one type of efficiency change possessing these properties for at least a particular type of production function. Hence, starting with any given family of production functions, one can always derive at least one type of efficiency change possessing these properties. This result is very important because it opens the possibility of studyingdifferent forms of efficiency or 'managerial ability' in the production function in addition to the measurement of ray-inefficiency. However, this issue is not tackled in this paper because a detailed presentation of the infinitesimal transformation, a main tool in Lie's theory, is essential in a full presentation of the 'efficiency production function', but is not needed for the analysis of scale problems.

When the impact of efficiency change on the production function is transformed into a scale effect, the production function is said to be holothetic under a given type of efficiency change. Holotheticity of the input oriented efficiency production function is the confusion of the isoquants under alternative combinations of inputs and input oriented efficiency, thus resulting in the same projections in the isoquant map. Thus, to avoid the identification problem, a production function that is non-holothetic under a given type of efficiency change must be used. The invariance is clearly recognized in the fact that the production function after efficiency change is always a function of the frontier production function. However, this is exactly the condition of the invariance of a family of curves under a group.

Note that input transformations under holotheticity are equivalent to output-oriented efficiency transformations. Given the results in the previous section, we can think of holotheticity of input oriented efficiency transformations and input homotheticity. This is just the result in Sato for Hicks-neutral (ray neutral in the taxonomy by Chambers and Färe (1994)) technical change[3]. Hence the following lemma.

Lemma 2. Radial (homothetic) efficiency transformations cannot be distinguished from scale effects when the production function is input-homothetic.

The non-identification result of homothetic technologies and radial transformations is only a particular case of more general holotheticity results. Given that the graph efficiency transformation is the composition of the input transformation and the output transformation, and given the invariance of the isoquant map in the output oriented component, we present the following lemma, that also applies to non-homothetic production functions:

Lemma 3. The graph efficiency production function belongs to the family of the input oriented production functions.

Summarizing our analysis, and following Sato (1980), given the properties of composition, inverse and identity, three basic theorems concerning holothetic technologies are presented without demonstration -only technical change is substituted for efficiency.

Existence and unity: If the efficiency production functions given by the transformation TE satisfy the Lie group properties then there is one and only one holothetic technology under TE.

[3] It is remarkable how Chambers and Färe apply tools from the efficiency literature to technical change issues, while this paper applies tools and results used for the analysis of technical change to efficiency characterization.

Possibility Theorem of Estimation of Efficiency: The effect of efficiency TE and the scale effect are independently identifiable if and only if the production function is not holothetic under a given type of efficiency TE.

Existence of a Lie Type of Efficiency: Given an isoquant map, there is at least one Lie type of efficiency production function under which the production function is holothetic.

4. Applications

It is remarkable that, in practice, many econometric models are identified under unidentification conditions in this paper. Thus, it is interesting to show that the problem considered in this paper is relevant in practice, and that it is not a mere theoretical nicety. Note that, in this paper, there is no assumption of the error structure or additional structure ofefficiency, which serve in practice as the identification tool. The estimation is possible by putting some separately identifiable structure into the characteristics of the efficiency. Calem (1990) analyzes how estimation of a Cobb-Douglas with variable returns to scale with technical change is possible depending on the stochastic structure of the input evolution. However, a practical problem could occur because of poorly identified models. As an example, there are many estimations of Cobb-Douglas with a constant rate of technical change (time trend) and the usual assumptions of the disturbances in linear regression. Autocorrelation problems are very common, probably due to the untenable assumption of constant growth rates. Calem (1990) also notes how false estimates of technical progress and returns to scale could be obtained when other aspects of the technology are mis-specified. As an example, he shows that a translog specification would measure a decline in economies of scale as a slowdown in the rate of technical change when the Zellner and Revankar (1969) functional form is the true stochastic technology.

A detailed analysis of the analysis of identification in stochastic frontier models is beyond the scope of this paper. Although recognizing the existence of conflicting results, Kumbhakar and Lovell (2000:107) conclude that different estimation methods are "likely to generate similar efficiency rankings, particularly at the top and at the bottom of the distribution, where managerial interest is concentrated." However, a simple numerical example will show that large changes in efficiency rankings, due to the scale-efficiency trade-off, appears in deterministic frontier models depending on the assumptions of the distribution of inefficiency. Aigner and Chu (1968) propose two methods of estimation of parametric production frontiers depending on minimization of the sum of the residuals or the minimization of the sum of the square of the residuals with respect to the efficient frontier. Both methods imply output oriented efficiency measures, and have been given some stochastic structure, such as log-likelihood of exponential or half-normal distributions in Schmidt (1976), although Greene (1980; 1997:94) has emphasized that the procedures violate the conditions for maximum likelihood estimations. The residual ($-u_i$) is the estimate of the logarithm of the parameter E^O_i.

Table 1a in the Appendix presents four observations that can be considered as the result of a 'true' technology y=x, or log(y)=log(x), with some inefficient firms[4]. It is clear that such a small and arbitrary sample and the failure of the statistical foundation of the estimation methods do not allow for valid statistical inferences. However, it is an easy example of the possibility of scale and efficiency trade-offs, involving changing efficiency rankings. A Cobb-Douglas production frontier is estimated using linear and quadratic programs. The estimations are detailed in Tables 1b and 1c. If the one-side sum of residuals ($-u_i = \log(E^O_i)$) is minimized by solving the linear program the estimated frontier is:

$$\log(y) = -0.25 + 1.25 \log(x) \tag{22}$$

There are increasing returns to scale (k=1.25) and, in addition to A, the smaller firm B is efficient. When the estimation method is the minimization of the sum of the square of the residuals the frontier is:

$$\log(y) = 0.056 + 0.944 \log(x) \tag{23}$$

[4] Note that this is, in fact, the frontier supporting observation A, estimated using DEA under constant returns to scale. Two comments are worth making here. 1 This is not to say that DEA is the relevant way of approaching this problem; as an example, only C is inefficient under variable returns to scale, increasing until A, and decreasing afterwards. 2 The analysis in this paper requires differentiability of the production function, hence its results have no direct translation to DEA models.

Now, the only efficient firm is A and there are decreasing returns to scale. The efficiency of the smaller firms B and C is lower and the efficiency of the larger firm D is higher than under the linear programming method.

The scale-efficiency trade-off is clear in the sum of the parameter related to the log-level of the efficiency frontier and the scale elasticity: (-0.25+1.25) = (0.056+0.944)=1. It is remarkable how some distributional assumptions about efficiency are instrumental in the estimation of scale and particular efficiencies. The problem is that there is not a theory of inefficiency to guide the choice of a particular estimation method. There is a plethora of empirical models in the literature and very limited work on selection models[5].

Given the potential relevance of the problem analyzed, there are some immediate practical consequences of the above theorems.

1. The solution to the indeterminacy over the role of returns to scale and efficiency is not to specify non-homothetic production with non-radial efficiency measures (say, Russell-like measures or one-parameter asymmetric efficiency measures). Thus, it is not clear that more general production functions are needed to separate non-neutral efficiency effects. The homogeneous Cobb-Douglas, without restrictions on the degree of returns to scale, is a valid functional form for non-neutral efficiency effects.

2. A perhaps surprising by-result is related to the figures used for illustration of alternative efficiency transformation, because the one-input one-output case is a collapse of the homotheticity property of the radial expansion path (the input axe). The graphs presented in the literature to explain the difference between output-oriented and input-oriented (and graph efficiency) measures are very useful for illustrating the idea, but, in practice, the situations depicted in them cannot be identified separately.

3. The hyperbolic efficiency measure introduced in Färe, Grosskopf and Lovell (1985), that simultaneously assumes the proportionate contraction of inputs and expansion of

[5] Greene (1997:88) remarks on "...a tendency … to equate *estimation technique* with a *model*.". It is worth noting that when distributional assumptions underlying estimation techniques have such deep implications for the economic structure, the consideration of different models is perhaps justified. On the other hand, more theoretical and, more important, empirical work is needed to test the distributional assumptions in stochastic frontier models, following Lee (1983) and Schmidt and Lin (1984).

outputs, belongs to the family of the input-oriented transformations. The isoquant map under the hyperbolic transformation is the same shape as the input oriented component. This means that identification of hyperbolic efficiency transformations and input oriented transformations is, at least, problematic. It is remarkable that the original graph efficiency measure is the square root of a revised definition presented here, in accordance with the structure of input and output changes.

4. Alvarez, Arias and Kumbhakar (2003), on the production function, based on Atkinson and Cornwell (1994) on the cost function, compare the effects of input- and output-oriented efficiencies, and ask whether the input-oriented and output-oriented technologies differ. Alvarez, Arias and Kumbhakar (2003) analyze empirically whether the features of the estimated production technology depend on the choice of the orientation of the technical efficiency measure. The answer is positive unless the technology is homothetic. The output-oriented efficiency production function $f(\mathbf{x},E^{O})$ and the input oriented efficiency production function $f(\mathbf{x},E^{I})$ do not belong to the same transformation group under non-homothetic technologies. It implies that both technologies differ in the shape of the isoquant maps. In principle, this fact allows for an econometric approach to the problem of choice of orientation as a problem of testing non-nested models. However, the consideration about distributional assumptions above remains.

5. The results can be extended to cover both efficiency and technical change production functions. Hicks neutral technical change, neutral efficiency and homothetic production functions are holothetic. This means that compositions of any of them cannot be identified separately. A more complete analysis of the conditions for separate identification of technical change and efficiency is needed, applying the tools in this paper to the issues in Chapter 8 of Kumbhakar and Lovell (2001). The difficulties of discriminating among models of efficiency and technical change has been reported in the literature, as in Kumbhakar, Heshmati and Hjalmarsson (1997). The theoretical analysis could help the empirical practice.

5. Conclusions

This paper introduces the concept of 'efficiency production function', based on both the 'managerial production function' (Griliches, 1957) and the 'technical change production function' (Sato, 1981). The efficiency production function is useful in the characterization of the identification of scale and efficiency effects, using the theory of transformation groups.

In order to identify both the scale and the efficiency effects empirically, one must know what type of efficiency characterization is completely separable from the scale effect. When the impact of efficiency on the production function is transformed into a scale effect, the production function is said to be holothetic under a given type of efficiency change. Thus, to avoid the problem a production function being non-holothetic under a given type of efficiency transformation is required. As a by-result, it is shown that the graph efficiency transformation belongs to the class of input-oriented efficiency transformations in the multiple input simple production function.

On the other hand, when non-neutral efficiency production functions are considered, a homothetic or even a homogeneous production function (such as Cobb-Douglas) can be used. On the contrary, more complicated non-homothetic production functions can be holothetic under non-neutral efficiency. Non-neutrality of efficiency can be formulated in the parameters (Russell measures, as an example) or in a stochastic structure.

The above results are deterministic. When a stochastic structure is introduced, the new parameters, such as means and variances, could allow for identification of different production structures. However, examples from the empirical literature suggest that, in practice, identification and model selection can be problematic. The joint analysis of functional structure and distributional assumptions is worth further research.

References

Aigner, D.J. and S. F. Chu (1968), "On estimating the industry production function", *American Economic Review*, 58, 826-39.

Alvarez, A., and C. Arias (2003), "Diseconomies of size with fixed managerial ability", *American Journal of Agricultural Economics*, 85, 134-42.

Alvarez, A., C. Arias and S. C. Kumbhakar (2003), "Empirical consequences of direction choice for technical efficiency analysis" *Efficiency Series Paper 02/2003*. University of Oviedo.

Atkinson, S.E. and C. Cornwell (1994), "Estimation of Output and Input Technical Efficiency Using a Flexible Functional Form and Panel Data", *International Economic Review*, 35, 245-256.

Calem, P. S. (1990), "On Estimating Technical Progress and Returns to Scale". 261-274, in R. Sato and R. V. Ramachandran (eds.): *Conservations Laws and Symmetry*. Kluwer Academic Publishers.

Chambers, R.G. and R. Färe (1994), "Hicks' Neutrality and Trade Biased Growth: A Taxonomy", *Journal of Economic Theory*, 64, 554-567.

Eisenhart, L. P. (1933), *Continuous Groups of Transformations*. Princeton University Press. (Reprinted in Dover, 1961).

Färe, R. and C.A.K. Lovell, (1978), "Measuring the Technical Efficiency of Production", *Journal of Economic Theory*, 19: 150-162.

Färe, R., S. Grosskopf and C.A.K. Lovell (1985), *The Measurement of Efficiency of Production*, Kluwer-Nijhoff Publishing.

Farrell, M.J. (1957), "The Measurement of Productive Efficiency" *Journal of the Royal Statistical Society, Series A*, 120, 253-81.

Good, D.H., M.I. Nadiri and R. C. Sickles (1997) "Index Numbers and Factor Demand Approaches to the Measurement of Productivity". 14-80, in M.H. Pesaran and P. Schmidt (eds.): *Handbook of Applied Econometrics. Volume II: Microeconomics*. Blackwell Publishers.

Greene, W. H. (1997), "Frontier production functions" 81-166, in M.H. Pesaran and P. Schmidt, eds. *Handbook of Applied Econometrics. Volume II: Microeconomics*. Blackwell Publishers.

Greene, W. H., (1980), "Maximum Likelihood Estimation of Econometrics Frontier Functions", *Journal of Econometrics*, 13, 101-115.

Griliches, Z. (1957), "Specification Bias in Estimates of Production Functions", *Journal of Farm Economics*, 39, 8-20.

Kopp, R.J. (1981), "The Measurement of Productive Efficiency: A Reconsideration", *Quarterly Journal of Economics*, 477-503.

Kumbhakar, S.C., A. Heshmati and L. Hjalmarsson (1997), "Temporal patterns of technicalefficiency: Results from competing models". *International Journal of Industrial Organization* 15, 597-616.

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

Kumbhakar, S.C. and E.G. Tsionas (2003), "Recent developments is stochastic frontier modeling". *Efficiency Series Paper 06/2003*. University of Oviedo.

Lee, L. (1983), "A Test for Distributional Assumptions for the Stochastic Frontier Function," *Journal of Econometrics*, 22, 245-267.

Millán, J.A. (2003), "On graph efficiency measurement". *Efficiency Series Paper 10/2003*. University of Oviedo.

Sato, R. (1980), “The impact of technical change on the holotheticity of the production functions” *Review of Economic Studie*s, 47, 767-776.

Sato, R. (1981), *Theory of Technical Change and Economic Invariance: Application of Lie Group*s. Academic Press.

Schmidt, P. (1976), “On the Statistical Estimation of Parametric Frontier Production Functions”. *Review of Economics and Statistic*s, 58, 238-239.

Schmidt, P. and T. Lin (1984), “ Simple Tests of Alternative Specifications in Stochastic Frontier Models”. *Journal of Econometric*s, 24, 349-361.

Shephard, R. W. (1953), *Cost and Production Functions.* Princeton University Press. (Reprinted in Springer-Verlag, 1981).

Solow, R.M. (1957), “Technological Change and Aggregate Production Function”, *Review of Economics and Statistic*s, 39: 312-320.

Timmer, C.P. (1971), “Using a Probabilistic Frontier Production Function to Measure Technical Efficiency”, *Journal of Political Econom*y, 79, 776-94.

Zellner, A. and N. S. Revankar (1969), “Generalized Production Functions”. *Review of Economic Studie*s, 36, 241-250.

Appendix: Table 1. Example

1a. Data

Firm	Output: y	Input: x	log(y)	log(x)
A	2.7183	2.7183	1.00	1.00
B	2.1170	2.2255	0.75	0.80
C	2.3396	2.4596	0.85	0.90
D	3.0042	3.3201	1.10	1.20

1b. Linear program

$\min \Sigma_i u_i$ $[u_i = a+b\cdot\log(x_i)-\log(y_i)]$

s.t. $a+b\cdot\log(x_i)^3\log(y_i)$ I = A, B, C, D

Estimated (log)efficiency production function

$\log(y_i) = -0.25 + 1.25 \log(x_i) - u_i$

Firm	Estimated log(y)	Residual u	Frontier Output y’	Efficiency E^O
A	1.0000	0.0000	2.7183	1.0000
B	0.7500	0.0000	2.1170	1.0000
C	0.8750	0.0250	2.3989	0.9753
D	1.2500	0.1500	3.4903	0.8607

1c. Quadratic program

$\min \Sigma_i (u_i)^2$ $[u_i=a+b\cdot\log(x_i)-\log(y_i)]$

s.t. $a+b\cdot\log(x_i)^3\log(y_i)$ i=A, B, C, D

Estimated (log)efficiency production function

$\log(y_i) = 0.0556 + 0.9444 \log(x_i) - u_i$

Firm	Estimated log(y)	Residual u	Frontier Output y’	Efficiency E^O
A	1.0000	0.0000	2.7183	1.0000
B	0.8111	0.0611	2.2504	0.9407
C	0.9056	0.0556	2.4733	0.9460
D	1.1889	0.0889	3.2834	0.9150