一般化最小二乗法†

担当: 長倉大輔

† 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。

一般化最小二乗法

- 一般化最小二乗法

一般化最小二乗法について説明する。
次の(行列で表現した) 回帰式を考えよう。

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

ここで $\mathbf{Y}$ は $N \times 1$ベクトル、$\mathbf{X}$ は $N \times K$ ベクトル、$\boldsymbol{\beta}$ は $K \times 1$ ベクトル、$\boldsymbol{\varepsilon}$ は $N \times 1$ベクトルである。

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}, \mathbf{X} = \begin{bmatrix} X_{11} & \cdots & X_{1K} \\ X_{21} & \cdots & X_{2K} \\ \vdots & \vdots & \vdots \\ X_{N1} & \cdots & X_{NK} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

一般化最小二乗法

- 一般化最小二乗法

通常、最小二乗法により推定する場合は誤差項 $\varepsilon_1, \ldots, \varepsilon_N$ は互いに無相関($\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$)で、均一な分散 ($\mathrm{var}(\varepsilon_i) = \sigma^2$) を持つと仮定される。

この時、$\boldsymbol{\varepsilon}$ の分散共分散行列 $\boldsymbol{\Sigma} = E[\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}^{\mathrm{T}}]$は

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_N$$

となる。ここで $\mathbf{I}_N$ は $N \times N$の単位行列である。

一般化最小二乗法

しかしながら実際のデータはこの仮定を満たさない場合も多い。このような時、最小二乗法は**一致性は持つ** (!)が、一般化最小二乗法の方が漸近的には**より効率的**である事が知られている。一般化最小二乗法は誤差項の分散共分散行列として、一般的な形

$$\boldsymbol{\Sigma} = \boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{1N} & \sigma_{2N} & \cdots & \sigma_N^2 \end{bmatrix}$$

を仮定する(分散共分散行列は**対称行列**である事に注意)

一般化最小二乗法

一般化最小二乗法を理解するには若干の行列に関する知識が必要である(とはいえ結果だけを知っておけばよいので難しくはない)。以下の事実が重要である。

「分散共分散行列 $\mathbf{\Omega}$ ($N \times N$行列) は正定値対称行列であるので、逆行列 $\mathbf{\Omega}^{-1}$が存在し、さらに $\mathbf{\Omega}^{-1} = \mathbf{\Psi}\mathbf{\Psi}^{\mathrm{T}}$となる$N \times N$行列 $\mathbf{\Psi}$ が存在する。$\mathbf{\Psi}$ は逆行列 $\mathbf{\Psi}^{-1}$を持つ。」

(演習問題2参照)

一般化最小二乗法

このような行列 $\mathbf{\Psi}^{\mathrm{T}}$ を先ほどの回帰式の左から掛けて、

$$\mathbf{\Psi}^{\mathrm{T}}\mathbf{Y} = \mathbf{\Psi}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} + \mathbf{\Psi}^{\mathrm{T}}\boldsymbol{\varepsilon}$$

を得る。$\tilde{\mathbf{Y}} = \mathbf{\Psi}^{\mathrm{T}}\mathbf{Y}$, $\tilde{\mathbf{X}} = \mathbf{\Psi}^{\mathrm{T}}\mathbf{X}$, $\tilde{\boldsymbol{\varepsilon}} = \mathbf{\Psi}^{\mathrm{T}}\boldsymbol{\varepsilon}$, とおくと上記の式は

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}$$

という回帰式を表している。

一般化最小二乗法

この回帰式の中の誤差項 $\tilde{\boldsymbol{\varepsilon}}$ の分散共分散行列は

$$
\begin{aligned}
E[\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^{\mathrm{T}}] = E[\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\Psi}] &= \boldsymbol{\Psi}^{\mathrm{T}}E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}]\boldsymbol{\Psi} \\
&= \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\Psi} \\
&= \boldsymbol{\Psi}^{\mathrm{T}}(\boldsymbol{\Omega}^{-1})^{-1}\boldsymbol{\Psi} \\
&= \boldsymbol{\Psi}^{\mathrm{T}}(\boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathrm{T}})^{-1}\boldsymbol{\Psi} \\
&= \boldsymbol{\Psi}^{\mathrm{T}}(\boldsymbol{\Psi}^{\mathrm{T}})^{-1}\boldsymbol{\Psi}^{-1}\boldsymbol{\Psi} \\
&= \mathbf{I}_N
\end{aligned}
$$

となり、**相関なし均一分散の最小二乗法の仮定を満たす！**

一般化最小二乗法

よって

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}$$

に最小二乗法を適用すれば $\boldsymbol{\beta}$ の効率的な推定量が得られる。

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{GLS} &= (\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{Y}} \\ &= (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathrm{T}}\mathbf{Y} \\ &= (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{Y}\end{aligned}$$

この $\hat{\boldsymbol{\beta}}_{GLS}$ を**一般化最小二乗推定量(generalized least square estimator; GLS)** という。

一般化最小二乗法

- 実行可能な一般化最小二乗法

一般化最小二乗推定量を計算するには $\mathbf{\Omega}^{-1}$ が分かっていなければならないが、実際にこれが前もってわかっている事は、いくつかの特殊ケースを除き、ほとんどない。このような場合 $\mathbf{\Omega}^{-1}$ は**その推定値** $\hat{\mathbf{\Omega}}^{-1}$ で置き換えられる。$\hat{\boldsymbol{\beta}}_{GLS}$ の$\mathbf{\Omega}^{-1}$をその推定値で置き換えたもの:

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}^{\mathrm{T}}\hat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{\Omega}}^{-1}\mathbf{Y}$$

を**実行可能な一般化最小二乗法 (feasible generalized least square; FGLS)** という。$\hat{\mathbf{\Omega}}^{-1}$ の推定法についてはここでは立ち入らない。

演習問題

問題1

次ページのスライドにあるように、Ωは三角分解によって $\mathbf{\Omega} = \mathbf{A}\mathbf{D}\mathbf{A}^{\mathrm{T}}$ と表すことができる。この時 $\mathbf{\Psi}$ は $\mathbf{A}$ と $\mathbf{D}^{1/2}$ を用いてどのように表す事ができるか答えなさい。

多変量時系列モデルその他

- 三角分解

正定値行列 $\mathbf{\Sigma}$ は $\mathbf{A} = [a_{ij}]$, $a_{ij} = 1\ (i = j)$, $a_{ij} = 0\ (i < j)$ であるような行列 $\mathbf{A}$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & 1 \end{bmatrix}$$

を用いて(このような行列を**下三角行列**という)、

$$\mathbf{\Sigma} = \mathbf{ADA}^{\mathrm{T}}$$

と書ける。ここで $\mathbf{D}$ はある対角行列である。
$\mathbf{\Sigma}$ をこのように書くことを**三角分解**するという。