

文章编号:1000-6788(2001)03-0031-07

# 相对生产率评价问题中选择DMU的关联系数分析方法

孙静春, 魏军胡, 徐寅峰

(西安交通大学管理学院, 陕西 西安 710049)

**摘要:** 在运用数据包络分析方法评价行业企业相对生产率时,企业的生产能力、技术水平、人员数量以及产出量的差距越大,参评企业从评价结果中得出值得借鉴的信息越少。针对这种情况,本文提出了DEA应用过程中从原始观测数据选择关联决策单元的方法,给出了算法步骤,并且比较了该方法与其它选择方案的差异,通过实例的计算说明了该方法的可行性与有效性。

**关键词:** 数据包络分析; 有效性; 决策单元

**中图分类号:** O 233 **文献标识码:** A

## Choice of DMUs by Relation Coefficients in Application of DEA

SUN Jing -chun, WEI Jun -hu, XU Yin -feng

(The Management School, Xi'an Jiaotong University, Xi'an 710049, China)

**Abstract** This paper puts forward an alternative method of choosing comparative DMUs from initial observed data, the algorithm and its difference from other algorithms are also discussed, finally an example is given out to show that the new method is feasible and efficient.

**Keywords** data envelopment analysis (DEA); efficiency; decision making unit

### 1 引言

企业生产率的评价与分析是企业生产率管理的重要环节,是生产率控制和改进、绩效管理等环节的基础。企业生产率评价的目的之一是与其它竞争企业比较,这实际上可以归结为行业企业的生产率优化问题。这个问题在1957年最先由Farrell讨论。针对用资本、人力两种资源生产单一产品行业厂商的生产率的优化问题,他提出了前沿生产函数法(以后被称为非参数生产前沿分析法),利用最有效率的生产厂商所形成的效率前沿,来确定各个生产厂商的生产率优化值。Farrell的方法在思路比较好,但是只能处理两种投入、一种产出的特殊情况,因此并没有得到广泛采用。另外,他将效率前沿定义为前沿生产函数,没有区分生产前沿与生产函数的差别,也是不妥当的。

1978年Charnes, Cooper及Rhodes提出了评价决策单元相对有效性的数据包络分析(Data Envelopment Analysis, DEA)方法。这是一种在Farrell方法的基础上发展起来的非参数生产前沿分析方法。该方法具有处理多输入、多输出指标的能力。利用该方法评价行业企业相对生产率具有比较突出的优点。

利用数据包络分析(DEA)方法处理行业企业相对生产率评价问题可以分为以下几个步骤:确定决策单元(DMU)以及投入产出指标;确定评价目标与评价约束;建立评价模型;模型计算;数据分析等。

对决策单元的选择是数据包络分析应用的第一个环节,也一直是薄弱环节。特别是在企业的生产能力、技术水平、人员数量以及产出量的差距较大的情况下,参评企业从评价结果中得出值得借鉴的信息比

收稿日期:1999-07-15

资助项目:本研究由中国科学院管理、决策及信息系统开放实验室与西安交通大学科研基金项目提供资助。

较少。

本文提出的依据部分投入、产出指标的决策单元关联分析方法给出了上述问题的一种解决途径。该方法可以通过计算原始数据中各决策单元之间的关联系数,根据对决策单元关联性的不同要求,对原始数据进行分组或者删选。处理以后的数据代表关联系数较大的决策单元,依此进行相对有效性评价,其结果有较高的参考价值,并且减少了相对有效性评价过程中的计算量;该方法还是面向评价问题的。通过对实际问题的分析,选择出对评价问题有重要影响的投入、产出指标,用作选择决策单元的依据,这可以弥补 DEA 方法面向实际观测数据的不足。

### 2 决策单元关联问题的提出

从数据包络分析得到的相对有效性评价可以向各个决策单元提供其投入产出效率的信息。如果各决策单元投入项、产出项的数量差距不大,换言之,决策单元之间较为相似,例如设备生产能力、员工数量、产品产出量等,那么,决策单元之间可以借助于评价结果互相学习管理方式、生产经营活动中的经验和教训,这样的决策单元之间具有较高参考价值;反之,如果各个决策单元的投入项、产出项的数量差距较大,则互相难以提供值得借鉴的经验和教训。因此,在应用数据包络分析方法时,应该注意选择决策单元,减少评价工作量,使评价结果更具有参考价值。因此,选择合适的决策单元具有重要意义。

在数据包络分析中,通常假定有  $n$  各决策单元(Decision Making Unit, DMU)参与评价,各决策单元用相同的投入、产出指标来描述其“生产过程”的投入产出状况,每个决策单元  $i$  的投入是一个  $m$  维的向量  $x(i)$ ,产出是一个  $s$  维的向量  $y(i)$ ,各变量之间的关系可以用图 1 说明:

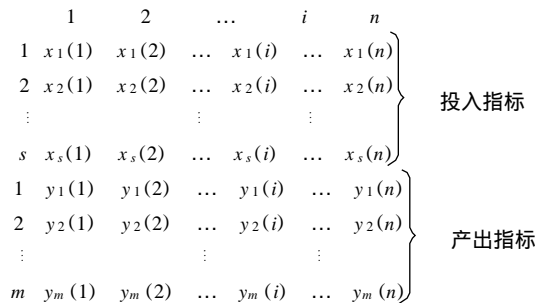


图 1 数据包络分析变量之间的关系

我们进一步假定这  $n$  个参评的决策单元是从  $l$  个原始决策单元(样本点)中选择出来的,因此,必然有  $l$  大于或等于  $n$ 。虽然各个决策单元是用  $s+m$  个投入、产出指标来描述的,对于具体问题而言,管理决策者不一定需要所有的投入、产出指标都很接近,即只要求部分投入产出指标是接近的,假设这样的指标数量为  $h$ ,那末,应有  $h$  小于或者等于  $s+m$ 。综合上述情况,我们可以看出,选择参评决策单元的指标向量应该是  $(x, y)$ ,向量维数是  $h$ 。  $x, y$  分别是投入向量、产出向量中一部分分量组成的新向量。由于在计算关联系数时不需要区分投入指标与产出指标,我们在  $(x, y)$  的基础上定义一个新向量  $p = (x, y)$ ,根据该向量来计算关联系数。

在选择关联系数的计算公式时,我们应该注意两个决策单元之间的关联性应该是不对称的。在研究两个决策单元之间的关联性时,第一个单元我们称之为基准单元,第二个单元我们称之为参考单元。对数据包络分析方法而言,选取参考单元的目的是为了建立生产可能性前沿,即建立相对有效性评价的参考系。因此对基准单元有参考价值的决策单元,不一定需要基准单元作为它的参考点。基于这种考虑,本文没有采用统计学中处理数据相关性的具备对称性的系数计算方法,如相关系数法、模糊聚类分析法等等,而是以不具备对称性的灰关联分析方法为基础来分析决策单元的关联性。

### 3 决策单元的关联系数选择方法

决策单元的关联分析方法步骤如下:



- 1) 确定对决策单元选择有重要影响的投入、产出指标  $(x_i, y_i)$ ;
- 2) 对所有单元构造向量

$$p_i = (p_i(1), p_i(2), \dots, p_i(h)) = (x_i, y_i)$$

并对所有单元的向量分量选择合适的计量单位, 使所有单元的每个指标范围都在 0 与 1 之间.

- 3) 以每个单元为基础计算与其它单元的关联系数  $\xi_i(k)$ ;
- 4) 选择基准单元, 建立关联领域  $B(\epsilon)$ ;
- 5) 取不同的值, 依据关联领域对单元进行分组(分类)或删除;
- 6) 分别选用各组单元作为参评决策单元, 运用数据包络分析方法评估相对有效性.

对上述算法需要进一步说明的是:

在第二步中的归一化不是在表述各个决策单元的向量内进行的, 而是对所有向量的同一分量进行归一化处理.

在第三个步骤中, 关联系数的计算公式如下:

$$\xi_i(k) = \frac{\min_i \min_k |p_0(k) - p_i(k)| + \rho \max_i \max_k |p_0(k) - p_i(k)|}{|\min_i \min_k |p_0(k) - p_i(k)| + \rho \max_i \max_k |p_0(k) - p_i(k)|}$$

表示基准单元  $i_0$  与单元  $i$  在指标  $k$  处的关联系数, 并且

$$\Delta_i(k) = |p_0(k) - p_i(k)|$$

成为第  $k$  个指标处单元  $i_0$  与单元  $i$  的绝对差.

$$\min_i \min_k |p_0(k) - p_i(k)|$$

称为两极最小差, 其中

$$\max_k |p_0(k) - p_i(k)|$$

是第一级最小差, 表示在单元  $i$  中找与基准单元的最小差距指标. 两极最小差是第二级最小差, 表示在最小差距指标的基础上, 找出所有单元中的最小差距.

$$\max_i \max_k |p_0(k) - p_i(k)|$$

是两极最大差, 其意义与两极最小差相似.  $\rho$  是分辨系数, 取值范围在 0 和 1 之间. 综合各个指标的关联系数, 得到基准单元  $i_0$  与单元  $i$  关联度.

$$r_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k)$$

假设所有的单元构成集合  $X$ , 那么, 关联度具有下列 4 个特征:

- 1) 规范性

$$r(p_0(k), p_i(k)) \in (0, 1]$$

$$r(p_0(k), p_i(k)) = 1 \Leftrightarrow p_0(k) = p_i(k)$$

$$r(p_0(k), p_i(k)) = 0 \Leftrightarrow p_0 \in \emptyset, p_i \in \emptyset, \emptyset \text{ 为空集}$$

- 2) 对称性

$$r(p_0(k), p_i(k)) = r(p_i(k), p_0(k)), p_0, p_i \in X \Leftrightarrow X = \{ p_j | j=0, i \}$$

- 3) 整体性

$$r(p_j(k), p_i(k)) \stackrel{\text{always}}{=} r(p_i(k), p_j(k)), p_j \in X \Leftrightarrow X = \{ p_j | j=1, 2, \dots, m \}, m \geq 3$$

- 4) 接近性  $r(p_i(k), p_j(k))$  随  $\Delta$  增大而减小

第四步关联领域是指由基准单元与关联单元构成的集合, 定义如下:

$$B(\epsilon) = \{ p_i | \epsilon \leq r(p_0, p_i) \leq 1, p_i \in X \}$$

邻域内所含单元的多少由  $\epsilon$  的取值决定.  $\epsilon$  值越大, 则关联单元与基准单元越近似, 关联单元越少;  $\epsilon$  值越小, 则关联单元与基准单元的差异程度越大, 关联单元越多. 上式定义的邻域只有一个基准单元. 一般情况下, 我们是选择与多个基准单元相关联的单元作为决策单元. 假定, 由基准单元构成的集合为  $X_0$ . 由被比

较单元构成的集合为  $X_i$ , 则有

$$X_0 \subset X, \quad X_i \subset X, \quad X = X_0 \cup X_i$$

$$B(X_0, \epsilon) = \{ p_i \mid \epsilon \leq r(p_0, p_i) \leq 1, p_0 \in X_0, p_i \in X_i \}$$

选择决策单元的关键在于基准单元的确定。

决策单元的关联系数选择方法与灰关联分析方法有明显的区别, 具体表现在: 其一, 两种方法解决的问题不同, 决策单元的关联分析方法是用来选择决策单元的, 决策单元是以向量形式表示, 向量的各个分量是独立投入、产出指标, 有着独立的含义; 灰关联分析方法是用来解决以函数形式表示的因素之间的关联测度问题, 因素也是以向量形式表示, 但是其意义是逼近函数的有限序列。其二, 数据的归一化处理方法不同, 对决策单元的关联分析方法而言, 由于效率评价量纲可以选用不同的形式, 因此, 决策单元的向量数据通过取不同的量纲, 可以转化为单位区间内的数据; 对于灰关联分析方法而言, 需要将因素向量的数据除以最大分量来使其落入单位区间内。

#### 4 算例

某地区分布着 16 个供热电厂, 为了了解这些电厂燃煤利用的潜力, 拟用数据包络分析方法对这 16 个供热电厂进行评价。评价指标是每千瓦时标准煤耗量。该评价指标主要受制于发电机组的性能, 发电机组越先进, 容量也越大, 每千瓦时标准煤耗率越低; 反之, 发电机组越陈旧, 容量也越小, 每千瓦时标准煤耗率越高。因此对管理者而言, 设备容量指标越是接近, 电厂之间的学习参考价值就越大。除了设备容量指标以外, 表中的投入产出指标也对每千瓦时标准煤耗量有很大影响。值得注意的是, 每千瓦时标准煤耗率的评价与工厂的人员数量关系很小。但是, 在评价火力发电厂的劳动生产率时, 评价指标与人员数量却有着密切的联系。因此, 在应用非参数生产前沿分析方法时, 应该注意选择决策单元, 减少评价工作量, 使评价结果更有价值。

表 1 供热电厂原始观测数据

序号	单位名称	发电消耗 标煤(吨)	供热消耗 标煤(吨)	发电设备 容量(千瓦)	供热设备 容量(千瓦)	发电供热合用电 量(万千瓦时)	供热量 (百万千焦)	发电量 (万千瓦时)
1	B01	415279	59893	150000	50000	10278.00	1424879	107236
2	B02	92306	30076	49000	24000	2730.00	751928	27294
3	B03	320695	535251	150000	250000	16747.00	13144638	95881
4	B04	548603	21329	200000	200000	11956.00	554820	150530
5	B05	138941	83306	50000	50000	5032.00	1951869	36769
6	B06	55320	29417	50000	50000	2729.00	698000	13629
7	B07	26606	197803	24000	24000	3772.00	6145022	11328
8	B08	4973	70808	12000	12000	923.00	1486976	2470
9	B09	12785	48572	9000	9000	864.00	979282	2014
10	B10	148288	173105	66000	66000	4758.00	3781317	32558
11	B11	420169	204941	172000	172000	13155.00	4735318	103352
12	B12	183999	264829	112000	112000	8828.00	6570384	53302
13	B13	4660	34787	9000	9000	477.00	888013	2512
14	B14	19532	209773	139000	56000	5164.00	5712736	3161
15	B15	22708	22140	6000	6000	1430.00	533925	14051
16	B16	204250	77750	29000	24000	7218.00	2014959	55494

表 2 供热电厂归一化数据

序号	发电标煤	供热标煤	发电容量	供热容量	合用电量	发热量	发电量
1	07570	01119	075	02	06137	01084	07124
2	01683	00562	0245	0096	01630	00572	01813
3	05846	1	075	1	1	1	06370
4	1	00398	1	08	07139	00422	1
5	02533	01556	025	02	03005	01485	02443
6	01008	00550	025	02	01630	00531	00905
7	00485	03696	012	0096	02252	04675	00753
8	00091	01323	006	0048	00551	01131	00164
9	00233	00907	0045	0036	00516	00745	00134
10	02703	03234	033	0264	02841	02877	02163
11	07659	03829	086	0688	07855	03602	06866
12	03354	04948	056	0448	05271	04999	03541
13	00085	00650	0045	0036	00285	00676	00167
14	00356	03919	0695	0224	03084	04346	00210
15	00414	00414	003	0024	00854	00406	00933
16	03723	01453	0145	0096	04310	01533	03687

表 3 供热电厂之间的关联系数

序号	1	2	3	4	5	6	7	8
1	1	06357	05911	06957	06920	06524	05350	06131
2	06466	1	04134	05587	08538	09302	07749	08120
3	06072	04213	1	05413	04520	04137	04328	03868
4	07178	05693	05467	1	05482	05705	04454	05100
5	06845	08410	04195	05135	1	08361	07211	07458
6	06626	09304	04065	05606	08490	1	07845	08192
7	05499	07761	04266	04342	07421	07853	1	07981
8	06316	08196	03906	05085	07719	08262	08044	1
9	06200	08341	03859	05248	07483	08418	07907	09650
10	06115	07771	04380	04682	08900	07632	07701	06689
11	07449	04689	06428	06874	05203	04658	05445	04500
12	05738	05457	04970	04581	06291	05321	06201	04783
13	06107	08372	03844	05323	07378	08444	07813	09558
14	06518	06892	05079	04942	07415	07437	08467	07243
15	06044	08609	03878	05502	07465	08773	07990	09104
16	07003	07861	04579	05430	08491	07354	07260	07569

续表 3 供热电厂之间的关联系数

序号	9	10	11	12	13	14	15	16
1	06017	06153	07411	06171	05908	06283	05844	06978
2	08282	07777	04904	06059	08309	06757	08543	07927
3	03833	04732	06617	05682	03807	05028	03866	04757
4	05274	05071	07111	05365	05341	04945	05528	05646
5	07208	08635	05155	06573	07084	07109	07176	08417
6	08366	07666	04882	05957	08387	07316	08714	07439
7	07853	07733	05602	06698	07750	08362	07929	07348
8	09658	06941	04824	05574	09564	07205	09092	07711
9	1	06786	04704	05477	09795	07104	09361	07454
10	06528	1	05882	07248	06432	07796	06604	07519
11	04387	05891	1	06438	04333	06167	04348	05401
12	04697	06898	06049	1	04637	06765	04754	06772
13	09796	06710	04664	05432	1	07020	09358	07358
14	07153	07908	06403	07328	07059	1	06942	06543
15	09385	06847	04659	05516	09380	06899	1	07364
16	07305	07493	05491	07045	07195	06312	07198	1

表 1 至表 3 给出了用关联系数方法删选决策单元(DMU)的计算数据. 如果取  $\xi=0.7$ , 那么决策单元 B03, B12 与其它各个单元的关联系数都小于 0.7. 从表 1 中 B03, B12 单元对应的数据可以看出, B03 的机组容量是最大的, 因此, 没有其它的单元与它处于同一个生产规模上, 互相之间的借鉴作用不大. B12 也是如此. 它们在分析数据包络分析的评价结果时参考价值较小, 可以删去. 如果取  $\xi=0.8$ , 那么决策单元 B11 也不符合要求. 从表 3 中还可以看出关联系数较大的决策单元, 表示单元之间有很大的参考价值. 从上述数据还可以看出, 与 B03 关联系数最大的单元是 B11, 但是与 B11 关联系数最大的单元是 B01, 这点反映了决策单元之间借鉴作用的不对称性. 总的来说, 上述 16 个供热电厂中, 多数生产规模都比较小, 因此机组容量排在前四位的电厂中有三个对其余电厂的参考作用不大, B04 单元是大容量机组, 与 B01 单元关联系数较大, 与其余的小容量机组关联系数较小, 这是由于 B04 单元的供热设备容量远远没有得到充分利用的缘故. 如果在选择删选指标时, 只选设备容量, 那末, 从关联系数反映出来 B04 单元的参考单元就不包括 B01.

表 4 供热电厂观测数据与相对生产率

序号	单位	发电消耗 标煤(吨)	供热消耗 标煤(吨)	发电设备 容量(千瓦)	供热设备 容量(千瓦)	供热量 (百万千焦)	发电量 (万千瓦时)	相对 生产率 <sup>1</sup>	相对 生产率 <sup>2</sup>
1*	B01	415279	59893	150000	50000	1424879	107236	1	*
2	B02	92306	30076	49000	24000	751928	27294	.9052	1
3*	B03	320695	535251	150000	250000	13144638	95881	0.7971	*
4*	B04	548603	21329	200000	200000	554820	150530	1	*
5	B05	138941	83306	50000	50000	1951869	36769	0.7979	0.8556
6	B06	55320	29417	50000	50000	698000	13629	0.7841	0.8588
7	B07	26606	197803	24000	24000	6145022	11328	1	1
8	B08	4973	70808	12000	12000	1486976	2470	1	1

续表4 供热电厂观测数据与相对生产率

序号	单位	发电消耗 标煤(吨)	供热消耗 标煤(吨)	发电设备 容量(千瓦)	供热设备 容量(千瓦)	供热量 (百万千焦)	发电量 (万千瓦时)	相对 生产率 <sub>1</sub>	相对 生产率 <sub>2</sub>
9	B09	12785	48572	9000	9000	979282	2014	0.6491	0.6500
10	B10	148288	173105	66000	66000	3781317	32558	0.7092	0.7377
11	B11	420169	204941	172000	172000	4735318	103352	0.7684	*
12	B12	183999	264829	112000	112000	6570384	53302	0.8097	0.8387
13	B13	4660	34787	9000	9000	888013	2512	0.8482	1
14	B14	19532	209773	139000	56000	5712736	3161	1	1
15	B15	22708	22140	6000	6000	533925	14051	1	1
16	B16	204250	77750	29000	24000	2014959	55494	1	1

表4给出了利用DEA方法计算相对生产率的结果,其中相对生产率<sub>1</sub>表示未删选以前的计算结果,相对生产率<sub>2</sub>表示取 $\xi=0.8$ 删选以后的计算结果.从中可以看出,在没有删选数据以前,相对生产率较低的几个决策单元B02、B03、B05、B06、B09、B10、B11、B12、B13中,B03与B11是大容量机组的电厂,难以从相对生产率较高的小容量机组的电厂B07、B08、B14、B15、B16获取提高燃煤效率的信息.同样,B02、B05、B06、B09、B10、B12、B13是小容量机组的电厂,也难以从相对生产率较高的大容量机组的电厂B01与B04获取提高燃煤效率的信息.相对生产率<sub>2</sub>表示利用关联系数分析方法删选以后相对生产率数据.在这一组数据中,相对生产率之间的差异更小,同时,由于机组容量更为接近,比较容易获取有价值的提高燃煤效率的信息.

## 5 结论

通过前面的分析和实例计算可以看出,依据部分投入、产出指标的决策单元关联分析方法具有如下的作用和优点:

1) 删选决策单元减少了计算量.在数据包络分析应用过程中,增加一个决策单元,至少要多进行一次数学规划模型的计算,对参评的决策单元进行删选可以减少计算量和分析问题的复杂程度.

2) 删选过程保留了原始数据的效率信息.数据包络分析是评价相对有效性(相对效率)的方法,准确的评价是建立在准确的原始观测数据的基础上的.如果删选过程不能保证保留原始数据的效率信息,将影响后面相对有效性的评价结果.在评价过程中的归一化处理由于是在不同的量纲单位下进行的数据变换,数据可以还原,因此,变换后的数据仍然保留着投入产出的效率信息.

3) 删选过程保留了原始数据的生产规模信息.生产规模对效率评价有重要的影响,生产规模的描述可以用投入、产出指标的绝对值,也可以用不同决策单元之间的相对值,在归一化处理时,由于是对所有决策单元生产向量的同一分量进行的,因此,该分量所表示的生产规模的相对信息没有改变.

显然,是否能成功地应用依据部分投入、产出指标的决策单元关联分析方法,这还取决于描述各个决策单元的投入、产出指标体系.只有在正确分析相对有效性评价问题的基础上给出适当的指标体系,才能保证准确地处理原始数据.

## 参考文献:

- [1] 魏权龄. 评价相对有效性的DEA方法[M]. 北京: 中国人民大学出版社,1988.
- [2] 邓聚龙. 多维灰色规划[M]. 武汉: 华中理工大学出版社,1990.
- [3] 邓聚龙. 灰色决策与预测[M]. 武汉: 华中理工大学出版社,1988.
- [4] 杨伦标, 高英仪. 模糊数学原理及应用[M]. 武汉: 华中理工大学出版社,1992.
- [5] Charnes A et al. Measuring the efficiency of Decision Making Units [J]. Euro J of Oper Res,1978, (2),429-444.