

DEA の図解

森田 浩

1. はじめに

データ包絡分析法 (DEA; Data Envelopment Analysis) はさまざまな活動のパフォーマンスを評価する手法として広く適用されている[1][5]. DEA は線形計画法をベースとした評価手法であり, 数理計画モデルとして, その拡張や統計的手法などとの親和性に優れていることから, 適用される状況に応じてさまざまなモデルを考えることができる. そのため, 事業体の効率性評価に始まって, ベンチマークや合意形成など, モデルや評価法が次々と開発されるとともに, その適用範囲はますます拡大しつつある.

多入力多出力をもつ活動体があるとき, それらの活動可能な集合を表す生産可能集合 (production possibility set) の有効な境界面を効率的フロンティア (efficient frontier) とよぶ. 各活動体がこのフロンティア上にあるかどうかによって効率的かどうかを判定するとともに, どれだけ離れているかによって非効率性の程度を測っているのが DEA である.

本稿では, 可変ウェイトによる効率性の定義からではなく, その双対系となっている効率的フロンティアと, 効率性を測るための効率性尺度 (efficiency measure) の図解を通して, DEA 評価法の仕組みを概観する. また階層型 DEA による階層フロンティアの構成法についても紹介する.

2. 効率的フロンティア

m 個の入力項目 x と s 個の出力項目 y によって表される活動をもった事業体 (DMU) が n 個あるものとする. 各事業体の活動は $(x_j, y_j), j=1, 2, \dots, n$ と表す. このとき実行可能な活動の集合を示すのが生産可能集合 P で, 一般的なモデルでは, 以下の仮定が置かれている.

- (i) $(x_j, y_j) \in P, j=1, 2, \dots, n$
- (ii) $(x, y) \in P \Rightarrow (tx, ty) \in P$ for any $t > 0$
- (iii) $(x, y) \in P \Rightarrow (\bar{x}, \bar{y}) \in P$ for $\bar{x} \geq x, \bar{y} \leq y$
- (iv) $(x_1, y_1), (x_2, y_2) \in P \Rightarrow$
 $(\lambda x_1 + \bar{\lambda} x_2, \lambda y_1 + \bar{\lambda} y_2) \in P$ for $0 < \lambda < 1$

このとき, 生産可能集合 P は行列 $X=(x_j), Y=(y_j)$ を用いて

$$P = \{(x, y) | x \geq X\lambda, y \leq Y\lambda, \lambda \geq 0\} \quad (1)$$

と表される.

生産可能集合の表す活動の中で, 効率的な活動の集合が効率的フロンティアである. 少ない入力で多くの出力があることが効率的であると評価されるため, 生産可能集合において, 入力が減少し出力が増加する方向における有効な境界面が効率的フロンティアとなる.

(ii)の仮定は規模の収穫が一定 (CRS; constant return to scale) であることを表しており, 規模の収穫が変動 (VRS; variable return to scale) する場合にはこの仮定は外される. このときの生産可能集合は

$$P = \{(x, y) | x \geq X\lambda, y \leq Y\lambda, e'\lambda = 1, \lambda \geq 0\} \quad (2)$$

となる. ここで $e=(1, 1, \dots, 1)$ である. 一般に, 生産可能集合は

$$P = \{(x, y) | x \geq X\lambda, y \leq Y\lambda, L \leq e'\lambda \leq U, \lambda \geq 0\} \quad (3)$$

と表され, $L=0, U=\infty$ のときが CRS, $L=U=1$ のときが VRS となり, L と U の設定によって規模の収穫に関する仮定が規定される (GRS; generalized return to scale).

これらの生産可能集合はいずれも観測されたデータの線形結合によって構成されているのに対し, 実際に観測されたデータのみから構成する FDH (free disposal hull) モデルもある. 線形結合の点は実際に観測されているわけではなく, そのような活動が可能であるという仮説の上に存在している点である. そこで (i)の仮定が成立する最小の集合が FDH であり,

$$P = \{(x, y) | x \geq X\lambda, y \leq Y\lambda, e'\lambda = 1, \lambda \in \{0, 1\}\} \quad (4)$$

と表される.

4つのモデルに対して, 1入力1出力の場合の生産可能集合と効率的フロンティアを図1に例示している.

もりた ひろし

大阪大学 大学院情報科学研究科
〒565-0871 吹田市山田丘2-1

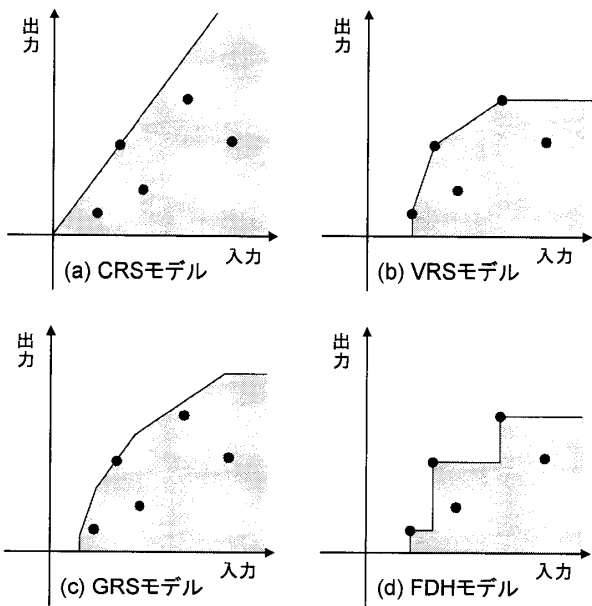


図1 生産可能集合と効率的フロンティア

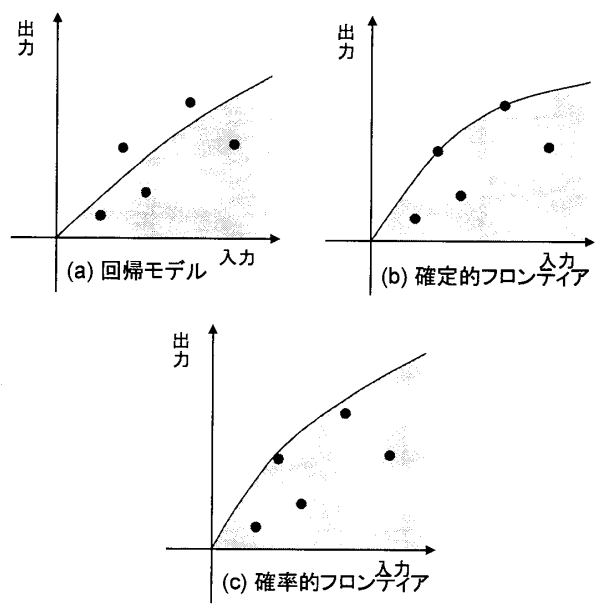


図2 パラメトリック手法のフロンティア

5つの活動があるときに構成される生産可能集合が示されているが、CRSが最も大きく、FDHが最も小さくなっている。線形性を仮定したときに最小となるのがVRSである。GRSは $L=0.8$ 、 $U=1.2$ のときを示しているが、CRSとVRSの間となっている。それぞれの効率的フロンティアは実線で表されていて、生産可能集合の有効な境界面となっている。これらはデータを包絡するようにノンパラメトリックに決められるもので、関数形を特定しているものではない。

確率的フロンティア法 (stochastic frontier analysis) は、生産関数 f によって入出力の関係を表すパラメトリックな手法である。統計的回帰モデルでは、入力 x と出力 y の間にある関数形 $y=f(x)$ を定め、これに誤差項 ε を含んで $y=f(x; \theta)+\varepsilon$ と観測されるデータから未知母数 θ を推定している。確率的フロンティア法では、誤差項を2つの要素に分解している。

$$\varepsilon = u - v \quad (5)$$

ここで、 u は通常の誤差項であるが、 v は非負の値を取るもので生産関数に達しなかった量を表しており、指数分布とか半正規分布などの分布を仮定している。これが非効率性の度合いを示すものになる。例えば、2入力の場合のCobb-Douglas型の生産関数では

$$y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} e^{u-v} \quad (6)$$

と表される。

$\varepsilon = u$ とすると、非効率性を考えない一般的回帰分析となる。 $\varepsilon = -v$ とすると、生産関数からの乖離はすべて非効率性とみなされ、確定的フロンティアとな

る。図2には、1入力1出力の場合のこれら3つの生産関数を例示している。

3. 効率性の尺度

一般に、効率的フロンティア上にあるDMUを効率的であるという。効率的フロンティア上になく生産可能集合の内部にあるDMUや、効率的フロンティア上にあっても入力の余剰や出力の不足(スラック)のあるDMUは、非効率性といわれる。効率的か非効率性か、あるいはその度合いを示すものが効率性の尺度である。

各々のDMUの効率性は、効率的フロンティアまでの距離をどう測るかで決まる。DEAにおける効率性尺度には、比率で表すradial尺度と差分で表すnon-radial尺度がある。

Radial尺度は、入力あるいは出力を何倍しても生産可能集合に含まれているかをみるもので、入力を減らす方向に見るか出力を増やす方向に見るかで、入力指向あるいは出力指向のモデルが考えられる。入力指向では入力を一律に $\theta(\leq 1)$ 倍しても生産可能集合 P に属するような θ の最大値を求めることになる。すなわち、入力を $\theta(\leq 1)$ 倍にしてもまだ活動できる余力があるということであり、これが非効率性に相当すると考えている。このことは、評価対象のDMU $_o$ の活動を (x_o, y_o) と表すと、

$$(\theta x_o, y_o) \in P \quad (7)$$

となる最小の θ の値 θ^* を求めることになり、

$$\min\{\theta | (\theta x_o, y_o) \in P\} \quad (8)$$

と表される。したがって、いろいろな生産可能集合 P に対して(8)式を展開すると DEA でよく知られている数理計画問題が得られる。例えば、CRS モデルにおける θ^* は、(1)式の P を用いて、次の線形計画問題によって求められる。

$$\begin{aligned} \min \quad & \theta \\ \text{subject to} \quad & \theta x_o \geq X\lambda, \\ & y_o \leq Y\lambda, \\ & \lambda \geq 0 \end{aligned} \quad (9)$$

$\theta^* < 1$ ならば、DMU は生産可能集合の内点になり、非効率的である。 $\theta^* = 1$ ならば、DMU は生産可能集合の効率的フロンティア上にあることになるが、スラックが存在していたら効率的とはいえない。そのため、(9)の最適解 θ^* を用いて

$$\begin{aligned} \max \quad & w = e's^- + e's^+ \\ \text{subject to} \quad & s^- = \theta^* x_o - X\lambda, \\ & s^+ = Y\lambda - y_o, \\ & \lambda \geq 0, s^- \geq 0, s^+ \geq 0 \end{aligned} \quad (10)$$

によって、スラックが存在しているかどうかを判断することになる。 $\theta^* = 1$ でかつスラックが存在しないとき、CCR 効率的という。

出力指向では、どれだけ出力を増やすことができるかを求めており、

$$(x_o, \varphi y_o) \in P \quad (11)$$

となる最大の φ の値 φ^* を求めることになる。

$$\max\{\varphi | (x_o, \varphi y_o) \in P\} \quad (12)$$

入力指向と出力指向を一つにまとめたものに Russell 尺度がある [4]。

$$(\theta x_o, \varphi y_o) \in P \quad (13)$$

となるようにしたとき、 θ と φ から計算される指標

$$\frac{1}{m+s} \left(\sum_{i=1}^m \theta_i + \sum_{r=1}^s \frac{1}{\varphi_r} \right) \quad (14)$$

を最小にする値を求めている。(13)式の θ と φ はベクトルであり、 θ_i と φ_r は各入出力項目に対する効率性を表すものである。このモデルは求解が容易ではない。

(8)式や(11)式のように radial 尺度では効率的フロンティアとの距離を比率によって測定しようとしている。定員を 20%削減するなどのような合理化案はこの比率尺度によって表現したものであるが、1,500 人の削減というように実数で表現することもある。データには比例スケールと間隔スケールがある。比には意味がなく、差分のみに意味があるのが間隔スケールで、摂氏で表現する温度はこれにあたる。長さや重さなどのデータは負の値は取らず、比を定義できるので比例

スケールである。間隔スケールを使用できる手法は比例スケールにも適用できるが、その逆は成り立たないため、radial 尺度では間隔スケールを適用することはできない。また、radial 尺度では比をとっているため、単位によらないという性質があるが、スラックの存在を確かめなければならない。さらに、スラックは間隔スケールになっているので、2つのスケールが混在していることにもなる。

Non-radial 尺度は、効率的フロンティアとの距離を L_1 距離によって測定し、それに基づいて効率性のスコアを算出したものである。Non-radial 尺度としてスラック基準型尺度 (SBM; slack-based measure) を示す。(10)式のスラックは入力を θ^* 倍に縮小した後に生じるスラックを考えているのに対し、SBM では入出力を縮小・拡大する前のデータと効率的フロンティアとの差分に基づいている。この差分が入出力データのうちのどれだけを占めているかを 0 から 1 までのスケールにしたものが SBM である。

$$\begin{aligned} \min \quad & \rho = \frac{1 - \frac{1}{m} \sum_{i=1}^m \frac{s_i^-}{x_{io}}}{1 + \frac{1}{s} \sum_{r=1}^s \frac{s_r^+}{y_{ro}}} \\ \text{subject to} \quad & s^- = x_o - X\lambda, \\ & s^+ = Y\lambda - y_o, \\ & \lambda \geq 0, s^- \geq 0, s^+ \geq 0 \end{aligned} \quad (15)$$

SBM 尺度は入力余剰と出力不足の両方を同時に考えており、入出力への指向性はない。スラックの大きさから効率性を算出しているため、(10)式のように改めてスラックの存在を調べる必要はない。 $\rho^* = 1$ であれば SBM 効率的といい、CCR 効率的にもなる。一般に、 $\rho^* \leq \theta^*$ となる。

図 3 に 2 つの尺度の図解を示す。 θ^* は OQ/OP で与えられるが、 ρ^* は s_1^-, s_2^- から計算される。

4. 階層型 DEA

効率的な DMU によって表される効率的フロンテ

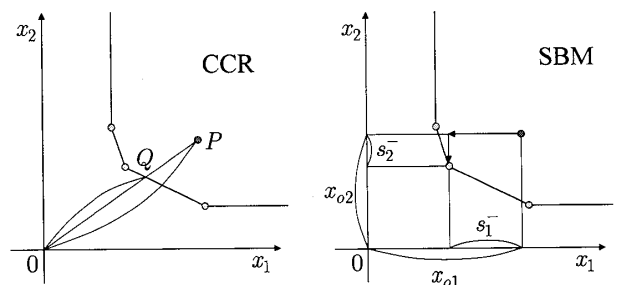


図 3 radial 尺度と non-radial 尺度

ニアからの距離によって効率性が測られる。そのため、非効率的な DMU の存在は、効率的な DMU の効率性評価に影響を与えないし、効率的な DMU はいずれも効率値が 1 という同じ評価をもつ。しかし、相対的な評価は周辺の DMU によっても影響を受けるはずである。同じ効率値をもつ DMU でも、同じような DMU の間で優れているものよりは際立って優れているものの方がより魅力的である。効率的フロンティアを階層的にいくつかのレベルのフロンティアで構成することにより、いろいろなフロンティアに基づいた効率値を見ることができるようになる[2][3]。

まず、すべての DMU を対象にした効率的フロンティアを第 1 レベルとする。次に第 1 レベルにある DMU を除いた DMU を対象にした効率的フロンティアを第 2 レベルとする。これを順次繰り返すことにより、効率的フロンティアの階層構造が得られる。第 l レベルが対象としている DMU の集合を J^l 、 J^l の効率的フロンティアにある DMU の集合を E^l とすると

$$J^{l+1} = J^l - E^l \quad (16)$$

となる。CRS モデルでは、 E^l は線形計画問題

$$\begin{aligned} & \text{minimize} \quad \theta_0 = \theta \\ & \text{subject to} \quad \sum_{j \in J^l} \lambda_j x_{ij} \leq \theta x_{i0}, \quad i=1, \dots, m \\ & \quad \quad \quad \sum_{j \in J^l} \lambda_j y_{rj} \geq y_{r0}, \quad r=1, \dots, s \\ & \quad \quad \quad \lambda_j \geq 0, \quad j \in J^l \end{aligned} \quad (17)$$

の最適値が 1 となる DMU の集合として求められる。VRS モデルや SBM モデルでも同様にして階層構造を構成することができる。

図 4 に CRS モデルと SBM モデルで構成される階層的フロンティアの一例を示す。CRS モデルでは 3 つのレベルのフロンティアが得られている。第 1 レベルにある 4 つの DMU はいずれも効率値は 1 であるが、これらを第 2 レベルのフロンティアに基づいて評価すると、効率値は順に 2.00, 1.58, 1.59, 1.50 となる。文献[3]ではこれを attractiveness score と呼んでいる。CRS モデルでは効率値にスラックが反映されないため、スラックのある DMU 4 や DMU 8 がより高いレベルの階層に入ってしまう。

非効率的な DMU には、(9)式の λ によって、改善目標となる DMU を示すことができる。CRS モデルでは DMU 10 の改善目標は DMU 2 と DMU 3 となるが、下位のレベルにある DMU にとっては、効率的な DMU を直接改善目標とするには差が大きすぎる

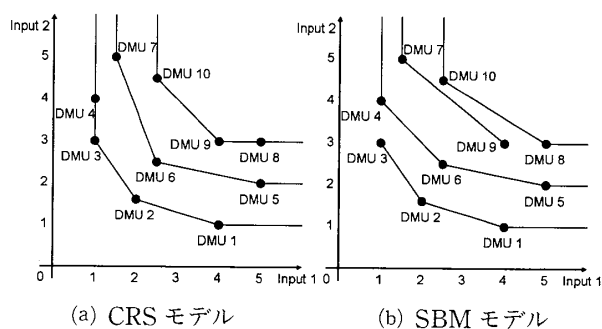


図 4 階層的フロンティア

こともありうる。そのような場合、階層構造を利用すると段階的な改善目標を設定することも可能となる。例えば、DMU 10 はまず、第 2 レベルにある DMU 6 や DMU 7 を目標とすることができる。

5. おわりに

DEA は数理計画の枠組みの中でも議論できるために、ネットワークモデル、確率的モデル、ゲーム理論的解釈などのさまざまな展開がなされている。

本稿では、DEA における効率性評価の基になっている生産可能集合とその効率的フロンティア、および効率性尺度について紹介した。さまざまなモデルが提案されている中で、それらの違いを理解し、適切に適用できる一助になれば幸いである。

参考文献

- [1] W. W. Cooper, L. M. Seiford and K. Tone: Data envelopment analysis, A comprehensive text with models, applications, references and DEA-solver software, Springer 2007.
- [2] H. Morita and J. Zhu: Context-dependent data envelopment analysis by slack-based measure, Omega, Vol. 33, No. 4, pp. 357-362 (2005).
- [3] L. M. Seiford and J. Zhu: Context-dependent data envelopment analysis: measuring attractiveness and progress, Omega, Vol. 31, No. 5, pp. 397-408 (2003).
- [4] T. Sueyoshi and K. Sekitani: Computational strategy for Russell measure in DEA: Second-order cone programming, European Journal of Operational Research, Vol. 180, pp. 459-471 (2007).
- [5] J. Zhu: Quantitative models for performance evaluation and benchmarking: Data envelopment analysis with spreadsheets and DEA Excel solver, Kluwer's International Series, 2002.