1996年度日本オペレーションズ・リサーチ学会
春季研究発表会

繰り返しデータによる確率的DEA

01604524　神戸大学　＊森田 浩　MORITA Hiroshi
神戸大学　岩永 浩史　IWANAGA Hiroshi
01501824　神戸大学　藤井 進　FUJII Susumu

1　確率的DEAと繰り返しデータ

一般的なDEAモデルでは観測して得られた実数値を入出力データとして用いているが、実際には入出力項目の値を正確に誤差のないように観測することが困難な場合や、値が不安定で変わりやすく観測しづらい入出力項目を代表値や予測値などで表現する場合、さらには数量化されていない客観的データとして与えられる場合などがある。このような不確実性のあるデータに対するDEAモデルの一つとして、確率的変動が含まれているデータによる確率的DEAモデルがある [2]。確率的DEAモデルでは確率的変動の大きさを分散成分として定量的にとらえて定式化しているが、この値は既知として取り扱われることが多く、実データへの適用には問題点が残されている。

統計的品質管理手法などでよく用いられる回帰分析では、データを繰り返し測定することで回帰へのあてはまりの悪さを抽出している。これと同様の考え方により、確率的変動のあるデータを複数回繰り返して観測することにより、母平均からばらつきである分散成分を分離させて推定し、これに基づいて効率性を評価することを試みる。データを繰り返し取るという状況は、例えば、一日の交通量を一週間観測したときの平均とか、一ヶ月間の利用者総数などを入出力データとして用いる場合が考えられる。一日の交通量とか一日の利用者数などの個々のデータをそれぞれ入出力データと見れば、繰り返しのあるデータとなり、平均とか総和だけでなく、個々のデータのばらつきを考慮に入れることが可能となる。

推定された分散成分の値から確率的データを乱数によって発生させ、効率値の確率分布を求めた。そして、実際に得られている繰り返しデータから効率値を計算するための効率的フロンティアの決定方法についての考察も行っている。

2　確率モデルとシミュレーション

2.1　分布の推定

N個のDMUを考える。k番目のDMUの入出力データを(X_k, Y_k)とし、r番目の入力X_k^rはn回繰り返し測定されるとする。出力データについても同様であるが簡単のため入力についてのみ示す。データX^rは正規分布に従う確率的変動を含んで観測されると仮定するとき、そのデータ構造は以下の確率モデルで表される。

$$\begin{aligned} f(x|\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(x-\mu)^2/2\sigma^2) \\ & (\mu,\sigma^2)\in\Theta \end{aligned}$$

ただし、パラメータ空間Θは

$$\Theta = \{(\mu_k,\sigma_k^2)|x\in X_k, k=1,2,\ldots,N\}$$

となる。これは自由パラメータ数$2N$の確率モデルである。各DMU毎に分散を推定できるだけの繰り返し数がとれればよいが、一般に多くの繰り返しデータが得られるとは限らないため条件付き確率モデルを考える。

- データの母平均やばらつきに大きな差がない場合には等分散と仮定して、

  $$\Theta_1 = \{(\mu_k,\sigma^2)|x\in X_k, k=1,\ldots,N\}$$

  と定義した自由パラメータ数$N+1$の確率モデル
- それぞれの入出力項目において各DMUの変動係数Cが等しいと仮定して、

  $$\Theta_2 = \{(\mu_k,\sigma_k^2)|\sigma_k = C\mu_k, x\in X_k, k=1,\ldots,N\}$$

  と定義した自由パラメータ数$N+1$の確率モデル。データのばらつきが±5%のように表されるときに適している。

それぞれのモデルに対してパラメータの最尤推定量を求め、入出力データの確率モデルを作成する。

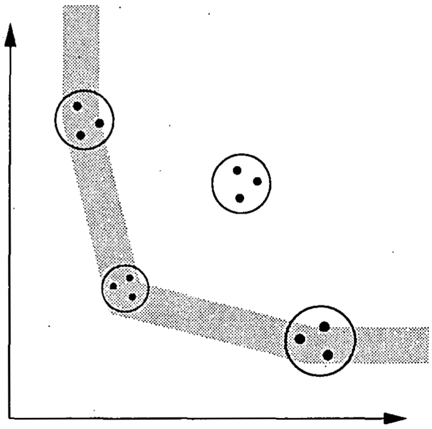

図 1: 繰り返しデータによる効率性分析

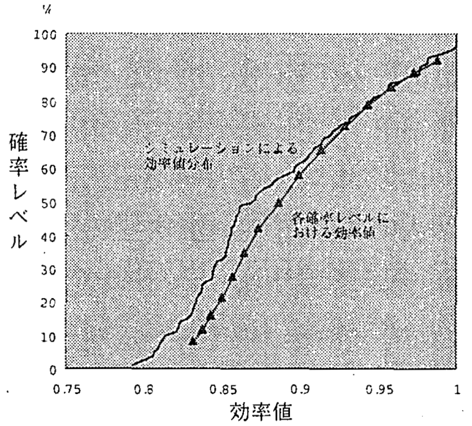

図 2: 効率値分布

2.2 シミュレーションによる効率値

前節で得られた最尤推定量を$\hat{\mu}, \hat{\sigma}^2$とする。$k$番目のDMUの$r$番目の入力$X_k^r$を確率変数$N(\hat{\mu}_k, \hat{\sigma}_k^2)$で置き換えることにより、得られた繰り返しデータはこの確率変数の一つの実現値であったと考える。そこで、この確率分布に従った正規乱数を発生させ、確率的データの組を擬似的に多数発生させ、各々の組み合わせに対して効率値を計算した。このとき得られる効率値の分布が、確率的データによる効率値を表すことになる。(図 2)

3 繰り返しデータによる効率値

繰り返しデータを用いる場合、図1のように複数のデータが1組となって入力あるいは出力データとなる。これらのデータ間に対応があったり、時系列的なあればウィンドウ分析法などが有効であるが、ここでは対応のないデータの組としている。このとき、効率的なDMUの集まりから構成される効率的フロンティアは従来のモデルでは決定できない。すべてのデータを包絡するフロンティアを構成することも可能であるが、これは最も厳しい評価を与えることになる。この他にも各DMUのデータの中で最大値、最小値あるいは平均値をもって代表値としてフロンティアを構成することもできるが、ばらつき成分に関する情報を取り入れているとはいえない。そこでシミュレーションによって求められた効率値の確率分布よって計算される効率値の代表値（平均効率値や上側80%点など）を推定できるフロンティアを構成することを試みる。

確率αで取りうる入力X_k^rの範囲は平均$\hat{\mu}_k$と分散$\hat{\sigma}_k^2$を用いて、$X_k^r \le \hat{\mu}_k + K_\alpha \hat{\sigma}_k^2$ と表すことができる。各αに対して、$X_k^r = \hat{\mu}_k + K_\alpha \hat{\sigma}_k^2$を入力値として効率的フロンティアを求めると、$\alpha$を大きくするほど生産可能集合が小さくなり、効率値は大きくなる。この効率的フロンティアに対して、繰り返しデータの効率値の平均値を求め確率水準αでのDMUの効率値とする。このときの効率値は1以上の効率値を許して安定性を評価している[1]の効率値を用いている。

適当な確率水準αを決定することができれば、確率的変動のある入出力データに対する効率的フロンティアを与えることが可能になる。図2はあるデータセットのDMUに対するシミュレーションによって求めた確率的効率値の累積分布と確率水準に対する繰り返しデータの平均効率値を表したものである。比較的よく推定できている例であるが、さらに、分散成分の推定値と真の値との違いによる検討やスラックのあるDMUに対する検討が必要である。

参考文献

[1] P. Andersen and N. C. Petersen, "A procedure for ranking efficient units in data envelopment analysis", *Management Science*, Vol. 39, pp. 1261–1264 (1993).

[2] 森田浩, "DEA の確率的側面について", 第 6 回 *RAMP* シンポジウム論文集, pp. 91–100 (1994).