

基于主成分分析的交叉环境 DEA 模型的应用

张恪渝¹ 周玲玲² 韩永明^{3,4*}

(1. 北京物资学院 经济学院, 北京 101149; 2. 对外经济贸易大学 国际经济贸易学院, 北京 100029;

3. 北京化工大学 信息科学与技术学院, 北京 100029; 4. 北京化工大学 智能过程系统工程教育部工程研究中心, 北京 100029)

摘 要: 定义一般环境交叉互评指数, 并针对现有交叉数据包络分析(DEA)模型计算指数时未能体现整体数据信息的缺点, 结合主成分分析(principal component analysis, PCA)提出一种改进型交叉互评指数方法, 使得每一个决策单元的评价向量依据其对评价矩阵中变异量的贡献度而赋予不同的权重; 并筛选一些影响较小的信息, 从而达到降维的目的, 改善了交叉环境模型的评估效率。最后基于真实的能源投入与产出数据, 对比传统交叉评价方法及改进型方法在计算能源效率结果上的异同。

关键词: 交叉环境 DEA 模型; 主成分分析; 降维; 能源效率

中图分类号: F830.0 **DOI:** 10.13543/j.bhxbzr.2018.06.015

引 言

Charnes 等^[1]于 1978 年首次提出数据包络分析(data envelopment analysis, DEA)方法, 该方法优势在于无需确定的函数形式即可对生产效率进行分解。目前该模型被广泛用于评价经济增长效率、企业绩效、资源环境配置等诸多方面。然而传统 DEA 方法存在两个重要的约束限制, 即强可处置性假设与决策单元(decision making unit, DMU)相互独立假设, 前者主要指投入与产出成正向关联关系, 后者主要指决策单元的投入或产出之间相互独立。

在基于传统 DEA 模型的强可处置性假设条件下, 分析资源环境效率问题时往往忽略非期望产出(负效应)的影响, 使得效率评价略显偏颇。因此, 如何对传统模型进行有效改进成为学术界关注的焦点。目前, 针对非期望产出的处理主要有以下两种思路: ①对非期望产出的原始数据进行有效处理, 即将诸如污染物等非期望产出视为投入, 然后运用传统 DEA 方法进行效率评价; ②对非期望产出进行弱处置性假设, 构建环境 DEA 模型或环境 DEA 技

术^[2]。Chambers 等^[3]首次在评价行业环境效率中引入非期望产出(如污染物、CO₂排放等), 在对其进行弱可处置性假设后, 基于径向 DEA 模型对行业环境效率进行测评。刘冰熙等^[4]借助修正的三阶段 Bootstrapped-DEA 方法评估地方政府环境治理效率, 结果发现我国地方政府环境治理存在比较严重的效率损失, 治理效率值呈波浪态势且日趋恶化。郭四代等^[5]运用三阶段 DEA 模型对中国区域层面的环境效率、变化趋势及其差异性等方面进行了全面评估。

传统 DEA 模型第二个重要假设表现在决策单元往往以自身效率最大化为目标约束, 并不考虑其他决策单元的信息, 因此造成效率评价过于片面。Sexton 等^[6]基于此问题, 首次提出了交叉效率模型, 其基本思想在于允许所有决策单元之间相互评价, 每个决策单元选择一组最优的权重依次评价其他的决策单元, 然后将每个决策单元的相互评价效率值集结成一个最终的评价结果。Doyle 和 Green^[7]提出了仁慈型和进取型的二次目标模型对决策单元的权重进行约束, 进而计算效率分数。刘学之等^[8]基于我国钢铁企业的投入产出数据, 运用 DEA 交叉效率模型评价钢铁产业的能源交叉效率。

尽管交叉 DEA 模型能够考量决策单元之间的关联, 但在计算交叉评价指数时会较为主观地给予每一个评价向量相同的权重($1/n$)。由于不同 DMU 间的相关性不尽相同, 且它们对整个评价系统的贡献度有所差异, 因此有些无效的评价信息会显著影

收稿日期: 2017-03-30

基金项目: 国家社会科学基金(14BTJ026); 国家自然科学基金青年基金(61603025)

第一作者: 男, 1984 年生, 博士, 讲师

* 通信联系人

E-mail: hanyun@mail.buct.edu.cn

响 DEA 模型的评估结果。目前,针对等权重赋值问题的处理方法,主要集中于对效率值进行模糊化处理,以此消除构建交叉互评矩阵的过度主观性问题。高明美等^[9-10]构造了改进型直觉模糊熵的计算公式,并利用改进的方法确定区间属性权重以判别区间多属性决策。范建平等^[11]定义了一种新的方法将三角模糊效率值转化为直觉模糊集,并用直觉模糊熵对模糊交叉效率进行计算,然后使用三元有向距离函数对全局模糊效率值排序。综上所述,已有文献虽然也对交叉互评指数的计算进行修正,但处理方法大都是引入新的参数并进行人为赋值,使得结果同样存在着一定的主观性。而交叉评价指数不准确的原因在于没有一个机制来区分信息有效程度。Adler 和 Golany^[12]使用主成分分析法(principal component analysis, PCA)从投入产出数据上对变量进行降维,以剔除一些无效变量,使得 DEA 评价更为有效。基于这个思想,本文将主成分分析方法应用到交叉互评矩阵中,筛除交叉评价系统中冗余的信息;在尽量保持矩阵完整信息的基础上,用少量不相关的主成分来代替众多原始评价向量,并根据对信息矩阵的贡献程度为评价单元赋值,从而改善交叉环境模型的评估效率。

1 基于主成分分析法的 DEA 交叉模型的构建

1.1 环境 DEA 交叉模型

如引言部分所述,传统的 DEA 模型没有考虑非期望产出物,与实际的生产情况相悖,如煤炭发电必然会产生二氧化碳,造纸行业必然会产生污水。在生产活动中,人们希望非期望产出越少越好,但这就与一般 DEA 模型假设产出为强可处置性相悖。因此,如果在产出中不将非期望产出与期望产出加以区分,则会使整个效率评价体系扭曲、无效。

假设在某个环境测评体系中有 n 个对象,则每一个对象 j 就是一个决策单元(z_j)。其中有 m 种投入要素指标, d_1 种期望产出指标以及 d_2 种非期望产出指标,则决策单元 j 的投入指标值 x_j 、期望产出指标值 y_j 和非期望产出指标值 u_j 分别为

$$\begin{aligned} x_j &= (x_{1j}, x_{2j}, \dots, x_{mj}), x \in R_+^{m \times n} \\ y_j &= (y_1, y_2, \dots, y_{d1}), y \in R_+^{d1 \times n} \\ u_j &= (u_1, u_2, \dots, u_{d2}), u \in R_+^{d2 \times n} \end{aligned}$$

依据一般环境 DEA 模型的设定^[13],假设期望产出 Y 及非期望产出 U 同时具有弱可处置性的特

征,即若 $(Y, U) \in P$, 则 $(\alpha Y, \alpha Z) \in P, \forall \alpha \in (0, 1]$, 其中, P 表示生产可能性集合(生产技术)。也就是说,要减少非期望产出的唯一方法就是同比例减少期望产出的数量。如果假设生产技术呈现规模报酬不变的特性(constant return to scale, CRS),则环境 Charnes, Cooper 和 Rhodes(CCR)模型的技术集合 P 可以表示为

$$P = \left\{ (X, Y, U) \mid \sum_{j=1}^m \lambda_j x_{mj} \leq x_m, \sum_{j=1}^{d_1} \lambda_j y_{jd_1} \geq y_m, \sum_{j=1}^{d_2} \lambda_j u_{jd_2} = u_m \right\} \quad (1)$$

即在生产技术 P 的条件下,投入 m 种要素 X , 可以产出 d_1 种期望产出 Y 以及 d_2 种非期望产出 U 。其中, λ_i 为决策单元向量的权重。如果生产技术呈现规模报酬可变(variable return to scale, VRS)的特性,则需要额外加入 $\sum_i \lambda_i = 1$ 的条件,环境 CCR 模型变成环境 Banker, Charnes 和 Cooper(BCC)模型。根据 Färe 等^[14]提出的处理方法,式(1)中生产技术可以转化为环境 CCR 模型

$$\begin{aligned} \min \quad & \theta \\ \text{s. t.} \quad & X\lambda + S^+ = x_{m0} \\ & Y\lambda - S^- = y_{m0} \\ & U\lambda = \theta u_{k0} \\ & \lambda_k, S^+, S^- \geq 0 \end{aligned} \quad (2)$$

式中, X, Y, U 分别为投入要素、期望产出以及非期望产出矩阵; λ 在这里作为要素投入的权重; S^-, S^+ 分别为产出松弛变量与投入松弛变量; θ 表示评价单元的环境效率,即环境评价指数(environmental performance index, EPI)。 θ 数值越大,说明该评价单元在控制非期望产出量方面没有更多的提升空间,则认定该单元的环境表现越好;反之则相反。可以看出,环境 CCR 模型的侧重点是在控制非期望产出 U 上。

接下来使用 Charnes Cooper 变换,将式(2)转换为其对偶问题

$$\begin{aligned} \text{Max } \rho &= \sum_{k=1}^{d_1} \alpha_k Y_k - \sum_{m=1}^M \beta_m X_m \\ \text{s. t.} \quad & \sum_{k=1}^{d_1} \alpha_{ki} Y_{ki} - \sum_{m=1}^M \beta_{mi} X_{mi} - \sum_{s=1}^{d_2} \gamma_{si} U_{si} \leq 0 \\ & \sum_{s=1}^{d_2} \gamma_{si} U_{si} = 1, i = 1, 2, \dots, N \\ & \alpha_k, \beta_m \geq 0; \gamma_s \text{ 为自由变量} \end{aligned} \quad (3)$$

式中,目标方程 ρ 的计算结果与式(2)中的 θ 相同,

都代表了决策单元的环境效率; α 、 β 和 γ 分别为投入要素、期望产出及非期望产出向量 X 、 Y 和 U 的权重。定义 ρ 达到最大值时投入产出向量的最优权重组合为 $L^* = (\alpha^*, \beta^*, \gamma^*)$, L^* 为 $m + d_1 + d_2$ 维的向量。

基于 Sexton 等^[6]的概念,定义交叉环境效率指数为

$$\rho_{ij}^{\text{cross}} = \frac{Y_{d_1}^T \alpha_{d_2}^*}{X_m^T \beta_m^* - U_{d_2}^T \gamma_{d_2}^*} \quad (4)$$

由式(4)可以看出,当 $j = i$ 时, $\rho_{ij} = \rho$; 交叉环境评价指数退化为一般环境效率指数,即部门 j 通过自己的最优权重 L_j^* 计算出 EPI 指数,记为环境自评价指数;当 $j \neq i$ 时,部门 j 的效率值是通过 i 部门的最优权重 L_i^* 计算,所计算出的 EPI 称为环境交叉评价指数, $i, j = 1, 2, 3, \dots, n$ 。

最后,由交叉指数可以构成一个环境交叉评价矩阵 C_{EM} (cross efficient matrix, CEM), 表示为

$$C_{EM} = \begin{pmatrix} \rho_{11} & \cdots & \rho_{1n} \\ \vdots & \ddots & \vdots \\ \rho_{n1} & \cdots & \rho_{nn} \end{pmatrix}$$

其中,对角线上的元素即为一般环境效率指数,而非对角线上元素为交叉环境评价指数,如 ρ_{ij} 表示从第 j 个决策单元的角度对第一个决策单元的评价。按照行方向计算每一个评价单元的算术平均值 $\bar{\rho}_i$, 则 $\bar{\rho}_i$ 就是一般概念下的环境交叉互评指数

$$\bar{\rho}_i = \frac{\sum_{j=1}^n \rho_{ij}}{n} \quad (5)$$

1.2 基于主成分分析的交叉评价方法

主成分分析方法是由 Hotelling^[15] 于 1933 年提出一种压缩高维矩阵,并从中提取数据分散特征 (variation) 的统计分析技术。该方法的优势在于将原始的存在着相关性的 n 维解释变量转化为 p 维直交的综合解释变量,且 $n \gg p$, 从而达到降维的目的。基于主成分分析法的特点,对交叉评价矩阵进行降维,用少量不相关的主成分来代替众多原始评价向量,从而改善交叉环境模型的评估效率。

在通常情况下,由于各解释变量的量纲不同,会造成主成分分析的偏差,因此需要对数据提前作预处理。然而,交叉环境 DEA 模型计算的结果,即交叉互评矩阵中的元素代表效率值,都是无量纲且分布于 $(0, 1]$ 之间,因此不需要额外进行归一化处理。

基于主成分分析的交叉评价指数构建具体步骤

如下。

1) 设环境 C_{EM} 矩阵中的互评向量为 $\rho = (\rho_1, \rho_2, \dots, \rho_n)$, 计算其相关系数矩阵 R 。

2) 将矩阵 R 进行分解,计算其特征值 $w_1 \geq w_2 \geq \dots \geq w_n \geq 0$ 以及相对应标准化的特征向量 e_1, e_2, \dots, e_n 。计算公式为

$$(R - w_j E) \times e_j = 0 \quad (6)$$

3) 使用评价指标 ρ 与特征向量 e , 计算主成分指标数据矩阵 ρ^{PC} 。则第 i 个主成分可以表示为

$$\rho_i^{PC} = e^T \rho = e_{1i} \rho_1 + e_{2i} \rho_2 + \dots + e_{pi} \rho_p \quad (7)$$

4) 根据各主成分的贡献率,选取前 p 个作为参考指标。假设它们已经基本反映了原评价体系的主要信息,之后计算环境综合评价指标

$$\rho^{PC} = \sum_{i=1}^p \rho_i^{PC} \quad (8)$$

可以看出,本文计算出的主成分由特征向量和原始的互评向量作线性组合而成,它包含了原来 C_{EM} 矩阵中的所有信息。此外,各主成分向量间是正交的关系,即 $(\rho_j^{PC})^T \times \rho_i^{PC} = 0, j \neq i$ 。各主成分的方差为 $\text{Var}(\rho_i^{PC}) = e_i^T R e_i, i = 1, 2, \dots, n$ 。按照方差的大小,依次称其为第一主成分、第二主成分、..., 这些主成分按照其对整个矩阵方差的贡献程度逐级递减。因此,在实际应用时,往往只使用前几个贡献度大的成分。如果要避免信息损失,可以按照累积贡献率 100% 的原则选取主成分变量。

2 算例分析

以往用来衡量产业 (部门) 环境绩效表现的指标有 3 个,即能源强度 (能源消耗量与 GDP 的比值)、排放强度 (CO_2 排放量与 GDP 的比值)、以及碳因子 (CO_2 排放量与能源消耗量的比值)。但是这些指标只能对投入、产出数据进行二维的考量,无法综合考虑所有因素。此外,化石能源是产生二氧化碳气体的主要原料,而中国是化石燃料消耗的大国。因此,本文采用北京市统计局发布的“2012 年北京市投入产出表”以及《中国能源统计年鉴 2013》中 42 个部门化石燃料的消耗和碳排放量,通过环境交叉 DEA 模型对行业的环境表现指数 (EPI) 进行度量,并将一般模型结果与改进模型结果进行对比。

2.1 数据来源及预处理

如表 1 所示,行业总量为 42,即决策单元数量 n 为 42。其中,包含第一产业部门 1 个,第二及第三产业部门分别为 27 个、14 个。

表 1 部门名称及代码
Table 1 Department name and code

代码	部门名称	代码	部门名称	代码	部门名称
01	农林牧渔产品和服务	15	金属制品	29	批发和零售
02	煤炭采选产品	16	通用设备	30	交通运输、仓储邮政
03	石油和天然气开采产品	17	专用设备	31	住宿和餐饮
04	金属矿采选产品	18	交通运输设备	32	信息传输、软件和信息技术服务
05	非金属矿和其他矿采选产品	19	电气机械和器材	33	金融
06	食品和烟草	20	通信设备、计算机和其他电子设备	34	房地产
07	纺织品	21	仪器仪表	35	租赁和商务服务
08	纺织服装鞋帽皮革羽绒及其制品	22	其他制造产品	36	科学研究和技术服务
09	木材加工品和家具	23	废品废料	37	水利、环境和公共设施管理
10	造纸印刷和文教体育用品	24	金属制品、机械和设备修理服务	38	居民服务、修理和其他服务
11	石油、炼焦产品和核燃料加工品	25	电力、热力的生产和供应	39	教育
12	化学产品	26	燃气生产和供应	40	卫生和社会工作
13	非金属矿物制品	27	水的生产和供应	41	文化、体育和娱乐
14	金属冶炼和压延加工品	28	建筑	42	公共管理、社会保障和社会组织

本文针对行业的环境表现效率进行评估,主要讨论的是在生产过程中对于非期望产出的控制。因此如表 2 所示,投入指标 X 选取为各部门生产所消耗的一次(化石)能源数量,包含煤炭、焦炭、汽油、煤油等 15 种主要能源。各部门的能源消耗量见图 1。

期望产出指标 Y 使用的是每一个部门的行业总产值,即投入 15 种一次能源后所产生的正向经济效益;而非期望产出指标 U 选取为每一个部门伴随着生产所排放出二氧化碳气体的数量。本文采用如下公式对各部门的 CO_2 排放总量进行估算

$$U_j = \sum_i X_{ij} \delta_i$$

(9)

式中, U_j 为 j 部门的排放总量; X_{ij} 为 j 部门的第 i 种能源的投入量(需要折算为标准煤); δ_i 为能源 i 的排放系数(以燃烧热值法为主,详见表 3)。

表 2 输入、输出数据
Table 2 Input and output data

变量名称	环境绩效指标	维度	单位
投入向量(X)	煤炭、石油、天然气等	15	万吨、亿
	15 种主要能源的消耗量		立方米等
期望产出向量(Y)	行业产出量	1	万元
非期望产出向量(U)	二氧化碳排放量	1	万吨

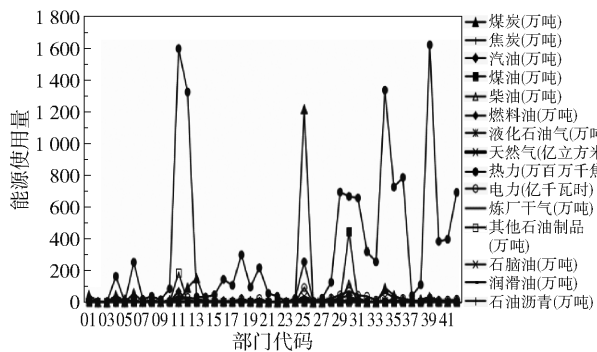


图 1 能源使用量
Fig. 1 Main energy uses

表 3 能源排放系数

Table 3 Energy emission coefficients

能源	能源排放系数	能源	能源排放系数
原煤	2.07	煤油	3.08
洗精煤	2.49	柴油	3.16
煤泥	0.89	炼干气	2.65
洗中煤	2.02	液化石油气	3.17
煤合计	2.23	其他焦化产品	3.04
焦炭	3.04	其他石油制品	2.95
原油	3.07	天然气	21.84
燃料油	3.24	焦炉煤气	7.71
汽油	3	其他煤气	5.92

数据来源:IPCC(2006);能源消耗系数定义为若某种能源消耗系数为 1,则表明消耗 1 万吨该能源所排放的 CO_2 气体数量为 1 万吨。

由于个别产业的投入数据不满足 DEA 模型的基本要求,因此需要进行预处理。

令 x_{ij} 表示第 j 个行业消耗第 i 种能源的数量,则转换后数据变为

$$x_{ij} = \frac{x_{ij} - \min(x_{1j}, x_{2j}, \dots, x_{nj})}{\max(x_{1j}, x_{2j}, \dots, x_{nj}) - \min(x_{1j}, x_{2j}, \dots, x_{nj})}$$

(10)

2.2 部门效率值评估

首先,使用一般环境 CCR 模型(假设行业规模收益不变),对“2012 年北京市投入产出表”中 42 个部门的环境表现指数进行自评;再利用一般交叉环境模型对 42 个部门进行交叉互评。两种模型的效率计算结果见图 2。

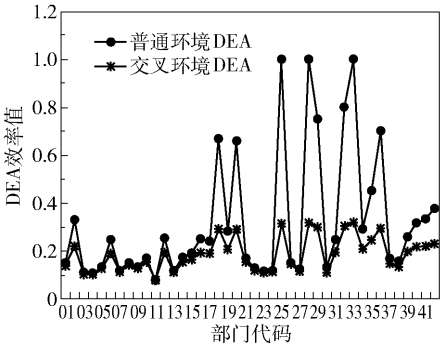


图 2 自评结果和互评结果对比

Fig. 2 Comparison of the results of self-evaluation and mutual evaluation

由图 2 可知,有 3 个部门的环境表现达到了 DEA 有效(效率值为 1),分别为电力、热力的生产和供应,建筑业以及金融业(部门代码为分别 25、28 和 33)。在引入互评机制后,点状线为交叉环境效率指数的结果。可以看出,行业效率值在经过其他决策单元的互评后,出现了显著的下降。排名在前三的依旧为电力、热力的生产和供应,建筑业以及金融业,但是已经可以明显区分它们之间的排名,且与达到 DEA 有效相距甚远。

随后,基于交叉评价矩阵(表 4)的计算结果,再通过式(6)、(7)对互评系统的主成分进行计算;其中,CEM 矩阵对角线上元素的数值为一般环境 DEA 模型所计算出的行业自评分数。由表 5 的结果可知,前两个主成分 ρ_1^{PC} 和 ρ_2^{PC} 对方差的贡献率已经达到 93.7%,后面成分的信息对互评矩阵的解释作用微乎其微。因此,本文保留原互评向量组 ρ 的前两个主成分。

表 4 42 部门交叉评价结果

Table 4 42 sector cross evaluation matrix

部门代码	数值				
	01	02	03	...	42
01	0.15	0.15	0.11	...	0.15
02	0.15	0.33	0.99	...	0.33
03	0.11	0.11	0.11	...	0.11
⋮	⋮	⋮	⋮	⋮	⋮
42	0.15	0.33	0.11	...	0.37

表 5 主成分结果

Table 5 Results of PCA

项目	数值		
	第一主成分	第二主成分	第三主成分
标准差	4.74	4.11	1.03
贡献率/%	53.4	40.3	2.5
累积贡献率/%	53.4	93.7	96.2

继续根据式(8)计算环境综合评价指标,然后将 3 个模型的计算结果进行对比。从图 3 可知,从效率值本身分析,基于主成分分析方法改进的环境绩效指数与一般交叉 DEA 模型结果相差较大。其中,有 21 个行业的 EPI 指数都在 0.5 以上,达到 DEA 有效的部门数量有 6 个,其行业代码分别为 25、28、29、32、33、36;而传统交叉模型计算出的结果中,只有 5 个行业 EPI 指数在 0.3 以上,数值上明显低估了行业效率。由于效率值只是反映行业间相对的变化,更为重要的是行业间的绩效排序情况。表 6 列出了环境 CCR 模型、交叉环境模型以及改进型交叉环境模型的计算结果对比。

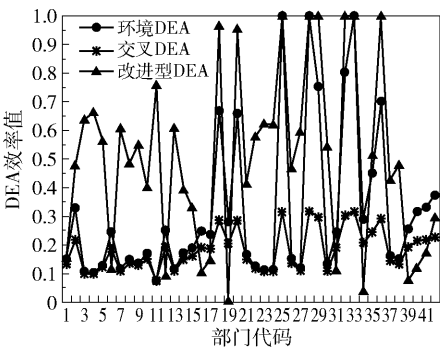


图 3 3 种模型环境表现指数

Fig. 3 Environmental performance improvement index

从表 6 取排名最前和最后的 6 个部门进行分析。首先观察一般环境模型(模型 a)与改进型交叉模型(模型 c)的结果,发现两者涵盖部门差异较小,排名前六位的全部涵盖。而排名后六位部门中前者包含废品废料业、金属制品修理服务业以及石油开采业(代码分别为 3、23、24);后者包含了交通设备业、交通运输与仓储服务业以及居民服务业(代码分别是 18、30、38)。其次,再对比交叉模型(模型 b)与改进型模型(模型 c),排名前六位的部门各有一个不同:交通运输业(代码 18,交叉模型)与金融业(代码 33,改进型模型);而排名后六位的则相差了 2 个部门:交叉模型中含有废品废料业(代码 23)与石油和天然气开采业(代码 3),而改进型模型中

包含了运输业及居民服务业(代码分别为 18、38)。为进一步验证 3 种方法测度下部门环境效率排序是否存在显著差异,运用非参数的 Wilcoxon rank-sum 检验,该检验的原假设为待检验的两组行业效率排名(rank)没有显著性差异。鉴于使用的 3 个序列的样本量都较大($n = 42$),因此本文假定 Wilcoxon 统计量服从标准正态分布。结果如表 7 所示。

表 6 3 种模型的排名对比

Table 6 Comparison of the three models						
排名	部门 代码 ^{a)}	EPI ^{a)}	部门 代码 ^{b)}	EPI ^{b)}	部门 代码 ^{c)}	EPI ^{c)}
1	25	1	25	0.32	25	1
2	28	1	28	0.31	28	1
3	33	1	32	0.30	29	1
4	32	0.80	29	0.30	32	1
5	29	0.75	36	0.29	33	1
6	36	0.70	18	0.28	36	0.97
⋮	⋮	⋮	⋮	⋮	⋮	⋮
37	13	0.12	13	0.11	11	0.12
38	24	0.11	30	0.11	5	0.11
39	23	0.11	23	0.10	30	0.094
40	3	0.11	3	0.10	38	0.078
41	5	0.10	5	0.09	13	0.039
42	11	0.07	11	0.07	18	0.009

a—环境 DEA 模型;b—一般交叉 DEA 模型;c—基于 PCA 的交叉环境模型。

表 7 Wilcoxon 检验结果

Table 7 Wilcoxon test results	
统计量	Z
模型 I 与模型 III 排名比较	-1.98 ** ($P = 0.04$)
模型 II 与模型 III 排名比较	-2.95 ** ($P = 0.00$)

**表示统计量在 $\alpha = 1\%$ 的水平下显著。

以基于 PCA 的交叉环境模型的计算结果作为基准组。如果选取 5% 的显著性水平下进行分析,环境 CCR 模型与一般交叉模型的计算结果都与基准组的行业效率结果有显著的不同(P 值分别为 0.04 与 0.00),且交叉模型的秩和统计量比环境模型的值要更大。

综上所述,基于 PCA 方法得到的改进型交叉评价指数,其计算结果无论与传统环境模型还是一般交叉模型对比,差异性都非常显著。究其原因,可能在于传统方法不是没有充分利用信息,就是对信息使用过于主观,造成了评价单元效率的扭曲。

3 结束语

传统交叉 DEA 模型通过交叉评价矩阵对决策单元效率进行评估时,往往是使用简单的算术平均方法。这种等权重的做法比较主观且没有充分利用整个评价系统的信息,造成 DEA 分数的无效性。本文提出一种改进型的互评指数计算方法,利用主成分分析方法筛除 CEM 矩阵中的冗余信息;在尽量保证互评矩阵信息完整性的基础上,用少量主成分分数来计算决策单元的效率值,从而改善交叉 DEA 模型的评估效率。

真实算例的结果表明,改进型方法计算出的行业效率值与传统交叉评价方法有着显著不同。由于没有充分利用 CEM 中的信息,环境 CCR 模型与一般交叉模型的计算结果出现了较大偏差,明显低估了行业的环境绩效指数。当然,本文提出的方法同样可以应用于非径向(non-radial)的 DEA 模型假设,还有待后续研究。

参考文献:

[1] CHARNES A, COOPER W W, RHODES E. Measuring the efficiency of decision making units [J]. European Journal of Operational Research, 1978, 2(6): 429-444.

[2] ZHOU P, ANG B W, POH K L. Measuring environmental performance under different environmental DEA technologies [J]. Energy Economics, 2008, 30(9): 1-14.

[3] CHAMBERS R G, FÄRE R, GROSSKOPF S. Productivity growth in APEC countries [J]. Pacific Economic Reviews, 1996, 1(3): 181-190.

[4] 刘冰熙, 王宝顺, 薛钢. 我国地方政府环境污染治理效率评价—基于三阶段 Bootstrapped DEA 方法[J]. 中南财经政法大学学报, 2016(1): 89-95.

LIU B X, WANG B S, XUE G. On assessment of public spending efficiency of environment protection in local China: based on three-stage Bootstrapped DEA[J]. Journal of Zhongnan University of Economics and Law, 2016(1): 89-95. (in Chinese)

[5] 郭四代, 全梦, 郭杰, 等. 基于三阶段 DEA 模型的真实环境效率测度与影响因素分析[J]. 中国人口·资源与环境, 2018, 28(3): 106-116.

GUO S D, TONG M, GUO J, et al. Measurement and influencing factors of inter-provincial real environmental efficiency based on three-stage DEA model [J]. China Population, Resources and Environment, 2018, 28(3): 106-116. (in Chinese)

- [6] SEXTON T R, SILKMAN R H. Data envelopment analysis: critique and extensions [M]. San Francisco: Jossey-Bass, 1986.
- [7] DOYLE J, GREEN R. Efficiency and cross-efficiency in DEA: derivations, meanings and uses[J]. Journal of the Operations Research Society, 1994, 45(5): 567–578.
- [8] 刘学之, 黄敬, 王玉. 基于 DEA 交叉效率模型的钢铁行业能源效率分析[J]. 管理世界, 2017(10): 182–183.
LIU X Z, HUANG J, WANG Y. A study of steel industry: based on the cross-efficiency DEA model [J]. Management World, 2017(10): 182–183. (in Chinese)
- [9] 高明美, 孙涛, 朱建军. 一种改进的直觉模糊熵公理化定义和构造公式[J]. 控制与决策, 2014, 29(3): 470–474.
GAO M M, SUN T, ZHU J J. Revised axiomatic definition and structural formula of intuitionistic fuzzy entropy[J]. Control and Decision, 2014, 29(3): 470–474. (in Chinese)
- [10] 高明美, 孙涛, 朱建军. 基于改进熵和新得分函数的区间直觉模糊多属性决策[J]. 控制与决策, 2016, 31(10): 1757–1764.
GAO M M, SUN T, ZHU J J. Interval-valued intuitionistic fuzzy multiple attribute decision-making method based on revised fuzzy entropy and new scoring function[J]. Control and Decision, 2016, 31(10): 1757–1764. (in Chinese)
- [11] 范建平, 薛坤, 吴美琴. 基于直觉模糊熵的交叉评价方法[J]. 计算机科学, 2018, 45(2): 280–286.
FAN J P, XUE K, WU M Q. Cross evaluation method based on intuitionistic fuzzy entropy [J]. Computer Science, 2018, 45(2): 280–286. (in Chinese)
- [12] ADLER N, GOLANY B. Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe [J]. European Journal of Operational Research, 2001, 132(2): 260–273.
- [13] TYTECA D. On the measurement of the environmental performance of firms—a literature review and a productive efficiency perspective [J]. Journal of Environmental Management, 1996, 46: 281–308.
- [14] FÄRE R, GROSSKOPF S. Modeling undesirable factors in efficiency evaluation; comment [J]. European Journal of Operational Research, 2004, 157: 242–245.
- [15] HOTELLING H. Analysis of a complex of statistical variables into principal components [J]. Journal of Educational Psychology, 1933, 24(6): 417–520.

Application of an environmental cross data envelopment analysis (DEA) model based on the principal component analysis (PCA) method

ZHANG KeYu¹ ZHOU LingLing² HAN YongMing^{3, 4*}

(1. School of Economics, Beijing Wuzi University, Beijing 101149; 2. School of International Economics and Business, University of International Business and Economics, Beijing 100029; 3. College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029; 4. Engineering Research Center of Intelligent PSE, Ministry of Education, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: In this work we define a cross environmental performance index (EPI), and propose a new method which integrates principal component analysis (PCA) into a cross evaluation matrix (CEM) in order to calculate the EPI; this overcomes the problem of the traditional cross data envelopment analysis (DEA) model being unable to utilize the information in a CEM. Moreover, our PCA-based method can give different weights to each decision making unit according to the contributions of variances in the CEM. In addition, the new method also removes redundant information in the CEM and improves the evaluation efficiency of the environmental cross-DEA model. Finally, by using the data for input and output of energy, we compare and analyze the results of the traditional cross-efficiency evaluation method and our revised method based on PCA in an assessment of energy efficiency.

Key words: environmental cross data envelopment analysis (DEA) model; principal component analysis (PCA); dimension reduction; energy efficiency

(责任编辑: 吴万玲)