平成 23 年 05 月 19 日

海洋大　　兵藤 哲朗

確率的フロンティアモデルのパラメータ推定式について

確率的フロンティアモデル（Stochastic Frontier Analysis: SFA）は，企業の生産性や費用構造の効率性を推計する方法論であり，誤差項にその効率性を帰着させるモデル構造を有する．

今，乗法型の生産関数を以下の通り仮定する．

$$\ln(y_i) = \beta_0 + \sum_k \beta_k \ln(x_{ik}) + v_i - u_i \tag{1}$$

ここで，$v_i \approx N(0, \sigma_v^2)$ の通常の正規分布，$u_i \approx N^+(0, \sigma_u^2)$ の positive half normal distribution とする．

すなわち，各々の確率密度関数は，

$$f(v_i) = \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left[-\frac{v_i^2}{2\sigma_v^2}\right] \tag{2}$$

$$f(u_i) = \frac{2}{\sqrt{2\pi}\sigma_u} \exp\left[-\frac{u_i^2}{2\sigma_u^2}\right] \quad \text{ただし，} u_i \geq 0 \tag{3}$$

で相互独立である．さて，最尤法でパラメータ推定を行う．それは $\varepsilon_i = v_i - u_i$ とおき，ε_i の尤度関数を最大化することを意味する．簡単のため，i の表記を省略する．まず，v と u は独立なので，

$$f(u, v) = \frac{2}{2\pi\sigma_u\sigma_v} \exp\left[-\frac{u^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2}\right] \tag{4}$$

である．$v = \varepsilon + u$ を代入し，

$$f(u, \varepsilon) = \frac{2}{2\pi\sigma_u\sigma_v} \exp\left[-\frac{u^2}{2\sigma_u^2} - \frac{(\varepsilon + u)^2}{2\sigma_v^2}\right] \tag{5}$$

なので，(5)式の周辺分布関数から，ε の確率密度関数を導出することができる．

$$\begin{aligned} f(\varepsilon) &= \int_0^\infty f(u, \varepsilon)\, du \\ &= \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v} \exp\left[-\frac{u^2}{2\sigma_u^2} - \frac{(\varepsilon + u)^2}{2\sigma_v^2}\right] du \end{aligned} \tag{6}$$

である．ここで，$\sigma^2 = \sigma_u^2 + \sigma_v^2$，$\lambda = \sigma_u / \sigma_v$ とすれば，(6)式は，

$$\begin{aligned} f(\varepsilon) &= \frac{2}{\sqrt{2\pi}\sigma}\left[1 - \Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right] \exp\left[-\frac{\varepsilon^2}{2\sigma^2}\right] \\ &= \frac{2}{\sigma} \cdot \phi\left(\frac{\varepsilon}{\sigma}\right) \cdot \Phi\left(-\frac{\varepsilon\lambda}{\sigma}\right) \end{aligned} \tag{7}$$

となる．$\phi(\ )$，$\Phi(\ )$は各々，標準正規密度関数，標準正規分布関数である．

しかしながら，(6)式から(7)式を得る過程は複雑であり，説明を要する．丁寧にその過程を確認しよう[1]．

まず，(6)式をそのまま展開していく．

$$\begin{aligned} f(\varepsilon) &= \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v}\exp\left[-\frac{u^2}{2\sigma_u^2}-\frac{(\varepsilon+u)^2}{2\sigma_v^2}\right]du \\ &= \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v}\exp\left[-\frac{u^2}{2\sigma_u^2}-\frac{\varepsilon^2}{2\sigma_v^2}-\frac{2\varepsilon u}{2\sigma_v^2}-\frac{u^2}{2\sigma_v^2}\right]du \\ &= \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v}\exp\left[-\left(\frac{1}{2\sigma_u^2}+\frac{1}{2\sigma_v^2}\right)u^2-\left(\frac{\varepsilon}{\sigma_v^2}\right)u-\left(\frac{\varepsilon^2}{2\sigma_v^2}\right)\right]du \end{aligned} \tag{8}$$

ここで，

$$\frac{1}{2\sigma_u^2}+\frac{1}{2\sigma_v^2}=\frac{\sigma_u^2+\sigma_v^2}{2\sigma_u^2\sigma_v^2}=\frac{\sigma^2}{2\sigma_u^2\sigma_v^2} \tag{9}$$

なので，

$$\begin{aligned} f(\varepsilon) &= \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v}\exp\left[-\frac{\sigma^2}{2\sigma_u^2\sigma_v^2}\left(u^2+\left(\frac{\varepsilon}{\sigma_v^2}\right)\left(\frac{2\sigma_u^2\sigma_v^2}{\sigma^2}\right)u\right)-\left(\frac{\varepsilon^2}{2\sigma_v^2}\right)\right]du \\ &= \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v}\exp\left[-\frac{\sigma^2}{2\sigma_u^2\sigma_v^2}\left(u+\frac{\sigma_u^2}{\sigma^2}\varepsilon\right)^2+\frac{\sigma_u^2\varepsilon^2}{2\sigma^2\sigma_v^2}-\left(\frac{\varepsilon^2}{2\sigma_v^2}\right)\right]du \\ &= \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v}\exp\left[-\frac{\sigma^2}{2\sigma_u^2\sigma_v^2}\left(u+\frac{\sigma_u^2}{\sigma^2}\varepsilon\right)^2-\frac{\varepsilon^2}{2\sigma^2}\right]du \\ &= \left\{\int_0^\infty \frac{1}{\sqrt{2\pi}\left(\frac{\sigma_u\sigma_v}{\sigma}\right)}\exp\left[-\frac{\left(u+\frac{\sigma_u^2}{\sigma^2}\varepsilon\right)^2}{2\left(\frac{\sigma_u\sigma_v}{\sigma}\right)^2}\right]du\right\}\cdot\left\{\frac{2}{\sqrt{2\pi}\sigma}\exp\left[-\frac{\varepsilon^2}{2\sigma^2}\right]\right\} \end{aligned} \tag{10}$$

である．標準正規密度関数は，標準化された，$z \approx N(0,1)$の密度関数であるため，例えば，$x \approx N(\mu,\sigma^2)$であれば，$z=(x-\mu)/\sigma$の標準化式を前提に，

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]=\frac{|\sigma|}{\sqrt{2\pi}\sigma}\exp\left[-\frac{z^2}{2}\right]=\phi(z) \tag{11}$$

と表すことができる．$|\sigma|$は x から z への変数変換に伴うヤコビアンである．すると，(10)式最下段の 2 番目の括弧内の式は，

$$\left\{\frac{2}{\sqrt{2\pi}\sigma}\exp\left[-\frac{\varepsilon^2}{2\sigma^2}\right]\right\}=2\cdot\phi\left(\frac{\varepsilon}{\sigma}\right) \tag{12}$$

であることが分かる．

さて，(10)式最下段の最初の括弧内であるが，$\sigma_u^2=\lambda^2\sigma^2/(\lambda^2+1)$なので，これを代入し，

[1] この式展開は，兵藤の手に余り，本学の小杉のぶ子先生にアドバイス頂いた．

$$\int_0^\infty \frac{1}{\sqrt{2\pi}\left(\frac{\sigma_u\sigma_v}{\sigma}\right)} \exp\left[-\frac{\left(u+\frac{\sigma_u^2}{\sigma^2}\varepsilon\right)^2}{2\left(\frac{\sigma_u\sigma_v}{\sigma}\right)^2}\right] du = \int_0^\infty \frac{1}{\sqrt{2\pi}\left(\frac{\sigma_u\sigma_v}{\sigma}\right)} \exp\left[-\frac{\left(u+\frac{\lambda^2}{\lambda^2+1}\varepsilon\right)^2}{2\left(\frac{\sigma_u\sigma_v}{\sigma}\right)^2}\right] du \tag{13}$$

なので，これは，

$$u \approx N\left(-\frac{\lambda^2}{\lambda^2+1}\varepsilon, \left(\frac{\sigma_u\sigma_v}{\sigma}\right)^2\right) \tag{14}$$

と見なすことができる．まず積分区間に留意して，

$$\int_0^\infty \frac{1}{\sqrt{2\pi}\left(\frac{\sigma_u\sigma_v}{\sigma}\right)} \exp\left[-\frac{\left(u+\frac{\lambda^2}{\lambda^2+1}\varepsilon\right)^2}{2\left(\frac{\sigma_u\sigma_v}{\sigma}\right)^2}\right] du = \int_{\frac{\lambda^2}{\lambda^2+1}\varepsilon}^\infty \frac{1}{\sqrt{2\pi}\left(\frac{\sigma_u\sigma_v}{\sigma}\right)} \exp\left[-\frac{u^2}{2\left(\frac{\sigma_u\sigma_v}{\sigma}\right)^2}\right] du \tag{15}$$

とおく．すると，これは

$$u \approx N\left(0, \left(\frac{\sigma_u\sigma_v}{\sigma}\right)^2\right) \tag{14}$$

に変換したことになる．(15)式の積分開始の値を標準化すると，

$$z = \frac{\lambda^2}{\lambda^2+1}\varepsilon \cdot \frac{\sigma}{\sigma_u\sigma_v} = \frac{\lambda\varepsilon}{\sigma} \tag{15}$$

であるため，

$$\int_{\frac{\lambda^2}{\lambda^2+1}\varepsilon}^\infty \frac{1}{\sqrt{2\pi}\left(\frac{\sigma_u\sigma_v}{\sigma}\right)} \exp\left[-\frac{u^2}{2\left(\frac{\sigma_u\sigma_v}{\sigma}\right)^2}\right] du = 1-\Phi\left(\frac{\lambda\varepsilon}{\sigma}\right) \tag{16}$$

を得る．

以上より，(10)，(12)，(16)式から(7)式が導かれることが理解できる．

では，SFA の挙動を R でプログラムを作成し，確認してみよう．

簡単に，右の 14 サンプルのデータを用いる．見て明らかなとおり，最初の 10 サンプルは線形に比例し，残りの 4 サンプルはその線を下回っている．すなわち，生産関数における非効率性が認められる．このデータについて，通常最小二乗法（OLS）と，SFA の二通りの結果を推定する R プログラムは以下の通り．

```
x,y
1,11
2,12
3,13
4,14
5,15
6,16
7,17
8,18
9,19
10,20
2,8
4,7
6,13
8,15
```

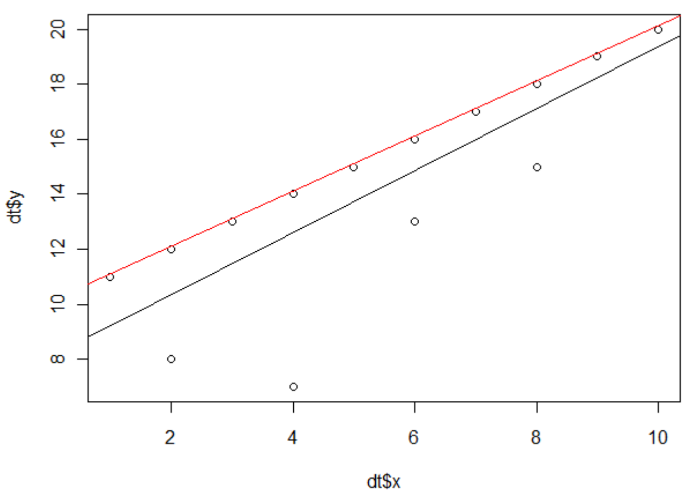
```
dt <- read.csv("c:/usr/frontier/data.csv",header=TRUE)
b1 <- c(1,8)
b2 <- c(1,8,2,1)

ols <- function(beta){
  res <- dt[,2]-(dt[,1]*beta[1]+beta[2])
  LL <- sum( log( dnorm(res,mean=0,sd=1) ) )
  return(LL)
}

sfa <- function(beta){
  res <- dt[,2]-(dt[,1]*beta[1]+beta[2])
  LL <- sum( log( 2/beta[3]*dnorm(res/beta[3],mean=0,sd=1)*
             pnorm(-res*beta[4]/beta[3],mean=0,sd=1) ) )
  return(LL)
}

est_ols <- optim(b1,ols,method="BFGS",hessian=TRUE,control=list(fnscale=-1) )
est_sfa <- optim(b2,sfa,method="BFGS",hessian=TRUE,control=list(fnscale=-1) )
plot(dt$x,dt$y)
abline(coef=c(est_ols$par[2],est_ols$par[1]))
abline(coef=c(est_sfa$par[2],est_sfa$par[1]),col="red")
```

推定結果を，OLS を黒線，SFA を赤線で示す．SFA では，綺麗に正比例部分が推定されており，非効率なサンプルとの分離推定がなされていることが確認できる．



以上