

基于遗传算法的前沿生产函数构造及其应用

王金祥, 吴育华, 张慎峰
(天津大学管理学院, 天津 300072)

摘要: 效率测度分析是评价企业经营管理的重要指标,其求解往往同构造前沿生产函数联系在一起.对采用遗传算法构造前沿生产函数进行了研究,电力工业的实证效率测度分析以及与非参数方法的比较结果验证了本方法的正确性和实用性,可为电力工业的决策者和这一领域的研究人员提供参考.

关键词: 前沿生产函数; 遗传算法; 效率测度; 电力工业; 非参数方法

中图分类号: F403.3 **文献标识码:** A **文章编号:** 1007 - 9807(2004)02 - 0013 - 05

测定决策单元(DMU)的效率往往同构造投入产出前沿生产函数联系在一起^[1].前沿生产函数的构造可采用参数方法和非参数方法.非参数方法^[2]是在假定投入不变的条件下对所观测样本构成的实际生产曲面向相对生产前沿面的最大扩展程度的测度,是现有样本产出向相对有效凸多面体包络面多维扩展程度的度量.而产出效率测度参数方法是在预先确定投入产出函数具体形式的前提下,对既定函数的生产曲面向样本总体决定的最大可能生产曲面的延伸程度的度量.参数方法较多,如 Aigner 和 Chu^[3]1968年提出的用线性规划或二次规划方法估计确定前沿生产函数(The Aigner-Chu deterministic frontier function)的参数方法,Aigner, Lovell 和 Schmidt, Meeusen 和 Van den Broeck^[4]1977年提出的随机前沿生产函数(stochastic frontier function)的参数方法等.各种方法构造前沿面思路不同,非参数方法由于有几个边和顶而不可微分,参数方法构造的前沿面是光滑的,在计算弹性等方面有优越性.

基于遗传算法的前沿生产函数应归类于参数方法,但其前沿生产函数构造过程与以往的参数方法有所区别,这种方法在弹性计算和效率测度

分析方面十分方便.

1 基于遗传算法的前沿生产函数构造思路

评价决策单元的效率往往采用决策单元的产出与有效前沿面的理想产出距离.而有效前沿面是未知的,各种方法都是以构造有效前沿面为出发点,寻找一个有效前沿面来求效率^[5],这些技术有两个特点:一是关于前沿面的假设,二是前沿面产生的过程.第一方面的讨论很多,如 Greene^[6]和 Lovell^[7].第二方面则想区分前沿面是由一个确定过程还是一个随机过程产生.确定性过程和随机性过程产生前沿面的区别在于:确定性过程产生前沿面可看作一个求解过程,不同方法最后都可到达相同的期望前沿面,即采用相同的数据、相同的方法和相同的限定会得到相同的结果;随机过程产生前沿面,即使起点相同,经过随机过程重复运行之后结果也不会相同,本文讨论的基于遗传算法构造前沿生产函数属于随机过程产生前沿面的方法.

考虑 K 个样本点, N 个投入变量,单个产出变量情形,假设 $y_i = f(X_i; \theta)$ ($i = 1, 2, \dots, K$), 这

里 y_i 表示第 i 个样本点的产出, $X_i = (x_{i1}, x_{i2}, \dots, x_{iN})$ 表示第 i 个样本点的投入, θ_0 是待估参数. 假设存在一个最优生产前沿函数 y^0 , 它是在所给样本点数据基础上获得的理想产出, 并且保证 $y_i - \theta_0 x_i = 0$, 则各个样本点的投入产出效率可用 $y_i / \theta_0 x_i$ 表示. 在单投入单产出情形下, 其直观含义如图 1 所示.

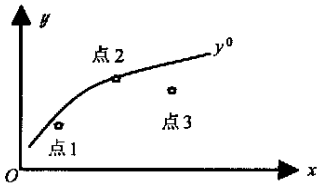


图 1 单投入单产出下不同样本点的效率
Fig. 1 Efficiency of different sample on the condition of single input and single output

图 1 中, 曲线 y^0 是理想产出前沿面, 要求所有的样本点产出不能超出这条曲线, 样本点 2 的实际产出十分接近理想前沿面, 因而其效率接近 1, 而点 1 和点 3 的产出与理想前沿面有一定距离, 其效率要小于 1, 距离前沿面的远近反映效率的大小.

求最优生产前沿函数 y^0 涉及两个问题: 函数的形式和参数估计. 不同的函数形式对效率有影响, 在假设函数形式时要进行数据拟合检验, 保证所假设的函数形式能够反映样本点的投入产出状况. 参数估计可以通过遗传算法解决.

遗传算法是 Holland 提出的基于自然遗传学和计算机科学的优化方法^[8], 通过遗传基因代码, 利用复制、杂交、变异 3 种基本算子, 进行全局寻优. 当染色体域足够大和遗传叠代次数足够多时, 理论上讲, 遗传算法可以给出原问题的最优解. 本文主要讨论采用遗传算法估计前沿生产函数的参数.

2 算例

以电力工业为例, 说明本方法的应用以及对应用结果进行分析.

2.1 数据及模型的构造

电力工业经营体制近年来发生了重大变革,

由于市场竞争力和经营管理水平差距明显, 各省电力公司资源配置效率差异逐渐加大. 根据 1999 年《中国电力工业统计年鉴》相关资料, 为方便分析问题, 保证样本企业的可比性和定量分析的有效性, 选择六大区中有代表性的八个省市电力公司为样本企业, 这些电力公司能够代表我国电力事业的各个层次, 它们是: 天津、吉林、江西、山东、湖北、广东、宁夏和贵州. 其投入产出数据如表 1 所示.

表 1 电力工业样本企业的投入产出状况

Table 1 Input and output state of electric power industry samples

样本	生产性投资 / 万元	生产费用 / 万元	发电量 / 万 kW · h
天津电力	180 652	164 700	1 460 400
吉林电力	168 154	364 500	2 939 800
江西电力	164 696	207 000	1 832 900
山东电力	672 380	936 675	7 930 300
湖北电力	407 996	271 125	4 672 400
广东电力	597 397	561 390.3	8 086 600
宁夏电力	109 751	125 127	1 115 400
贵州电力	141 980	184 812	2 040 500

资料来源: 1999 年《中国电力工业统计年鉴》

采用遗传算法构造前沿面首先要假设函数形式. 假设本投入产出函数形式为, $s_i = x_{i1} + bx_{i2}$, 其中, s_i 表示第 i 个样本点基于投入数据的理想发电量, x_{i1} 表示第 i 个样本企业生产性投资, x_{i2} 表示第 i 个样本企业生产费用, 第 i 个样本企业的实际产出用 y_i 表示. 其线形拟合检验相关系数为 $R^2 = 0.9473$, 因此做线性形式假设是合适的.

2.2 用遗传算法寻求前沿生产函数的参数

首先构造一系列群体, 每个群体都包含一个备选前沿面 $s(X; \theta)$, 每个备选前沿面对应不同的 θ 参数, 本算例中即 θ_0 和 b . 利用达尔文的思想 and 模拟进化理论, 提高群体的质量, 最终群体中最好的备选前沿面就是进化的结果, 对应的 θ_0 和 b 值即为所寻求的参数.

遗传算法步骤如下:

1) 编码. 编码一般采用二进制方式. 假设一

个变量 x 的取值范围为 $[1, 2]$, 则可将该区域线性映射到整数区域 $[0, 2^n]$, 其中 n 为变量 x 的编码长度. 在本算例中可令两个参数 a 和 b 取值范围为 $[0, 20]$, 并按照二进制方式编码.

2) 初始化. 选择一个整数 N 作为群体的规模参数, 随机地产生 N 个个体组成初始群体. 在本问题中, 通过遗传算法的群体生成程序产生多个 $s(X_i)$, 每个 $s(X_i)$ 都是一个备选前沿面. 本算例中令 $N = 10\ 000$.

3) 计算适应值. 计算每个备选前沿面 $s(X_i)$ 的生存概率和适应值, 适应值公式为

$$fitness = \frac{1}{\sum_{i=1}^K s(X_i)} \quad (1)$$

依照前沿面的定义, 叠代的过程要保证 $s(X_i) - \rho_i > 0$. $fitness$ 值越大, 对应的备选前沿面 $s(X_i)$ 生存的概率越大, 这将保证数据偏离程度小的备选前沿面生存下来.

4) 遗传算子. 利用选择策略从群体中选择进行繁殖的个体组成父代, 对其重组, 即用杂交和变异算子来形成下一代新的个体. 在前代染色体域的基础上产生新一代染色体域的工作称为遗传操作.

基本的遗传操作有三种:

复制. 根据前代染色体的适应值所确定的繁殖概率选择染色体并将其拷贝到下一代, 染色体不发生变化, 本算例有 40% 的备选前沿面复制到下一代.

杂交. 随机地选择两个已经进行繁殖的染色体为母体, 再随机地选择一个交叉位置, 将这两个母体位于交叉位置后的符号互换, 形成两个新的染色体, 本例让 40% 的备选前沿面的参数相互进行杂交.

变异. 将前代染色体的某位基因值由 1 变为 0, 由 0 变为 1. 本算例让 20% 的备选前沿面进行变异.

这种构造前沿面的方法能够依据每个备选前沿面 $s(X_i)$ 对提供数据的偏离程度给出适应值, 偏离程度小的个体适应值大, 下一代生存几率较大, 依据适应值通过复制、杂交和变异产生下一代, 然后计算下一代适应值, 如此反复, 直到满足

一定条件后停止. 最后, 当 $s(X_i)$ 十分接近 0 时, 对应的 s_i 值可以当作待估参数.

按照上述步骤, 最后的参数 a 的计算结果为 $a = 10.43$, $b = 3.32$, 即理想产出函数可近似写为 $s_i = 10.43x_{i1} + 3.32x_{i2}$, 然后计算各个样本点的投入产出效率, 计算过程和结果如表 2 所示.

表 2 基于遗传算法的前沿生产函数投入产出效率测度与非参数模型的效率测度

Table 2 Input and output efficiency measurement based on frontier production function constructed with genetic algorithm and non-parametric efficiency measurement

样 本	实际产出 / 万元	理想产出 s_i / 万元	效率测度 / s_i	非参数 效率测度
天津电力	1 460 400	2 431 004	0.600 74	0.606 72
吉林电力	2 939 800	2 963 986	0.991 84	1
江西电力	1 832 900	2 405 019	0.762 11	0.764 58
山东电力	7 930 300	10 122 684	0.783 42	0.786 64
湖北电力	4 672 400	5 155 533	0.906 29	1
广东电力	8 086 600	8 094 667	0.999 00	1
宁夏电力	1 115 400	1 560 125	0.714 94	0.716 70
贵州电力	2 040 500	2 094 427	0.974 25	0.977 69

2.3 模型求解与非参数方法测算结果比较分析

为了验证模型求解结果的合理性, 将计算结果同非参数模型的效率测度进行了比较. K 个样本点, N 个投入变量, 单个产出变量的面向投入的非参数模型的数学表达式为:

$$\min \theta_0$$

$$s. t. \begin{cases} \sum_{i=1}^K \lambda_i X_i \leq \theta_0 X_0 \\ \sum_{i=1}^K \lambda_i y_i \geq y_0 \\ \lambda_i \geq 0, i = 1, 2, \dots, K \end{cases} \quad (2)$$

式中: $X_i = (x_{i1}, x_{i2}, \dots, x_{iN})$ 表示第 i 个样本点的投入; y_i 表示第 i 个样本点的产出. 上述模型求解的经济含义: 某一决策单元 (X_0, y_0) 的产出 y_0 , 可由所有 K 个样本点产出线性组合替代的情况下, 投入 X_0 的可压缩程度. 压缩比例的大小为 θ_0 , 称为效率测度值. 其中 K 个样本点线性组合的“权重”为 λ_i . 根据实际情况, 可以对权重 λ_i 进行限制. 按照上述模型, 算例的非参数效率测度值的计算结果列于表 2 的最后一列.

两种方法测算结果排序, 结果如表 3 所示.

表 3 基于遗传算法的前沿生产函数投入产出效率测度与非参数模型效率测度排序
Table 3 Order of input and output efficiency measurement based on frontier production function constructed with genetic algorithm and non-parametric efficiency measurement

样 本	广东电力	吉林电力	贵州电力	湖北电力	山东电力	江西电力	宁夏电力	天津电力
遗传算法效率值	0.999 00	0.991 84	0.974 25	0.906 29	0.783 42	0.762 11	0.714 194	0.600 74
排列顺序	1	2	3	4	5	6	7	8
非参数效率值	1	1	0.977 69	1	0.786 64	0.764 58	0.716 70	0.606 72
排列顺序	1	2	4	3	5	6	7	8

产出效率表示每个样本企业产出的可扩展程度,效率值越大,现实产出面越接近于前沿面,效率也越高.从表 2 的测算结果可以发现,非参数测度方法由于没有具体生产函数形式限制,可以实现产出的最大可能扩张,其效率的相对最大产出值等于 1,而参数方法测算的产出扩张由于受到具体函数形式的限制,往往达不到相对最大产出值,其效率小于 1.以上的差别造成两种方法效率测算值的不同和顺序的略微差别.

通过测算可以将样本企业的效率分为 3 个档次,第一个档次是 0.9 以上,包括广东、吉林、贵州和湖北,效率值表明它们的生产技术水平和管理技术水平相对得到了充分的发挥;第二个档次是 0.7 ~ 0.9 之间的企业,包括山东、江西和宁夏,处于中等水平,有待通过管理充分发挥现有生产资源的利用效率,提高经济效益和市场竞争能力;0.7 以下为第三个档次,包括天津电力,说明其生产资源的利用效率相对较低,生产技术和管理工作水平亟待提高.

参 考 文 献:

- [1] Farrell M J. The measurement of productive efficiency[J]. Journal of the Royal Statistical Society(series A), 1957, (3): 253—281.
- [2] Charnes A, Cooper W W, Rhodes E. Measuring the efficiency of decision making units[J]. European Journal of Operational Research, 1978, (2): 429—444.
- [3] Aigner D J, Chu D S. On the industry production function[J]. American Economic Review, 1968, 58: 826—839.
- [4] Meeusen W, Van den Broeck J. Efficiency estimation from Cobb-Douglas production functions with composed error[J]. International Economic Review, 1977, 18: 435—444.
- [5] 孙 巍, 杨庆芳, 杨树绘. 产出资源配置的参数测度与非参数测度及其比较分析[J]. 系统工程理论与实践, 2000, (6): 118—122.
Sun Wei, Yang Qing-fang, Yang Shu-hui. The parametric and nonparametric measurement of the out allocation efficiencies and their comparison[J]. Systems Engineering - Theory & Practice, 2000, (6): 118—122. (in Chinese)
- [6] Greene W H. The Econometric Approach to Efficiency Analysis[M]. New York: Oxford University Press, 1993.
- [7] Lovell C A K, Kumbhakar S C. Stochastic Frontier Analysis[M]. London: Cambridge University Press, 2000.
- [8] Koza J. Genetic Programming: On the Programming of Computers by Means of Natural Selection[M]. Cambridge: MIT Press, 1992.

3 结 束 语

基于遗传算法的前沿生产函数的构造是参数构造前沿面方法的一种,这种方法不考虑误差项,在这一点上同确定生产前沿面构造方法 Aigner-Chu 算法非常相似,但其参数的寻求是随机的,而 Aigner-Chu 算法则是通过规划求解计算得到唯一确定的参数.同随机前沿面方法相比,这种算法在计算上非常方便,随机前沿面方法在误差项分解上十分麻烦,影响了该方法的应用.

通过算例可以看出,遗传算法在搜寻方程中参数的最优值方面是一个有效的方法.基于遗传算法构造的生产前沿面是现有投入的理想产出,这一点同非参数模型构造的前沿面道理是相同的,本方法由于给出了具体的函数形式,便于做评价和预测,在这一点上同非参数方法构造生产前沿面相比在应用上有优势.从效率测算结果以及与非参数方法的比较来看,本方法可以做为投入产出效率分析工具.

Construction of frontier production function with genetic algorithm and its application

WANG Jin-xiang, WU Yu-hua, ZHANG Shen-feng

School of Management, Tianjin University, Tianjin 300072, China

Abstract : Efficiency measurement is an important index to evaluate firms, the solving of it is connected with the construction of the frontier. Based on genetic algorithm, the frontier model with parameter character is constructed in this paper. The theory proved to be correct and useful with the empirical efficiency measurement of electric power industry and the comparison with nonparametric approach. It is the hope that this paper will be of interest to the decision-makers of electric power industry and the researchers of this field.

Key words : frontier production function; genetic algorithm; efficiency measurement; electric power industry; non-parametric approach

(上接第 6 页)

Control strategy to technology standard of firm based on network externality

WENG Yi-cong¹, CHEN Hong-min¹, KONG Xin-yu²

1. School of Management, Shanghai Jiaotong University, Shanghai 200030, China;

2. Department of World Economics, Fudan University, Shanghai 200437, China

Abstract : The paper analyzes how the dominant producer, to control its technology standard which is mainstream in the market which the network externality exists. The results show that the intensity of network externality and scale of the installed base have great influence on the decision of the dominant producer when the dominant producer faces the threat of the cooperative behavior of the fringes. Only if network externality is small and the scale of installed base of the dominant producer is large, the dominant producer should monopolize its standard. And when the network externality is large or the scale of installed base of the dominant producer is small, the dominant producer should fully or targetedly open its standard. When the government orders the dominant producer shouldn't open targetedly its standard, the social welfare might be impaired.

Key words : technology standard; merger; technology alliance; network externality