Enumerating Distance Functions for FDH technologies

Laurens Cherchye
Catholic University of Leuven
Center for Economic Studies
Naamsestraat 69
3000 Leuven, **BELGIUM**

Timo Kuosmanen
Helsinki School of Economics and Business Administration
Department of Economics and Management Science
P.O. Box 1210, FIN-00101 Helsinki, **FINLAND**

Thierry Post
Erasmus University Rotterdam
Department of Finance
Burg. Oudlaan 50
3062 PA Rotterdam, **THE NETHERLANDS**

Abstract:

We derive a general enumeration algorithm for computing directional distance functions relative to Free Disposable Hull (FDH) technologies. This general algorithm extends the enumeration schemes by Tulkens (1993) to alternative directions of measurement. The use of the proposed algorithm is illustrated by a simple numerical example.

Key Words: *Free Disposable Hull (FDH), directional distance function, enumeration*

1. Introduction

Free Disposable Hull (FDH) (Deprins et al., 1984) is a well-known empirical approximation of the production possibility set, which is based on minimal assumptions concerning the properties of the true but unobservable production set. In contrast to the popular Data Envelopment Analysis (DEA: Charnes et al., 1978) models, FDH is not restricted to convex technologies but applies to all monotonous ones. This is particularly convenient, given that it is frequently difficult to find a good theoretical or empirical justification for convexity (see e.g. Cherchye et al. (2000)).

One reservation often related to the practical implementation of FDH is the non-linear nature of the associated programming problems. DEA models can typically be solved through standard Linear Programming techniques. By contrast, FDH models generally require Mixed Integer Programming, which is more costly from a computational point of view. To remedy this problem, Tulkens (1993) introduced simple but effective enumeration algorithms for the classical Debreu-Farrell input and output measures.[1] These enumeration algorithms have proved useful and economical in many research situations, most notably in computationally hard simulation studies.

Recently, Chambers, Chung and Färe (1996,1998; see also Luenberger, 1992) have extended the usual notions of input and output distance functions by a generally formulated *directional distance function*. Attractively, this directional distance function encompasses a number of interesting efficiency gauges as special cases, most notably the classical Debreu-Farrell input and output measures and the McFadden (1978) gauge function. The general directional distance function formulation has already made inroads to various DEA applications (see e.g. Halme et al. (1999), and Halme and Korhonen (1999)). However, we have thus far not seen any application that uses the directional distance function with respect to the FDH set. To some extent this might be due to the computational burden associated with Mixed Integer Programming.

In this paper we extend the FDH enumeration possibilities by developing a general enumeration algorithm for computing values of the directional distance function relative to the FDH technology. Our formulation contains the earlier input and

[1] In a survey conducted among DEA experts by Seiford (1996), Tulkens' (1993) article was frequently mentioned as one of the most influential DEA papers.

output oriented enumeration schemes by Tulkens (1993) as its special cases, and hence can be viewed as a generalization of these schemes.

The rest of the paper unfolds as follows. In the next section we introduce the necessary notation and define the basic concepts. We also extend the dual interpretation of the directional distance function to the FDH approximation. In section 3 we derive the enumeration formula. In Section 4 we illustrate the usefulness of the proposed formula by a simple numerical example. Section 5 summarizes our conclusions.

2. Preliminaries

The standard input distance function by Shephard (1953) is defined as

(1) $$D_T(y,x) = \sup_{\theta}\left\{\theta \,\middle|\, (y, x/\theta) \in T\right\},$$

where T is the *production possibility set*

(2) $$T = \left\{(y,x) \in \Re_{+}^{m+l} \,\middle|\, \text{input } x \in \Re_{+}^{l} \text{ can produce output } y \in \Re_{+}^{m}\right\}.$$

The frequently employed Debreu-Farrell input efficiency measure is simply the inverse of (1). A more general distance metric, which contains (1) as its special case, is the *directional distance function* that has been introduced by Chambers et al. (1996, 1998). These authors propose to measure distance to the production frontier in some pre-assigned input-output direction $(g^y, g^x) \in \Re^{m+l}$. This directional distance function can be defined as

(3) $$d_T(y,x;g^y,g^x) = \sup_{\theta}\left\{\theta \,\middle|\, (y+\theta g^y, x-\theta g^x) \in T\right\}.$$

The input distance function (1) is obtained from (3) as $D_T(y,x) = (1-d_T(y,x;0,x))^{-1}$. Similarly, the corresponding output distance function is obtained as $1+d_T(y,x;y,0)$, whereas the McFadden (1978) gauge function is obtained as $(1+d_T(y,x;y,-x))^{-1}$.

In empirical studies, the production set T is typically unknown. In the nonparametric literature, empirical approximations for T are constructed from a set of observations drawn from the technology. The production data is henceforth represented by the output matrix $Y = (y^1 \dots y^n)^T$, with $y^j = (y_1^j \cdots y_l^j)$, and the input matrix $X = (x^1 \dots x^n)^T$, with $x^j = (x_1^j \cdots x_m^j)$. In addition, we use the index sets $L = \{1,\dots,l\}$ for inputs, $M = \{1,\dots,m\}$ for outputs, and $S = \{1,\dots,n\}$ for production units.

The FDH approximation is obtained from two minimal assumptions about T. First, T is assumed to envelop all observed data. Second, T is assumed monotonous, i.e. *all* inputs and outputs are freely (or strongly) disposable. In other words, the marginal products of inputs, marginal rates of substitution between inputs, and marginal rates of transformation between outputs are assumed to be non-negative. Of course, congestion of production factors can violate this assumption. But as Färe and Grosskopf (1983) have pointed out, imposing monotonicity can be interpreted as a congestion adjustment to the production set, i.e. distance functions for 'monotonized' technologies include both pure technical efficiency and congestion components. Alternatively, monotonicity can be motivated by the fact that it does not interfere with the Pareto-Koopmans classification of technical efficiency.

The minimal set that complies with these two assumptions gives the FDH approximation, formally defined as

$$\text{(4)} \qquad \hat{T}_{FDH}^{Y,X} = \left\{ (y,x) \middle| x \geq lX ; y \leq lY ; le = 1 ; l^j \in \{0,1\} \forall j \in S \right\}.$$

Note the similarity between $\hat{T}_{FDH}$ and the frequently employed convex monotone hull, which differs from FDH only in that it allows for convex combinations of production plans. Specifically, we obtain the convex monotone hull from (4) by substituting the constraint $l^j \in \{0,1\} \forall j \in S$ by $l \geq 0$.

Distance functions and efficiency gauges can be measured relative to the FDH approximation when the true technology is unknown. For instance, substituting T by the empirical FDH approximation in (3) gives the following empirical distance function for some direction $(g^y, g^x) \in \Re^{m+l}$

$$\text{(5)} \qquad \begin{aligned} &\hat{d}_{FDH}^{Y,X}(y,x;g^y,g^x) = \\ &\underset{q}{Sup} \left\{ q \middle| x - qg^x \geq lX ; y + qg^y \leq lY ; le = 1 ; l^j \in \{0,1\} \forall j \in S \right\}. \end{aligned}$$

Provided that the assumptions of envelopment and monotonicity are satisfied, the FDH set is contained within the true production set, i.e. $\hat{T}_{FDH}^{Y,X} \subseteq T$, and consequently $\hat{d}_{FDH}^{Y,X}(y,x;g^y,g^x) \leq d_T(y,x;g^y,g^x)$ for all (g^y,g^x) and $(y,x) \in T$. In other words, using the FDH approximation gives a lower bound for the true distance function value, which is an important property for many practical purposes.

One of the attractive features of the standard input distance function is its dual interpretation in terms of cost efficiency. Chambers et al. (1998) derived a similar dual interpretation of the directional distance function as a normalised profit difference. However, these authors restricted attention to convex technologies. We here extend this result for the monotonous but non-convex FDH set.

We start from expression (5), which can equivalently be expressed as

$$\text{(6)} \qquad \hat{d}_{FDH}^{Y,X}(y,x;g^y,g^x) = \underset{j \in S}{Sup} \left[\underset{q}{Sup} \left\{ q \middle| qg^x \leq x - x^j ; qg^y \leq y^j - y \right\} \right].$$

This clearly reveals that the directional distance function could also be computed by solving *n* linear programming problems (one for each $j \in S$). Using duality theory, one could also express the embedded problem for each $j \in S$ as

$$\begin{aligned}(7)\qquad & \sup_{q}\left\{q \,\middle|\, qg^{x} \le x - x^{j}; qg^{y} \le y^{j} - y\right\} \\ & = \inf_{(m,n)\in\Re_{+}^{m+l}}\left\{m\left(x - x^{j}\right) + n\left(y^{j} - y\right) \,\middle|\, mg^{x} + ng^{y} = 1\right\}.\end{aligned}$$

As a result, we can re-express (6) as

$$\begin{aligned}(8)\qquad & \hat{d}_{FDH}^{Y,X}(y, x; g^{y}, g^{x}) \\ & = \sup_{j\in S}\left[\inf_{(m,n)\in\Re_{+}^{m+l}}\left\{\left(ny^{j} - mx^{j}\right) - \left(ny - mx\right) \,\middle|\, mg^{x} + ng^{y} = 1\right\}\right].\end{aligned}$$

This expression yields an economic interpretation of the FDH distance function as a normalised difference between reference and actual profit.[2] It is to be recalled that FDH can account for a broad range of (possibly non-linear) objective functions (see e.g. Cherchye et al. (2000) and Wunsch (1996)).[3] Evidently, due to the specific construction of the FDH reference set, all shadow weight will usually be attributed to only one input or output (i.e. only one component of the vector (m,n) will be strictly positive). However, if some additional information about possible (relative) price, cost or revenue levels is available, one could integrate weights restrictions in problem (8) (see e.g. Pedraja-Chaparro, Salinas-Jimenez and Smith (1997) for a recent overview of alternatives).

Another attractive feature of the FDH set from a practical point of view is that for each inefficient producer a physically existing efficient reference unit can be identified that proves superior in all input and output dimensions, which is not generally true for the standard DEA models. However, the constraint $l^{j} \in \{0,1\} \forall j \in S$ that distinguishes FDH from convex DEA models also causes computational inconvenience. In general, computing FDH-based distance functions involves Mixed Integer Programming. To remedy this problem, Tulkens (1993) has provided a simple but effective enumeration algorithm for the input and output distance functions. In this procedure, the set of observations weakly dominating observation *k* is used as the reference set of *k*, i.e.

$$(9)\qquad R(y^{k}, x^{k}) = \left\{i \in S \,\middle|\, y^{i} \ge y^{k}; x^{i} \le x^{k}\right\}.$$

Following Tulkens (1993), the input distance function can be computed as

---

[2] Chambers et al. (1998) obtained an Inf-Sup structure whereas we obtain a Sup-Inf structure. Indeed, when the technology would be truly convex then the order of the Inf and Sup operators could be reversed.

[3] Of course, the precise interpretation of distance function in a more general setting depends on whether -and in what sense- the objective function value reacts in a *homogenous* way to input and output changes. The appropriate homogeneity condition further depends on the direction vector.

(10) $$\hat{D}^{Y,X}_{FDH}(y^k,x^k) = \underset{i\in R\left(y^k,x^k\right)}{Max}\left[\underset{j\in L}{Min}\left\{\frac{x^k_j}{x^i_j}\right\}\right].$$

This can be solved easily by enumeration.

For a similar formulation for the output distance function, see Tulkens (1993). In the next section we extend this enumerative principle to the more general directional distance function.

3. General enumeration formula

Note first that the value of the directional distance function for a production vector contained within the production set lies in the half-open interval $[0,\infty)$, and hence need not necessarily be finite. The following lemma characterizes the conditions for existence of a finite real-valued solution.

Lemma 1: Let Y and X represent input and output matrices composed of nonnegative real numbers. For any real-valued $(y,x)\in \hat{T}^{Y,X}_{FDH}$, the value of the distance function $\hat{d}^{Y,X}_{FDH}(y,x;g^y,g^x)$ will be a finite nonnegative real number, if at least one component of the direction vector (g^y,g^x) is strictly positive. Otherwise, $\hat{d}^{Y,X}_{FDH}(y,x;g^y,g^x)$ will be infinitely large.

PROOF: The reference set $\hat{T}^{Y,X}_{FDH}$ is bounded from above in output space by the vector $\overline{Y} \geq y^j\ \forall j\in S$ and is bounded from below in input space by the vector $\overline{X} \leq x^j\ \forall j\in S$. If there exists an output direction *k* with $g^y_k > 0$, then the condition $y_k + Jg^y_k \leq \overline{Y}_k$ is satisfied only for a finite nonnegative J. Alternatively, if there exists an input direction *i* with $g^x_i > 0$, then the condition $x_i - Jg^x_i \geq \overline{X}_i$ is satisfied only for a finite nonnegative J. As $\hat{T}^{Y,X}_{FDH}$ is contained within $\left(\overline{Y}+\Re^l_+, \overline{X}-\Re^m_+\right)$, we thus have that $\hat{d}^{Y,X}_{FDH}(y,x;g^y,g^x)\ \leq J$, and J was already shown to be a finite nonnegative real number if at least one component of the vector (g^y,g^x) is strictly positive.

By contrast, if $(g^y,g^x)\leq 0$, then the inequalities $y+qg^y\leq y$ and $x-qg^x\geq x$ hold for all $(y,x)\in \hat{T}^{Y,X}_{FDH}$ and $q>0$. Indeed, free disposability implies that $(y+qg^y,\ x-qg^x)\in \hat{T}^{Y,X}_{FDH}$ for all $q>0$, and $\hat{d}^{Y,X}_{FDH}(y,x;g^y,g^x)$ will consequently become infinitely large. *Q.E.D.*

Suppose that we evaluate a vector $(y,x)\in \hat{T}^{Y,X}_{FDH}$, and that at least one element of the direction vectors is strictly positive, which by Lemma 1 suffices to guarantee that the distance function value is computable. This allows us to write (6) equivalently as

(11) $\hat{d}^{Y,X}_{FDH}(y,x;g^y,g^x) = \underset{q\in\Re;\, j\in S}{Max}\left\{q \,\middle|\, y + qg^y \le y^j ; x - qg^x \ge x^j\right\}.$

We can now prove the following:

Proposition 1: Let Y and X represent input and output matrices composed of nonnegative real numbers, and let $(y,x)\in \hat{T}^{Y,X}_{FDH}$. If at least one component of the direction vector (g^y, g^x) is strictly positive, we have:

(i) $\hat{d}^{Y,X}_{FDH}(y,x;g^y,g^x) = \underset{j\in Q(y,x;g^y,g^x)}{Max}\, a^j,$

where

$$Q(y,x;g^y,g^x) = \left\{ j\in S \,\middle|\, y^j_k - y_k \ge 0\ \forall g^y_k = 0, k\in M; x_i - x^j_i \ge 0\ \ \forall g^x_i = 0, i\in L\right\},$$

$$a^j = \begin{cases} b^j, & b^j \ge g^j \\ 0, & b^j < g^j \end{cases},$$

$$b^j = Min\left\{ \underset{k\in M: g^y_k>0}{Min}\left(\frac{y^j_k - y_k}{g^y_k}\right), \underset{i\in L: g^x_i>0}{Min}\left(\frac{x_i - x^j_i}{g^x_i}\right)\right\},$$

and

$$g^j = \begin{cases} Max\left\{ \underset{k\in M: g^y_k<0}{Max}\left(\frac{y^j_k - y_k}{g^y_k}\right), \underset{i\in L: g^x_i<0}{Max}\left(\frac{x_i - x^j_i}{g^x_i}\right)\right\}, \text{if } \exists k\in M: g^y_k < 0 \vee i\in L: g^x_i < 0 \\ 0, \text{otherwise} \end{cases}.$$

PROOF: If $g^y_k = 0$, then the restriction $y_k + qg^y_k \le y^j_k$ reduces to $y_k \le y^j_k$. Similarly, if $g^x_i = 0$, then $x_i - qg^x_i \ge x^j_i$ reduces to $x_i \ge x^j_i$. Hence, the reference unit is restricted to the DMUs in $Q(y,x;g^y,g^x)$. In addition, the restrictions $y_k + qg^y_k \le y^j_k$, $k\in M: g^y_k > 0$, and $x_i - qg^x_i \ge x^j_i$, $i\in L: g^x_i > 0$, can be expressed as $q \le b^j$. Similarly, the restrictions $y_k + qg^y_k \le y^j_k$, $k\in M: g^y_k < 0$, and $x_i - qg^x_i \ge x^j_i$, $i\in L: g^x_i < 0$, can be expressed as $q \ge g^j$. Combining these insights gives

(ii) $\hat{d}^{Y,X}_{FDH}(y,x;g^y,g^x) = \underset{q\in\Re;\, j\in Q(x,y)}{Max}\left\{q \,\middle|\, g^j \le q \le b^j\right\}.$

If $g^j \le b^j$, the solution is simply b^j. By contrast, if $g^j > b^j$, there is no feasible solution for observation *j*. However, in this case a^j can harmlessly be set equal to zero since that is the minimal distance secured by the assumption $(y,x)\in \hat{T}^{Y,X}_{FDH}$. Combining these insights, we find that (ii) equals (i). *Q.E.D.*

The problem (i) in Proposition 1 can be easily solved by enumeration. Note that in case the direction vector (g^y, g^x) is strictly positive, the reference set

$Q(y, x; g^y, g^x) = S$ and $g^j = 0 \forall j \in S$, so the problem (i) reduces to $\underset{j \in S}{Max}\, b^j$. From this expression, the analogy with the Tulkens formulation (10) should be evident.

4. A Numerical Example

We illustrate the practical use of the general enumeration formulation (12) by a simple numerical example. Consider the single-input single-output data set of 9 DMUs reported in Table 1. It should be clear, however, that the enumeration algorithm applies with equal strength to situations with multiple inputs and outputs.

Table 1: The example data set

Observation *j*	A	B	C	D	E	F	G	H	I
Input *x*	3	6	4	6	5	8	12	14	18
Output *y*	4	5	6	7	8	9	11	13	14

For sake of brevity, we focus on observation B. To illustrate the use of enumeration for other measures than the standard Debreu-Farrell ones, we consider the direction vector $(g^y, g^x) = (5, -6)$. That is, we look for the maximal radial augmentation of both the input and the output of B. This direction corresponds to the McFadden gauge function (see Section 2). A GAUSS computer code for calculating the values of this gauge function by enumeration is reported in Appendix.

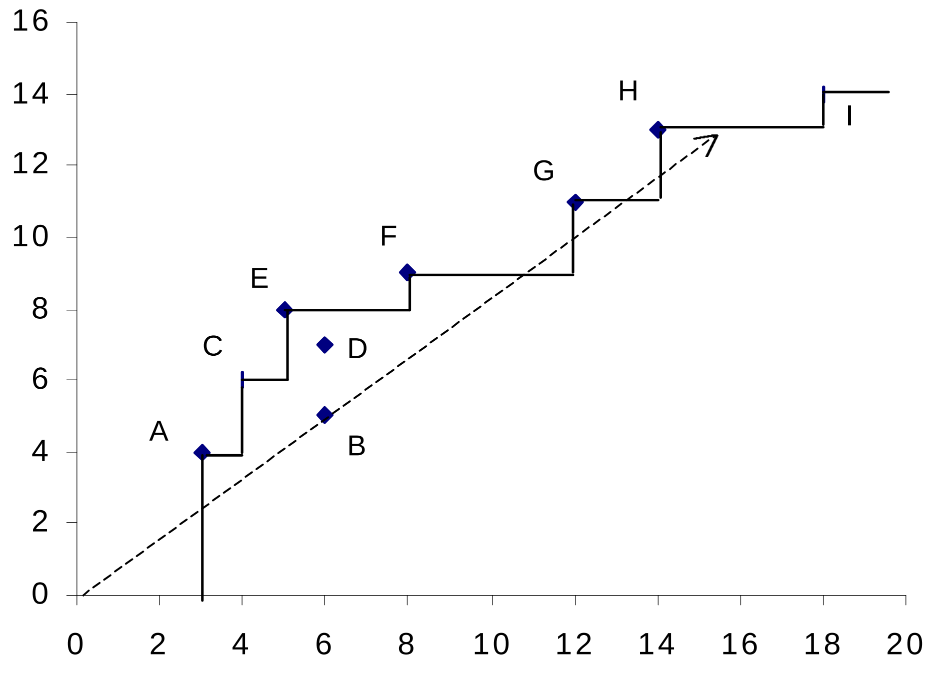

Figure 1: McFadden gauge of B relative to FDH

Figure 1 illustrates the McFadden projection procedure for observation B relative to the FDH set. Computation of the McFadden gauge function is a relatively challenging task. Figure 1 makes clear that the optimal reference production plan H does not dominate observation B in the input dimension. Observe further that the gauge function has two additional local optima (observations F and G) that are dominated by the global optimum H. This feature may render some interior point MILP algorithms useless.

As the direction vector is composed of non-zero elements, the reference set Q includes all observations. Table 2 reports the auxiliary indicators a^j, b^j, and g^j , $j = A,\ldots, I$ used in the enumeration. Recall that b involves the positive directions (here output) and g involves the negative directions (here input). Specifically, g^j is obtained as $(6 - x^j) / -6$, that is the normalized difference between the inputs of observation B and observation j. Similarly, b^j is obtained as $(y^j - 5) / 5$, that is the normalized difference between the outputs of observation j and observation B. In the bottom row, a^j equals b^j if $b^j \geq g^j$, otherwise a^j is set equal to zero. In this example, only for observation I $b^I < g^I$ (compare to Figure 1). The value of distance function is then simply computed as the maximum of a^j in the bottom row. It yields 8/5, obtained relative to observation H.

Table 2: Auxiliary indicators for enumeration

Observation *j*	A	B	C	D	E	F	G	H	I
g^j	-1/2	0	-1/3	0	-1/6	1/3	1	4/3	2
b^j	- 1/5	0	1/5	2/5	3/5	4/5	6/5	8/5	9/5
a^j	- 1/5	0	1/5	2/5	3/5	4/5	6/5	**8/5**	0

5. Conclusions

We have derived a general enumeration formula for computing directional distance functions relative to the FDH reference technology. This enumeration formula can compute distance functions in any meaningful direction by enumeration, and hence avoids having to solve computationally very demanding Mixed Integer Programming problems that are usually associated with the FDH approximation.

References

Chambers, R. G., Y. Chung, and R. Färe (1996): Benefit and Distance Functions, *Journal of Economic Theory* 70, 407-418

Chambers, R. G., Y. Chung, and R. Färe (1998): Profit, Directional Distance Functions, and Nerlovian Efficiency, *Journal of Optimization Theory and Applications* 98(2), 351-364

Cherchye, L., T. Kuosmanen, and G. T. Post (2000): What is the Economic Meaning of FDH? A Reply to Thrall, *Journal of Productivity Analysis* (forthcoming)

Deprins, D., L. Simar and H. Tulkens (1984): Measuring Labor Efficiency in Post Offices, p. 243 - 267 in M. Marchand, P. Pestieau and H. Tulkens (Eds.), *The Performance of Public Enterprises: Concepts and Measurements*, North Holland.

Färe, R. and S. Grosskopf. (1983). “Measuring Congestion in Production.” *Zeischrift fur Nationalökonomie* 43, 257-271

Halme, M,.T.Joro, P.Korhonen, S.Salo and J.Wallenius (1999):A Value Efficiency Approach to Incorporating Preference Information in Data Envelopment Analysis, *Management Science* 45(1), 103-115

Halme, M,.and P.Korhonen (1999): Restricting Weights in Value Efficiency Analysis, *European Journal of Operational Research* (forthcoming)

Luenberger, D.G. (1992): Benefit Functions and Duality, *Journal of Mathematical Economics* 21, 261-481

McFadden, D. (1978). “Cost, Revenue and Profit Functions.” In M. Fuss and D. McFadden (eds.), *Production Economics: A Dual Approach to Theory and Applications*. Amsterdam: North-Holland.

Pedraja-Chaparro, F., J. Salinas-Jimenez and P. Smith (1997): On the Role of Weight Restrictions in Data Envelopment Analysis, *Journal of Productivity Analysis* 8, 215-230.

Seiford, L.M. (1996): Data Envelopment Analysis: The Evolution and the State of the Art (1978-1995), *Journal of Productivity Analysis* 7, 99-137

Shephard, R.W. (1953): *Cost and Production Functions*, Princeton University Press, Princeton

Tulkens, H. (1993): On FDH Efficiency Analysis: Some Methodological Issues and Application to Retail Banking, Courts, and Urban Transit, *Journal of Productivity Analysis* 4(1), 183-210.

G.P. Wunsch (1996): *Peer Comparison and Regulation*, Doctoral dissertation, CORE, Louvain-la-Neuve.

APPENDIX

In this appendix we present a GAUSS computer code for computing
$d^{Y,X}_{FDH}\left(y^j, x^j; y^j, -x^j\right)\ \ j \in S$.

```
PROC(1)=McFadden(Y,X);

LOCAL a,b,c,d,e,j,n;

n=ROWS(X);
d=ONES(n,1)*9999;
j=1;
DO WHILE j .LE n;

    b=MINC(((Y-Y[j])./Y[j])');
    c=MAXC(((X-X[j])./X[j])');
    e=b~c;
    e=SELIF(e, e[.,1] .GE e[.,2]);
    a=e[.,1];
    d[j]=MAXC(a);
    j=i+1;

ENDO;
RETP(d);
ENDP;
```

The explanation proceeds as follows. The vector d contains the directional distance function value for each observation $j \in S$, i.e. $\mathrm{d} = (\hat{d}^{Y,X}_{FDH}(y^1, x^1; y^1, -x^1) \cdots \hat{d}^{Y,X}_{FDH}(y^n, x^n; y^n, -x^n))^T$. In addition, we use the following notation: $b = (b^1 \cdots b^n)^T$ and $c = (g^1 \cdots g^n)^T$. Finally, $e = \begin{pmatrix} b & c \end{pmatrix}$ is an auxiliary matrix, used to construct *a*, which is a column vector containing a^j for all observations for which $b^j \geq g^j$.