Nonparametric Technical Efficiency with K Firms, N Inputs, and M Outputs: A Simulation

Loren W. Tauer and John J. Hanchar

Monte-Carlo simulation of nonparametric efficiency shows that even when the number of firms is large, defining ten or more inputs results in most firms being measured as efficient. Comparison of the simulated results with any empirical results may suggest that the dimension of the problem, rather than actual efficiencies, determines computed efficiencies.

Nonparametric or data envelopment techniques using linear programming have become common tools to measure technical and cost efficiency of individual firms. The seminal work was by Farrell (1957), with the data envelopment technique presented by Charnes, Cooper, and Rhodes (1978), and recent developments reported by Färe, Grosskopf and Lovell (1985). The attraction of the nonparametric approaches is that a functional form need not be specified for the technology of the firm. Although flexible functional forms are available, it is believed by many that complete flexibility is preferred.

One characteristic of data envelopment analysis (DEA) procedures is that computed firm efficiencies appear to be dependent on the number of comparison firms used and the number of defined outputs and inputs—that is, the dimension of the problem. Nunamaker (1985) reported that variable set expansion through disaggregation or addition of new factors produces an upward trend in efficiency scores. Thrall (1988) shows the conditions under which Nunamaker's proposition is true, and supplies transition theorems for output and input expansion while holding the number of firms constant. As Leibenstein and Maital (1992) recently state, given enough inputs, all (or most) of the firms are rated efficient. They state that this is a direct result of the dimensionality of the input/output space relative to the number of observations (firms). In fact, Button and Weyman-Jones (1992) state it is well known that measured DEA efficiency in small samples is sensitive to the difference between the number of firms and the sum of inputs and outputs used. Although this may be common knowledge among DEA practitioners, there does not appear to be an analytical discussion of the problem.

The purpose of this paper is to determine the role of dimensionality in determining measured firm efficiency by performing a Monte-Carlo simulation of nonparametric efficiency using various combinations of firms, inputs, and outputs. For each firm, the quantities of individual inputs and outputs are randomly drawn from univariate distributions. Thus, there is no relationship between inputs and outputs, implying no production structure. Any change in computed efficiency as the number of inputs, outputs, or firms is varied should be strictly a function of the dimensionality of the problem. These simulated results can be used to test whether the empirical results of an efficiency study are different from the results using random data. Differences would lend more credibility that empirical results measure efficiencies rather than simply measuring the impact of dimensionality. A recent application by Thomas and Tauer (1994) uses this approach to separate the impact of linear input aggregation versus input dimensionality on measured technical efficiency.

Procedure

The underlying concept of the nonparametric approach is the existence of a bounding technology characterized by an input requirement set $L(Y)$, which can be constructed from observed input-output data from K firms. This set is specified as

---

The authors are Professor and Extension Associate, Department of Agricultural, Resource, and Managerial Economics, Cornell University. The authors thank R.N. Boisvert, A. Weersink, and two anonymous reviewers for providing useful comments.

$$L(y_1, \ldots, y_m) \equiv \left\{(x_1, \ldots, x_n): y_i \leq \sum_{k=1}^{K} \mu_k y_{ik},\ i = 1, \ldots, m;\ x_j \geq \sum_{k=1}^{K} \mu_k x_{jk},\ j = 1, \ldots, n;\ \mu_k \geq 0,\ k = 1, \ldots K\right\}, \quad (1)$$

where $\mu = (\mu_1, \ldots, \mu_k)$ is an intensity vector that forms linear combinations of the observed input vectors x_j and output vectors y_i. The technical efficiency of any of the K firms can be measured relative to this set by determining how much a firm can increase its output and remain in this set (output distance function), or decrease its use of inputs and remain in this set (input disance function).

Empirically, the technical input efficiency of firm k via the output distance function is calculated by solving the linear programming problem

$$D_o^{-1} = \max_{\mu_k} \theta_k$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \mu_k y_{ik} \geqslant \theta_k y_{ik}, \qquad i = 1, \ldots, m,$$

$$\sum_{k=1}^{K} \mu_k x_{jk} \leq x_{jk}, \qquad j = 1, \ldots, n,$$

$$\mu_k \geq 0,\ k = 1, \ldots, K,$$

where y_{ik} is the output i produced by firm k, and x_{jk} is the input j used by firm k, with m outputs and n inputs. This specification assumes radial technical inefficiency, strong disposability of inputs and outputs, and constant returns to scale, since the summation of the intensity vector μ is not constrained to be equal to one (variable returns to scale) or less than one (non-increasing returns to scale). The solution value θ_k shows the fraction by which a firm can expand its output and use no more input. The solution value $\theta_k = 1$ determines the firm as technically efficient. Any value $\theta_k > 1$ determines the firm as technically inefficient in its production of output. The inverse of θ_k shows the degree of efficiency, bounded between 0 and 1. Since constant returns to scale are imposed, this output-based efficiency measure is the inverse of the input-based efficiency measure (Färe et al. 1985). To solve for the technical efficiency of all K firms, it is necessary to solve K linear programs where the y_{ik} and x_{jk} on the RHS of the LP are replaced with the outputs and inputs of the k^{th} firm for each LP solution.

By defining various combinations of inputs, outputs, and firm numbers, the technical efficiency of each of K firms was computed from data randomly generated. The quantities of output i and input j for firm k were randomly drawn from univariate uniform distributions [0, 1]. By specifying the input-output data set this way, the chance that any one firm will lie on the bounding technology is strictly random. The simulations were performed for total inputs of 3, 5, 10 and 15, outputs of 1 and 3, with the number of firm combinations of 25, 50, 100 and 200 (50 and 200 for the three-output case). This spans most empirical combinations of inputs, outputs, and firms. Forty complete replications were completed at each of the firm-input-output number combinations.

It should be noted that the specification used to measure technical efficiency here is commonly used but is by no means unique. Other specifications may yield different results.

Results

The results for a single output and various combinations of inputs and firms are summarized in Table 1 by the percentage of firms measured as being completely technically efficient ($\theta_k = 1$). These results are also plotted in Figure 1. With 3 inputs and 25 firms, on average, over the forty replications 21.8% of the firms were technically efficient. The range of firms efficient over the forty replications went from a low of 8% to a high of 32 percent. With 3 inputs and 200 firms, on average, 4.8% of the firms were technically efficient.

As the number of firms increases, the computed efficiencies decrease, since it becomes more likely that any firm would then be dominated. What is more striking is the relationship between the number of defined inputs and the computed efficiencies. There is a dramatic increase in the number of firms that are efficient as the number of inputs increase. When 15 inputs are used, in all cases, over half of the firms were measured as technically efficient.

Table 1. Percentage of Firms Technically Efficient from Data Envelopment Simulation; One Output*

		Number of Inputs			
Number of Firms		3	5	10	15
25	mean	21.8	40.1	62.5	74.3
	s.d.	5.4	10.1	11.3	9.0
	range	(8–32)	(20–64)	(40–80)	(56–92)
50	mean	13.8	25.4	51.9	66.0
	s.d.	3.8	5.8	8.5	5.7
	range	(6–22)	(12–36)	(26–66)	(54–76)
100	mean	5.1	19.2	45.8	59.8
	s.d.	1.3	3.5	5.0	4.2
	range	(3–8)	(9–26)	(29–56)	(49–67)
200	mean	4.8	13.2	37.8	55.2
	s.d.	1.2	2.6	3.1	3.9
	range	(2–7)	(9–18)	(32–44)	(47–66)

*Inputs and the output were randomly generated from univariate uniform distributions [0, 1]; forty replications at each cell.

The results for three outputs and various combinations of inputs and firms are summarized in Table 2. The pattern as the number of inputs or firms increases is similar to the earlier results with one output; an increase in the number of inputs causes more firms (percentage) to be measured efficient, and an increase in the number of firms causes a smaller percentage of firms to be measured as efficient. The stronger effect is the number of inputs.

Comparing results across one and three outputs (Tables 1 and 2) shows that with a given number of inputs and firms, defining three outputs rather than one output, causes more firms to be measured as efficient. This is not surprising since increasing the number of outputs raises the dimensionality of the input/output spaces as stated by Leibenstein and Maital.

Comparing Simulated and Empirical Results

A comparison of the results of empirical efficiency studies and the simulated nonparametric efficiency results here is useful to determine whether the empirical studies replicate the simulated results. The results of four efficiency studies will be reported. Farrell's seminal article included technical efficiency computation of agricultural production in each of the then 48 United States. Using six different combinations of two inputs (ignoring two

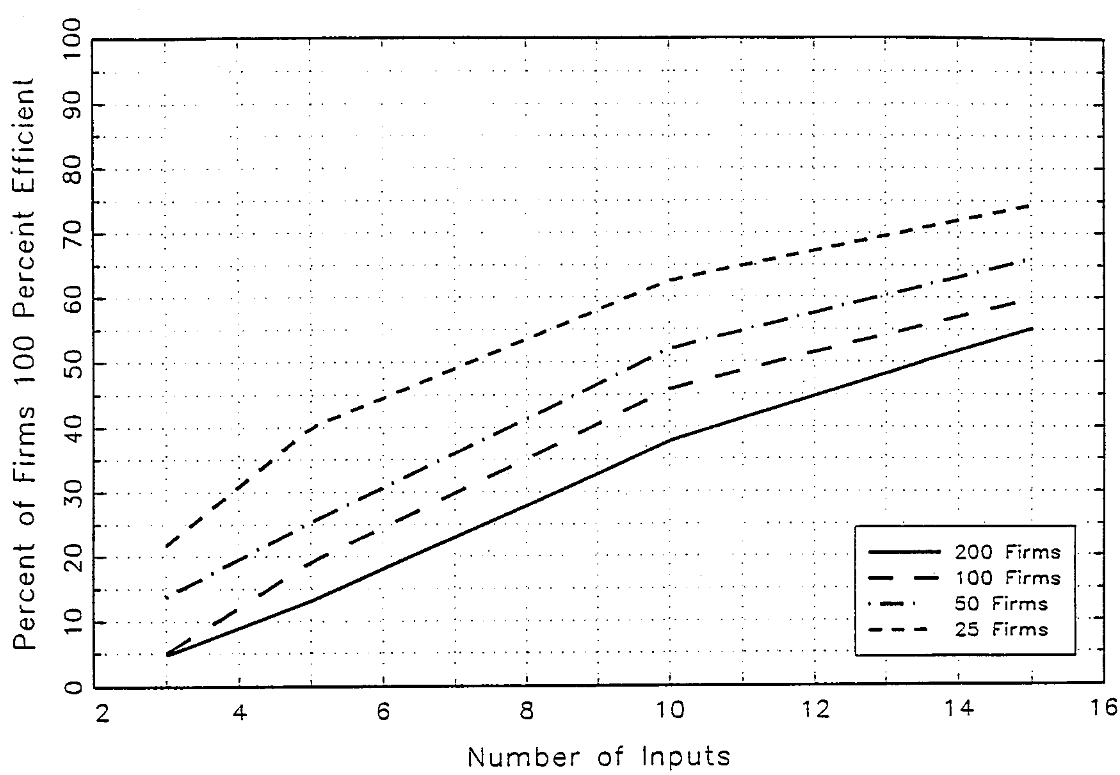

Figure 1. Graph of Table 1

Table 2. Percentage of Firms Technically Efficient from Data Envelopment Simulation; Three Outputs*

Number of Firms		Number of Inputs 3	5	10	15
50	mean	24.2	48.1	82.2	92.0
	s.d.	6.9	6.0	5.4	4.2
	range	(10–40)	(34–62)	(72–96)	(78–98)
200	mean	9.3	25.7	65.5	84.3
	s.d.	2.2	2.8	3.8	2.5
	range	(4–13)	(20–32)	(58–76)	(79–90)

*Inputs and the outputs were randomly generated from univariate uniform distributions [0, 1]; forty replications at each cell.

inputs at a time), he reported that between 4 and 12% of the observations were efficient. Using four different combinations of three inputs (ignoring the fourth input), he reported as efficient 8.3, 12.5, 14.6, and 16.7% of the observations. These results are within the ranges of 6 to 22% found in Table 1 under 3 inputs and 50 firms. Using all four inputs, 18.8% of the observations were efficient. In Table 3 a simple statistical test compares the sample mean percentage of the simulated results to the percentage reported for Farrell's empirical study of the three and four input cases. In four of the five situations, the simulated results differ from the empirical results.

Defining four outputs, six inputs, and using 92 firms, Grabowki *et al.* found 39.1% of their firms technically efficient. This differs statistically from a simulated 54.8% using the same number of outputs, inputs, and firms.

Sitoras used the Farrell approach to examine the agriculture sector in the Philippines using 1960 census data. A subsample of 58 agricultural municipalities taken from a total of 431 indicated that for four and eight inputs, 17.2 and 31% of the observations were technically efficient, respectively. In both cases, the simulation result does not replicate the empirical result. The empirical results also illustrate the trend that the percent of technically efficient firms increases as the number of inputs increases, holding the number of observations constant.

Thompson *et al.* (1990) applied efficiency analysis to Kansas farming. Results reported for 32 dryland wheat farms indicated that for one output and four inputs, 18.8% of farms were technically efficient. With two outputs and four inputs, 26.1% of 23 firms were technically efficient, and with threes outputs, 39.3% of 28 firms were technically efficient. Based upon the test statistics, the empirical results differ from the simulated results. Their results show that empirical estimates of technical efficiency increase with the number of outputs defined.

Weersink, Turvey and Godak computed technical efficiency measures for 105 Ontario dairy farms using one output and seven inputs. They reported that approximately 43% of the farms in the sample were technically efficient. This result is statistically different from the simulated percentage.

Table 3. Percentage of Firms Technically Efficient by Study, Simulated Efficiencies, and Tests of Significance

	Number of Outputs	Number of Inputs	Number of Firms	Percentage of Firms	Simulated Percentage of Firms Technically Efficient[a]	Standard Deviation	Test Statistic[b]
Farrell (1957)	1	3	48	8.3	14.9	3.0	13.7
	1	3	48	12.5	14.9	3.0	5.0
	1	3	48	14.6	14.9	3.0	0.6
	1	3	48	16.7	14.9	3.0	−3.7
	1	4	48	18.8	21.8	5.4	3.5
Grabowski et al. (1988)	4	6	92	39.1	54.8	5.2	18.9
Sitoras (1966)	1	4	58	17.2	18.9	4.5	2.4
	1	8	58	31.0	43.4	6.1	12.7
Thompson et al. (1990)	1	4	32	18.8	28.1	6.1	9.5
	2	4	23	26.1	42.9	12.9	8.1
	3	4	28	39.3	50.3	10.6	6.4
Weersink et al. (1990)	1	7	105	42.9	31.4	4.4	16.3

[a]Computed using DEA model, with number of outputs, inputs, and firms in columns 2, 3, and 4; 40 replications.
[b]A t-statistic is used to test the null hypothesis that the mean percent of efficient firms from the simulation is equal to the percent reported for the empirical study. The critical t value for alpha equal to 0.01 and 39 degrees of freedom is approximately 2.4.

Conclusions

Some researchers have observed that use of the nonparametric approach or data envelopment analysis to measure firm efficiency is sensitive to the difference between the number of firms and the sum of inputs and outputs used. This paper explores the severity of this problem by simulating nonparametric efficiency computations based on randomly generated data using various combinations of inputs, outputs, and firms. It was found that the number of inputs and outputs had a greater impact on calculated technical efficiency than the number of firms. Use of more than ten inputs caused the majority of firms to be measured as efficient.

If a researcher finds that the percentage of firms technically efficient in a study is not statistically different from simulated results given the same number of firms, inputs, and outputs, then those empirical results might simply be due to the dimensionality of the input/output space relative to the number of firms, rather than actual efficiencies. If empirical results are statistically different from the simulated results, then it could be more likely that actual differences in efficiencies exist among firms.

References

Button, K.J. and T.G. Weyman-Jones. "Ownership Structure, Institutional Organization and Measured X-Efficiency." *American Economic Review* 82(1992):439–45.

Charnes, A., W.W. Cooper, and E. Rhodes. "Measuring Efficiency of Decision Making Units." *European Journal of Operations Research* 6(1978):429–44.

Färe, R., S. Grosskopf and C.A.K. Lovell. *The Measurement of Efficiency of Production*. Boston: Kluwer-Nijhoff Publishing, (1985).

Farrell, M.J. "The Measurement of Productive Efficiency." *Journal of the Royal Statistical Society* 120(1957):253–81.

Grabowski, R., S. Kraft, S. Mehdian and C. Pasurka. "Technical Change in Illinois Agriculture, 1982–1984." *Agricultural Economics* 2(1988):303–18.

Leibenstein, H. and S. Maital. "Empirical Estimation and Partitioning of X-Inefficiency: A Data-Envelopment Approach." *The American Economic Review* 82(1992):428–33.

Nunamaker, Thomas R. "Using Data Envelopment Analysis to Measure the Efficiency of Non-profit Organizations: A Critical Evaluation." *Managerial and Decision Economics* 6(1985):50–8.

Sitoras, B. "Productive Efficiency and Redundant Factors of Production in Traditional Agriculture of Underdeveloped Countries." *Western Farm Economics Association Proceedings of 1966* 39(1966):153–8.

Thomas, A.C. and L.W. Tauer. "Linear Input Aggregation Bias in Nonparametric Technical Efficiency Measurement." *Canadian Journal of Agricultural Economics* 42(1994):77–86.

Thompson R.G., L.N. Langemeier, C.T. Lee, E. Lee and R.M. Thrall. "The Role of Multiplier Bounds in Efficiency Analysis with Application to Kansas Farming." *Journal of Econometrics* 46(1990):93–108.

Thrall, R.M. "Classification Transitions under Expansion of Inputs and Outputs in Data Envelopment Analysis." *Managerial and Decision Economics* 10(1989):159–62.

Weersink, A., C.G.. Turney and A. Godak. "Decomposition Measures of Technical Efficiency for Ontario Dairy Farms." *Canadian Journal of Agricultural Economics* 38(1990):439–56.