SNF Working Paper No. 23/05

Does Scale Efficiency Tell Us Anything About Optimal Scale?

Rögnvaldur Hannesson

SNF Project No. 5255: "Strategic Program in Resource Management"
The project is funded by The Research Council of Norway

INSTITUTE FOR RESEARCH IN ECONOMICS AND BUSINESS ADMINISTRATION

BERGEN, APRIL 2005

ISSN 1503-2140

Abstract

In DEA, scale efficiency is routinely calculated. This measure may, however, tell us very little about whether a production unit is over- or undersized. An empirical case is used to illustrate that, under some circumstances, scale inefficiency may simply reflect that a production unit is producing too little, given its use of factors of production, and not that is over- or undersized. A fictitious sample based on a production function with variable returns to scale is used for demonstrating that in small samples with large deviations from the efficiency frontier and limited variability between units in terms of factor proportions scale efficiency may not reflect very well how far the production units are from being of an optimal size.

Introduction

"Returns to scale" is an old concept in economics. It relates to whether production units are of an optimal size or not. We have increasing returns to scale if the output changes relatively more than the inputs as the size (scale) of the production unit is increased, and decreasing returns to scale if the opposite holds. The optimal size of the production unit is where a marginal increase in all inputs (scale) leads to the same relative increase in output. It is of obvious importance to identify the optimal scale of units in any industry.

Data envelopment analysis is now widely used to analyze the efficiency of firms and industries. One of the methods on offer distinguishes between scale inefficiencies and other sources of inefficiencies. Unfortunately, it appears that this method may give misleading results. Instead of identifying which production units are optimally scaled and which not, it may simply tell us by how much the output of an inefficient production unit can be increased, for given inputs. This appears most likely to happen when the sample of production units are not very different in terms of size or factor proportions but vary a great deal in production efficiency. While it is certainly true that it would not be easy to identify the optimally scaled units from a sample that contains units of similar size and factor proportion, the risk is that apparent differences in scale efficiencies might be mistaken for indicators of optimal or non-optimal scale. These points will be illustrated by an empirical example from the Norwegian fisheries and thereafter by a hypothetical example.

Scale efficiency for Norwegian factory trawlers

Norwegian factory trawlers are subject to individual quotas for the most important species they fish and severe restrictions on catching other types of fish, so an input oriented analysis appears most adequate for this fleet. For the year 2000 we have data on total catch value (V), the assessed replacement value of the vessel (K), and the number of people employed (L). Figure 1 shows the input coefficients (K/V and L/V) for the boats in the sample. One boat is clearly exceptional, producing much less per unit of labor and capital than the other boats. This boat is not very different from the others in terms of the labor and capital it employs but, as it turns out, the DEA attributes most of the inefficiency of this boat to scale.

An input-oriented data envelopment analysis can be formulated as follows[1]

$$\min_{\theta,\lambda} \theta$$

subject to
$$y_i \le \sum_{j=1}^{N} \lambda_j y_j \quad j = 1,..., N$$
$$\theta x_i \ge \sum_{j=1}^{N} \lambda_j x_{kj} \quad k = 1,..,M; j = 1, ..., N$$
$$\lambda \ge 0$$

This pertains to the case of constant returns to scale (CRTS). To deal with variable returns to scale (VRTS), the following constraint is added[2]

[1] See, e.g., Coelli, Rao and Battese (1998).
[2] This method was proposed by Färe and Grosskopf (1985), but an antecedent is a paper by Banker, Charnes and Cooper (1984).

$$\sum_{j=1}^{N} \lambda_j = 1$$

The data sample consists of 15 boats employing from 24 to 33 people, with an estimated replacement value varying from 104 to 151 million kroner. As can be seen from Figure 1 there are two peer candidates for this fleet. The boat using much more capital and labor per unit output than the other boats is in the middle of the above size range, with a replacement value of 126 million kroner and employing 28 people.

Table 1 shows the efficiency scores θ of the 15 boats in the sample under the CRTS versus VRTS assumptions. The table also shows the implied scale efficiency, i.e., SE = $\theta_{CRTS}/\theta_{VRTS}$. Four boats (1, 2, 8 and 13) are efficient with respect to scale or very nearly so. These boats differ markedly in size. Boat 1 is one of the smallest; it has the smallest crew and the second lowest replacement value. Boat 13 is one of the largest, with the highest replacement value, and second in terms of people employed. Boat 2 is of an intermediate size, while Boat 8 has the lowest replacement value but employs relatively many people. Hence it is difficult to see how these results could tell us anything of interest in terms of returns to scale, except perhaps that we have constant returns to scale over the size range of the sample.

To get a handle on what is involved, it is useful to look at the results from the above minimization problem for the individual boats, particularly boat 7 which differs so markedly from the other boats. Tables 2 and 3 show the λ-values obtained for the CRTS versus VRTS assumptions. Now look at Boat 7 in Table 2. Boat 1 is peer for Boat 7, with $\lambda = 0.1414$. What the LP-program is doing is scaling back the production of the peer (y_1) to match the production of Boat 7 (y_7), cf. the first constraint in the minimization problem above; from the primary data it is possible to verify that $y_7/y_1 = 0.1414$. The use of factors of production (K and L) by Boat 1 is also scaled back in a similar way, cf. the second constraint. The LP-problem asks by how much it would be possible to reduce the input of the factors of production by Boat 7 to match the required inputs by Boat 1, if the scaled-back output were in fact being produced as if by Boat 1. Boat 7 is not using inputs in the same proportion as Boat 1, and from the primary data we can find that $\lambda K_1 = 0.1179K_7$, and $\lambda L_1 = 0.1212L_7$. The value of θ must be the higher of those two ($\theta = 0.1212$), as we would otherwise violate the second constraint in the minimization problem.

Now consider Table 3 and the column for Boat 7 again. The constraint that the λs sum to one simply increases the λ value obtained for the peer (Boat 1) to one. Our reference values are now the output produced by Boat 1 and the inputs used by that boat. The second constraint in the LP-problem tells us that we can reduce the inputs of Boat 7 to match those levels. From the primary data we find that $K_1/K_7 = 0.8338$ and $L_1/L_7 = 0.8571$, and again we must set θ equal to the higher of those values, so as not to violate the resource constraint. Our efficiency measure θ is now telling us how much inputs we can save by using the same input combination as Boat 1, the peer boat. What the scale efficiency then tells us by how much we can expand the production of Boat 7 if in fact it were as efficient as Boat 1. Obviously, this has nothing to do with the scale of Boat 7 in the sense of its size; i.e., its replacement value and the number of men it employs.

As can be seen from Table 1 there is an enormous difference between the efficiency scores for Boat 7 in the CRTS and the VRTS cases. Despite the fact that Boat 7 uses many times more inputs per unit of output than the other boats it has an efficiency score in the VRTS case which is slightly higher than the scores of Boat 9 and 13. The inefficiency of this boat is ascribed to low scale efficiency (SE), but the difference in scale is associated only with an unusually low level of production, not with the size of the production unit, which is what we normally associate with scale.

A hypothetical example

Is the above result just due to a slight pathology in the sample; i.e., little variability in size and factor proportions and large inefficiencies in production for some units? To throw some light on this, the following hypothetical case was constructed.

Assume the following variable returns to scale production function:

$$Y = \left[K^{\alpha} L^{1-\alpha}\right]^{\bar{Y}/(K^{\alpha} L^{1-\alpha})}$$

where $\bar{Y}$ is the level at which the scale of operations is optimal. For an efficient production process we can express the optimal scale in terms of output rather than inputs, even if scale is associated with the use if inputs. From this we can verify that the scale elasticity is greater than one for $Y < \bar{Y}$ and less than one in the opposite case. Figure 2 shows the production function for a given ratio *K/L* as both factors are increased proportionally, with $\bar{Y} = 1$.

Now consider a "sample" where *K* and *L* are increased from 0.5 in steps of 0.25 up to a maximum of 1.5. All combinations give us 25 cases. Let there be disturbances in each observation determined by drawing random, evenly distributed variables ε between 0 and 1. Let these random disturbances translate into excessive use of *K* and *L* as

$$K^* = K + \delta\varepsilon$$

$$L^* = L + \delta\varepsilon$$

where K^* and L^* are the "observed" inputs. Table 4 shows the "observations" and Figure 3 the input coefficients implied by this "sample", for $\delta = 0.1$. It may be noted that there is considerable variability in these "observations" in terms of factor proportions and size of production units; there are many peer candidates.

From this sample the efficiency parameters θ_{CRTS}, θ_{VRTS}, and the scale efficiency SE = $\theta_{CRTS}/\theta_{VRTS}$ were estimated. How well does SE reflect the true scale efficiency? In the one-dimensional case in Figure 2 the scale inefficiency is the horizontal distance from a point on the production function to the straight line from the origin, which would be the production function if there were constant returns to scale. Since the slope of this line is one, we can measure by the vertical distance; i.e., the difference between the actual production and what the actual use of factors would produce if the returns to scale were constant. Hence, the true scale efficiency (SE^*) is

$$SE^* = \left[K^{\alpha} L^{1-\alpha}\right]^{\bar{Y}/(K^{\alpha} L^{1-\alpha})-1}$$

which, for $Y = \bar{Y}$, is one and for all other *Y*s less than one.

Figure 4 shows the relationship between the true scale efficiency and the measured scale efficiency as estimated by the input-oriented DEA. We see that the relationship is nearly one-to-one. In this case the estimated scale efficiency comes very close to estimating the true scale efficiency. Increasing the deviation parameter δ to one gives a slightly disturbed relationship, but the optimal scale is still very close to being identified.

Now make the sample slightly pathological. Let K and L assume values of 0.8, 1 and 1.1, giving us nine observations. Again we draw random variables to generate deviations of K and L and use a deviation parameter of one. The resulting “sample” is shown in Table 5 and the input coefficients in Figure 5. This new “sample” reminds us of the trawler example in the previous section. There is not much variability in the factor proportions, all units are close to being on optimal scale, there appear to be large inefficiencies for some units, and there is only one peer unit in the sample.

The relationship between the true and the estimated scale elasticity for this restricted, hypothetical sample is shown in Figure 6. The unit with optimal scale is correctly identified, but we see that some units are characterized as severely scale inefficient even if their true scale efficiency is quite high and close to one.

Conclusion

Scale efficiency in DEA is not necessarily what it purports to be. In samples with sufficient variability of data in terms of size and factor proportions it is likely to correctly identify which production units are close to being of optimal size and which are not, and how far the individual production units are from being scale efficient. In samples of similar units which differ primarily in terms of technical efficiency, however, the estimated scale inefficiency may reflect insufficient production due to inefficient use of inputs rather than a non-optimal size of the production unit. The scale efficiency results from DEA-models therefore need to be interpreted with caution and perhaps supplemented by alternative estimates of scale properties, such as recommended by Førsund and Hjalmarsson (2004), and Cooper, Seiford and Tone (2000).

References

Banker, R.D., A. Charnes and W.W. Cooper (1984): Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, Vol. 5, pp. 1078-1092.

Coelli, T., D.S.P. Rao and G.E. Battese (1994): *An Introduction to Efficiency and Production Analysis*. Kluwer, Dordrecht.

Cooper, W.W., L.M. Seiford and K. Tone (2000): *Data Envelopment Analysis*. Kluwer, Dordrecht.

Färe, R. and S. Grosskopf (1985): A Nonparametric Cost Approach to Scale Efficiency. *Scandinavian Journal of Economics*, Vol. 87, pp. 594-604.

Førsund, F. (1996): On the Calculation of the Scale Elasticity in DEA Models. *Journal of Productivity Analysis*, Vol. 7, pp. 283-302.

Førsund, F. and L. Hjalmarsson (2004): Are all Scales Optimal in DEA? Theory and Empirical Evidence. *Journal of Productivity Analysis*, Vol. 21, pp. 25-48.

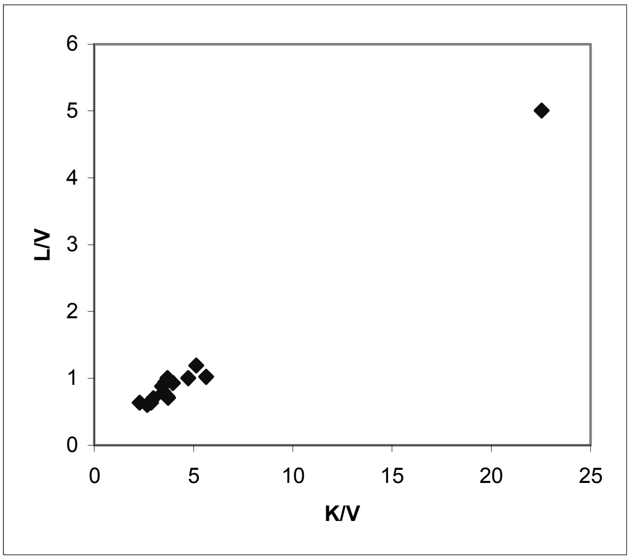

Figure 1

Input coefficients for 15 Norwegian factory trawlers 2000.

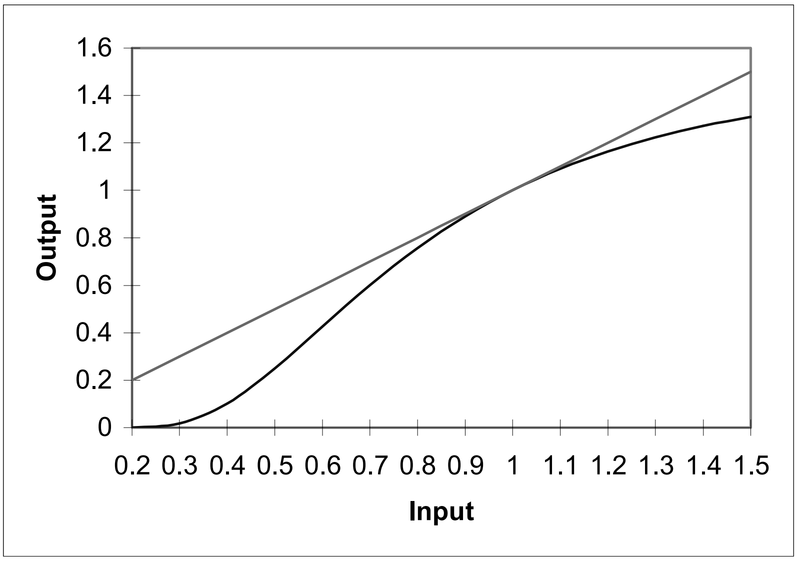

Figure 2

A variable returns to scale production function and a constant returns to scale production function (straight line).

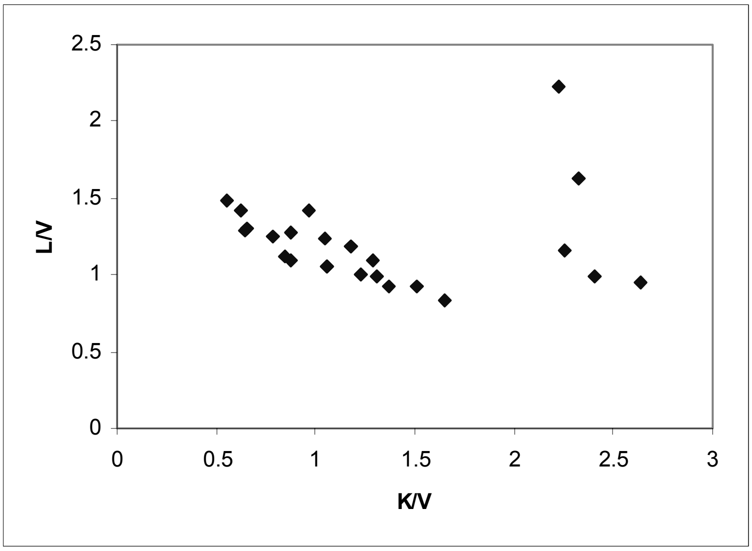

Figure 3

Input coefficients for hypothetical example.

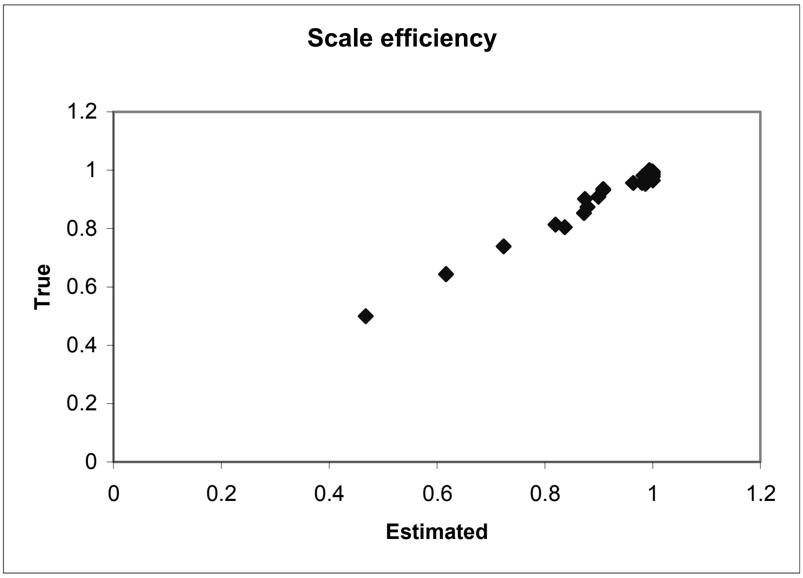

Figure 4

True and estimated scale efficiency in hypothetical sample.

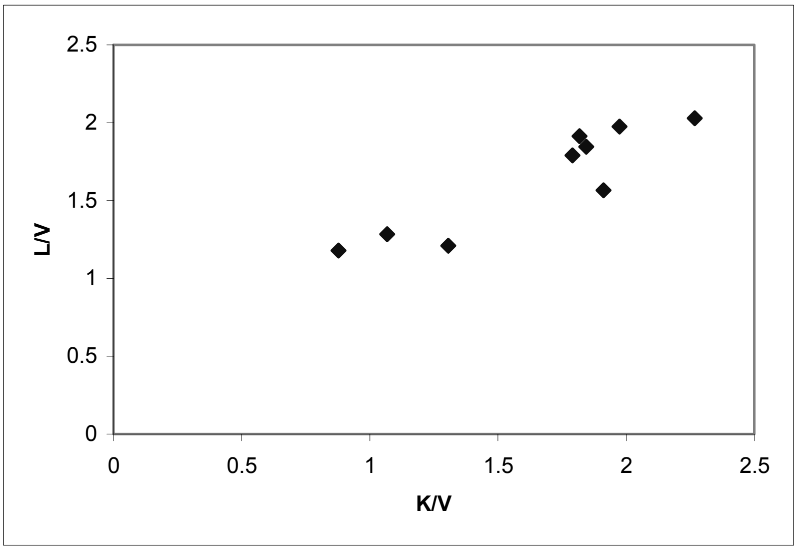

Figure 5

Input coefficients in the restricted, hypothetical sample.

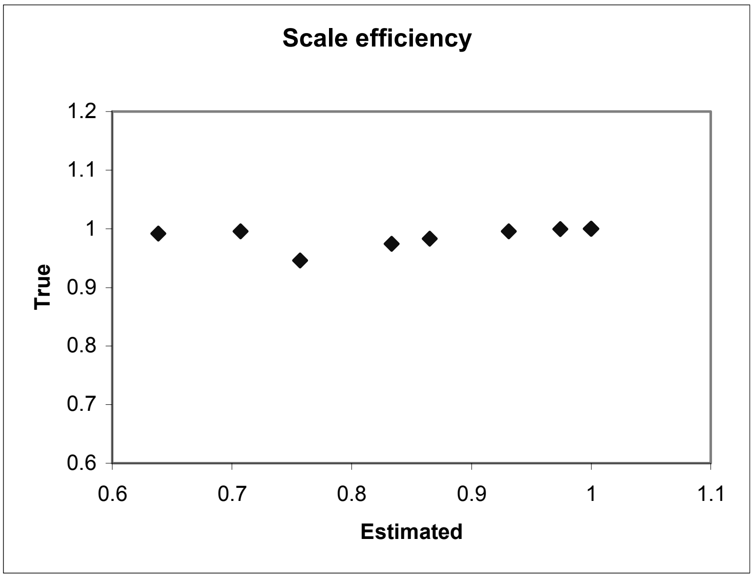

Figure 6

True and estimated scale efficiency in the small hypothetical sample.

Table 1

Efficiency scores for 15 factory trawlers and the input of labor and capital.

Boat	θ_{CRTS}	θ_{VRTS}	**SE**	***L***	***K***
1	1	1	1	24	105
2	0.8687	0.8729	0.9952	29	122
3	0.6052	0.9600	0.6304	25	118
4	0.7601	0.8588	0.8851	28	122
5	0.6326	0.8632	0.7328	33	121
6	0.6678	0.9912	0.6737	28	105
7	0.1212	0.8571	0.1414	28	126
8	1	1	1	29	104
9	0.6580	0.8509	0.7733	29	123
10	0.8524	0.9231	0.9235	26	136
11	0.5122	0.8130	0.6300	30	129
12	0.7125	0.9046	0.7876	30	116
13	0.8305	0.8430	0.9851	30	151
14	0.9530	0.9924	0.9603	28	126
15	0.5935	0.9600	0.6183	25	138

.

Table 2

Values of λ in the CRTS case. Each column shows values obtained for one boat's LP-problem.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	1	0.8334	0.6304	0.8767	0.0753	0.1679	0.1414	0	0.6689	0.9235	0.5808	0.2933	1.0381	1.1118	0.6183
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0.1790	0	0.0084	0.6575	0.5058	0	1	0.1044	0	0.0492	0.4943	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Σλ	1	1.0124	0.6304	0.8851	0.7328	0.6737	0.1414	1	0.7733	0.9235	0.6300	0.7876	1.0381	1.1118	0.6183

Table 3

Values of λ in the VRTS case. Each column shows values obtained for one boat's LP-problem.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	1	0.7373	1	0.9905	0.1028	0.2492	1	0	0.8650	1	0.9219	0.3723	0.7420	0.2425	1
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0.2627	0	0.0095	0.8972	0.7508	0	1	0.1350	0	0.0781	0.6277	0.2580	0.7575	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\Sigma\lambda$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 4

"Observations" of output and inputs in hypothetical example. *K* and *L* are inputs given technical efficiency. ε is a random variable, and K^* and L^* are "observed" inputs.

K	***L***	***Y***	ε	K^*	L^*
0.5	0.5	0.250	0.555	0.555	0.555
0.5	0.75	0.540	0.194	0.519	0.769
0.5	1	0.774	0.077	0.508	1.008
0.5	1.25	0.947	0.889	0.589	1.339
0.5	1.5	1.073	0.975	0.597	1.597
0.75	0.5	0.363	0.935	0.844	0.594
0.75	0.75	0.681	0.554	0.805	0.805
0.75	1	0.910	0.243	0.774	1.024
0.75	1.25	1.067	0.911	0.841	1.341
0.75	1.5	1.176	0.101	0.760	1.510
1	0.5	0.455	0.247	1.025	0.525
1	0.75	0.782	0.261	1.026	0.776
1	1	1.000	0.560	1.056	1.056
1	1.25	1.143	0.029	1.003	1.253
1	1.5	1.238	0.848	1.085	1.585
1.25	0.5	0.530	0.245	1.275	0.525
1.25	0.75	0.857	0.409	1.291	0.791
1.25	1	1.065	0.608	1.311	1.061
1.25	1.25	1.195	0.166	1.267	1.267
1.25	1.5	1.280	0.862	1.336	1.586
1.5	0.5	0.593	0.661	1.566	0.566
1.5	0.75	0.917	0.105	1.510	0.760
1.5	1	1.114	0.242	1.524	1.024
1.5	1.25	1.234	0.941	1.594	1.344
1.5	1.5	1.310	0.475	1.548	1.548

Table 5

"Observations" of output and inputs in a restricted, hypothetical example. *K* and *L* are inputs given technical efficiency. ε is a random variable, and K^* and L^* are "observed" inputs.

K	***L***	***Y***	$\boldsymbol{\varepsilon}$	$\boldsymbol{K^*}$	$\boldsymbol{L^*}$
0.8	0.8	0.757	0.555	1.355	1.355
0.8	1	0.931	0.194	0.994	1.194
0.8	1.1	1.000	0.077	0.877	1.177
1	0.8	0.833	0.889	1.889	1.689
1	1	1.000	0.975	1.975	1.975
1	1.1	1.064	0.935	1.935	2.035
1.1	0.8	0.865	0.554	1.654	1.354
1.1	1	1.028	0.243	1.343	1.243
1.1	1.1	1.091	0.911	2.011	2.011