

1. 引言

长期以来，生产函数理论定义了给定一组输入和固定技术条件下能够获得的最大产出量。尽管计量经济学家多年来一直在估算平均生产函数，但直到Farrell（1957年）的开创性工作，估计所谓的生产前沿函数，即理论与实证研究之间差异的桥梁，才受到重视。然而，由于多种原因，这些尝试并未完全达成目标。本文介绍了一种新的生产前沿函数估计方法，通过假设误差项由一个正态分布和一个单侧分布组成，克服了先前方法的一些关键限制。文本结构如下：第2部分回顾了以往的参数化前沿模型估计方法；第3部分提出我们的新模型；第4部分讨论了我们模型的统计特性；第5部分报告了蒙特卡罗实验结果；第6部分提供了一个实证示例；第7部分总结了本文。

2. 先前的参数化前沿模型

以往关于参数化前沿生产函数估计的研究，特别是Aigner和Chu（1968）、Afriat（1972）和Richmond（1974）的工作，主要是假设一个函数能给出给定输入下的最大可能输出。对于某个特定企业，假设为第*i*家，我们有：

$$y_i = f(x_i; \beta) \quad (1)$$

这里 y_i 代表从 x_i （一组非随机输入）能得到的最大输出， β 是待估计的未知参数向量。

Aigner和Chu（1968）建议通过对给定行业内*N*家企业的横截面数据进行数学规划方法估计 β 。具体而言，他们提出最小化：

$$\sum_{i=1}^N (y_i - f(x_i; \beta)),$$

受约束 $y_i \leq f(x_i; \beta)$ ，若 $f(x_i; \beta)$ 对 β 是线性的，则构成线性规划问题。或者，他们提出最小化：

$$\sum_{i=1}^N \|y_i - f(x_i; \beta)\|_2$$

在相同的约束下，若 $f(x_i; \beta)$ 是线性的，则构成二次规划问题。

显然，在从（1）到这些“估计”方法的转变中，发生了一些神奇的事情：为了解释具有相同输入向量的企业之间的产出差异，或解释给定企业的产出为何低于“前沿” $f(x_i, \beta)$ ，隐含地假设了扰动项。

这些方法的一个问题是对异常值极其敏感。这导致了所谓的“概率”前沿的发展[Timmer（1971），Dugger（1974）]，它们是通过上述讨论的相同类型的数学算法估计的，除了允许某些指定比例的观察值位于前沿之上。选择这个比例本质上是任意的，缺乏明确的经济或统计论证。另一个问题涉及如何使位于前沿之上的观察值与作为最大可能产出的前沿概念相协调。通常，这是通过将极端观察值归因于测量误差来解决的。

然而，以尽量上的方式将测量误差和其他不可观测冲击的可能性纳入似乎更为合适。

因此，按照之前的应用，数学规划技术未能提供具有已知统计特性的估计。为了在统计上正当化这些技术，Schmidt（1976）在模型中显式地引入了一个单侧扰动项，得到了以下模型：

$$y_i = f(x_i; \beta) + \varepsilon_i, i = 1, \dots, N, \quad (2)$$

其中 $\varepsilon_i \leq 0$ 。给定扰动项的分布假设，然后通过最大似然技术估计模型。特别是，假设 $-\varepsilon_i$ 有指数分布导致线性规划技术应用，而假设 $-\varepsilon_i$ 有半正态分布导致二次规划技术应用。因此，Aigner和Chu的估计可以被视为在特定误差规格下的最大似然估计。

不幸的是，即使模型可以通过最大似然方法估计，并且在适当假设下线性和二次规划可以视为最大似然方法，这一发现的实用价值也很小。这是因为应用最大似然的通常“常规条件”被违反了。特别是，由于 $y_i \leq f(x_i; \beta)$ ，随机变量 y 的范围依赖于待估计的参数。**因此，无法使用通常的定理来判断参数估计的渐近分布。**在这种情况下，我们对估计后的前沿了解多少并不清楚。

在另一篇论文中，Aigner、Amemiya和Poirier (1976) 提出了一种比单纯单侧更为合理的误差结构。具体而言，他们假设：

$$\varepsilon_i = \begin{cases} \varepsilon_i^* / \sqrt{1 - \theta} & \text{if } \varepsilon_i^* > 0, i = 1, \dots, N, \\ \varepsilon_i^* / \sqrt{\theta} & \text{if } \varepsilon_i^* \leq 0, \end{cases} \quad (3)$$

其中误差 ε_i^* 是独立正态分布的随机变量，对于 $0 < \theta < 1$ ，均值为0，方差为 σ^2 ；否则， ε_i^* 具有负的或正的截断正态分布，当 $\theta = 1$ 或 $\theta = 0$ 时，分别如此。

他们对这个误差结构的论证是，假定企业在给定一组‘输入’值的情况下在生产 y 方面的差异，根据（1）他们利用‘最佳实践’技术的能力的随机变化，一个误差源是单侧的（ $\varepsilon_i \leq 0$ ），以及（2）输入数量或 y 的测量误差，一个对称误差。参数 β 被解释为这两个误差源的‘相对变异性’的度量，其值界定了‘完全’前沿函数（ $\theta = 1$ ），‘平均’函数（ $\theta = \frac{1}{2}$ ）和一些有趣的中间情况。¹

这个误差结构对文献的主要贡献是，通过参数 θ 来估计拟合函数的位置，同时可以估计其他感兴趣参数。因此，之前对平均函数的批判和对使用生产前沿或包络函数作为“适当”行业生产函数的严格批评得到了缓解，这得益于这个更为灵活的误差规范。

然而，将 θ 解释为误差源相对变异性的度量只是在Aigner, Amemiya, Poirier的公式中隐含假设的，一个更直接的方法是明确地模拟由上述行为考虑所暗示的误差过程。

3. 随机前沿

现在我们回到方程(2)中给出的模型，但采用以下误差结构

$$\varepsilon_i = v_i + u_i, i = 1, \dots, N. \quad (4)$$

误差成分 v_i 代表对称扰动：假设 $\{v_i\}$ 是独立同分布的，服从 $N(0, \sigma_v^2)$ 。误差成分 u_i 假设与 v_i 独立分布，并满足 $u_i \leq 0$ 。我们特别关注 u_i 来源于在零以上截断的 $N(0, \sigma_u^2)$ 分布的情况。然而，其他单侧分布也是可行的，我们还将简要考虑 $-u_i$ 具有指数分布的情况。

当 $\sigma_v^2 = 0$ 时，这个模型简化为一个确定性前沿模型；当 $\sigma_u^2 = 0$ 时，它简化为Zellner, Kmenta, 和 Dreze (1966)的随机生产函数模型。注意 $y_i \leq f(x_i; \beta) + v_i$ ，所以前沿本身现在显然是随机的。

这种框架背后的经济逻辑是，生产过程受到两种经济上可区分的随机扰动，它们有不同的特征。我们相信，尽管我们的解释显然是新的，但文献中有充分的先例支持这种观点。² 并且从实用的角度来看，这种区分极大地促进了前沿的估计和解释。非正扰动 u_i 反映了每个公司的产出必须位于其前沿 $f(x_i; \beta) + v_i$ 上或以下。任何这样的偏差都是公司可控因素的结果，如技术和经济效率低下、生产者及其员工的意愿和努力，以及可能的因素，如缺陷和损坏的产品。但是，前沿本身可以在公司之间或对于同一公司随时间随机变化。在这种解释下，前沿是随机的，随机扰动 $v_i \geq 0$ 或 $v_i \leq 0$ 是由于有利及不利的外部事件造成的，例如运气、气候、地形和机器性能。对 y 的观察和测量误差构成了 $v_i \geq 0$ 或 $v_i \leq 0$ 的另一个来源。

这种方法的一个有趣副产品是，我们可以估计 v_i 和 u_i 的方差，从而获得关于它们相对大小的证据。这种方法的另一个含义是，生产效率应该原则上通过比率

$$\frac{y_i}{f(x_i; \beta) + v_i} \quad (5)$$

来衡量，而不是通过比率

$$\frac{y_i}{f(x_i; \beta)} \quad (6)$$

这只是将生产效率低下与公司控制之外的其他扰动源区分开。例如，因干旱或风暴而作物受损的农民，在我们的测量(5)上被认为是不幸的，但按通常的测量(6)是低效的。³

如果我们考虑一个线性模型，我们的估计讨论将被简化。因此，我们以明显的矩阵形式写作

$$y = X\beta + s, \quad (7)$$

代替(2)，其中现在 $s = v + u$ 。

随机前沿模型的估计

对称正态随机变量和截断正态随机变量之和的累积分布函数的显式表达式是由M.A. Weinstein (1964) 首次推导出来的。 ε 的密度函数的推导是直接的，所以我们在这里不展开。结果是

$$f(\varepsilon) = \frac{2}{\sigma} f^* \left(\frac{\varepsilon}{\sigma} \right) [1 - F^*(\varepsilon \lambda \sigma^{-1})], -\infty \leq \varepsilon \leq +\infty \quad (8)$$

其中 $\sigma^2 = \sigma_v^2 + \sigma_u^2$, $\lambda = \frac{\sigma_u}{\sigma_v}$, $f^*(\cdot)$ 和 $F^*(\cdot)$ 分别是标准正态的概率密度和累积分布函数。这个概率密度函数在0为中心是不对称的，其均值和方差由下式给出：

$$\begin{aligned} E(\varepsilon) &= E(u) = -\frac{\sqrt{2}\sigma_v}{\sqrt{\Pi}}, \\ V(\varepsilon) &= V(v) + V(u) = \left(\frac{\Pi - 2}{\Pi} \right) \sigma_u^2 + \sigma_v^2, \end{aligned} \quad (9)$$

可以从基本的考虑和 u 的矩的计算中轻松得出。

(8) 中的特定参数化方便地解释了 λ 作为区分不同公司的两种随机误差来源的相对变异性的指标。⁴ $\lambda^2 \rightarrow 0$ 意味着 $\sigma_v \rightarrow \infty$ 或 $\sigma_u \rightarrow 0$ ，即在对称误差在 ε 的确定中占主导地位，此时Eq.(8)则变为服从 $N(0, \sigma^2)$ 随机变量的概率密度。同样，当 $\sigma_u \rightarrow 0$ 时，单侧误差成为模型中随机变异的主要来源，并且(5)采取了负半正态的形式。⁵

估计问题是通过假设我们有 N 个观测值的随机样本，然后形成相关的对数似然函数，

$$\ln L(y|\beta, \lambda, \sigma^2) = N \ln \left(\frac{\sqrt{2}}{\sqrt{\Pi}} \right) + N \ln \sigma^{-1} + \sum_{i=1}^N \ln [1 - F^*(\varepsilon_i \lambda \sigma^{-1})] - \frac{1}{2\sigma^2} \sum_{i=1}^N \varepsilon_i^2, \quad (10)$$

这与Amemiya (1973年, 第1015页, 方程式(10.2))考虑的对数似然函数形式几乎完全相同。

求偏导后，我们有

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^N (y_i - \beta' x_i)^2 + \frac{\lambda}{2\sigma^3} \sum_{i=1}^N \frac{f_i^*}{(1 - F_i^*)} (y_i - \beta' x_i), \quad (11)$$

$$\frac{\partial \ln L}{\partial \lambda} = -\frac{1}{\sigma} \sum_{i=1}^N \frac{f_i^*}{(1 - F_i^*)} (y_i - \beta' x_i), \quad (12)$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{\sigma^2} \sum_{i=1}^N (y_i - \beta' x_i) x_i + \frac{\lambda}{\sigma} \sum_{i=1}^N \frac{f_i^*}{(1 - F_i^*)} x_i, \quad (13)$$

其中 x_i 是 X 中第 i 行的元素组成的 $k \times 1$ 向量， f_i^* 和 F_i^* 分别是在 $(y_i - \beta' x_i) \lambda \sigma^{-1}$ 处评估的标准正态分布的概率密度和累积分布函数。

给定(12)，我们知道在最优情况下有

$$\sum_{i=1}^N \frac{f_i^*}{(1 - F_i^*)} (y_i - \beta' x_i) = 0,$$

通过将这个结果插入(11)，最大似然估计量 β' 通过以下方式确定

$$-\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^N (y_i - \beta' x_i) = 0, \quad (14)$$

这导出了

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \beta' x_i)^2, \quad (15)$$

这是回归模型中残差方差的常用最大似然估计器的基础。但是， β 的确定并不独立于其他方程中的 β' 。

无论如何，这个结果可以用作迭代解决方案的基础。将 β' 预乘进(13)得到

$$\frac{1}{\sigma^2} \sum_{i=1}^N (y_i - \beta' x_i) \beta' x_i + \frac{\lambda}{\sigma} \sum_{i=1}^N \frac{f_i^*}{(1 - F_i^*)} \beta' x_i = 0. \quad (16)$$

将 $-\lambda$ 乘以(12)加到这个结果上并简化, 我们得到

$$\frac{1}{\sigma^2} \sum_{i=1}^N (y_i - \beta' x_i) \beta' x_i + \frac{\lambda}{\sigma} \sum_{i=1}^N \frac{f_i^*}{(1 - F_i^*)} y_i = 0, \quad (17)$$

这与(13)一起给出了一个与所谓的“Tobit”模型中遇到的一阶方程系统非常接近的 $k + 1$ 方程系统。⁶ 由于我们的密度函数在 ε 的范围内是连续的, 所以预计不会出现在Tobit模型中遇到的困难, 因此我们声称, 将(11), (12), 和(13)同时等于零的 β , λ 和 σ^2 的值具有一般的最大似然属性。这一主张的正式证明和对支持它的常规条件的检验遵循Amemiya (1973)的分析。

对于寻找 β , λ 和 σ^2 的优化值, 有多种解决算法可用。大多数这些算法 (例如Fletcher-Powell算法) 除了似然函数本身外, 还需要一阶或二阶导数以在计算机时间上以合理的成本达到最佳性能。由于这样的算法现在已经很容易获得,⁷ 我们不会花费空间讨论最大似然计算问题, 只需注意, 基于我们的经验, 这个似然函数似乎表现良好。二阶导数在本文的附录中给出, 用于那些用途以及计算最大似然估计的渐近标准误差的基础。

顺便提一下, 如果仅仅想估计 β , 除了对应于 X 中一列全为1的系数外, 其他的 β 通过最小二乘法估计是无偏且一致的。此外, 基于最小二乘结果, 可以利用等式(9)为 $V(\varepsilon)$ 关于 σ_u^2 和 σ_v^2 以及一个类似的关系对 e 的更高阶矩, 从而提取 σ^2 的组成部分 (即, 可以找到它们的一致估计器), 因为 $V(\varepsilon)$ 和更高阶的校正均值的 ε 矩本身可以从计算得到的最小二乘残差中一致估计。⁸

对于 u_i 的其他分布假设, 类似的评论和推导也适用。例如, 我们可以为 $-u_i$ 选择简单的单参数指数分布,

$$f(u) = \frac{1}{\phi} \exp(u/\phi), \quad u \leq 0, \quad (18)$$

其中 $\phi \geq 0$ 是 $-u_i$ 的均值。(方差是 ϕ^2 。) 通过一点代数操作, 我们可以发现 $\varepsilon_i = v_i + u_i$ 的分布由密度

$$f(\varepsilon) = \frac{1}{\phi} \left[1 - F^* \left(\frac{\varepsilon}{\sigma_v} + \frac{\sigma_v}{\phi} \right) \right] \exp \left(\frac{\varepsilon}{\phi} + \frac{\sigma_v^2}{2\phi^2} \right), \quad (19)$$

给出, 其中 $F^*(\cdot)$ 代表标准正态分布的累积分布函数。模型的似然函数立即随之而来。

1. 随着 $\theta \rightarrow 1$ ，正误差成分的方差变大（因此在似然函数中的影响较小），而负误差成分占主导。这导致了作为极限情形的“完全”前沿（ $\theta = 1$ ）。对于 $\theta \rightarrow 0$ 的情形，有类似的解释，尽管这种情况缺乏行为学解释。当 $\theta = \frac{1}{2}$ 时，似然函数形式为两个半正态分布的混合，每个都有相等的影响。 [↵](#)
2. Marschak 和 Andrews (1944) 建议 $v_i + u_i$ 的总和反映了生产者的“技术效率”和“意愿、努力和运气”。Zellner, Kmenta, 和 Dreze (1966) 建议它反映了“诸如天气、机器或劳动性能的不可预测变化等因素”，他们可能是第一批提出随机生产函数的人，尽管他们显然没有考虑到前沿。误差项的其他表述存在：Aigner 和 Chu (1968) 通过技术和经济效率低下以及生产过程中可能由于粗心处理和缺陷或损坏产出导致的纯随机冲击来解释它。Timmer (1971) 引用了技术和经济效率低下，以及“变量中的定义和测量问题”。而且，农业经济学家经常引用诸如气候、地形和土壤类型等环境条件在农场之间的变化，作为随机生产函数的指示。 [↵](#)
3. 如定义，(5) 是一个奇怪的效率测量，因为它是随机的，并且依赖于一个不可观察的 v_i 。这里只是为了支持论点而提供。 [↵](#)
4. 尽管 σ_u^2 不是 $((\pi - 2)/\pi)\sigma_u^2$ 的方差，我们也更倾向于使用这种对于 λ 的解释。另一个有用的 σ^2 参数化是使用 $\alpha = \sigma_u^2/\sigma^2$ 。 [↵](#)
5. 对于 $\sigma_v^2 = 0$ ，进而 $\lambda = \infty$ ，此时 (5) 变为 $f(\varepsilon) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\pi}\sigma_u} \exp(-\frac{1}{2\sigma_u^2}\varepsilon^2) & \text{for } \varepsilon_i \leq 0, \\ 0 & \text{otherwise.} \end{cases}$ [↵](#)
6. 见 Amemiya (1973, 1011 页，方程式 (7.2)，(7.3))。 [↵](#)
7. 对于几年前可用的算法的良好讨论，请参见 Goldfeld-Quandt (1971)。 [↵](#)
8. 例如， ε 的三阶矩是 $E[\varepsilon - E[\varepsilon]]^3 = \frac{2\sigma_u^3}{\sqrt{2\pi}}(1 - \frac{4}{\pi})$ 。在随后的蒙特卡罗结果中，没有尝试评估这些估计量的性质。 [↵](#)