非線形最小二乗法†

担当: 長倉 大輔
（ながくらだいすけ）

† この資料は私のゼミにおいて使用するために作成した資料です。ゼミのWEBページ上で公開しており、自由に参照して頂いて構いません。ただし、内容について、一応検証してありますが、もし間違いがあった場合でもそれによって生じるいかなる損害、不利益について責任を負いかねますのでご了承ください。間違いは発見次第、継続的に直していますが、まだ存在する可能性があります。

非線形最小二乗法

- 非線形モデル

スカラーの被説明変数 y_i , $i = 1,\ldots, n$ に対して、次のようなモデルを考える。

$$y_i = m\,(\mathbf{x}_i, \boldsymbol{\theta}) + \varepsilon_i\,, \quad \varepsilon_i \sim \text{i.i.d.}\ (0, \sigma^2)$$

ここで $\mathbf{x}_i$ は $l \times 1$ 説明変数ベクトル、$\boldsymbol{\theta}$ は $s \times 1$ 未知パラメーターベクトル、$m\,(.\,,\,.)$ は説明変数ベクトルと未知パラメーターベクトルの関数で、スカラーであり、誤差項 ε_i は期待値 0, 分散 σ^2 の i.i.d. 確率変数とする。

非線形最小二乗法は観測値$\{y_i, \mathbf{x}_i\}$, $i =1,\ldots,n$ をもとに未知パラメーターベクトル $\boldsymbol{\theta}$ を推定する方法。

非線形最小二乗法

- 非線形モデルのいくつかの例

「コブ･ダグラス型生産関数 + 加算的誤差項 」

y_i を 生産量、L_i を労働投入量、K_i を資本投入量とする。
y_i は コブ･ダグラス型生産関数(+ 加法的誤差項)によって決定されるとすると

$$y_i = \beta_1 K_i^{\beta_2} L_i^{\beta_3} + \varepsilon_i$$

というモデルが考えられる。これは先ほどのモデルにおいて $\mathbf{x}_i = [L_i, K_i]^{\mathrm{T}}$ 、$\boldsymbol{\theta} = [\beta_1, \beta_2, \beta_3]^{\mathrm{T}}$、$m(\mathbf{x}_i, \boldsymbol{\theta}) = \beta_1 K_i^{\beta_2} L_i^{\beta_3}$ とした場合に相当する。

非線形最小二乗法

「CES (constant elasticity of substitution ; 代替の弾力性一定) 生産関数」

先ほどと同様に y_i , $\mathbf{x}_i$ を定義する。y_i が CES生産関数(+加算的誤差項)によって決定されるのであれば

$$y_i = \beta_1[\beta_2 K_i^{-\beta_3} + (1-\beta_2)L_i^{-\beta_3}]^{-\beta_4/\beta_3} + \varepsilon_i$$

というモデルとなる。これは先ほどのモデルにおいて、$\boldsymbol{\theta} = [\beta_1, \beta_2, \beta_3, \beta_4]^{\mathrm{T}}$、$m(\mathbf{x}_i, \boldsymbol{\theta}) = \beta_1[\beta_2 K_i^{-\beta_3} + (1-\beta_2)L_i^{-\beta_3}]^{-\beta_4/\beta_3}$ とした場合に相当する。

またその他にもいろいろな非線形な生産関数が考えられる。

非線形最小二乗法

「誤差項が AR(1) モデルに従う 線形回帰モデル」

$$
\begin{aligned}
y_t &= \beta x_t + u_t , \\
u_t &= \rho u_{t-1} + \varepsilon_t , \quad \varepsilon_t \sim \text{i.i.d.}\ (0, \sigma^2)
\end{aligned}
$$

というモデルを考える。u_t が AR(1) モデルに従っている。
ここで、$u_t = y_t - \beta x_t$ を下の式に代入して整理すると

$$
y_t = \beta x_t + \rho y_{t-1} - \rho\beta x_{t-1} + \varepsilon t
$$

となる。これは先程のモデルにおいて $\mathbf{x}_i = [x_t, y_{t-1}]^{\mathrm{T}}$,
$\boldsymbol{\theta} = [\beta, \rho]^{\mathrm{T}}$, $m(\mathbf{x}_i, \boldsymbol{\theta}) = \beta x_t + \rho y_{t-1} + \rho\beta x_{t-1}$
である場合に相当する。

非線形最小二乗法

- 何が非線形なのか?

非線形最小二乗法で推定する必要があるモデルというのはモデルが**パラメーターに関して非線形**となっているモデルである。例えば次のモデルを考えよう。

$$y_i = \alpha + \beta \log x_i + \varepsilon_i, \quad \varepsilon_i \sim \text{i.i.d.}\,(0, \sigma^2)$$

このモデルにおいて y_i と x_i と ε_i は非線形の関係にある。しかしながら、$z_i = \log x_i$ と置けば、

$$y_i = \alpha + \beta z_i + \varepsilon_i$$

と書け、y_i と z_i と ε_i は線形の関係であるから、z_i を説明変数とすれば、これは通常の最小二乗法で推定できる。

非線形最小二乗法

では次に以下のモデルを考えよう。

$$y_i = \alpha + x_i^{\beta} + \varepsilon_i, \quad \varepsilon_i \sim \text{i.i.d.}\ (0, \sigma^2)$$

この場合、β と y_i, ε_i とは非線形な関係にあり、また線形になるように変換もできない。この場合、最小二乗法では β と α を(少なくとも同時に)推定することはできない。

このような場合は非線形最小二乗法で α と β を推定する。

非線形最小二乗法

- 変換によって線形にする

もともとのモデルの形によっては、簡単な変換でパラメーターについて線形なモデルにすることができる。

例えば、最初に出てきた コブ･ダグラス型生産関数のモデルにおいて、誤差項が加算的ではなく**乗算的**な場合、すなわち

$$y_i = \varepsilon_i \, \beta_1 K_i^{\beta_2} L_i^{\beta_3}$$

のような形で入っている場合、もし $\varepsilon_i > 0$ が常に成り立てば両辺の対数をとって

非線形最小二乗法

$$\log y_i = \log \beta_1 + \beta_2 \log K_i + \beta_3 \log L_i + u_i,$$
$$u_i = \log \varepsilon_i$$

と表す事が出来る。これはパラメータを $\alpha = \log \beta_1$ と定義しなおせば、パラメーター α、β_2, β_3 について線形なモデルなので、最小二乗法で推定できる。

上記はまた「推定モデル」というのは**誤差項がどのようにモデル入るのかも含めて記述する必要がある**ということを示唆している。誤差項を含めて記述されていないと、その論文を読んでいる読者がどのようなモデルがどのように推定されたかわからなくなり混乱をきたす可能性が高まる。

非線形最小二乗法

- 非線形最小二乗法

非線形最小二乗法とは以下の**残差平方和**を $\boldsymbol{\theta}$ について最小にするような推定方法のことである。

(非線形最小二乗法の残差平方和 SSR)

$$SSR(\boldsymbol{\theta}) = \sum_{i=1}^{n} (y_i - m(\mathbf{x}_i, \boldsymbol{\theta}))^2$$

この残差平方和を最小にする $\boldsymbol{\theta}$ が**非線形最小二乗推定量(値)**である。線形モデルの最小二乗法の場合と異なり、この解は一般的には明示的な解をもたないので、通常、数値的に解を求める。

非線形最小二乗法

- 最小化のための1階の条件

先ほどの SSR を最小にするための1階の条件は方程式

$$\frac{\partial SSR(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 2\sum_{i=1}^{n}(y_i - m(\mathbf{x}_i, \boldsymbol{\theta}))\frac{\partial m(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$$

である。この非線形方程式を満たす $\boldsymbol{\theta}$ が非線形最小二乗推定量(値)である。

非線形最小二乗法

- 数値計算

ある(スカラー)目的関数 $g(\boldsymbol{\alpha})$ を変数 $\boldsymbol{\alpha}$ について最小化する場合、そのような $\boldsymbol{\alpha}$ を求める数値計算のアルゴリズムはだいたい次のようなステップを踏む。

(ステップ1)　$\boldsymbol{\alpha}$ の**初期値** $\boldsymbol{\alpha}^{(0)}$ を**適当に**決める。
(ステップ2)　ある数値計算アルゴリズムにより $\boldsymbol{\alpha}^{(0)}$を $\boldsymbol{\alpha}^{(1)}$ に更新する(通常、$g(\boldsymbol{\alpha}^{(1)})<g(\boldsymbol{\alpha}^{(0)})$が満たされ $\boldsymbol{\alpha}^{(1)}$は $\boldsymbol{\alpha}^{(0)}$より(何らかの意味で)より解に近くなるように更新する)。
(ステップ3)　(通常 ステップ2 と 同じ)数値計算アルゴリズムで $\boldsymbol{\alpha}^{(1)}$ から $\boldsymbol{\alpha}^{(2)}$、さらに同様に $\boldsymbol{\alpha}^{(2)}$から $\boldsymbol{\alpha}^{(3)}$ ….. , $\boldsymbol{\alpha}^{(j)}$へとどんどん更新していく。

非線形最小二乗法

(ステップ4) このように更新を繰り返していき、ある基準をみたしたところで更新を停止し、その値を解とする

ステップ4 において、更新を停止する基準はそこでの値 $\boldsymbol{\alpha}^{(j)}$ が真の解に十分近くなったということが保障されるように決める。例えば、非常に小さな e に対して ($e = 10^{-6} \sim 10^{-15}$ など)

(i) $g(\boldsymbol{\alpha}^{(j-1)}) - g(\boldsymbol{\alpha}^{(j)}) < e$

(ii) $\mathrm{dis}(\boldsymbol{\alpha}^{(j-1)}, \boldsymbol{\alpha}^{(j)}) < e$

(iii) $\boldsymbol{\alpha}^{(j)}$において $\partial g/\partial \boldsymbol{\alpha} \approx \mathbf{0}$

などが用いられる。ここで dis(. , .) は2つのベクトルの(何らかの意味での)近さ(距離)を測るものとする。

非線形最小二乗法

- σ^2 の推定

$\boldsymbol{\theta}$ をこのように推定し、σ^2はこの推定値 $\hat{\boldsymbol{\theta}}$ を用いて

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - m(\mathbf{x}_i, \hat{\boldsymbol{\theta}}))^2$$

と推定できる。