EBM模型

Epsilon-based measure model

1.模型提出的背景

- DEA模型主要分为径向和非径向两大类，它们都有各自的优点和缺陷。
- 为了更好地体现径向和非径向模型的长处，尽可能少地避免两者的缺陷，可以使用包含径向和非径向的混和模型（Hybrid)，EBM模型事实上也是一种混和模型。
- 但Hybrid模型只是从投入或产出要素的径向或非径向上做文章，但有时决策者是很难搞清楚哪个投入或产出是径向或非径向的。
- EBM模型试图把径向和非径向整合到一个统一的框架中，它提供了两个参数，一个标量（scalar)一个向量（vector),通过标量$\varepsilon_x(\varepsilon_y)$（Epsilon）把径向和非径向联系起来，通过PCA（主成份分析）来决定投入和产出的权重$\mathbf{w}^-(\mathbf{w}^+)$

2.模型的具体思路

EBM—I—C

$$\gamma^* = \min_{\theta,\boldsymbol{\lambda},\mathbf{s}^-} \theta - \varepsilon_x \sum_{i=1}^{m} \frac{w_i^- s_i^-}{x_{io}}$$

$$\text{subject to} \quad \theta \mathbf{x}_o - \mathbf{X}\boldsymbol{\lambda} - \mathbf{s}^- = \mathbf{0}$$

$$\mathbf{Y}\boldsymbol{\lambda} \ge \mathbf{y}_o$$

$$\boldsymbol{\lambda} \ge \mathbf{0}$$

$$\mathbf{s}^- \ge \mathbf{0}.$$

w_i^- 是投入i的重要程度的权重，满足 $\sum_{i=1}^{m} w_i^- = 1 \quad (w_i^- \ge 0\ \forall i)$

ε_x 是联接径向效率θ和非径向松弛变量的关键参数。

这两个参数必须在求解效率之前确定，EBM模型的关键部分就是确定合适的两参数问题。

3.参数确定的思路

异化指数和亲和指数（**Diversity index and affinity index** ）

$\mathbf{a} \in R_+^n$ and $\mathbf{b} \in R_+^n$ 表示两个n维正向量，指n个DMU的投入值

定义a和b间亲和指数$S(\mathbf{a},\mathbf{b})$，满足下列性质

$S(\mathbf{a},\mathbf{a}) = 1 \quad (\forall \mathbf{a})$ 同一性

$S(\mathbf{a},\mathbf{b}) = S(\mathbf{b},\mathbf{a})$ 对称性

$S(t\mathbf{a},\mathbf{b}) = S(\mathbf{a},\mathbf{b}) \ (\forall t > 0)$ 单位不变性

$1 \geq S(\mathbf{a},\mathbf{b}) \geq 0 \quad (\forall \mathbf{a},\mathbf{b})$

3.参数确定的思路

$$c_j = \ln \frac{b_j}{a_j} \quad (j = 1, \dots, n)$$

$$\bar{c} = \frac{1}{n} \sum_{j=1}^{n} c_j$$

$$c_{\max} = \max_j \{c_j\} \text{ and } c_{\min} = \min_j \{c_j\}$$

定义如左

定义异化指数为

$$D(\mathbf{a}, \mathbf{b}) = \frac{\sum_{j=1}^{n} \left| c_j - \bar{c} \right|}{n(c_{\max} - c_{\min})}$$
$$= 0 \quad \text{if } c_{\max} = c_{\min}$$

$$0 \le D(\mathbf{a}, \mathbf{b}) = D(\mathbf{b}, \mathbf{a}) \le \frac{1}{2}$$

当且仅当a和b成比例时， D=0

3.参数确定的思路

定义亲和指数

$$S(\mathbf{a},\mathbf{b}) = 1 - 2D(\mathbf{a},\mathbf{b})$$

如果亲和指数位于【0，1】，则可证明它满足以上4条性质

$$r(\mathbf{a},\mathbf{b}) = \frac{\sum_{j=1}^{n}(a_j - \bar{a})(b_j - \bar{b})}{\sqrt{\sum_{j=1}^{n}(a_j - \bar{a})^2}\sqrt{\sum_{j=1}^{n}(b_j - \bar{b})^2}}$$

注：为什么用亲和指数代替Pearson相关系数？（1）a和b的绝对值对r(a,b)影响较大，而在DEA中我们主要关心相对度量。（2）Pearson相关系数 $-1 \le r(\mathbf{a},\mathbf{b}) \le 1$ 这与PCA分析相矛盾。（3）用对数形式代替非对数形式，在于不取对数违反第2性质（对称性）。

3.VRS的扩展

第一步，在VRS前沿下，求解加法模型，得到最优的松弛，得到相应的目标值为

$$\bar{x}_{io} = x_{io} - s_i^{-*} \quad (i = 1, \ldots, m)$$
$$\bar{y}_{io} = y_{io} + s_i^{+*} \quad (i = 1, \ldots, s).$$

形成VRS有效的数据集，即

$$\begin{bmatrix} \bar{\mathbf{X}} \\ \bar{\mathbf{Y}} \end{bmatrix} = \begin{bmatrix} \bar{x}_{11} \cdots \bar{x}_{1n} \\ \cdots\cdots\cdots\cdots \\ \bar{x}_{m1} \cdots \bar{x}_{mn} \\ \bar{y}_{11} \cdots \bar{y}_{1n} \\ \cdots\cdots\cdots\cdots \\ \bar{y}_{s1} \cdots \bar{y}_{sn} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{x}}_1 \\ \cdots \\ \bar{\mathbf{x}}_m \\ \bar{\mathbf{y}}_1 \\ \cdots \\ \bar{\mathbf{y}}_s \end{bmatrix}$$

第二步，亲和矩阵。以投入角度为例，

$$s_{ij} = S(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j)$$

所有元素满足 $1 \geq s_{ij} \geq 0 \ (\forall (ij))$

第三步，求解最大特征值 和特征向量。由于S矩阵是对称、非负且对角元素都为1，根据Perron-Frobenius理论，S一定有最大的特征根 ρ_x 和特征向量$\mathbf{w}_x (\geq 0)$ 且$m \geq \rho_x \geq 1$

第四步，计算ε_x **and** $\mathbf{w}^-$

$$\varepsilon_x = \frac{m-\rho_x}{m-1} \quad (\text{if } m > 1 \text{ and } m > \rho_x)$$
$$= 0 \quad (\text{if } m = 1 \text{ or } m \leq \rho_x).$$
$$\mathbf{w}^- = \frac{\mathbf{w}_x}{\sum_{i=1}^{m} w_{xi}}.$$

第五步，把上述参数代入EBM中。如果已经设定了权重，则直接使用上述参数。

PCA（主成份分析）过程略。

产出角度和非角度过程与此类似。

主要模型类别

EBM-I-C
EBM-I-V
EBM-O-C
EBM-O-V

EBM-C
EBM-V