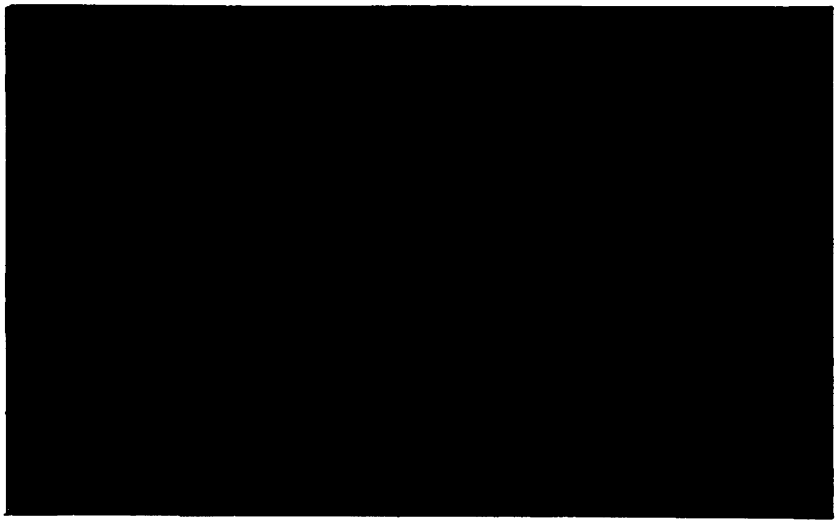
DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS
SYMONS HALL
UNIVERSITY OF MARYLAND
COLLEGE PARK 20742



USING DOMINANCE IN FORMING DEA MODELS:

The Case of Experimental Agricultural Data

by

Robert G. Chambers, Rolf Färe, and

Edward Jaenicke

WP Working Paper
94-08

I. Introduction

Data Envelopment Analysis (DEA) or activity analysis models are frequently formed from an input-output data set as its convex disposable hull. In forming DEA models, Banker, Charnes and Cooper (1984) suggest the minimum extrapolation postulate which requires that all input-output observations be used to form the reference technology. However, the minimum extrapolation postulate was formulated in a context where the focus was on efficiency measurement relative to a "best-practice" frontier. While this postulate is entirely reasonable in many circumstances, in others it might lead to construction of an overly optimistic reference technology. Consider, for example, the use of DEA-related methods to characterize agricultural production technologies where the data are drawn from experimental field trials rather than from observations on actual farming operations. Such field trials are notoriously optimistic in their predictions about fertilizer responsiveness. Relying on an outer frontier approximation would only exacerbate this tendency.

Our purpose here is to present and compare pessimistic and optimistic reference technologies in order to create some bounds within which one can reasonably expect the "true" technology to be. The criteria we use to create the optimistic and pessimistic technologies are dominance in the sense of efficiency and reversed efficiency. Given a set of $\mathcal{K}$ observations of inputs and outputs, the optimistic technology is formed from the subset $\mathcal{E}$ of efficient elements. The first pessimistic, or rather the first conservative technology, is formed from the subset $\mathcal{C}$ of reversed efficient elements of $\mathcal{K}$. In order to compare the optimistic technology and this conservative technology, both formed as convex disposable hulls, we prove a result showing that the BCC model and our optimistic technology coincide. The result, interesting in its own right, allows us to conclude that the conservative reference technology is a subset of the optimistic.

Next we move on to an even more conservative reference technology that is formed by exluding from $\mathcal{C}$ those elements which dominate convex combinations of their complements in $\mathcal{C}$. Once these points are removed, the most conservative reference technology is found by enveloping the remaining elements of $\mathcal{C}$.

II. The Models

We introduce the different reference technologies in this section. In particular we make use of the idea of dominance in the sense of efficiency and reversed efficiency to eliminate data points. Dominance in the above sense has also been used by Fried et al. (1993), Tulkens and Vander Eeckaut (1991), and Hougaard and Tvede (1993).

Assume that we are given a set $\mathcal{K} = \{1, ..., k, ..., K\}$ of input vectors $\mathbf{x}^k \in \mathbb{R}_+^N$ and output vectors $\mathbf{y}^k \in \mathbb{R}_+^M$. The convex disposable hull technology formed from these may be written as (see Banker, Charnes and Cooper, 1984 or Färe, Grosskopf and Lovell, 1994),

$$T(\mathcal{K}) = \{(x, y) : \sum_{k \in K} z_k y_{km} \geqq y_m, \quad m = 1, ..., M,$$
$$\sum_{k \in K} z_k x_{kn} \leqq x_n, \quad n = 1, ..., N,$$
$$z_k \geqq 0, \ k \in \mathcal{K} \sum_{k \in K} z_k = 1 \}.$$

For each $k \in \mathcal{K}$, $(\mathbf{x}^k, \mathbf{y}^k) \in T(\mathcal{K})$, and $T(\mathcal{K})$ is convex with inputs and outputs being freely disposable, i.e., if $(\mathbf{x}, -\mathbf{y}) \leq (\mathbf{x}^0, -\mathbf{y}^0)$ and if, $(\mathbf{x}^0, \mathbf{y}^0)$ belongs to $T(\mathcal{K})$, so does $(\mathbf{x}, \mathbf{y})$.

The efficient subset of $\mathcal{K}$ is defined as

$$\mathcal{E} = \{ k \in \mathcal{K} : (\mathbf{x}^i, -\mathbf{y}^i) \leq (\mathbf{x}^k, -\mathbf{y}^k) \Rightarrow i \notin \mathcal{K} \}.$$

The efficient subset consists of those $k \in \mathcal{K}$ which are not dominated in the sense of less than or equal to by any element in $\mathcal{K}$.[1] The reference technology associated with $\mathcal{E}$ is given by

$$T(\mathcal{E}) = \{(x, y) : \sum_{k \in \mathcal{E}} z_k y_{km} \geqq y_m, \ m = 1, ..., M,$$
$$\sum_{k \in \mathcal{E}} z_k x_{kn} \leqq x_n, \ n = 1, ..., N,$$
$$z_k \geqq 0, k \in \mathcal{E}, \sum_{k \in \mathcal{E}} z_k = 1 \}.$$

The *conservative subset* of $\mathcal{K}$ is defined as

$$\mathcal{C} = \{ k \in \mathcal{K}: (x^i, -y^i) \geq (x^k, -y^k) \Rightarrow i \notin \mathcal{K} \}.$$

In words, if i dominates any k in the sense of $\geq$, then i is not in $\mathcal{K}$. The conservative technology is given by

$$T(\mathcal{C}) = \{(x, y) : \sum_{k \in \mathcal{C}} z_k y_{km} \geqq y_m, \ m = 1, ..., M,$$
$$\sum_{k \in \mathcal{C}} z_k x_{kn} \leqq x_n, \ n = 1, ..., N,$$
$$z_k \geqq 0, k \in \mathcal{C}, \sum_{k \in \mathcal{C}} z_k = 1 \}.$$

Figure 1 illustrates the reference technologies $T(\mathcal{E})$ and $T(\mathcal{C})$. There are four observations and $\mathcal{K} = \{1, 2, 3, 4\}$. The efficient subset of $\mathcal{K}$ is $\mathcal{E} = \{1, 2, 3\}$, and the conservative subset is $\mathcal{C} = \{3, 4\}$.

For this figure it is clear that the conservative technology is a subset of the efficient. To prove this generally, we use the following lemma, the proof is in the appendix.

<u>Lemma</u>: $T(\mathcal{K}) = T(\mathcal{E})$.

From this lemma it follows from the observation that $\mathcal{K} \supseteq \mathcal{C}$ that and the fact that reference technologies are constructed as convex disposable hulls that

[1]As usual $\mathbf{x} \leq \mathbf{y}$ means that each element of $\mathbf{x}$ is no larger than the corresponding element of $\mathbf{y}$ and at least one element of $\mathbf{x}$ does not equal the corresponding element of $\mathbf{y}$. $\mathbf{x} \leqq \mathbf{y}$ means that each element of $\mathbf{x}$ is no larger than the corresponding element of $\mathbf{y}$.

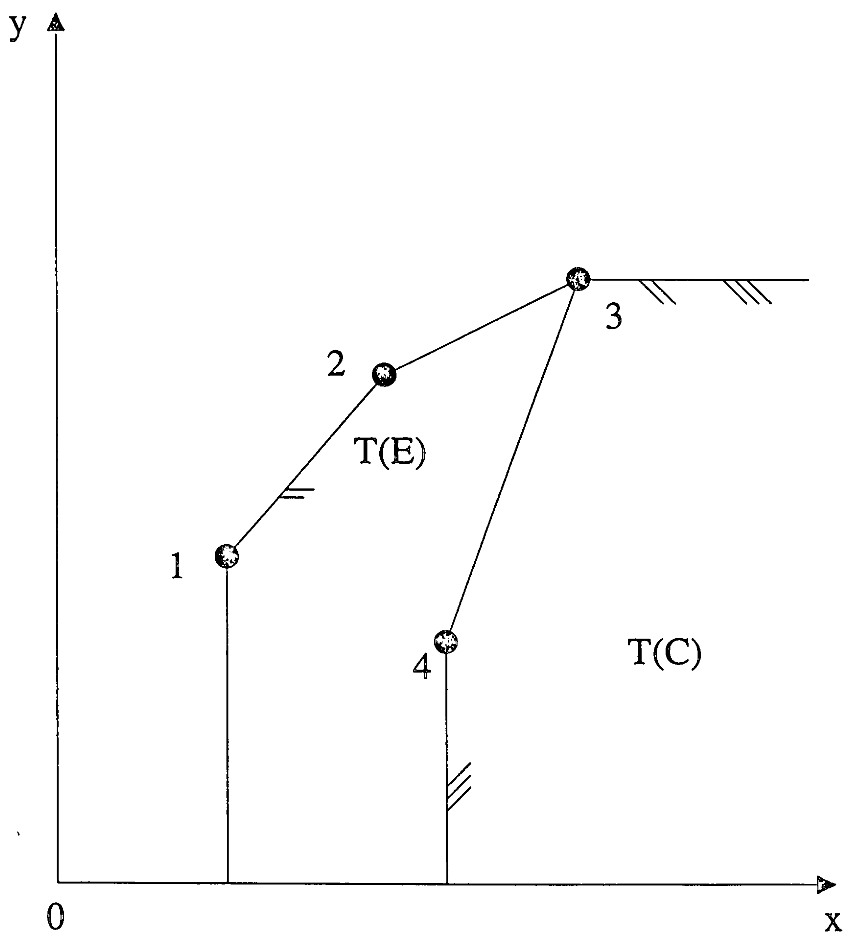


Figure 1. The Efficient and Conservative
Reference Technology

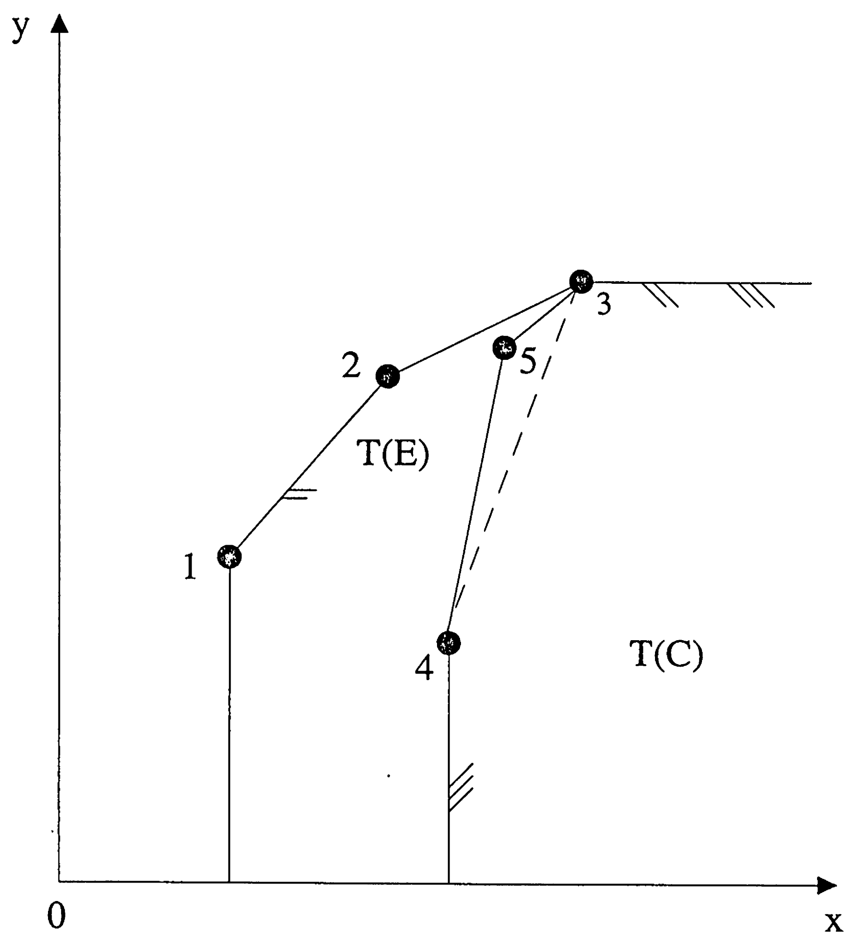


Figure 2. The Most Conservative Reference Technology

Proposition: $T(\mathcal{S}) \supseteq T(\mathcal{C})$.

To measure the differences between the optimistic and pessimistic technologies $T(\mathcal{S})$ and $T(\mathcal{C})$ we choose the inverse of the output oriented Farrell measures of technical efficiency, i.e., we compute the ratios

$$F(\mathbf{x}^k, \mathbf{y}^k \,|\, \mathcal{S}) \;/\; F(\mathbf{x}^k, \mathbf{y}^k \,|\, \mathcal{C})$$

for each $k \in \mathcal{C}$. In particular we calculate

$$F(\mathbf{x}^k, \mathbf{y}^k \,|\, \mathcal{S}) = \min \{ \; \theta \in \mathbb{R}_+ : (\mathbf{x}^k, \mathbf{y}^k/\theta) \in T(\mathcal{S}) \; \}$$

and

$$F(\mathbf{x}^k, \mathbf{y}^k \,|\, \mathcal{C}) = \min \{ \; \theta \in \mathbb{R}_+ : (\mathbf{x}^k, \mathbf{y}^k/\theta) \in T(\mathcal{C}) \; \}.$$

$T(\mathcal{C})$, however, does not always provide the most conservative representation of the technology consistent with free disposability of inputs and outputs. To see that this is true, consider Figure 2 which replicates the observations Figure 1 while adding a new observation labelled "5". For this data set, observation 5 belongs to $T(\mathcal{C})$, whereas in Figure 1 the point corresponding to 5 would not have been in $T(\mathcal{C})$. A more conservative representation of the reference technology for the data set in Figure 2 would be given by the same $T(\mathcal{C})$ as in Figure 1.

This more conservative version of the technology is isolated by eliminating from $\mathcal{C}$ all elements of $\mathcal{C}$ that dominate convex combination of their complements in $\mathcal{C}$. (Notice in Figure 2 that observation 5 dominates a range of convex combinations of 4 and 3.) Formally, we form $\mathcal{I} \subseteq \mathcal{C}$ by

$$\mathcal{I} = \{ k \in \mathcal{C} : \sum_{\substack{i \in \mathcal{C} \\ i \neq k}} z_i \mathbf{y}^i \leq \mathbf{y}^k$$

$$\sum_{\substack{i \in \mathcal{C} \\ i \neq k}} z_i \mathbf{x}^i \geq \mathbf{x}^k$$

$$\text{for any } z_i \geqq 0,\ i \in \mathcal{C},\ i \neq k,\ \sum_{\substack{i \in \mathcal{C} \\ i \neq k}} z_i = 1 \ \}.$$

The *most conservative subset* of $\mathcal{K}$ is then defined by

$$\mathcal{M} = \mathcal{C} - \mathcal{I},$$

and the most conservative reference technology is given by

$$\begin{aligned} T(\mathcal{M}) = \{\ (\mathbf{x}, \mathbf{y}) : & \sum_{k \in M} z_k y_{km} \geqq y_m,\ \ m = 1, \ldots, M, \\ & \sum_{k \in M} z_k x_{kn} \leqq x_k,\ \ n = 1, \ldots, N \\ & z_k \geqq 0,\ k \in \mathcal{M},\ \sum_{k \in M} z_k = 1 \ \}. \end{aligned}$$

To measure the differences between the three reference technologies, we again compute the ratios of the inverses of the output-oriented Farrell measures of technical efficiency:

$$F(\mathbf{x}^k, \mathbf{y}^k | \mathcal{E}) \ / \ F(\mathbf{x}^k, \mathbf{y}^k | \mathcal{M})$$

for $k \in \mathcal{M}$ and

$$F(\mathbf{x}^k, \mathbf{y}^k |\ \mathcal{C}) \ / \ F(\mathbf{x}^k, \mathbf{y}^k |\ \mathcal{M})$$

$k \in \mathcal{M}$ where

$$F(\mathbf{x}^k, \mathbf{y}^k |\ \mathcal{M}) = \min\ \{\ \theta : (\mathbf{x}^k, \mathbf{y}^k/\theta) \in T(\mathcal{M})\ \}.$$

III. An Algorithm for Finding the Elements of I

While the definition of I is straightforward, in practice it is useful to have a simple computational procedure for isolating the elements of I. Here we demonstrate that I can be isolated by solving several simple linear programs. Consider the set $R(\mathcal{C})$ defined by

$$\begin{aligned} R(\mathcal{C}) = \{(\mathbf{x}, \mathbf{y}) : & \ x_n \leqq \sum_{i \in C} \lambda_i\, x_{in} \ \ n = 1, \ldots, N \\ & \ y_m \geqq \sum_{i \in C} \lambda\, y_{im} \ m = 1, \ldots, M \\ & \sum_{i \in C} \lambda_i = 1,\ \lambda_i \geqq 0,\ i \in \mathcal{C}\}. \end{aligned}$$

It is immediate that $I \subset R(\mathcal{C})$. To see this suppose that $k \in I$, then by definition

$$y_{km} \geqq \sum_{\substack{i \in C \\ i \neq k}} \lambda_i \, y_{im} \qquad m = 1, ..., M$$

$$x_{kn} \leq \sum_{\substack{i \in C \\ i \neq k}} \lambda_i \, x_{in} \qquad n = 1, ..., N$$

$$\sum_{\substack{i \in C \\ i \neq k}} \lambda_i = 1, \lambda_i \geq 0, i \in C, i \neq k.$$

Now add zero in the form of $\lambda_k \, y_{km}$, $\lambda_k \, x_{kn}$, λ_k (with $\lambda_k = 0$) to each respective inequality to establish that $(\mathbf{x}^k, \mathbf{y}^k) \in R(C)$.

Define the input-oriented and output-oriented radical measure of the distance to the frontier of $R(C)$ by

$$F^*(\mathbf{x}^k, \mathbf{y}^h \mid R(C)) = \min \{\theta \in \mathbb{R}_+ : (\mathbf{x}^k, \theta \mathbf{y}^k) \in R\,(C)\}$$

$$H^*(\mathbf{y}^h, \mathbf{x}^k \mid R(C)) = \max \{\lambda \in \mathbb{R}_+ : (\lambda \mathbf{x}^k, \mathbf{y}^k) \in R\,(C)\}.$$

Because $R(C)$ is a closed convex set both of these linear programs are well defined so long as $k \in C$. Now suppose that $F^*(\mathbf{x}^k, \mathbf{y}^k \mid R(C)) < 1$, then it must be true that

$$y_{km} \geqq \frac{\sum_{i \in C} \lambda_i \, y_{im}}{F^*} \geqq \sum_{i \in C} \lambda_i \, y_{iM} \qquad m = 1, ..., M$$

where the second inequality is strict if the far right-hand term is not zero. There are two cases to consider: $\lambda_k = 1$ and $\lambda_k < 1$. In the latter it follows immediately that

$$y_{kM} > \sum_{\substack{i \in C \\ i \neq k}} \lambda_i^* \, y_{iM}$$

where $\lambda_i^* = \lambda_i / (1 - \lambda_k)$. Thus, $k \in I$. The former is easily seen to be impossible for $y^k \neq 0^M$. Similarly it is easy to show that $H^*(y^k, x^k \mid R(C)) > 1 \Rightarrow R \in I$.

Now suppose that both $H^*(y^h, x^k \mid R(C)) = 1$ and $F^*(x^k, y^h \mid R(C)) = 1$. The only interesting case to consider is where $\lambda_k < 1$ in the solution to either of the respective linear programs. In that case, we only need to check whether there exists one strong inequality satisfying when

$$y_{km} > \sum_{\substack{i \in C \\ i \neq k}} \lambda_i^* y_{im}$$

for any m

$$x_{kn} < \sum_{i \in C} \lambda_i^* x_{kin}$$

at the $\lambda_i^* = \lambda_i / (1 - \lambda_k)$ defined by the solution value. If there is that observation will belong to I.

IV. Data

We demonstrate our approach using data from a three-year field study comparing yields of no-till corn following four cover crops (hairy vetch, crimson clover, Austrian peas and winter wheat) and winter fallow in the Maryland Coastal Plain (for a complete description see Decker et al.). The experiment examined only applied nitrogen use, so our study is for a single variable input. Fixed inputs are yield of the cover crop (to measure the organic nitrogen content of the soil), precipitation, and temperature. Weather records were used to construct two variables measuring precipitation and temperature: total precipitation during the early growing season and the number of days during the late growing season between 70° and 86°F.

Four different nitrogen fertilizer rates were used on each winter cover crop. Fertilization rates of 0, 40, 80 and 120 pounds of nitrogen per acre were used on the vetch system, rates of 0, 60, 120 and 180 pounds of nitrogen per acre on the clover and peas systems and rates of 0, 120, 180 and 240 pounds per acre on the wheat and winter fallow systems. The experiment was conducted on different plots on the farm each year. Each cover crop followed no-till corn. Corn was planted in the spring between 5 and 15 days after the covers were killed with a knockdown herbicide, depending upon the condition of the killed cover crop growth, soil condition and rainfall. Samples of corn grain were dried to 15.5 percent moisture and used to estimate yield per acre. Samples of the cover crop were also harvested and dried and used to estimate yield per acre.

V. Results

The differences between the three representations for two of the cover crops (crimson clover and hairy vetch) are summarized in Tables 1-4. All cover crops yielded similar differences, but to conserve space we only report on two.

For crimson clover 14 out of 48 of the elements of $\mathcal{K}$ were in $\mathcal{C}$ (see Table 1) while for the hairy vetch cover crop 15 out of 48 elements of $\mathcal{K}$ were in $\mathcal{C}$ (see Table 2). For crimson clover, we find that there is considerable difference between the optimistic and conservative reference technologies. For example, Table 1 indicates that for the input level for observation 44 the output frontier of the conservative technology was only approximately one quarter of the output frontier for the optimistic technology. Even more dramatic differences are found when comparing the optimistic and conservative reference technologies for the hairy vetch cover crop. For observation 34, Table 2 indicates that the conservative frontier was only approximately one tenth of the optimistic technology. Because we are considering a technology with a scalar output, these differences in the technologies may be made more intuitive by recognizing that the ratio of the inverses of the output-oriented Farrell measure of technical efficiency corresponds to the ratios of the scalar production function for the conservative and optimistic technology. Thus, for example, for observation 34 in the hairy vetch cover crop, an appropriate interpretation is that the maximum output attainable with the conservative technology is only one-tenth of that obtainable within the optimistic technology.

Moving from the conservative technology to the most conservative technology, we find that 9 elements of C are also in I for the crimson clover cover crop while 7 elements of C are also in I for the hairy vetch cover. Thus, in both instances we significantly reduce the number of observations that are included in the reference technology by moving from the conservative to the most conservative technology. Moreover, the earlier pattern experienced with the move from the optimistic to the conservative technology is preserved—for both cover crops while differences are uncovered between the reference technologies.

VI. Conclusion

Dominance techniques have been used to develop more conservative free disposal hull reference technologies for data sets. The potential differences between these technologies have been examined by comparing the reference technologies for data drawn from agricultural field trials. Our results indicate that there are significant differences between the three reference technologies.

Table 1: Crimson Clover
(using days within 70-86° F as temp variable)

Obs.	y	F(x,y/E)	F(x,y/C)	F(E)/F(C)
3	35.69	0.3869	1.0000	0.3869
5	61.68	0.4911	0.7587	0.6472
13	104.32	0.8305	1.0000	0.8305
21	179.30	0.8909	1.0000	0.8909
23	113.67	0.7289	1.0000	0.7289
24	161.42	0.9038	1.0000	0.9038
28	165.76	0.8254	0.9864	0.8368
33	101.48	0.7594	1.0000	0.7594
35	148.79	0.9097	1.0000	0.9097
36	132.81	0.8120	0.9984	0.8133
37	142.50	0.8712	1.0000	0.8712
40	103.36	0.6844	0.9605	0.7126
41	88.31	0.6040	0.9531	0.6337
44	34.16	0.2474	1.0000	0.2474

Note: there are 48 observations in T(K)

Table 3: Crimson Clover

y	F(x,y/E)	F(x,y/R)	Ratio F((E)/F(R)
35.69	0.3869	1.0000	0.3869
61.68	0.4911	1.0000	0.4911
179.30	0.8909	1.0000	0.8909
113.67	0.7289	1.0000	0.7289
34.16	0.2474	1.0000	0.2474

Table 2: Vetch (using days within 70-86° F as temp variable) — **Table 4: Vetch**

	Table 2				Table 4			
		Effic. set Dist. to E	Nondom. set Dist. to C	Ratio				
Obs.	y	F(x,y/E)	F(x,y/C)	F(E)/F(C)	y	F(x,y/E)	F(x,y/RC)	F(E)/F(RC)
4	82.132	0.59631	1.00000	0.5963	82.132	0.59631	1.00000	0.5963
17	186.773	0.94593	1.00000	0.9459				
18	155.194	0.86300	0.99488	0.8674				
19	140.603	1.00000	1.00000	1.0000	140.603	1.00000	1.00000	1.00000
20	187.279	0.94243	1.00000	0.9424	187.279	0.94243	1.00000	0.9424
22	186.008	0.93603	1.00000	0.9360				
23	139.185	1.00000	1.00000	1.0000				
24	185.535	0.93966	1.00000	0.9397				
25	0.137	0.00077	0.00098	0.7857	0.137	0.00077	1.00000	0.0008
28	182.485	0.92193	1.00000	0.9219	182.485	0.92193	1.00000	0.9219
32	93.922	0.67508	1.00000	0.6751				
33	124.571	0.72686	1.00000	0.7269				
34	0.135	0.00090	1.00000	0.0009	0.135	0.00090	1.00000	0.0009
36	127.766	0.74550	1.00000	0.7455	127.766	0.74550	1.00000	0.7455
44	91.869	0.54168	0.75849	0.7142	91.869	0.54168	0.75892	0.7138

Note: there are 48 observations in $T(\mathcal{K})$

Appendix

Proof of lemma: That $T(\mathscr{E}) \subseteq T(\mathscr{K})$ is obvious. To prove the lemma we only need show that $T(\mathscr{K}) \subseteq T(\mathscr{E})$. Consider the case where $\mathscr{E}$ contains all but one element of $\mathscr{K}$, call it (x^k, y^k). Now consider any $(x, y) \in T(\mathscr{K})$: By definition

$$x_n \geqq \sum_{i \in K} \lambda_i x_{in} \qquad n = 1, ..., N$$

$$y_m \leqq \sum_{i \in K} \lambda_i y_{im} \qquad m = 1, ..., M$$

$$\sum_{i \in K} \lambda_i = 1, \qquad \lambda_i \geq 0 \quad i \in \mathscr{K}.$$

Because $(\mathbf{x}^k, \mathbf{y}^k) \notin \mathscr{E}$, there must exist an element of $\mathscr{E}$, call it $(\mathbf{x}^j, \mathbf{y}^j)$ such that $(\mathbf{x}^j, -\mathbf{y}^j) \leq (\mathbf{x}^k, -\mathbf{y}^k)$ whence $\sum_{\substack{i \in K \\ i \neq k}} \lambda_i x_{in} + \lambda_k x_{kn} \geqq \sum_{\substack{i \in K \\ i \neq k}} \lambda_i x_{in} + \lambda_k x_{jn} = \sum_{i \in \mathscr{E}} \lambda_i^* x_{in}$ $(n = 1, ..., N)$

where $\lambda_i^* = \lambda_i$ $(i \neq j)$ $\lambda_j^* = \lambda_j + \lambda_k$. Moreover, it also follows that $\sum_{i \in \mathscr{E}} \lambda_i^* y_{im} \leqq \sum_{i \in K} \lambda_i y_{im}$, $m = 1, ..., M$. Hence we have established that

$$x_n \geqq \sum_{i \in \mathscr{E}} \lambda_i^* x_{in} \quad n = 1, ..., N$$

$$y_m \leq \sum_{i \in \mathscr{E}} \lambda_i^* y_{im} \qquad m = 1, ..., M$$

$$\sum_{i \in \mathscr{E}} \lambda_i^* = 1$$

which implies $(\mathbf{x}, \mathbf{y}) \in T(\mathscr{E})$ and hence $T(\mathscr{K}) \subseteq T(\mathscr{E})$. It follows by induction that $T(\mathscr{K}) \subseteq T(\mathscr{E})$ when $\mathscr{K}$ contains an arbitrary number of elements that are not in $\mathscr{E}$.

Q.E.D.

References

Banker, R.D., A. Charnes, and W.W. Cooper, "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," Management Sci. 30, 9 (September 1984), 1078-1092.

Färe, R., S. Grosskopf, and C.A.K. Lowell, Production Frontiers, Cambridge University Press, New York, 1994.

Fried, H.O., C.A.K. Lovell, and P. Vander Eeckaut, "Evaluating the Performance of U.S. Credit Unions," J. of Banking and Finance, 17 (1993), 251-265.

Hougaard, J.L., and M. Tvede, "Intertemporal Dominance Analysis," Working Paper 5-93, Institute of Economics, Copenhagen Business School (1993).

Tulkens, H. and Vanden Eeckaut, "Non-Frontier Measures of Efficiency, Progress and Regress," CORE Discussion Paper No. 9155, Center for Operations Research and Economics, Universite Catholique de Louvain (December 1991).

DOMINAN.WP/mek/7-12-94