CAPíTULO 9

Mínimos cuadrados generalizados y máxima verosimilitud

9.1. Introducción

En el marco del modelo clásico, los supuestos de homocedasticidad, $E(u_i^2) = \sigma_u^2$ $(i = 1,\ 2,\ \dots\ n)$, y ausencia de autocorrelación, $E(u_iu_j) = 0\ \forall i \neq j\ (i,j = 1,\ 2,\ \dots\ n)$, implican que la matriz de varianzas y covarianzas de $\mathbf{u}$ es escalar

$$
\begin{aligned}
V(\mathbf{u}) =& E[(\mathbf{u} - E(\mathbf{u}))(\mathbf{u}' - E(\mathbf{u}'))] = E\left[\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \begin{pmatrix} u_1 & u_2 & \dots & u_n \end{pmatrix}\right] \\
=& \begin{pmatrix} E(u_1^2) & E(u_1u_2) & \dots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \dots & E(u_2u_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(u_nu_1) & E(u_nu_2) & \dots & E(u_n^2) \end{pmatrix} = \begin{pmatrix} \sigma_u^2 & 0 & \dots & 0 \\ 0 & \sigma_u^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_u^2 \end{pmatrix} = \sigma_u^2 \mathbf{I}_n
\end{aligned}
$$

El supuesto de homocedasticidad implica que todos los elementos de la diagonal principal de $V(\mathbf{u})$, las varianzas, son iguales a un escalar σ_u^2, mientras que el de autocorrelación implica que los elementos situados fuera de la diagonal principal de $V(\mathbf{u})$, las covarianzas, son iguales a cero. Si relajamos estos dos supuestos, entonces la matriz de varianzas y covarianzas deja de ser escalar

$$
V(\mathbf{u}) = \begin{pmatrix} E(u_1^2) & E(u_1u_2) & \dots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \dots & E(u_1u_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(u_nu_1) & E(u_nu_2) & \dots & E(u_n^2) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{pmatrix} = \mathbf{\Sigma}
$$

Definición 81. *El modelo lineal general con perturbaciones no esféricas es*

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, \dots, n$$

en donde $E(u_i) = 0$, $E(u_i^2) = \sigma_i^2$ y $E(u_iu_j) = \sigma_{ij}$ $\forall i \neq j$, que podemos escribir en notción matricial como $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, con $E(\mathbf{u}) = \mathbf{0}$ y $E(\mathbf{uu}') = \mathbf{\Sigma}$.

Observación 55. *Conviene escribir $\mathbf{\Sigma} = \sigma_u^2\mathbf{\Omega}$, para obtener el modelo lineal general con perturbaciones esféricas como un caso especial del modelo con perturbaciones no esféricas, $\mathbf{\Omega} = \mathbf{I}$.*

En este capítulo vamos a demostrar que en el modelo lineal general con perturbaciones no esféricas, hay un estimador alternativo y superior al estimador de mínimos cuadrados ordinarios: el estimador de mínimos cuadrados generalizados, que es equivalente al estimador de máxima verosimilitud cuando el vector de errores sigue una distribución normal multivariante.

9.2. El estimador de mínimos cuadrados ordinarios

Proposición 85. *En el modelo lineal general con heterocedasticidad y/o autocorrelación, el estimador de mínimos cuadrados ordinarios es*

$$\hat{\boldsymbol{\beta}}_{MCO} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

Demostración. El método de mínimos cuadrados ordinarios no tiene en cuenta la matriz de varianzas y covarianzas de los errores al minimizar la suma de cuadrados de los residuos:

$$Q = \hat{\mathbf{u}}'\hat{\mathbf{u}} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = (\mathbf{y}' - \hat{\boldsymbol{\beta}}'\mathbf{X}')(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

□

Proposición 86. *En el modelo lineal general con heterocedasticidad y/o autocorrelación, el estimador MCO es insesgado.*

Demostración. Por definición

$$E(\hat{\boldsymbol{\beta}}_{MCO}) = E\left[\boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}\right]$$

Como $\boldsymbol{\beta}$ es un parámetro y $\mathbf{X}$ es una matriz no estocástica,

$$E(\hat{\boldsymbol{\beta}}_{MCO}) = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E(\mathbf{u}) = \boldsymbol{\beta}$$

porque $E(\mathbf{u}) = 0$. □

Observación 56. *La proposición anterior es lógica porque que la propiedad de insesgadez se basa en los supuestos de regresores no estocásticos y $E(\mathbf{u}) = \mathbf{0}$, pero no tiene en cuenta la matriz de varianzas y covarianzas de los errores.*

Proposición 87. *En el modelo lineal general con heterocedasticidad y/o autocorrelación, la matriz de varianzas y covarianzas de $\hat{\boldsymbol{\beta}}_{MCO}$ es*

$$V(\hat{\boldsymbol{\beta}}_{MCO}) = \sigma_u^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

Demostración. Por definición

$$V(\hat{\boldsymbol{\beta}}_{MCO}) = E\left[\left(\hat{\boldsymbol{\beta}}_{MCO} - E(\hat{\boldsymbol{\beta}}_{MCO})\right)\left(\hat{\boldsymbol{\beta}}_{MCO} - E(\hat{\boldsymbol{\beta}}_{MCO})\right)'\right]$$

Como el estimador es insesgado, $\hat{\boldsymbol{\beta}}_{MCO} - E(\hat{\boldsymbol{\beta}}_{MCO}) = \hat{\boldsymbol{\beta}}_{MCO} - \boldsymbol{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}$, y su traspuesta $\left(\hat{\boldsymbol{\beta}}_{MCO} - E(\hat{\boldsymbol{\beta}}_{MCO})\right)' = \left(\hat{\boldsymbol{\beta}}_{MCO} - \boldsymbol{\beta}\right)' = \mathbf{u}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}$, de modo que

$$\begin{aligned} V(\hat{\boldsymbol{\beta}}_{MCO}) =& E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right] = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E\left[\mathbf{u}\mathbf{u}'\right]\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ =& \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left[\sigma_u^2\boldsymbol{\Omega}\right]\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \sigma_u^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1} \end{aligned}$$

□

Proposición 88. *En el modelo lineal general con heterocedasticidad y/o autocorrelación, el estimador MCO es consistente si*

$$\lim_{n\to\infty}\left(\frac{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}}{n}\right) = \mathbf{R}$$

es una matriz finita.

Demostración. Un estimador es consistente si su error cuadrático medio tiende a un vector de ceros cuando n tiende a infinito. Como el estimador de mínimos cuadrados es insesgado, el error cuadrático medio es igual a la matriz de varianzas y covarianzas, y

$$\lim_{n\to\infty} V(\hat{\boldsymbol{\beta}}) = \lim_{n\to\infty} \frac{\sigma^2}{n}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}}{n}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} = 0\mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}^{-1} = \mathbf{O}$$

□

Proposición 89. *Suponiendo* $\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2\boldsymbol{\Omega})$*, el estimador MCO tiene una distribución normal*

$$\hat{\boldsymbol{\beta}}_{MCO} \sim N(\boldsymbol{\beta}, \sigma_u^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1})$$

Proposición 90. *El estimador* $\hat{\sigma}_u^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k)$ *es un estimador sesgado.*

Demostración.

$$E(\hat{\mathbf{u}}'\hat{\mathbf{u}}) = E(\mathbf{u}'\mathbf{M}\mathbf{u}) = E(tr\mathbf{u}'\mathbf{M}\mathbf{u}) = E(tr\mathbf{M}\mathbf{u}\mathbf{u}') = tr\mathbf{M}E(\mathbf{u}\mathbf{u}') = \sigma_u^2 tr\mathbf{M}\boldsymbol{\Omega} \neq \sigma_u^2(n-k)$$

□

9.3. El estimador de mínimos cuadrados generalizados

En esta sección nos planteamos la siguiente pregunta: ¿es posible transformar un modelo lineal general con perturbaciones no esféricas en un modelo lineal general con perturbaciones esféricas? Si la respuesta es afirmativa, entonces el modelo transformado cumplirá las hipótesis básicas y todos los resultados establecidos en los temas anteriores serán de aplicación directa. El estimador de mínimos cuadrados ordinarios (MCO) en el modelo transformado se denomina estimador de mínimos cuadrados generalizados (MCG). Este estimador será ELIO.

Para encontrar un modelo transformado con las hipótesis básicas, premultiplicamos el modelo lineal general con perturbaciones no esféricas por una matriz $\mathbf{P}$ no estocástica

$$\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\mathbf{u}$$

Este modelo transformado puede escribirse como

$$\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \mathbf{u}_*$$

en donde $\mathbf{y}_* = \mathbf{P}\mathbf{y}$, $\mathbf{X}_* = \mathbf{P}\mathbf{X}$ y $\mathbf{u}_* = \mathbf{P}\mathbf{u}$.

El término de error en el modelo transformado $\mathbf{u}_*$ cumple las siguientes propiedades:

1. $E(\mathbf{u}_*) = E(\mathbf{P}\mathbf{u}) = \mathbf{P}E(\mathbf{u}) = \mathbf{0}$
2. $E(\mathbf{u}_*\mathbf{u}_*') = E(\mathbf{P}\mathbf{u}\mathbf{u}'\mathbf{P}') = \mathbf{P}E(\mathbf{u}\mathbf{u}')\mathbf{P}' = \sigma_u^2\mathbf{P}\boldsymbol{\Omega}\mathbf{P}$

Si la matriz $\mathbf{P}$ es tal que $\sigma_u^2\mathbf{P}\boldsymbol{\Omega}\mathbf{P}' = \sigma_u^2\mathbf{I}$, entonces el modelo transformado:

1. contiene los parámetros de interés $\boldsymbol{\beta}$ y σ_u^2
2. cumple las hipótesis básicas.

De aquí, el estimador de mínimos cuadrados ordinarios en el modelo transformado proporciona el estimador lineal, insesgado y eficiente de $\boldsymbol{\beta}$ y el estimador insesgado de σ_u^2.

Proposición 91. *Existe una matriz* $\mathbf{P}$ *tal que* $\mathbf{P}\boldsymbol{\Omega}\mathbf{P}' = \mathbf{I}$

Demostración. De la definicición de autovalores y autovectores, podemos escribir

$$\mathbf{\Omega C} = \mathbf{C\Lambda}$$

en donde $\mathbf{\Lambda} = diag(\lambda_1, \ldots, \lambda_n)$ es la matriz diagonal de autovalores y $\mathbf{C}$ es la matriz autovectores. Además, por ser $\mathbf{\Omega}$ una matriz simétrica, la matriz $\mathbf{C}$ es ortogonal o unitaria $\mathbf{C}^{-1} = \mathbf{C}'$. De aquí, podemos escribir

$$\mathbf{\Omega} = \mathbf{C\Lambda C}'$$

Definiendo $\mathbf{\Lambda}^{1/2} = diag(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$, tenemos que

$$\mathbf{\Omega} = \mathbf{C\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{C}'$$

Premultiplicando $\mathbf{\Omega}$ por $\mathbf{\Lambda}^{-1/2}\mathbf{C}'$ y postmultiplicando por $\mathbf{C\Lambda}^{-1/2}$

$$\mathbf{\Lambda}^{-1/2}\mathbf{C}'\mathbf{\Omega C\Lambda}^{-1/2} = \mathbf{\Lambda}^{-1/2}\mathbf{C}'\mathbf{C\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{C}'\mathbf{C\Lambda}^{-1/2} = \mathbf{I}$$

De aquí, vemos que la matriz buscada es

$$\mathbf{P} = \mathbf{\Lambda}^{-1/2}\mathbf{C}'$$

. □

De la demostración anterior, se derivan las dos siguientes relaciones que serán de interés más adelante:

1. $\mathbf{\Omega}^{-1} = \mathbf{P}'\mathbf{P}$
2. $\mathbf{\Omega} = \mathbf{P}^{-1}\mathbf{P}'^{-1}$

Proposición 92. *El estimador lineal, insesgado y óptimo de $\boldsymbol{\beta}$ es*

$$\hat{\boldsymbol{\beta}}_{MCG} = \left(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}$$

que se denomina estimador de Mínimos Cuadrados Generalizados o estimador de Aitken.

Demostración. Como el modelo transformado cumple los supuestos del modelo clásico, el estimador el estimador de mínimos cuadrados ordinarios

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}_*'\mathbf{X}_*\right)^{-1}\mathbf{X}_*'\mathbf{y}_*$$

será el estimador lineal, insesgado y óptimo, que podemos expresarse en términos de los datos originales

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{y} = \left(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}$$

□

Proposición 93. *La matriz de varianzas y covarianzas del estimador de MCG es*

$$Var(\hat{\boldsymbol{\beta}}_{MCG}) = \sigma_u^2(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}$$

Demostración. La matriz de varianzas y covarianzas del estimador de MCO de $\boldsymbol{\beta}$ en el modelo transformado es

$$Var(\hat{\boldsymbol{\beta}}_{MCG}) = \sigma_u^2(\mathbf{X}_*'\mathbf{X}_*)^{-1} = \sigma_u^2\left(\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{X}\right)^{-1} = \sigma_u^2\left(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1}$$

□

Proposición 94. *El estimador insesgado de σ_u^2 es*

$$\hat{\sigma}^2_{MCG} = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MCG})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MCG})}{n-k}$$

Demostración. El estimador insesgado de σ_u^2 en el modelo transformado es

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}_*'\hat{\mathbf{u}}_*}{n-k} = \frac{(\mathbf{y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}}_{MCG})'(\mathbf{y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}}_{MCG})}{n-k} = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MCG})'\mathbf{P}'\mathbf{P}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MCG})}{n-k}$$

□

9.4. Contraste de hipótesis

Los contrastes de hipótesis se realizan en el modelo transformado aplicando los procedimientos establecidos en los temas anteriores. A modo de resumen, se presentan las siguientes proposiciones cuya demostración es trivial.

Proposición 95. *Bajo el supuesto $\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2\boldsymbol{\Omega})$, el estimador MCG de $\boldsymbol{\beta}$ tiene una distribución normal*

$$\hat{\boldsymbol{\beta}}_{MCG} \sim N\left(\boldsymbol{\beta}, \sigma_u^2(\mathbf{X}_*'\mathbf{X}_*)^{-1}\right) \equiv N\left(\boldsymbol{\beta}, \sigma_u^2(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\right)$$

Proposición 96. *Bajo el supuesto $\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2\boldsymbol{\Omega})$, el estadístico $(n-k)\hat{\sigma}^2_{MCG}/\sigma_u^2$ tiene una distribución Chi-cuadrado con $n-k$.*

Proposición 97. *La hipótesis $H_0 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$ se rechaza al nivel de significación α si*

$$F \equiv [\mathbf{R}\hat{\boldsymbol{\beta}}_{MCG} - \mathbf{r}]'[\hat{\sigma}^2_{MCG}\mathbf{R}(\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{R}']^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}}_{MCG} - \mathbf{r}]/q > c$$

o bien

$$F \equiv [\mathbf{R}\hat{\boldsymbol{\beta}}_{MCG} - \mathbf{r}]'[\hat{\sigma}^2_{MCG}\mathbf{R}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}}_{MCG} - \mathbf{r}]/q > c$$

en donde c es el valor crítico para el cual $Prob(F_{q,n-k} > c) = \alpha$.

9.5. Bondad de ajuste

En la estimación MCG podemos definir dos residuos:

1. los calculados en el modelo de interés

$$\hat{\mathbf{u}}_{MCG} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MCG}$$

2. los calculados en el modelo transformado

$$\hat{\mathbf{u}}_* = \mathbf{y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}}_{MCG} = \mathbf{P}\hat{\mathbf{u}}_{MCG}$$

Los residuos $\hat{\mathbf{u}}_*$ derivados de la estimación MCO del modelo transformado cumplen la propiedad $\mathbf{X}_*'\hat{\mathbf{u}}_* = \mathbf{0}$. Cuando el modelo transformado incluye término constante, la media de estos residuos es igual cero. Sin embargo, en la mayoría de las situaciones el modelo transformado no incluye término constante, por lo que la media de los residuos $\hat{\mathbf{u}}_*$ es distinta de cero. De aquí, en el modelo transformado, la descomposición de la suma de cuadrados total en explicada y residual no se cumple siempre

$$SCT_* \neq SCE_* + SCR_*$$

En consecuencia, el coeficiente de determinación R_*^2 no está acotado entre 0 y 1. Pero, aún cuando el modelo transformado incluya término constante y el R_*^2 esté comprendido entre 0 y 1, no tiene mucho sentido usar este estadístico como medida de bondad de ajuste, porque no estamos interesados en explicar $\mathbf{y}_*$ sino los datos observados $\mathbf{y}$.

Por otro lado, los residuos calculados en el modelo de interés $\hat{\mathbf{u}}_{MCG}$ no tienen media cero, y el R^2 basado en estos residuos no está acotado.

9.6. Mínimos cuadrados generalizados factibles

El cálculo del estimador MCG de $\boldsymbol{\beta}$ requiere conocer la matriz $\boldsymbol{\Omega}$. Como los errores aleatorios no son observables, la matriz $\boldsymbol{\Omega}$ es desconocida y no es posible obtener el estimador MCG. En la práctica, tenemos que estimar la matriz $\boldsymbol{\Omega}$.

Definición 82. *El estimador de mínimos cuadrados generalizados factibles de* $\boldsymbol{\Omega}$ *es*

$$\hat{\boldsymbol{\beta}}_{MCGF} = \left(\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y}$$

en donde $\hat{\boldsymbol{\Omega}}$ *es una estimación de* $\boldsymbol{\Omega}$.

Observación 57. Las propiedades en pequeñas muestras del estimador $\hat{\boldsymbol{\beta}}_{MCGF}$ *son desconocidas, por lo que no es claro si es un estimador mejor que el de MCO.*

El cálculo del estimador MCG requiere invertir la matriz $\boldsymbol{\Omega}$ de orden $n \times n$. La inversión de esta matriz supone una gran coste computacional y puede evitarse cuando la matriz $\boldsymbol{\Omega}$ tiene una determinada estructura. En los temas de heterocedasticidad y autocorrelación, estudiaremos las formas más comunes de la matriz $\boldsymbol{\Omega}$ y los procedimientos más convenientes para obtener el estimador de mínimos cuadrados generalizados sin invertir $\boldsymbol{\Omega}$.

9.7. Método de máxima verosimilitud

El método de mínimos cuadrados no requiere conocer la distribución de las observaciones. En 1921 R.A. Fisher propuso un método de estimación basado en la función de verosimilitud.

Definición 83. *El vector de variables aleatorias* $\mathbf{y} = (y_1\ y_2\ \dots\ y_n)'$ *sigue una* **distribución normal multivariante** *con vector de medias* $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ *y matriz de covarianzas* $V(\mathbf{y}) = \sigma_u^2\boldsymbol{\Omega}$, $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma_u^2\boldsymbol{\Omega})$, *si tiene una función de densidad conjunta de la forma*

$$p(\mathbf{y}) = \left(\frac{1}{2\pi}\right)^{n/2} |\sigma_u^2\boldsymbol{\Omega}|^{-1/2} exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\sigma_u^2\boldsymbol{\Omega})^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right) \tag{9.1}$$

where $|\sigma_u^2\boldsymbol{\Omega}|$ *es el determinante de la matriz de covarianzas y* $exp()$ *indica el número e elevado a ese argumento.*

La función de densidad conjunta nos dice cuál es la *probabilidad* de observar una muestra particular de la variable aleatoria $\mathbf{y}$. Para calcular esta probabilidad necesitamos conocer los parámetros $(\boldsymbol{\beta};\ \sigma_u^2)$. Usando unas estimaciones de estos parámetros, junto con los valores conocidos de las matrices $\mathbf{X}$ y $\boldsymbol{\Omega}$, podríamos estimar la probabilidad de obtener una muestra observada simplemente evaluando el determinante y el exponente de la función.

Definición 84. *La función de densidad conjunta contemplada como una función de los parámetros desconocidos*

$$L(\boldsymbol{\beta},\ \sigma_u^2|\mathbf{y},\mathbf{X},\boldsymbol{\Omega}) = p(\mathbf{y})$$

se denomina **función de verosimilitud***.*

El método de estimación de máxima verosimilitud consiste en encontrar los valores de los prarámetros que maximizan la probabilidad de obtener la muestra observada.

Definición 85. *Los estimadores de máxima verosimilitud de los parámetros desconocidos $\boldsymbol{\beta}$ y σ_u^2 son los valores $\tilde{\boldsymbol{\beta}}$ y $\tilde{\sigma}_u^2$ que maximizan la función de verosimilitud $L(\boldsymbol{\beta},\ \sigma_u^2|\mathbf{y},\mathbf{X},\boldsymbol{\Omega})$.*

Puesto que la probabilidad siempre es positiva y el logaritmo es una transformación monótona, maximinzar $L(\boldsymbol{\beta},\sigma_u^2|\mathbf{y},\mathbf{X},\boldsymbol{\Omega})$ es equivalente a maximizar su logaritmo neperiano, $\ell(\boldsymbol{\beta},\sigma_u^2) = \ln(L(\boldsymbol{\beta},\ \sigma_u^2|\mathbf{y},\mathbf{X},\boldsymbol{\Omega}))$. Tomando logaritmos neperianos en (9.1) tenemos

$$\ell(\boldsymbol{\beta},\sigma_u^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma_u^2) - \frac{1}{2}\ln(\boldsymbol{\Omega}) - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$$

en donde hemos usado los resultados $|\sigma_u^2\boldsymbol{\Omega}| = (\sigma_u^2)^n|\boldsymbol{\Omega}|$ y $\ln(e^z) = z$.

Proposición 98. *Los estimadores de máxima verosimilitud de $\boldsymbol{\beta}$ y σ_u^2 son*

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$$

$$\tilde{\sigma}_u^2 = \frac{(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}})}{n}$$

Demostración. Las derivadas parciales de $\ell(\boldsymbol{\beta},\sigma_u^2)$ respeto de $\boldsymbol{\beta}$ y σ_u^2 son

$$\frac{\partial\ell(\boldsymbol{\beta},\sigma_u^2)}{\partial\boldsymbol{\beta}} = -\frac{1}{\sigma_u^2}(-\mathbf{X}'\boldsymbol{\Omega}^{-\mathbf{1}}\mathbf{y} + \mathbf{X}'\boldsymbol{\Omega}^{-\mathbf{1}}\mathbf{X}\boldsymbol{\beta})$$

$$\frac{\partial\ell(\boldsymbol{\beta},\sigma_u^2)}{\partial\sigma_u^2} = -\frac{n}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$$

Igualando estas dos derivadas parciales a cero y resolviendo simultaneamente las ecuaciones resultantes encontramos los estimadores buscados. □

Observación 58. *En la demostración anterior, al igualar las derivadas parciales a cero tenemos que reemplazar los parámetros desconocidos por sus estimaciones. Estas derivadas no tienen porqué anularse cuando se evalúan para los valores verdaderos de los parámetros.*

Los estimadores de máxima verosimilitud son invariantes a transformaciones de los parámetros. Es equivalente maximizar la función de verosimilitud respecto de σ_u^2 que respecto de σ_u^2.

9.8. Resumen

1. Al relajar los supuestos de homocedasticidad y autocorrelación, el estimador de mínimos cuadrados de $\boldsymbol{\beta}$ es lineal e insesgado, pero ineficiente.

2. El estimador lineal, insesgado y óptimo de $\boldsymbol{\beta}$ en el modelo lineal general con perturbaciones no esféricas es el estimador de mínimos cuadrados generalizados

$$\hat{\boldsymbol{\beta}}_{MCG} = \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$$

3. El estimador de MCG supone que la matriz $\boldsymbol{\Omega}$ es conocida.
4. Las medidas de bondad de ajuste asociadas a la estimación por MCG no son muy informativas.
5. Bajo el supuesto normalidad, es estimador de mínimos cuadrados generalizados coincide con el estimador de máxima verosimilitud.

Palabras clave

Heterocedasticidad	Mínimos cuadrados generalizados
Autocorrelación	MCG factibles
Perturbaciones no esféricas	Máxima verosimilitud

9.9. Ejercicios

1. Demuestre que el estimador MCG de $\boldsymbol{\beta}$ minimiza la suma de cuadrados generalizada

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MCG})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MCG})$$

2. Demuestre que el estimador $\hat{\boldsymbol{\beta}}_{MCG}$ es más eficiente que el estimador $\hat{\boldsymbol{\beta}}_{MCO}$. Pista: demuestre que la diferencia entre $Var(\hat{\boldsymbol{\beta}}_{MCG})^{-1}$ y $Var(\hat{\boldsymbol{\beta}}_{MCO})^{-1}$ es una matriz semidefinida positiva y use la relación $\boldsymbol{\Omega}^{-1} = \mathbf{P}'\mathbf{P}$.
3. Obtenga los estimadores de máxima verosimilitud en el marco del modelo clásico con normalidad.