MAT-266: Errores correlacionados y estimación de funciones de varianza

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Considere el modelo

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

donde $\mathsf{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ y

$$\mathsf{Cov}(\boldsymbol{\epsilon}) = \mathsf{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) = \sigma^2\boldsymbol{\Omega}, \qquad \boldsymbol{\Omega} > \mathbf{0}.$$

Primeramente vamos a suponer que $\boldsymbol{\Omega}$ es conocida y sea

$$\boldsymbol{\Omega} = \boldsymbol{B}\boldsymbol{B}^\top,$$

con $\boldsymbol{B}$ matriz no singular $n \times n$ y note que

$$\boldsymbol{B}^{-1}\boldsymbol{Y} = \boldsymbol{B}^{-1}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{B}^{-1}\boldsymbol{\epsilon}.$$

haciendo $\boldsymbol{Y}_* = \boldsymbol{B}^{-1}\boldsymbol{Y}$, $\boldsymbol{X}_* = \boldsymbol{B}^{-1}\boldsymbol{X}$ y $\boldsymbol{\epsilon}_* = \boldsymbol{B}^{-1}\boldsymbol{\epsilon}$. Entonces $\mathsf{E}(\boldsymbol{\epsilon}_*) = \mathbf{0}$, y

$$\mathsf{Cov}(\boldsymbol{\epsilon}_*) = \boldsymbol{B}^{-1}\,\mathsf{Cov}(\boldsymbol{\epsilon})\boldsymbol{B}^{-\top} = \sigma^2\boldsymbol{B}^{-1}\boldsymbol{B}\boldsymbol{B}^\top\boldsymbol{B}^{-\top} = \sigma^2\boldsymbol{I}.$$

Es decir, el modelo transformado

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

satisface las condiciones A1-A4. Así,

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{\mathsf{GLS}} &= (\boldsymbol{X}_*^\top \boldsymbol{X}_*)^{-1}\boldsymbol{X}_*^\top \boldsymbol{Y}_* \\
&= (\boldsymbol{X}^\top \boldsymbol{B}^{-\top}\boldsymbol{B}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{B}^{-\top}\boldsymbol{B}^{-1}\boldsymbol{Y} \\
&= (\boldsymbol{X}^\top (\boldsymbol{B}\boldsymbol{B}^\top)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top (\boldsymbol{B}\boldsymbol{B}^\top)^{-1}\boldsymbol{Y} \\
&= (\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{Y}.
\end{aligned}$$

Es fácil mostrar que

$$\begin{aligned}
\mathsf{E}(\widehat{\boldsymbol{\beta}}_{\mathsf{GLS}}) &= (\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1}\,\mathsf{E}(\boldsymbol{Y}) = \boldsymbol{\beta} \\
\mathsf{Cov}(\widehat{\boldsymbol{\beta}}_{\mathsf{GLS}}) &= \sigma^2(\boldsymbol{X}_*^\top \boldsymbol{X}_*)^{-1} = \sigma^2(\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}.
\end{aligned}$$

Además

$$\boldsymbol{e}_* = \boldsymbol{Y}_* = \boldsymbol{X}_*\widehat{\boldsymbol{\beta}}_{\mathsf{GLS}} = \boldsymbol{B}^{-1}\boldsymbol{Y} - \boldsymbol{B}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathsf{GLS}} = \boldsymbol{B}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathsf{GLS}}),$$

luego

$$\begin{aligned}\|\boldsymbol{e}_*\|^2 &= (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathsf{GLS}})^\top \boldsymbol{B}^{-\top}\boldsymbol{B}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathsf{GLS}}) \\ &= (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathsf{GLS}})^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathsf{GLS}}).\end{aligned}$$

Adicionalmente

$$\begin{aligned}\boldsymbol{e}_* &= \boldsymbol{B}^{-1}\boldsymbol{Y} - \boldsymbol{B}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y} \\ &= \boldsymbol{B}^{-1}\boldsymbol{Y} - \boldsymbol{B}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{B}^{-\top}\boldsymbol{B}^{-1}\boldsymbol{Y}.\end{aligned}$$

De este modo,

$$\boldsymbol{e}_* = (\boldsymbol{I} - \boldsymbol{H}_\Omega)\boldsymbol{Y}_*,$$

con[1]

$$\boldsymbol{H}_\Omega = \boldsymbol{B}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{B}^{-\top}.$$

[1] Evidentemente $\boldsymbol{H}_\Omega^\top = \boldsymbol{H}_\Omega$ y $\boldsymbol{H}_\Omega^2 = \boldsymbol{H}_\Omega$.

Desafortunadamente, en general la matriz $\boldsymbol{\Omega}$ no es conocida y requiere ser estimada. Si $\widehat{\boldsymbol{\Omega}}$ es un estimador de $\boldsymbol{\Omega}$, entonces

$$\widehat{\boldsymbol{\beta}}_{\mathsf{EGLS}} = (\boldsymbol{X}^\top \widehat{\boldsymbol{\Omega}}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \widehat{\boldsymbol{\Omega}}^{-1} \boldsymbol{Y}.$$

Se debe notar que las propiedades de $\widehat{\boldsymbol{\beta}}_{\mathsf{EGLS}}$ son difíciles de caracterizar.

Suponga que

$$\boldsymbol{\Omega} = (1 - \rho)\boldsymbol{I} + \rho \boldsymbol{1}\boldsymbol{1}^\top,$$

con ρ un parámetro desconocido y considere $\boldsymbol{X} = (\boldsymbol{1}, \boldsymbol{Z})$ tal que $\boldsymbol{Z}^\top \boldsymbol{1} = \boldsymbol{0}$. Fijando $\rho = \widehat{\rho}$ con $\widehat{\rho}$ algún estimador de ρ, podemos notar que

$$\widehat{\boldsymbol{\beta}}_{\mathsf{EGLS}} = (\boldsymbol{X}^\top \widehat{\boldsymbol{\Omega}}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \widehat{\boldsymbol{\Omega}}^{-1} \boldsymbol{Y} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y},$$

donde

$$\widehat{\boldsymbol{\beta}} = (\widehat{\delta}, \widehat{\boldsymbol{\theta}}^\top)^\top, \qquad \widehat{\delta} = \overline{Y}, \quad \widehat{\boldsymbol{\theta}} = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{Y}.$$

Un caso particular importante, corresponde a mínimos cuadrados ponderados en cuyo caso $\text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{W}^{-1}$ donde $\boldsymbol{W} = \operatorname{diag}(\omega_1, \dots, \omega_n)$ con $\omega_i > 0$, $\forall\, i$. De este modo,

$$\widehat{\boldsymbol{\beta}}_{\mathsf{WLS}} = (\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{Y}.$$

Note que este estimador es solución del problema

$$\min_{\beta}\ Q_W(\boldsymbol{\beta}), \qquad Q_W(\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}).$$

Además,

$$Q_W(\boldsymbol{\beta}) = \sum_{i=1}^n \omega_i (Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2.$$

Observación:

Consideraciones computacionales sobre GLS y WLS son dadas por ejemplo en el Capítulo 4 de Björk (1996)[2]

[2] Numerical Methods for Least Squares Problems, SIAM.

El objetivo de esta sección es considerar modelos de regresión heterocedásticos, tal que

$$\mathsf{E}(Y_i) = \mu_i, \qquad \mathsf{var}(Y_i) = \sigma^2 g^2(\boldsymbol{z}_i; \mu_i, \boldsymbol{\phi}), \qquad i = 1, \dots, n,$$

donde $\mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, $\boldsymbol{x}_i$ y $\boldsymbol{z}_i$ representan vectores de covariables (que podrían ser iguales), $\boldsymbol{\beta}$ son coeficientes de regresión, g es función de varianza (que permite modelar la heterogeneidad), $\sigma^2 > 0$ y $\boldsymbol{\phi}$ son parámetros de escala desconocidos.

Ejemplo:

Considere

$$\mathsf{var}(Y_i) = \sigma^2 \{g(\boldsymbol{x}_i; \boldsymbol{\beta})\}^{2\phi}, \qquad \phi > 0.$$

Si suponemos $g(\boldsymbol{x}_i; \boldsymbol{\beta}) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$ y $\phi = \frac{1}{2}$ tenemos la estructura de varianza Poisson, mientras que $\phi = 1$ es de tipo-gama.

Ejemplo:

Ejemplos habituales de funciones de varianza son los siguientes:

(a) Función de varianza cuadrática

$$\sigma g(\boldsymbol{z}_i; \mu_i, \phi) = 1 + \phi_1 z_{1i} + \phi_2 z_{2i}^2.$$

(b) Modelo potencia extendido

$$\mathsf{var}(Y_i) = \sigma^2(\phi_1 + \phi_2 \mu_i^{\phi_3}).$$

(c) Modelo exponencial

$$\mathsf{var}(Y_i) = \sigma^2 \exp(2\phi\mu_i).$$

(d) También puede depender de $\boldsymbol{\phi}$ según un predictor lineal

$$\mathsf{var}(Y_i) = \sigma^2 \exp(2\boldsymbol{z}_i^\top \boldsymbol{\phi}).$$

Funciones de varianza

Considere el modelo

$$Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \qquad i = 1, \dots, n,$$

tal que

$$\mathsf{E}(Y_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}, \qquad \mathsf{var}(Y_i) = \sigma^2/\omega_i,$$

para constantes conocidas $\omega_i > 0$, $i = 1, \dots, n$.

Bajo el supuesto $\boldsymbol{\epsilon} \sim \mathsf{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{W}^{-1})$, con $\boldsymbol{W} = \operatorname{diag}(\omega_1, \dots, \omega_n)$. Es fácil notar que el estimador $\widehat{\boldsymbol{\beta}}_{\mathsf{WLS}}$ minimiza la función

$$Q_W(\boldsymbol{\beta}) = \sum_{i=1}^n \omega_i (Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2.$$

Además, el estimador $\widehat{\boldsymbol{\beta}}_{\mathsf{WLS}}$ resuelve las siguientes ecuaciones de estimación

$$\boldsymbol{X}^\top \boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = \mathbf{0},$$

o equivalentemente

$$\sum_{i=1}^n \omega_i (Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}) \boldsymbol{x}_i = \mathbf{0}.$$

El estimador ML para σ^2 asume la forma:

$$\widehat{\sigma}^2_{\mathsf{WLS}} = \frac{1}{n}\sum_{i=1}^{n}\omega_i(Y_i - \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}_{\mathsf{WLS}})^2.$$

Observación:

El problema anterior puede ser resuelto usando OLS en el siguiente 'problema modificado'

$$\boldsymbol{W}^{1/2}\boldsymbol{Y}, \qquad \boldsymbol{W}^{1/2}\boldsymbol{X}.$$

En la práctica es bastante improbable conocer las constantes ω_i, $i = 1, \dots, n$.

El método WLS sugiere un procedimiento natural para la estimación de varianzas heterogéneas.

Suponga el modelo

$$\mathsf{E}(Y_i) = \mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}, \qquad \mathsf{var}(Y_i) = \sigma^2 g^2(\boldsymbol{z}_i; \mu_i, \boldsymbol{\phi}),$$

con $\boldsymbol{\phi}$ conocido.

Una estrategia de estimación es usar un algoritmo de mínimos cuadrados iterativamente ponderados (IWLS).

Funciones de varianza

Algoritmo 1: IWLS para estimación de varianzas

1 **begin**

2 Considerar una estimación inicial para $\boldsymbol{\beta}$, digamos $\boldsymbol{\beta}^{(0)}$, resolviendo

$$\sum_{i=1}^{n}(Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})\boldsymbol{x}_i = \mathbf{0}.$$

y hacemos $k = 0$.

3 Construir los "pesos"

$$\omega_i^{(k)} = 1/g^2(\boldsymbol{z}_i; \mu_i^{(k)}, \boldsymbol{\phi}), \quad \mu_i^{(k)} = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(k)}.$$

4 Actualizar $\boldsymbol{\beta}^{(k+1)}$, resolviendo

$$\sum_{i=1}^{n}\omega_i^{(k)}(Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})\boldsymbol{x}_i = \mathbf{0},$$

hacer $k = k + 1$ y volver a **Paso 3**.

5 **end**

Evidentemente el **Paso 4** del Algoritmo 1, debe ser resuelta usando mínimos cuadrados ponderados.

En el algoritmo anterior σ^2 puede ser estimado por analogía a WLS. Específicamente, podemos considerar

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\widehat{\omega}_i(Y_i - \boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}})^2,$$

donde $\widehat{\omega}_i$ $(i = 1, \dots, n)$ y $\widehat{\boldsymbol{\beta}}$ representan los valores de ω_i y $\boldsymbol{\beta}$ a la convergencia del Algoritmo 1.

Alternativamente, es posible incorporar una etapa adicional al Algoritmo 1 como:

5′
$$\sigma^{2\,(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\omega_i^{(k)}(Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}^{(k+1)})^2.$$

Claramente, no siempre es posible especificar valores para $\boldsymbol{\phi}$. Para simplificar la exposición considere

$$Y_i \overset{\text{ind}}{\sim} \mathsf{N}(\mu_i, \sigma^2 g^2(\boldsymbol{z}_i; \mu_i, \boldsymbol{\phi})), \qquad i = 1, \dots, n,$$

con $\mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$.

Defina la matriz diagonal

$$\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{\beta}, \boldsymbol{\phi}) = \operatorname{diag}\big(g^2(\boldsymbol{z}_1; \mu_1, \boldsymbol{\phi}), \dots, g^2(\boldsymbol{z}_1; \mu_n, \boldsymbol{\phi})\big).$$

De este modo, el modelo anterior puede ser escrito como

$$\boldsymbol{Y} \sim \mathsf{N}_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{G}),$$

con función de densidad conjunta

$$f(\boldsymbol{y}) = |2\pi\sigma^2 \boldsymbol{G}|^{-1/2} \exp\Big\{ -\frac{1}{2\sigma^2} (\boldsymbol{Y} - \boldsymbol{X\beta})^\top \boldsymbol{G}^{-1} (\boldsymbol{Y} - \boldsymbol{X\beta}) \Big\}.$$

La parte relevante de la función de log-verosimilitud es

$$\ell_n(\boldsymbol{\psi}) = -\frac{1}{2}\log|\sigma^2\boldsymbol{G}| - \frac{1}{2\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top\boldsymbol{G}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}),$$

con $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\phi}^\top)^\top$. Por la estructura diagonal de $\boldsymbol{G}$ tenemos que

$$\boldsymbol{G}^{-1} = \operatorname{diag}\big(g^{-2}(\boldsymbol{z}_1;\mu_1,\boldsymbol{\phi}),\dots,g^{-2}(\boldsymbol{z}_1;\mu_n,\boldsymbol{\phi})\big),$$

$$\log|\sigma^2\boldsymbol{G}| = \sum_{i=1}^n \log\sigma^2 g^2(\boldsymbol{z}_i;\mu_i,\boldsymbol{\phi}).$$

Esto permite escribir la log-verosimilitud como:

$$\begin{aligned}\ell_n(\boldsymbol{\psi}) &= -\frac{1}{2}\sum_{i=1}^n \log\sigma^2 g^2(\boldsymbol{z}_i;\mu_i,\boldsymbol{\phi}) - \frac{1}{2\sigma^2}\sum_{i=1}^n \frac{(Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})^2}{g^2(\boldsymbol{z}_i;\mu_i,\boldsymbol{\phi})}\\ &= -\frac{1}{2}\sum_{i=1}^n\Big\{\frac{(Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})^2}{\sigma^2 g^2(\boldsymbol{z}_i;\mu_i,\boldsymbol{\phi})} + \log\sigma^2 g^2(\boldsymbol{z}_i;\mu_i,\boldsymbol{\phi})\Big\}\end{aligned}$$

La estimación de parámetros se puede desarrollar como:

1. Para una estimación preliminar $\boldsymbol{\beta}^{(k)}$ de $\boldsymbol{\beta}$ minimizar con relación a $\boldsymbol{\phi}$ y σ^2, la función perfilada:

$$\ell_*(\boldsymbol{\beta}^{(k)}, \sigma^2, \boldsymbol{\beta}) = -\frac{1}{2}\sum_{i=1}^n \Big\{\frac{(Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(k)})^2}{\sigma^2 g^2(\boldsymbol{z}_i; \mu_i^{(k)}, \boldsymbol{\phi})} + \log \sigma^2 g^2(\boldsymbol{z}_i; \mu_i^{(k)}, \boldsymbol{\phi})\Big\},$$

con $\mu_i^{(k)} = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(k)}$. Diferenciando con relación a $\boldsymbol{\phi}$ y σ^2 lleva a las ecuaciones de estimación:

$$\sum_{i=1}^n \frac{1}{g^4(\boldsymbol{z}_i; \mu_i^{(k)}, \boldsymbol{\phi})}\{r_i^2 - \sigma^2 g^2(\boldsymbol{z}_i; \mu_i^{(k)}, \boldsymbol{\phi})\}\, \boldsymbol{q}(\mu_i^{(k)}, \sigma, \boldsymbol{\phi}) = \boldsymbol{0},$$

donde

$$\boldsymbol{q}(\mu_i^{(k)}, \sigma, \boldsymbol{\phi}) = g^2(\mu_i^{(k)}, \boldsymbol{\phi}) \begin{pmatrix} 1/\sigma \\ \boldsymbol{u}(\mu_i^{(k)}, \boldsymbol{\phi}) \end{pmatrix},$$

y $\boldsymbol{u}(\mu_i^{(k)}, \boldsymbol{\phi})$ representa el vector de derivadas de $\log g(\boldsymbol{z}_i, \mu_i, \boldsymbol{\phi})$ con relación a $\boldsymbol{\phi}$, mientras que $r_i = Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(k)}$.

Funciones de varianza

Inspección de la ecuación anterior permite notar que la estimación de σ y $\boldsymbol{\phi}$ corresponde a WLS con "respuesta" r_i^2, "función de regresión" $\sigma^2 g^2(\boldsymbol{z}_i; \mu_i^{(k)}, \boldsymbol{\phi})$, "parámetros de regresión" $(\sigma, \boldsymbol{\phi}^\top)^\top$, "pesos" $g^{-4}(\boldsymbol{z}_i, \mu_i^{(k)}, \boldsymbol{\phi})$ y gradiente (matriz de diseño cuyas filas están dadas por) $\boldsymbol{q}(\mu_i^{(k)}, \sigma, \boldsymbol{\phi})$.

Finalmente este paso debe ser alternado con

2. Actualizar $\boldsymbol{\beta}^{(k+1)}$ como la solución del problema mínimos cuadrados ponderados

$$\boldsymbol{X}^\top \boldsymbol{G}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{0},$$

donde $\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{\beta}^{(k)}, \boldsymbol{\phi}^{(k+1)})$.