On the Concept of Returns to Scale: Revisited

Parvez Azim[1]

Abstract

This paper shows why it is that in Economics text books and literature we invariably consider constant returns to scale (linearly homogeneous production functions) and not increasing returns to scale or decreasing returns to scale production . It has been demonstrated in this paper by using both cost elasticity output approach and Euler's theorem that the constant returns to scale production functions enable us to achieve productive efficiency and equilibrium. Production under increasing returns to scale or decreasing returns to scale are not at equilibrium. Only under constant returns to scale technology do we achieve productive efficiency and equilibrium.

Keywords: Returns to scale, linearly homogeneous production function, cost elasticity of output, productive efficiency

JEL classification: D24, O47, C23

1. Introduction

In the literature, the terms such as constant returns to scale (constant economies of scale), increasing returns to scale (economies of scale) and decreasing returns to scale (diseconomies of economies of scale) have been used quite frequently. However, it is has not been explicitly explained why we make use of constant returns to scale (linear homogeneous production function) and not increasing and decreasing returns to scale. This paper will shed light on this, classification will be done and the relationship between them will be elaborated upon. Understanding of these issues is important to understand various studies dealing with economies and diseconomies and pedagogical purposes.

2. Methodology

To start with, let there be a production function of the constant elasticity of substitution (CES) form whose associated cost function is derived here. It is trivial to obtain the ratio between the marginal cost (MC) and average cost (AC). The ratio is represented by the following expression which is called cost elasticity of output.

It is used to measure the returns to scale. Constant returns to scale (CRS) is when $\eta = 1$, MC=AC, increasing returns to scale (IRS) is when $\eta < 1$, MC < AC and decreasing returns to scale (DRS) is when $\eta > 1$, MC > AC. It is important to give definition of each of these. CRS ($\eta = 1$) means that a 1 percent increase in output results in exactly 1 percent increase in the total cost. IRS ($\eta < 1$) means that a 1 percent increase in output results in less than 1 percent increase in the total cost. DRS ($\eta > 1$) means that a 1 percent increase in output results in more than 1 percent increase in the total cost.

[1] *Professor of Economics, GC University Faisalabad*

The relationship between AC and output level y could also be used to explain CRS, IRS and DRS.

Let $AC = \theta(y) \equiv \frac{C(w,y)}{y}$ *where w is assumed to be fixed vector, and* $\theta(y)$ *signifies the average cost. By partial differentiation, we get*

$$\frac{\partial\theta}{\partial y} = \frac{y\frac{\partial C}{\partial y} - C}{y^2} = \frac{y\left\{\frac{\partial C}{\partial y} - \frac{C}{y}\right\}}{y^2} = \frac{\{MC - AC\}}{y}$$

$\frac{\partial\theta}{\partial y} = 0$ *when* $MC = AC$ (*CRS prevails if AC is constant with an increase in output*)

$\frac{\partial\theta}{\partial y} < 0$ *when* $MC < AC$ (*IRS prevails if AC is falling with an increase in output*)

$\frac{\partial\theta}{\partial y} > 0$ *when* $MC > AC$ (*DRS prevails if AC is rising with an increase in output*)

It is evident that CRS (η) *prevails if and only if CRS* (θ) *prevails.*

It is evident that IRS (η) *prevails if and only if IRS* (θ) *prevails.*

It is evident that DRS (η) *prevails if and only if DRS* (θ) *prevails.*

Let $\frac{\partial\theta}{\partial y} = \theta'$ *Thus it can be concluded that*

$\theta' = 0$ *if and only if* $\eta = 1$, $\theta' < 0$ *if and only if* $\eta < 1$ *and* $\theta' > 0$ *if and only if* $\eta > 1$.

$$\frac{\partial\left(\frac{C}{y}\right)}{\partial y} = \frac{y\left\{\frac{\partial C}{\partial y} - \frac{C}{y}\right\}}{y^2} = \frac{C\left(\frac{\frac{\partial C}{\partial y}}{\frac{C}{y}} - 1\right)}{y^2} = \frac{C(\eta - 1)}{y^2}$$ *shows that*

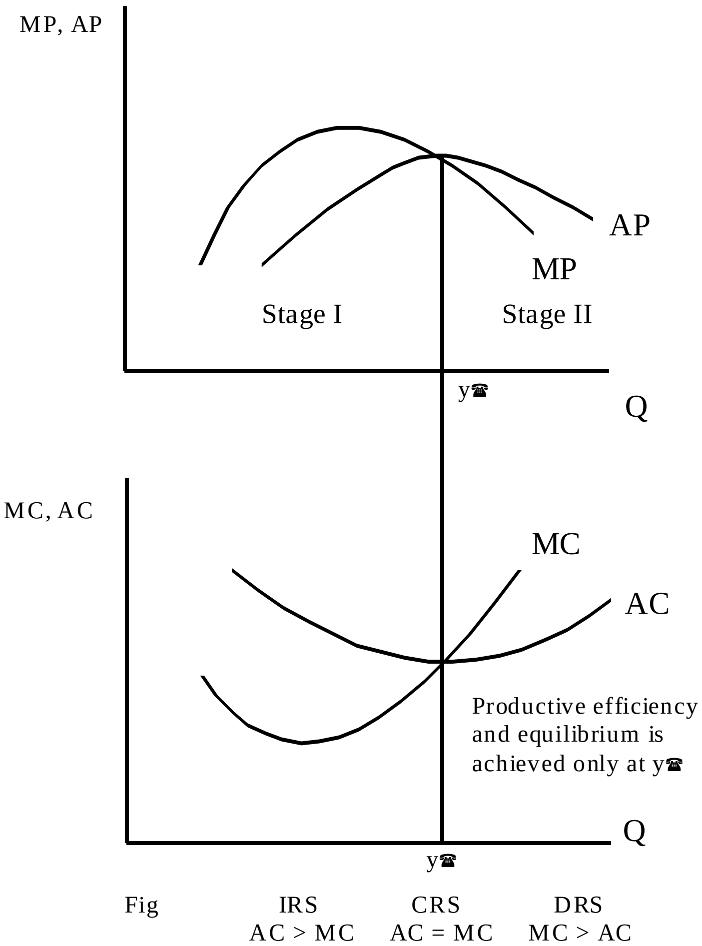
when $MC = AC$, *AC does not change with an increase in ouput* y. *as shown in the figure below.*

$$\frac{\partial (AC)}{\partial y} = 0 \Rightarrow \eta - 1 = 0 \Rightarrow MC = AC \ (\eta = 1) \ \ MC = AC$$

$$\frac{\partial (AC)}{\partial y} < 0 \Rightarrow \eta - 1 < 1 \Rightarrow MC < AC \ (\eta < 1) \ \ MC \text{ lies below } AC$$

$$\frac{\partial (AC)}{\partial y} > 0 \Rightarrow \eta - 1 > 1 \Rightarrow MC > AC \ (\eta > 1) \ \ AC \text{ lies below } MC$$

Looking at the usual U-shaped AC curve, the AC curve is decreasing to a certain level of output (in which $\frac{\partial \theta}{\partial y} < 0 \text{ and } \eta < 1)$ and then increases as output increases (in which $\frac{\partial \theta}{\partial y} > 0 \text{ and } \eta > 1)$.



Fig

It is also to be noted that MC curve intersects the AC curve from below when the AC curve is at its lowest point. It means that it is important to restrict the concept of returns to scale to a certain range of output levels.

Therefore, the productive efficiency is achieved at the output level that corresponds to the minimum point of AC curve where MC=AC. Any output level that is less than or more than y* is not a productive efficiency point. It is evident that both IRS when $\frac{\partial \theta}{\partial y} < 0 \; and \; \eta < 1$, and DRS when $\frac{\partial \theta}{\partial y} > 0 \; and \; \eta > 1$, don't correspond to the point where productive efficiency is achieved. This appears to be the most plausible reason that why in the literature we use the concept of CRS and not IRS or DRS. For example, the linearly homogeneous Cobb-Douglas production function and constant elasticity of substitution (CES) production function.

With reference to the above given figure; if a firm/industry is operating to the left of the minimum AC point, then it can increase profits by increasing production until it reaches the minimum AC point . Similarly, if a firm/industry is operating to the right of the minimum AC point, then it can increase profits by decreasing production until it reaches minimum AC point which is y* in the accompanied figure. It can be observed that when the MC curve is below an average cost curve the AC curve is falling. This relation holds true regardless of whether the MC curve is falling or rising. When the MC curve is above an AC curve the AC is rising. The MC curve intersects an AC curve at its minimum. It is important to mention that when average product (AP) is rising AC is falling and when AP is falling, AC is rising. It is also to be noted that when marginal product (MP) starts to fall, MC starts to rise. There is inverse relationship between AP and AC and MP and MC as shown in the figure. A rational producer will not produce in stage I where MP lies above AP. The production will take place in the economic region in stage II.

The downward sloping portion of the long run average cost (LRAC) curve corresponds to IRS (economies of scale). The horizontal portion of the LRAC curve corresponds to CRS (neither economies nor diseconomies of scale). An upward sloping portion of the LRAC curve corresponds to DRS (diseconomies of scale). In a long-run perfectly competitive (in the input markets) environment, the productively efficient and equilibrium level of output corresponds to the minimum efficient scale marked as y* in the figure. This is due to the zero profit requirement of a perfectly competitive equilibrium. This result implies production is at a level corresponding to the lowest possible AC. In fact all points along the LRAC are productively efficient by definition, but not all are equilibrium points in a long-run perfectly competitive environment. To elaborate on the point, let us consider a Constant Elasticity of Substitution (CES) production function and derive its associated cost function. The cost function will give us a relationship between MC and AC. It is evident as shown below that MC < AC when we have Increasing Returns to Scale, MC < AC when we have Decreasing Returns to Scale, and MC =AC when we have Constant Returns to Scale, Let us consider the CES *production function* $y = A\left[\delta L^{-\beta} + (1-\delta)K^{-\beta}\right]^{\frac{-h}{\beta}}$

whose associated cost function is $C = \left[\frac{y}{A}\right]^{\frac{1}{h}}\left[\delta^{\frac{1}{1+\beta}} w^{\frac{\beta}{1+\beta}} + (1-\delta)^{\frac{1}{1+\beta}} r^{\frac{\beta}{1+\beta}}\right]^{\frac{1+\beta}{\beta}}$ *and its*

cost elasticity of output is $\left(\frac{\partial C}{\partial Y}\frac{Y}{C}\right) = \frac{1}{h}$ *where h is the degree of homogeneity*

of the given CES production function.

Derivation of the cost function associated with the given CES production function

From $y = A\left[\delta L^{-\beta} + (1-\delta)K^{-\beta}\right]^{\frac{-h}{\beta}}$ *we get*

$$MP_L = w = \frac{\partial y}{\partial L} = h\delta L^{-(\beta+1)} A\left[\delta L^{-\beta} + (1-\delta)K^{-\beta}\right]^{\frac{-h}{\beta}-1}$$

$$MP_K = r = \frac{\partial y}{\partial K} = h(1-\delta)K^{-(\beta+1)} A\left[\delta L^{-\beta} + (1-\delta)K^{-\beta}\right]^{\frac{-h}{\beta}-1} \text{ dividing them we get}$$

$$\frac{w}{r} = \frac{\delta L^{-(\beta+1)}}{(1-\delta)K^{-(\beta+1)}} \text{ transposition leads to}$$

$$K^{-(\beta+1)} = \frac{r\delta L^{-(\beta+1)}}{w(1-\delta)} \Rightarrow K = L\left[\frac{r\delta}{w(1-\delta)}\right]^{\frac{1}{-(\beta+1)}} \quad (i)$$

$$\Rightarrow K^{-\beta} = L^{-\beta}\left[\frac{r\delta}{w(1-\delta)}\right]^{\frac{\beta}{(\beta+1)}} \quad (ii)$$

substitution of (ii) *into the given production function yields*

$$y = A\left[\delta L^{-\beta} + (1-\delta)L^{-\beta}\left[\frac{r\delta}{w(1-\delta)}\right]^{\frac{\beta}{(\beta+1)}}\right]^{\frac{-h}{\beta}} \text{ factoring out } L^{-\beta} \text{ gives us}$$

$$y = L^h A\left[\delta + (1-\delta)\left[\frac{r\delta}{w(1-\delta)}\right]^{\frac{\beta}{(\beta+1)}}\right]^{\frac{-h}{\beta}} \Rightarrow L = \left[\frac{y}{A}\right]^{\frac{1}{h}}\left[\delta + (1-\delta)\left[\frac{r\delta}{w(1-\delta)}\right]^{\frac{\beta}{(\beta+1)}}\right]^{\frac{1}{\beta}} \quad (iii)$$

substituting (iii) into (i) gives us

$$K = \left[\frac{y}{A}\right]^{\frac{1}{h}} \left[\delta + (1-\delta)\left[\frac{r\delta}{w(1-\delta)}\right]^{\frac{\beta}{(\beta+1)}}\right]^{\frac{1}{\beta}} \left[\frac{r\delta}{w(1-\delta)}\right]^{\frac{1}{-(\beta+1)}} \qquad (iv)$$

Substitution of (iii) and (iv) into the cost equation $C = wL + rK$ *gives us*

$$C = \left[\frac{y}{A}\right]^{\frac{1}{h}} \left[\delta + (1-\delta)^{\frac{1}{1+\beta}}\left[\frac{w}{r\delta}\right]^{\frac{-\beta}{(\beta+1)}}\right]^{\frac{1}{\beta}} \left[w + r\left(\frac{r\delta}{w(1-\delta)}\right)^{\frac{1}{-(\beta+1)}}\right]$$

$$\Rightarrow C = \left[\frac{y}{A}\right]^{\frac{1}{h}} \left[\frac{\delta^{\frac{1}{1+\beta}} w^{\frac{\beta}{1+\beta}} + (1-\delta)^{\frac{1}{1+\beta}} r^{\frac{\beta}{1+\beta}}}{\delta^{\frac{-\beta}{1+\beta}} w^{\frac{\beta}{1+\beta}}}\right]^{\frac{1}{\beta}} \left[\frac{\delta^{\frac{1}{1+\beta}} w^{\frac{\beta}{1+\beta}} + (1-\delta)^{\frac{1}{1+\beta}} r^{\frac{\beta}{1+\beta}} \left(\frac{\delta}{w}\right)^{\frac{-1}{1+\beta}}}{\delta^{\frac{1}{1+\beta}} w^{\frac{-1}{1+\beta}}}\right]$$

cancelling out the denominators we end up with

$$C = \left[\frac{y}{A}\right]^{\frac{1}{h}} \left[\delta^{\frac{1}{1+\beta}} w^{\frac{\beta}{1+\beta}} + (1-\delta)^{\frac{1}{1+\beta}} r^{\frac{\beta}{1+\beta}}\right]^{\frac{1+\beta}{\beta}} \qquad QED$$

(ii) The cost elasticity of output $= \dfrac{\partial C}{\partial y}\dfrac{y}{C} = \dfrac{1}{h}\left[\dfrac{y}{A}\right]^{\frac{1}{h}-1} \dfrac{1}{A}\left[\delta^{\frac{1}{1+\beta}} w^{\frac{\beta}{1+\beta}} + (1-\delta)^{\frac{1}{1+\beta}} r^{\frac{\beta}{1+\beta}}\right]^{\frac{1+\beta}{\beta}} \dfrac{y}{C}$

$$\left(\frac{\partial C}{\partial y}\right)\left(\frac{y}{C}\right) = \frac{1}{h}\left(\frac{C}{\frac{y}{A}}\frac{1}{A}\right)\left(\frac{y}{C}\right) = \frac{1}{h} \quad or \quad \Rightarrow \left(\frac{\partial C}{\partial y}\right) = \frac{1}{h}\left(\frac{C}{y}\right)$$

$$MC = \frac{1}{h} AC \quad \Rightarrow h = \frac{AC}{MC}$$

where h is the degree of homogenity of the given CES production function.

When $h = 1$ *we have CRS* $(MC = AC)$ *production function*

When $h > 1$ *we have IRS* $(MC < AC)$ *production function*

When $h < 1$ *we have DRS* $(MC > AC)$ *production function*

The concept of returns to scale could also be explained with the help of the Euler's theorem as explained below.

According to the Euler's Theorem

$K\,MP_K + L\,MP_L = hQ$ *where 'h' is the degree of homogeneity of the production function. Multiplying throughout by P gives us*

$K\,PMP_K + L\,PMP_L = h\,(PQ)$ *where P is the price of output and PQ is total value of output.*

$$\underbrace{K \overbrace{PMP_K}^{\text{wage of capital (VMP)}=r}}_{\text{total cost of input } K} + \underbrace{L \overbrace{PMP_L}^{\text{wage of labour (VMP)}=w}}_{\text{total cost of input } L} = h \underbrace{PQ}_{\text{Total value of output or revenue}}$$

Total cost of producing Q

If $h = 1$ the firm spends whole revenue on inputs.

If $h > 1$ the firm spends more on inputs than revenue, so incurring loss

If $h < 1$ the firm spends less than revenue, so the firm makes a positive profit.

When $h = 1$, it is termed as constant returns to scale (CRS), $h > 1$ increasing returns to scale (IRS) and $h < 1$ decreasing returns to scale (DRS).

It is obvious that costs will increase equal to, more than, or less than linearly with output as 'h' is equal to, greater than, or less than one. It is so because the CES technology exhibits constant, increasing or decreasing returns to scale depending on the value of 'h'.

3. Conclusion

In the production theory, linearly homogeneous production functions (also called constant returns to scale) are used because they depict constant returns to scale technology which shows productive efficiency and equilibrium point shown by the lowest AC curve point y* in the above given figure. Output less than or more than that point does not give us equilibrium point of production. Productive efficiency plus equilibrium is achieved only at the lowest AC curve where MC is cutting AC curve from below.

Compliance with Ethical Standards: The author declares that he has no conflict of interest, financially, non - financially, directly or indirectly related to the work.

References

- Basu, Susanto and John G. Fernall. “Returns to Scale in U.S. Production. Estimations and Implications”, The Journal of Political Economy. Vol.105,No.2 (april,1997), 249-283.
- Hanoch,G.1975. “The Elasticity of Scale and the Shapes of Average Costs.” *American Economic Review* 65 (June): 492-97
- Shephard,R.W. 1953. *Cost and Production Functions*. Princeton, N.J. Princeton University Press.
- Shephard,R.W. 1970. *Theory of Cost and Production Functions*. Princeton, N.J. Princeton University Press.
- Silberberg, E. 1990. *The Structure of Economics*.2nd ed. New York: McGraw-Hill.
- Takayama, A. 1985. *Mathematical Economics*. 2d. ed. New York: Cambridge University Press.

- Takayama, A. 1994. *Analytical Methods in Economics*. 1st.ed. Hertfordshire: Harvester Wheatsheaf, USA.
- Uzawa,H. 1962. “Production Functions with Constant Elasticities of Substitution.” Review of Economic Studies 29 (October): 291-99.