

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

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM



Considere el modelo

$$Y = X\beta + \epsilon,$$

donde $E(\epsilon) = 0$ y

$$\text{Cov}(\epsilon) = E(\epsilon\epsilon^\top) = \sigma^2\Omega, \quad \Omega > 0.$$

Primeramente vamos a suponer que Ω es conocida y sea

$$\Omega = BB^\top,$$

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

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

haciendo $Y_* = B^{-1}Y$, $X_* = B^{-1}X$ y $\epsilon_* = B^{-1}\epsilon$. Entonces $E(\epsilon_*) = 0$, y

$$\text{Cov}(\epsilon_*) = B^{-1} \text{Cov}(\epsilon) B^{-\top} = \sigma^2 B^{-1} B B^\top B^{-\top} = \sigma^2 I.$$



Es decir, el modelo transformado

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

satisface las condiciones A1-A4. Así,

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

Es fácil mostrar que

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



Además

$$e_* = Y_* = X_* \hat{\beta}_{\text{GLS}} = B^{-1}Y - B^{-1}X\hat{\beta}_{\text{GLS}} = B^{-1}(Y - X\hat{\beta}_{\text{GLS}}),$$

luego

$$\begin{aligned}\|e_*\|^2 &= (Y - X\hat{\beta}_{\text{GLS}})^\top B^{-\top} B^{-1} (Y - X\hat{\beta}_{\text{GLS}}) \\ &= (Y - X\hat{\beta}_{\text{GLS}})^\top \Omega^{-1} (Y - X\hat{\beta}_{\text{GLS}}).\end{aligned}$$

Adicionalmente

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

De este modo,

$$e_* = (I - H_\Omega)Y_*,$$

con¹

$$H_\Omega = B^{-1}X(X^\top \Omega^{-1}X)^{-1}X^\top B^{-\top}.$$

¹Evidentemente $H_\Omega^\top = H_\Omega$ y $H_\Omega^2 = H_\Omega$.

Errores correlacionados

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

$$\hat{\beta}_{\text{EGLS}} = (\mathbf{X}^\top \hat{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\Omega}^{-1} \mathbf{Y}.$$

Se debe notar que las propiedades de $\hat{\beta}_{\text{EGLS}}$ son difíciles de caracterizar.

Suponga que

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

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

$$\hat{\beta}_{\text{EGLS}} = (\mathbf{X}^\top \hat{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\Omega}^{-1} \mathbf{Y} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y},$$

donde

$$\hat{\beta} = (\hat{\delta}, \hat{\theta}^\top)^\top, \quad \hat{\delta} = \bar{Y}, \hat{\theta} = (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{Y}.$$



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

$$\hat{\beta}_{\text{WLS}} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{Y}.$$

Note que este estimador es solución del problema

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

Además,

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

Observación:

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

²Numerical Methods for Least Squares Problems, SIAM.



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

$$E(Y_i) = \mu_i, \quad \text{var}(Y_i) = \sigma^2 g^2(z_i; \mu_i, \phi), \quad i = 1, \dots, n,$$

donde $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$, \mathbf{x}_i y \mathbf{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 ϕ son parámetros de escala desconocidos.

Ejemplo:

Considere

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

Si suponemos $g(\mathbf{x}_i; \boldsymbol{\beta}) = \mathbf{x}_i^\top \boldsymbol{\beta}$ 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

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

- (b) Modelo potencia extendido

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

- (c) Modelo exponencial

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

- (d) También puede depender de ϕ según un predictor lineal

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



Funciones de varianza

Considere el modelo

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

tal que

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

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

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

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

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

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

o equivalentemente

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



El estimador ML para σ^2 asume la forma:

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

Observación:

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

$$\mathbf{W}^{1/2} \mathbf{Y}, \quad \mathbf{W}^{1/2} \mathbf{X}.$$

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

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

con ϕ conocido.

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



Algoritmo 1: IWLS para estimación de varianzas

1 **begin**

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

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

y hacemos $k = 0$.

3 Construir los “pesos”

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

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

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

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

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

donde $\hat{\omega}_i$ ($i = 1, \dots, n$) y $\hat{\boldsymbol{\beta}}$ 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'
$$\sigma^{2(k+1)} = \frac{1}{n} \sum_{i=1}^n \omega_i^{(k)} (Y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(k+1)})^2.$$



Claramente, no siempre es posible especificar valores para ϕ . Para simplificar la exposición considere

$$Y_i \stackrel{\text{ind}}{\sim} N(\mu_i, \sigma^2 g^2(z_i; \mu_i, \phi)), \quad i = 1, \dots, n,$$

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

Defina la matriz diagonal

$$\mathbf{G} = \mathbf{G}(\boldsymbol{\beta}, \phi) = \text{diag}(g^2(z_1; \mu_1, \phi), \dots, g^2(z_n; \mu_n, \phi)).$$

De este modo, el modelo anterior puede ser escrito como

$$\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{G}),$$

con función de densidad conjunta

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



La parte relevante de la función de log-verosimilitud es

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

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

$$\mathbf{G}^{-1} = \text{diag} (g^{-2}(z_1; \mu_1, \phi), \dots, g^{-2}(z_n; \mu_n, \phi)),$$

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

Esto permite escribir la log-verosimilitud como:

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



La estimación de parámetros se puede desarrollar como:

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

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

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

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

donde

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

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



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

Finalmente este paso debe ser alternado con

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

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

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

