

DOI: 10.19327/j.cnki.zuaxb.1007-9734.2018.01.010

随机边界模型: 进展及 Stata 应用

连玉君

(中山大学 岭南学院, 广东 广州 510275)

摘 要: 本文首先介绍了 SFA 的基本思想、估计方法、效率估算和分析方法, 进而分别介绍异质性 SFA、面板 SFA 以及双边 SFA、考虑内生性和门槛效应的 SFA 模型, 最后梳理出这些模型在 Stata 中的实现方法和应用中面临的一些问题。

关键词: 随机边界; 随机前沿; 面板随机边界; 双边随机边界; Stata 应用

中图分类号: F562

文献标识码: A

文章编号: 1007-9734(2018)01-0097-16

一、随机边界模型简介

随机边界模型(Stochastic Frontier Analysis, 后文简称 SFA) 由 Aigner 等(1977) 和 Meeusen 和 Broeck(1977) 提出, 目前已经在经济学和金融学等领域得到了广泛的应用。Kumbhakar 和 Lovell(2000), Murillo-Zamorano(2004), Coelli 等(2005), Fried 等(2008), Greene(2008), Kumbhakar 和 Tsionas(2011) 进行了非常细致的综述。Belotti 等(2013), Kumbhakar 等(2015) 详述了多种 SFA 模型在 Stata 中的实现方法。

(一) SFA 的基本思想

从理论上讲, 任何经济个体的“实际产出”都无法超出“产出边界”, 二者的偏离程度便可视作其效率损失或非效率。假设一个厂商在理想状况下(没有任何效率损失)的产出水平为 $f(z)$ ——给定投入要素 z 的最大产出, 亦称“理论产出”或“产出边界”。在现实中, 总是存在各种因素导致效率损失(如工人偷懒、管理不当等), 从而使实际产出 q 低于产出边界 $f(z)$, 即 $q < f(z)$ 。此时, 可以将产出效率定义如下:

$$TE(q, z) = \frac{q}{f(z)} \leq 1 \quad (1)$$

以上述思想为基础, 可以定义如下模型:

$$q_i = f(z_i, \beta) \cdot TE_i \quad (2)$$

收稿日期: 2017-12-08

基金项目: 国家自然科学基金项目(71672206)

作者简介: 连玉君, 男, 河南淮阳人, 副教授, 博士生导师, 研究方向为公司金融、金融计量。

其中, $0 < TE_i(q_i, z_i) \leq 1$ 为样本中第 i 家厂商的产出效率; β 为生产函数中的待估参数。在 SFA 中, 重点关注的是 TE_i , 包括效率的高低以及影响效率的因素, 而 β 则处于次要地位。

在上述设定中, $f(z_i, \beta)$ 是第 i 家厂商的“产出边界”, 但并非“随机边界”, 因为前者并未考虑随机因素的影响。例如, 对于两家各方面特征(包括技术效率)都相同的厂商, A 厂商的产出略低可能只是因为它的运气差一点(如遭遇雷电、CEO 车祸意外身亡等)。然而, 按照模型(2)的设定, 这种坏运气都会被归入 TE 之中。又如, 在实证分析过程中, 产出变量通常都会存在衡量偏误, 等式右侧的模型设定也可能存在偏误, 若这些随机偏误不能得到妥善的处理, 都有可能被归入 TE 。为此, 需要在模型中增加一个随机干扰项, 以便“吸收”这些厂商自身无法控制的随机因素, 使得 TE 的设定更加“干净”:

$$q_i = f(z_i, \beta) \cdot TE_i \cdot \exp(v_i) \quad (3)$$

这里, v_i 表示随机干扰, 假设其服从正态分布, 即 $v_i \sim N(0, \sigma_v^2)$ 。为了保证产出为正, 对其执行指数转换, 采用 $\exp(v_i)$ 的形式进入模型。由于 $\exp(0.01) \approx 1.01$, $\exp(-0.01) \approx 0.99$, 所以正面(负面)的干扰会使实际产出略微偏高(偏低)。换言之, 若将(3)式重新表述如下:

$$y_i = [f(z_i, \beta) \exp(v_i)] \cdot TE_i \quad (4)$$

则“ $[\]$ ”中的部分其实就是所谓的“随机边界”, 而所谓“随机”是指对产出边界的设定中考虑了随机因素的影响。换个角度来看, 与模型(2)相比, 模型(3)将影响效率的因素分成两类: 一类是由生产效率决定的(如管理能力、激励机制等); 另一类则是由那些我们无法控制的随机因素导致的。这种分离的好处在于能使我们对于产出效率的分析更加有针对性。

对(3)式两边同时取对数可得:

$$\ln(q_i) = \ln\{f(z_i, \beta)\} + v_i - u_i \quad (5)$$

其中, $u_i = -\ln(TE_i)$, 由于 $0 < TE_i \leq 1$, 所以 $u_i \geq 0$ 。也就是说, 由于 u_i 的存在, 厂商的实际产出 $\ln(q_i)$ 始终在边界下方。 u_i 通常也被称为“技术无效率”(technical inefficiency), 或简称为“无效率项”(inefficiency term), 这是因为 $u_i = -\ln(TE_i) \approx 1 - TE_i$, 而 TE 亦可表述如下:

$$TE_i = \exp(-u_i) \quad (6)$$

需要说明的是, 多数情况下, 我们会施加约束条件 $\text{cov}(u_i, v_i) = 0$, 即, 影响产出边界周围的那些随机干扰因素 v 与厂商的技术无效率 u 是彼此独立的。换言之, 我们假设生产过程中的好运气或坏运气完全是随机因素, 与技术无效率无关。

图1非常直观地呈现了随机边界模型的核心思想。厂商1在生产中遭遇了坏运气, 若使用 $[f(z_1, \beta) - y_1]$ 衡量其非效率产出, 显然会高估其非效率水平。同理, 对于厂商2, 不考虑其好运气带来的随机偏差, 则会导致我们低估其非效率水平。

(二) 模型设定和估计方法

假设有 k 个投入要素, 且产出函数 $f(z_i, \beta)$ 是对数线性化的(如 C-D 生产函数或超越对数生产函数), 则:

$$\ln(q_i) = \beta_0 + \sum_{j=1}^k \beta_j \ln(z_{ji}) + v_i - u_i \quad (7)$$

在实证分析过程中, SFA 模型通常采用如下设定形式:

$$y_i = x_i \beta + \varepsilon_i, \varepsilon_i = v_i - u_i \quad (8)$$

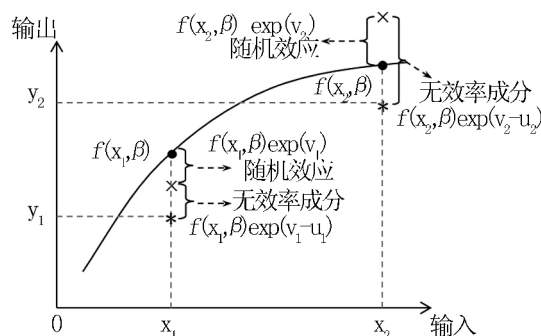


图1 随机边界模型释义 (Porcelli, 2009)

其中, $y_i = \ln(q_i)$, x_i 表示解释变量构成的向量, 其中第 j 个解释变量(要素投入)为: $x_{ji} = \ln(z_{ji})$, β 为相应的参数向量。 ε_i 称为“复合干扰项”, 由两部分组成: 常规意义上的干扰项 v 和非效率项 u 。因此, 我们亦可换个视角来看待 SFA 模型。相对于传统的线性回归模型($y_i = x_i'\beta + \varepsilon_i$), SFA 将干扰项分成了两个部分, 一部分具有对称分布, 而另一部分则具有单边分布($u_i \geq 0$)。同时, 假设 $\text{cov}(x_i, \varepsilon_i) = 0$, 即解释变量既不能与 v 相关, 也不能与非效率项 u 相关, 此时, 采用 OLS 估计模型(8)可以得到 β 的无偏估计。

然而, 前文已经提到, 对于 SFA 而言, 我们重点关注的并非 β , 而是 TE, 为了获得模型中的参数估计值, 需要对干扰项的分布形式进行设定。 v_i 的设定较为简单, 通常假设其服从正态分布, 即 $v_i \sim N(0, \sigma_v^2)$ 。由于 u_i 无效率项的取值必须大于零, 文献中多将其设定成单边分布(one-side distribution), 典型的分布形式包括: 指数分布($u_i \sim \text{Exp}(\sigma_u)$)、半正态分布($u_i \sim N^+(0, \sigma_u^2)$)、截断型半正态分布($u_i \sim N^+(\varpi, \sigma_u^2)$)。^①

以此为基础可以定义出如下三类典型的 SFA 模型:

(1) 正态-半正态模型(Normal/half-normal model):

$$y_i = x_i\beta + v_i - u_i, v_i \sim \text{iid } N(0, \sigma_v^2), u_i \sim \text{iid } N^+(0, \sigma_u^2) \quad (9)$$

其中, 半正态分布 $u_i \sim N^+(0, \sigma_u^2)$ 定义为 $u_i = |U_i|$, $U_i \sim N(0, \sigma_u^2)$ 。

(2) 正态-截断型半正态模型(Normal/truncated-normal model):

$$y_i = x_i\beta + v_i - u_i, v_i \sim \text{iid } N(0, \sigma_v^2), u_i \sim \text{iid } N^+(\varpi, \sigma_u^2) \quad (10)$$

其中, 截断型半正态分布 $u_i \sim N^+(\varpi, \sigma_u^2)$ 定义为 $u_i = |U_i|$, $U_i \sim N(\varpi, \sigma_u^2)$ 。使用这个模型的好处在于我们可以很方便地研究影响非效率项的因素。^②

(3) 正态-指数模型(Normal/exponential model):

$$y_i = x_i\beta + v_i - u_i, v_i \sim \text{iid } N(0, \sigma_v^2), u_i \sim \text{iid } \text{Exp}(\sigma_u) \quad (11)$$

其中, 指数分布 $u_i \sim \text{Exp}(\sigma_u)$ 的密度函数为 $f_u(u_i) = \theta \exp(-\theta u_i)$, $\theta = 1/\sigma_u$, $u_i \geq 0$ 。需要说明的是, u_i 的标准差为 $\sigma_u = 1/\theta$ 。

最后需要说明的是, 为了获得参数的无偏一致估计量, 在上述模型中都会附加约束条件, 即 u 和 v 不相关, 而二者与 x 亦不相关。无论对于上述何种设定, 虽然复合干扰项 ε_i 的分布都是非对称的, 但其密度函数却可以很容易地推导出来, 这使得上述模型都可以采用 MLE 进行估计。^③

(三) 效率\非效率的估计

根据(4),我们可以将效率定义如下:

$$TE_i = \frac{y_i}{y_i^*} = \frac{y_i}{f(x_i, \beta) \exp(v_i)} \quad (12)$$

其中, y_i^* 表示产出的随机边界。也就是说,技术效率等于实际产出与理论上的“随机产出边界”的比值。(6)式便是以此为基础得到的,即 $TE_i = \exp(-u_i)$ 。这个表达式中包含了不可观测的非效率项 u_i ,使之无法直接估计。即使已经获得了(8)中的各个参数的MLE估计值,也只能观察到复合干扰项($\varepsilon_i = v_i - u_i$)对应的残差。Jondrow等(1982)建议采用 u 的条件期望估计无效率项,即

$$E(u_i | \varepsilon_i) = \tilde{\mu}_i + \tilde{\sigma} \left\{ \frac{\phi(-\tilde{\mu}_i/\tilde{\sigma})}{\Phi(-\tilde{\mu}_i/\tilde{\sigma})} \right\} \quad (13)$$

Jondrow等(1982)认为可以直接用 $1 - E(u_i | \varepsilon_i)$ 估计效率。Battese和Coelli(1988)指出,对于正态-半正态分布模型而言,技术效率(TE_i)的最佳估计式为:^④

$$TE_i = E\{\exp(-u_i) | \varepsilon_i\} = \left\{ \frac{1 - \Phi(\tilde{\sigma} - \tilde{\mu}_i/\tilde{\sigma})}{1 - \Phi(-\tilde{\mu}_i/\tilde{\sigma})} \right\} \exp\left(-\tilde{\mu}_i + \frac{1}{2}\tilde{\sigma}^2\right) \quad (14)$$

其中, $\tilde{\mu}_i = -\varepsilon_i \sigma_u^2 / \sigma^2$, $\tilde{\sigma} = \sigma_u \sigma_v / \sigma$, $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$, $\Phi(\cdot)$ 表示标准正态分布的累积分布函数。一旦获得式中各个参数的MLE估计值,便可估计出TE。这是文献中使用最为广泛的效率计算方法。

当 u 服从指数分布或截断型半正态分布时,仍然可以用(14)式估计效率,但其中的参数会有所不同。具体而言,对于正态-指数分布模型而言, $\tilde{\mu}_i = -\varepsilon_i - \sigma_v^2 / \sigma_u$, $\tilde{\sigma} = \sigma_v$; 对于正态-截断型半正态分布模型而言, $\tilde{\mu}_i = (-\varepsilon_i \sigma_u^2 + \mu \sigma_v^2) / \sigma^2$, $\tilde{\sigma} = \sigma_u \sigma_v / \sigma$ 。

二、异质性随机边界模型

在传统的线性回归模型中,异方差的影响比较有限,此时估计量仍然是无偏且一致的,但缺乏有效性。然而,在SFA模型中异方差的影响要严重得多。在前文介绍的SFA模型中,复合干扰项通常都较为复杂。例如,在模型(15)中,复合干扰项中的常规的干扰项 v 和非效率项 u 都有可能存在异方差,从而影响系数统计推断的有效性。

考虑如下“正态-截断型半正态”SFA模型:

$$\begin{aligned} y_i &= x_i \beta + v_i - u_i, \\ v_i &\sim N(0, \sigma_{vi}^2), \\ u_i &\sim N^+(\varpi_i, \sigma_{ui}^2) \end{aligned} \quad (15)$$

细心的读者会注意到,相对于(8),此处的 v 和 u 对应的参数都附加了下标 i ,如,常规干扰项 v_i 的方差由 σ_v^2 设定为 σ_{vi}^2 ,也就是说公司 i 和公司 j 的干扰项具有不同的方差。显然,对于只有 N 个观察值的截面数据而言,我们无法估计出 N 个 σ_{vi}^2 。Kumbhakar等(1991)、Huang和Liu(1994)建议将 σ_{vi}^2 设定为一系列可能导致异方差的变量的函数,以便大幅减少待估参数的个数。一个直觉的想法是设定 $\sigma_{vi}^2 = s_i \lambda$,以便减少待估参数,但这无法保证 $\sigma_{vi}^2 \geq 0$,因此,文献中普遍使

用的是如下设定:^⑤

$$\sigma_{vi}^2 = \exp(s_i \lambda) \quad (16)$$

其中, s 是一系列可能导致异方差的变量, 例如公司规模、行业归属等。按照相似的思路, 我们也可以对 u 的方差, 即 σ_{ui}^2 做相似的设定:

$$\sigma_{ui}^2 = \exp(w_i \varphi) \quad (17)$$

事实上, 异质性设定的主要目的是研究非效率(平均值)及其不确定性(方差)的影响因素。核心参数是 ω_i , 它决定了偏离产出边界的位置(无效率水平的期望值), 文献中通常将其设定为一系列可能影响非效率水平的变量的线性函数(ω_i 的取值没有特定限制):

$$\omega_i = z_i \gamma \quad (18)$$

例如, 设定 $\omega_i = \gamma_0 + \gamma_1 z_{1i} + \gamma_2 z_{2i}$, 其中, z_{1i} 和 z_{2i} 是可能影响非效率的变量, 如公司规模、所有权性质等。^⑥

对于面板数据模型而言, 亦可采用上述设定方式。例如, Battese 和 Coelli (1995)、Habib 和 Ljungqvist (2005) 设定 $\gamma = 0, \varphi = 0$, 重点分析了影响效率水平的因素; Hadri (1999) 则设定 $\gamma = 0, \lambda = 0$, 重点研究了影响效率不确定性的因素; Wang (2003)、连玉君和苏治 (2009) 则仅设定 $\gamma = 0$, 采用了更为一般化的异质性设定。

多数研究都集中于对 ω_i 的异质性设定。Wang (2007) 应用 Battese 和 Coelli (1992, 1995) 模型研究了 30 个国家的 R&D 效率, 其中, ω_i 设定为国内 R&D 支出、PC 使用率以及经济自由度的线性函数。白俊红等 (2009) 采用相似的方法研究了 1998 ~ 2007 年中国 30 个省份研发创新效率。Kumbhakar 等 (2012) 应用异质性 SFA 研究了欧洲公司研发效率的影响因素。

有关 SFA 异质性设定的更详细的介绍可以参考 Greene (2008)、Kumbhakar 和 Lovell (2000) 以及 Kumbhakar 等 (2014)。

(一) 模型设定问题

在上述异质性设定中, w_i 和 z_i 中的变量可以有重合或者完全相同, 因为二者分别是影响效率水平和效率不确定性的因素。例如, Wang (2003) 在研究融资约束对台湾上市公司投资行为的影响时, 便设定 $w_i = z_i$, 包括现金流与资本存量的比值和总资产的自然对数两个变量。

然而, 对于 (w_i, z_i) 与 s_i 中的变量是否能重合则没有明确的判断依据。一方面, 在本文所提及的 SFA 模型中, 都需要假设 u 和 v 不相关以便能识别参数, 这意味着 (w_i, z_i) 与 s_i 中的变量要尽可能不重合; 另一方面, 由于 s_i 只是出现在 v 的方差中(二阶矩), 而非其均值部分(一阶矩), 所以即使 (w_i, z_i) 与 s_i 中的变量重合也不必然导致 u 和 v 相关。当然, 目前尚未有文献对此问题进行专门讨论。

按照相似的思路, 便会很自然地想到另一个更为重要的问题: 那些出现在产出边界函数中的变量能否出现在非效率项或普通干扰项的异质性设定中, 即, x 和 w, z, s 中的变量能否重合? 如果可以的话, 就相当于违反了干扰项与 x 不相关的原假设; 若不能的话, 又与经济常识相矛盾, 因为很多变量是同时影响产出和效率的。这是尚待解决的问题。

(二) 一步估计和两步估计

使用 SFA 模型的一个主要目的是研究影响效率的因素, 早期的文献主要使用两步法 (two

- stage approach), 而近期的文献则主要使用一步法。

两步法的第一步是估计 SFA 模型并得到 TE 的估计值, 在第二步中再以 TE 为被解释变量与理论上可能影响效率的因素进行回归分析。^⑦ 这种处理方式主要存在如下两个方面的问题: 其一, 研究假设前后矛盾。在第一阶段的分析中, 需要假设无效率项 u 是独立同分布的, 即 iid, 唯有如此才能使用 Jondrow 等(1982)提出的方法估算效率值 TE。然而, 在第二步中, TE 被设定为一系列公司特征变量的函数, 这意味着 TE 并不是独立同分布的, 这与第一步中的假设相矛盾(Coelli 等, 1998; Kumbhakar 和 Lovell, 2000; Greene, 2005b)。Wang 和 Schmidt(2002) 的 Monte Carlo 模拟分析表明, 相对而言, 一步估计法更为可靠。

三、面板随机边界模型

使用面板数据资料进行 SFA 分析有几个主要的优点: 其一, 由于拥有两期以上的资料, 我们可以控制那些不可观测的个体效应, 而采用截面数据则无法实现这一目标, 从而导致效率的估计有偏; 其二, 利用面板资料可以分析效率的时序变化特征, 这也是效率分析的一个重要研究主题, 而使用截面数据则只能分析个体之间的差异; 其三, 面板资料的样本量较大, 是我们可以获得更为有效的统计推断。

(一) 早期模型

前面介绍的基于截面数据的模型可以直接用在面板数据上, 类似于线性回归模型的 Pooled OLS, 但这并未充分利用面板数据提供的信息。考虑如下面板 - SFA 模型:

$$\begin{aligned} y_{it} &= y_{it}^* - u_{it}, \\ y_{it}^* &= \alpha + x_{it}'\beta + v_{it} \end{aligned} \quad (19)$$

假设 u 和 v 都是 iid 的, 且彼此不相关, $v_{it} \sim \text{iid } N(0, \sigma_v^2)$, $u_{it} \sim \text{iid } N^+(\varpi, \sigma_u^2)$ 。早期的面板 - SFA 模型将重点集中于无效率项 u 的设定上, 大体上可以分为两类。

1. 效率非时变模型

此类模型假设无效率项 u 不随时间变化, 即 $u_{it} = u_i$ 。例如, Pitt 和 Lee(1981) (简称 PL81) 将 u_i 视为随机变量, 提出了 RE - SFA 模型; 而 Schmidt 和 Sickles(1984) (简称 SS84) 则将 u_i 视为非随机变量 (即 $N-1$ 个虚拟变量), 提出了 FE - SFA 模型。无论是 FE - SFA 还是 RE - SFA, 本身都只能捕捉非效率的长期变化 (persistent component), 而无法反映随时间变化的非效率项。^⑧

2. 效率时变模型。

此类模型允许 u 随时间发生变化。通常是将 u_{it} 设定为 u_i 和一个反映时间变化特征的函数 $g(t)$ 的乘积, 即 $u_{it} = g(t) \cdot u_i$ 。不同模型之间的差别在于对 $g(t)$ 函数的设定形式不同, 如 Cornwell 等(1990) (简称 CSS90), Lee 和 Schmidt(1993) (简称 LS93), Kumbhakar(1990) (简称 Kumb90), 以及 Battese 和 Coelli(1992) (简称 BC92)。

例如, “CSS90 模型”设定如下:

$$\begin{aligned} y_{it} &= \alpha + x_{it}'\beta + v_{it} - u_{it}, i = 1, 2, \dots, N, t = 1, 2, \dots, T_i \\ \alpha_{it} &= \omega_i + \omega_{11}t + \omega_{12}t^2, \end{aligned} \quad (20)$$

其中, $\alpha_{it} = \alpha - u_{it}$ 。该模型的基本思想是把非效率项 u_{it} 设定为时间 t 的二次函数, 以便捕捉效率的时间变化特征。

“LS93 模型”则将 u_{it} 设定成如下形式, 以便减少待估参数:

$$u_{it} = g(t) \cdot u_i \quad (21)$$

其中, $g(t)$ 表示一组时间虚拟变量。这种模型设定方式仅需估计较少的参数, 但灵活性却降低了很多, 因为上述设定意味着在同一年度上所有个体的效率都相同。

“Kumb90 模型”只需两个参数, 但对非效率项的设定却非常灵活:

$$u_{it} = g(t) \cdot u_i = [1 + \exp(\gamma t + \delta t^2)]^{-1} \cdot u_i \quad (22)$$

其中, γ 决定了效率水平的高低, δ 决定了效率变化速度的快慢。

“BC92 模型”亦称为“时间衰减(time decay)”模型, 在文献中应用最为广泛。^⑨

$$u_{it} = g(t) \cdot u_i = \exp[-\eta(t - T_i)] \cdot u_i \quad (23)$$

其中, T_i 表示第 i 个个体中的最大时间长度。 η 称为延迟参数, 用以衡量非效率项随时间的下降程度。若 $\eta < 0$ 则表示随着时间的推移, 非效率程度在增强。由于在最后一期 $t = T_i$, 因此, 对于公司 i 而言, 最后一期的非效率程度即为该公司非效率程度的比较基准。

(二) TFE - SFA 模型

早期提出的所谓的面板 - SFA 模型有很明显的局限性——在产出边界函数 y_{it}^* 的设定中, 都没有考虑不可观测的个体效应(α_i), 这相当于把异质性的公司个体效应都归入了非效率项 u_{it} , 从而导致 TE 的估计值存在偏误。为了控制不同公司之间的差异, 可以在设定边界函数 y_{it}^* 时, 在 x_{it} 中加入不随时间变化的因素, 如公司的行业归属、经理人的性别、教育背景等。然而, 这仍然无法控制那些不可观测的个体效应, 如公司文化、经理人的个性等。此时, 可以将 (19) 式中的产出边界函数设定为:

$$y_{it}^* = \alpha_i + x_{it}'\beta + v_{it} \quad (24)$$

其中, α_i 表示不随时间变化且不可观测的个体效应。

该模型由 Greene(2005a) 正式提出, 为了区别于早期的 FE - SFA 模型, 他将该模型命名为“真正的固定效应模型”, 简称 TFE - SFA(True Fixed Effect SFA)。为了克服 TFE 模型存在的冗余参数问题^⑩, Wang 和 Ho(2010) 将非效率项 u_{it} 设定为 u_i 和一个公司特征变量函数的乘积, 称为缩放因子模型, 该模型的优点在于可以事先通过组内去心或一阶差分去除个体效应, 进而使用 MMLE(Marginal MLE) 进行估计。^⑪Chen 等(2014), Belotti 和 Ilardi(2017) 则在更为宽松的假设条件下提出了“一阶差分一致估计量”。但是由于差分后的边际极大似然函数没有解析解, 需要使用蒙特卡罗模拟进行参数估计, 称之为模拟边际极大似然估计法(MMSLE 方法), 缺点是计算所耗时间较长。^⑫

四、其他模型

(一) 双边随机边界模型

在传统的 SFA 模型中, 实际产出始终是小于理论产出的。但在有些情况下, 我们关注的 y

变量可以高于,也可能低于其理论上的最优值 y^* 。如一个人的体重、公司的投资支出、政府的财政支出等。以公司投资为例, $y > y^*$ 称为过度投资,反之则为投资不足。此时可以使用双边随机边界模型(two-tier SFA)来估算投资不足率和过度投资率。该模型由 Kumbhakar 和 Christopher(2009) 提出^⑬,并将应用于研究工资议价问题。模型设定如下:

$$\begin{aligned} y_i &= x_i \beta + \varepsilon_i \\ \varepsilon_i &= v_i + w_i - u_i \end{aligned} \quad (25)$$

复合干扰项 ε_i 由三部分组成: v_i 为常规干扰项, $v_i \sim i. i. d. N(0, \sigma_v^2)$; w_i 和 u_i 都具有单边分布特征,假设其服从指数分布, $w_i \sim i. i. d. \text{Exp}(\sigma_w, \sigma_w^2)$, $u_i \sim i. i. d. \text{Exp}(\sigma_u, \sigma_u^2)$ 。同时,假设三者彼此不相关,且它们与 x 也不相关。使用 MLE 得到参数估计值后,可以进一步计算非效率指标:投资不足率 $\text{NTEu} = E(1 - e^{-u_i} | \varepsilon_i)$, 过度投资率 $\text{NTEw} = E(1 - e^{-w_i} | \varepsilon_i)$ 和净效率损失 $\text{NS} = \text{NTEu} - \text{NTEw}_i$ 。

双边随机边界模型在诸多领域得到了应用: Lian 和 Chung(2008) 用该模型研究了中国上市公司的投资效率; 卢洪友等(2011) 用该模型估算了中国医疗市场中的信息不对称程度; 刘海洋等(2013) 运用 2006 年海关进出口交易数据,估算了中国国有企业的国际议价能力。

(二) 考虑内生性和门槛特征的 SFA 模型

在多数面板 SFA 模型中,都假设样本中的所有公司具有相同的生产函数。然而,个体异质性(包括技术水平、生产模式、经理人特征等),以及外部环境(如监管制度、法律保护、文化等)的差异都会导致这一假设无法满足,从而导致效率估计值有偏。前面介绍的 TFE, TRE, G - TRE 以及对应的异质性设定模型都在尝试刻画上述公司内部和外部的异质性。

一个较为常用的方法是根据一些先验的指标(如公司规模、所有权性质、行业归属等)把样本公司分成两组或多组,分别针对每个子样本组估计 SFA 模型,并对比组间系数或效率差异。这是处理组间异质性的一个典型做法。然而这种做法的局限性也很明显。其一,分组界点的确定往往是根据经验外生确定的。^⑭其二,子样本的特征可能与分组变量密切相关,这意味着分组变量本身可能是内生的,与最初的外生性假设相矛盾(Lai, 2013)。

针对第一个问题,文献中提出了马尔科夫状态转换 SFA 模型(Tsionas 和 Kumbhakar, 2004)、门槛 SFA 模型(Tsionas 等, 2017; Yéou 等, 2010; Wang 和 Huang, 2009)。^⑮对于第二个问题, Greene(2010)、Kumbhakar 等(2009)以及 Horrace 等(2016)提出了考虑样本选择偏误的 SFA 模型。^⑯Lai(2013)则进一步将上述两种特征都纳入同一个模型中,提出了考虑自我选择偏误的面板门槛 SFA 模型,模型的设定思路非常类似于 Hansen(1999, 2000),但进一步考虑样本选择偏误。

一般化的内生性问题在 SFA 文献中也逐渐得到重视,例如, Kutlu(2010)将文献中广泛应用的 Battese 和 Coelli(1992)模型扩展为允许包含内生变量的情形,并提供了基于 MLE 的估计方法。Tran 和 Tsionas(2013)则进一步给出了基于 GMM 的估计方法。

在最近的研究中, Griffiths 和 Hajargasht(2016)基于贝叶斯分析,提供了一种较为一般化的处理 SFA 中内生性的方法。该模型设定中,允许无效率项和随机干扰项与解释变量相关。主要的想

法是通过取对数转换将无效率项转换为对数正态分布,进而将其期望值设定为解释变量的函数。

Karakaplan(2017)提供了 **sfkk** 命令,用于估计包含内生性问题的 SFA 模型。文中提供了多个范例来说明该命令的使用方法。

(三) 零无效 SFA 模型

Kumbhakar 等(2013)提出了一个新的模型:零无效 SFA 模型(zero inefficiency SFA model,简称 ZISF)。这一模型的基本思想类似于零膨胀泊松模型(ZIP, zero inflated poisson model)。传统的 SFA 模型假设样本中的所有公司都存在不同程度的效率损失,然而,在 Kumbhakar 等(2013)的模型中,则假设一部分公司是“零效率损失(zero inefficiency)”的,而另一部分公司则存在非零的效率损失。利用该模型可以检验样本是否存在零无效公司。在传统 SFA 模型中,由于设定非效率项服从半正态分布或指数分布,相当于先验地假设样本公司都存在一定程度的效率损失,因为在这些分布函数设定下,无效率项为零的概率为 0。ZISF 是介于 OLS 和 SFA 的一般化模型。因为,如果样本中的所有公司都是完全有效率的,则 ZISF 便转变成传统的 OLS 回归;若所有公司都存在一定程度的效率损失,则 ZISF 便转变成传统的 SFA 模型。

除了上述模型,文献中还提出了 SUR-SFA 模型(Lai 和 Huang, 2011)、考虑衡量偏误的 SFA 模型(Chang 等, 2012)。

五、随机边界分析的 Stata 实现

(一) Stata 中的 SFA 命令

Stata 中提供了一系列命令,可以很方便地进行 SFA 估计和分析。包括官方命令:**frontier**, **xtfrontier**^⑦, 外部命令:**sfcross**, **sfpanel**, **sftfe**, **sfkk**, **SFA2tier** 等。

外部命令可以使用 **ssc install cmdname, replace** 命令安装最新版。^⑧或使用 **findit cmdname** 搜索,按照 Stata 的提示下载,这种方式的好处是可以同时下载作者提供的范例数据 Do 文档。

sfcross 和 **sfpanel** 命令可以完全实现 Stata 官方命令 **frontier**, **xtfrontier** 的功能,能够估算的模型和效率的估算方法也更为丰富。Belotti 等(2013)对这两个命令的选项和具体使用方法做了非常细致的介绍。

sftfe 用于估计 Chen 等(2014), Belotti 和 Ilardi(2017)新近提出的“一阶差分一致估计量”,其帮助文件提供了非常详细的 Stata 范例。而 **sfkk** 则用于估计 Karakaplan 和 Kutlu(2017)提出的内生性 SFA 模型,具体语法格式和范例说明参见 Karakaplan(2017)。

SFA2tier 用于估计第 3 节介绍的双边随机边界模型,由笔者编写,相关应用参见卢洪友等(2011)。^⑨

下面通过几个简单的例子进行说明。对于截面数据,可以使用 **sfcorss** 命令实现本文介绍的所有模型;对于面板数据,主要使用 **sfpanel** 和 **sftfe**,虽然原理较为复杂,但操作上与 **sfcross** 命令无异。我们以 Stata 手册中的范例数据 **frontier1.dta** 为例进行说明。

```

L1 .webuse frontier1.dta, clear
L2 .global y "lnoutput" //y
L3 .global x "lnlabor lncapital" //x1 x2
L4 .sfcross $y $x, distribution(hnormal) //Eq.(9)
L5 * -效率估计1:Jondrow et al.(1982, JLMS1982)
L6 .predict te_jlms, jlms //TE = 1 - E(u|e), Eq.(13)
L7 * -效率估计2:Battese and Coelli(1988, BC88)
L8 .predict te_BC88, bc //TE = E[exp(-u|e)], Eq.(14)
L9 * -超越对数生产函数
L10 .sfcross $y c.($x)##c.($x) //Eq.(26)
L11 * -异质性 SFA Eq.(15) - (18)
L12 .sfcross $y $x, d(tn) emean(size) vsigma(lncapital) usigma(ln-
labor)

```

第 L4 行的命令用于估计(9)式中的“正态 - 半正态”SFA 模型。其中,非效率项的分布函数是由选项 `distribution(hnormal)` 来确定的,可以简写为 `d(hn)`。若将其改为 `d(tn)` 或 `d(e)`,则分别对应模型(10)和(11)。完成估计后,可以使用 `predict` 命令获得效率估计值,参见 L6 和 L8。若将使用超越对数生产函数(26),则可用 Stata 中的因子变量语法(`help fvvarlist`)进行设定,写法见 L10。^②

在 L12 中,我们设定了一个一般化的异质性 SFA 模型,对应(15) - (18)式。选项 `emean(size)` 用于设定(18)式中的 z 变量,即 $\varpi_i = \gamma_0 + \gamma_1 \text{size}_i$ 。^② `vsigma()` 和 `usigma()` 选项用于设定(16)和(17)式中的 s 和 w 变量。至于这三个选项该如何组合使用,请参阅第2节。

(二)实操中需要考虑的问题

1. 模型设定问题

在多数 SFA 模型设定中,被解释变量和解释变量都应采用对数形式(Stata15 Manual[R])。这是因为,虽然模型(7)是文献中惯用的模型设定形式,但它是模型(3)的对数线性化。^②唯有如此,才能保证(6)式定义的效率表达式具有正确的经济含义。

诚如 Greene(2008)所言,在 SFA 文献中,函数形式的选择通常被视为技术层面的问题,并未得到太多关注。以文献中广泛使用的 Cobb - Douglas 生产函数为例,虽然经过对数转换后该模型具有很好的线性形式,但也隐含着非常严格的假设条件,如在给定要素价格的前提下,要素的需求弹性(demand elasticities)和要素份额(factor shares)均为常数,且要素的替代弹性均为 -1。为了放松这些假设条件,我们可以使用更为灵活的“超越对数生产函数”,通过在模型中加入二次项来捕捉要素的交互影响和潜在的非线性关系(Kumbhakar, 1989)^③:

$$\ln y = \alpha + \sum_{k=1}^K \beta_k \ln x_k + \frac{1}{2} \sum_{k=1}^K \sum_{m=1}^K \gamma_{km} \ln x_k \ln x_m \quad (26)$$

2. 分布函数的选择

对于 SFA 分析而言,由于非效率项通常都被视为干扰项的一部分,因此非效率项的分布函数的设定虽然会在一定程度上影响效率的估计值,但起决定性作用的仍然是对模型主体部分的设定(即对产出函数的设定),这需要非常深入的理论分析,并结合前期文献选择模型的形式

和核心变量的衡量方法(Greene, 2005a)。在实证分析过程中,将无效率项 u_i 设定为半正态分布或指数分布并不会产生太大的影响(Battese 和 Coelli, 1992)。

3. 不收敛问题

在 Stata 中, SFA 模型是通过极大化对数似然函数得到参数估计值的, 主要采用牛顿-莱布尼兹方法进行数值求解(Stata15 Manual[R])。虽然 Stata 的数值求解功能已经非常强大, 使用户基本不用自行设定初始值和步长, 但不收敛也是家常便饭。遇到这种情况, 要从如下几个方面入手。第一, 模型设定。模型的形式越复杂, 涉及的参数也就越多, 也就越不容易收敛。因此, 在保证理论分析合理的前提下, 可以先从简单的模型设定入手, 在保证收敛的前提下, 再逐步提高模型的复杂度。第二, 查验关键变量中是否包含了严重的离群值, 若有则可以考虑使用对数转换或 winsor 缩尾处理。第三, 变量的精度。在执行 SFA 估计前, 要把所有变量都设定为双精度型。^④

最后, 需要特别说明的是, 所谓“收敛”事实上有一定的主观性。Stata 中的收敛条件是: 两次拟合的对数似然函数值之差小等于 10^{-7} , 或两次拟合的参数向量之差小于 10^{-6} 。如果上述两个条件都无法得到满足, 则 Stata 在默认迭代 16000 次后即呈现估计结果, 同时提示用户尚未达到收敛条件。因此, 在初步测试阶段, 我们可以先设定较为宽松的收敛标准, 以便初步确认模型设定和程序无误。这可以使用 `tolerance(#)` 或 `ltolerance(#)` 选项来设定。^⑤

4. 估计过程中的提示信息

MLE 拟合过程中, 在有些区间内, 对数似然函数的二阶偏导是一个非常平坦的函数, Stata 便会提示“Not Concave”信息, 形如:

```
Iteration #: log likelihood = ... (not concave)
```

有时 Stata 会改变搜索的步长, 以调整搜索精度, 屏幕上则会提示:

```
Iteration #: log likelihood = ... (backed up)
```

除非上述信息在迭代的最后一步出现, 通常不用在意它。^⑥

六、简要评述

有关 SFA 模型的相关研究还在不断深入。主要集中于如下几个方面: 其一, 内生性问题的研究尚处于起步阶段; 其二, 新近提出的动态 SFA 模型在分析非效率的长期和短期效应方面非常有用, 但相关的应用研究还非常有限; 其三, 鉴于生产单元之间往往存在较强的互动行为, 在 SFA 框架下纳入截面相关或空间相关就显得尤为必要。

就实际应用而言, Stata 提供的 `sfcross`, `sfpanel`, `sftfe`, `sfkk`, `SFA2tier` 命令能够估计文献中使用最为广泛的 SFA 模型, 但还无法很好地支持内生性-SFA, 门限-SFA, 动态-SFA 和空间-SFA 的估计。有如下几点建议: 其一, 虽然使用 SFA 的初衷是估计效率及其影响因素, 但基本模型的设定(即产出边界函数的设定)仍然是非常重要的, 这需要做扎实的理论分析, 在不遗漏关键变量的前提下尽可能使用精简的模型。其二, 不可观测的个体效应到底归入个体的异质性(产出边界函数部分)还是无效率项(干扰项部分)是目前文献中颇具争议的问题(Greene, 2005a), 诚如 Greene 在多篇论文中强调的, 或许二者无法进行严格的区分; 其三, 在第 2 小节中介绍的异质性设定方法同样适用于面板 SFA 模型, 相对于两步法, 采用这种一步法分析效率的影响因素更为合理。

注 释:

- ①也可以将 u_i 设定为形式上更为灵活的伽马分布,详情可参见 Greene(2008)。我们可以借助 Belotti 等(2013)编写的 `sfcross` 命令估计“正态-伽马”SFA 模型。
- ②详情参见第2小节有关异质性 SFA 模型的介绍。
- ③详情参见 Greene(2008)以及 Stata15 手册[R]。对于有兴趣自行编写程序的读者而言,可以采用 Stata 提供的 `ml` 命令定义似然函数并执行估计。Stata 官方命令 `frontier` 对应的 `ado` 文档(`frontier.ado`),以及相应的似然函数文档(`fron_hn.ado`)都存储于 stata 安装目录下 `ado\base\l` 文件夹下。Gould 等(2010)非常细致地介绍了如何在 Stata 中执行 MLE 估计。
- ④一个令人遗憾的事实是,这个应用最为广泛的 TE 估计式并不是一致估计量,因为 $(u_i | \varepsilon_i)$ 的方差独立于 i ,意味着其方差不会随着样本数的增加而减小。
- ⑤有些文献会采用另一种等价的表述方式: $\ln(\sigma_{vi}^2) = z_{vi}\gamma$ 。
- ⑥相关应用参见 Habib 和 Ljungqvist(2005)、Wang(2003)、连玉君和苏治(2009)。
- ⑦Koutsomanoli - Filippaki 和 Mamatzakis(2010)采用两步法估计了效率的调整速度。第一步采用传统的 SFA 模型估计出 TE 值,第二步则以 TE 为被解释变量,采用动态面板模型估计调整速度。
- ⑧另需说明的是,这里所谓的“个体效应”指的是不随时间变化的非效率项,而非产出函数中的个体效应。因此,严格而言,这里的 FE 或 RE 并不是面板数据文献中所言的真正意义上的 FE 或 RE。
- ⑨Kumbhakar 和 Wang(2005)采用该模型研究了技术进步的追赶效应。Kumbhakar 和 Peresetsky(2013)综合应用了多种 Panel - SFA 模型分析了俄罗斯银行业的效率。
- ⑩由于每个公司中包含的样本区间(T)都比较短, σ_i 的方差并不会随着 N 的增大而降低。
- ⑪Chen 等(2011)应用缩放因子模型研究了融资约束对资产定价的影响;林伯强和杜克锐(2013)应用该模型研究了中国的能源效率问题。
- ⑫孙焱林和温湖炜(2017)使用该方法估算了我国的产能过剩率。
- ⑬这个模型的思想源泉来自于 Polachek 和 Yoon(1996),前面介绍的 TFE 和 TRE 也将根源归于此模型。
- ⑭例如,很多文献根据公司总资产对数的中位数或第 33th 和 66th 分位数来划分大公司和小公司。
- ⑮Mastromarco 等(2012)应用门槛 SFA 模型研究了全球化是否能提升效率。
- ⑯Su 和 He(2012)应用 Greene(2010)模型研究了中国上市公司所有权结构与产出效率。
- ⑰Stata15 手册提供了非常细致的范例和解释。
- ⑱`cmdname` 表示命令名称,如执行 `ssc install sfpanel, replace`,即可下载 `sfpanel` 程序包。
- ⑲`SFA2tier` 命令发布于 Stata 学术论文视频教程,参见 <http://baoming.pinggu.org/Default.aspx?id=134>。
- ⑳`c.x` 表示将 x 按连续变量处理;`c.x##c.y` 表示在模型中加入 x, y ,以及二者的平方项和交乘项,换言之,`c.x##c.y` 等价于 $x \ y \ x^2 \ y^2 \ x * y$ 。

- ⑳ `emmean()` 选项中可以放置多个变量。本例中, `size` 变量的系数为 -0.002 , 表明大规模公司的效率损失相对较低。
- ㉑ 当然, 有些情况下, 产出函数并不必然是 C-D 生产函数的形式, 上述要求也就不再那么严格了。例如, Habib 和 Ljungqvist (2005) 在使用 SFA 研究上市公司的管理层薪酬激励与公司价值关系时, 采用 Tobin's Q (没有取对数) 作为被解释变量用以衡量公司价值, 解释变量则包含了各种财务指标, 如负债率、投资支出比率, 此类变量都未进行对数转换, 原因是多数财务比率中都会包含小于零的数值。
- ㉒ Altunbas 等 (2000) 使用 cost-SFA 估算了日本银行的效率和风险, 使用了高阶项, 非常灵活的模型设定。Wang (2007) 在研究 30 个国家 R&D 效率时, 也采用了高阶项的方式进行模型设定。王德祥和李建军 (2009) 基于超越对数生产函数估算了我国的税收流失率。
- ㉓ 生成新变量时, Stata 默认将其存储为浮点型。浮点型仅精确到小数点后 7 位, 而双精度型可以精确到小数点后 16 位。可以使用 `recast` 命令改变变量的存储精度, 亦可在生成新变量时指定其存储精度, 例如 `gen double lnx = ln(x)`。参见 Gould 等 (2010)。
- ㉔ 输入 `help maximize##tolerance` 可以参看完整帮助文件。
- ㉕ 对于这两条信息的详细解释, 参见 Gould 等 (2010)。

参考文献:

- [1] Aigner D, Lovell C A, Schmidt P. Formulation and Estimation of Stochastic Frontier Production Function Models [J]. *Journal of Econometrics*, 1977, 6(1): 21-37.
- [2] Meeusen W, Broeck J V D. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error [J]. *International Economic Review*, 1977, 18(2): 435-444.
- [3] Kumbhakar S C, Lovell C. *Stochastic Frontier Analysis* [M]. Cambridge: Cambridge University Press, 2000.
- [4] Murillo-Zamorano L R. Economic Efficiency and Frontier Techniques [J]. *Journal of Economic Surveys*, 2004, 18(1): 33-77.
- [5] Coelli T J, Rao D S P, O'Donnell C J, Battese G E. *An Introduction to Efficiency and Productivity Analysis* [M]. New York: Springer, 2005.
- [6] Fried H O, Lovell C A K, Schmidt S S. *The Measurement of Productive Efficiency and Productivity Change* [M]. New York: Oxford University Press, 2008.
- [7] Greene W H. The Econometric Approach to Efficiency Analysis [J]. *Measurement of Productive*, 2008, 1(5): 92-251.
- [8] Kumbhakar S C, Tsionas E G. Some Recent Developments in Efficiency Measurement in Stochastic Frontier Models [J]. *Journal of Probability and Statistics*, 2011, forthcoming.
- [9] Belotti F, Daidone S, Ilardi G, Atella V. Stochastic Frontier Analysis Using Stata [J]. *Stata Journal*, 2013, 13(4): 719-758.
- [10] Kumbhakar S C, Wang H J, Horncastle A P. *A Practitioner's Guide to Stochastic Frontier Analysis Using Stata* [J]. Cambridge University Press, 2015.
- [11] Porcelli F. Measurement of Technical Efficiency. A Brief Survey on Parametric and Non-Parametric Techniques [J]. Working Paper, 2009.
- [12] Gould W, Pitblado J, Poi B. *Maximum Likelihood Estimation with Stata* [M]. Stata Press, 2010.
- [13] Jondrow J, Lovell K, Materov I, Schmidt P. On the Estimation of Technical Inefficiency in the Stochastic Fron-

- tier Production Function Model[J]. *Journal of Econometrics*, 1982, 19(2-3): 233-238.
- [14] Battese G E, Coelli T J. Prediction of Firm-level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data[J]. *Journal of Econometrics*, 1988, 38(3): 387-399.
- [15] Kumbhakar S C, Ghosh S, McGuckin J T. A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U. S. Dairy Farms[J]. *Journal of Business & Economic Statistics*, 1991, 9(3): 279-286.
- [16] Huang C J, Liu J T. Estimation of a Non-Neutral Stochastic Frontier Production Function[J]. *Journal of Productivity Analysis*, 1994, 5(2): 171-180.
- [17] Habib M, Ljungqvist A. Firm Value and Managerial Incentives: A Stochastic Frontier Approach[J]. *Journal of Business*, 2005, 78(6): 2053-2094.
- [18] Wang H J. A Stochastic Frontier Analysis of Financing Constraints on Investment: The Case of Financial Liberalization in Taiwan[J]. *Journal of Business and Economic Statistics*, 2003, 21(3): 406-419.
- [19] 连玉君, 苏治. 融资约束、不确定性与上市公司投资效率[J]. *管理评论*, 2009, 21(1): 19-26.
- [20] Battese G E, Coelli T J. A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data[J]. *Empirical Economics*, 1995, 20(2): 325-332.
- [21] Hadri K. Estimation of a Doubly Heteroscedastic Stochastic Frontier Cost Function[J]. *Journal of Business & Economic Statistics*, 1999, 17(3): 359-363.
- [22] Wang E C. R&D Efficiency and Economic Performance: A Cross-Country Analysis Using the Stochastic Frontier Approach[J]. *Journal of Policy Modeling*, 2007, 29(2): 345-360.
- [23] Battese G E, Coelli T J. Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India[J]. *Journal of Productivity Analysis*, 1992, 3(1): 153-169.
- [24] 白俊红, 江可申, 李婧. 应用随机前沿模型评测中国区域研发创新效率[J]. *管理世界*, 2009, (10): 51-61.
- [25] Kumbhakar S C, Ortega-Argilés R, Potters L, Vivarelli M, Voigt P. Corporate R&D and Firm Efficiency: Evidence from Europe's Top R&D Investors[J]. *Journal of Productivity Analysis*, 2012, 37(2): 125-140.
- [26] Kumbhakar S C, Lien G, Hardaker J B. Technical Efficiency in Competing Panel Data Models: A Study of Norwegian Grain Farming[J]. *Journal of Productivity Analysis*, 2014, 41(2): 1-17.
- [27] Koutsomanoli-Filippaki A, Mamatzakis E C. Estimating the Speed of Adjustment of European Banking Efficiency under a Quadratic Loss Function[J]. *Economic Modelling*, 2010, 27(1): 1-11.
- [28] Coelli T J, Rao D S P, O'Donnell C J, Battese G E. *An Introduction to Efficiency and Productivity Analysis* [M]. Boston: Kluwer Academic Publishers, 1998.
- [29] Greene W. Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model[J]. *Journal of Econometrics*, 2005b, 126(2): 269-303.
- [30] Wang H J, Schmidt P. One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels[J]. *Journal of Productivity Analysis*, 2002, 18(2): 129-144.
- [31] Pitt M M, Lee L F. The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry[J]. *Journal of Development Economics*, 1981, 9(1): 43-64.
- [32] Schmidt P, Sickles R. Production Frontiers and Panel Data[J]. *Journal of Business & Economic Statistics*, 1984, 2(4): 367-374.
- [33] Cornwell C, Schmidt P, Sickles R C. Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels[J]. *Journal of Econometrics*, 1990, 46(1-2): 185-200.

- [34] Lee Y H, Schmidt P. A Production Frontier Model with Flexible Temporal Variation in Technical Inefficiency [M]. Measurement of Productive Efficiency: Techniques and Applications. 1993.
- [35] Kumbhakar S C. Production Frontiers, Panel Data, and Time - Varying Technical Inefficiency[J]. Journal of Econometrics, 1990, 46(1): 201 - 211.
- [36] Kumbhakar S C, Wang H J. Estimation of Growth Convergence Using a Stochastic Production Frontier Approach [J]. Economics Letters, 2005, 88(3): 300 - 305.
- [37] Kumbhakar SC, Peresetsky A. Cost Efficiency of Kazakhstan and Russian Banks: Results from Competing Panel Data Models[J]. Macroeconomics and Finance in Emerging Market Economies, 2013, 6(1): 88 - 113.
- [38] Greene W. Fixed and Random Effects in Stochastic Frontier Models[J]. Journal of Productivity Analysis, 2005a, 23(1): 7 - 32.
- [39] Wang H J, Ho C W. Estimating Fixed - Effect Panel Stochastic Frontier Models by Model Transformation[J]. Journal of Econometrics, 2010, 157(2): 286 - 296.
- [40] Chen N K, Chen Y Y, Wang H J. Asset Prices and Capital Investment - a Panel Stochastic Frontier Approach [J]. Working Paper, 2011.
- [41] 林伯强, 杜克锐. 要素市场扭曲对能源效率的影响[J]. 经济研究, 2013(9): 125 - 136.
- [42] Chen Y Y, Schmidt P, Wang H J. Consistent Estimation of the Fixed Effects Stochastic Frontier Model[J]. Journal of Econometrics, 2014, 181(2): 65 - 76.
- [43] Belotti F, Ilardi G. Consistent Inference in Fixed - Effects Stochastic Frontier Models[J]. Journal of Econometrics, 2017, forthcoming.
- [44] 孙焱林, 温湖炜. 我国制造业产能过剩问题研究[J]. 统计研究, 2017, 34(3): 76 - 83.
- [45] Kumbhakar S C, Christopher F. The Effects of Bargaining on Market Outcomes: Evidence from Buyer and Seller Specific Estimates[J]. Journal of Productivity Analysis, 2009, 31(1): 1 - 14.
- [46] Polachek S W, Yoon B J. Panel Estimates of a Two - Tiered Earnings Frontier[J]. Journal of Applied Econometrics, 1996, 11(2): 169 - 178.
- [47] Lian Y, Chung, C F. Are Chinese Listed Firms Over - Investing? [J]. SSRN working paper, 2008, Available at SSRN: <http://ssrn.com/abstract=1296462>.
- [48] 卢洪友, 连玉君, 卢盛峰. 中国医疗服务市场中的信息不对称程度测算[J]. 经济研究, 2011(4): 94 - 106.
- [49] 刘海洋, 逯宇铎, 陈德湖. 中国国有企业的国际议价能力估算[J]. 统计研究, 2013, 30(5): 47 - 53.
- [50] Lai H P. Estimation of the Threshold Stochastic Frontier Model in the Presence of an Endogenous Sample Split Variable[J]. Journal of Productivity Analysis, 2013, 40(2): 227 - 237.
- [51] Tsionas E G, Kumbhakar S C. Markov Switching Stochastic Frontier Model[J]. Econometrics Journal, 2004, 7(2): 398 - 425.
- [52] Tsionas E, Tran K, Michaelides P. Bayesian Inference in Threshold Stochastic Frontier Models[J]. Working Paper, 2017.
- [53] Yélou C, Larue B, Tran K C. Threshold Effects in Panel Data Stochastic Frontier Models of Dairy Production in Canada[J]. Economic Modelling, 2010, 27(3): 641 - 647.
- [54] Wang M H, Huang T H. Threshold Effects of Financial Status on the Cost Frontiers of Financial Institutions in Nondynamic Panels[J]. Applied Economics, 2009, 41(26): 3389 - 3401.
- [55] Mastromarco C, Serlenga L, Shin Y. Is Globalization Driving Efficiency? A Threshold Stochastic Frontier Panel Data Modeling Approach[J]. Review of International Economics, 2012, 20(3): 563 - 579.

- [56] Greene W. A Stochastic Frontier Model with Correction for Sample Selection[J]. *Journal of Productivity Analysis*, 2010, 34(1): 15–24.
- [57] Kumbhakar S C, Tsionas E G, Sipilainen T. Joint Estimation of Technology Choice and Technical Efficiency: An Application to Organic and Conventional Dairy Farming[J]. *Journal of Productivity Analysis*, 2009, 31(3): 151–161.
- [58] Horrace W C, Liu X, Patacchini E. Endogenous Network Production Functions with Selectivity[J]. *Journal of Econometrics*, 2016, 190(2): 222–232.
- [59] Su D, He X. Ownership Structure, Corporate Governance and Productive Efficiency in China[J]. *Journal of Productivity Analysis*, 2012, 38(3): 303–318.
- [60] Hansen B E. Threshold Effects in Non – Dynamic Panels: Estimation, Testing, and Inference[J]. *Journal of Econometrics*, 1999, 93(2): 345–368.
- [61] Hansen B E. Sample Splitting and Threshold Estimation[J]. *Econometrica*, 2000, 68(3): 575–603.
- [62] Kutlu L. Battese – coelli Estimator with Endogenous Regressors[J]. *Economics Letters*, 2010, 109(2): 79–81.
- [63] Tran K C, Tsionas E G. GMM Estimation of Stochastic Frontier Model with Endogenous Regressors[J]. *Economics Letters*, 2013, 118(1): 233–236.
- [64] Griffiths W E, Hajargasht G. Some Models for Stochastic Frontiers with Endogeneity[J]. *Journal of Econometrics*, 2016, 190(2): 341–348.
- [65] Karakaplan M U. Fitting Endogenous Stochastic Frontier Models in Stata[J]. *Stata Journal*, 2017, 17(1): 39–55.
- [66] Kumbhakar S C, Parmeter C F, Tsionas E G. A Zero Inefficiency Stochastic Frontier Model[J]. *Journal of Econometrics*, 2013, 172(1): 66–76.
- [67] Lai H P, Huang C J. Maximum Likelihood Estimation of Seemingly Unrelated Stochastic Frontier Regressions [J]. *Journal of Productivity Analysis*, 2013, 40(1): 1–14.
- [68] Chang S K, Chen Y Y, Wang H J. A Bayesian Estimator for Stochastic Frontier Models with Errors in Variables [J]. *Journal of Productivity Analysis*, 2012, 38(1): 1–9.
- [69] Karakaplan M U, Kutlu L. Handling Endogeneity in Stochastic Frontier Analysis[J]. *Economics Bulletin*, 2017, 37(2): 889–901.
- [70] Kumbhakar S C. Estimation of Technical Efficiency Using Flexible Functional Forms and Panel Data[J]. *Journal of Business and Economic Statistics*, 1989, 7: 253–258.
- [71] Altunbas Y, Liu M H, Molyneux P, Seth R. Efficiency and Risk in Japanese Banking[J]. *Journal of Banking & Finance*, 2000, 24(10): 1605–1628.
- [72] 王德祥, 李建军. 我国税收征管效率及其影响因素——基于随机前沿分析(SFA)技术的实证研究[J]. *数量经济技术经济研究*, 2009(4): 152–160.

责任编辑:张 静,罗 红