



University of
Connecticut

Department of Economics Working Paper Series

**Nonparametric Measures of Scale Economies and Capacity Utilization:
An Application to U.S. Manufacturing**

Subhash Ray
University of Connecticut

Working Paper 2013-09

March 2013

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

This working paper is indexed on RePEc, <http://repec.org>

**NONPARAMETRIC MEASURES OF SCALE ECONOMIES AND
CAPACITY UTILIZATION: AN APPLICATION TO U.S. MANUFACTURING**

Subhash C Ray

Department of Economics

University of Connecticut

Storrs CT 06269-1063

subhash.ray@uconn.edu

Abstract

An economic measure of scale efficiency is the ratio of the minimum average cost to the average cost at the actual output level of a firm. It is easily measured by the ratio of the total cost of this output under the constant and variable returns to scale assumptions. This procedure does not identify the output level where the average cost reaches a minimum. This paper proposes a nonparametric method of measuring this output level using DEA. The relation between this efficient production scale, the short run physical capacity output, and the most productive scale size (MPSS) is also discussed. An empirical application using state level data from U.S. manufacturing is used to illustrate the procedure.

Keywords: Efficient output; Most Productive Scale Size; Data Envelopment Analysis

JEL Classification: C61, L25, D24

**NONPARAMETRIC MEASURES OF SCALE ECONOMIES AND
CAPACITY UTILIZATION: AN APPLICATION TO U.S. MANUFACTURING**

Subhash C Ray

Department of Economics

University of Connecticut

Storrs CT 06269-1063

subhash.ray@uconn.edu

In standard microeconomic theory, the *capacity output* of a firm has been defined in several different ways. The simplest of them is the maximum level of output that can be produced from a given level of quasi-fixed inputs (like plant and machinery) even when variable inputs (like labor or materials) are available without restriction. By definition, the actual output produced cannot exceed this maximum quantity. This is a physical measure of capacity that is technologically determined. First proposed by Johansen (1968) it has been subsequently popularized in empirical applications by Färe, Grosskopf, and Kokkelenberg (1991). An *economic* interpretation of capacity is the output level where the average cost curve of the firm reaches a minimum. Here again, one needs to distinguish between the short run, where some inputs are fixed and the long run, where all inputs are variable. The presence of fixed costs associated with the quasi-fixed inputs of the firm justifies the U-shaped average cost curve and the output level where the short run average (total) cost reaches the minimum is the capacity level for the given bundle point of the (quasi) fixed inputs¹. In the long run, there are no fixed inputs and all inputs are freely adjusted in order to minimize the cost of producing a given output level. The presumed U-shape of the long run average cost curve results not from the presence of any fixed inputs but from

¹ In fact, Cassels (1937) argues “It is generally agreed that, since the absolute technical upper limit of output obtainable from the fixed factors is likely to lie far beyond the realm of practical economic operations, their capacity output should be taken as that at which the average full costs of production are at their minimum”.

economies of scale at smaller levels of output followed by diseconomies of scale at higher output levels. In the standard textbook analysis of long run equilibrium in a constant cost perfectly competitive industry, free entry and exit drive the market price to the level of the minimum long run average cost and each firm that remains in the industry produces the corresponding level of output. The long run competitive output level is considered to be the capacity output. This is an *economic* interpretation of capacity, which is determined by the position and curvature of the average cost curve and, for nonhomothetic technologies, on input prices as well. Economies of scale exist at all output levels below this capacity output. Diseconomies of scale set in once this benchmark output level is exceeded². Monopolistic competition is regarded as an inefficient market structure relative to perfect competition because although firms earn zero profit in the long run in both cases, only in the perfectly competitive market is the output produced at the level where the long run average cost is minimized³. It is considered to be socially wasteful because there remains *excess capacity* in the sense that further economies of scale remain unexploited⁴. In parametric models, one can determine the efficient output level from the estimated cost function by solving for the condition that the output elasticity of total cost should be unity at the efficient output scale⁵. In nonparametric analysis, however, this is not a feasible approach because there is no explicit cost function that can be differentiated. However, at the efficient output level, average cost attains a minimum and, hence, average and marginal costs are equal at this point. This, in its turn, implies *locally* constant returns to scale (CRS). Hence, even when variable returns to scale (VRS) holds across different levels of output, the minimum cost of producing this output level would be the same whether or not one assumed constant returns to scale *globally*.

² The presumed U-shape of the long run average cost curve has been questioned by many writers (e.g. Kaldor (1936) and Klein (1962)). When neither economies nor diseconomies of scale prevail at different output levels, the long run average cost curve is horizontal and there is no unique minimum point. Presence of fixed costs would still ensure the U-shape of the short run average cost curve. The capacity output level in such cases is defined by the point of tangency between the short run and the long run average cost curves. (Klein (1962), Berndt and Morrison (1981), Segerson and Squires (1990)). Even when scale economies and diseconomies are present at different output levels, it is possible that the average cost curve may have a 'flat bottom' in which case capacity output corresponds to an interval rather than a point on the average cost curve.

³ Of course, the concept of average cost is meaningful only in the context of a single output technology. For multiple outputs one must consider the minimum *ray average cost* for a given output-mix. In this paper we consider the single output case only.

⁴ Cassels (1937).

⁵ For multiple outputs, the partial elasticities of total cost with respect to the individual outputs must add up to unity.

In the Data Envelopment Analysis (DEA) literature, one compares the optimal solutions of the cost minimization problems under alternative returns to scale assumptions to determine whether or not a particular output level represents the efficient output scale. Of course, the minimum cost under the CRS assumption is never higher than that under the VRS assumption. The ratio of the two measures the *scale efficiency* at the observed output level. The measured level of scale efficiency provides useful information in two ways. First, when it is different from unity, one concludes that the output level under consideration does not constitute the efficient production scale. Second, its numerical value shows how much the average cost of the firm would be lowered if it changed its output to the optimal scale. But it does not identify the *output level*⁶ where the average cost curve attains a minimum.

This paper develops a simple procedure in DEA that lets one determine not only the minimum long run average cost but also the *level of output* where the average cost attains a minimum. When the actual output is smaller than this efficient output level, *economic* capacity utilization is less than 100%. The opposite is true, when actual output exceeds this benchmark. The paper unfolds as follows. Section 2 presents the methodological background for DEA and introduces the procedure that determines the efficient output level that can be used as benchmark for measuring capacity utilization. Section 3 provides an empirical application using state level data for total manufacturing production in the US from the 2007 Census of Manufactures. Section 4 is the summary.

2. The Methodological Background

Consider an industry where firms produce a scalar output y using input bundles $x \in R^n_+$. An input-output bundle (x, y) is considered feasible when y can be produced from x . The set of all feasible input-output bundles constitutes the production possibility set

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

Suppose that (x^j, y_j) is the observed input-output bundle⁷ of firm j ($j = 1, 2, \dots, N$). Obviously, every observed input-output bundle is feasible. Under the fairly non-restrictive assumptions that the

⁶ There may be a range of outputs where the average cost remains a minimum. In such cases, one would like to identify the smallest and the largest output levels where the minimum average cost is attained.

⁷ In this paper vectors are identified by superscripts and scalars by a subscripts.

production possibility set is convex and that both inputs and outputs are freely disposable, one can approximate the set T by the free disposal convex hull of the observed bundles

$$S = \left\{ (x, y) : x \geq \sum_1^N \lambda_j x^j; y \leq \sum_1^N \lambda_j y_j; \sum_1^N \lambda_j = 1; \lambda_j \geq 0 (j = 1, 2, \dots, N) \right\}. \quad (2)$$

If, additionally, one assumes CRS, for any $(x, y) \in T$, for any $k \geq 0$, $(kx, ky) \in T$.

Correspondingly the empirical construct of the CRS production possibility set would be

$$S^C = \left\{ (x, y) : x \geq \sum_1^N \lambda_j x^j; y \leq \sum_1^N \lambda_j y_j; \lambda_j \geq 0 (j = 1, 2, \dots, N) \right\}. \quad (3)$$

An alternative characterization of the technology is possible in terms of the input requirement sets. For any specific output level (y_0) , the input requirement set consists of all input bundles that can produce y_0 . This set can be expressed as

$$V(y_0) = \{x : (x, y_0) \in T\}. \quad (4)$$

Empirically, the input requirement sets will be

$$\hat{V}(y_0) = \left\{ x : x \geq \sum_1^N \lambda_j x^j; y_0 \leq \sum_1^N \lambda_j y_j; \sum_1^N \lambda_j = 1; \lambda_j \geq 0 (j = 1, 2, \dots, N) \right\} \quad (5)$$

under the assumption of variable returns to scale (VRS) and

$$\hat{V}^C(y_0) = \left\{ x : x \geq \sum_1^N \lambda_j x^j; y_0 \leq \sum_1^N \lambda_j y_j; \lambda_j \geq 0 (j = 1, 2, \dots, N) \right\} \quad (6)$$

if CRS is assumed.

When VRS is assumed, the minimum cost of producing any output y at input price (vector) w^0 is

$$C^{VRS}(w^0, y) = \min w^{0'} x : x \in V(y). \quad (7a)$$

Similarly, under CRS, the minimum cost is

$$C^{CRS}(w^0, y) = \min w^{0'} x : x \in V^C(y). \quad (7b)$$

In empirical applications, the minimum cost of output y at input prices w^0 under the VRS assumption is estimated as⁸

⁸ Under the VRS assumption the minimum cost can be estimated only if y does not exceed the largest observed output quantity.

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

The corresponding average cost is

$$AC^{VRS}(w^0, y) = \frac{C^{VRS}(w^0, y)}{y}. \tag{9}$$

In order to find the output level where the VRS average cost reaches a minimum one needs to solve the problem:

$$\begin{aligned}
\min \frac{w^{0'} x}{y} \\
s.t. \sum_j \lambda_j x^j &\leq x; \\
\sum_j \lambda_j y_j &\geq y; \\
\sum_j \lambda_j &= 1; \\
\lambda_j &\geq 0; (j = 1, 2, \dots, N).
\end{aligned} \tag{10}$$

The optimal input-output bundle (x^*, y_*) from (10) above yields the minimum average cost

$$AC^{VRS}(w^0, y_*) = \frac{C^{VRS}(w^0, y_*)}{y_*} = \frac{w^{0'} x^*}{y_*} \tag{11}$$

along with the efficient output level (y_*) .

Unlike standard DEA problems, the minimization problem in (10) above is a non-linear programming problem and can be computationally quite burdensome. However, one can utilize the following two lemmas from production economics to determine both the output level y_* and the minimum average cost $AC^{VRS}(w^0, y_*)$ by solving a simple DEA LP problem.

Lemma 1. Locally Constant Returns to Scale holds at the input-output bundle (x^, y_*) where the average cost reaches a minimum.*

Proof: Suppose this is not true. Then there exists a pair of non-negative scalars (α, β) such that $(\beta x^*, \alpha y_*) \equiv (x^{**}, y^{**}) \in T$ and $\alpha > \beta$. Note that the average cost at (x^{**}, y^{**}) is

$$AC^{VRS}(w^0, y_{**}) = \frac{w^{0'} x^{**}}{y_{**}} = \frac{\beta w^{0'} x^*}{\alpha y_*} = \frac{\beta}{\alpha} AC^{VRS}(w^0, y_*).$$

But, because $\alpha > \beta$ by assumption, this would imply that $AC^{VRS}(w^0, y_{**}) < AC^{VRS}(w^0, y_*)$. This contradicts the assumption that average cost reaches a minimum at (x^*, y_*) . Hence it is proven.

Lemma 2 If the technology exhibits Constant Returns to Scale globally, average cost is a constant at all output levels.

Proof: Consider two different output levels, y_0 and y_1 . Suppose that at some vector of input prices w the input bundle x_0^* minimizes the cost of producing y_0 . Hence, $C^{CRS}(w, y_0) = w'x_0^*$ and $AC^{CRS}(w, y_0) = \frac{w'x_0^*}{y_0}$. Similarly, x_1^* minimizes the cost of producing y_1 . $C^{CRS}(w, y_1) = w'x_1^*$ and $AC^{CRS}(w, y_1) = \frac{w'x_1^*}{y_1}$. Now suppose that $AC^{CRS}(w, y_0) \neq AC^{CRS}(w, y_1)$ and assume (arbitrarily) that $AC^{CRS}(w, y_0) > AC^{CRS}(w, y_1)$. Define $\rho = \frac{y_0}{y_1}$. We know that (x_1^*, y_1) is a feasible input-output bundle. Now, because CRS holds, $(\rho x_1^*, \rho y_1)$ is also feasible. Denote the bundle ρx_1^* as x^2 . Further, ρy_1 equals y_0 . That is, (x^2, y_0) is feasible. Now, if $AC^{CRS}(w, y_0) > AC^{CRS}(w, y_1)$, $\frac{w'x_0^*}{y_0} > \frac{w'x_1^*}{y_1} = \frac{w'\rho x_1^*}{\rho y_1} = \frac{w'x^2}{y_0}$. That is, the input bundle x^2 can produce the same output level y_0 at a lower cost than the bundle x_0^* . In that case, x_0^* could not have minimized the cost of producing the output y_0 . Hence it is proven.

Finding the Efficient Output Scale and Measuring Capacity Utilization

Suppose that firm k faces the input prices w^k and produces the output quantity y_k . Our objective is to find the output level $y_k^* = y(w^k)$ that minimizes $AC^{VRS}(w^k, y)$.

For this we first consider the CRS cost minimization problem

$$\begin{aligned} \min \quad & w^{k'} x \\ \text{s.t.} \quad & \sum_j \lambda_j x^j \leq x; \\ & \sum_j \lambda_j y_j \geq y_k; \\ & \lambda_j \geq 0. \end{aligned} \tag{12}$$

Suppose that at the optimal solution

$$\sum_j \lambda_j^* = q. \text{ and } \sum_j \lambda_j^* x^j = x^*.$$

This implies that $w^{k'} x^* = C^{CRS}(w^0, y_k)$ is the minimum cost of producing y_k at input prices w^k under the CRS assumption. Then, by lemma 2, $\frac{1}{q} w^{k'} x^* = \frac{1}{q} C^{CRS}(w^k, y_k) = C^{CRS}(w^0, \frac{1}{q} y_k)$.

Define the output level $\bar{y} = \frac{1}{q} y_k$ and the weights $\mu_j^* = \frac{1}{q} \lambda_j^*$. Clearly, $\sum_j \mu_j^* = 1$ and each $\mu_j^* \geq 0$.

Further, define the input vector $\tilde{x} = \frac{1}{q} x^* = \sum_j \mu_j^* x^j$.

Now consider the VRS cost minimization problem

$$\begin{aligned} C^{VRS}(w^k, \bar{y}) &= \min w^{k'} x \\ \text{s.t. } \sum_j \lambda_j x^j &\leq x; \\ \sum_j \lambda_j y_j &\geq \bar{y}; \\ \sum_j \lambda_j &= 1; \\ \lambda_j &\geq 0. \end{aligned} \quad (13)$$

Suppose that the optimal solution of (13) yields a minimum cost $w^{k'} \bar{x}$. It can be seen from above that the μ_j^* s and \tilde{x} from (12) constitute a feasible solution for (13). Hence,

$$w^{k'} \bar{x} \leq w^{k'} \tilde{x}. \quad (14)$$

At the same time, $w^{k'} \tilde{x} = \frac{1}{q} w^{k'} x^* = C^{CRS}(w^k, \bar{y})$ is the optimal solution of the less restrictive CRS problem

$$\begin{aligned} \min w^{k'} x \\ \text{s.t. } \sum_j \lambda_j x^j &\leq x; \\ \sum_j \lambda_j y_j &\geq \bar{y}; \\ \lambda_j &\geq 0. \end{aligned} \quad (15)$$

Hence,

$$w^{k'} \tilde{x} \leq w^{k'} \bar{x}. \quad (16)$$

The inequalities together imply

$$w^{k'} \tilde{x} = w^{k'} \bar{x}. \quad (17)$$

Therefore, at the output level \bar{y} the total and average costs are the same under both CRS and VRS assumptions. That is at the output level \bar{y} the average cost $AC^{VRS}(w^k, y)$ reaches a minimum. In other words, \bar{y} is an efficient production scale.

Recall, however, that $\bar{y} = \frac{1}{q} y_k$. Hence, an economic measure of capacity utilization is

$$CU = \frac{y_k}{\bar{y}} = q. \quad (18)$$

In order to obtain the capacity output one only needs to solve the CRS cost minimization problem and from the optimal values of the λ s compute

$$\bar{y} = \frac{y_k}{q} = \frac{y_k}{\sum_j \lambda_j^*}. \quad (19)$$

Comparison with Banker's Approach for Determining Returns to Scale

The method proposed in (12) and (19) above is very similar to Banker's (1984) primal approach to returns to scale determination. However, there is a subtle difference. In Banker's approach, for specific input-output combination (x^k, y_k) , one solves the following input-oriented CCR problem

$$\begin{aligned} & \min \theta \\ & s.t. \sum_j \lambda_j x^j \leq \theta x^k; \\ & \quad \sum_j \lambda_j y_j \geq y_k; \\ & \quad \lambda_j \geq 0 \quad (j = 1, 2, \dots, N). \end{aligned} \quad (20)$$

From the optimal solution of (20), one computes $q = \sum_j \lambda_j^*$, $\beta = \frac{\theta^*}{q}$, and $\alpha = \frac{1}{q}$, then, as shown

by Banker (1984), $(\beta x^k, \alpha y_k)$ is an MPSS. Note that this optimization problem is totally independent of input prices. Comparing (12) with (20) one can easily see that there is no reason why αy_k would have to be the same as the output level \bar{y} obtained from (19). The fundamental difference is that for Banker's MPSS, the input-mix in the observed bundle x^k is held constant and only radial changes are considered⁹. By contrast, in (12) the input-mix is not predetermined and is obtained from the optimal solution of the problem.

The Short Run Capacity Output as a Physical Limit

⁹ Minor adjustments for slacks at the optimal solution of the MPSS problem do not alter α .

It would be useful at this point to compare this economic measure of the capacity output with the physical capacity measure defined by Färe, Grosskopf, and Kokkelenberg (FGK) (1991).

Following Johansen (1968), FGK define the capacity output as the maximum quantity of output that a firm can produce given its (vector of) quasi-fixed inputs¹⁰ *even if unlimited quantities of the variable inputs were available*. Suppose that input vector is partitioned as $x = (v, f)$, where v is the sub-vector of variable inputs while f is the sub-vector of fixed inputs. Then for any given bundle of fixed inputs f^0 , capacity out is defined by FGK as

$$y^{**}(f^0) = \max y : (v, f^0, y) \in T. \quad (21)$$

In empirical applications, the FGK capacity output is measured as

$$\begin{aligned} y^{**} &= \varphi^{**} y_0 \text{ where} \\ \varphi^{**} &= \max \varphi \\ \text{s.t. } \sum_j \lambda_j y_j &\geq \varphi y_0; \\ \sum_j \lambda_j v^j &\leq v; \\ \sum_j \lambda_j f^j &\leq f^0; \\ \sum_j \lambda_j &= 1; \\ v \geq 0; \lambda_j &\geq 0 (j = 1, 2, \dots, N). \end{aligned} \quad (22)$$

Note that apart from non-negativity the variable inputs are unconstrained and, in consequence, play no role in the optimization problem in (21). Recall that the technically efficient output producible from the given levels of variable and fixed inputs (v^0, f^0) is

$$y^* = \varphi^* y_0 \text{ where}$$

¹⁰ Ray, Mukherjee, and Wu (2006) consider working capital (measured by a firms's total expenditure on variable inputs) as an additional quasi-fixed input.

$$\begin{aligned}
\varphi^* &= \max \varphi \\
s.t. \sum_j \lambda_j y_j &\geq \varphi y_0; \\
\sum_j \lambda_j v^j &\leq v^0; \\
\sum_j \lambda_j f^j &\leq f^0; \\
\sum_j \lambda_j &= 1; \\
v \geq 0; \lambda_j &\geq 0 (j = 1, 2, \dots, N).
\end{aligned} \tag{23}$$

FGK defined capacity utilization as $\frac{y^*}{y^{**}} = \frac{\varphi^*}{\varphi^{**}}$. It is clear from above that $\varphi^{**} \geq \varphi^*$ and, therefore, their measure of capacity utilization can never exceed unity. By contrast, the capacity utilization rate defined above in (18) can be greater than unity. This is because, while FGK's capacity output represents a *physical upper bound* that cannot be exceeded given a specific bundle of fixed inputs f^0 , the efficient output level where average cost reaches a minimum can easily be exceeded by a firm because there are no fixed inputs and all inputs are choice variables. As Cassels (1937) pointed out the economic capacity output refers to capacity of all inputs rather than any specific set of fixed inputs.

Minimum and Maximum Efficient Production Scales

The capacity output level \bar{y} define above will not be unique unless the DEA LP problem (12) above has a unique optimal solution. When multiple optimal solutions exist, even though the objective function would have the same minimum value, the λ s (and q , the sum of these λ s) would differ from one solution to another. In that case, we would need to determine the smallest and the largest value of q across the alternative optimal solutions. Let q_L be the smallest and q_H the largest value of the sum of the λ s across these alternative solutions. Then, $\bar{y}_L = \frac{y_k}{q_H}$ is the smallest efficient scale of output. On the other hand, $\bar{y}_H = \frac{y_k}{q_L}$ is the largest efficiency output scale beyond which diseconomies of scale will apply. Over the interval $[\bar{y}_L, \bar{y}_H]$ the average cost curve would be flat.

In order to determine the smallest efficient output scale $\bar{y}_L = \frac{y_k}{q_H}$ one needs to solve the following auxiliary LP problem:

$$\begin{aligned}
q_H &= \max \sum_j \lambda_j \\
s.t. \sum_j \lambda_j x^j &= x; \\
\sum_j \lambda_j y_j &\geq y_k; \\
w^{0'} x &= w^{0'} x^*; \\
\lambda_j &\geq 0
\end{aligned} \tag{24}$$

In this problem the constraint $w^0 kx = w^{k'} x^*$ ensures that only the λ s from the optimal solution of (12) would be eligible. In an analogous manner the largest efficient production scale $\bar{y}_L = \frac{y_k}{q_L}$ can be obtained from the solution of the following LP problem:

$$\begin{aligned}
q_L &= \min \sum_j \lambda_j \\
s.t. \sum_j \lambda_j x^j &= x; \\
\sum_j \lambda_j y_j &\geq y_k; \\
w^{0'} x &= w^{0'} x^*; \\
\lambda_j &\geq 0; (j = 1, 2, \dots, N).
\end{aligned} \tag{25}$$

Capacity Utilization in U.S. Manufacturing

In this empirical application we use output quantity and input quantity and price data for each of the 48 continental states in the U.S. constructed from the Census of Manufactures for the year 2007. For this, we conceptualize a 1-output, 5-input production technology and assume that there is no difference in the technology across the states in any given year.

One problem with the Census of Manufactures data is that the reported output and input quantities are aggregates over all establishments covered by the survey in that state. This total input-output bundle is *not an observed data point*. However, the unit-level input-output quantities (although not reported individually) are all feasible input-output bundles. The total input-output bundle is necessarily feasible only if the underlying production technology is assumed to be additive. This, in its turn, implies constant returns to scale. But, in that case, the question of the efficient production scale becomes

irrelevant. However, the average (or per establishment) input-output quantities for that state is, in effect, an equally weighted convex combination of the actual (though unreported) input-output bundles of the individual units and is feasible even under the VRS assumption so long as the production possibility set is convex. In this empirical application *per establishment* data from each state is treated as the *observation* from that state. The production possibility set is constructed as the free disposal convex hull of these points.

Given the fact that the market for manufactured goods is nationally integrated, we assume that the output price does not vary across states so that the value of output is a reasonable measure of the quantity produced. Input prices, however, do vary across states. The inputs included are (a) production labor (L_1), (b) non-production labor (L_2), (c) capital (K), (d) energy (E), and (e) materials (M). Output is measured by the gross value of production. Production labor is measured by the number of hours worked (in thousands). The corresponding input price is wage paid per hour to production workers (w_1). The other labor input is the number of non-production employees. The corresponding wage rate (w_2) is total annual emolument per employee. The capital input is the average of beginning-of-the year and end-of the year (nominal) values of gross fixed assets (millions of dollars). The capital input price (i.e. user cost), p_K , is measured by the sum of depreciation, rent, and interest expenses per dollar of gross value of capital. The quantity of the energy input (E) is constructed by deflating the expenditure on purchased fuels and electricity by state specific energy price (p_E). Total expenditure on materials, parts, and containers is used as a measure of the materials input quantity (M). By implication, materials price (p_M) was set equal to unity (when measured in current dollars) for every state.

Summary statistics about the input and output quantities along with the input prices are reported in Table 1. Gross value of output per establishment varies between a maximum of \$47.466 million in Louisiana and a minimum of \$ 5.224 million in Rhode Island. The average of gross output per establishment across the 48 continental states was \$13.280 million. The coefficient of variation was 0.544. Number of production workers per establishment was the highest in Mississippi (96.65) and the lowest in Montana (20.23). The US average was 58.67. The highest number of non-production employees was

highest in Kansas (15.69) and lowest in Montana (4.33). The US average was 11.67. Consistent with the size of gross output per establishment, energy consumption was the highest in Louisiana and the lowest in Rhode Island. Hourly wage rate of production workers was a maximum of \$22.39 in Michigan and a minimum of \$15.62 in Arkansas with the US average of \$18.62. For non-production workers, the annual salary (including benefits) per employee was \$76.27 thousand in New Jersey at the high end and \$48.74 thousand in Montana at the low end. Price of energy per million btu was the highest (\$21.61) in Massachusetts and the lowest (\$7.56) in North Dakota. User cost of capital (per dollar of gross value of fixed assets) had an average of 17 cents across the states with a maximum of 21 cents per dollar in Nevada and 13 cents per dollar in Wyoming.

Table 2 reports the actual output (y_j), the scale efficient output $\bar{y}(w^j)$, capacity utilization rate (CU), scale efficiency (SE) and related information about each individual state (j). The efficient output level ($\bar{y}(w^j)$) was found to be \$9.415 million for every state. Differences in the input prices did alter the long run average cost curves but did not change the output level where these different average cost curves reach their respective minima. For a parametric and differentiable cost function, this would suggest that the total cost function is multiplicatively separable in the input prices and the output quantity¹¹. Such multiplicative separability of the cost function implies that the underlying technology is *homothetic*. In the case of DEA, the empirically constructed isoquants are

¹¹ Suppose that the total cost curve can be written as

$$C(w; y) = a(w).b(y),$$

where w is the vector of input prices and y is the output quantity. In that case,

$$AC(w, y) = \frac{C(w, y)}{y} = a(w) \cdot \left[\frac{b(y)}{y} \right].$$

It is clear that the first order condition for a minimum of the average cost is

$$\frac{yb'(y) - b(y)}{y^2} = 0.$$

This, in turn, implies

$$b'(y) = \frac{b(y)}{y}.$$

This last equation is totally independent of w . Hence, changes in input prices would not change the output level where the average cost reaches a minimum..

piece-wise linear and do not have the smooth convex shape found in textbook illustrations. As a result, the optimal point is not sensitive to (even moderate) changes in relative input price. Hence, any conclusion about homotheticity of the technology remains merely speculative.

Comparison of the actual output per establishment with the benchmark output of \$9.415 million yields the capacity utilization rate (CU) reported in Table 2. Connecticut is one state where output per establishment is at the efficient scale. Of the 16 states with capacity utilization rates below unity, 6 states (Rhode Island, Florida, Nevada, Montana, Colorado, and New York) show capacity utilization rates below 75%. At the other extreme, the output per establishment in Louisiana is over 5 times the efficient production scale and in Delaware it is over three times as large as the optimal scale.

It is important to note that although the optimal production scale ($\bar{y}(w^j)$), where the average cost attains a minimum at the applicable input prices is the same for all of the states, there is considerable variation in the minimum average cost itself across states due to input price differences. At the lower end it is 57.3 cents per dollar of gross output in Mississippi and at the upper end it is slightly over 64.9 cents per dollar in Massachusetts. The actual output (y_j) (in millions of dollars), the efficient total cost $C(w^j, y_j)$ and average cost $C(w^j, y_j)$ at the observed output level, and the minimum average cost $C(w^j, y_j)$ are also reported in Table 2. Comparison of the average cost at the optimal and the actual output levels yields the measure of scale efficiency (SE). While the capacity utilization rate varies considerably across the states, far less variation can be found in levels of scale efficiency. An interesting point to note is that the same level of scale efficiency is observed for two states located at two completely different segments of their respective average cost curves. Take the example of Florida and Delaware. Both states have scale efficiency of 0.88. That is, the minimum average cost (at the optimal scale) would be about 88% of the average cost at their observed levels of output. But this does not reveal that fact that the actual output in Florida (\$5.795 million) is only 61.5% of the optimal output (\$9.415 million) whereas in Delaware the actual output (\$29.842 million) is 317% of the optimal output. By the same token, the capacity utilization rate only identifies how far below or above the efficient production scale does a firm operate. It does not measure

the extent of scale economy that can be exploited by altering the output level appropriately. In this sense, CU and SE provide complementary information about the operational efficiency of a firm. Together they also provide a (gross) measure of the elasticity of the average cost with respect to a change in output. Take the case of Florida. At the actual output of \$5.795 million the average cost is 67.7 cents per dollar whereas at the optimal scale (\$9.415 million) the average cost would fall to 59.7 cents per dollar. This is a \$3.62 million (62.46%) increase in output and a 8 cent (11.8%) decline in average cost. Broadly, it implies an output elasticity of average cost equal to -0.189. For Delaware, the change in output would be a decline of \$20.427 million (68.45%) and the decline average cost would be 8.6 cents (11.91%). The resulting elasticity of average cost would be 0.174. As a general rule, let y_0 and y^* be the actual and optimal levels of output and, correspondingly, AC_0 and AC^* be the actual and optimal levels of average cost. Then the proportionate changes in average cost and output are

$$\frac{\Delta AC}{AC_0} = \frac{AC^* - AC_0}{AC_0} = (SE - 1) \text{ and}$$

$$\frac{\Delta y}{y_0} = \frac{y^* - y_0}{y_0} = \left(\frac{1}{CU} - 1\right) = \frac{1 - CU}{CU}.$$

Thus the output elasticity of average cost is

$$\varepsilon_y^{AC} = \frac{SE - 1}{\frac{1 - CU}{CU}}. \quad (26)$$

The numerator is less than or equal to 0 by definition. When CU is less than unity, the denominator is positive and elasticity is negative. This is consistent with positive scale economies at output levels below the capacity output. Similarly, the elasticity is positive for output above the capacity level signifying that output and average cost both change in the same direction (i.e., decline).

Table 3 reports two other measures of the capacity output along with the associated measures of capacity utilization. The first is the short run capacity output (y^{**}) from (22) due to FGK. The ratio of this actual output (y_0) and this capacity output is the *gross* capacity utilization measure (fgk_cu) shown in Table 3. It should be noted that it incorporates both technical inefficiency (wasted capacity) and under-utilization of capacity (unutilized capacity). For 7 out of the 48 states (Colorado, Florida, Montana,

Nevada, New Mexico, New York, and Rhode Island) the short run (maximum) capacity output is smaller than the economic (efficient) capacity. It needs to be remembered, however, that the FGK capacity output for any state is the maximum output producible given the quantity of its quasi-fixed input whereas for the efficient output level all inputs are optimally chosen and the quantity of quasi-fixed input may exceed the observed level if that leads to a lower average cost. In 4 cases (Idaho, Maine, West Virginia, and Wyoming) the FGK capacity utilization rate was less than 50% and in 2 others (Rhode Island and Vermont) it was below 60%. At the other end, in 5 cases (Delaware, Kansas, Louisiana, Montana, and New Mexico) short run capacity utilization was 100%. In 2 other cases (Nebraska and South Dakota) the rate of capacity utilization exceeded 90%. The other measure of capacity output reported in Table 3 is the output level at Banker's MPSS (obtained from (20) above). The corresponding capacity utilization rate is reported for the individual states in the column identified as 'Ratio'. For 4 states (Connecticut, Maine, Oregon, and Rhode Island), Banker's MPSS output level is smaller than the efficient output scale. However, in all of these cases, the difference between the two is really small. Especially for Connecticut, the two are almost identical. In 8 cases (Alabama, Indiana, Iowa, Kentucky, South Carolina, Tennessee, and West Virginia) the actual output exceeded the MPSS. For Alabama, however, the two were almost identical. In 8 other states (Connecticut, Delaware, Kansas, Louisiana, Massachusetts, North Carolina, Oregon, and Wyoming) the actual output was at the MPSS. On average, states were utilizing 75.2% of their short run capacity (the FGK measure) and were producing 83.5% of their respective MPSS output levels. However, they were producing 41% above the optimal output level and were experiencing diseconomies of scale. Correlation between the economic capacity utilization rate and the MPSS ratio was about 52%. Correlation between the economic and physical (short run) capacity utilization rates was even lower (32.7%). As already noted, the FGK measure of capacity utilization is lower than 1 by construction. The other two measures, by contrast, allow capacity utilization beyond 100%. A closer look at the results reported in Table 2 shows that 16 of the 48 states were operating in the downward sloping segment of their average cost curves and a single state (Connecticut) was at the efficient production scale. The remaining 31 states were experiencing diseconomies of scale. As can be found from Table 3, many of these

states were found to be operating well below their MPSS output levels. For example, the rate of economic capacity utilization (CU in Table 2) in Texas was 2.38. Thus, it was way up in the region of diseconomies of scale along the average cost curve. However, as shown by the 'Ratio' in Table 3, it was found to be producing only 82% of the MPSS output level and was in the region of increasing returns to scale along the VRS frontier. Such conflicting evidence arises because of the fact that while the MPSS analysis keeps the observed input-mix unchanged, in determining the economic capacity level one changes the input-mix to attain full allocative efficiency. For practical policy, the economic capacity utilization measure is a better guide because when input prices are available, changing the production scale while keeping a sub-optimal input mix makes no sense. A state where the average firm is experiencing diseconomies of scale, the right policy would be to encourage a reorganization of the industry with an increased number of smaller firms. The opposite would be true where the average firm is found to be experiencing economies of scale. There the policy should be to provide incentives for consolidation of firms so that at a larger output scale, the average firm can exploit the economies of scale to the fullest extent.

One needs to end with a note of caution. The purpose of the empirical analysis is primarily to provide an application of the methodology proposed in the paper rather than to serve as an in-depth analysis of the U.S. manufacturing sector. For any serious policy prescription, the conclusions drawn from the empirical analysis should be validated by a more disaggregated analysis with establishment level data.

Summary

There are alternative ways to measure the rate of capacity utilization. The short run capacity output level as a physical upper bound on the output producible from given level of quasi-fixed inputs is easily conceptualized but has limited economic content. It is not clear why a firm would want to reach this capacity output without regard to the cost of the variable inputs that might be required to attain this goal. The MPSS output level does maximize ray average productivity and as a benchmark has an optimality property. However, it keeps the input mix unchanged and, hence, ignores allocative inefficiency. This can be particularly important for non-homothetic technologies where the expansion

path of a firm differs from a ray through the origin. The capacity output where the average cost reaches a minimum is economically meaningful and is well accepted as a benchmark in the literature. The simple empirical example suggests that in 2007 in the U.S. the average manufacturing firm in most states were operating in the region of diseconomies of scale and it would be economically efficient to break them into smaller firms reducing the average firm size.

Table 1 Summary Statistics

Variables	Mean	minimum	maximum
Output (Y) (dollars millions)	13.281	5.225	47.466
Production Labor (L1) (thousands of hrs)	58.67	20.23	96.65
Non-Production Labor (L2) (number of persons)	11.67	4.33	15.69
Energy (E) (Billions of Btu)	31.91	5.87	186.12
Materials (M) (\$ millions)	6.70	2.16	31.03
Capital (K) (\$ millions)	6.26	2.45	19.04
Production Labor wage (\$/hr)	18.62	15.62	22.39
Non-production Labor salary (thousands of dollars per year)	62.69	48.74	76.27
Price of energy (\$/million Btu)	12.41	7.56	21.61
Materials Price (\$)	1.00	1.00	1.00
User cost of capital (per \$ of gross value of fixed assets)	0.17	0.13	0.21

Table 2 Scale Efficiency and Capacity Utilization Rates

State	y_j	$C(w^j, y_j)$	$AC(w^j, y_j)$	$y^*(w^j)$	$C^*(w^j, y_j^*)$	$AC(w^j, y_j^*)$	CU	SE
AL	18.209	11.417	0.627	9.415	10.714	0.588	1.934	0.938
AZ	9.098	5.732	0.630	9.415	5.707	0.627	0.966	0.996
AR	15.573	9.249	0.594	9.415	8.946	0.574	1.654	0.967
CA	8.833	5.591	0.633	9.415	5.545	0.628	0.938	0.992
CO	6.972	4.541	0.651	9.415	4.346	0.623	0.741	0.957
CT	9.415	6.087	0.647	9.415	6.087	0.647	1.000	1.000
DE	29.842	21.552	0.722	9.415	18.968	0.636	3.170	0.880
FL	5.795	3.924	0.677	9.415	3.459	0.597	0.615	0.882
GA	13.155	7.840	0.596	9.415	7.715	0.586	1.397	0.984
ID	7.426	4.681	0.630	9.415	4.522	0.609	0.789	0.966
IL	13.038	8.116	0.622	9.415	8.061	0.618	1.385	0.993
IN	19.663	12.718	0.647	9.415	11.814	0.601	2.089	0.929
IA	20.445	13.367	0.654	9.415	12.209	0.597	2.172	0.913
KS	19.302	12.603	0.653	9.415	11.716	0.607	2.050	0.930
KY	22.637	15.180	0.671	9.415	13.575	0.600	2.404	0.894
LA	47.466	38.108	0.803	9.415	29.437	0.620	5.042	0.772
ME	7.101	4.412	0.621	9.415	4.218	0.594	0.754	0.956
MD	8.877	5.587	0.629	9.415	5.545	0.625	0.943	0.992
MA	8.818	5.775	0.655	9.415	5.724	0.649	0.937	0.991
MI	13.576	8.649	0.637	9.415	8.578	0.632	1.442	0.992
MN	10.701	6.624	0.619	9.415	6.601	0.617	1.137	0.997
MS	18.277	11.294	0.618	9.415	10.474	0.573	1.941	0.927
MO	13.213	8.123	0.615	9.415	7.997	0.605	1.403	0.984
MT	6.455	4.083	0.632	9.415	3.819	0.592	0.686	0.935

Table 2 Scale Efficiency and Capacity Utilization Rates (continued)

State	y_j	$C(w^j, y_j)$	$AC(w^j, y_j)$	$\bar{y}(w^j)$	$C^*(w^j, y_j^*)$	$AC(w^j, y_j^*)$	CU	SE
NE	16.050	9.678	0.603	9.415	9.351	0.583	1.705	0.966
NV	6.125	4.138	0.676	9.415	3.842	0.627	0.651	0.928
NH	7.066	4.727	0.669	9.415	4.531	0.641	0.751	0.959
NJ	10.025	6.401	0.639	9.415	6.398	0.638	1.065	1.000
NM	8.706	5.374	0.617	9.415	5.312	0.610	0.925	0.988
NY	6.921	4.432	0.640	9.415	4.225	0.610	0.735	0.953
NC	16.021	9.662	0.603	9.415	9.408	0.587	1.702	0.974
ND	11.787	6.930	0.588	9.415	6.835	0.580	1.252	0.986
OH	14.443	9.011	0.624	9.415	8.877	0.615	1.534	0.985
OK	12.222	7.378	0.604	9.415	7.281	0.596	1.298	0.987
OR	9.262	5.655	0.611	9.415	5.642	0.609	0.984	0.998
PA	12.120	7.456	0.615	9.415	7.382	0.609	1.287	0.990
RI	5.225	4.000	0.766	9.415	3.168	0.606	0.555	0.792
SC	17.207	10.645	0.619	9.415	10.186	0.592	1.828	0.957
SD	9.840	5.767	0.586	9.415	5.748	0.584	1.045	0.997
TN	16.490	10.051	0.610	9.415	9.682	0.587	1.751	0.963
TX	22.478	15.344	0.683	9.415	13.956	0.621	2.388	0.910
UT	10.065	6.097	0.606	9.415	6.079	0.604	1.069	0.997
VT	7.668	4.834	0.630	9.415	4.686	0.611	0.814	0.969
VA	12.678	7.682	0.606	9.415	7.606	0.600	1.347	0.990
WA	11.743	7.372	0.628	9.415	7.353	0.626	1.247	0.997
WV	14.133	8.611	0.609	9.415	8.470	0.599	1.501	0.984
WI	13.431	8.307	0.619	9.415	8.213	0.612	1.427	0.989
WY	11.887	7.219	0.607	9.415	7.146	0.601	1.263	0.990

Table 3 Alternative Measures of Capacity Output and Capacity Utilization

State	actual y	y^{**} (FGK)	fgk cu	α	β	MPSS y	Ratio
AL	18.21	27.5	0.66	0.99	0.84	18.04	1.01
AZ	9.1	10.75	0.85	1.23	1.20	11.17	0.81
AK	15.57	20.48	0.76	1.20	1.02	18.62	0.84
CA	8.83	10.27	0.86	1.44	1.36	12.71	0.70
CO	6.97	9.21	0.76	1.60	1.52	11.17	0.62
CT	9.41	11.04	0.85	1.00	1.00	9.41	1.00
DE	29.84	29.84	1	1.00	1.00	29.84	1.00
FL	5.79	7.23	0.8	1.84	1.64	10.69	0.54
GA	13.15	18.99	0.69	1.23	1.04	16.13	0.82
ID	7.43	18.24	0.41	1.41	1.18	10.43	0.71
IL	13.04	17.56	0.74	1.12	0.99	14.61	0.89
IN	19.66	33.01	0.6	0.69	0.62	13.63	1.44
IA	20.45	24.2	0.85	0.81	0.78	16.52	1.24
KS	19.3	19.3	1	1.00	1.00	19.30	1.00
KY	22.64	31.17	0.73	0.93	0.81	21.11	1.07
LA	47.47	47.47	1	1.00	1.00	47.47	1.00
ME	7.1	14.75	0.48	1.31	1.21	9.30	0.76
MD	8.88	13.02	0.68	1.13	1.06	10.00	0.89
MA	8.82	12.03	0.73	1.00	1.00	8.82	1.00
MI	13.58	19.15	0.71	1.35	1.15	18.32	0.74
MN	10.7	14.02	0.76	1.34	1.14	14.30	0.75
MS	18.28	24.63	0.74	1.08	0.91	19.70	0.93
MO	13.21	15.81	0.84	1.33	1.22	17.58	0.75
MT	6.46	6.46	1	4.02	3.75	25.92	0.25

Table 3 (continued) Alternative Measures of Capacity Output and Capacity Utilization

State	actual y	y** (FGK)	fgk_cu	α	β	MPSS_y	Ratio
NE	16.05	16.75	0.96	1.38	1.34	22.22	0.72
NV	6.12	8.37	0.73	1.65	1.60	10.12	0.61
NH	7.07	9.66	0.73	1.39	1.16	9.79	0.72
NJ	10.03	13.81	0.73	1.25	1.17	12.51	0.80
NM	8.71	8.74	1	2.49	2.48	21.64	0.40
NY	6.92	9.33	0.74	1.45	1.34	10.01	0.69
NC	16.02	18.09	0.89	1.00	1.00	16.02	1.00
ND	11.79	13.7	0.86	1.92	1.69	22.67	0.52
OH	14.44	20.03	0.72	1.08	0.93	15.60	0.93
OK	12.22	13.3	0.92	1.60	1.50	19.51	0.63
OR	9.26	14.98	0.62	1.00	1.00	9.26	1.00
PA	12.12	16.04	0.76	1.18	1.04	14.31	0.85
RI	5.22	9.24	0.57	1.73	1.48	9.04	0.58
SC	17.21	29.38	0.59	0.90	0.75	15.51	1.11
SD	9.84	10.85	0.91	1.72	1.62	16.90	0.58
TN	16.49	23.64	0.7	0.90	0.76	14.80	1.11
TX	22.48	26.3	0.85	1.22	1.16	27.34	0.82
UT	10.07	15.3	0.66	1.19	1.02	12.01	0.84
VT	7.67	15.09	0.51	1.50	1.30	11.48	0.67
VA	12.68	18.36	0.69	0.91	0.84	11.47	1.10
WA	11.74	13.59	0.86	1.26	1.22	14.84	0.79
WV	14.13	29.87	0.47	0.94	0.80	13.28	1.06
WI	13.43	17.57	0.76	1.19	1.03	16.01	0.84
WY	11.89	30.22	0.39	1.00	1.00	11.89	1.00

References:

- Banker, R.D. (1984), "Estimating the Most Productive Scale Size Using Data Envelopment Analysis", *European Journal of Operational Research* 17: 1 (July) 35-44.
- Berndt, E. and C. Morrison (1981) "Capacity Utilization measures: Underlying Theory and an Alternative Approach"; *American Economic Review*, 71, pp 48-52.
- Cassels, J.M. (1937) "Excess Capacity and Monopolistic Competition"; *Quarterly Journal of Economics*, Vol 51, No 3; pp 426-443.
- Färe, R., S. Grosskopf, and E.C. Kokkelenberg (1989) "Measuring Plant Capacity, Utilization and Technical Change: A Nonparametric Approach"; *International Economic Review*, Vol 30, No. 3; pp 655-666.
- Johansen, L. (1968) "Production Functions and the Concept of Capacity"; Reprinted in F.R. Førsund (ed.) *Collected Works of Leif Johansen*, Volume 1, Amsterdam, North Holland.
- Kaldor, N. (1935) "Market Imperfection and Excess Capacity"; *Economica*, New Series, Vol 2, No. 5; pp 33-50.
- Klein, L.R. (1960) "Some Theoretical Issues in the Measurement of Capacity"; *Econometrica*, Vol 28,2 pp 272-286.
- Klein, L.R. and V. Long (1973) "Capacity Utilization: Concept, Measurement, and Recent Estimates"; *Brookings Papers on Economic Activity*, Issue 3 (1973), pp 743-756.
- Paine, C.L. (1936) "Rationalisation and Theory of Excess Capacity"; *Economica*, New Series, Vol 3, No. 9; pp 46-60.
- Ray, S.C., K. Mukherjee, and Y. Wu (2006) "Direct and Indirect Measures of Capacity Utilization: A Nonparametric Analysis of U.S. Manufacturing"; *The Manchester School* 74, pp 526-548.
- Segerson, K. and D. Squires (1990) "On the Measurement of Economic Capacity Utilization for Multiproduct Industries"; *Journal of Econometrics*, 75, pp 76-85.