対応のない繰り返しデータによるリコース付きＤＥＡ

(申請中)　神戸大学　＊　道田 英雄　MICHIDA Hideo
01604524　神戸大学　森田 浩　MORITA Hiroshi
01501824　神戸大学　藤井 進　FUJII Susumu

1. はじめに

多入力多出力システムの効率性評価の手法であるDEA (Data Envelopment Analysis) は従来、対応関係をもつ確定したデータの評価に用いられてきた。しかし、データがもつ不確実性を考慮した評価を行うために新たな確率的DEAのモデルが必要となる。確率的データを数理計画問題に組み込んで最適化を図る手法として確率計画法があり、機会制約条件のモデルに関しての研究などがなされている。

本研究ではデータがもつ不確実性を考慮するため、繰り返し測定によりデータの測定を行う。そして確率的要素を含むデータから効率性評価を行うための確率計画法をとりいれたDEAの定式化から、対応のないデータの組を考慮した場合の効率値の振る舞いを考察する。

2. リコース付きDEAの定式化と解法

本研究でDEAの評価モデルとして用いるCCRモデルの双対問題を以下に示す。

$$\text{CCR}\left|\begin{array}{ll}\text{Minimize} & \theta-\epsilon(s_x+s_y)\\ \text{subject to} & \theta X_o - X\lambda - s_x = 0\\ & Y_o - Y\lambda + s_y = 0\\ & \theta\ge 0, \lambda\ge 0, s_x\ge 0, s_y\ge 0\end{array}\right.$$

入出力データ (X, Y) が不確実性を含む場合、CCRモデルだけでは効率性評価ができないため、確率計画法のリコース問題にあてはめる。簡略化のため入力 $X(w)$ が確率的、出力 Y は確定的としてCCRモデルをリコース問題にあてはめる。出力を確率的としてもよい。リコース最適値 $Q(\theta,\lambda,s_x,\omega)$ は

$$Q(\theta,\lambda,s_x,\omega)=$$

$$\min_{z\ge 0}\{q'(w)z\,|q'(w)z=\theta X_o(w)-X(w)\lambda-s_x(w)\}$$

で表現され、その期待値 $Q(\theta,\lambda,s_x)$ を目的関数に組み込むことで、リコースを用いたDEAモデルの定式化ができる。ここで θ は $0\le\theta\le 1$ の変数であるため、$q'(w)=\frac{1}{X_o(w)}$ としてリコース値が大きくなりすぎないように正規化している。

$$\text{REC}\left|\begin{array}{ll}\text{Minimize} & \theta+Q(\theta,\lambda,s_x)-\epsilon(s_x+s_y)\\ \text{subject to} & Y_o - Y\lambda + s_y = 0\\ & \theta\ge 0, \lambda\ge 0, s_x\ge 0, s_y\ge 0\end{array}\right.$$

このモデルにより効率性評価を行うと、効率値以外にデータのばらつきの指標を表すリコース値が得られる。リコース値の存在が効率性評価に何らかの影響を及ぼしていると考えられるが、その検討は今後の課題である。

リコース問題の確率変数が離散的である場合、L 型法 (L-shaped method) と呼ばれる効率的な解法で解くことができる。確率変数の観測値が $(X^l,Y^l), l=1,\ldots,L$ という離散的な値をとるとき、それぞれを実現する確率を p_l とする。すると

$$Q(\theta,\lambda,s_x)=\sum_{l=1}^{L}p_lQ(\theta,\lambda,s_x,\omega^l)$$

として目的関数に組み込むことができる。L 型法による解法は、2種類のカット (feasibility cut と optimality cut) を順次加えていく切除平面法であり、有限回の繰り返しの後、最適解が得られる。

3. 繰り返しデータの対応関係に応じた効率性評価

従来、DEAの効率性評価は1回の測定データや、複数回の測定の平均値といった代表値を用いて行ってきた。しかし、これではデータが不確実性をもつ場合、そのばらつきを知ることができない。データのばらつきを考慮した評価を行うためには繰り返し測定したデータを代表値などに変換して評価するのではなく、そのまま利用することが有効である。

入出力データに不確実性をもつ K 個のDMUが存在するとき、DMU k に関して L_k $(k=1,\ldots,K)$ 回の測定を行い入出力データが $(X_k^l, Y_k^l), l=1,\ldots,L_k$ と得られたとき、DMU k の入出力値は確率 $p_k=1/L_k$ で生起す

る離散的確率変数とみなすことができる。このとき、リコースを用いたDEAの定式化にデータをあてはめて効率性評価を行うことができる。DMU k の測定回数 L_k はDMUごとに異なっても、同じでもよい。しかし、必ずしもデータの対応関係を一意に定めることはできない。データに対応がある場合とない場合についてデータセットの構成方法を検討する必要がある。

・データの対応がある場合

対応関係が定まっているとき、$L_k = L$ の条件が成立していれば L 通りのデータセットが一意に決まる。

・データの対応がない場合

$L_k \neq L$ であるときは常にデータの対応関係は定まらない。また、$L_k = L$ であってもデータのとり方次第ではデータの対応関係が定まらないこともある。このとき、入出力データの組合せは $N(= L_1 \times L_2 \times \ldots \times L_K)$ 個存在し、それぞれ確率 $1/N$ で生起するものと考えることができる。N 組のデータセットに対して効率性評価を行うのは多くの計算上望ましくないため、その部分集合をからデータセットをランダムに構成することができる。すなわち、l 組目のデータセットを

$$X^{(l)} = (X_1^{(l)}, X_2^{(l)}, \ldots, X_K^{(l)})$$

とする。ここで、$X_k^{(l)}$ は $(X_k^1, X_k^2, \ldots, X_k^{L_k})$ からのランダムサンプルである。この部分集合の大きさを n としたとき、n 組のデータセットから効率性評価するのを1回の試行とすると、各試行ごとに選ばれるデータセットは異なるため、効率性評価も各試行ごとに変化する。この方法で効率値を推定するためには、n の増加と効率値のばらつきを調べる必要がある。

4. 数値例

2入力1出力の6つのDMU群を考え、図1のように各DMUで3回繰り返し観測した($L_k = 3$, $k = 1, \ldots, 6$)。各観測データに対応関係がないなら、$N = 3^6$ 通りの組合せが考えられる。データセットの大きさを変えて効率性評価を行った。この試行をそれぞれの n について10回ずつ行い、DMU Aについての結果を図2と表1に示す。図2の縦軸の θ はリコースを考えたときの効率値、横軸は対応づけたデータセットの大きさである。繰り返しのデータセットの数を増加させると効率値は徐々に収束しているのがわかる。もし、$n = 3^6$ 通りのデータセットを用いて同様の試行を10回行った場合、全ての対応パターンを網羅してるため、10回とも同じ効率値が得られる。n の値を大きくすれば、図2や表1の標準偏差より効率値のばらつきは小さくなることが確認できた。推定精度とデータセットの大きさならびに得られた効率値とリコース値の関係に関する解析は今後の課題である。

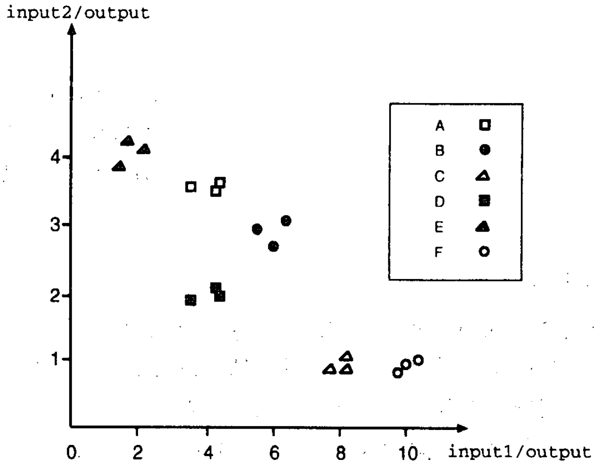

図 1: 繰り返しデータをもつDMU群

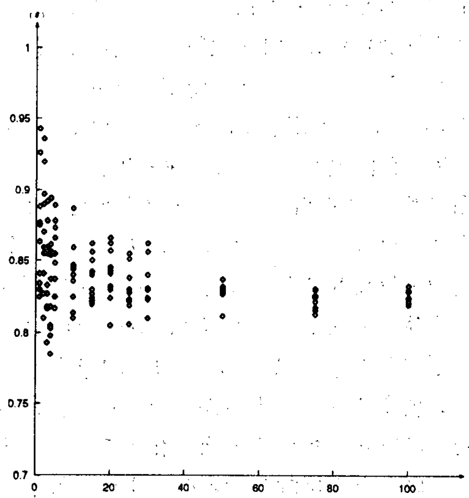

図 2: DMU A の測定結果

表 1. 平均値と標準偏差

n	1	10	25	50	100
平均値	0.870	0.841	0.830	0.827	0.825
標準偏差	0.040	0.022	0.015	0.008	0.005

参考文献

J. R. Birge, and F. Louveaux (1997), "Introduction to stochastic programming", Springer.

刀根 薫,"経営効率性の測定と改善–包絡分析法DEAによる–", 日科技連,1993.

森田，道田，藤井，"リコースを用いた確率的ＤＥＡ"，日本OR学会春季研究発表会アブストラクト集，1998.