双边随机边界模型的STATA应用

中山大学 刘畅
liuch288@mail2.sysu.edu.cn

内容概述

- 随机边界模型概述与应用
- 基于分布假设的模型设定与估计
- 利用缩放性质的模型设定与估计
- STATA实际操作演示

双边随机边界模型的发展历程

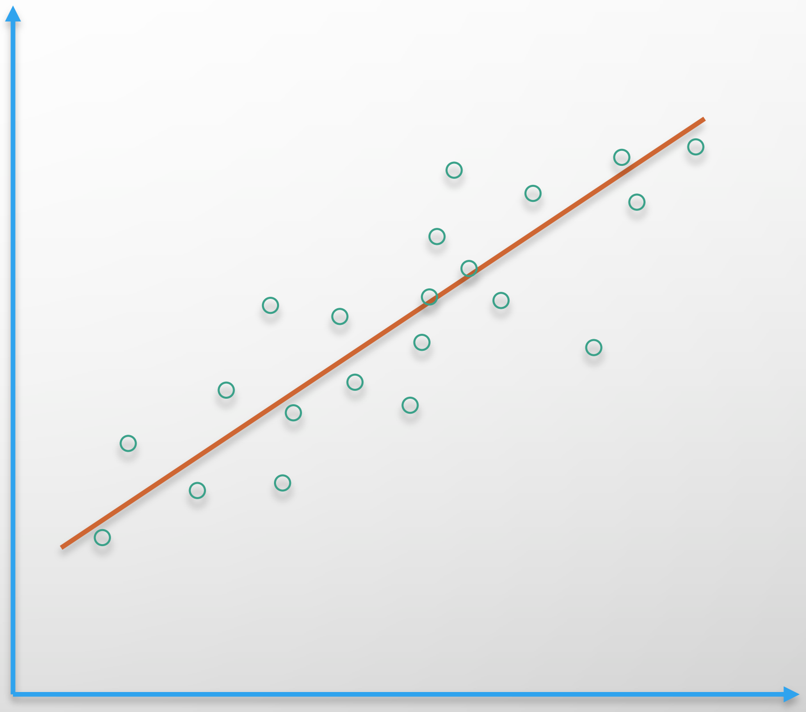
- 随机边界模型 (Aigner et al., 1977; Meeusen and roeck, 1977)
- 双边随机边界模型 (Polachek and Yoon, 1987)
- 非效率项分析 (Kumbhakar and Parmeter, 2009)
- 其他模型设定 (Papadopoulos, 2015; Parmeter, 2018)

模型概述

- 经典线性/广义线性模型：
  - 关注被解释变量的条件均值
  - 扰动项为对称分布

$$y = X \cdot \beta + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

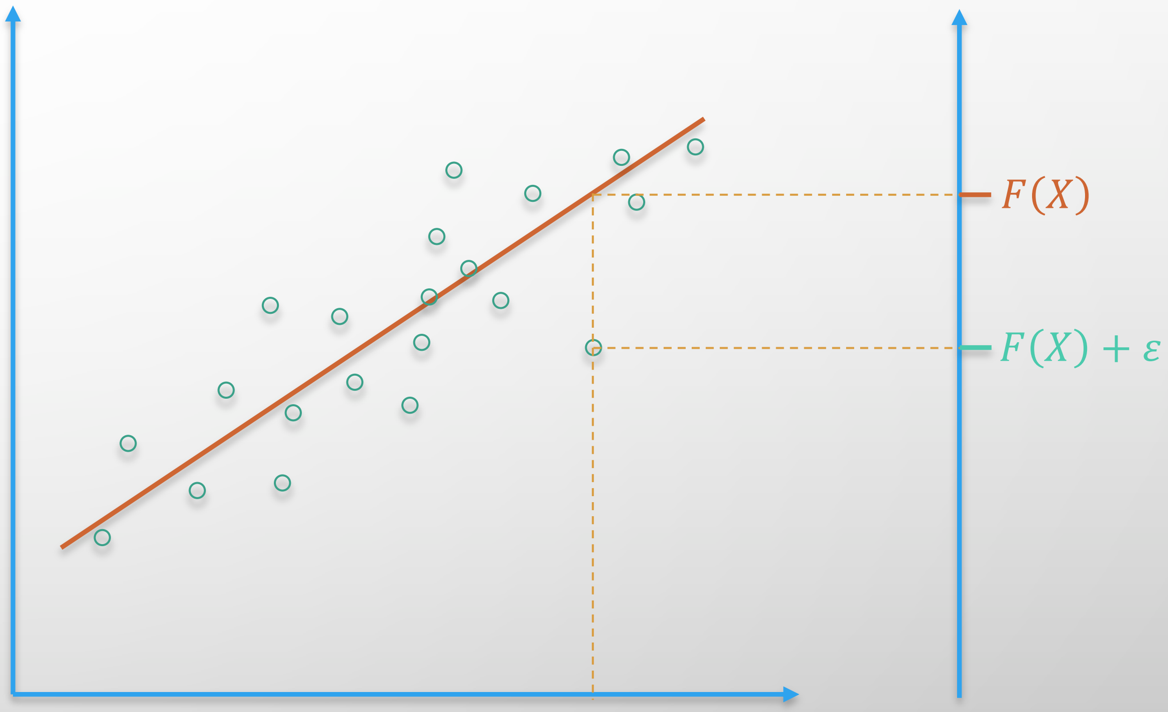


模型概述

- 经典线性/广义线性模型：
  - 关注被解释变量的条件均值
  - 扰动项为对称分布

$$y = X \cdot \beta + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$



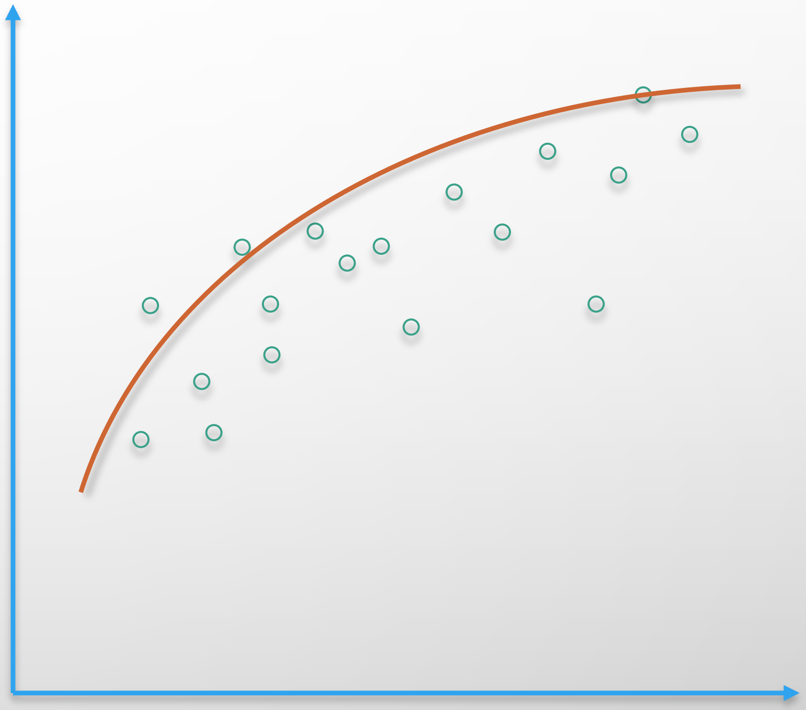
模型概述

- 随机边界模型：
  - 对被解释变量的最大值、最小值建模
  - 允许扰动项非对称

$$y = X \cdot \beta + \varepsilon$$

$$\varepsilon = -u + v$$

$$u > 0, v \sim N(0, \sigma_v^2)$$



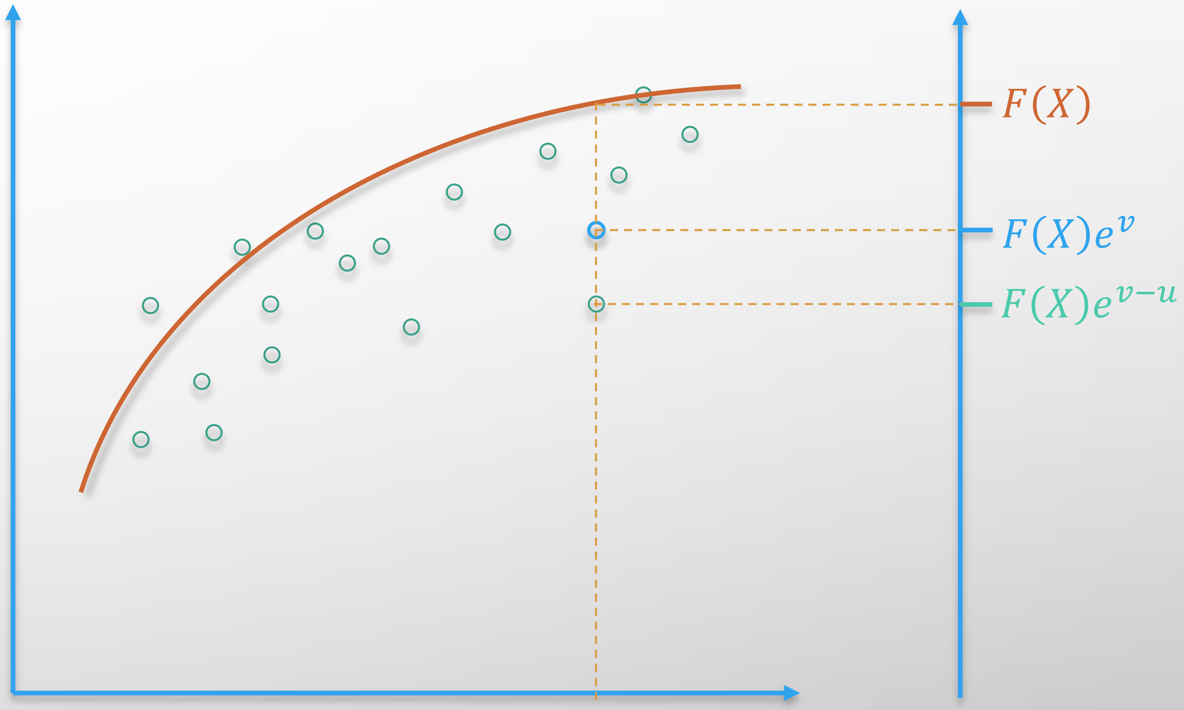
模型概述

- 随机边界模型：
  - 对被解释变量的最大值、最小值建模
  - 允许扰动项非对称

$$y = X \cdot \beta + \varepsilon$$

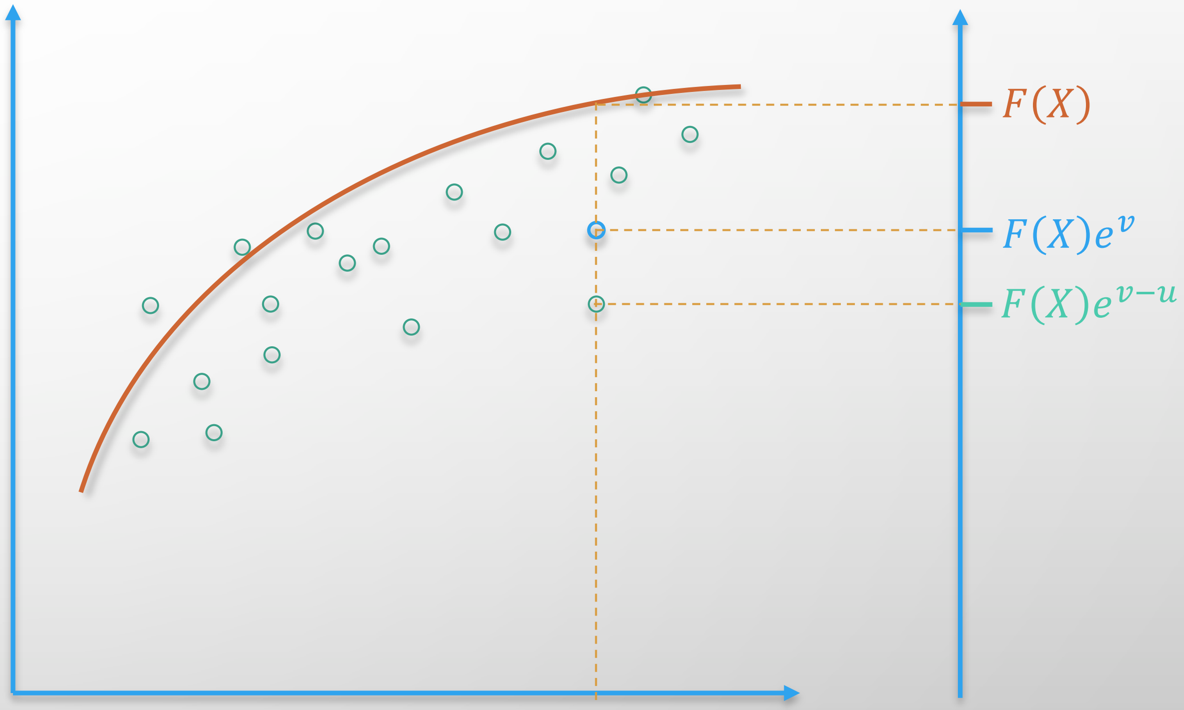
$$\varepsilon = -u + v$$

$$u > 0, v \sim N(0, \sigma_v^2)$$



模型概述

- 随机边界模型：
  - 对被解释变量的最大值、最小值建模
  - 应用场景：
    - 生产效率分析
      (Fenn et al., 2008)
    - 产能过剩水平测度
      (Artez and Pope, 2018)



模型概述

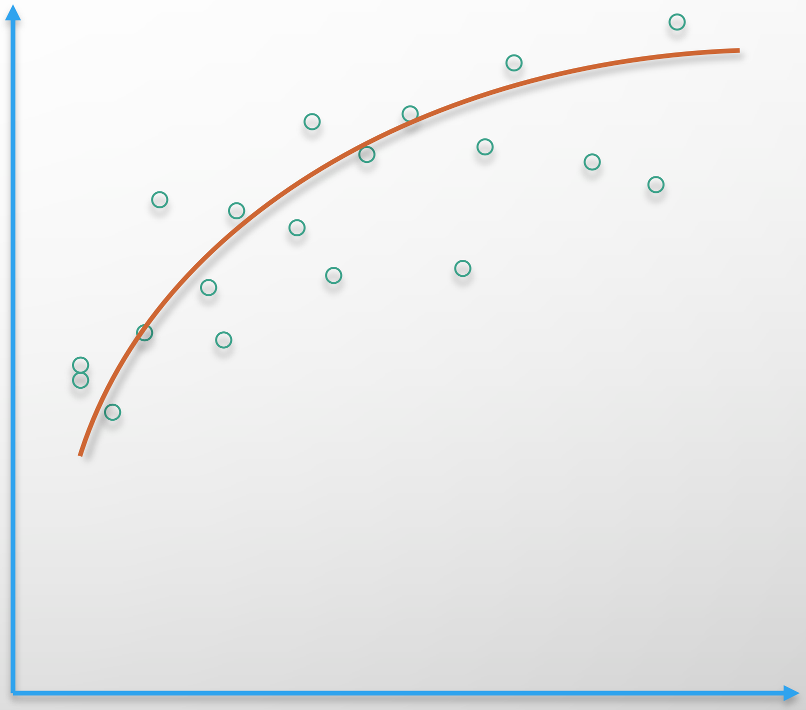
- 双边随机边界模型：
  - 对被解释变量的最优值建模
  - 单边扰动项分布于两个方向

$$y = X \cdot \beta + \varepsilon$$

$$\varepsilon = v - u + w$$

$$u, w > 0,$$

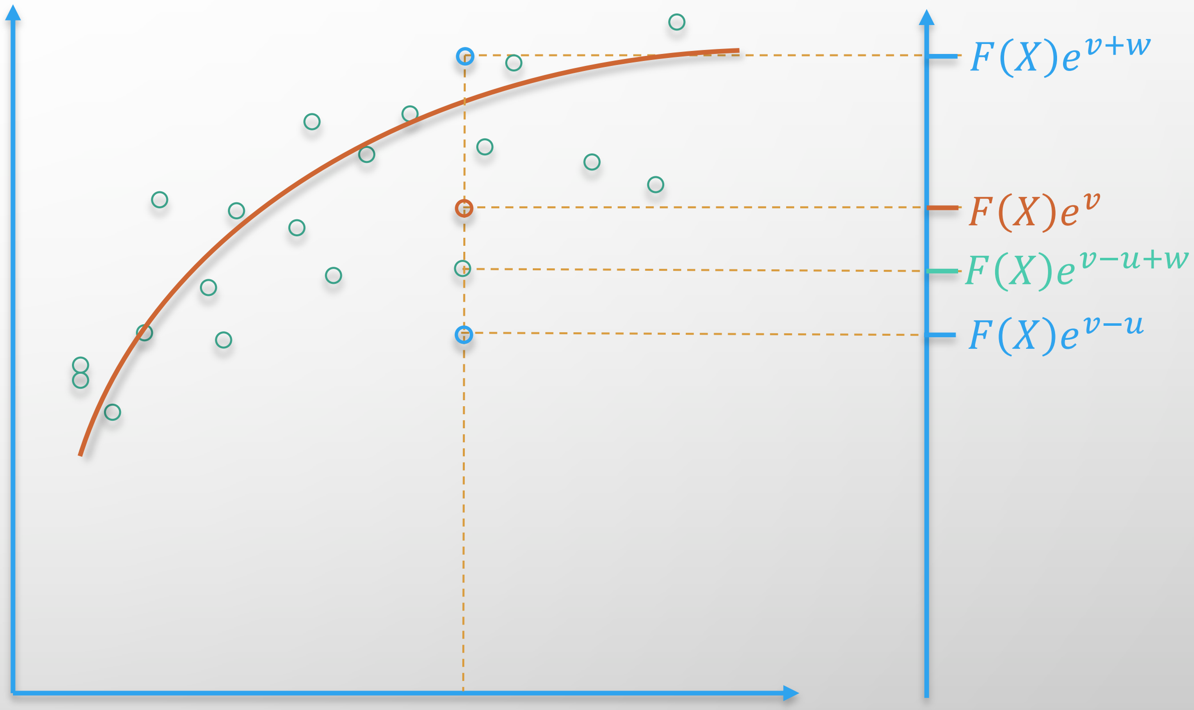
$$v \sim N(0, \sigma_v^2)$$



模型概述

- 双边随机边界模型：
  - 对被解释变量的最优值建模
  - 单边扰动项分布于两个方向

$$y = X \cdot \beta + \varepsilon$$
$$\varepsilon = v - u + w$$
$$u, w > 0,$$
$$v \sim N(0, \sigma_v^2)$$



模型概述

- 双边随机边界模型：
  - 对被解释变量的最优值建模
  - 单边扰动项分布于两个方向
  - 应用场景：
    - 价格博弈(Kumbhakar and Parmeter, 2009; Blanco 2017; Fried and Tauer 2019)
    - 信息不对称（卢洪友等，2011；Liu et al., 2019)

双边随机边界模型设定

- 考虑如下双边随机边界模型：

$$
\begin{aligned}
y_i &= x_i{}'\delta + \varepsilon_i, \\
\varepsilon_i &= v_i - u_i + w_i
\end{aligned}
$$

- 通常，双向随机扰动项v_i服从正态分布。
- 单边扰动项应如何设定？

基于分布假设的模型设定与估计

- 考虑如下双边随机边界模型：

$$
\begin{aligned}
y_i &= {x_i}'\delta + \varepsilon_i, \\
\varepsilon_i &= v_i - u_i + w_i
\end{aligned}
$$

- 单边扰动项应如何设定？
  - 分布假设法：假设单边扰动项符合某种分布

基于分布假设的模型设定与估计

- 考虑如下双边随机边界模型：

$$\begin{aligned} y_i &= x_i{}'\delta + \varepsilon_i, \\ \varepsilon_i &= v_i - u_i + w_i \end{aligned}$$

- 参考Kumbhakar and Parmeter (2009)以及卢洪友等(2011)，假设单边扰动项服从指数分布：

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

基于分布假设的模型设定与估计

- 使用最大似然法进行估计
- 在给定扰动项的分布后，可以推导出概率密度函数：

$$
\begin{aligned}
f(\varepsilon_i) &= \frac{e^{\alpha_i}}{\sigma_u+\sigma_w}\Phi(\beta_i)+\frac{e^{a_i}}{\sigma_u+\sigma_w}\int_{-b_i}^{\infty}\phi(z)dz \\
&= \frac{e^{\alpha_i}}{\sigma_u+\sigma_w}\Phi(\beta_i)+\frac{e^{a_i}}{\sigma_u+\sigma_w}\Phi(b_i)
\end{aligned}
$$

其中， $a_i=\frac{\sigma_v^2}{2\sigma_w^2}-\frac{\varepsilon_i}{\sigma_w}$; $b_i=\frac{\varepsilon_i}{\sigma_v}-\frac{\sigma_v}{\sigma_w}$; $\alpha_i=\frac{\varepsilon_i}{\sigma_u}+\frac{\sigma_v^2}{2\sigma_u^2}$; $\beta_i=-\left(\frac{\varepsilon_i}{\sigma_v}+\frac{\sigma_v}{\sigma_u}\right)$

基于分布假设的模型设定与估计

- 使用最大似然法进行估计
- 进而推导出最大似然函数：

$$\ln L(x;\theta) = -n\ln(\sigma_u + \sigma_w) + \sum_{i=1}^{n} \ln\left[e^{\alpha_i}\Phi(\beta_i) + e^{a_i}\Phi(b_i)\right]$$

其中， $\theta = \{\delta, \sigma_v, \sigma_u, \sigma_w\}$。

基于分布假设的模型设定与估计

- 通过最大似然法，我们可以获取模型参数
- 对于两个单边扰动项，还可以进一步解读：
  - 绝对水平的条件预期值

$$E(u_i|\varepsilon_i) = \frac{1}{\lambda} + \frac{e^{\alpha_i - a_i}\sigma_v[\phi(-\beta_i) + \beta_i\Phi(\beta_i)]}{\chi_{1i}},$$

$$E(w_i|\varepsilon_i) = \frac{1}{\lambda} + \frac{\sigma_v[\phi(-b_i) + b_i\Phi(b_i)]}{\chi_{1i}}$$

基于分布假设的模型设定与估计

- 通过最大似然法，我们可以获取模型参数
- 对于两个单边扰动项，还可以进一步解读：
  - 对数水平的条件预期值

$$E(e^{-u_i}|\varepsilon_i)=\frac{\lambda}{1+\lambda}\frac{1}{\chi_{1i}}\left[\Phi(b_i)+e^{\alpha_i-a_i}\cdot e^{\frac{\sigma_v^2}{2}-\sigma_v\beta_i}\Phi(\beta_i-\sigma_v)\right]$$

$$E(e^{-w_i}|\varepsilon_i)=\frac{\lambda}{1+\lambda}\frac{1}{\chi_{2i}}\left[\Phi(\beta_i)+e^{a_i-\alpha_i}\cdot e^{\frac{\sigma_v^2}{2}-\sigma_v b_i}\Phi(b_i-\sigma_v)\right]$$

基于分布假设的模型设定与估计

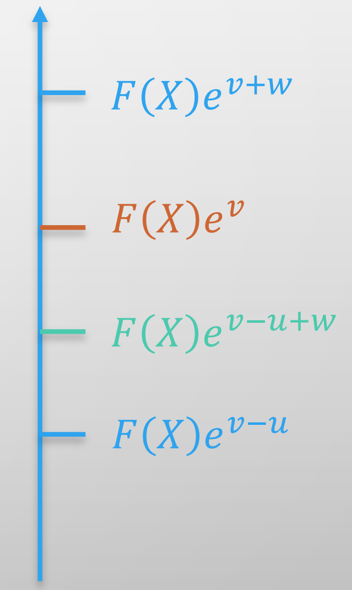
- 通过最大似然法，我们可以获取模型参数
- 对于两个单边扰动项，还可以进一步解读：
  - 对数水平的条件预期值

$$E(e^{w_i}e^{-u_i}|\varepsilon_i) = \exp\left\{\frac{\sigma_w^2+\sigma_u^2}{s^2}\left(\frac{\sigma_v^2}{2}+\varepsilon_i\right)\right\}\times$$

$$\frac{\Phi_2\left(\frac{\varepsilon_i+\sigma_v^2}{\omega_1},0;\rho=\frac{\lambda_1}{\sqrt{1+\lambda_1^2}}\right)-\Phi_2\left(\frac{\varepsilon_i+\sigma_v^2}{\omega_2},0;\rho=\frac{-\lambda_2}{\sqrt{1+\lambda_2^2}}\right)}{\Phi_2\left(\frac{\varepsilon_i}{\omega_1},0;\rho=\frac{\lambda_1}{\sqrt{1+\lambda_1^2}}\right)-\Phi_2\left(\frac{\varepsilon_i}{\omega_2},0;\rho=\frac{-\lambda_2}{\sqrt{1+\lambda_2^2}}\right)}$$

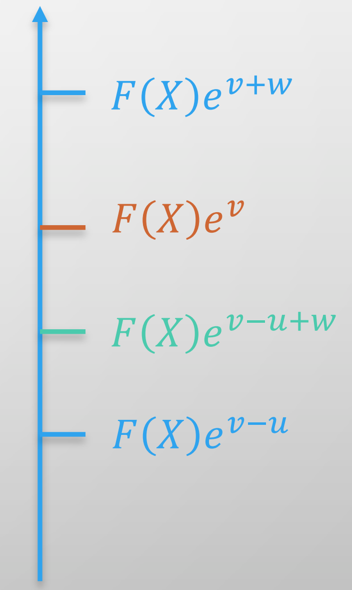
基于分布假设的模型设定与估计

- 单边扰动项的解读
- 在对数价格的模型当中，实际价格为$F(X)e^{v-u+w}$，其中$F(x)=e^{X'\delta}$。
  - 消费者净收益：$\frac{F(X)e^{v}-F(X)e^{v-u+w}}{F(X)e^{v+w}}=e^{-w}-e^{-u}$ (Net gain on consumer surplus)



基于分布假设的模型设定与估计

- 单边扰动项的解读
- 在对数价格的模型当中，实际价格为$F(X)e^{v-u+w}$，其中$F(x)=e^{X'\delta}$。
  - 消费者净收益：$\frac{F(X)e^{v}-F(X)e^{v-u+w}}{F(X)e^{v+w}}=e^{-w}-e^{-u}$
  - 价格偏离程度：$\frac{F(X)e^{v-u+w}-F(X)e^{v}}{F(X)e^{v}}=e^{w}e^{-u}-1$



基于分布假设的模型设定与估计

- 考虑如下双边随机边界模型：

$$\begin{aligned} y_i &= x_i{}'\delta + \varepsilon_i, \\ \varepsilon_i &= v_i - u_i + w_i \end{aligned}$$

- 参考Papadopoulos (2015)，还可以假设单边扰动项服从半正态分布：

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

基于分布假设的模型设定与估计

- 在半正态的分布假设下，最大似然函数和相应的条件均值也会发生改变：

$$\ln L(x;\theta) = -n\ln(\sigma_u + \sigma_w) + \sum_{i=1}^{n} \ln\left[e^{\alpha_i}\Phi(\beta_i) + e^{a_i}\Phi(b_i)\right]$$

$$E(e^{-u_i}|\varepsilon_i) = \frac{\lambda}{1+\lambda}\frac{1}{\chi_{1i}}\left[\Phi(b_i) + e^{\alpha_i - a_i}\cdot e^{\frac{\sigma_v^2}{2}-\sigma_v\beta_i}\Phi(\beta_i - \sigma_v)\right]$$

$$E(e^{-w_i}|\varepsilon_i) = \frac{\lambda}{1+\lambda}\frac{1}{\chi_{2i}}\left[\Phi(\beta_i) + e^{a_i - \alpha_i}\cdot e^{\frac{\sigma_v^2}{2}-\sigma_v b_i}\Phi(b_i - \sigma_v)\right]$$

利用缩放性质的模型设定与估计

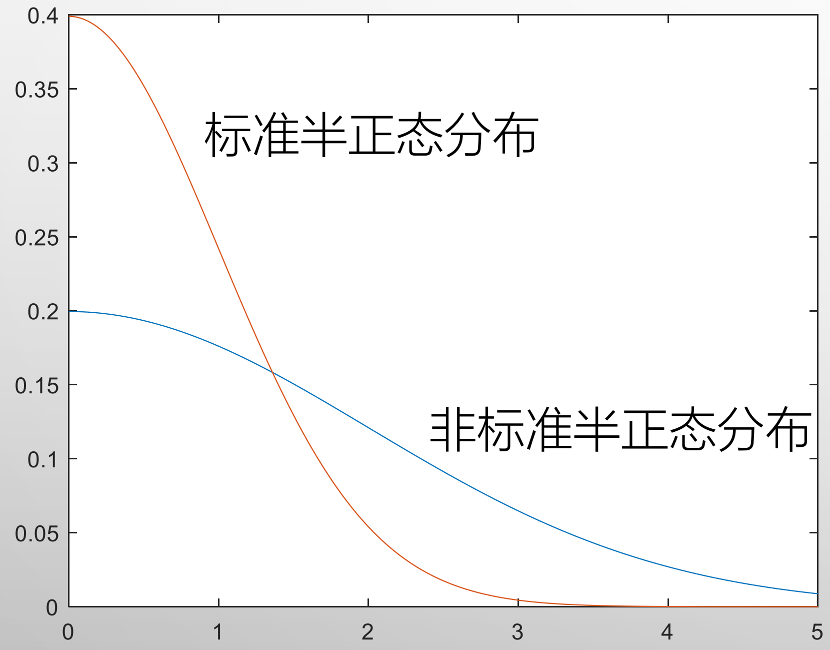
- 考虑如下双边随机边界模型：

$$\begin{aligned} y_i &= x_i{}'\delta + \varepsilon_i, \\ \varepsilon_i &= v_i - u_i + w_i \end{aligned}$$

- 单边扰动项应如何设定？
  - 分布假设法：假设单边扰动项符合某种分布
  - 缩放性质法：利用缩放性质规避分布假设

利用缩放性质的模型设定与估计

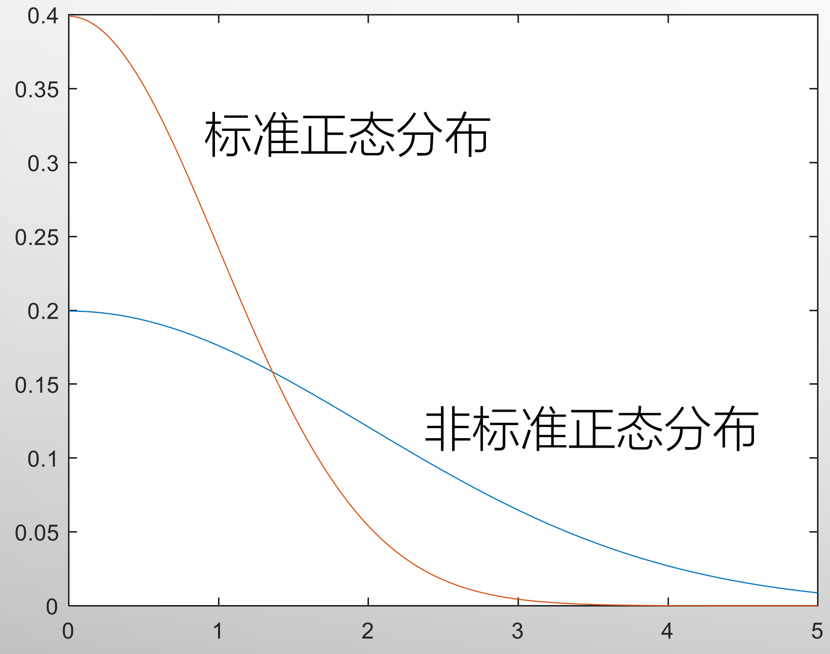
- 缩放性质的思路：
  - 单参数分布的“形状”不会发生改变

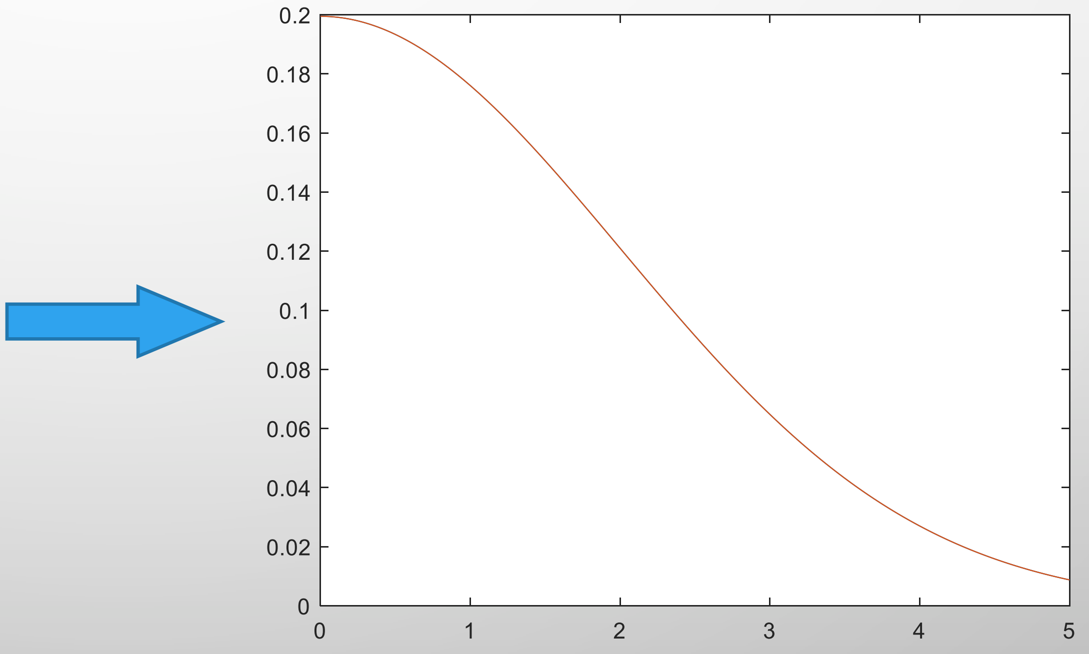


$$
\begin{aligned}
y_i &= x_i'\delta + \varepsilon_i, \\
\varepsilon_i &= v_i - u_i + w_i
\end{aligned}
$$

利用缩放性质的模型设定与估计

- 缩放性质的思路：
  - 单参数分布的“形状”不会发生改变





利用缩放性质的模型设定与估计

- 参考Parmeter (2018)，建立如下双边随机边界模型：

$$y_i = X_i \cdot \delta - u(z_{ui};\delta_u) + w(z_{wi};\delta_w) + v_i$$

- 对于单边扰动项：

$$\begin{aligned} u(z_{ui};\delta_u) &= g_u(z_{ui};\delta_u) \cdot u_i^*, \\ u(z_{wi};\delta_w) &= g_w(z_{wi};\delta_w) \cdot w_i^* \end{aligned}$$

其中，$g(\cdot) > 0$。

利用缩放性质的模型设定与估计

- 令$g(z;\delta)=e^{z'\delta}$：

$$y_i=x_i\delta-e^{z_u\delta_u}u_i^*+e^{z_w\delta_w}w_i^*+v_i$$

- 对两边同时取均值：

$$E[y|x,z_u,z_w]=x'\beta-E(e^{z_u\delta_u}u_i^*)+E(e^{z_w\delta_w}w_i^*)$$

利用缩放性质的模型设定与估计

- 对两边同时取均值：

$$E[y|x,z_u,z_w] = x^{'}\delta - E(e^{z_u\delta_u}u_i^*) + E(e^{z_w\delta_w}w_i^*)$$

- 由于u_i^*和w_i^*与其他变量独立，因此可以提取出来：

$$E[y|x,z_u,z_w] = x^{'}\delta - e^{z_u\delta_u}E(u_i^*) + e^{z_w\delta_w}E(w_i^*)$$

利用缩放性质的模型设定与估计

- 令$\mu_u^* = E(u_i^*), \mu_w^* = E(w_i^*)$：

$$E[y|x, z_u, z_w] = x' \delta - \mu_u^* e^{z_u \delta_u} + \mu_w^* e^{z_w \delta_w}$$

- 使用非线性最小二乘法进行估计

$$(\hat{\delta}, \hat{\delta}_u, \hat{\delta}_w, \hat{\mu}_u^*, \hat{\mu}_w^*) = \min_{\delta, \delta_u, \delta_w, \mu_u^*, \mu_w^*} n^{-1} \sum_{i=1}^{n} [y_i - x_i' \delta + \mu_u^* e^{z_{ui} \delta_u} - \mu_w^* e^{z_{wi} \delta_w}]^2$$

利用缩放性质的模型设定与估计

- 使用缩放性质模型的注意事项：
  - 在样本量较小时，系数估计值偏高
  - 无法计算单边扰动项的条件预期
  - 尚未在文献中得到广泛应用

小结

- 双边随机边界模型
  - 基于分布假设估计模型
    - 指数分布 (Kumbhakar and Parmeter, 2009; 卢洪友等, 2011)
    - 半正态分布 (Papadopoulos, 2015)
    - 分析流程：系数解读、方差分析、单边扰动项分析
  - 利用缩放性质进行分析
    - 方法参考Parmeter (2018)
    - 分析流程：系数解读、方差分析

Stata命令介绍：sftt

- sftt基于连玉君老师SFA2tier命令修改而来
- 在原有功能基础上:
  - 增加了对Papadopoulos (2015)半正态分布模型的支持
  - 增加了对Parmeter (2018)缩放性质模型的支持
  - 改善了命令的使用方法
  - 增加了Owen's T 函数的计算命令

Stata命令介绍：sftt

- 分布假设-指数分布

  sftt y x

- 分布假设-半正态分布

  sftt y x, hnorm

- 缩放性质

  sftt y x, scal

Stata命令介绍：sftt

- 固定效应

```
sftt y x i.year, fe hnormal
```

- 截距项

```
sftt y x, nocons
```

- 随机数种子

```
sftt y x, scal seed(8)
sftt y x, hnorm findseed
```

Stata命令介绍：owenst

- 计算Owen’s T Function

  owenst h a t

Stata命令介绍：sftt

- Stata操作演示

Stata命令介绍：sftt

- 命令的不足之处：
  - 基于半正态分布和缩放性质的模型需要较大样本量，且有可能无法一次得出结果；
  - 半正态情形下可能出现某个扰动项几乎为0的情况。

谢谢！