

Universidad de Chile
Departamento de Ingeniería Industrial
Profesor: Mattia Makovec (mmakovec@dii.uchile.cl)
Auxiliar: Dario Zuñiga

MÍNIMOS CUADRADOS GENERALIZADOS

(Versión preliminar, no difundir)

Mínimos cuadrados generalizados (MCG) con regresores no estocásticos

Dado el modelo lineal general $y = X\beta + \varepsilon$, relajamos el supuesto que la matriz de varianza y covarianza de los errores sea una matriz escalar (es decir: $E(\varepsilon\varepsilon') = \sigma^2 I_N$). Los supuestos H1-H3 y H5 del modelo lineal general siguen siendo validos, el unico supuesto que cambia es H4: $E(\varepsilon\varepsilon') = \sigma^2 \Omega$; además, en este momento no es necesario ningún supuesto particular sobre la distribución de los errores. Los supuestos basicos entonces ahora quedan:

H1 Linealidad: el modelo sigue siendo lineal en el vector de parametros β : la variable dependiente y es una función lineal de β .

H2 Condición de identificación: X es una matriz de rango completo de columnas y $\text{rango}(X) = k$.

H3 $E(\varepsilon) = 0$.

H4 $E(\varepsilon\varepsilon') \neq \sigma^2 I_N$, y $E(\varepsilon\varepsilon') = \sigma^2 \Omega$ donde Ω es una matriz $(N \times N)$ simetrica y definida positiva.

H5 X es una matriz de regresores no estocásticos.

En este contexto, se analizan las propiedades principales del estimador de mínimos cuadrados ordinarios (linealidad, insesgadez, matriz de varianza y covarianza) en presencia de una matriz de varianza y covarianza de los errores no escalar.

Propiedades del estimador de Mínimos cuadrados ordinarios $\hat{\beta}_{MCO}$ cuando $E(\varepsilon\varepsilon') = \sigma^2 \Omega$

- **Linealidad:**

$$\begin{aligned}\hat{\beta}_{MCO} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + \varepsilon) \\ &= \underbrace{(X'X)^{-1}X'X}_{I}\beta + (X'X)^{-1}X'\varepsilon \\ &= \beta + (X'X)^{-1}X'\varepsilon\end{aligned}$$

El estimador $\hat{\beta}_{MCO}$ sigue siendo lineal en β, y y ε .

- **Insesgadez:**

$$\begin{aligned}E(\hat{\beta}_{MCO}) &= E[\beta + (X'X)^{-1}X'\varepsilon] \\ &= E(\beta) + (X'X)^{-1}X'\underbrace{E(\varepsilon)}_{=0} \\ &= \beta\end{aligned}$$

$\hat{\beta}_{MCO}$ es todavia estimador insesgado de β .

• **Matriz de varianza y covarianza de $\hat{\beta}_{MCO}$:**

$$\begin{aligned}
Var(\hat{\beta}_{MCO}) &= E[(\hat{\beta}_{MCO} - E(\beta))(\hat{\beta}_{MCO} - E(\beta))'] \\
&= E[(\hat{\beta}_{MCO} - \beta)(\hat{\beta}_{MCO} - \beta)'] \quad \text{sustituyendo } \hat{\beta}_{MCO} - \beta = (X'X)^{-1}X'\varepsilon \text{ se obtiene:} \\
&= E[((X'X)^{-1}X'\varepsilon)((X'X)^{-1}X'\varepsilon)'] \\
&= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}] \\
&= (X'X)^{-1}X'\underbrace{E(\varepsilon\varepsilon')}_{=\sigma^2\Omega}X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \neq \sigma^2(X'X)^{-1}
\end{aligned}$$

La varianza del estimador $\hat{\beta}_{MCO}$ cuando $E(\varepsilon\varepsilon') = \sigma^2\Omega$ no es la varianza minima de $\hat{\beta}_{MCO}$ cuando $E(\varepsilon\varepsilon') = \sigma^2I$. Si $E(\varepsilon\varepsilon') = \sigma^2\Omega$, no se cumplen todos los supuestos del teorema de Gauss-Markov y el estimador $\hat{\beta}_{MCO}$ no sigue siendo el estimador optimo en el sentido de la varianza minima. Una implicación importante de este resultado es que los test de hipótesis basados sobre $Var(\hat{\beta}_{MCO})$ derivados en las clases anteriores no son validos.

Transformación del modelo lineal general cuando $E(\varepsilon\varepsilon') = \sigma^2\Omega$

Es posible transformar el modelo inicial $y = X\beta + \varepsilon$ de una forma conveniente, de modo que se cumplan todos los supuestos basicos del teorema de Gauss-Markov en el modelo transformado. Dado que Ω es una matriz definida positiva, puede ser factorizada como:

$$\Omega = C\Lambda C'$$

donde C es la matriz de los N vectores caracteristicos de la matriz Ω y Λ la matriz diagonal y simetrica de las N raices características correspondientes. Esa factorización se define tambien *decomposición espectral* de la matriz Ω . La matriz C tiene las siguientes propiedades:

$$\begin{aligned}
C'C &= I \\
C' &= C^{-1} \\
CC' &= CC^{-1} = I
\end{aligned}$$

Se define la matriz no singular ($N \times N$), $P = C\Lambda^{1/2}$ con las siguientes propiedades: $PP' = \Omega$, $[P^{-1}]'[P^{-1}] = \Omega^{-1}$.

$$\begin{aligned}
PP' &= C\Lambda^{1/2}(C\Lambda^{1/2})' \quad \text{dado que } (\Lambda^{1/2})' = \Lambda^{1/2} \\
&= C\Lambda^{1/2}\Lambda^{1/2}C' \\
&= C\Lambda C' = \Omega
\end{aligned}$$

$$\begin{aligned}
[P^{-1}]'[P^{-1}] &= [(C\Lambda^{1/2})^{-1}]'[(C\Lambda^{1/2})^{-1}] \\
&= [(\Lambda^{-1/2}C^{-1})'(\Lambda^{-1/2}C^{-1})] \\
&= (C^{-1})'(\Lambda^{-1/2})'\Lambda^{-1/2}C^{-1} \quad \text{dado que } C^{-1} = C', \quad (C^{-1})' = (C')' = C \\
&= C(\Lambda')^{-1/2}\Lambda^{-1/2}C' \quad \text{dado que } \Lambda' = \Lambda \\
&= C\Lambda^{-1/2}\Lambda^{-1/2}C' = C\Lambda^{-1}C' = \Omega^{-1}
\end{aligned}$$

Se considera la siguiente transformación del modelo lineal general, premultiplicando y por P^{-1} :

$$P^{-1}y = P^{-1}X\beta + P^{-1}\varepsilon$$

Se definen ahora las variables transformadas de la forma siguiente:

$$P^{-1}y = y^*, \quad P^{-1}X = X^*, \quad P^{-1}\varepsilon = \varepsilon^*$$

El modelo inicial transformado es igual a:

$$y^* = X^*\beta + \varepsilon^*$$

Se puede verificar que el modelo transformado de esta forma cumple todos los supuestos del Teorema de Gauss-Markov:

H1 Linealidad: el modelo sigue siendo lineal dado que y^* es función lineal de β .

H2 Condición de identificación: X^* sigue siendo una matriz de rango completo de columnas y $\text{rango}(X^*) = k$.

H3 $E(\varepsilon^*) = 0$ dado que: $E(\varepsilon^*) = E(P^{-1}\varepsilon) = \underbrace{P^{-1}E(\varepsilon)}_{=0} = 0$

H4 $E(\varepsilon^* \varepsilon^{*'}) :$

$$\begin{aligned}
E(\varepsilon^* \varepsilon^{*'}) &= E(P^{-1}\varepsilon(P^{-1}\varepsilon)') \\
&= E(P^{-1}\varepsilon\varepsilon'[P^{-1}]') \\
&= P^{-1} \underbrace{E(\varepsilon\varepsilon')}_{=\sigma^2\Omega} [P']^{-1} \\
&= \sigma^2 P^{-1} \underbrace{\Omega}_{=PP'} [P']^{-1} \\
&= \sigma^2 \underbrace{P^{-1}P}_{=I} \underbrace{P' [P']^{-1}}_{=I} \\
&= \sigma^2 I
\end{aligned}$$

H5 X^* sigue siendo una matriz de regresores no estocásticos.

El modelo transformado entonces satisface todos los supuestos básicos del teorema de Gauss-Markov, en particular la matriz de varianzas y covarianza de los errores transformados es escalar. Esta transformación nos permite derivar el estimador de Mínimos cuadrados generalizados.

Estimador de Mínimos cuadrados generalizados $\hat{\beta}_{MCG}$

El estimador de mínimos cuadrados generalizados, $\hat{\beta}_{MCG}$, se define como el *estimador de mínimos cuadrados ordinarios de β en el modelo transformado*:

$$\begin{aligned}
\hat{\beta}_{MCG} &= (X^{*'} X^*)^{-1} X^{*'} y^* \\
&= ((P^{-1}X)'(P^{-1}X))^{-1} (P^{-1}X)'(P^{-1}y) \\
&= X' \underbrace{(P^{-1})'(P^{-1})}_{=\Omega^{-1}} X^{-1} \underbrace{X'(P^{-1})'(P^{-1})}_{=\Omega^{-1}} y \\
&= (X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} y)
\end{aligned}$$

$\hat{\beta}_{MCG}$ se puede escribir también como:

$$\begin{aligned}
\hat{\beta}_{MCG} &= (X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} (X\beta + \varepsilon)) \\
&= \underbrace{(X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} X)}_{=I} \beta + (X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} \varepsilon) \\
&= \beta + (X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} \varepsilon)
\end{aligned}$$

Propiedades de $\hat{\beta}_{MCG}$ en muestras finitas:

- **Insesgadez:**

$$\begin{aligned}
E(\hat{\beta}_{MCG}) &= E[\beta + (X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} \varepsilon)] \\
&= E(\beta) + (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \underbrace{E(\varepsilon)}_{=0} \\
&= \beta
\end{aligned}$$

• **Matriz de varianza y covarianza:**

$$\begin{aligned}
Var(\hat{\beta}_{MCG}) &= E[(\hat{\beta}_{MCG} - E(\hat{\beta}_{MCG}))(\hat{\beta}_{MCG} - E(\hat{\beta}_{MCG}))'] \\
&= E[(\hat{\beta}_{MCG} - \beta)(\hat{\beta}_{MCG} - \beta)'] \quad \text{sustituyendo } \hat{\beta}_{MCG} - \beta = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\varepsilon) \text{ se obtiene:} \\
&= E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon)((X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon)'] \\
&= E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon\varepsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}] \\
&= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\underbrace{E(\varepsilon\varepsilon')}_{=\sigma^2\Omega}\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\
&= \sigma^2(X'\Omega^{-1}X)^{-1}X'\underbrace{\Omega^{-1}\Omega}_{=I}X(X'\Omega^{-1}X)^{-1} \\
&= \sigma^2(X'\Omega^{-1}X)^{-1}\underbrace{X'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}}_{=I} \\
&= \sigma^2(X'\Omega^{-1}X)^{-1}
\end{aligned}$$

- **Eficiencia:** dado que el modelo transformado satisface todos los supuestos del teorema de Gauss-Markov, ya sabemos que el estimador $\hat{\beta}_{MCG}$ es el estimador lineal, insesgado y optimo de β en el sentido de la varianza minima. En particular, se demuestra que:

$$Var(\hat{\beta}_{MCO}) - Var(\hat{\beta}_{MCG}) = A \geq 0 \text{ (semidefinida positiva)}$$

Prueba:

$$\begin{aligned}
Var(\hat{\beta}_{MCO}) - Var(\hat{\beta}_{MCG}) &= \sigma^2(X'X)^{-1}X'\underbrace{\Omega}_{=PP'}X(X'X)^{-1} - \sigma^2(X'\Omega^{-1}X)^{-1} \\
&= \sigma^2[\underbrace{(X'X)^{-1}X'PP'X(X'X)^{-1}}_{\substack{=D' \\ (k \times N)}} - (X'\Omega^{-1}X)^{-1}] \\
&= \sigma^2[D'D - (X'\Omega^{-1}X)^{-1}]
\end{aligned}$$

Dado que $D'P^{-1}X = I_k$, premultiplicamos el segundo termino en la parentesis cuadrada por $D'P^{-1}X$ y lo postmultiplicamos por $(D'P^{-1}X)' = X'(P^{-1})'D$

$$\begin{aligned}
Var(\hat{\beta}_{MCO}) - Var(\hat{\beta}_{MCG}) &= \sigma^2[D'D - D'P^{-1}X(X'\Omega^{-1}X)^{-1}X'(P^{-1})'D] \\
&= \sigma^2[D'(I - P^{-1}X(X'\Omega^{-1}X)^{-1}X'(P^{-1})')D] \\
&= \sigma^2[D'\underbrace{(I - X^*(X^{*'}X^*)^{-1}X^{*'})}_{=M^*}_{(N \times N)}D] \\
&= \sigma^2D'M^*D \geq 0 \text{ semidefinida positiva}
\end{aligned}$$

dado que M^* es idempotente. La matriz M^* en el modelo transformado es la matriz que “produce” los residuos estimados en una regresión lineal de MCO de la variable dependiente transformada y^* sobre los regresores transformados X^* .

Estimación de mínimos cuadrados generalizados de σ^2 :

El estimador de mínimos cuadrados ordinarios de σ^2 , $\hat{\sigma}_{MCO}^2$, es sesgado si $E(\varepsilon\varepsilon') = \sigma^2\Omega$:

$$\begin{aligned}
E(\hat{\sigma}_{MCO}^2) &= E\left(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{N-k}\right) \\
&= E\left(\frac{\varepsilon'M\varepsilon}{N-k}\right) \\
&= E\left(\frac{tr(\varepsilon'M\varepsilon)}{N-k}\right) \\
&= \frac{tr(ME(\varepsilon\varepsilon'))}{N-k} \\
&= \frac{\sigma^2 tr(M\Omega)}{N-k} \neq \sigma^2
\end{aligned}$$

No obstante, el estimador de σ^2 obtenido utilizando los residuos estimados del modelo transformado es insesgado. Sea $(\hat{\sigma}_{MCO}^*)^2 = \frac{\hat{\varepsilon}^{*'} \hat{\varepsilon}^*}{N-k}$:

$$\begin{aligned}
E\left(\frac{\hat{\varepsilon}^{*'} \hat{\varepsilon}^*}{N-k}\right) &= E\left(\frac{\varepsilon^{*'} M^* \varepsilon^*}{N-k}\right) \\
&= E\left(\frac{tr(\varepsilon^{*'} M^* \varepsilon^*)}{N-k}\right) \\
&= \frac{tr(M^* E(\varepsilon^* \varepsilon^{*'}))}{N-k} \\
&= \frac{\sigma^2 tr(M^*)}{N-k} = \frac{\sigma^2 (N-k)}{N-k} = \sigma^2
\end{aligned}$$