Journal of Quantitative Economics, Vol. 12, No.1, January 2014

NEOCLASSICAL CHARACTERIZATION OF RETURNS TO SCALE IN NONPARAMETRIC PRODUCTION ANALYSIS

BIRESH K. SAHOO[1]
JATI K. SENGUPTA[2]

Abstract

From an empirical perspective, this paper critically examines the concepts of returns to scale (RTS) and economies of scale (EOS), and argues that the concept of EOS is more relevant and broader enough to exhibit proper scale economies behavior of firms. Two approaches, i.e., production function and total cost function are available in the literature that are generally taken to be a satisfactory way of empirically verifying scale economies behavior of firms without mentioning whether they are taken exhibit the same causal factors. As argued in this paper that because of the fundamental shortcomings associated with the production function approach where output is narrowly related to the inputs only by defining the input-mix in a special way, the concept of RTS seems to be very restrictive in exhibiting any relevant scale behavior. By exercising prudence in pending more thorough exploration of the sources of scale increases; a strong case is made in this paper supporting the use of a nonparametric cost frontier approach to estimate scale economies behavior.

Keywords: Returns to Scale, Economies of Scale, Data Envelopment Analysis

JEL Classification: C14, C61, D21, D24

1. Introduction

The information on the 'scale economies' behavior of real-life firms can provide important insights not only to firm managers making operational decisions in strengthening their competitive position, but also to policy makers debating on regulatory issues concerning the restructuring of any specific-sector in the economy. Therefore, this concept needs to be examined with more prudence in the face of empirical realities.

In the literature, the concept of either returns to scale (RTS) or economies of scale (EOS) is generally taken to be a satisfactory way treating scale economies without bothering much about whether they are taken to exhibit the same causal factors. The relationship between the two may not be that quite clear-cut as those often claimed in the literature. Much of earlier literature, in fact, treats RTS as synonymous with EOS indicating local point-wise equivalence

[1] Xavier Institute of Management, Xavier Square, Bhunaneswar 751 013, India, e-mail: biresh@ximb.ac.in
[2] Department of Economics, University of California, Santa Barbara, CA 93106, USA

between the two for any cost minimizing firm. That is, any point on the firm's long-run average cost curve exhibiting local economies (or diseconomies) of scale is associated with a corresponding point on the underlying production function exhibiting local increasing (decreasing) RTS. This may be true under two conditions: 1) production structure is *homothetic*; 2) input prices are *exogenous*.

To understand the nature of relationship between RTS and EOS, consider an example of a *non-homothetic* technology where the assumption of *constant factor proportion* is violated. If an activity involving one man with one shovel yields one acre of ploughed land, then the activity involving 10,000 men with 10,000 shovels should yield 10,000 acres of ploughed land. This augmented activity is, however, quite possible, but may not be efficient, and hence, may not be a point on the production frontier. The augmented activity might not necessarily involve 10,000 shovels but may be one tractor if the capital embodied in the tractor is that of 10,000 shovels. Similarly, a few skilled labors may be substituted for 10,000 men. This new input combination could possibly yield more than 10,000 acres of ploughed land, a case of increasing returns to scale.

In any production process, capital can be augmented in two ways: one by mere 'replication' and other by 'reconfiguration'. Replication means increases of original capital good by integer multiples (creating units identical to those already in use) whereas reconfiguration of capital means altering its physical specifications. In response to the need to alter the rate of output, a differently constituted machine (tractor, say) may be used. However, in case of mere replication, the production function will always exhibit constant returns to scale. It can, therefore, be argued that the use of the concept of RTS to explaining shape of LRAC curve (i.e., economies/diseconomies of scale), a practice usually followed in the literature, can be potentially misleading.

Second, the local equivalence between RTS and EOS may not hold when factor prices are endogenous, because there are pecuniary effects associated with the firm's employment of factors that might be sufficient enough to cause economies (or diseconomies) of scale even though firm's technology exhibits decreasing (or increasing) returns to scale.

Researchers involved in the empirical estimation of scale economies, generally, uses either a production function approach to examine RTS, or a total cost function approach to test for EOS. One can, however, argue that the shortcomings of the former are too strong enough to measure any relevant scale effects. This is because output is related to inputs only by defining the input-mix in a special way, e.g., as a replication measure, as a size measure, or as a long-run measure of only one input such as plant and machinery or capital. The replication measure is purely statistical in nature, often, used in statistical theory of design of experiments. But this lacks any economic meaning as our tractor-shovel example shows, the techniques and inputs used at higher scale are very different from those used at lower scale. The size measure in terms of inputs is not unequivocal. In agricultural economics it is natural to measure size by acres, but in industrial manufacturing there is no such natural measure. The measure of plant and machinery is difficult due to heterogeneity except through costs. However, if the current input mix can be represented by a size measure, then the 'size' elasticity of output is a good measure of RTS. So is for the plant and machinery measure. Therefore, in the light of the aforementioned problems, the cost measure scores well over the output-based measure to estimate scale economies, i.e., EOS.

The nonparametric data envelopment analysis (DEA) cost approach has been very popular in estimating the productive efficiency of firms (Sengupta 2004; Tone and Sahoo 2005, 2006; Sahoo and Tone 2013; Sahoo *et al*. 2014). Note that Färe and Grosskopf (1985) are the first to introduce the cost frontier approach as a way of estimating scale efficiency. Our contribution will be viewed as an extension of their work in measuring EOS.

The remainder of this paper proceeds as follows. Section 2 first deals with the discussion on the relevance of RTS vis-à-vis EOS in terms of their underlying causal factors, and then argues for the latter, which can be estimated using cost-based scale elasticity (SE) measure. Section 3 identifies and sorts through competing accounts of EOS. Section 4 deals with the discussion on the relevant nonparametric methods for estimating scale economies followed by some concluding remarks in Section 5.

2. Relevance of RTS vis-à-vis EOS

The broader sources of scale economies, which have long been recognized by the Classicists, are in the form of cost advantages due to *division of labor* by Adam Smith, *effect of cooperation and team work* by Karl Marx, technological improvement by Alfred Marshall and *technical and managerial improvement* by Clark and Robinson (Tone and Sahoo 2003, p.167). It can be observed that the benefits of expansion that flow from the many diverse components of a 'firm' are based not only on technology, but more on the entire gamut of organization, learning by doing, reorganization of inputs and the other capabilities of the firm.

The most often cited explanation for EOS is due to indivisibilities of inputs such as specialized equipments that are available in certain discrete sizes only; and if production is carried out at levels that are not at the designed optimum capacity levels, then average costs would be higher, which would essentially mean that there would be a fall in the average costs if output were expanded.

For a continuous and smoothly declining long-run average cost curve, it is usually assumed that *plant possibilities* are numerous. But the explanation of scale is by considering the indivisibility of technique associated with certain plant size. The question then remains as to whether this treatment of scale will be in confirmation to constant factor proportions. Reflection upon the physical nature of the real-life production processes suggests that the assumption of constant factor proportion is hardly met.

Another way in which scale effects may emerge is to consider the use of capital equipments, which have the characteristics of incorporating less *capital* than its contribution to *capacity*. Physical capital equipments in the form of cylinders, pipes, vessels, etc., all exhibit the well-known engineers' 0.6 rule of thumb. Such scale effects are purely due to the physical properties of materials, and should be treated as *natural* sources of scale.

One can, therefore, argue that the concept of EOS is more appropriate to describe the scale economies behavior in several situations, e.g., (a) when input proportions change in the long run, the firm's expansion path is no longer a straight line and the concept of RTS no longer applies. EOS includes IRS as a special case but it is more general because it reflects varying input proportions over output, (b) when inputs are heterogeneous, (c) cost is more in line with the accounting view of firm's balance sheet, where 'cost of goods sold' is usually used to compute net income or net receipts, and (d) cost can be more easily related to other concepts like EOS and size-specific costs due to indivisibility.

In the immediately following section we sort through competing accounts of EOS.

3. Sources of economies of scale

Four broad major sources of EOS can be easily distinguished: (a) economies of scope (b) indivisibility comprising size, (c) learning by doing through cumulative experience, and (d) reduced input costs due to power over suppliers.

Economies of scope (ES) are said to exist if the cost of producing goods in combination by a diversified firm is less than the cost of producing them in isolation by the specialized firms. Let the cost function of a diversified firm producing two outputs be $C(y_1,y_2)$ and the cost of functions of two specialized firms producing outputs y_1 and y_2 be $C(y_1,0)$ and $C(0,y_2)$ respectively. ES are said to exist if $C(y_1,y_2) < C(y_1,0) + C(0,y_2)$ where the degree of ES (*DES*) of any firm *h* is defined as

$$DES_h = \frac{C(y_1,0) + C(0,y_2) - C(y_1,y_2)}{C(y_1,y_2)} \qquad \text{... (1)}$$

$DES_h > 0$ implies that firm h exhibits economies of scope, $DES_h < 0$ implies diseconomies of scope, and $DES_h = 0$ implies that the cost function $C(y_1,y_2)$ is additive in nature.

Now, turning to the case of size-based cost, let the long-run cost function of firm *h* be:

$$c_h = as_h^{\alpha};\ \ a > 0,\ \ 0 < \alpha < 1 \qquad \text{... (2)}$$

where s_h is the scale of the plant *h* of a given size measured by *capacity output*. Here doubling the size causes long-run costs c_h to be less than double. Here, the *scale elasticity* of firm $h(\varepsilon_h)$, defined as the ratio of average cost (AC_h) over marginal cost (MC_h), is

$$\varepsilon_h = \frac{1}{\alpha} > 1 \text{ since } 0 < \alpha < 1.$$

The third source of EOS is derived from the log-linear relation of *cumulative costs* of firm *h* (K_h) and *cumulative output* (Z_h):

$$K_h = aZ_h^{\alpha};\ \ 0 < \alpha < 1, a > 0 \qquad \text{... (3)}$$

Here, as the cumulative experience represented by output Z_h rises, cumulative costs represented by K_h declines. Note that we need not assume here that current output doubles or increases threefold. Current cost is then dK_h/dt and current output dZ_h/dt if the costs and output can vary continuously.

Finally, the fourth source of EOS is the reduced input costs that are due to power over suppliers. Pecuniary economies in the firms' employment of inputs such as monopsonistic employment of labor and monopsonistic purchase of specialized inputs explain the downward slopping average cost behavior of firms.

4. Empirical estimation of scale economies

Let us consider the following cost efficiency model:

$$C(y;w) = \min_x \sum_{i=1}^{m} w_i x_i \text{ subject to } f(x_1, x_2, ..., x_m) \geq y^0,\ x_i \geq 0\ (\forall i) \qquad ... (4)$$

where the production function takes a Cobb-Douglas form, i.e., $f(x_1, x_2, ..., x_m) = A\prod_{i=1}^{m} x_i^{b_i}$.

For any fixed output $y^0 > 0$, one may derive the cost frontier as

$$\ln c^* = k_0 + \sum_{i=1}^{m} k_i \ln w_i + a \ln y^* \qquad ... (5)$$

where $y^* = f(x_1, x_2, ..., x_m) = A\prod_{i=1}^{m} x_i^{b_i}$, $k_i = b_i / \sum_{i=1}^{m} b_i$ and $a = \left(\sum_{i=1}^{m} b_i\right)^{-1}$ and k_0 is a suitable constant depending on *A* and $b_1, b_2, ..., b_m$. If 0 < *a* < 1, then we have IRS. This is equivalent to EOS by duality. Charnes *et al*. (1991) have generalized this formulation further to include more generalized production functions.

Sueyoshi (1997) is the first who measured 'scale economies' (scale elasticity) in terms of EOS using the scheme (4) in the following cost efficiency DEA model for firm *h*:

$$C(y_h; w_h) = \min_{x,\lambda} \sum_{i=1}^{m} w_{ih} x_i$$

$$\text{subject to } \sum_{j=1}^{n} x_{ij}\lambda_j \leq x_i (\forall\ i),\ \sum_{j=1}^{n} y_{rj}\lambda_j \geq y_{rh} (\forall\ r),\ \sum_{j=1}^{n} \lambda_j = 1,\ \ \lambda_j \geq 0 \qquad ... (6)$$

Here $C(y_h; w_h)$ represents firm *h*'s *minimum* cost to produce y_h with the input price vector w_h. Firm *h*'s cost efficiency (CE_h), defined as the ratio of *minimum* cost to *actual* cost, is

$$CE_h = C(y_h; w_h)/c_h = \sum_{i=1}^{m} w_{ih} x_i^* \Big/ \sum_{i=1}^{m} w_{ih} x_{ih} \qquad ... (7)$$

where x_i^* is the optimal solution obtained from (6). The following dual program of (6) can be used to compute the scale elasticity of firm *h* (ε_h) as follows:

$$\begin{aligned} C(y_h; w_h) \ &= \ \max_{u,\omega} \sum_{r=1}^{s} u_r y_{rh} + \omega_h \\ \text{subject to } &\sum_{r=1}^{s} u_r y_{rj} + \omega_h \leq c_h,\ (\forall j),\ u_r \geq 0\ (\forall r),\ \omega_h : \text{ free} \end{aligned} \qquad ... (8)$$

If firm h is efficient, the 'minimum' and 'actual' costs are the same, and it also holds that

$$c_h = C(y_h; w_h) \ = \ \sum_{r=1}^{s} u_r^* y_{rh} + \omega_h^* \qquad ... (9)$$

Following Baumol *et al*. (1982), ε_h can be obtained as

$$\varepsilon_h(y;w) = \frac{C(y_h;w_h)}{\sum_{r=1}^{s} y_{rh} \frac{\partial C(y_h;w_h)}{\partial y_{rh}}} = \frac{C(y_h;w_h)}{\sum_{r=1}^{s} u_r^* y_{rh}} = \frac{1}{1-[\omega_h^* / C(y_h;w_h)]} \quad \dots (10)$$

There are economies of scale if $\omega_h^* > 0$, no economies/diseconomies if $\omega_h^* = 0$, and diseconomies of scale if $\omega_h^* < 0$. Note that Sueyoshi (1997) has implicitly presumed throughout scale elasticity in terms of RTS only (but not EOS).

The CE model (6) can be of limited use in actual applications as this model is based on a number of simplifying assumptions. First, not only factor inputs are homogeneous, but also that their prices are exogenous. As a result, the *SE*s in both production and cost environments are equal, thus giving the illusion that RTS and EOS are the one and same.

With an expansion in production, firms experience *changes* in the organization of their processes or in the characteristics of their inputs that are economically more attractive than the replicated alternatives of those already in use. Therefore, the technique and inputs used at higher scale are very different from those used at lower scale. Hence, the inputs are heterogeneous, and as a result of which, their prices may vary across firms. Since the input resources vary in their quality, the construction of *technology* in (6) becomes problematic.

Input prices are also not exogenous, but they vary according to actions by the firms. Firms often face *ex ante* price uncertainty while making production decisions. Economic theory suggests that firms enjoying some degree of monopoly power should charge different prices if there is productivity heterogeneity in their inputs. This is empirically valid since most firms face upward slopping supply curve in their input purchase decisions. This observation also suggests that the assumption of common prices by firms, i.e., the law of one price, which has long been maintained as a necessary and sufficient condition for Pareto efficiency in competitive markets, is not at all justified in revealing proper scale economies behavior of firms when market imperfections exist in any form.

Second, the factor-based technology employed in (6) is convex. Convexity, as argued by Farrell (1959), assumes away some important technological features such as *indivisible* production activities, *economies of scale* and *economies of specialization*, which all, in fact, result from *concavities* in production.

Third, the CE model (6) can also be of limited value in actual applications even when inputs are homogeneous. This is because as pointed out by Camanho and Dyson (2005, 2008), the CE reflects only input inefficiencies, i.e., *technical inefficiency* and/or *allocative efficiency*) but not *market* (*price*) *inefficiencies*. Therefore, they suggested a very comprehensive scheme to measure CE that can be attributed to both inputs and market inefficiencies.

Note that when market imperfections exist, $\varepsilon_h(y;w)$ is not very comprehensive as it involves the cost effects of output expansion only. To make it comprehensive, one need to link $\varepsilon_h(y;w)$ with further cost reduction due to other sources like pecuniary economies. Therefore, when inputs are heterogeneous, in order to account for varying input prices, the alternative CE model of Tone (2002) should be used where the technology is defined in cost-output space so as to account for varying input heterogeneity.

In order to measure EOS, this alternative CE DEA model of firm *h* can be set up as:

$$\theta^* = \min_{\theta,\lambda} \theta$$

$$\text{s.t.} \quad \sum_{j=1}^{n} c_j\lambda_j \le \theta c_h\ ,\ \sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{rh}(\forall\ r),\ \sum_{j=1}^{n}\lambda_j = 1,\ \lambda_j \ge 0 \qquad \text{... (11)}$$

For the cost efficient firm *h*, $\theta^* = 1$, and then it holds that

$$\sum_{r=1}^{s}\beta_r^* y_{rh} - \beta_1^* c_h + \beta_0^* = 0;\ \ \beta_r^* \ge 0\ (\forall\ r),\ \beta_1^* \ge 0\ ,$$

and the cost frontier is then derived as

$$c_h^* = \left(\beta_0^* / \beta_1^*\right) + \sum_{r=1}^{s}(\beta_r^* / \beta_1^*) y_{rh} \qquad \text{... (12)}$$

where β_r^*, β_o^* and β_1^* are dual multipliers obtained from the following Langrange function:

$$L = \text{-}\theta + \beta_1\left(\theta c_h - \sum_{j=1}^{n} c_j\lambda_j\right) + \sum_{r=1}^{s}\beta_r\left(\sum_{j=1}^{n} y_j\lambda_j - y_h\right) + \beta_0\left(\sum_{j=1}^{n}\lambda_j - 1\right) \qquad \text{... (13)}$$

From equation (12), we can obtain the marginal cost of the r^{th} output ($\partial c_h^* / \partial y_{rh}$) as

$$\frac{\partial c_h^*}{\partial y_{rh}} = \frac{\beta_r^*}{\beta_1^*}\ (\forall r)$$

Therefore, the alternative SE measure (ε_h^A) for firm *h* can be derived as

$$\varepsilon_h^A(y,w) = \frac{c_h^*}{\sum_{r=1}^{s} y_{rh}(\partial c_h^* / \partial y_{rh})} = \frac{c_h^*}{\sum_{r=1}^{s}(\beta_r^* / \beta_1^*) y_{rh}} = \frac{1}{1 - \left((\beta_0^* / \beta_1^*) / c_h^*\right)} \qquad \text{... (14)}$$

There are economies of scale (i.e., $\varepsilon_h^A > 1$), or no economies/diseconomies of scale (i.e., $\varepsilon_h^A = 1$) or diseconomies of scale (i.e., $\varepsilon_h^A < 1$) if (β_0^* / β_1^*) is, respectively, greater than, or equal to, and less than zero.

The CE model (11) is free from unrealistic assumptions such as price exogeneity, convexity, and the availability of unit market price data, which were all maintained in the CE model (6), and is therefore considered flexible in exhibiting EOS. Note that $\varepsilon_h^A(y,w) = \varepsilon_h(y,w)$ when all the firms face common unit input prices.

However, when the input market is competitive, and inputs are homogeneous, the cost model (6) is preferred to (11) with its potential limitation that its resultant CE measures only input inefficiency but not price efficiency. This is because the CE projection in (11) is made on a completely different technology where one sometimes finds the target over-demanding for some firms and not sufficiently challenging for others, given the properties of the factor-based technology. That is, given inputs, the target input prices are well below the observed prices, and given input prices, the target inputs are beyond the factor-based technology.

When inputs are heterogeneous, model (11) is preferred to model (6). The setup in (11) assumes that firms not only control over on the mix and quantities of inputs used but also exercise control over input prices. Using several different aspects of production planning process, this model indeed imputes a multifactor perspective in its SE estimates to track overall performance. Increasingly, companies are also too discovering the competitive power of undertaking each of these aspects - or the dangers of not doing so.

5. Concluding remarks

While many real-life manufacturing industries such as cement, iron, steel, etc. readily present themselves as contributing to scale, it is not clear at all, how many instances would actually result in there being ‘scale economies’ for constant factor proportion. The concept of RTS is argued to be too restrictive to capture any relevant scale effects of a firm. From an empirical perspective, this paper critically examines the concepts of RTS and EOS, and then argues that EOS is more relevant and broader enough to exhibit a firm’s actual scale behavior. Because of the fundamental shortcomings associated with the measure of RTS where output is narrowly related to the inputs only by defining the input-mix in a special way, RTS seems to be restrictive in exhibiting any relevant scale behavior. By exercising prudence in pending more thorough exploration of the sources of scale increases; a strong case is made in this contribution supporting the use of a nonparametric cost frontier approach to estimate EOS.

References

Baumol, W.J., J.C. Panzar and R.D. Willig (1982), *Contestable Markets and Theory of Industrial Structure*, Harcourt Brace Jovanovich, New York.

Camanho, A.S. and R.G. Dyson (2005), “Cost Efficiency Measurement with Price Uncertainty: A DEA Application to Bank Branch Assessments”, *European Journal of Operational Research*, 161(2), 432-446.

Camanho, A.S. and R.G. Dyson (2008), “A Generalization of the Farrell Cost Efficiency Measure Applicable to Non-fully Competitive Settings”, *Omega*, 36(1), 147-162.

Charnes, A., W.W. Cooper and A. Schinnar (1991), “An Extended Cobb-Douglas Form for Use in Production Economies” in Kaul, T. and J.K. Sengupta (Eds.) *Economic Models, Estimation and Socio-economic Systems*, North Holland, Amsterdam.

Färe, R. and S. Grosskopf (1985), “A Non-parametric Cost Approach to Scale Efficiency”, *Scandinavian Journal of Economics*, 87(4), 594-604.

Farrell, M.J. (1959), “Convexity Assumptions in Theory of Competitive Markets”, *Journal of Political Economy*, 67(4), 377-391.

Sahoo, B.K., M. Mehdiloozad and K. Tone (2014), “Cost, Revenue and Profit Efficiency Measurement in DEA: A Directional Distance Function Approach”, *European Journal of Operational Research*, 237(3), 921-931.

Sahoo, B.K. and K. Tone (2013), “Non-parametric Measurement of Economies of Scale and Scope in Non-competitive Environment with Price Uncertainty”, *Omega*, 41(1), 97-111.

Sengupta, J.K. (2004), “The Survivor Technique and the Cost Frontier: A Nonparametric Approach”, *International Journal of Production Economics*, 87(2), 185-193.

Sueyoshi, T. (1997), "Measuring Efficiencies and Returns to Scale in Nippon Telegraph and Telephone in Production and Cost Analysis", *Management Science*, 43(6), 779-796.

Tone, K. and B.K. Sahoo (2003), "Scale, Indivisibilities and Production Function in Data Envelopment Analysis", *International Journal of Production Economics*, 84(2), 165-192.

Tone, K. and B.K. Sahoo (2005), "Evaluating Cost Efficiency and Returns to Scale in the Life Insurance Corporation of India using Data Envelopment Analysis", *Socio-Economic Planning Sciences*, 39(4), 261-285.

Tone, K. and B.K. Sahoo (2006), "Re-examining Scale Elasticity in DEA", *Annals of Operations Research*, 145(1), 69-87.

Tone, K. (2002), "A Strange Case of Cost and Allocative Efficiencies in DEA", *Journal of the Operational Research Society*, 53(11), 1225-1231.

♦ ♦ ♦