

October 2002

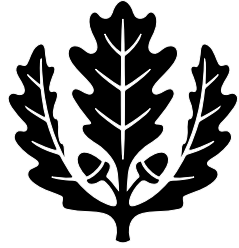
Nonparametric Measures of Capacity Utilization: A Dual Approach

Subhash C. Ray
University of Connecticut

Follow this and additional works at: https://opencommons.uconn.edu/econ_wpapers

Recommended Citation

Ray, Subhash C., "Nonparametric Measures of Capacity Utilization: A Dual Approach" (2002). *Economics Working Papers*. 200235.
https://opencommons.uconn.edu/econ_wpapers/200235



University of
Connecticut

Department of Economics Working Paper Series

Nonparametric Measures of Capacity Utilization: A Dual Approach

Subhash C. Ray
University of Connecticut

Working Paper 2002-35

October 2002

341 Mansfield Road, Unit 1063
Storrs, CT 06269-1063
Phone: (860) 486-3022
Fax: (860) 486-4463
<http://www.econ.uconn.edu/>

Abstract

This paper develops a nonparametric method of obtaining the minimum of the long run average cost curve of a firm to define its capacity output. This provides a benchmark for measuring of capacity utilization at the observed output level of the firm. In the case of long run constant returns to scale, the minimum of the short run average cost curve is determined to measure short run capacity utilization. An empirical application measures yearly rates of capacity utilization in U.S. manufacturing over the period 1968-1998. Nonparametric determination of the short run average cost curve under variable returns to scale using an iterative search procedure is described in an appendix to this paper.

NONPARAMETRIC MEASURES OF CAPACITY UTILIZATION: A DUAL APPROACH

Subhash C Ray
Department of Economics
University of Connecticut
Storrs CT 06269-1063

The capacity output of a firm can be defined in alternative ways. As a physical upper limit, it measures the maximum quantity of output that a firm can produce from a given bundle of quasi-fixed inputs *even when other (variable) inputs are available without any restriction*. This definition, due to Johansen (1968), is intuitively quite appealing. After all, even when labor, material, and energy are available in unlimited quantities, a firm can produce only a finite quantity of output from its plant and equipment of any given size. The actual output produced must be less than or equal to this capacity output. The rate of capacity utilization, then, is merely the ratio of its actual output and the capacity output level. Less than 100% capacity utilization may be due either to insufficient demand faced by the firm inducing it to restrict production to a level below capacity or due to shortage of some critical input (e.g., energy) holding back production even when there is sufficient demand for the product. A different, and economically more meaningful, definition of the capacity output due to Cassels (1937) is the level of production where the firm's long run average cost curve reaches a minimum. Because we consider the long run average cost, no input is held fixed. For a firm with the typical U-shaped average cost curve, at this capacity level of output, economies of scale have been exhausted but diseconomies have not yet set in. The physical limit defines the capacity of one or more quasi-fixed input. On the other hand, the economic measure pertains to capacity utilization of *all inputs*¹.

In production economics, considerable interest centers on the output level where the U-shaped average cost curve of a firm reaches a minimum. In deed, it is the only output level of the firm that is consistent with long run equilibrium in a constant cost competitive industry. Also, the market demand is supplied at the lowest cost when produced by the requisite number of identical firms each producing this specific level of output. In this sense, it is a socially efficient output scale for a typical firm in the industry. If its actual scale of production is smaller than this efficient scale, a firm can exploit economies of scale and lower its average cost by increasing its output. Conversely, when actual output exceeds the optimal, the firm is experiencing diseconomies of scale and reducing output would lower its average cost. Thus, identifying the optimal economic scale is clearly useful for policy².

Klein (1960) argued that the long run average cost curve may not have a minimum and proposed the output level where the short run average cost curve is tangent to the long run average cost curve as an alternative measure of the capacity output. This is also the approach adopted by Berndt and Morrison (1981).

In the short run, the economic measure of capacity output would correspond to the level where the short run average cost curve of the firm reaches a minimum. Of course, when the technology exhibits

constant returns to scale, the long run average cost curve is horizontal and the capacity level of output is not defined. In this case, however, at the minimum point the short run average cost curve is tangent to the long run average cost curve. This helps to determine the economic capacity output level in the short run and yields a measure of the rate of capacity utilization of the fixed input.

In the nonparametric literature, Färe, Grosskopf, and Kokkelenberg (1989) provide a nonparametric model for measuring the physical capacity output and the associated rate of capacity utilization in the presence of fixed inputs³. By contrast, there is no comparable parametric model. An apparent reason for this lack of a parametric measure of (physical) capacity utilization is that except for Leontief-type production functions, the marginal product of any variable input always remains strictly positive. Therefore, even when some inputs remain fixed, the output continues to increase when the variable inputs are increased⁴. For the economic measure of the capacity output level and the rate of capacity utilization, however, the story is just the opposite. In the case of a parametrically specified cost function, one typically examines the appropriate first order condition for minimizing the average cost in order to determine the relevant level of output. For example, Nelson (1989) estimated a Translog variable cost function to measure capacity utilization in a sample of privately owned electrical utilities in the U.S. This is not possible in nonparametric analysis. In the existing Data Envelopment Analysis (DEA) literature considerable attention is devoted to measuring scale efficiency (e.g., Färe, Grosskopf, and Lovell (1994) Chapter 3). One can solve appropriate DEA models to determine whether the actual output level of a firm minimizes average cost and, if not, how far its average cost could be lowered by an appropriate change in the level of output. The literature does not explain, however, how to determine *the output level* that minimizes the average cost of a firm facing a given vector of input prices.

The present paper extends the DEA literature by formulating a procedure for determining this efficient output level in the single output case⁵. We consider both the long run and the short run average cost curves. For a determinate minimum of the average cost curve in the long run we assume variable returns to scale. Both constant and variable returns to scale are considered for the short run when some inputs are held fixed. In the empirical application using annual aggregate data from the Bureau of Labor Statistics for input quantities and prices as well as output quantities for U.S. manufacturing for the years 1968-98. For this empirical application, a production technology characterized by long run constant returns to scale and non-regressive technical change is conceptualized. We measure both the physical and the economic capacity output levels and the corresponding rates of capacity utilization in the short run.

The paper unfolds as follows. Section 2 provides a brief theoretical background from production economics. Section 3 presents the relevant DEA models for finding the minimum point on the long run average cost curve under variable returns to scale, the minimum of the short run average cost curve under long run constant returns to scale, and also the model for finding the physical capacity output level for given level of the quasi-fixed inputs.. Section 4 contains an empirical application using the BLS data for U.S. manufacturing. Section 5 concludes. A search procedure for finding the minimum of the short run average cost function under variable returns to scale is proposed in an appendix.

2. The Theoretical Background

Consider a firm in some industry producing the output vector $y \in R_m^+$ using the input vector $x \in R_n^+$. Let the technology faced by the industry be described by the production possibility set

$$T = \{(x, y): x \text{ can produce } y\}. \quad (1)$$

If we assume that both inputs and outputs are freely disposable and the production possibility set is convex, then

- (a) if $(x^0, y^0) \in T$ and $x^1 \geq x^0$, then $(x^1, y^0) \in T$;
- (b) if $(x^0, y^0) \in T$ and $y^1 \leq y^0$, then $(x^0, y^1) \in T$; and,
- (c) if $(x^0, y^0) \in T$ and $(x^1, y^1) \in T$, then $(\lambda x^0 + (1-\lambda)x^1, \lambda y^0 + (1-\lambda)y^1) \in T$ for $0 \leq \lambda \leq 1$.

If we additionally assume globally constant returns to scale, then

- (d) if $(x, y) \in T$, then $(kx, ky) \in T$ for $k \geq 0$.

For any specific output bundle y^0 , the input set is

$$V(y^0) = \{x: (x, y^0) \in T\}. \quad (2)$$

Thus, $V(y^0)$ consists of all input bundles x from which it is possible to produce the output bundle y^0 . Given an input price vector w^0 , the minimum cost of producing the output y^0 is

$$C(y^0, w^0) = \min w^0 x$$

subject to $x \in V(y^0)$. (3)

Suppose that x^* is the cost minimizing input bundle and

$$C^* = w^0 x^* = C(y^0, w^0). \quad (4)$$

For a scalar output, the average cost at the output level y_0 for input price w^0 is

$$AC(y_0, w^0) = \frac{C(y_0, w^0)}{y_0}. \quad (5)$$

The optimal output level, then, is one that minimizes $AC(y, w)$ for a given w .⁶

From a parametric specification of the cost function, one can derive the output in the single output case and the optimal scale in the multiple output case from the first order condition for minimizing the (ray) average cost, when an interior solution exists for a minimum. By contrast, a nonparametric procedure like DEA yields a measure of the minimum cost only at a specific level (or scale) of output. Hence, analytical derivation of the optimal level or scale of output is not possible. One can solve the problem indirectly, however, by utilizing a number of useful results from production economics.

Consider, first, the *most productive scale size (MPSS)* of a given input-output mix (x, y) . Frisch (1965) defined the technically optimal production scale of an input bundle in the single output scale as one where the ray average productivity reaches a maximum. Banker (1984) generalized Frisch's concept of the technical optimal production scale of a given input mix to the multiple-output case. A feasible input-output combination (x^0, y^0) is an MPSS for the specific input- and output-mix if *for every feasible input-output combination* (x, y) *satisfying* $x = \beta x^0$ *and* $y = \alpha y^0$, $\frac{\alpha}{\beta} \leq 1$. Further, locally constant return

s to scale holds at (x^0, y^0) if it is an MPSS (See proposition 1 in Banker (1984)).

Next, note that if the input bundle x^* minimizes the average cost at the output level y^* , then (x^*, y^*) is an MPSS.. Suppose this is not true. Then, by the definition of an MPSS there exist non-negative scalars (α, β) such that $(\beta x^*, \alpha y^0)$ is a feasible input-output combination satisfying $\frac{\alpha}{\beta} > 1$. Define $x^{**} = \beta x^*$ and $C^{**} = w^0 x^{**}$. Then, at input price w^0 , the minimum cost of producing the output bundle (αy^0) cannot be any greater than C^{**} . This implies that

$$AC(\alpha y^0, w^0) = \frac{C(\alpha y^0, w^0)}{\alpha y^0} \leq \frac{C^{**}}{\alpha y^0} = \frac{\beta w^0 x^*}{\alpha y^0} = \frac{\beta}{\alpha} AC(y^0, w^0).$$

But, by assumption $\frac{\beta}{\alpha} < 1$.

Thus,

$$AC(\alpha y^0, w^0) < AC(y^0, w^0).$$

Hence, y_0 cannot be the output level where average cost reaches a minimum. This shows that the average cost minimizing input-output combination must be an MPSS and, therefore, exhibit locally constant returns to scale.

When the technology exhibits globally constant returns to scale, the cost function is homogeneous of degree 1 in output. Thus,

$$C(ty, w^0) = t \cdot C(y, w^0), \text{ for all } t > 0.$$

Setting $t = \frac{1}{y}$, yields $C(y, w^0) = y \cdot C(1, w^0)$ so that

$$AC(y, w^0) = \frac{C(y, w^0)}{y} = C(1, w^0) \text{ for all } y.$$

Therefore, the minimum average cost attained at the efficient output level for a variable returns to scale technology is the same as what would be the average cost at *every output level (scale)* if the technology exhibited constant returns everywhere.

Finally, a cost indirect representation of the technology is in terms of the indirect output set:

$$IP(w, C) = \{y: (x, y) \in T, w^0 x \leq C\}. \quad (6a)$$

It is the set of all output quantities that can be produced from input bundles that cost no more than C at price w . The indirect output isoquant may be defined as the set

$$\overline{IP}(w, C) = \{y: y \in IP(w, C), \phi y \notin IP(w, C), \phi > 1\}. \quad (6b)$$

Neoclassical duality in production ensures that

$$C = C(y, w) \text{ if and only if } y \in \overline{IP}(w, C).$$

In the single output case, $C(y, w)$ is the minimum cost of producing output y at price w of the inputs and at the same time y is the maximum output that can be produced from any bundle that a firm can buy if it incurs an expenditure C .

3. The DEA Methodology

Suppose that we observe the input-output bundles of N firms. Let (x^j, y^j) be the observed input-output bundle for firm j . Clearly, every actually observed bundle is feasible. Hence, under the assumption of free disposability of inputs and outputs and convexity, the production possibility set under the VRS assumption is

$$T^V = \{(x, y) : x \geq \sum_{j=1}^N \lambda_j x^j; y \leq \sum_{j=1}^N \lambda_j y^j; \sum_{j=1}^N \lambda_j = 1; \lambda_j \geq 0, (j = 1, 2, \dots, N)\} \quad (7a)$$

and if CRS is assumed to hold everywhere

$$T^C = \{(x, y) : x \geq \sum_{j=1}^N \lambda_j x^j; y \leq \sum_{j=1}^N \lambda_j y^j; \sum_{j=1}^N \lambda_j \geq 0, (j = 1, 2, \dots, N)\} \quad (7b)$$

3.1 Long run Cost Minimization

For any scalar output y_0 and input price w^0 the minimum cost under the CRS assumption is

$$\begin{aligned} C^{**}(y^0, w^0) &= \min w^{0'} x \\ \text{s.t. } \sum_{j=1}^N \lambda_j x^j &\leq x; \\ \sum_{j=1}^N \lambda_j y_j &\geq y_0; \\ \lambda_j &\geq 0; (j = 1, 2, \dots, N). \end{aligned} \quad (8)$$

Define

$$\alpha^* = \frac{C^{**}(y_0, w^0)}{y_0}. \quad (9)$$

Now consider variable returns to scale. Suppose that the y^* is the output level where the average cost reaches a minimum under VRS and x^* is the corresponding input bundle that minimizes the *total* cost for y^* . Then, as argued above, locally constant returns to scale hold at (x^*, y^*) . Hence the total or the average cost of producing y^* at input price w^0 is the same whether VRS or CRS is assumed. Now, because under the

assumption of CRS the average cost is the same at all levels of the output, α^* is also the average cost at output y^* . Therefore the total cost of producing output y^* at price w^0 (both under CRS and VRS) is

$$C(y^*, w^0) = \alpha^* y^* \equiv \bar{C}. \quad (10)$$

Now we can determine y^* as the largest y in $\bar{IP}(w^0, \bar{C})$ by solving the following LP problem:

$$\begin{aligned} y^* &= \max y \\ \text{s.t.} \quad &\sum_{j=1}^N \lambda_j x^j \leq x; \\ &\sum_{j=1}^N \lambda_j y_j \geq y; \\ &\sum_{j=1}^N \lambda_j = 1; \\ &w^{0j} x \leq \alpha^* y; \\ &\lambda_j \geq 0; (j = 1, 2, \dots, N). \end{aligned} \quad (11)$$

The method outlined above can be explained for the 1-output, 2-input case diagrammatically as follows. In Figure 1, the quadrant to the left of the origin shows the isoquant map of the firm with input x_1 measured along the horizontal axis and input x_2 up the vertical axis. Tangency of the firm's expenditure line(s) with slope equal to the ratio of the input prices with the various isoquants determines the cost-minimizing input bundle for the relevant output level. Suppose that the scale of input x_2 has been normalized so that its price equals unity. In that case, the vertical intercept of the tangent line also measures the minimum cost of producing that output level. Thus, the vertical axis is utilized to measure both the quantity of input x_2 and the cost, C . In the quadrant to the right of the origin in the same diagram, we measure the quantity of output y . A point on the curve labeled TC in this quadrant shows the combinations of the output level y (that corresponds to any particular isoquant) and the vertical intercept of the (cost-minimizing) line tangent to that isoquant. Thus, it is the usual total cost curve. It may be noted that just as for any isoquant the (tangent) expenditure line shows the minimum cost of producing the corresponding output level, the isoquant tangent to the line defines the maximum output level that can be produced from any input bundle that costs no more than the specified amount of expenditure. Consider, for example, the pair (y_1, C_1) on the TC curve. Measured vertically, C_1 is the minimum cost of producing output y_1 at the given input prices. At the same time, measured horizontally, y_1 is the maximum output producing from an expenditure of C_1 at these prices. As one moves along the TC curve, the average cost first declines and then increases with the output level. At output y_2 in this diagram the average cost reaches a minimum and locally constant returns to scale holds at this point. If, instead, the technology exhibited constant returns to scale everywhere, the average cost would have remained constant at all levels of output (including y^*) so that the total cost would have been

the dotted line TC^* which is a ray through the origin. We first solve the DEA problem for cost minimization at a set of given input prices for any arbitrary output level under the assumption of constant returns. The resulting average cost is defined as α^* . Next we search for the output level where the average cost corresponding to the TC curve is α^* . Thus, at the output level, the total cost would equal α^* times y . This is accomplished by solving the DEA-LP problem in (11).

3.2 Long run Constant Returns and Short run Capacity Utilization

When the technology exhibits long run constant returns, the long run average cost curve is horizontal and the capacity level of output is not defined. But, in this case, finding the minimum point of the short run average cost (SAC) curve is quite simple.

Assume that the input vector x of the firm can be partitioned in the short run as

$$x = (v, K)$$

where v is the vector of its variable inputs while K is the vector of quasi-fixed inputs. Correspondingly, w_v and w_r are the price vectors of variable and fixed inputs, respectively. For a given bundle of quasi-fixed inputs, K_0 , the firm's fixed cost ($w_r'K_0$) is given and the firm's optimization problem is to minimize the variable cost for the given output level, y_0 . The firm's minimum variable cost function is

$$VC(w_v, K_0, y_0) = \min w_v'v : (v, K_0) \in V(y_0). \quad (12)$$

Thus the firm's short run total cost is

$$SRTC = VC(w_v, K_0, y_0) + w_r'K_0 \quad (13a)$$

and the short run average cost is

$$SAC = \frac{VC(w_v, K_0, y_0) + w_r'K_0}{y_0}. \quad (13b)$$

Because we assume long run CRS, as explained before, this SAC at its minimum will be equal to the constant long run average cost (LAC). Therefore, we first solve the long run cost minimization problem (8) and compute the LAC (α^*).

Thus, if y_s^* minimizes the SAC for $K=K_0$, then the short run total cost is $\alpha^* y_s^*$ and the variable cost is $\alpha^* y_s^* - w_r'K_0$.

We can, therefore determine y_s^* as

$$\begin{aligned} y_s^* &= \max y \\ \text{s.t.} \quad &\sum_{j=1}^N \lambda_j v^j \leq v; \\ &\sum_{j=1}^N \lambda_j y_j \geq y; \\ &\sum_{j=1}^N \lambda_j K \leq K_0; \\ &\alpha^* y - w_v'v \geq w_r'K_0; \lambda_j \geq 0; (j = 1, 2, \dots, N). \end{aligned} \quad (14)$$

The optimal solution of this problem yields the short run capacity output level. The economic measure of capacity utilization rate is

$$\rho = \frac{y_o}{y_s^*}. \quad (15)$$

3.3 Physical Measure of Capacity Utilization

Finally, we recall the DEA procedure proposed by Färe, Grosskopf, and Lovell for measuring the maximum output producible from the quasi-fixed input K_0 without any restriction on the variable inputs:

$$\begin{aligned} & \max \varphi_F \\ \text{s.t.} \quad & \sum_{j=1}^N \lambda_j v^j - v \leq 0; \\ & \sum_{j=1}^N \lambda_j K_j \leq K_0; \\ & \sum_{j=1}^N \lambda_j y_j \geq \varphi_F y_0; \\ & \lambda_j \geq 0; j = 1, 2, \dots, N. \end{aligned} \quad (16)$$

Note that in this problem the variable input constraints are irrelevant and can be deleted. The (physical) capacity output level is $y_F = \varphi_F^* y_0$. It represents the output level that a firm operating at full technical efficiency can produce from the given quasi-fixed input bundle K_0 when there is no restriction on the availability of the variable inputs. As Färe, Grosskopf, and Kokkelenberg argued, the actual output y_0 would be even lower than what is maximally producible from K_0 and the observed bundle of its variable inputs if it is not technically efficient.

In order to measure the technical efficiency of the firm producing output y_0 from its input bundle $x^0 = (v^0, K_0)$ we solve the DEA model due to Charnes, Cooper, and Rhodes (CCR):

$$\begin{aligned} & \max \varphi_C \\ \text{s.t.} \quad & \sum_{j=1}^N \lambda_j v^j \leq v^0; \\ & \sum_{j=1}^N \lambda_j K_j \leq K_0; \\ & \sum_{j=1}^N \lambda_j y_j \geq \varphi_C y_0; \\ & \lambda_j \geq 0; j = 1, 2, \dots, N. \end{aligned} \quad (17)$$

The technical efficiency of the firm is

$$TE = \frac{1}{\varphi_C^*} \quad (18)$$

and the maximum output producible from the firm's observed bundle of variable and fixed inputs is

$$y_C^* = \varphi_C^* y_0. \quad (19)$$

It should be emphasized that the physical measure of capacity utilization is

$$\rho_F = \frac{y_C^*}{y_F} = \frac{\varphi_C^*}{\varphi_F^*}. \quad (20)$$

In particular, the rate of physical capacity utilization is not measured by the ratio of the actual output y_0 and the capacity output y_F because the ratio incorporates technical inefficiency on top of under-utilization of capacity.

4. The Empirical Application

In this paper we examine annual data for total manufacturing in the U.S for the years 1968-98. A production technology involving a single output and 5 inputs is considered. Output is measured by a quantity index of gross output. The inputs are (a) labor, (b) capital, (c) energy, (d) materials, and (e) purchased services. All inputs are measured by the appropriate quantity indexes. We treat capital as quasi-fixed in the short run. Price indexes of the individual inputs were used as the relevant input prices in the cost-minimization problems. It is assumed that constant returns to scale hold in the long run. Further, technical change is assumed to be non-regressive. Therefore, all input-output combinations from previous years along with the current input-output bundle in any year are considered feasible during that year. Thus, in effect, we consider a sequential frontier.

A problem one faces in applying DEA to a time series data set of inputs and output is that we have only one observed input-output combination for each period. Even when non-regressive technical change is allowed only the current and past data can be used to construct the production possibility set for any particular year. An implication of this is that for the earlier years in the sample there will be too few observations to use for any meaningful approximation of the production frontier. To overcome this shortcoming we decided to use the data for the years 1948 through 1967 as initial values and started constructing the sequential frontier only from the year 1968.

Table 1 reports the actual output quantity indexes along with the physical and the economic measures of the capacity output for individual years. The implied rates of capacity utilization for the alternative measures are also reported in separate columns. Additionally, the official measure of capacity utilization in manufacturing from the U.S. Federal Board of Governors⁷ is included for comparison.

For the different years, the actual output level is shown in the column Q. The columns named Q_A and Q_B are, respectively, the economic and the physical capacity levels of output. The corresponding

capacity utilization rates are shown as $CAPU_A$ and $CAPU_B$, respectively. The Federal Reserve's measures of capacity utilization for the different years are reported in the column named $CAPU_{FRB}$. Several interesting points may be noted. Overall, the physical capacity utilization measure and the official measure are in broad agreement with a correlation coefficient of 0.712. By contrast, the economic measure has a much lower correlation (0.438) with the official measure. However, during the stagflation years (1974-82), the different measures are in much closer agreement. For this sub-period, the pair-wise correlations increase to 0.80 and 0.71, respectively. Second, the economic capacity utilization rates are, in general, much higher than either the physical or the official utilization rate. In 12 of the 31 years the actual output in manufacturing is where short run average cost reaches a minimum.⁸ Except for 6 years (1975-76, 1979-82) the actual output was no less than 90% of the economic capacity output. On the other hand, The physical capacity utilization rate was below 90% in all years except 1968-69 and 1973. In deed, it was below 80% in 22 out of the 31 years considered falling even below 70% during 1981-83. Another interesting point is that the physical capacity utilization rate was below the official rate for every year during 1974-98.

The empirical evidence suggests that during most years the level of production in U.S. manufacturing was quite close to the optimal scale for the given level of the capital input. Thus, any potential reduction the average cost could come only from improvement in efficiency rather than from economies of scale. At the same time, the physical capacity utilization measure shows that the available quantity of the capital input was not a bottleneck and the manufacturing output could have meet substantially additional demand. This is especially true for the years since 1974.

5. Conclusion

In this paper, we propose a nonparametric procedure for measuring the output level where the long run average cost curve of a firm reaches a minimum under variable returns to scale. This is the long run capacity output of the firm. Comparing its actual output with this capacity output we obtain a dual measure of capacity utilization. In the case of long run CRS, we find the minimum point of its short run average cost curve of a firm to measure its short run capacity utilization. The empirical application to U.S. manufacturing shows that in most years the actual output was close to its economic capacity level. By contrast, the physical capacity measure shows considerable under utilization.

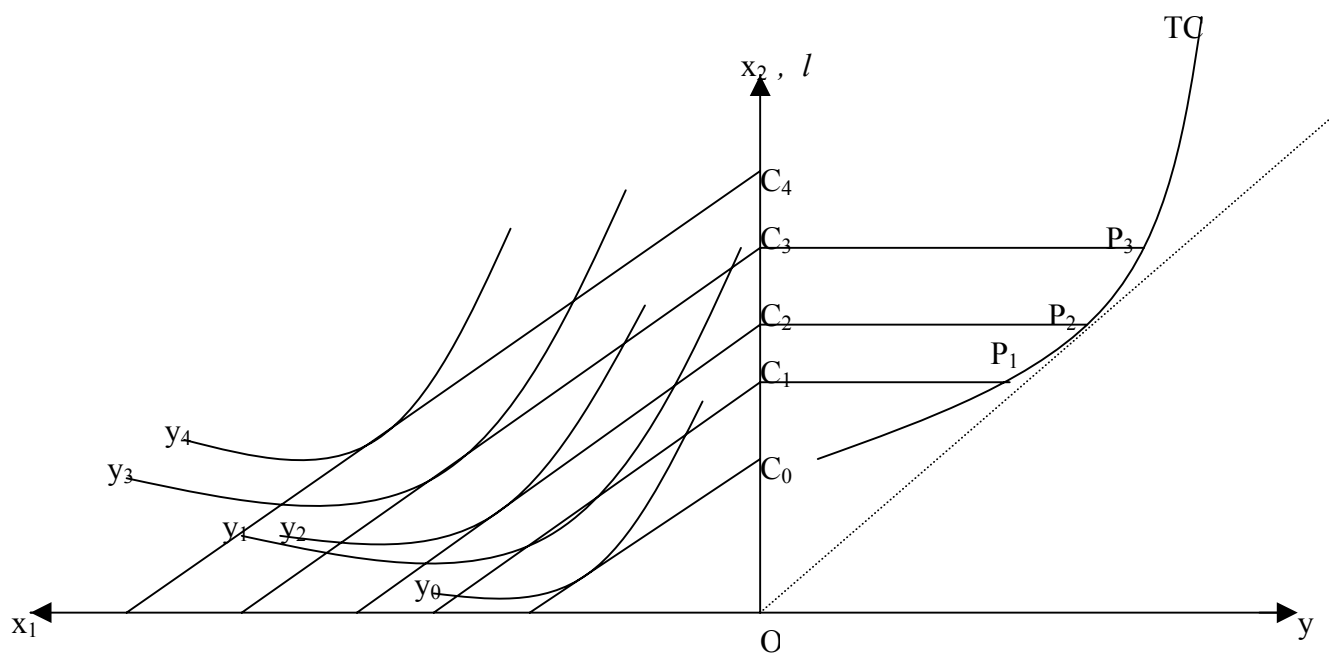


Figure 1

Appendix

Short run Capacity Utilization under Variable Returns to Scale

Suppose that we consider a firm in the short run when one or more inputs are treated as quasi-fixed. For simplicity, let x be the vector of variable inputs and w the corresponding vector of input prices. Further, let K be a single quasi-fixed input and r the price of this input. Thus, the variable cost of the firm is

$$VC = w'x \quad (A1)$$

and the fixed cost is

$$FC = rK. \quad (A2)$$

Suppose that the observed input bundles are (x^j, K_j) $j=1,2,\dots,N$. Then, under variable returns to scale, the short run variable cost minimization problem of a firm using K_0 units of the quasi-fixed input to produce output y_0 is:

$$\begin{aligned} \min VC &= w'x \\ \text{s.t.} \quad & \sum_{j=1}^N \lambda_j x^j; \\ & \sum_{j=1}^N \lambda_j K_j \leq K_0; \\ & \sum_{j=1}^N \lambda_j y_j \geq y_0; \\ & \sum_{j=1}^N \lambda_j = 1; \\ & \lambda_j \geq 0; j = 1,2,\dots,N. \end{aligned} \quad (A3)$$

Let $VC^* = w'x^*$ be the optimal solution and the short run cost of the firm be

$$SRTC = w'x^* + rK_0 \quad (A4).$$

The short run average cost of the firm is

$$SAC = \frac{w'x^* + rK_0}{y_0}. \quad (A5)$$

The marginal cost (MC) at the output level y_0 is given by the value of the dual variable associated with the output constraint at the optimal solution of this variable cost minimization problem. At the output level where the short run average cost curve of the firm reaches a minimum, SAC equals the marginal cost. If the short run average cost curve has the usual U-shape, SAC is declining at an output level where MC is lower than the SAC and rising where MC exceeds SAC . Unlike in the case of the long run average cost function,

in the case of the *SAC* function we actually need to employ an iterative search procedure consisting of the following steps.

Step 1: Select some output level y_0 , a positive step length δ , and a small positive number ε (e.g., 0.00001).

Step 2. Solve the DEA-LP problem (A3) at $y = y_0$ for the given w and K_0 .

Compute *SAC* at y_0 and compare it with the *MC* obtained from the optimal solution of (A3).

Step 3: If $SAC = MC$, then stop. *SAC* reaches a minimum at y_0 .

If $SAC > MC$, go to Step 4.

If $SAC < MC$, go to Step 5.

Step 4. Set $y = y + \delta$. Repeat steps 2-3. Keep increasing y by δ so long as $SAC > MC$.

If $SAC < MC$, set $\delta = \frac{1}{2}\delta$. If δ is less than ε , then stop. Otherwise set $y = y - \delta$.

Go to Step 2.

Step 5: Set $y = y - \delta$. Repeat steps 2-3. Keep reducing y by δ so long as $SAC < MC$.

If $SAC < MC$, set $\delta = \frac{1}{2}\delta$. If δ is less than ε , then stop. Otherwise set $y = y + \delta$.

Go to Step 2.

Figure 2 illustrates this search procedure. We begin with the output level y_0 . Here *SAC* exceeds *MC* and we move to output $y_1 (= y_0 + \delta)$. Here again, $SAC > MC$ and we increase output further by δ to move up to y_2 . At this point *MC* exceeds *SAC* signifying that we have gone past the minimum of the *SAC* curve. We now reduce output by $\frac{1}{2}\delta$ and reach y_3 where again *MC* is less than *SAC*. This implies that the (short run) average cost minimizing output lies between y_3 and y_2 . We move further to the right by adding $\frac{1}{4}\delta$ to y_3 and reach y_4 . Here again, *MC* exceeds *SAC* and we have overshoot the target. The desired output level lies between y_3 and y_4 . We select the mid point by subtracting $\frac{1}{8}\delta$ from y_4 . Because the step size has become very small, we terminate the process here.

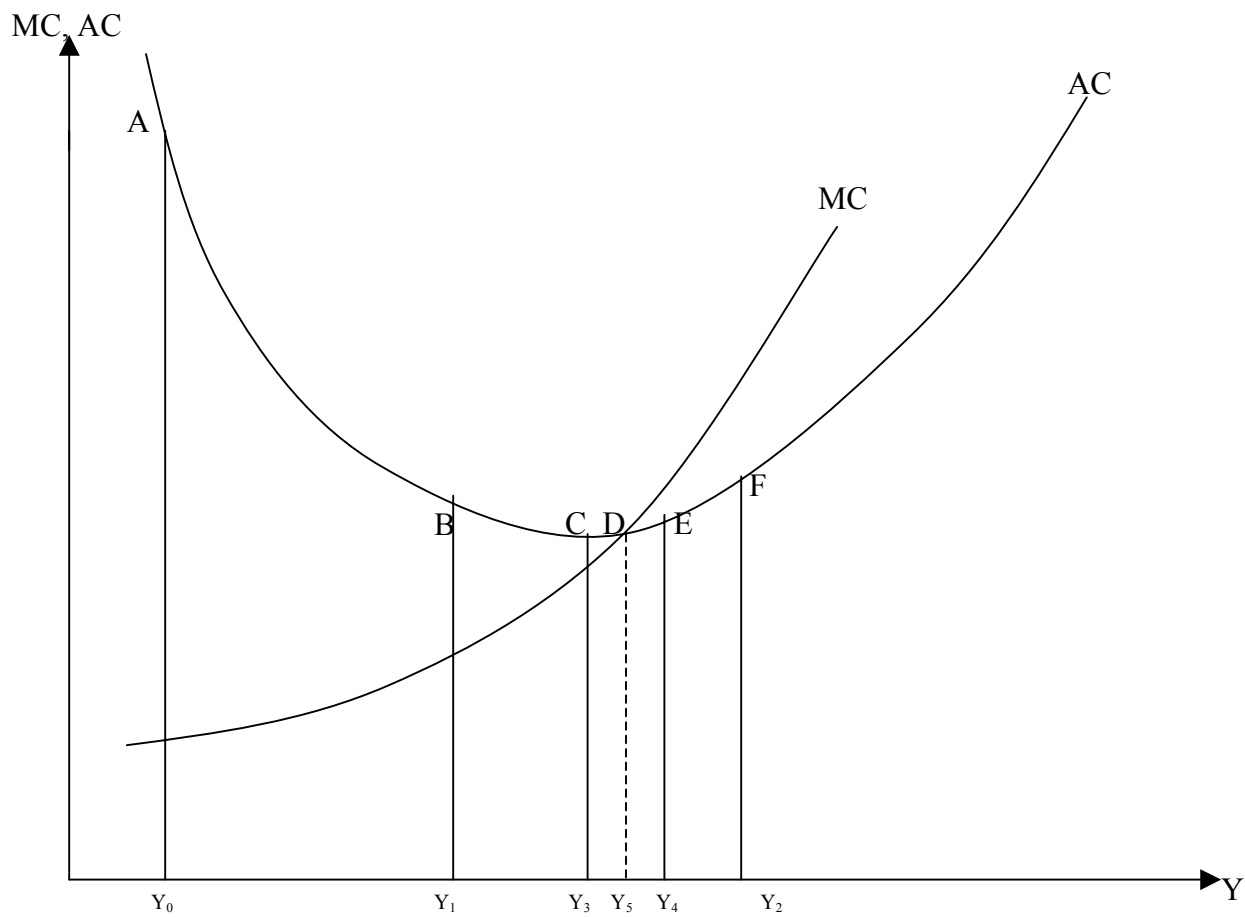


Figure 2.

Table 1. Actual and Capacity Levels of Output

YEAR	Q	Q _A	Q _B	CAPU _A	CAPU _B	CAPU _{FRB}
1968	57.7	60.562	61.227	0.95274	0.9424	87.1
1969	59.4	63.713	64.413	0.9323	0.92218	86.6
1970	56.5	62.164	67.183	0.90888	0.84099	79.4
1971	58.1	63.923	69.261	0.9089	0.83886	77.9
1972	63.3	63.3	71.339	1	0.88732	83.4
1973	67.8	67.8	75.079	1	0.90305	87.7
1974	66.1	71.803	79.511	0.92057	0.83133	83.4
1975	62.5	75.431	83.528	0.82857	0.74825	72.9
1976	68.2	78.434	86.853	0.86952	0.78524	78.2
1977	73.9	81.939	90.732	0.9019	0.81449	82.6
1978	77.8	85.439	94.61	0.91059	0.82232	85.2
1979	78.7	89.191	98.766	0.88238	0.79683	85.3
1980	75.3	93.444	103.476	0.80583	0.72771	79.5
1981	75.6	107.796	108.324	0.70132	0.69791	78.3
1982	72.7	100.699	111.51	0.72195	0.65196	71.8
1983	75.9	75.9	112.341	1	0.67562	74.4
1984	83.7	83.7	115.388	1	0.72538	79.8
1985	86	86	119.267	1	0.72107	78.8
1986	88.5	88.5	122.453	1	0.72273	78.7
1987	91.6	91.6	124.946	1	0.73311	81.3
1988	96.1	96.1	127.163	1	0.75572	83.8
1989	96.6	97.88	129.518	0.98692	0.74585	83.6
1990	97.3	100.288	132.703	0.97021	0.73322	81.4
1991	95.4	102.486	135.613	0.93086	0.70347	77.9
1992	100	104.684	138.521	0.95525	0.72191	79.4
1993	103.3	106.987	141.569	0.96553	0.72968	80.5
1994	108.7	109.395	144.755	0.99365	0.75092	82.5
1995	113.4	113.4	149.603	1	0.75801	82.6
1996	116.9	116.9	155.006	1	0.75417	81.5
1997	123.5	123.5	161.932	1	0.76267	82.4
1998	130.7	130.7	170.104	1	0.76835	80.9

¹ See footnote 5 in Klein's paper.

² Klein (1960) emphasized the welfare theoretic foundation of this measure of the capacity output. See also Chamberlin (1933), Kaldor (1935) and Paine (1936).

³ Färe, Grosskopf, and Kokkelenberg proposed their physical measure as an alternative to the cost-based measure.

⁴ An example of a parametric production function with a physical upper limit of the output quantity producible from a fixed input (K_0) even with unlimited quantities of the variable input (L) is

$$y = \frac{AK_0^\alpha}{1 + e^{-\beta L}}; A, \alpha, \beta > 0.$$

Here the output level asymptotically approaches (AK_0^α) as L goes to infinity.

⁵ One can easily apply the same analytical format to determine the optimal scale of production for any output vector in the multiple output case.

⁶ In the multiple output case, we look at the *ray average cost*

$$RAC(ty^0, w^0) = \frac{C(ty^0, w^0)}{t}.$$

The optimal output scale (t^*) is one that minimizes $RAC(ty^0, w^0)$.

⁷ For an assessment of the Federal Reserve's measure of capacity utilization, see Shapiro (1989) and also Klein and Long (1973).

⁸ The *actual average cost* need not be minimum, however, when allocative and technical inefficiencies exist.

References:

Berndt, E.R. and C.J. Morrison. "Capacity Utilization Measures: Underlying Economic Theory and an Alternative Approach"; *American Economic Review*, 71, 1981, pp 48-52.

Cassels, J.M. "Excess Capacity and Monopolistic Competition"; *Quarterly Journal of Economics*, May 1937, pp 426-443.

Chamberlin, E.H. *The Theory Of Monopolistic Competition* (Cambridge, MA: Harvard University Press; 1933).

Färe, R., S. Grosskopf, and E. Kokkelenberg. "Measuring Plant Capacity, Utilization, and Technical Change: A Nonparametric Approach"; *International Economic Review*, August 1989, pp 655-666.

Klein, L.R. "Some Theoretical Issues in the Measurement of Capacity"; *Econometrica*, Apr 1960; pp 272-286.

Klein, L.R. and V. Long. "Capacity Utilization: Concept, Measurement, and Recent Estimates"; *Brookings Papers on Economic Activity*; Issue 3 (1973) ; pp 743-756.

Kaldor, N. "Market Imperfection and Excess Capacity"; *Economica*; No. 5, Vol 2; 1935; pp 33-50.

Nelson, R.A. "On the Measurement of Capacity Utilization"; *Journal of Industrial Economics*, Vol 37, No. 3, March 1989; pp 273-286.

Paine, C.L. "Rationalisation and the Theory of Excess Capacity"; *Economica*, N. S. No. 9, Vol III, 1936; pp 46-55.

Shapiro, M. "Assessing the Federal Reserve's Measures of Capacity and Utilization"; *Brookings Papers on Economic Activity*; Issue 1 (1989); pp 181-225.