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

DEAにおける質的変数の扱い方

01001600 成蹊大学 *上田 徹 UEDA Tohru
西田 英之 NISHIDA Hideyuki

1．まえがき

DEA法の特徴を示すために回帰分析が比較対象として取り上げられることがよくある。その比較例の一つとして、回帰分析では質的変数（カテゴリカル・データ）を用いた林の数量化理論１類がよく知られているように質的変数を扱えるが、既存のDEA法では基本的には質的変数を扱えない。ここではDEAの入力として質的変数を取り込む方法を提案する。

2．多変量解析法における質的変数を用いた個体（DMU）評価

林の数量化理論１類では目的（基準）変数yの説明変数として質的変数（カテゴリカル・データ）を用いており、質的変数$\delta_{ij}(k)$と量的変数$x_h(k)$が混在する場合には

$$y(k)=\sum_{i,j}a_{ij}\delta_{ij}(k)+\sum_h b_h x_h(k)+\varepsilon_k \tag{1}$$

$\delta_{ij}(k)=1$：個体kがアイテムiのカテゴリjに属するとき

0：その他

$x_h(k)$：個体kのh番目の量的変数の値

で表現される。ε_kが正であれば個体kは平均よりも高い水準にあり、負であれば平均よりも低い水準にあるといえる。$y(k)$が出力、$x_h(k)$が入力であれば

$$y(k)=\sum_h b_h x_h(k)+\eta_k \tag{2}$$

におけるη_kが正であれば個体kは入力$x_h(k)$の割には高い水準にあり、負であれば低い水準にあるということになる。これに質的変数$\delta_{ij}(k)$を加えることにより更に平均水準の把握が精密になる。ここで式(1)におけるε_kがあまりに大きい個体は回帰分析では異常値として除去されたり、別の説明変数の導入が図られたりするであろうが、個体評価に用いるとすると、むしろε_kが大きいほど属性の割には高い出力を持っているとしてその個体の能力を高く評価することになるであろう。そのときにはモデルの適合度は問題ではなく、むしろ比較の土俵に上げる属性として何を取り上げるのかが問題となる。規模が大きくなればそれなりにε_kの振れる幅も大きくなると考えれば式(1)のε_kの代りに

$$y(k)/\{\sum_{i,j}a_{ij}\delta_{ij}(k)+\sum_h b_h x_h(k)\} \tag{3}$$

を用いることにすればほぼDEAと同様の定式化に到達する。

出力が１種類だけであればこのように式(1)のε_kあるい式(3)を用いて評価できる。複数出力であれば正準相関分析を用いて同様の議論ができる。すなわち、入力変数$\delta_{ij}(k)$、$x_h(k)$と出力変数$y_m(k)$間の相関が最も高くなるように係数a_{ij},b_k,c_mを求め、

$$\sum_m c_m y_m(k)-\{\sum_{i,j}a_{ij}\delta_{ij}(k)+\sum_h b_h x_h(k)\} \tag{4}$$

あるいは

$$\sum_m c_m y_m(k)/\{\sum_{i,j}a_{ij}\delta_{ij}(k)+\sum_h b_h x_h(k)\} \tag{5}$$

を用いれば、平均よりは高い水準にあるのかどうかの議論が可能になる。

3．DEA法における質的変数を用いた個体評価

上に述べた方法ではDEA法の自分にとって最も有利なウェイト付けをするという特徴を生かすことができない。そこで質的変数のアイテムごとに出力変数との正準相関分析を行い、それによって決定されたa_{ij}を用いてカテゴリ間のウェイト付けには縛りを入れ、

$$\sum_m c_m y_m(k)/\{\sum_i d_i\sum_j a_{ij}\delta_{ij}(k)+\sum_h b_h x_h(k)\} \tag{6}$$

としてc_m,d_i,b_hを通常のDEA法により決定する。また、質的変数が１種類だけなら乗法型の

$$\sum_m c_m y_m(k)/\{\sum_j a_{1j}\delta_{1j}(k)\sum_h b_h x_h(k)\} \tag{7}$$

とすることも考えられる。

しかし、式(6)のCCRモデルでは、原点の移動（平行移動）に対して不変な結果を生まない。また、式(7)の場合には

$$\sum_j a_{1j}\delta_{1j}(k)$$

の最小値を1とするなど比例尺度としての妥当性を確保する必要がある。

そこで式(6)の代りに

$$y(k)/\{\sum_{i,j} a_{ij}\delta_{ij}(k)+\sum_{h} b_h x_h(k)+c\} \tag{8}$$

を用いてc 自身も推定すべき対象とするか、単位調整をした加法モデルを用いる必要がある。加法モデルでは、正準相関分析から得られた正準変量

$$z_i(k)=\sum_{j} a_{ij}\delta_{ij}(k) \tag{9}$$

の標準偏差σ_{zi} により基準化された変数

$$z_i(k)'=z_i(k)/\sigma_{zi} \tag{10}$$

および量的変数$x_h(k)$の標準偏差σ_{xh} により基準化された変数

$$x_h(k)'=x_h(k)/\sigma_{xh}$$

を用いて単位を揃える。

4. 分析例

文献[2]に掲載されている図書館デー多を使わせていただく。通常入力として床面積、蔵書数、職員数を用い、出力として登録者数、貸出冊数を用いる。また、カテゴリカル・データとしてはやはり文献[2]に従い、人口を基準とした3カテゴリを用いる。表1に効率値の計算結果を示す。ただし、CCRはカテゴリ変数を入れずに通常のCCRモデルから得られた効率値であり、モデルA、モデルBの効率値は次のように求められている。

【モデルA】式(8)を用いた場合の効率値であり、c は正負どちらでもよい。原理的には単位の影響を受けないが、ここでは式(9),(10)の基準化された変数を用いている。

【モデルB】式(9),(10)の基準化された変数を用いた加法モデルにおけるT効率値([3],p.685のθ_1^*)である。

この計算ではモデルAとモデルBの効率的DMUは一致した。今後、計算結果の更に詳細な検討を進める予定である。

参考文献

[1]上田:「多変量解析法を用いたDEA法入出力変換法の検討」1994年OR学会秋季大会

[2]刀根:「経営効率性の測定と改善」日科技連, 1993

[3]刀根:「DEAのモデルをめぐって一再論一」、OR学会誌、Vol.40, No.12,1995

表1 効率値

i	CCR	モデルA	モデルB
0	0.226	1.000	1.000
1	0.638	1.000	1.000
2	0.539	1.000	1.000
3	0.593	1.000	1.000
4	0.912	1.000	1.000
5	0.745	1.000	1.000
6	0.650	0.799	0.753
7	0.539	0.692	0.695
8	0.907	1.000	1.000
9	0.705	0.906	0.809
10	0.538	0.648	0.670
11	0.719	0.844	0.839
12	0.657	0.875	0.810
13	0.715	0.844	0.788
14	0.845	1.000	1.000
15	0.582	0.688	0.672
16	1.000	1.000	1.000
17	0.786	0.912	0.705
18	1.000	1.000	1.000
19	0.848	0.902	0.746
20	0.787	0.877	0.753
21	0.681	0.746	0.632
22	1.000	1.000	1.000