MAT-266: Estimación sujeto a restricciones lineales

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Estimación bajo restricciones lineales

Problema:

Abordar la estimación de $\boldsymbol{\beta}$ y σ^2 sujeto a restricciones lineales del tipo:

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

donde $\boldsymbol{G} \in \mathbb{R}^{q\times p}$ con $\operatorname{rg}(\boldsymbol{G}) = q$ y $\boldsymbol{g} \in \mathbb{R}^{q}$.

Objetivo:

Consideraremos dos procedimientos para obtener estimadores restrigidos

- Método de reducción.
- Método de multiplicadores de Lagrange.

Además, estudiaremos las propiedades estadísticas de tales estimadores.

Estimación bajo restricciones lineales

Sea $\boldsymbol{G} = (\boldsymbol{G}_r, \boldsymbol{G}_q)$ donde $\boldsymbol{G}_q \in \mathbb{R}^{q\times q}$ de rango q. De este modo, podemos escribir las restricciones en (1) como:

$$\boldsymbol{G}\boldsymbol{\beta} = (\boldsymbol{G}_r, \boldsymbol{G}_q)\begin{pmatrix}\boldsymbol{\beta}_r\\ \boldsymbol{\beta}_q\end{pmatrix} = \boldsymbol{G}_r\boldsymbol{\beta}_r + \boldsymbol{G}_q\boldsymbol{\beta}_q = \boldsymbol{g},$$

como $\boldsymbol{G}_q$ es no singular, tenemos

$$\boldsymbol{\beta}_q = \boldsymbol{G}_q^{-1}(\boldsymbol{g} - \boldsymbol{G}_r\boldsymbol{\beta}_r).$$

Particionando $\boldsymbol{X}$ del mismo modo que $\boldsymbol{\beta} = (\boldsymbol{\beta}_r^\top, \boldsymbol{\beta}_q^\top)^\top$, sigue que

$$\begin{aligned}\boldsymbol{X}\boldsymbol{\beta} = (\boldsymbol{X}_r, \boldsymbol{X}_q)\begin{pmatrix}\boldsymbol{\beta}_r\\ \boldsymbol{\beta}_q\end{pmatrix} &= \boldsymbol{X}_r\boldsymbol{\beta}_r + \boldsymbol{X}_q\boldsymbol{\beta}_q\\ &= \boldsymbol{X}_r\boldsymbol{\beta}_r + \boldsymbol{X}_q\boldsymbol{G}_q^{-1}(\boldsymbol{g} - \boldsymbol{G}_r\boldsymbol{\beta}_r)\\ &= (\boldsymbol{X}_r - \boldsymbol{X}_q\boldsymbol{G}_q^{-1}\boldsymbol{G}_r)\boldsymbol{\beta}_r + \boldsymbol{X}_q\boldsymbol{G}_q^{-1}\boldsymbol{g}\end{aligned}$$

Así, podemos escribir el modelo lineal

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

como

$$\boldsymbol{Y} = (\boldsymbol{X}_r - \boldsymbol{X}_q\boldsymbol{G}_q^{-1}\boldsymbol{G}_r)\boldsymbol{\beta}_r + \boldsymbol{X}_q\boldsymbol{G}_q^{-1}\boldsymbol{g} + \boldsymbol{\epsilon},$$

es decir, obtenemos el modelo reducido, dado por

$$\boldsymbol{Y}_R = \boldsymbol{X}_R\boldsymbol{\beta}_r + \boldsymbol{\epsilon},$$

donde

$$\boldsymbol{Y}_R = \boldsymbol{Y} - \boldsymbol{X}_q\boldsymbol{G}_q^{-1}\boldsymbol{g}, \qquad \boldsymbol{X}_R = \boldsymbol{X}_r - \boldsymbol{X}_q\boldsymbol{G}_q^{-1}\boldsymbol{G}_r.$$

En cuyo caso, sabemos que

$$\widetilde{\boldsymbol{\beta}}_r = (\boldsymbol{X}_R^\top\boldsymbol{X}_R)^{-1}\boldsymbol{X}_R^\top\boldsymbol{Y}_R,$$
$$s_r^2 = \frac{1}{n-r}Q_R(\widetilde{\boldsymbol{\beta}}_r),$$

con

$$Q_R(\widetilde{\boldsymbol{\beta}}_r) = \boldsymbol{Y}_R^\top(\boldsymbol{I} - \boldsymbol{X}_R(\boldsymbol{X}_R^\top\boldsymbol{X}_R)^{-1}\boldsymbol{X}_R^\top)\boldsymbol{Y}_R$$

Además,

$$\widetilde{\boldsymbol{\beta}} = \begin{pmatrix} \widetilde{\boldsymbol{\beta}}_r \\ \widetilde{\boldsymbol{\beta}}_q \end{pmatrix} = \begin{pmatrix} \widetilde{\boldsymbol{\beta}}_r \\ \boldsymbol{G}_q^{-1}(\boldsymbol{g} - \boldsymbol{G}_r\widetilde{\boldsymbol{\beta}}_r) \end{pmatrix}$$
$$= \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{G}_q^{-1}\boldsymbol{g} \end{pmatrix} + \begin{pmatrix} \boldsymbol{I} \\ -\boldsymbol{G}_q^{-1}\boldsymbol{G}_r \end{pmatrix}\widetilde{\boldsymbol{\beta}}_r, \qquad (2)$$

Mientras que

$$\boldsymbol{Y}_R - \boldsymbol{X}_R\widetilde{\boldsymbol{\beta}}_r = \boldsymbol{Y} - \boldsymbol{X}_q\boldsymbol{G}_q^{-1}\boldsymbol{g} - (\boldsymbol{X}_r - \boldsymbol{X}_q\boldsymbol{G}_q^{-1}\boldsymbol{g})\widetilde{\boldsymbol{\beta}}_r$$
$$= \boldsymbol{Y} - (\boldsymbol{X}_r, \boldsymbol{X}_q)\begin{pmatrix} \widetilde{\boldsymbol{\beta}}_r \\ \widetilde{\boldsymbol{\beta}}_q \end{pmatrix} = \boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}},$$

de este modo

$$\|\boldsymbol{Y}_R - \boldsymbol{X}_R\widetilde{\boldsymbol{\beta}}_r\|^2 = \|\boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}\|^2. \qquad (3)$$

Resultado 1:

Para el modelo lineal $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ sujeto a las restricciones $\boldsymbol{G}\boldsymbol{\beta} = \boldsymbol{g}$ con $\boldsymbol{\epsilon} \sim \mathsf{N}_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$. el MLE restringido de $\boldsymbol{\beta}$ es dado por (2) y tenemos que

$$\widetilde{\boldsymbol{\beta}} \sim \mathsf{N}_p(\boldsymbol{\beta}, \mathsf{Cov}(\widetilde{\boldsymbol{\beta}})),$$

donde

$$\mathsf{Cov}(\widetilde{\boldsymbol{\beta}}) = \sigma^2 \begin{pmatrix} \boldsymbol{I} \\ -\boldsymbol{G}_q^{-1}\boldsymbol{G}_r \end{pmatrix} (\boldsymbol{X}_R^\top \boldsymbol{X}_R)^{-1} \begin{pmatrix} \boldsymbol{I} \\ -\boldsymbol{G}_q^{-1}\boldsymbol{G}_r \end{pmatrix}^\top .$$

Mientras que

$$s_r^2 = \frac{1}{n-r} Q_R(\widetilde{\boldsymbol{\beta}}_r) = \frac{1}{n-r} Q(\widetilde{\boldsymbol{\beta}}),$$

donde

$$\frac{(n-r)s_r^2}{\sigma^2} \sim \chi^2(n-r),$$

y $\widetilde{\boldsymbol{\beta}}$, $Q(\widetilde{\boldsymbol{\beta}})$ son independientes.

Demostración:

Sabemos que

$$\widetilde{\boldsymbol{\beta}}_r \sim \mathsf{N}_r(\boldsymbol{\beta}_r, \sigma^2(\boldsymbol{X}_R^\top \boldsymbol{X}_R)^{-1}),$$

$$\frac{(n-r)s_r^2}{\sigma^2} = \frac{Q_R(\widetilde{\boldsymbol{\beta}}_r)}{\sigma^2} = \frac{\boldsymbol{Y}_R^\top(\boldsymbol{I}-\boldsymbol{H}_R)\boldsymbol{Y}_R}{\sigma^2} \sim \chi^2(n-r).$$

Así, por (2), tenemos

$$\begin{aligned}\mathsf{E}(\widetilde{\boldsymbol{\beta}}) &= \begin{pmatrix}\mathbf{0}\\ \boldsymbol{G}_q^{-1}\boldsymbol{g}\end{pmatrix} + \begin{pmatrix}\boldsymbol{I}_r\\ -\boldsymbol{G}_q^{-1}\boldsymbol{G}_r\end{pmatrix}\mathsf{E}(\widetilde{\boldsymbol{\beta}}_r)\\ &= \begin{pmatrix}\boldsymbol{\beta}_r\\ \boldsymbol{G}_q^{-1}(\boldsymbol{g}-\boldsymbol{G}_r\boldsymbol{\beta}_r)\end{pmatrix} = \begin{pmatrix}\boldsymbol{\beta}_r\\ \boldsymbol{\beta}_q\end{pmatrix} = \boldsymbol{\beta}\end{aligned}$$

Además,

$$\begin{aligned}\mathsf{Cov}(\widetilde{\boldsymbol{\beta}}) &= \begin{pmatrix} \boldsymbol{I}_r \\ -\boldsymbol{G}_q^{-1}\boldsymbol{G}_r \end{pmatrix} \mathsf{Cov}(\widetilde{\boldsymbol{\beta}}_r) \begin{pmatrix} \boldsymbol{I}_r \\ -\boldsymbol{G}_q^{-1}\boldsymbol{G}_r \end{pmatrix}^\top \\ &= \sigma^2 \begin{pmatrix} \boldsymbol{I}_r \\ -\boldsymbol{G}_q^{-1}\boldsymbol{G}_r \end{pmatrix} (\boldsymbol{X}_R^\top \boldsymbol{X}_R)^{-1} \begin{pmatrix} \boldsymbol{I}_r \\ -\boldsymbol{G}_q^{-1}\boldsymbol{G}_r \end{pmatrix}^\top.\end{aligned}$$

como $\widetilde{\boldsymbol{\beta}}$ es una función lineal de $\widetilde{\boldsymbol{\beta}}$ la normalidad sigue. La independencia entre $\widetilde{\boldsymbol{\beta}}$ y $Q(\widetilde{\boldsymbol{\beta}})$ sigue por el Resultado 1 en Slides 9.

Considere

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathsf{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}).$$

La función Langrangiana asociada a la restricción lineal $\boldsymbol{G}\boldsymbol{\beta} = \boldsymbol{g}$ es dada por:

$$F(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \ell(\boldsymbol{\theta}) + \frac{1}{\sigma^2}\boldsymbol{\lambda}^\top(\boldsymbol{G}\boldsymbol{\beta} - \boldsymbol{g}),$$

con $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$. De este modo,

$$\begin{aligned}
\frac{\partial F(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}\boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \frac{1}{\sigma^2}\boldsymbol{G}^\top\boldsymbol{\lambda} \\
\frac{\partial F(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\{Q(\boldsymbol{\beta}) - 2\boldsymbol{\lambda}^\top(\boldsymbol{G}\boldsymbol{\beta} - \boldsymbol{g})\} \\
\frac{\partial F(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} &= \boldsymbol{G}\boldsymbol{\beta} - \boldsymbol{g}
\end{aligned}$$

Desde la condición de primer orden, obtenemos las ecuaciones de estimación,

$$\begin{aligned} &\boldsymbol{X}^\top(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})+\boldsymbol{G}^\top\boldsymbol{\lambda}=\boldsymbol{0}, \\ &n\sigma^2-\{Q(\boldsymbol{\beta})-2\boldsymbol{\lambda}^\top(\boldsymbol{G}\boldsymbol{\beta}-\boldsymbol{g})\}=0, \\ &\boldsymbol{G}\boldsymbol{\beta}=\boldsymbol{g}, \end{aligned}$$

es decir,

$$\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{X}^\top\boldsymbol{Y}+\boldsymbol{G}^\top\boldsymbol{\lambda}, \tag{4}$$

$$\sigma^2=\frac{1}{n}\,Q(\boldsymbol{\beta}) \tag{5}$$

$$\boldsymbol{G}\boldsymbol{\beta}=\boldsymbol{g}, \tag{6}$$

Resolviendo la Ecuación (4) con relación a $\boldsymbol{\beta}$ tenemos

$$\widetilde{\boldsymbol{\beta}}=(\boldsymbol{X}^\top\boldsymbol{X})^{-1}(\boldsymbol{X}^\top\boldsymbol{Y}+\boldsymbol{G}^\top\boldsymbol{\lambda})$$

Substituyendo este resultado en (6) y resolviendo para $\boldsymbol{\lambda}$, sigue que

$$\boldsymbol{G}\widetilde{\boldsymbol{\beta}} = \boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}(\boldsymbol{X}^\top\boldsymbol{Y} + \boldsymbol{G}^\top\boldsymbol{\lambda}) = \boldsymbol{g},$$

es decir,

$$\boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y} + \boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top\boldsymbol{\lambda} = \boldsymbol{g},$$

por tanto,

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

Reemplazando este resultado en $\widetilde{\boldsymbol{\beta}}$ resulta

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}} &= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}(\boldsymbol{X}^\top\boldsymbol{Y} + \boldsymbol{G}^\top\widetilde{\boldsymbol{\lambda}}) \\
&= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y} + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top(\boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top)^{-1}(\boldsymbol{g} - \boldsymbol{G}\widehat{\boldsymbol{\beta}}) \\
&= \widehat{\boldsymbol{\beta}} + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top(\boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top)^{-1}(\boldsymbol{g} - \boldsymbol{G}\widehat{\boldsymbol{\beta}})
\end{aligned}$$

Que puede ser reorganizado como:

$$\widetilde{\boldsymbol{\beta}} = \boldsymbol{A}\widehat{\boldsymbol{\beta}} + \boldsymbol{B}\boldsymbol{g} = \widehat{\boldsymbol{\beta}} - \boldsymbol{B}(\boldsymbol{G}\widehat{\boldsymbol{\beta}} - \boldsymbol{g}),$$

donde $\widehat{\boldsymbol{\beta}}$ corresponde al MLE no restringido para $\boldsymbol{\beta}$, con

$$\boldsymbol{B} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{G}^\top(\boldsymbol{G}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{G}^\top)^{-1} \tag{7}$$

$$\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{B}\boldsymbol{G} \tag{8}$$

y el estimador insesgado para σ^2 es dado por

$$s_r^2 = \frac{1}{n-r}\, Q(\widetilde{\boldsymbol{\beta}}).$$

Para estudiar las propiedades de este MLE restringido, considere primeramente el siguiente lema.

Lema 1:

La matriz $\boldsymbol{A}$ definida en (8) tiene las siguientes propiedades:

(i) $\boldsymbol{A}$ es idempotente con $\operatorname{rg}(\boldsymbol{A}) = r$.

(ii) $\boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ es idempotente y simétrica con rango r.

(iii) $\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^\top = \boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^\top$.

Demostración:

Para mostrar que $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{B}\boldsymbol{G}$ es idempotente, basta mostrar que $\boldsymbol{B}\boldsymbol{G}$ es idempotente. En efecto,

$$\boldsymbol{G}\boldsymbol{B}\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top(\boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{G}^\top)^{-1}\boldsymbol{G} = \boldsymbol{G},$$

de ahí que $\boldsymbol{B}\boldsymbol{G}$ es idempotente.

Además,
$$\mathrm{rg}(\boldsymbol{B}\boldsymbol{G}) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{G}) = \mathrm{tr}(\boldsymbol{G}\boldsymbol{B}) = q,$$
así $\mathrm{rg}(\boldsymbol{I} - \boldsymbol{B}\boldsymbol{G}) = \mathrm{tr}(\boldsymbol{A}) = p - q = r$, lo que muestra la parte (i).

Para notar la parte (ii), sea $\boldsymbol{C} = \boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$, luego
$$\boldsymbol{C}^2 = \boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top = \boldsymbol{X}\boldsymbol{A}^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top,$$
como $\boldsymbol{A}$ es idempotente, sigue que $\boldsymbol{C}^2 = \boldsymbol{C}$. Esto permite escribir
$$\mathrm{rg}(\boldsymbol{C}) = \mathrm{tr}(\boldsymbol{C}) = \mathrm{tr}(\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{A}) = r.$$
Por otro lado,
$$\boldsymbol{C}^\top = (\boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)^\top = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^\top\boldsymbol{X}^\top.$$

Tenemos que,

$$\boldsymbol{A}^\top = \boldsymbol{I} - \boldsymbol{G}^\top \boldsymbol{B}^\top = \boldsymbol{I} - \boldsymbol{G}^\top (\boldsymbol{G}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{G}^\top)^{-1} \boldsymbol{G}(\boldsymbol{X}^\top \boldsymbol{X})^{-1},$$

luego premultiplicando por $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ y factorizando lleva a,

$$\begin{aligned}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{A}^\top &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} - (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{G}^\top (\boldsymbol{G}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{G}^\top)^{-1} \boldsymbol{G}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \\ &= (\boldsymbol{I} - \boldsymbol{B}\boldsymbol{G})(\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \boldsymbol{A}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}, \qquad (9)\end{aligned}$$

así

$$\boldsymbol{C}^\top = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{A}^\top \boldsymbol{X} = \boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X} = \boldsymbol{C}.$$

Finalmente, Ecuación (9) permite notar la primera igualdad de la parte (iii). Ahora, por (9), sigue que

$$\boldsymbol{A}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{A}^\top = \boldsymbol{A}^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \boldsymbol{A}(\boldsymbol{X}^\top \boldsymbol{X})^{-1},$$

pues $\boldsymbol{A}$ es idempotente y esto termina la prueba.

Resultado 2:

Para el modelo lineal

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathsf{N}_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}).$$

El MLE de $\boldsymbol{\beta}$ bajo las restricciones lineales $\boldsymbol{G}\boldsymbol{\beta} = \boldsymbol{g}$, es dado por[1]

$$\widetilde{\boldsymbol{\beta}} = \boldsymbol{A}\widehat{\boldsymbol{\beta}} + \boldsymbol{B}\boldsymbol{g},$$

con distribución

$$\widetilde{\boldsymbol{\beta}} \sim \mathsf{N}_p(\boldsymbol{\beta}, \boldsymbol{A}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{A}^\top).$$

El MLE restringido de σ^2 es

$$s_r^2 = \frac{1}{n-r}\, Q(\widetilde{\boldsymbol{\beta}}),$$

donde

$$\frac{Q(\widetilde{\boldsymbol{\beta}})}{\sigma^2} \sim \chi^2(n-r),$$

y $\widetilde{\boldsymbol{\beta}}$ es independiente de $Q(\widetilde{\boldsymbol{\beta}})$.

[1] También $\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{G}^\top(\boldsymbol{G}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{G}^\top)^{-1}(\boldsymbol{g} - \boldsymbol{G}\widehat{\boldsymbol{\beta}})$.

Demostración:

La normalidad sigue desde la linealidad con relación a $\widehat{\boldsymbol{\beta}}$. Ahora,

$$\mathsf{E}(\widetilde{\boldsymbol{\beta}}) = \boldsymbol{A}\,\mathsf{E}(\widehat{\boldsymbol{\beta}}) + \boldsymbol{B}\boldsymbol{g} = (\boldsymbol{I} - \boldsymbol{B}\boldsymbol{G})\boldsymbol{\beta} + \boldsymbol{B}\boldsymbol{g} = \boldsymbol{\beta} - \boldsymbol{B}(\boldsymbol{G}\boldsymbol{\beta} - \boldsymbol{g}) = \boldsymbol{\beta},$$

y

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

Notando que

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}} &= \boldsymbol{A}\widehat{\boldsymbol{\beta}} + \boldsymbol{B}\boldsymbol{g} = \boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y} + \boldsymbol{B}\boldsymbol{g} \\
&= \boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) + \boldsymbol{B}\boldsymbol{g} \\
&= \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\epsilon} + \boldsymbol{B}\boldsymbol{g} \\
&= \boldsymbol{\beta} + \boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\epsilon}
\end{aligned}$$

De este modo, podemos escribir

$$\begin{aligned} \boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}} &= \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\epsilon} \\ &= \boldsymbol{\epsilon} - \boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\epsilon} \\ &= (\boldsymbol{I} - \boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{\epsilon}, \end{aligned}$$

por la parte (ii) del Lema 1, sigue que

$$\frac{Q(\widetilde{\boldsymbol{\beta}})}{\sigma^2} = \frac{\boldsymbol{\epsilon}^\top(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{\epsilon}}{\sigma^2} \sim \chi^2(n-r).$$

Para notar la independencia entre $\widetilde{\boldsymbol{\beta}}$ y $Q(\widetilde{\boldsymbol{\beta}})$ debemos tener[2]

$$\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top) = \boldsymbol{0}$$

[2] lo que es consecuencia de escribir $\widetilde{\boldsymbol{\beta}}$ y $Q(\widetilde{\boldsymbol{\beta}})$ en términos de $\boldsymbol{\epsilon}$.

En efecto,

$$\begin{aligned}
&\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)\\
&\qquad = \boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} - \boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top.
\end{aligned}$$

Notando que $\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^\top$, obtenemos

$$\begin{aligned}
\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top &= \boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^\top\boldsymbol{X}^\top\\
&= \boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^\top\boldsymbol{X}^\top = \boldsymbol{A}^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top = \boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top,
\end{aligned}$$

lo que concluye la demostración.