European Journal of Operational Research 220 (2012) 571–576

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

# Measuring returns to scale in *DEA* models when the firm is regulated

Pierre Ouellette [a], Jean-Patrice Quesnel [a], Stéphane Vigeant [b,c,*]

[a] Department of Economics, Université du Québec à Montréal, P.O. Box 8888, Station Centre-Ville, Montréal, Québec, Canada H3C 3P8
[b] Laboratoire ÉQUIPPE, Université de Lille 1, 59655 Villeneuve d'Ascq Cedex, France
[c] IÉSEG School of Management, 3 rue de la Digue, 59000 Lille, France

ARTICLE INFO

*Article history:*
Received 20 July 2010
Accepted 24 January 2012
Available online 31 January 2012

*Keywords:*
Data envelopment analysis
Regulation
Returns to scale
Quasi-fixed inputs

ABSTRACT

In the standard framework of data envelopment analysis (DEA) models, the returns to scale are fully characterized using the multiplier on the convexity constraint of inefficient decision making units (DMU) using the projection of the input–output vector on the frontier. In this note, we investigate how the returns to scale measurements in DEA models are affected by the presence of regulatory constraints. These additional constraints change the role played by the convexity constraint. In order to avoid biased estimation of the returns to scale, we show that the interaction between the regulatory and the convexity constraints has to be taken into account.



## 1. Introduction

One of the key properties of production structure is the scale of operations. It receives particular interest because the existence of economies or diseconomies of scale may have different implications for the market structure and conduct. Examples abound to illustrate their importance. Increasing returns to scale in an industry may lead to its monopolization while the adoption and application of trade measures by governments to promote openness may rest on the maturity of the industry with respect to its scale. Adjustment of firms to their optimal size is slow due to quasi-fixed inputs, introducing a new complication as firms cannot choose the optimal level of some inputs instantaneously. This fact must also be accounted for by the scale measure used to characterize the firms' technology. While the previous examples are concerned with the private sector firms, returns to scale measures are also relevant for public sector firms. For example, understanding scale relationships facilitate the work when searching for the optimal size of hospitals and schools in an environment where market signals are of little or no help. A common denominator to these examples is the presence of government intervention and this intervention plays a crucial role. Governments usually respond to scale signals with regulations, price control and the likes. Furthermore, these policies and regulations are often already part of the environment of the firms and already integrated in the production process, irrespective of the scale considerations. Consequently, quality returns to scale measurements are essential because of the role they play in the design of policies. Those measurements must include the regulation when it is already a component of the environment of the firms. Consequently, returns to scale measurement must be adapted to the proper circumstances. The focus of this study is on returns to scale measurements for DEA models with a piecewise linear frontier when the decision making units (DMU) face a complex environment that includes regulation and quasi-fixed inputs. Returns to scale can be measured with different methodologies from DEA to regression methods, including stochastic frontier models. Banker and Natarajan (2008) have shown, however that DEA models perform as well as the other methods if not better. For this reason we believe that improving on RTS measurement using DEA methods is a crucial issue.

Attempts to understand the relationship between the technology of a *DMU*, its efficiency and environment, and the measurement of returns to scale (*RTS*) are not new to *DEA*. Banker et al. (1984) and Banker and Thrall (1992) have shown that *DEA* programs need an additional constraint in order to obtain variable *RTS*. The core of the *BCC* model is a convex combination of the DMUs used to infer *RTS*. When this constraint is not included, as it is the case in the standard CCR model, *RTS* are constant. The foundation of *RTS* measurements is to exploit the structure of the production sets to deduce the nature of the economies of scale. The focus is then on scale efficiency, leading to qualitative information on the economies of scale (e.g. Banker et al., 1984). Another approach computes directly the elasticity of scale from efficiency measures and the multiplier on the convexity constraint. Examples of this approach are Banker et al. (1984), Banker and Thrall (1992), Førsund (1996) and Førsund and Hjalmarsson (2004). An important feature of this framework is its natural compatibility with

* Corresponding author. Address: Faculté des Sciences Économiques et Sociales, Université de Lille 1: Sciences et technologies, 59655 Villeneuve d'Ascq Cedex, France. Tel.: +33 (0) 3 20 33 63 56; fax: +33 (0) 3 20 43 66 55.
*E-mail address:* Stephane.Vigeant@univ-lille1.fr (S. Vigeant).

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2012.01.048

multi-output productions. The methodology consists in projecting the inefficient *DMU* on the frontier and then calculating the *RTS* at this projection point. It is necessary to use the projection on the frontier because *RTS* is a meaningless concept for inefficient firms. There is another important problem, this time with efficient *DMU*s. Because they are either on vertices or on ridges, the multiplier associated to the convexity constraint can take on multiple values and consequently the *RTS* are not uniquely determined for these firms. Authors working on *RTS* have all suggested particular ways to calculate boundaries on the value that *RTS* can assume when multiple solutions exist.

These approaches are the starting point to study the measurement of returns to scale. The point we raise in this paper is that the institutional framework used to calculate the *RTS* in these approaches is too simple, considering the environment usually faced by firms. In a more realistic environment of the *DMUs*, not all inputs are fully discretionary and the environment in which they operate is regulated. We show here how to introduce these refinements of the firm's environment into the calculation of the *RTS*. Quasi-fixed inputs (non-discretionary inputs) are introduced and we show that *RTS* calculations are not substantively affected by this change as the role played by the convexity constraint in the *BCC* model remains unchanged. Introducing regulation has important consequences, however. The *RTS* formula is different than the one used in the standard case. Furthermore, the constraints in the *CCR* model have to be modified in order to impose constant *RTS*. In fact, some constraints must be combined to form a new set of constraints in order to achieve constant *RTS*.

## 2. The standard approach to returns to scale measurement

The calculation of *RTS* or the elasticity of scale amounts to measuring a relationship between inputs and outputs in a production structure. There are many approaches and methodologies to measure *RTS*, from local to global measures, in the primal space (input–output) or the dual space (prices). The focus here is on local calculations of *RTS* in the primal space. In order to characterize the various *RTS* measures considered in this paper, we suppose that firms produce $m$ outputs, $y = (y_1,\ldots,y_m)$, using $n$ variable inputs, $x = (x_1,\ldots,x_n)$, with a technology given by a twice differentiable transformation function, $f(y,x) \leqslant 0$ with $\partial f(y,x)/\partial y_i \geqslant 0$ for $i = 1,\ldots,m$ and $\partial f(y,x)/\partial x_j \leqslant 0$ for $j = 1,\ldots,n$. To find *RTS*, we have to measure the required change in outputs to keep the transformation function equal to zero when inputs are increased proportionally. In other words, we let the inputs be expanded proportionally along a ray by a constant factor, $\mu$, and then we find the factor, $\gamma$, by which we multiply the outputs so that $f(\gamma y, \mu x) = 0$. Obviously, $\gamma$ depends on the reference point $(y,x)$ and the input scaling factor, $\mu$. For a local measure, we take the ratio of differentials[1]:

$$RTS = \left.\frac{d\ln\gamma}{d\ln\mu}\right|_{\gamma=\mu=1} = \left.\frac{d\gamma}{d\mu}\right|_{\gamma=\mu=1} = -\frac{\sum_{i=1}^{n} f_{x_i} x_i}{\sum_{j=1}^{m} f_{y_j} y_j}.$$

The elasticity of scale is clearly a concept that makes sense only for points on the frontier, that is for $f(y,x) = 0$. Otherwise, there is no clear meaning to a change in output following a ray-variation in inputs. When $f(y,x) < 0$, the change in $y$ can be attributed to either RTS or efficiency or both, with no means to discriminate between cases. It does not mean however, that points inside the frontier (inefficient observation) do not carry valuable information on *RTS*. Førsund and Hjalmarsson (2004) addressed the *RTS* measurement problem when decision units are not efficient. They did so using the *DEA* method. As mentioned above, the equation for *RTS* is meaningless for an inefficient unit *at* the observed point, but it does not prevent us from measuring RTS using this point, however. The trick is to use the projection of the unit on the efficient frontier, instead of the point itself. This procedure does not characterize the *RTS* of the unit, but those at the projected point. To implement this procedure we use the equivalence between the marginal productivities and the dual variables in a *DEA* program. The input-oriented model to compute technical efficiency for unit $h$ is given by the following problem[2]:

$$\min_{\theta^h}\{\theta^h : f(y, \theta^h x) \leqslant 0\}. \tag{1}$$

The Lagrangian of this problem is:

$$L_h^{TE} = \theta^h + \varphi f(\theta^h x, y), \tag{2}$$

where $\varphi$ is the Lagrange multiplier and $\theta^h$ is Farrell's (1957) input-oriented technical efficiency measure. Unless we add some structure to the technology, this problem has no empirical content. To go along this path, the usual procedure consists in approximating the production set by a convex hull of the data. The favored procedure uses a linear program, so in order to establish a connection between the "true" problem and the empirical approximation we use a first-order Taylor approximation of $L^{TE}$ around $(x^0,y^0)$, to obtain:

$$L^{TE} \approx \theta^h + \varphi f(x^0,y^0) + \varphi \sum_{j=1}^{m} f_{y_j}(x^0,y^0)\left(y_{jh} - y_{jh}^0\right) + \varphi \sum_{i=1}^{n} f_{x_i}(x^0,y^0)\left(x_{ih} - x_{ih}^0\right). \tag{3}$$

This equation is not implementable in practice because $f$ is unknown. To estimate the unknown technology we use a DEA approach.

Suppose that some form of free disposal (for both outputs and inputs) is satisfied and that the production set is convex. Then the problem given by Eq. (1) for firm $h$ can be approximated in practice by:

$$\min_{\theta^h}\left\{\theta^h : \sum_{d=1}^{D} \lambda_d y_{jd} \geqslant y_{jh},\ \forall j = 1,\ldots,m; \sum_{d=1}^{D} \lambda_d x_{id} \leqslant \theta^h x_{ih},\ \forall i = 1,\ldots,n; \sum_{d=1}^{D} \lambda_d = 1, \lambda_d \geqslant 0,\ \forall d = 1,\ldots,D\right\},$$

where $D$ is the number of units or firms included in the optimization problem. The convexity constraint on the combination of the units, i.e. $\sum_{d=1}^{D} \lambda_d = 1$, creates the smallest convex envelop of data and gives the *BCC* model. The Lagrangian corresponding to this problem, that is the empirical equivalent of $L^{TE}$, is:

$$L^{DEA} = \theta^h + \sum_{j=1}^{m} v_j^y \left[y_{jh} - \sum_{d=1}^{D} \lambda_d y_{jd}\right] + \sum_{i=1}^{n} v_i^x \left[\sum_{d=1}^{D} \lambda_d x_{id} - \theta^h x_{ih}\right] + v_0\left[1 - \sum_{d=1}^{D} \lambda_d\right].$$

Now, a correspondence between the terms in the linearized version of $L^{TE}$ (i.e. Eq. (3)) and the terms in $L^{DEA}$ can be established as follows: $f_{y_j} \approx v_j^y/\varphi$ and $f_{x_i} \approx -\theta^h v_i^x/\varphi$.[3] These equations relate the dual variables in the *BCC* model to the partial derivatives of the

[1] The conventional RTS measures presented here have been developed and used by many authors, in particular Nadiri (1993), Baumol et al. (1982) among others. For a complete and thorough discussion of RTS measurement, the reader is referred to Førsund and Hjalmarsson (2004).

[2] Every result presented here have an equivalent counterpart using the output-oriented model, i.e. $\max_{\theta^o}\{\theta^o : f(x, \theta^o y) \leqslant 0\}$, as it is done in Førsund and Hjalmarsson (2004).

[3] The approximation follows from the fact that we have used a first-order Taylor approximation of the production function in Eq. (2) to get Eq. (3). Consequently, the marginal productivities are also related to the Lagrange multipliers of $L^{DEA}$ through the same Taylor approximation. For a detailed derivation of this equivalence, see Ouellette and Vigeant (2011).