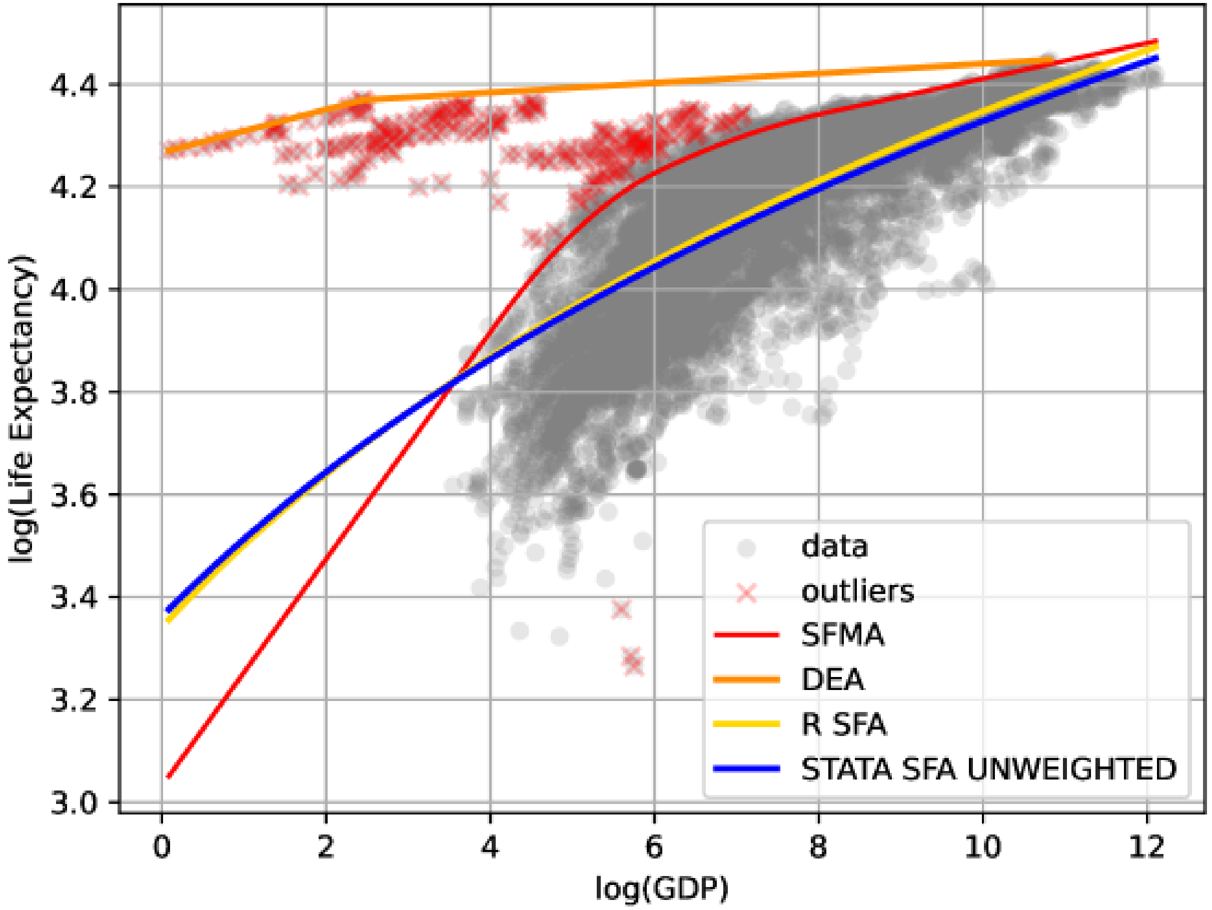
B6. 随机边界模型

- 异质性随机边界模型 `sfcross`, `sfpanel`
- 双边随机边界模型 `sftt`
- 稳健非参数随机前沿分析 `sfma`



连玉君 (中山大学)
arlionn@163.com

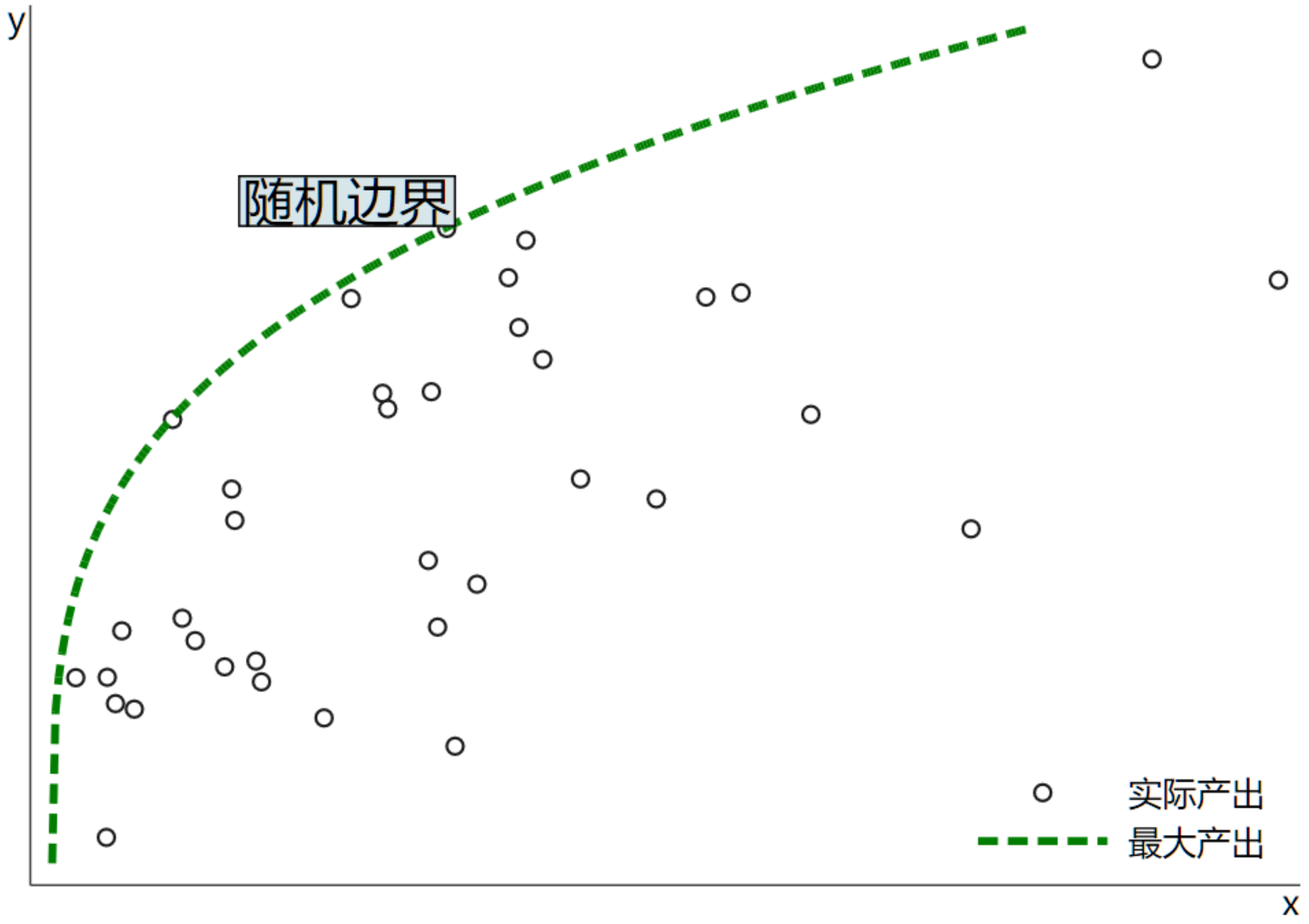
1. SFA 简介

1.1 SFA 的基本思想

理论上，任何经济个体的“实际产出”不可能超过“产出边界”，两者的偏离可以视为 **无效率损失**。

统计上，该思想可建模为包含“复合干扰项”的回归模型：

- 一项为正态误差项 v_i，用于捕捉测量误差与其他统计偏差；
- 一项为单边分布误差项 u_i，反映无效率。



假设第 i 个厂商在理想状态下（无效率损失）的最大产出为 $f(z_i)$，则其实际产出 $q_i < f(z_i)$。定义其技术效率为：

$$TE_i = \frac{q_i}{f(z_i)} \leq 1 \tag{18.1}$$

$$q_i = f(z_i, \beta) \cdot TE_i, \quad 0 < TE_i \leq 1 \tag{18.2}$$

若 $TE_i = 1$，表示完全效率；若 $TE_i < 1$，存在效率损失。SFA 关注的是 TE_i 及其决定因素。

为区分随机误差和效率因素，在模型中引入随机项 v_i：

$$q_i = f(z_i, \beta) \cdot TE_i \cdot \exp(v_i) \tag{18.3}$$

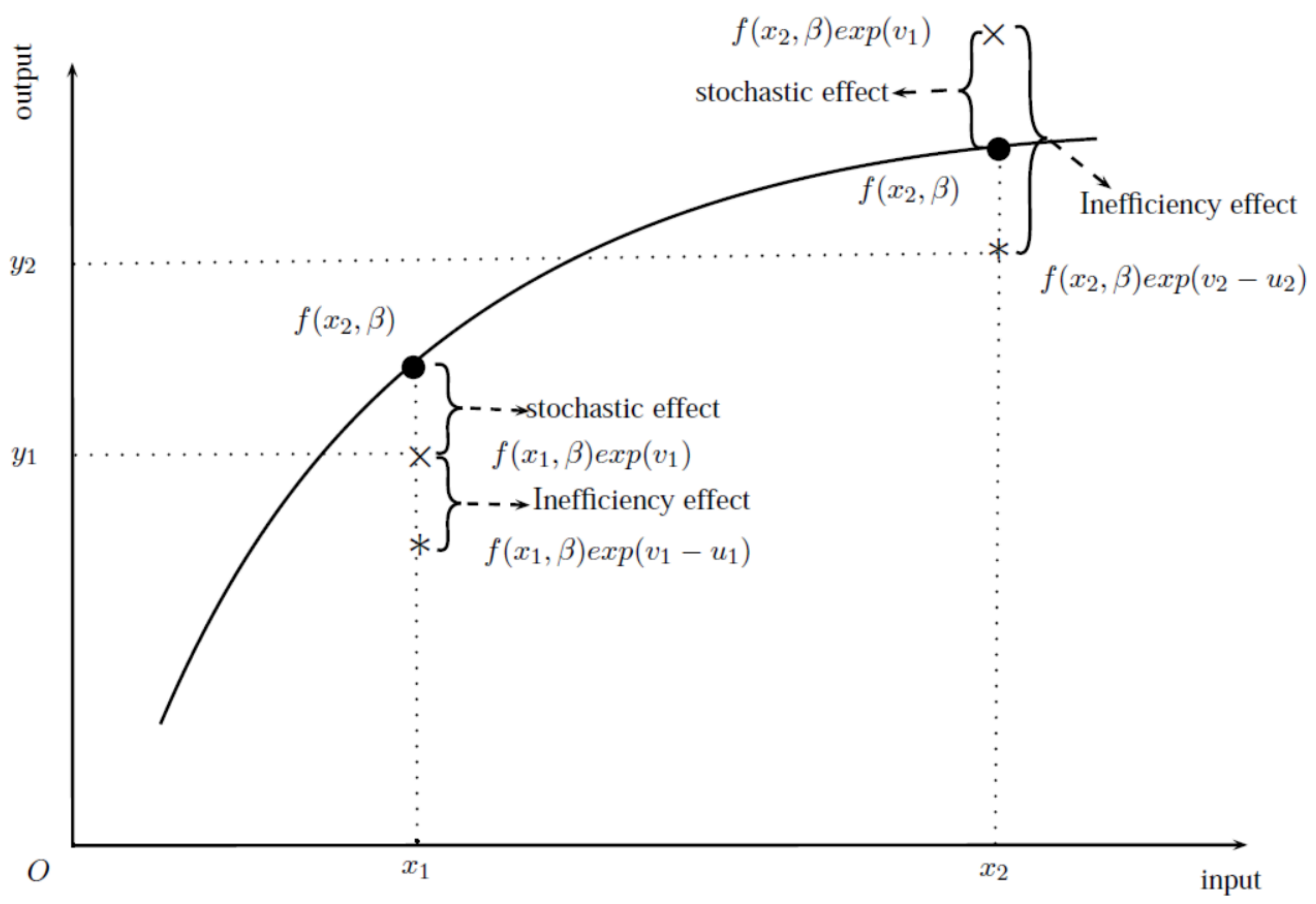
其中，$v_i \sim N(0, \sigma_v^2)$，通过对数变换，得：

$$\ln q_i = \ln f(z_i, \beta) + v_i - u_i \tag{18.5}$$

令 $u_i = -\ln(TE_i)$，$u_i \geq 0$，称为“技术无效率项”。因此：

$$TE_i = \exp(-u_i) \tag{18.6}$$

多数情况下假设 $\mathrm{Cov}(u_i, v_i) = 0$，即随机误差与无效率项独立。



Source: Porcelli, F. (2009) Measurement of Technical Efficiency. A Brief Survey on Parametric and Non-Parametric Techniques. University of Warwick 11 (527), 1-27, 2009. Figure 8. -PDF-.

1.2 模型设定

若采用对数线性生产函数（如 Cobb-Douglas）：

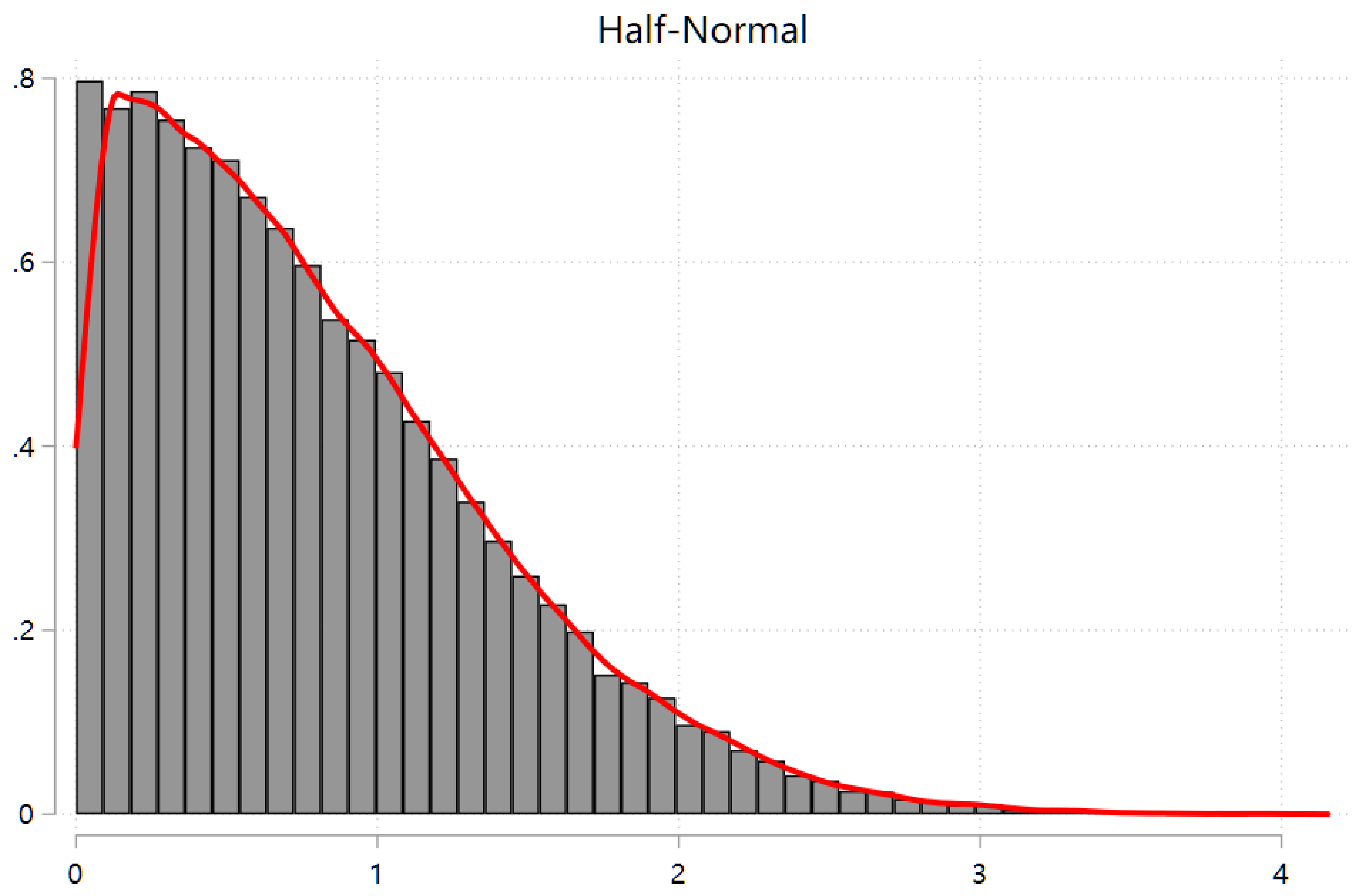
$$\ln q_i = \beta_0 + \sum_{j=1}^{k} \beta_j \ln z_{ji} + v_i - u_i \tag{18.7}$$

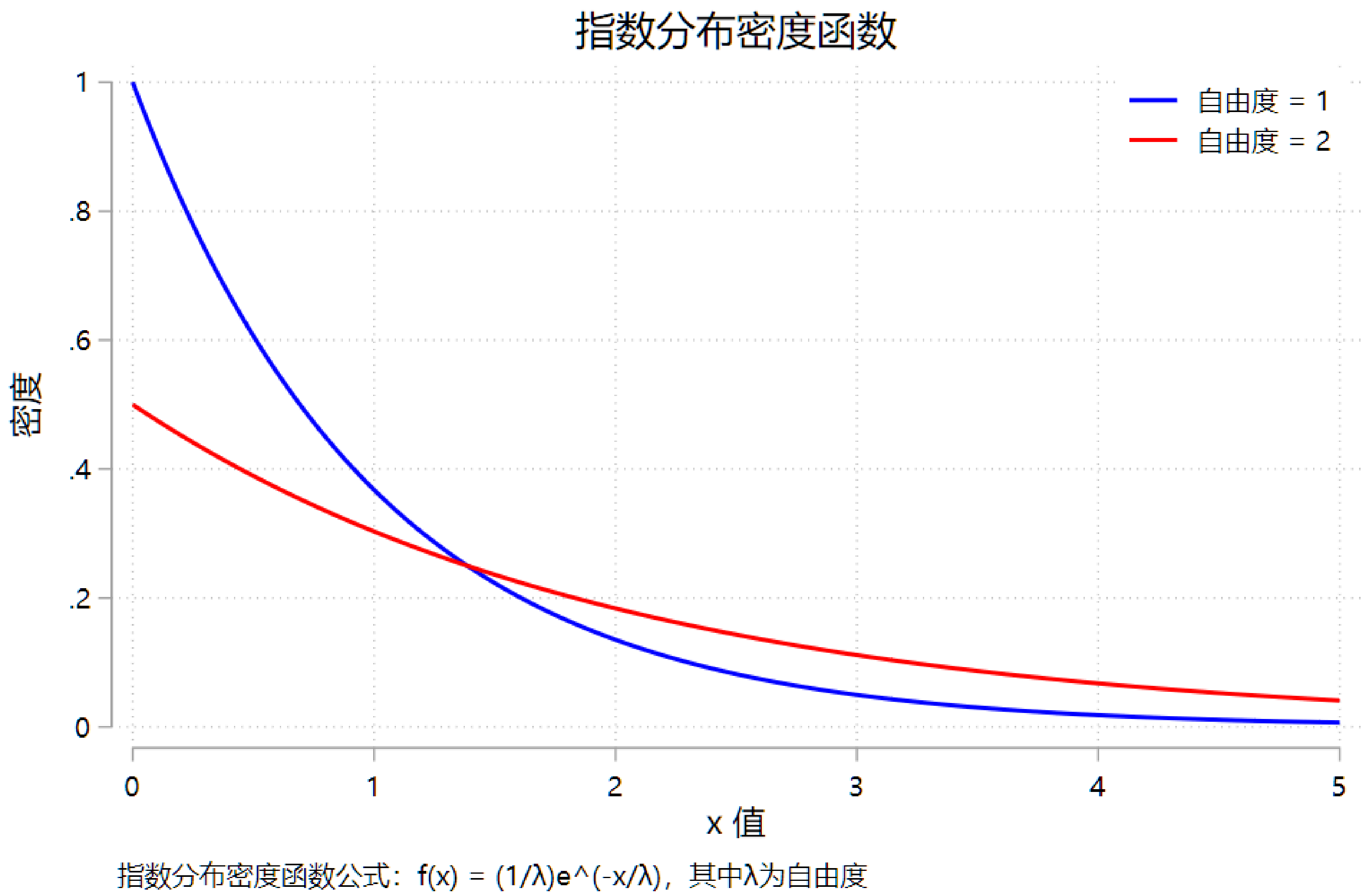
$$y_i = x_i'\beta + v_i - u_i = x_i'\beta + \varepsilon_i, \quad \varepsilon_i = v_i - u_i \tag{18.8}$$

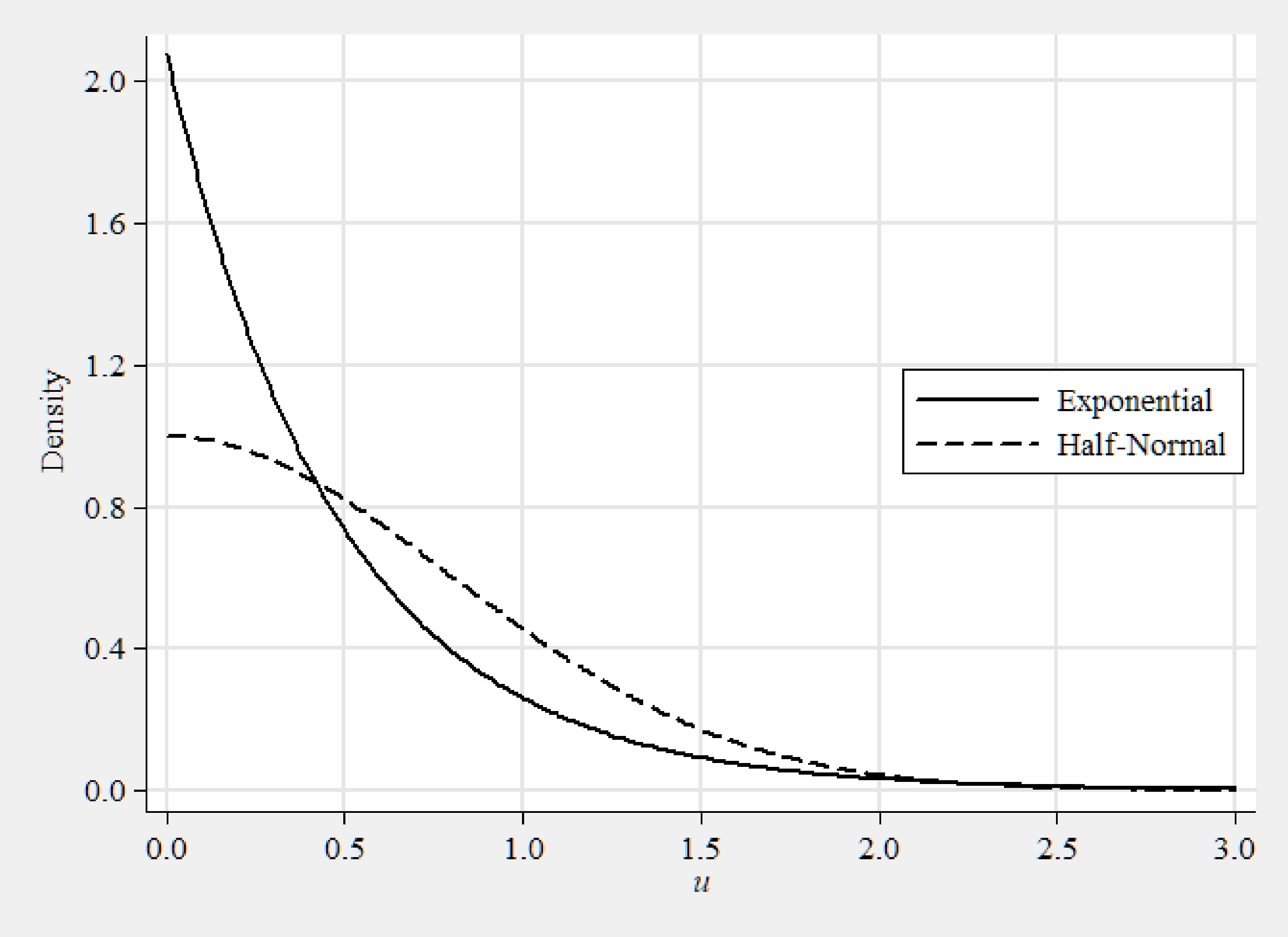
- $y_i = \ln q_i$，$x_{ji} = \ln z_{ji}$；ε_i 为复合误差项；假设 $\mathrm{Cov}(x_i, \varepsilon_i) = 0$，则 OLS 得到的 $\hat{\beta}$ 是一致的。

为估计 TE_i，需对 v_i 和 u_i 的分布做进一步假设：

- $v_i \sim N(0, \sigma_v^2)$；
- u_i 单边分布：
  - 半正态分布 $u_i \sim |N(0, \sigma_u^2)|$；
  - 截断正态 $u_i \sim N^+(\mu, \sigma_u^2)$；
  - 指数分布 $u_i \sim \mathrm{Exp}(\theta)$。


Half-Normal
.8
.6
.4
.2
0
0
1
2
3
4


指数分布密度函数
1
.8
.6
.4
.2
0
密度
0
1
2
3
4
5
x 值
自由度 = 1
自由度 = 2
指数分布密度函数公式：f(x) = (1/λ)e^(-x/λ)，其中λ为自由度


Density
2.0
1.6
1.2
0.8
0.4
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
u
Exponential
Half-Normal

1.4 效率 / 非效率的估计

使用 SFA 模型的主要目的，在于研究“效率”或“非效率”。一般分两类用途：

- 对比不同公司或行业的效率水平；
- 探究影响效率的因素。

根据模型 (18.4)，可以定义技术效率如下：

$$TE_i = \frac{y_i}{f(x_i,\beta)\cdot\exp(v_i)} = \exp(-u_i) \tag{18.24}$$

其中 $y_i^* = f(x_i,\beta)\cdot\exp(v_i)$ 表示第 i 个个体的“随机边界产出”，效率即为实际产出与该边界的比值。

1.5 假设检验和模型筛选

$$y_i = x_i'\beta + v_i - u_i = x_i'\beta + \varepsilon_i, \quad \varepsilon_i = v_i - u_i \tag{18.8}$$

- $v_i \sim N(0, \sigma_v^2) \qquad u_i \sim N^+(\mu, \sigma_u^2);$

非效率项显著性检验

检验企业是否存在效率损失，本质是检验 u_i 是否显著。对应假设为：

- H_0: $\sigma_u^2 = 0$（即无效率项不存在）
- H_1: $\sigma_u^2 > 0$

若不能拒绝 H_0，则 SFA 模型简化为 OLS 模型。

由于 σ_u^2 位于参数空间边界（不能小于零），传统 LR 检验不再适用，需使用一般化 LR 检验，其统计量服从**混合卡方分布**，详见 Greene (2008)、Kumbhakar and Lovell (2000)。

嵌套模型之间的对比

若两个模型具有嵌套关系，则可使用 LR 检验。以正态-半正态模型 (hN) 和截断正态模型 (tN) 为例：

- hN 模型：$u_i \sim N^+(0, \sigma_u^2)$
- tN 模型：$u_i \sim N^+(\omega, \sigma_u^2)$

检验 $H_0 : \omega = 0$。定义：

$$LR = 2(\ln L_1 - \ln L_0) \sim \chi^2(d_1 - d_0) \tag{18.29}$$

其中 L_1, L_0 分别为两个模型的对数似然值，d_1, d_0 为参数个数。

1.5 生产函数的设定

- Cobb-Douglas 函数虽形式简单，但隐含严格假设：
  - 要素份额和需求弹性为常数；
  - 要素间替代弹性为 −1；
- 若需放松这些限制，可引入二次项，设定更灵活模型形式（Kumbhakar, 1989）：

$$\ln y = \alpha + \sum_k \beta_k x_k + \frac{1}{2} \sum_k \sum_m \gamma_{km} x_k x_m \tag{18.30}$$

文献实例：

- Altunbas et al. (2000) 在 cost-SFA 模型中使用高阶项分析日本银行；
- Wang (2007) 在研究 R&D 效率时也采用该设定。

2. 异质性 SFA

SFA 模型的核心目的是分析企业效率及其决定因素。

- 在前文中我们主要假定非效率项 u_i 或 u_{it} 是同质的，即不同个体的非效率项服从相同分布。
- 但现实中，非效率项通常会受到公司特征、管理制度、行业属性等影响。

为此，文献提出可将非效率项的分布参数设定为异质性的函数，构成 **异质性 SFA 模型**（Heteroscedastic SFA）。

2.1 模型设定问题

最常见的方式是将非效率项的分布参数（如截断正态的均值 ω_i）设定为公司特征变量的函数。例如，在正态-截断型半正态模型中：

$$u_i \sim N^+(\omega_i, \sigma_u^2), \quad \omega_i = z_i'\gamma \tag{18.20}$$

其中：

- z_i 为与无效率相关的解释变量，如公司规模、行业、产权性质等；
- γ 为待估系数；
- ω_i 决定非效率项分布的中心。

这样设定可将非效率建模为函数 $u_i = f(z_i;\gamma) + \text{扰动项}$。

当 u_{it} 服从时变模型时，也可设：

$$u_{it} = z_{it}'\gamma + \eta_{it}, \quad \eta_{it} \sim N^+(0, \sigma^2)$$

2.2 一步估计与两步估计

两步估计法（传统做法）

1. 第一步：估计基础 SFA 模型，得到效率估计值 $T\hat{E}_i = \exp(-\hat{u}_i)$；
2. 第二步：以 $T\hat{E}_i$ 或 $\hat{u}_i$ 为被解释变量，回归于公司特征变量 z_i，分析影响因素。

该方法的缺点：

- $\hat{u}_i$ 是估计值，带有误差，第二步 OLS 回归标准误不能正确估计；
- 此外，$\hat{u}_i$ 的分布偏离正态，标准 t 检验不可靠。

一步估计法（推荐做法）

直接在极大似然估计中嵌入非效率项的异质性结构（例如设 $\omega_i = z_i'\gamma$），最大化如下对数似然函数：

$$\ln L = \sum_i \ln\left\{\frac{1}{\sigma}\phi\left(\frac{\varepsilon_i}{\sigma}\right)\Phi\left(\frac{\mu_i}{\sigma^*}\right)\right\}$$

- 其中 $\mu_i = z_i'\gamma$，其余参数如前所定义。
- 🍎 Stata： `frontier` 命令 + `het()` 选项，或 `sfcross` 命令配合 `hetmean()` 和 `hetsd()` 选项。

小结

- 异质性 SFA 模型将非效率项的分布参数建模为公司特征的函数；
- 推荐使用一步估计法，避免传统两步法的误差传导问题；
- 通过模拟方法可以得到效率的边际效应估计；
- Stata 中 `frontier`，`sfcross`，`sfpanel` 命令都支持异质性设定。

3. 双边随机边界模型

- 双边随机边界模型（Two-Tier SFA）是对传统 SFA 模型的扩展，主要用于研究工资议价等双边市场问题。
- 该模型由 Kumbhakar & Parmeter (2009) 提出，并将其应用于工资议价问题。
- Papadopoulos & Parmeter (2025) 的专著，从议价理论、信息不对称、遗漏变量 (如生产能力、管理能力等)、潜变量等角度提供设定双边随机边界模型的多种可能的理论框架。
- Lian, Liu and Parmeter et al. (2023) Papadopoulos & Parmeter (2025) 对传统的 TT-SFA 模型进行如下拓展：
  - 非效率项的分布不再局限于指数分布，允许半正态分布等更为灵活的设定，使我们可以更好地捕捉数据中的异质性特征。
  - 支持面板数据：可以设定双向固定效应
  - 允许干扰项截面相关，这是审稿人经常质疑的一个问题。

3.1 模型设定

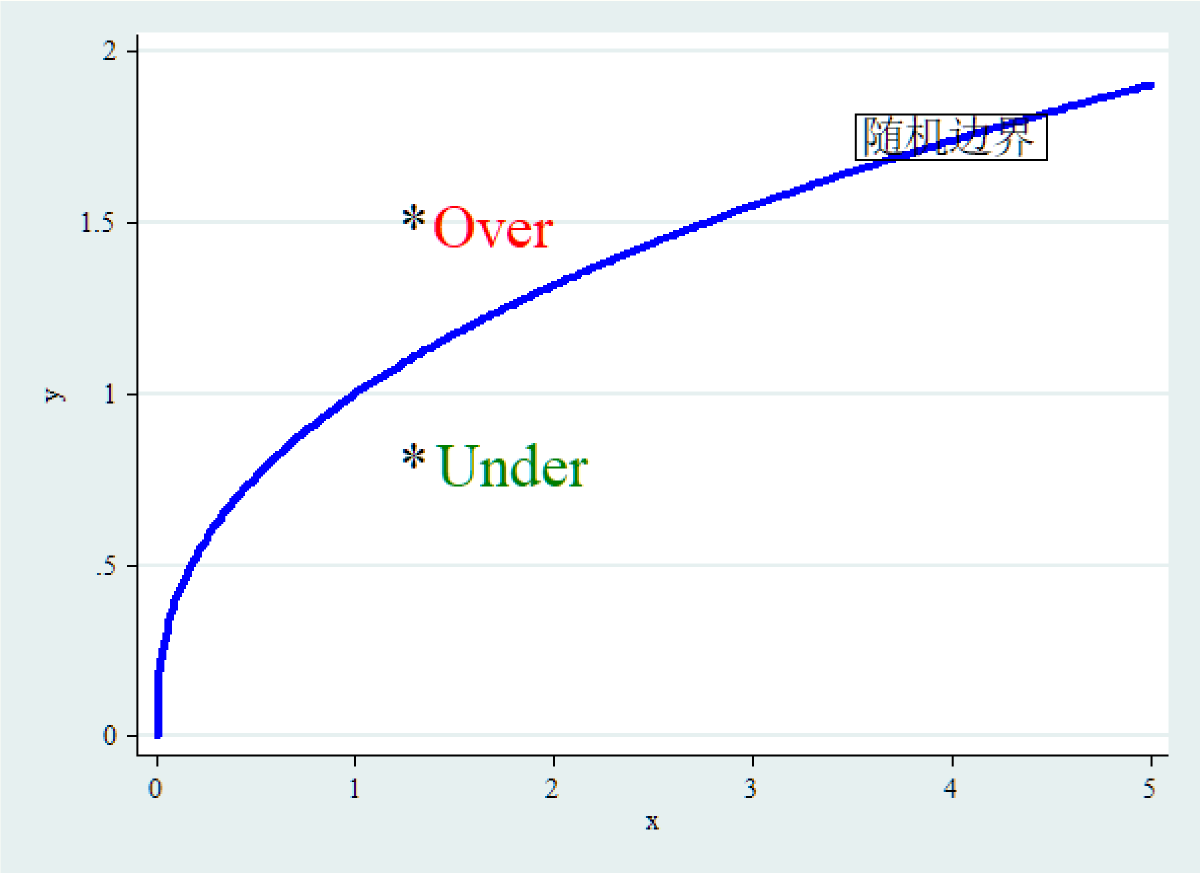
$$y_i = x_i\beta + v_i + w_i - u_i$$

$$\begin{aligned} v_i &\sim i.i.d.\ N(0,\sigma_v^2) \\ w_i &\sim i.i.d.\ \mathrm{Exp}(\sigma_w,\sigma_w^2) \\ u_i &\sim i.i.d.\ \mathrm{Exp}(\sigma_u,\sigma_u^2) \end{aligned} \quad (18.60)$$

该模型的复合干扰项有三部分组成：

- v_i 表示常规干扰项，假设其服从正态分布；
- w_i 和 u_i 都具有单边分布，此处假设它们服从指数分布。

在这种设定下，w 和 u 的期望值都是大于零的，即 $E(w_i) \geq 0$ 且 $E(u_i) \geq 0$。



3.2 理论基础：以工资议价为例

- 在劳动力市场 (Kumbhakar & Parmeter, 2009, JPA)
  - 工人的最低要求工资（保留工资）：$\underline{y}_i$
  - 雇主可支付的最高工资：$\bar{y}_i$
  - 市场合理工资水平（均衡工资）：y^*
- 三者之间满足：$\underline{y}_i \leq y^* \leq \bar{y}_i$
- **工人剩余**：$\left(y_i^* - \underline{y}_i\right)$ **厂商剩余**：$(\bar{y}_i - y_i^*)$
- 假定观测到的实际工资为 y_i，工人的议价能力为 θ，则

$$y_i = \underline{y}_i + \theta\left(\bar{y}_i - \underline{y}_i\right)$$

- 其中 $\theta\left(\bar{y}_i - \underline{y}_i\right)$ 表示工人可获得的总剩余。

工资表达式的进一步变形

$$y_i = y_i^* + \theta\left(\bar{y}_i - y_i^*\right) - (1-\theta)\left(y_i^* - \underline{y}_i\right)$$

议价能力与模型参数化

- 设工人剩余 $w_i = \theta(\bar{y}_i - y_i^*) \geq 0$，无法直接观测
- 设厂商剩余 $u_i = (1-\theta)(y_i^* - \underline{y}_i) \geq 0$，无法直接观测
- 均衡价格 y^* 通常无法直接观测，可设定为厂商和工人特征变量的线性函数：

$$y^* = \mathbf{x}_i\beta + v_i$$

- 其中 v_i 为干扰项，$\mathbf{x}_i$ 可包含行业、教育、经验、年龄、婚否、种族、IQ 等
- w_i 和 u_i 也可以设定为厂商和工人特征的线性函数。
- 则上述工资表达式变为二元 SFA 结构：

$$y_i = \mathbf{x}_i\beta + v_i + w_i - u_i$$

3.3 效率衡量及含义 (Lian et al., 2023, SJ)

厂商议价能力： $1-e^{-u}$

$$\frac{\text{Maximum price - Actual price}}{\text{Maximum price}}=\frac{\bar{y}_i-y_i^*}{\bar{y}_i}=1-e^{-u}$$

工人议价能力： $1-e^{-w}$

$$\frac{\text{Actual price }-\text{ Minimum price}}{\text{Actual price}}=\frac{y_i^*-\underline{y}_i}{y_i^*}=1-e^{-w}$$

4. SFMA 模型：稳健非参数随机前沿分析

4.1 SFMA 简介

随机前沿元分析（SFMA） 是一种半参数前沿估计方法，旨在解决传统方法的核心局限：

- 突破预设函数形式限制（如 SFA 的参数假设）
- 处理输入数据的报告误差（支持异质性方差）
- 增强对异常值的鲁棒性（内置修剪策略）

核心创新：

- 采用 B 样条与形状约束建模前沿函数
- 整合元分析思想处理数据不确定性
- 基于似然的修剪策略自动剔除异常值

> 软件实现：开源 Python 包 `sfma`，Github：https://github.com/ihmeuw-msca/sfma

4.2 模型设定

基本模型表达式：

$$y_i = \langle x_i, \beta \rangle + u_i - v_i + \epsilon_i$$

$$\begin{aligned} u_i &\sim N(0, \gamma) \quad &\text{（随机效应，非抽样误差）} \\ v_i &\sim HN(0, \eta) \quad &\text{（无效率项，单边分布）} \\ \epsilon_i &\sim N(0, \sigma_i^2) \quad &\text{（抽样误差，已知方差）} \end{aligned}$$

参数含义：

- y_i：产出变量
- $\langle x_i, \beta \rangle$：基于协变量的生产前沿（样条基函数构建）
- u_i：随机效应项（非抽样误差）
- v_i：技术无效率项（$v_i \geq 0$）
- ϵ_i：抽样误差（可含已知异质性方差）

4.3 数据生成过程 (Next Page)

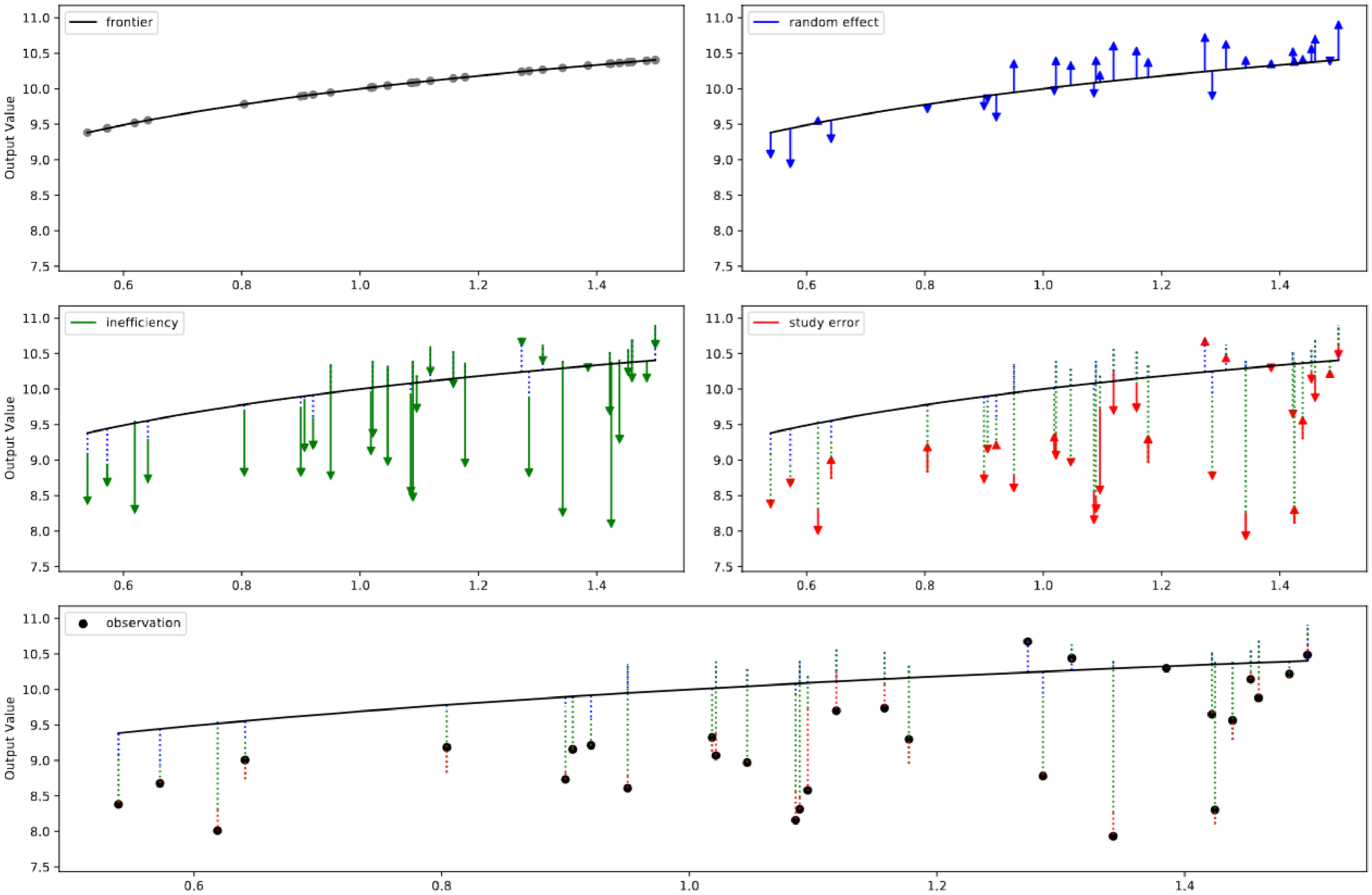

frontier
random effect
inefficiency
study error
observation
Output Value
11.0
10.5
10.0
9.5
9.0
8.5
8.0
7.5
0.6
0.8
1.0
1.2
1.4

图 1 展示了如何生成随机前沿元分析数据。其中观测数据点以黑色显示于底部。数据生成过程中的参数设定如下：

- 前沿函数设定为 $f(x) = \log(x) + 10$；
- 协变量 x_i 在 0.5 至 1.5 间均匀抽取；
- 随机效应 $u_i \sim N(0, \gamma = 0.25)$；
- 技术无效率项 $v_i \sim HN(0, \eta = 1)$；
- 抽样误差 $\epsilon_i \sim N(0, \sigma_i^2)$，其中 σ_i^2 在 0 至 0.75 间均匀抽取。

图 1 的具体制作过程如下：

1. 左上角：
   - 生成 x_i 的值，范围在 0.5 至 1.5 之间；
   - 计算前沿函数 $f(x_i) = \log(x_i) + 10$；
   - 绘制散点图 $f(x_i) \sim x_i$。
2. 右上角：
   - 生成随机效应 u_i；
   - 生成实际产出 y_i，计算公式为 $y_i = f(x_i) + u_i$。
   - 绘制散点图 $y_i \sim x_i$。
3. 后续图形：
   - 依次生成技术无效率项 v_i (左中)、抽样误差 ϵ_i (右中)；
   - 计算最终观测值 $y_i = f(x_i) + u_i - v_i + \epsilon_i$；
   - 绘制最终观测数据点。

4.4 样条基函数的基本思想

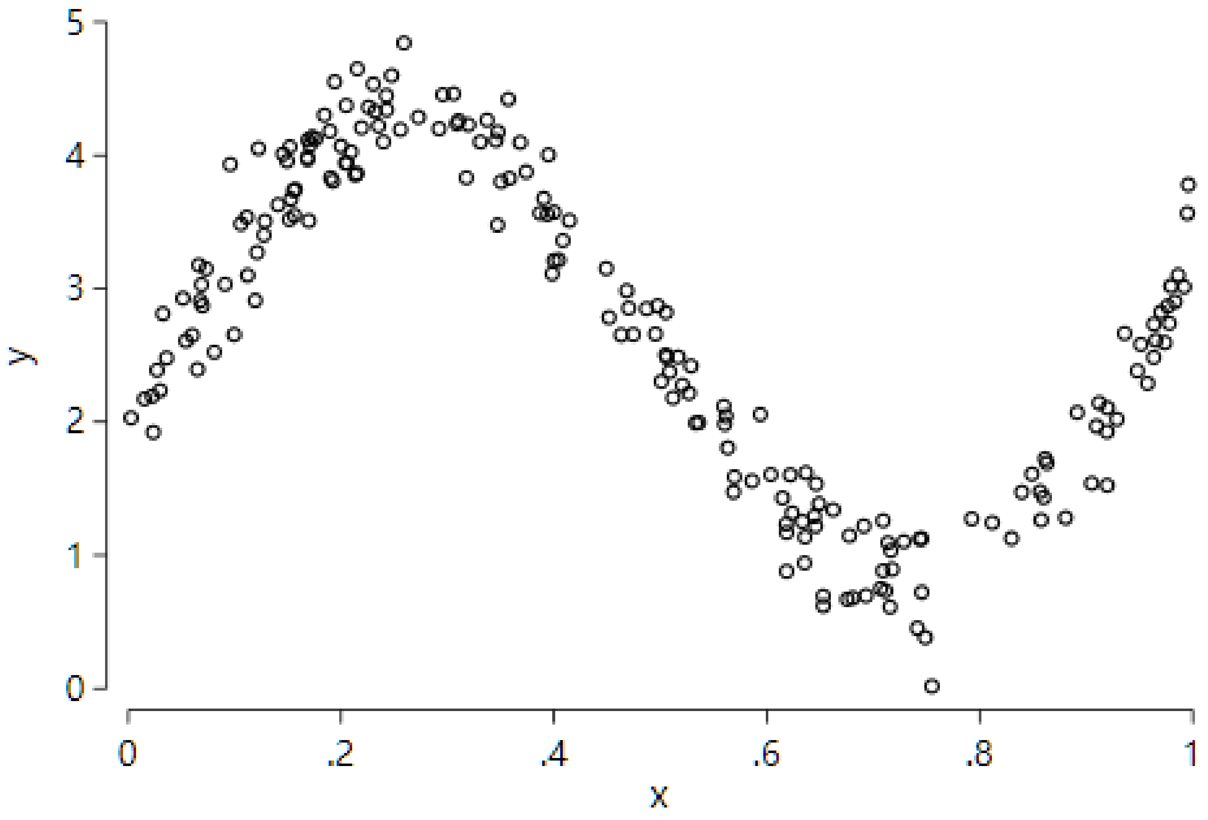
本例设计一个高度非线性的生产前沿函数：

$y = 2 + 1.5x - 0.8x^2 + 0.5x^3 + 2\sin(2\pi x) + \epsilon$

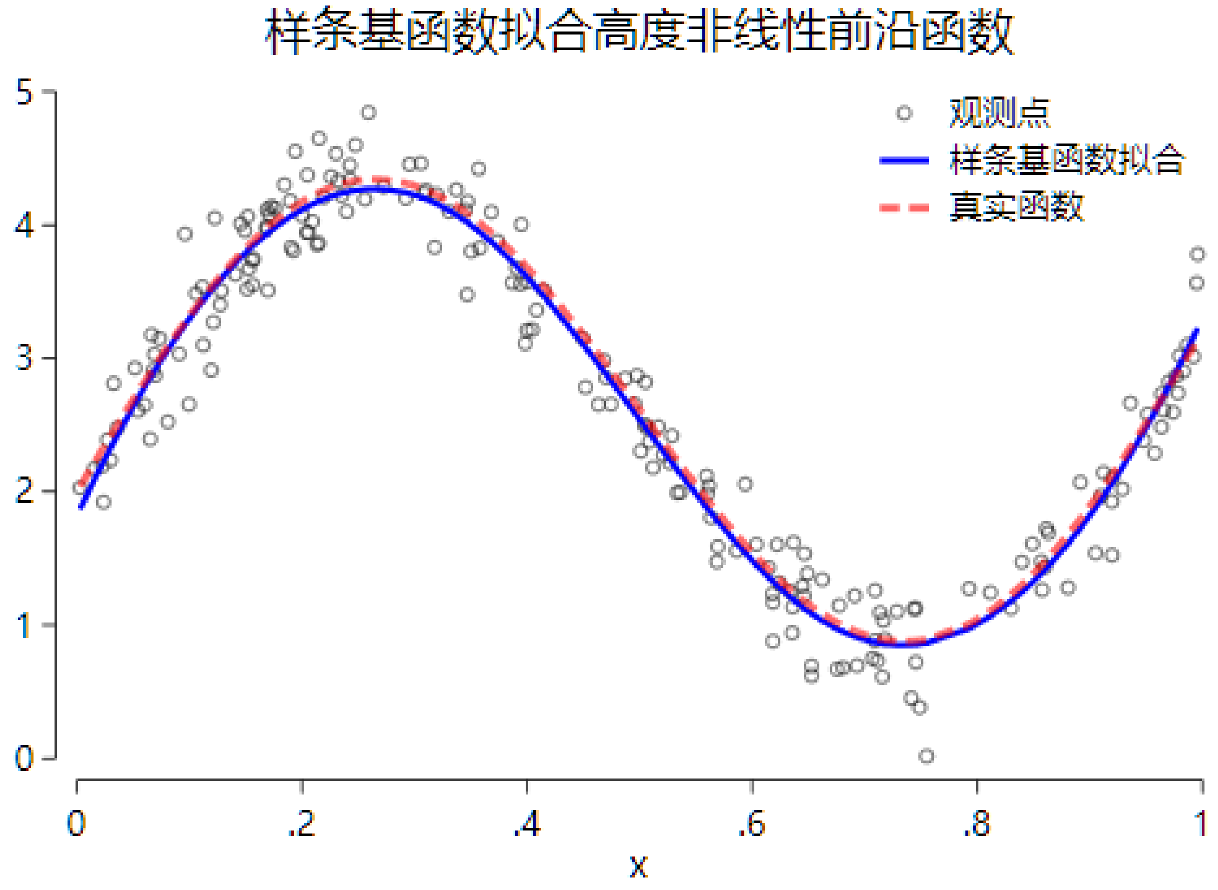
- 其中 $\epsilon \sim N(0, 0.25^2)$， $x \in [0, 1]$

```
clear
set obs 200

gen x = runiform()

gen y_true = 2 + 1.5*x - 0.8*x^2
       + 0.5*x^3 + 2*sin(2*_pi*x)

gen y = y_true + rnormal(0, 0.25)

scatter y x, msymbol(Oh) msize(small)
```

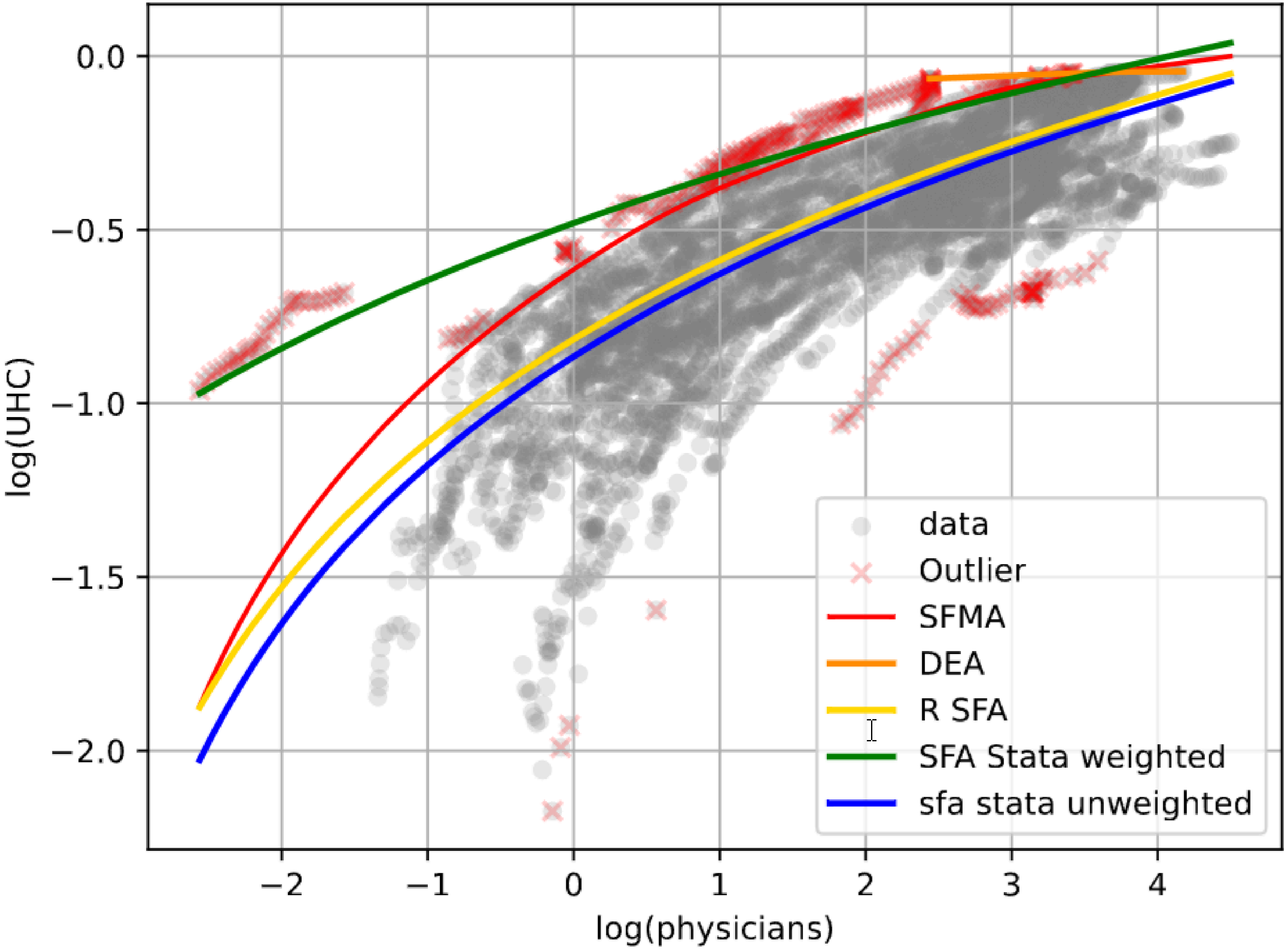
散点图：高度非线性前沿函数

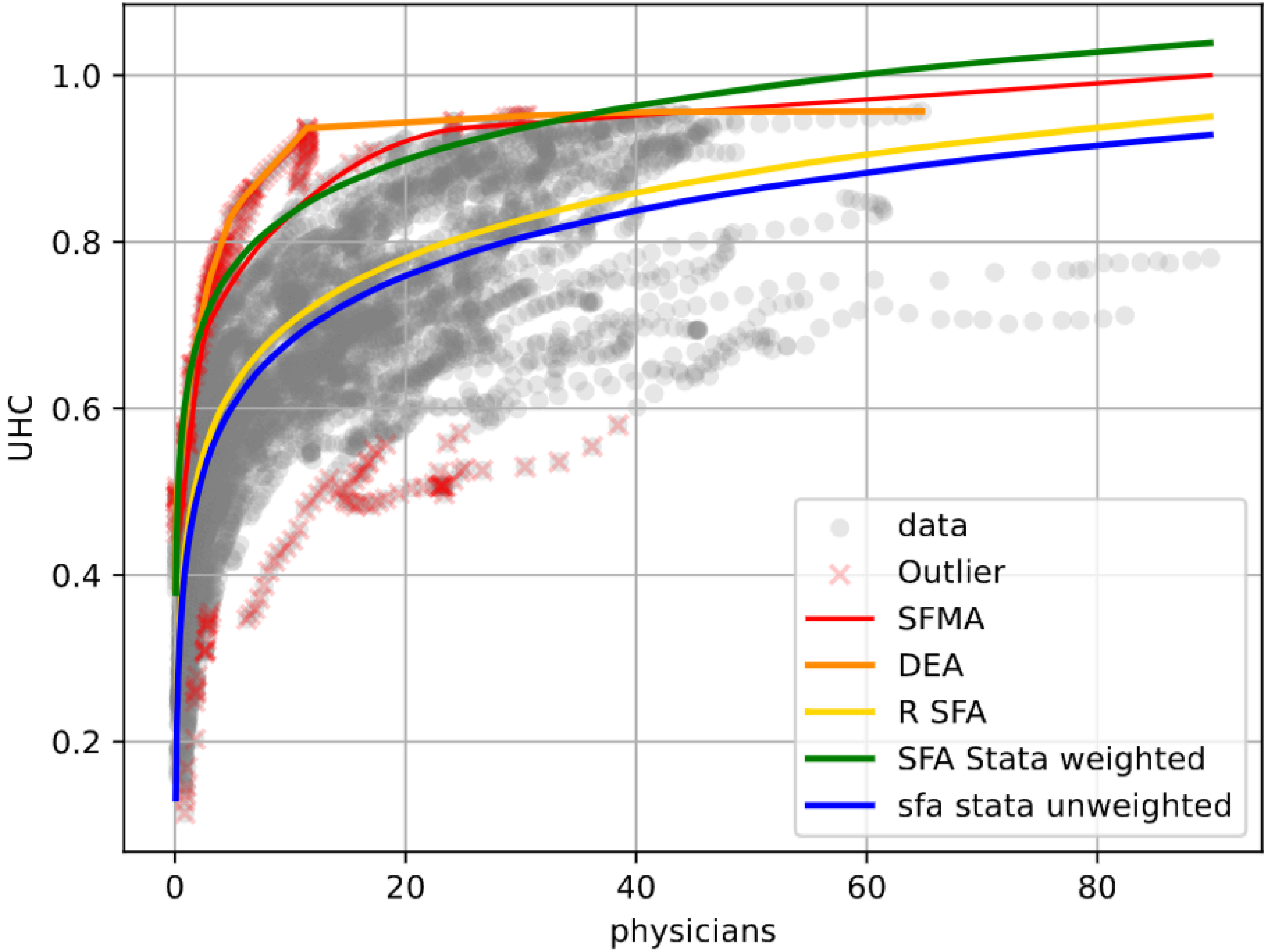


```
* 6. 手动构造样条基函数
gen x2 = x^2
gen x3 = x^3
gen s1 = (x > 0.33) * (x - 0.33)^3
gen s2 = (x > 0.66) * (x - 0.66)^3

* 7. 拟合 OLS
reg y x x2 x3 s1 s2

* 8. 预测拟合值
predict y_hat

* 9. 三线合一作图
#delimit ;
twoway
  (scatter y x, msym(Oh) mc(black%50))
  (line y_hat x, sort lc(blue)
        lw(*2.0) lp(solid))
  (line y_true x, sort
        lc(red%60)  lw(*2.5) lp(dash)),
  legend(order(1 "观测点"
               2 "样条基函数拟合"
               3 "真实函数")
         ring(0) position(1))
  xsize(4) ysize(3);
#delimit cr
```



4.5 SFMA 与传统方法的对比




1.0
0.8
0.6
0.4
0.2
UHC
0
20
40
60
80
physicians
data
Outlier
SFMA
DEA
R SFA
SFA Stata weighted
sfa stata unweighted

4.6 Python 实现

参见

- 作者：Github
- 课件：【B6_SFA/sfma/notebooks】 文件夹
  - `Data Simulations-Fig1-5.ipynb`
  - `GDP and LE-Fig6-7.ipynb`
  - `UHC-Current-Fig8-9.ipynb`

Thanks

lianxh.cn