MAT-266: Restricciones lineales estocásticas

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

El estimador de $\boldsymbol{\beta}$ sujeto a las restricciones lineales:

$$\boldsymbol{g} = \boldsymbol{G}\boldsymbol{\beta}, \tag{1}$$

adopta la forma,[1]

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{G}) = \widehat{\boldsymbol{\beta}} + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top(\boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top)^{-1}(\boldsymbol{g} - \boldsymbol{G}\widehat{\boldsymbol{\beta}})$$

Sabemos que cuando (1) se satisface, tenemos:

$$\mathsf{E}(\widehat{\boldsymbol{\beta}}(\boldsymbol{G})) = \boldsymbol{\beta} + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top(\boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top)^{-1}(\boldsymbol{g} - \boldsymbol{G}\boldsymbol{\beta}),$$

y

$$\mathsf{Cov}(\widehat{\boldsymbol{\beta}}(\boldsymbol{G})) = \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1} - \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top(\boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top)^{-1}\boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}.$$

[1] Anteriormente anotamos el MLE restringido como $\widetilde{\boldsymbol{\beta}}$.

Es posible notar que:

$$\begin{aligned}&\mathsf{Cov}(\widehat{\boldsymbol{\beta}}) - \mathsf{Cov}(\widehat{\boldsymbol{\beta}}(\boldsymbol{G}))\\ &\qquad = \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top(\boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top)^{-1}\boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \geq \mathbf{0}.\end{aligned}$$

Observación:

Es decir, usar las restricciones lineales en (1) lleva a una ganancia en eficiencia.

Ahora considere,

$$\boldsymbol{g} = \boldsymbol{G}\boldsymbol{\beta} + \boldsymbol{u}, \qquad \boldsymbol{u} \sim \mathsf{N}_q(\boldsymbol{0}, \sigma^2 \boldsymbol{V}),$$

con $\boldsymbol{V} > \boldsymbol{0}$ matriz conocida.

Asumiremos también que $\mathsf{Cov}(\boldsymbol{u}, \boldsymbol{\epsilon}) = \boldsymbol{0}$ $(= \mathsf{E}(\boldsymbol{u}\boldsymbol{\epsilon}^\top))$. Podemos notar además,

$$\mathsf{E}(\boldsymbol{g}) = \boldsymbol{G}\boldsymbol{\beta}.$$

Así podemos escribir el modelo mixto:

$$\begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{g} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{G} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{u} \end{pmatrix},$$

donde

$$\begin{pmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{u} \end{pmatrix} \sim \mathsf{N}_{n+q}\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \sigma^2 \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V} \end{pmatrix} \right).$$

Considere las matrices aumentadas:

$$\boldsymbol{Y}_a = \begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{g} \end{pmatrix}, \qquad \boldsymbol{X}_a = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{G} \end{pmatrix}, \qquad \boldsymbol{\epsilon}_a = \begin{pmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{u} \end{pmatrix}.$$

De este modo, podemos escribir el modelo mixto como:

$$\boldsymbol{Y}_a = \boldsymbol{X}_a\boldsymbol{\beta} + \boldsymbol{\epsilon}_a, \qquad \boldsymbol{\epsilon}_a \sim \mathsf{N}_{n+q}(\boldsymbol{0}, \sigma^2\boldsymbol{W}), \tag{2}$$

con

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V} \end{pmatrix}.$$

Note que el modelo en (2) tiene función de log-verosimilitud

$$\ell(\boldsymbol{\theta}) = -\frac{n+q}{2}\log\sigma^2 - \frac{1}{2}\log|2\pi\boldsymbol{W}| - \frac{1}{2\sigma^2}(\boldsymbol{Y}_a - \boldsymbol{X}_a\boldsymbol{\beta})^\top\boldsymbol{W}^{-1}(\boldsymbol{Y}_a - \boldsymbol{X}_a\boldsymbol{\beta}).$$

De este modo,

$$\mathsf{d}_\beta\,\ell(\boldsymbol{\theta}) = \frac{1}{\sigma^2}(\boldsymbol{Y}_a - \boldsymbol{X}_a\boldsymbol{\beta})^\top\boldsymbol{W}^{-1}\boldsymbol{X}_a\,\mathsf{d}\,\boldsymbol{\beta}.$$

Resolviendo la condición de primer orden, obtenemos

$$\widehat{\boldsymbol{\beta}}_{\mathsf{ME}}(\boldsymbol{G}) = (\boldsymbol{X}_a^\top\boldsymbol{W}^{-1}\boldsymbol{X}_a)^{-1}\boldsymbol{X}_a^\top\boldsymbol{W}^{-1}\boldsymbol{Y}_a$$

Es fácil notar que

$$\boldsymbol{X}_a^\top \boldsymbol{W}^{-1}\boldsymbol{X}_a = (\boldsymbol{X}^\top, \boldsymbol{G}^\top)\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}^{-1}\end{pmatrix}\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{G}\end{pmatrix} = \boldsymbol{X}^\top\boldsymbol{X} + \boldsymbol{G}^\top\boldsymbol{V}^{-1}\boldsymbol{G},$$

$$\boldsymbol{X}_a^\top \boldsymbol{W}^{-1}\boldsymbol{Y}_a = (\boldsymbol{X}^\top, \boldsymbol{G}^\top)\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}^{-1}\end{pmatrix}\begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{g}\end{pmatrix} = \boldsymbol{X}^\top\boldsymbol{Y} + \boldsymbol{G}^\top\boldsymbol{V}^{-1}\boldsymbol{g}.$$

Es decir,

$$\widehat{\boldsymbol{\beta}}_{\mathsf{ME}}(\boldsymbol{G}) = (\boldsymbol{X}^\top\boldsymbol{X} + \boldsymbol{G}^\top\boldsymbol{V}^{-1}\boldsymbol{G})^{-1}(\boldsymbol{X}^\top\boldsymbol{Y} + \boldsymbol{G}^\top\boldsymbol{V}^{-1}\boldsymbol{g}).$$

Esto lleva al siguiente resultado.

Resultado 1 (estimador mixto):

En el modelo mixto (2), el mejor estimador lineal e insesgado es:

$$\widehat{\boldsymbol{\beta}}_{\mathsf{ME}}(\boldsymbol{G}) = \widehat{\boldsymbol{\beta}} + \boldsymbol{S}^{-1}\boldsymbol{G}^{\top}(\boldsymbol{V} + \boldsymbol{G}\boldsymbol{S}^{-1}\boldsymbol{G}^{\top})^{-1}(\boldsymbol{g} - \boldsymbol{G}\widehat{\boldsymbol{\beta}}),$$

con matriz de covarianza

$$\mathsf{Cov}(\widehat{\boldsymbol{\beta}}_{\mathsf{ME}}(\boldsymbol{G})) = \sigma^2(\boldsymbol{S} + \boldsymbol{G}^{\top}\boldsymbol{V}^{-1}\boldsymbol{G})^{-1},$$

donde $\boldsymbol{S} = \boldsymbol{X}^{\top}\boldsymbol{X}$.

Demostración:

Usando la fórmula de Sherman-Morrison-Woodbury, tenemos:

$$(\boldsymbol{S} + \boldsymbol{G}^\top \boldsymbol{V}^{-1}\boldsymbol{G})^{-1} = \boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{G}^\top(\boldsymbol{V} + \boldsymbol{G}\boldsymbol{S}^{-1}\boldsymbol{G}^\top)^{-1}\boldsymbol{G}\boldsymbol{S}^{-1}.$$

Lo que permite escribir

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{\mathsf{ME}}(\boldsymbol{G}) &= (\boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{G}^\top(\boldsymbol{V} + \boldsymbol{G}\boldsymbol{S}^{-1}\boldsymbol{G}^\top)^{-1}\boldsymbol{G}\boldsymbol{S}^{-1})(\boldsymbol{X}^\top\boldsymbol{Y} + \boldsymbol{G}^\top\boldsymbol{V}^{-1}\boldsymbol{g}) \\
&= \boldsymbol{S}^{-1}\boldsymbol{X}^\top\boldsymbol{Y} + \boldsymbol{S}^{-1}\boldsymbol{G}^\top\boldsymbol{V}^{-1}\boldsymbol{g} - \boldsymbol{S}^{-1}\boldsymbol{G}^\top(\boldsymbol{V} + \boldsymbol{G}\boldsymbol{S}^{-1}\boldsymbol{G}^\top)^{-1}\boldsymbol{G}\boldsymbol{S}^{-1}\boldsymbol{X}^\top\boldsymbol{Y} \\
&- \boldsymbol{S}^{-1}\boldsymbol{G}^\top(\boldsymbol{V} + \boldsymbol{G}\boldsymbol{S}^{-1}\boldsymbol{G}^\top)^{-1}\boldsymbol{G}\boldsymbol{S}^{-1}\boldsymbol{G}^\top\boldsymbol{V}^{-1}\boldsymbol{g} \\
&= \widehat{\boldsymbol{\beta}} + \boldsymbol{S}^{-1}\boldsymbol{G}^\top(\boldsymbol{V} + \boldsymbol{G}\boldsymbol{S}^{-1}\boldsymbol{G}^\top)^{-1}\{(\boldsymbol{V} + \boldsymbol{G}\boldsymbol{S}^{-1}\boldsymbol{G}^\top)\boldsymbol{V}^{-1}\boldsymbol{g} - \boldsymbol{G}\widehat{\boldsymbol{\beta}} \\
&- \boldsymbol{G}\boldsymbol{S}^{-1}\boldsymbol{G}^\top\boldsymbol{V}^{-1}\boldsymbol{g}\},
\end{aligned}$$

y el resultado sigue.

Observación:

Podemos notar que

$$\lim_{V \to 0} \widehat{\boldsymbol{\beta}}_{\mathsf{ME}}(\boldsymbol{G}) = \widehat{\boldsymbol{\beta}}(\boldsymbol{G}),$$

y

$$\mathsf{Cov}(\widehat{\boldsymbol{\beta}}) - \mathsf{Cov}(\widehat{\boldsymbol{\beta}}_{\mathsf{ME}}(\boldsymbol{G})) = \sigma^2 \boldsymbol{S}^{-1}\boldsymbol{G}^\top(\boldsymbol{V} + \boldsymbol{G}\boldsymbol{S}^{-1}\boldsymbol{G}^\top)^{-1}\boldsymbol{G}\boldsymbol{S}^{-1} \geq \boldsymbol{0}.$$

Producción textil en los Países bajos (Theil y Nagar, 1961)[2]

Ejemplo (datos de producción textil):

Datos sobre la demanda de la producción textil en los Países bajos durante el periodo 1923-1939.

La respuesta (Y) es el logaritmo del consumo textil per cápita anual, que se obtiene dividiendo el valor monetario del consumo textil por cada hogar por kN, donde k es el índice de precios al por menor de la ropa para la ciudad de Ámsterdam y N es el tamaño de la población de los Países Bajos.

Suponga los regresores, X_1 el logaritmo del índice de precios deflactado de la ropa, que se obtiene como la relación k/γ, y X_2 el logaritmo de la renta real per cápita, que se calcula dividiendo el valor monetario de la renta de los hogares por γN, donde γ es el índice general de precios al por menor.

[2] Journal of the American Statistical Association **56**, 793-806.

En Theil (1963)[3] se proporciona información sobre las restricciones estocásticas para las variables X_1 y X_2. Quien, basado en argumentos asintóticos propuso usar:

$$\boldsymbol{g} = \begin{pmatrix} -0.7 \\ 1.0 \end{pmatrix}, \qquad \boldsymbol{G} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{V} = \begin{pmatrix} 0.0225 & -0.0100 \\ -0.0100 & 0.0225 \end{pmatrix}.$$

[3] Journal of the American Statistical Association **58**, 401-414.

Producción textil en los Países bajos (Theil y Nagar, 1961)

```
# base de datos
> textile <- read.csv("textile.csv")

> textile
   year consumption clothing  income
1  1923     1.99651  2.00432 1.98543
2  1924     1.99564  2.00043 1.99167
3  1925     2.00000  2.00000 2.00000
4  1926     2.04766  1.95713 2.02078
5  1927     2.08707  1.93702 2.02078
6  1928     2.07041  1.95279 2.03941
7  1929     2.08314  1.95713 2.04454
8  1930     2.13354  1.91803 2.05038
9  1931     2.18808  1.84572 2.03862
10 1932     2.18639  1.81558 2.02243
11 1933     2.20003  1.78746 2.00732
12 1934     2.14799  1.79588 1.97955
13 1935     2.13418  1.80346 1.98408
14 1936     2.22531  1.72099 1.98945
15 1937     2.18837  1.77597 2.01030
16 1938     2.17319  1.77452 2.00689
17 1939     2.21880  1.78746 2.01620
```

Producción textil en los Países bajos (Theil y Nagar, 1961)

```
> library(fastmatrix)
> fm <- ols(consumption ~ clothing + income, data = textile)

# Salida
> fm

Call:
ols(formula = consumption ~ clothing + income, data = textile)

Coefficients:
(Intercept)     clothing       income
     1.3739      -0.8289       1.1432

Degrees of freedom: 17 total; 14 residual
Residual standard error: 0.01354036
```

Producción textil en los Países bajos (Theil y Nagar, 1961)

```
> summary(fm)

Call:
ols(formula = consumption ~ clothing + income, data = textile)

Residuals:
      Min         1Q     Median         3Q        Max
-0.024126 -0.011643  0.002971  0.008565  0.021566

Coefficients:
            Estimate Std. Error t value  Pr(>|t|)
(Intercept)   1.3739    0.3061     4.4892   0.0005
clothing     -0.8289    0.0361   -22.9562   0.0000
income        1.1432    0.1560     7.3289   0.0000

Residual standard error: 0.01354 on 14 degrees of freedom
Log-likelikood: 50.66
```

Producción textil en los Países bajos (Theil y Nagar, 1961)

```
# ingresando 'g', 'G' y 'V'
> g <- c(-0.7, 1)
> g
[1] -0.7  1.0

> G <- matrix(c(0,0,1,0,0,1), ncol = 3)
> G
     [,1] [,2] [,3]
[1,]    0    1    0
[2,]    0    0    1

> V <- matrix(c(0.0225, -0.01, -0.01, 0.0225), ncol = 2)
> V
        [,1]    [,2]
[1,]  0.0225 -0.0100
[2,] -0.0100  0.0225
```

```
# extrayendo información desde el objeto 'fm'
> cf <- fm$coef
> Sinv <- fm$cov.unscaled

# Salida:
> cf
(Intercept)    clothing      income
  1.3739214  -0.8288617   1.1431750

> Sinv
            (Intercept)   clothing      income
(Intercept)  510.8912078  0.4166848 -254.252176
clothing       0.4166848  7.1105753   -6.824195
income      -254.2521765 -6.8241953  132.704345

# calculando mixed-estimator
> rhs <- g - G %*% cf
> M <- V + tcrossprod(G %*% Sinv, G)
> me <- cf + crossprod(G %*% Sinv, solve(M, rhs))
> me
(Intercept)    clothing      income
  1.4211080  -0.7004047   1.0001827
```

Producción textil en los Países bajos (Theil y Nagar, 1961)

```
# estimador insesgado de sigma^2
> RSS <- fm$RSS
> n <- fm$dims[1]
> p <- fm$dims[2]
> s2 <- RSS / (n - p)

# calculando error estandar del mixed-estimator
> cov.me <- Sinv - crossprod(G %*% Sinv, solve(M, G %*% Sinv))
> cov.me <- s2 * cov.me
> se <- sqrt(diag(cov.me))

> se
(Intercept)    clothing      income
0.005302874 0.002027800 0.002030361
```

Producción textil en los Países bajos (Theil y Nagar, 1961)

Resumen de estimación, bajo el modelo sin y con restricciones estocásticas:

Parámetro	OLS		ME	
	estimación	error estándar	estimación	error estándar
β_0	1.3739	0.3061	1.4211	0.0053
β_1	-0.8289	0.0361	-0.7004	0.0020
β_2	1.1432	0.1560	1.0002	0.0020