文章编号: 1003- 207(2008) 05- 0153- 04

两种改进的超效 DEA 模型

李宇红[1], 赵二帅[2], 关忠诚[2]

(1. 北京财贸职业学院, 北京　101101; 2. 中国科学院科技政策与管理科学研究所, 北京　100080)

摘　要: 对 DEA 有效决策单元不能排序是传统 DEA 模型的一大缺点。该问题也成为 DEA 领域研究的热点。AP 模型是最常用的区分 DEA 有效决策单元效率值大小的超效模型, 但在实际应用中却出现了无解或解不稳定的情况。国外学者新近提出的 MAJ 模型、LJK 模型则有效解决了这两个问题, 但却存在模型区分能力不强的缺点。本文分析了这两种模型缺点产生的原因、予以改进并给出理论证明。标准数据集的实验显示改进效果显著, 尤其是改进后的 LJK 模型表现优异。

关键词: 超效模型; MAJ 模型; LJK 模型; 改进

中图分类号: C931　　文献标识码: A

1　引言

数据包络分析(Data Envelopment Analysis, DEA) 是一种数据导向的评价具有多个输入和多个输出的决策单元(DMUs) 间相对有效性的有效方法。自 Charnes 等人于 1978 年创建以来, DEA 领域的研究吸引了众多的学者, DEA 方法也被成功应用于许多领域, 如银行业[1]、保险业[2]、教育业[3]、证券业[4] 等。

然而, DEA 方法常常会将许多决策单元评为有效(效率值为 1) , 而标准 DEA 模型对于如何区分这些有效决策单元却无能为力。于是近些年来, 如何区分有效决策单元效率值的大小逐渐成为一个研究热点[5]。总的来说, 该领域的研究主要分为六种方法[6]。第一种方法由 Sexton 等[7] 最早开创, 他们通过构建一个交叉效率矩阵(cross- efficiency ratio matrix) 来给有效决策单元排序; 第二种方法是一种基于标杆(benchmark) 的思想。Torgersen 等[8] 认为, 一个有效决策单元, 只有当它被许多有效决策单元选为参考后, 才应该有较高排名; 第三种方法采用了多变量统计工具, 如典型相关分析, 判别分析等同时为有效和非有效决策单元排序。该方法由 Friedman and Sinuany- Stern[9] 开创; 第四种方法聚焦于对无效决策单元的排序, Bardhan 等[10] 称之为“无效优势度量”(measure of inefficiency dominance) 方法; 第五种方法将多准则决策模型(MCDM) 与 DEA 模型相结合来为有效决策单元排序。

最后一种也是应用最为广泛的方法, 是由 Andesen 和 Petersen[11] 提出的一般被称为 AP 模型的一种超效 DEA 模型(super- efficiency model)。超效 DEA 模型实际上是将被评价决策单元排除在该决策单元的评价参考集之外而使其效率值可能大于 1(输入型) 或小于 1(输出型) 的模型[12]。AP 模型由于其操作简单, 效率值含义明确而迅速流行, 不少学者还将其功能拓展到了离群值检测(outlier detection) , 敏感性分析(sensitivity analysis) 和规模分类(scale classification)。然而, AP 方法的弊端也逐渐显现。Thrall[13] 指出如果一些输入数据是 0 或接近于 0, AP 模型可能会导致无解或解不稳定现象。随后, Seiford 和 Zhu[14] 研究了不同超效模型无解的充要条件, 有兴趣的读者可以参考。

在这种背景下, 新的超效模型被陆续提出, 典型的如 MAJ 模型(1999) [15], LJK 模型(2007) [16]。但笔者发现这两种模型都存在求得的有效决策单元的效率值相近, 区分能力不强的缺点。本文通过将两种模型应用于标准数据集实际计算, 对此缺点进行了实验说明。针对该缺点, 在不变规模效益(CRS) 假设下, 本文对 MAJ 模型、LJK 模型分别进行改进。改进后的两种模型其具有可行解的条件与原模型相同, 本文对此给出了理论证明。数据集实验显示模型改进效果显著, 改进模型对有效决策单元的区分能力有明显提高。

收稿日期: 2007- 12- 05; 修订日期: 2008- 09- 19
作者简介: 李宇红(1964-) , 女(汉族) , 北京人. 北京财贸职业学院, 副教授, 研究方向: 管理科学、计算机应用.

2 MAJ 模型和 LJK 模型

2.1 MAJ 模型

MAJ 模型由 Mehrabian 等[15] 于 1998 年提出。具体模型如下:

$$\min \quad \omega_p + 1$$

$$st: \quad \sum_{j=1}^{n} \lambda_j X_j \leqslant X_p + \omega_p I$$

$$\sum_{j=1}^{n} \lambda_j Y_j \geqslant Y_p \quad \lambda_j \geqslant 0, \quad j = 1, \cdots, n$$

其中, I 是一个非零向量。实际使用该模型时, 投入约束常采用:

$\sum_{j=1, j\neq p}^{n} \lambda_j x_{ij} \leqslant x_{ip} + \omega_p x_{ik_i}, \quad i = 1, \cdots, m$, 其中, $x_{ik_i} = \max_{j=1,\cdots,n} x_{ij}$。

MAJ 模型解决了 AP 模型无解及解不稳定的问题。文献[15] 给出了 MAJ 模型有可行解的充要条件:

定理 1: 被评价决策单元 DMU_P 的输出向量 $Y_p \geqslant 0$, 则 MAJ 模型是有可行解的当且仅当对于任意 r, $r = 1, \cdots, s$, 或者 $y_{rp} = 0$, 或者存在一个 DMU_j, $j \neq p$, 满足 $y_{rj} \neq 0$。

值得一提的是, MAJ 模型提出了自己的新的生产可能集的概念(production possibility set, PPS), 所以 MAJ 模型实际上不是一种基于 CCR 模型的超效模型, 这与 AP 模型、LJK 模型不同。

2.2 LJK 模型

LJK 模型由 Li, Jahanshahloo 和 Khodabakhshi[16] 于 2006 年提出。具体模型如下:

$$\min \quad 1 + \frac{1}{m} \sum_{i=1}^{m} \frac{s_{i2}^{+}}{R_i^{-}}$$

$$st: \quad \sum_{j=1, j\neq 0}^{n} \lambda_j x_{ij} \leqslant x_{i0} + s_{i2}^{+}, \quad i = 1, \cdots, m$$

$$\sum_{j=1, j\neq 0}^{n} \lambda_j y_{rj} \geqslant y_{r0}, \quad r = 1, \cdots, s$$

$$\lambda_j, \quad s_{i2}^{+} \geqslant 0$$

其中 R_i^- 是所有决策单元中第 i 项输入的最大值, 即 $R_i^- = \max_{j=1,\cdots,n} x_{ij}$。

Li 等人也给出了 LJK 模型有可行解的充要条件:

定理 2: LJK 模型有可行解的充要条件与 MAJ 模型相同[16]。

LJK 模型同样解决了 AP 模型无解及解不稳定的问题, 同时, Li 等人指出了 MAJ 模型存在的问题: (1). MAJ 模型是基于新的生产集的模型, 可靠性堪疑; (2). 采用 MAJ 模型评价时, 所有 DMUs 的输入必须增加相等的数量, 这一数量的意义却难以解释, 而 LJK 模型不存在这一问题; (3). MAJ 模型最大的问题是区分能力不足, 某些情况下, 该模型不能够区分不同的有效决策单元, 而 LJK 模型可以。然而, 事实上模型区分能力不强同样也是 LJK 模型存在的问题, 下文我们会以实例说明。

3 数据实验及模型改进

以下, 我们将采用两个数据集实验来说明 MAJ 模型、LJK 模型相对于 AP 模型的优势, 并指出两种模型共有的缺点, 给出改进办法。

数据集 1 在文献[15]、[16] 中都有使用。它包含了 5 个决策单元, 每个决策单元都有两个输入、两个输出数据。为了对比说明, DMUA 分别用 $DMUA_1$, $DMUA_2$, $DMUA_3$ 依次与余下 4 个决策单元进行计算。具体数据见表 1。

表 1 对比实验数据集 1

	A_1	A_2	A_3	B	C	D	E
Input 1	2	0	0.1	5	10	10	2
Input 2	8	8	8	5	4	6	12
Output 1	1	1	1	1	2	2	1
Output 2	2	2	2	1	1	1	2

数据集 2 在[16] 中被使用。它包含 19 个决策单元, 每个决策单元都有两个输入、两个输出数据。由于数据集较大, 这里不再给出, 具体数据可参考文献[16]。

采用 CCR、AP、MAJ、LJK 模型分别计算两个数据集, 计算结果列于表 2、表 3 中(其中表 3 只给出了 CCR 有效的决策单元的结果)。

表 2 对比实验数据集 1 结果

	AP	MAJ	LJK	改进 MAJ	改进 LJK
A1	1.47	1.28	1.17	1.47	1.57
A2	Infeasible	1.31	1.27	1.533	1.92
A3	20	1.31	1.26	1.530	1.90
标准差	*	0.02	0.05	0.03	0.16

表 3 对比实验数据集 2 结果

	AP	MAJ	LJK	改进 MAJ	改进 LJK
1	1.15	1.05	1.03	1.119	1.163
2	1.74	1.09	1.07	1.25	1.38
5	1.3	1.10	1.07	1.29	1.41
9	Infeasible	1.04	1.02	1.123	1.12
15	1.33	1.06	1.03	1.16	1.157
19	Infeasible	1.28	1.24	1.73	2.16
标准差	*	0.08	0.07	0.21	0.36

首先看表2、表3 的 AP 列。我们发现两个实验中, AP 模型均有无解的情况发生, 并且表 2 中 A_3

效率值居然达到了 20! 以数据集 1 为例究其原因, 是因为 A_2 的第一项输入数据为 0, 而 A_3 的第一项输入数据接近于 0。这也验证了前文所说的 AP 模型的无解和解不稳定的缺点。再看 MAJ、LJK 列, 两个表中均没出现无解现象, 可是在区分能力上, 两者表现都不尽如意人意。例如数据集 1 的实验中, A_2 与 A_3 的第一项输入有 0. 1 的差别, 其它输入、输出都相同, LJK 模型区分出了 A_2 与 A_3 效率值的差别, 而 MAJ 模型则没能区分出; 而数据集 2 的实验中, LJK 模型却在 DMU 1 和 DMU 15、DMU 2 和 DMU 5 之间两度未能区分出其差别。为了定量比较这两种模型的区分能力, 我们可引入标准差这一指标。因为标准差是一种衡量数据离散程度的统计量, 直观的讲, 模型区分能力越强, 其计算结果的离散程度应越大, 标准差就应越大。从表 2、表 3 的最后一行可看出, 两个模型在两个数据集上的表现可谓平分秋色。所以, Li[16] 所说的“LJK 模型比 MAJ 模型优越”并不能完全站得住脚。

MAJ 模型区分能力不强的原因在于模型中输入约束项的不等号右边采用了 $\omega_p x_{ik_i}$, (x_{ik_i} 是所有决策单元中第 i 项输入的最大值) , x_{ik_i} 值过大, 导致 ω_p 的取值空间过小, 因而使目标函数 1+ ω_p 的取值空间小, 从而导致模型区分能力不强; LJK 模型的原因也类似, R_i^- 值过大, 使目标函数第二项取值空间小, 以致模型区分能力不强。所以, 我们对两种模型的改进如下: 将 x_{ik_i} , R_i^- 都取所有决策单元第 i 项输入的平均值。以下我们将给出改进后的 MAJ、LJK 模型有可行解的条件及其证明。

命题 1: 改进后的 MAJ 模型有可行解的条件与原 MAJ 模型(定理 1) 相同。

证明: 文献[15] 中, I 是一个非零向量, 定理 1 的证明过程中不涉及 I 的具体取值。所以命题 1 自然成立。有兴趣的读者可参看文献[5, 15] 相关证明。

命题 2: 改进后的 LJK 模型有可行解的条件与 MAJ 模型(定理 1) 相同。

证明: 首先看改进后的 LJK 模型的对偶模型 D – LJK(改进) :

$$\max \quad 1+\sum_{r=1}^{s}\mu_r y_{r0}-\sum_{i=1}^{m}v_i x_{i0}$$

$$st: \quad \sum_{r=1}^{s}\mu_r y_{rj}-\sum_{i=1}^{m}v_i x_{ij}\leqslant 0, j\neq 0, j=1,\cdots,n$$

$$v_i\leqslant\frac{1}{mR_i^-}, i=1,\cdots,m (R_i^- \text{ 是第} i \text{ 项输入的均值})$$

$$\mu_r, v_i\geqslant 0, i=1,\cdots,m,\quad r=1,\cdots,s$$

显然, 对于 $\forall r, \forall i, \mu_r=0, v_i=0$ 是上述模型的一个解。为了方便证明, 我们可以将目标函数中的 1 去掉。以下采用反正法证明。如果 D– LJK (改进) 的解是无界的, 则必定 $\exists d=\begin{pmatrix} dv\\ d\mu\end{pmatrix}>0$, 因此有:

$$\begin{cases} d_\mu Y_0-d_v X_0>0,\\ d_\mu Y_j-d_v X_j\leqslant 0, j\neq 0, j=1,\cdots,n,\\ d_v\leqslant 0,\\ d_\mu, d_v\geqslant 0\end{cases}$$

等价于: $\begin{cases} d_\mu Y_0>0,\\ d_\mu Y_j\leqslant 0, j\neq 0, j=1,\cdots,n,\\ d_v=0, d_\mu\geqslant 0\end{cases}$ 定理 1 中有可行解的条件是对于所有 r , $r=1,\cdots,s$, 或者 $y_{r0}=0$, 或者存在一个 $DMU_j(j\neq 0)$, 满足 $y_{rj}>0$, 不妨假设 $y_{r0}=0$, $r=1,\cdots,k$; $y_{r0}>0$, $r=k+1,\cdots,s$ 。因此, 对于每一个 r , $r=k+1,\cdots,s$,

存在一个 $j_r\neq 0$ 满足 $y_{rj_r}>0$, 因此, $d_\mu y_{rj_r}=0$ 可推出 $(d_\mu)_r=0, r=k+1,\cdots,s$, 因此: $d_\mu Y_0=\sum_{r=1}^{s}(d_\mu)_r y_{r0}=\sum_{r=1}^{k}(d_\mu)_r * 0+\sum_{r=k+1}^{s}0 * y_{r0}=0$。这显然与上文 $d_\mu Y_0>0$ 矛盾。

因此, D– LJK(改进) 模型解无界的假设是错误的, 因此在定理 1 条件下 LJK(改进)

模型是有可行解的。□

Li[16] 给出的 LJK 模型中 R_i^- 的原取值的优点在于如此 R_i^- 必将大于 0(若 R_i^- 为 0, 意味着所有决策单元的该项输入均小于 0, 即没有决策单元采用该项输入) 。而将 R_i^- 换为平均值后, 这一优点依然可以保持。从表 2、表 3 的第 5、6 列可以看出, 改进后的 MAJ、LJK 模型的区分能力确实有了较大提升。此外, 从相应的标准差的大小也可看出这种改进的效果, 尤其是改进后的 LJK 模型, 区分能力提升的幅度极为显著。

4　结语

本文首先介绍了用于 DEA 有效决策单元排序的两种超效模型 MAJ 模型和 LJK 模型。用例证的方式证明它们确实解决了 AP 模型的两个缺点: 无解或解不稳定。同时指出了这两个模型共有的问题: 区分能力不强。我们分析了原因并给出了改进。改进的模型保持了原模型的优点, 不会出现无解或解不稳定的情况, 同时本文从理论上证明了改进模

型与原模型拥有可行解的条件相同。最重要的是，改进的模型可以将有效决策单元更加清晰的区分开来。改进模型的优越性可以从上文的两个数据集实验来检验，改进 LJK 模型的表现尤为优异。事实上，笔者还曾将改进前后的模型应用于规模更大的实际数据集，发现改进模型的优越性更为显著。

值得一提的是，虽然改进的 MAJ 模型、LJK 模型有以上所述的优点，但它们的作用仅仅是将有效决策单元按效率值的高低进行了排序，其效率值本身却不具有实际含义，这一点与 AP 模型不同。因此，理论上改进的 MAJ 模型、LJK 模型可应用于所有数据集，但当采用 AP 模型有解时，还是首选 AP 模型。可否研究出一种既不会出现无解情况，效率值又具有实际背景意义的模型，依然需要进一步的探索。

参考文献:

[1] 张渝. 基于 DEA 的典型相关分析在商业银行信用风险评估中的应用[J]. 数学的实践与认识, 2007, 37(24): 33 - 39.

[2] 宋增基, 李春红. 中国保险业 DEA 效率实证分析[J]. 系统工程学报, 2007, 22(1): 93- 97.

[3] 王亚雄, 王红悦, 李洋波. 高校教育资源配置效率的实证分析[J]. 财经理论与实践, 2007, 28(2): 113- 116.

[4] 赵秀娟, 汪寿阳. 中国证券投资基金运行效率的一个实证分析[J]. 系统工程理论与实践, 2007, 27(3): 1- 11.

[5] G. R. Jahanshahloo, L. Pourkarimi, M. Zarepisheh. Modified MAJ model for ranking decision making units in data envelopment analysis[J]. Applied Mathematics and Computation, 2006, 174(2): 1054- 1059.

[6] Zilla S S, Lea F. Review of ranking methods in the data envelopment analysis context[J]. European Journal of Operational Research, 2002, 140: 249- 256.

[7] Sexton T R, Silkman R H, Hogan A. Data envelopment analysis: critique and extensions[C]. Silkman R H. Measuring efficiency: an assessment of data envelopment analysis. San Francisco, CA: Jossey—Bass, 1986: 73 - 105.

[8] A. M. Torgersen, F. R. Forsund, S. A. C. Kittelsen. Slack- adjusted efficiency measures and ranking of efficient units[J]. The Journal of Productivity Analysis, 1996, 7: 379 - 398.

[9] L. Fridman, Z. Sinuany- Stern. Scaling units via the canonical correlation analysis and the data envelopment analysis[J]. European Journal of Operational Research, 1997, 100(3): 629 - 637.

[10] I. Bardhan, , W. F. Bowlin, W. W. Cooper, T. Sueyoshi. Model for efficiency dominance in data envelopment analysis[J]. Journal of the Operations Research Society of Japan 1996, 39: 322 - 332.

[11] P. Andersen, N. C. Petersen. A procedure for ranking efficient units in data envelopment analysis[J]. Management Science, 1993, 39(10): 1261 - 1264.

[12] Yao Chen. Ranking efficient units in DEA[J]. The International Journal of Management Science, 2004, 32: 213- 219.

[13] R. M. Thrall. Duality, classification and slacks in data envelopment analysis[J]. The Annals of Operations Research, 1996, 66: 109 - 138.

[14] Seiford LM, Joe Zhu. Infeasibility of super- efficiency data envelopment analysis models[J]. INFOR, 1999: 37(5): 174- 87.

[15] S. Mehrabian, A. Alirezaee, G. R. Jahanshahloo. A complete efficiency ranking of decision making units in DEA[J]. Computational Optimization and Applications (COAP), 1999, 14: 261- 266.

[16] Shanling Li, G. R. Jahanshahloo, M. Khodabakhshi. A super- efficiency model for ranking efficient units in data envelopment analysis[J]. Applied Mathematics and Computation, 2007, 184(2): 638- 648.

Two Improved Super- Efficient DEA Models

LI Yu hong[1], ZHAO Er shuai[2], GUAN Zhong cheng[2]

(1. Beijing Vocational College Of Finance and Commerce, Beijing 101101, China;
2. Institute of Policy & Management, Chinese Academy of Sciences, Beijing 100080, China)

Abstract: The traditional Data Envelopment Analysis (DEA) models are unable to discriminate between efficient DMUs. This issue has become the interest of many DEA researchers. AP model is a popular super - efficient model for ranking efficient DMUs, but it has infeasibility and instability problems. MAJ model and LJK model recently proposed can overcome these deficiencies but their ability to distinguish efficient DMUs is not strong. This paper analyzes the reason of these two models' shortcomings, then improves them and provides theoretical proofs. Through two standard numerical examples, we find the improved models perform better than old ones, especially the improved LJK model.

Key words: supper- efficient model; MAJ model; LJK model; improvement