

生产函数中参数估计方法的进一步改进

宋海洲

(华侨大学数学系, 福建泉州 362021)

摘要 在原有生产函数参数的估计方法基础上, 提出两种新的估计方法. 这两种新的估计方法具有最小的回归残差平方和, 文中给出实现这两种参数估计方法的具体算法.

关键词 生产函数, 残差平方和, 非线性最小二乘法, 遗传算法

中图分类号 O 241.82

文献标识码 A

1928年, 由美国数学家 Cobb 和经济学家 Douglas 提出的生产函数这个概念以来, 生产函数在生产实际中得到了广泛的应用. 生产函数反映了投入要素与产出量之间的关系. 建立生产函数模型, 我们可以分析资金、劳动力等生产要素在经济发展中的作用, 可以分析科技进步及信息对经济发展产生的重要作用. 此外, 利用生产函数可以合理安排有限的资源, 使经济发展更快提供十分有用的信息. 利用生产函数也可以为我们国家的领导进行决策提供有力的理论依据. 因此, 合理选择估计生产函数参数的方法, 具有一定的意义.

1 生产函数模型及其常见参数估计法

1.1 生产函数模型

在多种投入要素下, 常见的生产函数模型为

$$y = ax_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} e^{\mu} \quad (1)$$

在式(1)中, y 为产出量, x_i 为第 i 种生产要素投入量, a 为效率系数, α_i 为第 i 种投入要素的产出弹性, $a, \alpha_i (i=1, \dots, n)$ 均为待估计参数, μ 为随机扰动项, $\mu \sim N(0, \sigma^2)$.

1.2 生产函数模型中参数常见的估计方法

(1) 方法 1. 对公式(1)两边取对数, 得到

$$\ln y = \ln a + \alpha_1 \ln x_1 + \alpha_2 \ln x_2 + \cdots + \alpha_n \ln x_n + \mu \quad (2)$$

令 $y^* = \ln y, a^* = \ln a, x_1^* = \ln x_1, \dots, x_n^* = \ln x_n$, 代入公式(2)得

$$y^* = a^* + \alpha_1 x_1^* + \alpha_2 x_2^* + \cdots + \alpha_n x_n^* + \mu \quad (3)$$

在样本数据(假设样本个数为 m)的支持下, 利用多元线性回归的最小二乘法, 可得式(3)中参数 $a^*, \alpha_1, \dots, \alpha_n$ 的估计值 $\hat{a}^*, \hat{\alpha}_1, \dots, \hat{\alpha}_n$, 并使残差

$$S_e = \sum_{j=1}^m (y_j^* - \hat{y}_j^*)^2 \quad (4)$$

最小. 再得式(1)中的估计值为

$$\hat{a} = e^{\hat{a}^*}, \quad (5)$$

而 $\alpha_1, \dots, \alpha_n$ 的估计值为 $\hat{\alpha}_1, \dots, \hat{\alpha}_n$ (即保持不变). 从而, 得到多种投入要素的生产函数的回归形式为

$$y = e^{\hat{a}^*} x_1^{\hat{\alpha}_1} x_2^{\hat{\alpha}_2} \cdots x_n^{\hat{\alpha}_n} \quad (6)$$

(2) 方法 2. 由文[1]知, 公式(5)估计出的期望 $E(\hat{a}) \neq a$, 即方法 1 估计的参数是有偏的. 因此, 文[1]

改进了参数 a 的估计量 \hat{a} . \hat{a} 重新定义为

$$\hat{a} = e^{\hat{a}^*} \left[1 + \frac{1}{2} D(\hat{a}^*) \right]^{-1}, \tag{7}$$

其中 $D(\hat{a}^*)$ 为 \hat{a}^* 的方差, 其它参数的估计量不变. 此时多种投入要素的生产函数的回归形式为

$$y = e^{\hat{a}^*} \left[1 + \frac{1}{2} D(\hat{a}^*) \right]^{-1} x_1^{\hat{\alpha}_1} x_2^{\hat{\alpha}_2} \dots x_n^{\hat{\alpha}_n}. \tag{8}$$

(3) 方法 3. 文[2]提出一种新方法, 是先根据方法 1 利用最小二乘法估计出 $\hat{\alpha}_1, \dots, \hat{\alpha}_n$. 然后将 $\hat{\alpha}_1, \dots, \hat{\alpha}_n$ 代入模型(1), 并忽略随机扰动项 e^u . 设产出为 y , 再直接利用最小二乘法求出 \hat{a} , 使残差平方和 S_{el} 达到最小. 即 $S_{el} = \sum_{j=1}^m (y_j - \hat{y}_j)^2 \Rightarrow \min$, 可得

$$\hat{a} = \frac{\sum_{j=1}^m y_j \prod_{i=1}^n x_{ji}^{\hat{\alpha}_i}}{\sum_{j=1}^m \left(\prod_{i=1}^n x_{ji}^{\hat{\alpha}_i} \right)^2}. \tag{9}$$

此时, 多种投入要素的生产函数的回归形式为

$$y = \hat{a} x_1^{\hat{\alpha}_1} x_2^{\hat{\alpha}_2} \dots x_n^{\hat{\alpha}_n}. \tag{10}$$

2 参数估计方法的进一步改进

2.1 参数估计的两种方法

上述 3 种方法均是对非线性模型利用变换化为线性模型的方法, 再利用最小线性二乘法估计其中的参数. 显然, 这些方法估计出来的参数并未使得回归残差平方和

$$S_{el} = \sum_{j=1}^m (y_j - \hat{y}_j)^2 \triangleq g(a, \alpha_1, \alpha_2, \dots, \alpha_n) \tag{11}$$

最小. 本文提出参数 $a, \alpha_1, \dots, \alpha_n$ 估计的两种新方法. 其中一个新方法(本文称之为方法 4)是利用求解无约束优化的拟牛顿法^[3](BFGS 公式、DFP 公式), 直接求 $a, \alpha_1, \dots, \alpha_n$ 的估计量 $\hat{a}, \hat{\alpha}_1, \dots, \hat{\alpha}_n$, 使得式(11)达到最小. 另一个新方法(本文称为方法 5)是采用改进的遗传算法, 求 $a, \alpha_1, \dots, \alpha_n$ 的估计量 $\hat{a}, \hat{\alpha}_1, \dots, \hat{\alpha}_n$, 使得式(11)达到最小.

2.2 方法 4 的实现

即利用拟牛顿法估计参数的实现. 这种方法在数学软件十分发达的今天, 实现之非常容易. 例如, 在 Matlab 中就可利用 Leastsq 函数或利用 Nlinfit 函数, 实现 $a, \alpha_1, \dots, \alpha_n$ 的估计. 由于式(11)为非线性函数, 故这种方法我们称之为非线性最小二乘法.

2.3 方法 5 的实现

遗传算法^[4](Genetic Algorithms, 简称 GA), 它是模拟达尔文的遗传选择和自然淘汰的生物进化过程的优化方法. 它只要求被优化的函数是可计算的, 不要求目标函数具有连续可微性. 这是一种并行的多点搜索过程. 因此, 遗传算法具有简单通用、鲁棒性强、适于并行处理以及高效、实用等显著特点. 但由于简单遗传算法中存在着一些问题, 如收敛速度太慢, 容易陷入局部最优等缺陷. 为此, 我们可以在简单遗传算法中作如下改进(图 1). (1) 人工方法产生初始种群. 先将优化问题的一个近似最优解作为初始种群的一个个体. 一种

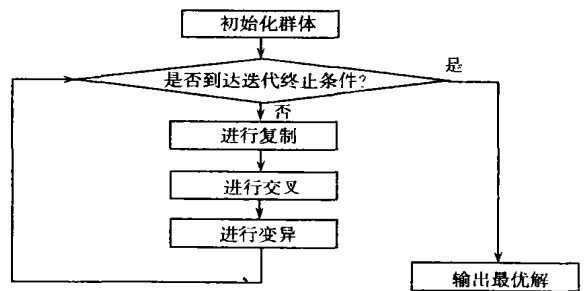


图 1 改进的遗传算法流程图

方法是先利用基本遗传算法求其近似最优解, 该近似最优解作为初始种群的一个个体. (2) 进行适应度变换. 在遗传算法的早(后)期阶段, 对个体的适应度进行适当的缩小或(放大). 改变个体之间的差异程度, 以维护群体的多样性(提高个体之间的竞争性). (3) 优选父代进行交叉. 进行“优”、“优”交叉或者进行“差”、“差”交叉, 即适应度较高的个体之间进行交叉或适应度较低的个体之间进行交叉. 产生更好的新个体代替这部分个体. (4) 采取精英法则. 父代的最优个体直接进入子代. (5) 自适应调整交叉概率

与变异概率. 随着遗传算法在线性能的提高, 相应增大交叉概率. 随着遗传算法在线性能的下降, 相应减少变异概率. (6) 采用自适应变焦技术. 在稳定的极值点附近进行控制变焦, 具体方法为当算法进入局部微调时, 缩小设计变量的范围. 利用改进的遗传算法求 $a, \alpha_1, \dots, \alpha_n$ 的估计量 $\hat{a}, \hat{\alpha}_1, \dots, \hat{\alpha}_n$ 的流程图, 如图 1 所示.

3 实例及 5 种参数估计方法的比较

现有美国马萨诸塞州 1820~ 1926 年的产值、资金投入和劳动力投入等 3 个经济指数统计数据^[3], 如表 1 所示. Douglas 生产函数模型中, 产值 y 、资金投入 x_1 和劳动力投入 x_2 之间关系为

$$y = ax_1^{\alpha_1}x_2^{\alpha_2}e^{\mu}.$$
 (12)

表 1 美国马萨诸塞州 1820~ 1926 年的产值 y 及资金投入 x_1 和劳动力投入 x_2 统计数据(以 1899 年为 1.00)

t	y	x_1	x_2	t	y	x_1	x_2	t	y	x_1	x_2
1890	0.72	0.95	0.78	1903	1.30	1.22	1.22	1915	2.00	3.24	1.62
1891	0.78	0.96	0.81	1904	1.30	1.27	1.17	1916	2.09	3.61	1.86
1892	0.84	0.99	0.85	1905	1.42	1.37	1.30	1917	1.96	4.10	1.93
1893	0.73	0.96	0.77	1906	1.50	1.44	1.39	1918	2.20	4.36	1.96
1894	0.72	0.93	0.72	1907	1.52	1.53	1.47	1919	2.12	4.77	1.95
1895	0.83	0.86	0.84	1908	1.46	1.57	1.31	1920	2.16	4.75	1.90
1896	0.81	0.82	0.81	1909	1.60	2.05	1.43	1921	2.08	4.54	1.58
1897	0.93	0.92	0.89	1910	1.69	2.51	1.58	1922	2.24	4.54	1.67
1898	0.96	0.92	0.91	1911	1.81	2.63	1.59	1923	2.56	4.58	1.82
1899	1.00	1.00	1.00	1912	1.93	2.74	1.66	1924	2.34	4.58	1.60
1900	1.05	1.04	1.05	1913	1.95	2.82	1.68	1925	2.45	4.58	1.61
1901	1.18	1.06	1.08	1914	2.01	3.24	1.65	1926	2.58	4.54	1.64
1902	1.29	1.16	1.18								

公式 (12) 两边取对数, 得

$$\ln y = \ln a + \alpha_1 \ln x_1 + \alpha_2 \ln x_2 + \mu.$$
 (13)

现令 $y^* = \ln y, a^* = \ln a, x_1^* = \ln x_1, x_2^* = \ln x_2$, 代入公式 (13) 得

$$y^* = a^* + \alpha_1 x_1^* + \alpha_2 x_2^* + \mu.$$
 (14)

经过上机求解得到二元线性回归方程为

$$y^* = 0.0068 + 0.2182x_1^* + 0.8457x_2^*.$$
 (15)

由此, 可得

$$\hat{a}^* = 0.0068, \quad \hat{\alpha}_1 = 0.2182, \quad \hat{\alpha}_2 = 0.8457.$$

根据公式 (5), 有 $\hat{a} = e^{\hat{a}^*} = e^{0.0068} = 1.0068$. 所以, 生产函数的回归方程为

$$y = 1.0068x_1^{0.2182}x_2^{0.8457}.$$
 (16)

由计算机算出 \hat{a}^* 的方差为 $D(\hat{a}^*) = 0.0069$, 利用公式 (7) 可求 a 近似的无偏估计量 \hat{a} 为

$$\hat{a} = e^{\hat{a}^*} \left[1 + \frac{1}{2} D(\hat{a}^*) \right]^{-1} = e^{0.0068} (1 + 0.5 \times 0.0069)^{-1} = 1.0034.$$

所以, 生产函数的另一回归方程为

$$y = 1.0034x_1^{0.2182}x_2^{0.8457}.$$
 (17)

利用公式 (9) 可以给出 a 的另外一个估计值 $\hat{a} = 1.0039$. 由此可以得到生产函数的第 3 个回归方程为

$$y = 1.0039x_1^{0.2182}x_2^{0.8457}.$$
 (18)

现利用非线性最小二乘法估计 a, α_1, α_2 , 利用数学软件 Matlab 的 Leastsq 函数编程直接求得 $\hat{a} = 1.0316, \alpha_1 = 0.3609, \alpha_2 = 0.4398$. 得到生产函数的第 4 个回归方程为

$$y = 1.0316x_1^{0.3609}x_2^{0.4398}.$$
 (19)

利用上述实例和改进的遗传算法, 求 $a, \alpha_1, \dots, \alpha_n$ 的估计量 $\hat{a}, \hat{\alpha}_1, \dots, \hat{\alpha}_n$. 利用 Matlab 进行编程, 采用浮

点式编码. 适应值函数采用进行了适应度变换的函数 $f = \exp(6 \times (1 / (1 + g(a, \alpha_1, \alpha_2))))$. 遗传算法中的复制采用轮盘赌法, 并采用精英法则强行保留每一代最优个体. 遗传算法中的交叉采用“优”、“优”交叉或者进行“差”、“差”交叉. 遗传算法中的变异采用单点变异, 并采用自适应调整交叉概率与变异概率方法, 进行多次遗传算法求近似最优解. 在得到多个近似最优解之后, 对变量进行变焦. 将变量缩小为包含这些近似最优解的小区域. 在这些近似最优解中采用最好的一个作为初始群体的一个个体, 再进行一次遗传算法. 我们得到生产函数的第 5 个回归方程为

$$y = 1.0316x_1^{0.3609}x_2^{0.4398}. \quad (20)$$

由式(16), (17), (18), (19) 和式(20) 给出的生产函数的回归方程, 从中选择一种残差平方和最小的. 对这些回归方程的残差平方和进行比较, 有

$$Se_2 = \sum_{j=1}^m (y_j - \hat{y}_j)^2 \Rightarrow \min.$$

经过计算, 式(16), (17), (18), (19), (20) 所决定生产函数的残差平方和, 依次为 1.0178, 1.0170, 1.0170, 0.8241, 0.8241. 可见, 由式(19), (20) 所决定的生产函数的残差平方和最小. 因此, 应该选用式(19), (20) 作为生产函数的回归方程.

4 结 束 语

最小线性二乘法的方法 1、方法 2 和方法 3, 它们估计参数的优点是计算量相对简单, 而缺点是回归方程的回归残差平方和的相对较大. 直接利用求解无约束优化的拟牛顿法(BFGS 公式、DFP 公式) 估计参数的方法 4, 其优点是可以使得回归方程具有最小的回归残差平方和; 缺点是该法有一个较大的限制条件——要求目标函数具有连续可微性. 利用遗传算法估计参数的方法 5, 不仅使得回归方程具有最小的回归残差平方和, 而且还可利用来进行估计参数. 它对目标函数没有特定的要求, 也可以用于其它模型的参数估计, 因而具有较强的通用性. 在数学软件 Matlab 如此发达的今天, 编程计算相当容易, 则此方法应用更加方便. 比如, 我们就利用 Matlab 编制了一个遗传算法工具箱(将另文发表). 因此, 在对生产函数进行参数估计时, 建议大家尽量采用改进的遗传算法.

参 考 文 献

- 1 陈鹤琴. 经济计量学[M]. 北京: 中国商业出版社, 1989. 6~ 198
- 2 王志江. 生产函数中参数估计方法的改进[J]. 运筹与管理, 1999, 8(4): 80~ 84
- 3 姜启源. 数学实验[M]. 北京: 高等教育出版社, 1989. 180~ 204
- 4 王小平, 曹立明. 遗传算法的理论、应用与软件实现[M]. 西安: 西安交通大学出版社, 2002. 18~ 87

Further Improvement of the Method for Estimating Parameters in Production Function

Song Haizhou

(Department of Mathematics, Huaqiao University, 362021, Quanzhou, China)

Abstract Starting from previously existing method for estimating parameters in production function, the author puts forward two new methods of estimation which possesses minimum regression residual sum of squares; and gives specific algorithm for realizing these two methods of parameter estimation.

Keywords production function, residual sum of squares, nonlinear least square method, genetic algorithm