SUR モデルにおける偏回帰係数ベクトルの
リスク行列の下での最適な推定量について

慶應義塾大・理工　松浦　峻　　　東京大・総合文化　倉田 博史

以下のような互いに相関がある p 個の重回帰式を想定する．

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i,\ i=1,\ldots,p \quad \text{with} \quad \begin{aligned} &E[\boldsymbol{\varepsilon}_i] = \boldsymbol{0}_m,\ V[\boldsymbol{\varepsilon}_i] = \sigma_i^2\boldsymbol{I}_m,\ i=1,\ldots,p, \\ &E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j'] = \sigma_{ij}\boldsymbol{I}_m = \sigma_i\sigma_j r_{ij}\boldsymbol{I}_m,\ i \neq j. \end{aligned}$$

各重回帰式の説明変数の数（$\boldsymbol{\beta}_i$ の次元）を $k_i,\ i=1,\ldots,p$ とおく．これは，見かけ上無関係な回帰（Seemingly Unrelated Regression, SUR）モデルと呼ばれるものである．p 個の重回帰式をまとめることで，以下のようにシンプルに表すことができる．

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{with} \quad E[\boldsymbol{\varepsilon}] = \boldsymbol{0}_{mp},\ V[\boldsymbol{\varepsilon}] = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_m.$$

ただし，$\boldsymbol{y} = (\boldsymbol{y}_1',\ldots,\boldsymbol{y}_p')'$，$\boldsymbol{X} = \mathrm{diag}\{\boldsymbol{X}_1,\ldots,\boldsymbol{X}_p\}$，$\boldsymbol{\beta} = (\boldsymbol{\beta}_1',\ldots,\boldsymbol{\beta}_p')'$，$\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1',\ldots,\boldsymbol{\varepsilon}_p')'$，$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_d\boldsymbol{\Lambda}\boldsymbol{\Sigma}_d$，$\boldsymbol{\Sigma}_d = \mathrm{diag}\{\sigma_1,\ldots,\sigma_p\}$，$\boldsymbol{\Lambda} = (r_{ij})_{1\le i,j\le p}$（$r_{ii}=1,\ i=1,\ldots,p$）である．

Kurata and Matsuura (2016) は，誤差ベクトル $\boldsymbol{\varepsilon}$ が楕円対称分布の確率密度関数

$$g(\boldsymbol{\varepsilon}) = |\boldsymbol{\Sigma} \otimes \boldsymbol{I}_m|^{-1/2} h\left(\boldsymbol{\varepsilon}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_m)\boldsymbol{\varepsilon}\right) \quad \text{for some known function} \quad h : [0,\infty) \to [0,\infty) \tag{1}$$

を持ち，相関係数行列 $\boldsymbol{\Lambda}$ が既知であるとき，損失関数 $L(\hat{\boldsymbol{\beta}},(\boldsymbol{\beta},\boldsymbol{\Sigma})) = (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{I}_m)\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ の下での偏回帰係数ベクトル $\boldsymbol{\beta}$ の最良共変推定量を与えている．ここでの共変推定量は次を満たす推定量である．

$$\left\{\begin{array}{ll} \boldsymbol{y}_i \to a_i\boldsymbol{y}_i + \boldsymbol{X}_i\boldsymbol{c}_i, & i=1,\ldots,p, \\ \hat{\boldsymbol{\beta}}_i \to a_i\hat{\boldsymbol{\beta}}_i + \boldsymbol{c}_i, & i=1,\ldots,p, \end{array}\right. \quad \text{for any} \quad a_i \in (0,\infty),\ \boldsymbol{c}_i \in \Re^{k_i},\ i=1,\ldots,p. \tag{2}$$

この変換の下で $\boldsymbol{\Lambda}$ は最大不変量パラメータとなり，従って，これは最大不変量パラメータが既知のモデルにおける推定問題とみることができる．また，Matsuura and Kurata (2019) は，分散共分散行列 $\boldsymbol{\Sigma}$ に関する共変推定について議論している．

本発表では，Kurata and Matsuura (2016) と同様に，誤差ベクトル $\boldsymbol{\varepsilon}$ が楕円対称分布の確率密度関数 (1) を持ち，相関係数行列 $\boldsymbol{\Lambda}$ が既知であるという仮定をおく．一つの重回帰式，ここでは一般性を失うことなく第 1 番目の重回帰式の偏回帰係数ベクトル $\boldsymbol{\beta}_1$ について，リスク行列

$$\boldsymbol{R}_1(\hat{\boldsymbol{\beta}}_1) = \frac{1}{\sigma_1^2}E\left[\left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\right)\left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\right)'\right]$$

の下での最良共変推定量を導出する．具体的には，共変性 (2) を満たす任意の推定量 $\hat{\boldsymbol{\beta}}_1(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_p) : \Re^{mp} \to \Re^{k_1}$ に対し，

$$\boldsymbol{R}_1(\hat{\boldsymbol{\beta}}_1) \ge \boldsymbol{R}_1(\hat{\boldsymbol{\beta}}_1^*)$$

が成り立つような推定量 $\hat{\boldsymbol{\beta}}_1^*$ を導出する．ただし，ここでの $\ge$ は，左辺の行列から右辺の行列を引くと非負定符号行列になることを意味している．発表当日は，数値例を合わせて紹介する．

参考文献

Kurata, H., Matsuura, S. (2016). Best equivariant estimator of regression coefficients in a seemingly unrelated regression model with known correlation matrix, *Annals of the Institute of Statistical Mathematics*, Vol. 68, No. 4, 705-723.

Matsuura, S., Kurata, H. (2019). Covariance matrix estimation in a seemingly unrelated regression model under Stein's loss, *Statistical Methods and Applications* (in press).