Tests de Hipotesis en Base al Principio de Verosimilitud

Walter Sosa-Escudero
Universidad de San Andres, Argentina

Conceptos basicos

$Y \sim f(y;\theta)$, $\theta \in \Theta \subseteq \Re^K$.

- θ es un vector de K parametros.
- Θ es el 'espacio de parametros': conjunto de los valores que puede tomar θ

Una hipotesis es una aseveracion acerca de los valores que puede tomar θ. El objetivo es, en base a una muestra, 'aprender' algo acerca de si la hipotesis es verdadera o falsa.

Si $H_0 : \theta \in \Theta_0$, la hipotesis es anidada si $\Theta_0 \subseteq \Theta$. En general, son hipotesis de *restriccion* (ej, $\beta = 0$, $\Theta = \Re, \Theta_0 = \{0\}$).

- Un estadistico para el test (test statistic) es una variable aleatoria que tiene una distribucion cuando H_0 es cierta y otra distinta cuando es falsa.
- La hipotesis alternativa, H_A, es un subconjunto de los valores que puede tomar θ cuando la H_0 es falsa.
  Ejemplo: $K = 1$, $\Theta = \Re$, $H_0 : \theta = 0$, $H_A : \theta > 0$.
- La region de rechazo es el conjunto de los valores que puede tomar el estaditistico para el cual la hipotesis sera rechazada.
- Un test es un estadistico combinado con una regla de rechazo.

- La probabilidad de que el test rechace H_0 cuando es cierta (error de Tipo I) es el nivel del test. Depende estrictamente de la hipotesis nula.
- La potencia de un test es la probabilidad de que el test rechace H_0. Cuando H_0 es falsa, esta probabilidad depende de cual es H_A.
- La probabilidad de que el test no rechace H_0 cuando es falsa (error de tipo II) es 1 menos la potencia.

- Existe en general un trade-off entre las probabilidades de cometer errores de tipo I y II: un test que tiene nivel 0 (acepta siempre) tiende a tener potencia cero (no rechaza nunca).
- El enfoque clasico (Neyman-Pearson) propone fijar exogenamente un nivel, y buscar tests que maximicen la potencia para alguna hipotesis alternativa relevante.
- Un test es consistente si cuando la hipotesis nula es falsa, la potencia del test tiende a 1 cuando la muestra tiende a infinito.

Hipotesis no-lineales

$H_0 : h(\theta) = 0, \;\; h(\theta) : \Re^K \rightarrow \Re^r,$
$h(\theta)$ es un vector de r restricciones sobre K parametros'
En este caso, $\Theta = \Re^k$, y

$$\Theta_0 = \left\{\theta \in \Theta \mid h(\theta) = 0\right\}, \Rightarrow \Theta_0 \subseteq \Theta$$

,de modo que la hipotesis es anidada.
$D(\theta)$, matriz $(r \times K)$ con elemento carateristico:

$$D(\theta)_{ij} = \frac{\partial h(\theta)_i}{\partial \theta_j}$$

Supondremos que $\rho(D(\theta)) = r$

Ejemplo: $\theta = (\theta_1, \theta_2, \theta_3)'$. Consideremos $H_0 : h(\theta) = 0$, con

$$h(\theta) = \begin{bmatrix} \theta_1 - \theta_2 \\ \theta_2\theta_3 - 1 \end{bmatrix}$$

Esto es equivalente a $H_0 : \theta_1 = \theta_2 \ , \ \theta_2\theta_3 = 1$. Son $p = 2$ restricciones sobre K parametros.

En este caso:

$$D(\theta) = \begin{bmatrix} 1 & -1 & 0 \\ 0 & \theta_3 & \theta_2 \end{bmatrix}$$

- El estimador MV sin restringir, $\hat{\theta}$ satisface $\hat{\theta} \equiv \text{argmax}_{\theta \in \Theta} \; l(\theta; \mathbf{y})$, y bajo condiciones de regularidad:

$$s(\hat{\theta}, \mathbf{y}) = 0$$

- El estimador MV restringido $\hat{\theta}_R$ satisface:
  $\hat{\theta} \equiv \text{argmax}_{\theta \in \Theta_0} \; l(\theta; \mathbf{y})$
  El Lagraneano del problema es:

$$LG(\theta, \lambda) = l(\theta; \mathbf{y}) - \hat{\lambda} \, h(\theta)$$

y las condiciones de primer orden son:

$$\left.\begin{array}{rcl} s(\hat{\theta}_R; \mathbf{y}) - D(\hat{\theta}_R)'\lambda & = & 0 \\ h(\hat{\theta}_R) & = & 0 \end{array}\right\}$$

Tres tests

Supongamos que V_n es un estimador consistente de $I(\theta_0, Y)$.
Consideremos los siguientes principios para evaluar $H_0 : h(\theta_0) = 0$

1. Razon de verosimilitud (LR):

$$LR = n\ 2\left[\frac{1}{n}\ l(\hat{\theta}; \mathbf{y}) - \frac{1}{n} l(\hat{\theta}_R; \mathbf{y})\right]$$

2. Wald (W):

$$W = n\ h(\hat{\theta})'\ \left[D(\hat{\theta}) V_n(\hat{\theta})^{-1} D(\hat{\theta})'\right]^{-1} h(\hat{\theta})$$

3. Score (o Multiplicador de Lagrange) (LM):

$$LM = n\ \left[\frac{s(\hat{\theta}_R)}{n}\right]'\ V_n(\hat{\theta}_R)^{-1} \left[\frac{s(\hat{\theta}_R)}{n}\right]$$

Resultado: Si $y_i \sim f(y;\theta_0), i = 1, 2, \ldots, n$, bajo $H_0 : h(\theta_0) = 0$, LR, W y LM tienen distribucion asintotica $\chi^2(p)$ y bajo $H_A : h(\theta_0) = \delta$ tienden a infinito para cualquier constante $\delta \neq 0$.

- Los tres tests son asintoticamente equivalentes.
- Los tres tests son consistentes. (porque?)

Intuicion: $\hat{\theta}$ siempre es consistente para θ_0. $\hat{\theta}_R$ es consistente para θ_0 solo si H_0 es cierta. Entonces, bajo H_0:

- $\hat{\theta}$ y $\hat{\theta}_R$ son consistentes, entonces $l(\hat{\theta}) - l(\hat{\theta}_R) \cong 0$ (*razon de verosimilitud*).
- $h(\hat{\theta}) \cong 0$ (*principio de Wald*)
- El 'precio sombra' de imponer la restriccion es cero: $\lambda \cong 0$. De la primer condicion de primer orden:

$$s(\hat{\theta}_R; \mathbf{y}) - D(\hat{\theta}_R)'\lambda = 0$$

Como $\rho(H(\theta)) = r$, entonces $s(\hat{\theta}_R) \cong 0$ (*principio de score o multiplicador de lagrange*).

- La idea es rechazar H_0 cuando LR,W o LM toman valores altos de acuerdo a la distribucion $\chi^2(p)$.

Distribuciones asintoticas bajo $H_0 : h(\theta_0) = 0$

Para las pruebas necesitamos algunos resultados. Ya probamos que:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I(\theta)^{-1})$$

- Tambien necesitamos el siguiente resultado: Si Z es un vector $K \times 1$ tal que $Z \sim N(0, D)$, entonces:

$$Z' D^{-1} Z \sim \chi^2(K)$$

- Recordar que: $I(\theta_0, \mathbf{y}) = nI(\theta_0, y)$

Test de Wald

$$W = n\, h(\hat{\theta})' \left[D(\hat{\theta}) V_n^{-1} D(\hat{\theta})' \right]^{-1} h(\hat{\theta})$$

Tomemos una expansion de Taylor de $h(\hat{\theta})$ en θ_0:

$$h(\hat{\theta}) \cong h(\theta_0) + D(\theta_0) \left(\hat{\theta} - \theta_0 \right)$$

Notar que la aproximacion es asintoticamente valida ya que $\hat{\theta}$ es consistente. Multiplicando por $\sqrt{n}$ y bajo $H_0 : h(\theta_0) = 0$

$$\sqrt{n}\, h(\hat{\theta}) \cong D(\theta_0) \sqrt{n} \left(\hat{\theta} - \theta_0 \right)$$

Por normalidad asintotica de $\hat{\theta}$:

$$\sqrt{n}\, h(\hat{\theta}) \xrightarrow{d} N\Big(0, D(\theta_0) I(\theta_0)^{-1} D(\theta_0)'\Big)$$

Como $V_n \xrightarrow{p} I(\theta_0)$ y $\hat{\theta} \xrightarrow{p} \theta_0$, por teorema de Slutzky:

$$\sqrt{n}\, h(\hat{\theta}) \xrightarrow{d} N\left(0, D(\hat{\theta}) V_n^{-1} D(\hat{\theta})'\right)$$

Entonces, por el resultado anterior:

$$W = n\, h(\hat{\theta})' \left[D(\hat{\theta}) V_n^{-1} D(\hat{\theta})'\right]^{-1} h(\hat{\theta}) \xrightarrow{d} \chi^2(r)$$

Mas resultados...

Para probar los otros dos casos, necesitamos dos resultados adicionales:

- $$\sqrt{n}(\hat{\theta}_R - \theta_0) \xrightarrow{d} N(0, \tilde{V}(\theta_0)^{-1})$$

  con $\tilde{V}(\theta_0)^{-1} \equiv I^{-1} - I^{-1}D'[DI^{-1}D']^{-1}DI^{-1}$, $I \equiv I(\theta_0, Y)$, $D \equiv D(\theta_0)$

- $$\frac{1}{\sqrt{n}}\hat{\lambda} \simeq -R(\theta_0)^{-1}D(\theta_0)I(\theta_0)^{-1}\frac{s(\theta_0)}{\sqrt{n}} \xrightarrow{d} N(0, R(\theta_0)^{-1})$$

  con $R(\theta_0) \equiv D(\theta_0)I(\theta_0)^{-1}D(\theta_0)'$

Prueba: de las condiciones de primer orden

$$\left.\begin{array}{rcl} \frac{1}{\sqrt{n}} s(\hat{\theta}_R; \mathbf{y}) - \frac{1}{\sqrt{n}} D(\hat{\theta}_R)' \hat{\lambda} & = & 0 \\ \sqrt{n}\, h(\hat{\theta}_R) & = & 0 \end{array}\right\} \quad (1)$$

Tomemos una expansion de Taylor en θ_0 para ambas ecuaciones:

$$\begin{aligned} s(\hat{\theta}_R) &\simeq s(\theta_0) + H(\theta_0)(\hat{\theta}_R - \theta_0) \\ \frac{1}{\sqrt{n}} s(\hat{\theta}_R) &\simeq \frac{1}{\sqrt{n}} s(\theta_0) + \sqrt{n} \frac{H(\theta_0)}{n} (\hat{\theta}_R - \theta_0) \\ &\simeq \frac{1}{\sqrt{n}} s(\theta_0) - \sqrt{n} I(\theta_0)(\hat{\theta}_R - \theta_0) \qquad \text{(Porque?)} \quad (2) \end{aligned}$$

$$\begin{aligned}\sqrt{n}h(\hat{\theta}_R) &\simeq \sqrt{n}h(\theta_0) + \sqrt{n}D(\theta_0)(\hat{\theta}_R - \theta_0) \\ &\simeq \sqrt{n}D(\theta_0)(\hat{\theta}_R - \theta_0) \qquad \text{(Porque?)} \quad (3)\end{aligned}$$

Ahora reemplazo (2) y (3) en (1) y tomo $D(\hat{\theta}_R) \simeq D(\theta_0)$

$$\left.\begin{aligned}\frac{1}{\sqrt{n}}s(\theta_0) - \sqrt{n}I(\theta_0)(\hat{\theta}_R - \theta_0) - \frac{1}{\sqrt{n}}D(\theta_0)'\hat{\lambda} &= 0 \\ \sqrt{n}D(\theta_0)(\hat{\theta}_R - \theta_0) &= 0\end{aligned}\right\} \quad (1)$$

Despejando (es denso...):

$$\sqrt{n}(\hat{\theta}_R - \theta_0) = A(\theta_0)\frac{1}{\sqrt{n}}s(\theta_0)$$

con $A(\theta_0) \equiv I^{-1} - I^{-1}D'R^{-1}DI^{-1}$

Notar que:

$$\frac{1}{\sqrt{n}} s(\theta_0) = \sqrt{n} \frac{s(\theta_0)}{n} = \sqrt{n} \frac{\sum_{i=1}^{n} s(\theta_0, y_i)}{n} \xrightarrow{d} N\big(0, I(\theta_0)\big)$$

,por TCL. Por Slutzky,

$$\sqrt{n}(\hat{\theta}_R - \theta_0) = A(\theta_0) \frac{1}{\sqrt{n}} s(\theta_0) \quad \xrightarrow{d} \quad N(0, A(\theta_0) I(\theta_0) A(\theta_0)')$$

que es el resultado buscado.

Dejamos como ejercicio el correspondiente a $\hat{\lambda}$.

Test LM

$$LM = \frac{1}{n}\, s(\hat{\theta}_R)' V(\hat{\theta}_R)^{-1} s(\hat{\theta}_R)$$

De la condicion de primer orden:

$$\frac{1}{\sqrt{n}} s(\hat{\theta}_R; \mathbf{y}) - \frac{1}{\sqrt{n}} D(\hat{\theta}_R)' \hat{\lambda} = 0$$

Reemplazando:

$$\frac{\hat{\lambda}'}{\sqrt{n}} D(\hat{\theta}_R) V(\hat{\theta}_R)^{-1} D(\hat{\theta}_R)' \frac{\hat{\lambda}}{\sqrt{n}} \simeq \frac{\hat{\lambda}'}{\sqrt{n}} R(\theta_0) \frac{\hat{\lambda}}{\sqrt{n}} \xrightarrow{d} \chi^2(r)$$

,dado que $\hat{\lambda}/\sqrt{n} \xrightarrow{d} N(0, R(\theta)^{-1})$, y que es una forma cuadratica.

Test de razon de verosimilitud (LR)

$$LR = 2\left[l(\hat{\theta};\mathbf{y}) - l(\hat{\theta}_R;\mathbf{y})\right]$$

Consideremos una expansion de segundo orden de $l(\hat{\theta}_R;\mathbf{y})$ en $\hat{\theta}$ bajo H_0:

$$\begin{aligned} l(\hat{\theta}_R) &\simeq l(\hat{\theta}) + s(\hat{\theta})'\left(\hat{\theta}_R - \hat{\theta}\right) + \frac{1}{2}\left(\hat{\theta}_R - \hat{\theta}\right)' H(\hat{\theta})\left(\hat{\theta}_R - \hat{\theta}\right) \\ &\simeq l(\hat{\theta}) + \frac{1}{2}\left(\hat{\theta}_R - \hat{\theta}\right)' H(\hat{\theta})\left(\hat{\theta}_R - \hat{\theta}\right) \qquad \text{(Porque?)} \end{aligned}$$

Reemplazando y dividiendo y multiplicando por n:

$$\begin{aligned} LR &\simeq -\sqrt{n}\left(\hat{\theta}_R - \hat{\theta}\right)'\left(\frac{1}{n}H(\hat{\theta})\right)\sqrt{n}\left(\hat{\theta}_R - \hat{\theta}\right) \\ &\sqrt{n}\left(\hat{\theta}_R - \hat{\theta}\right)' I(\theta_0)\sqrt{n}\left(\hat{\theta}_R - \hat{\theta}\right) \qquad \text{(Porque?)} \end{aligned}$$

Recordar:

$$\begin{aligned}\sqrt{n}(\hat{\theta}-\theta_0) &\simeq I^{-1}\frac{s(\theta_0)}{\sqrt{n}}\\ \sqrt{n}(\hat{\theta}_R-\theta_0) &\simeq \left(I^{-1}-I^{-1}D'R^{-1}DI^{-1}\right)\frac{s(\theta_0)}{\sqrt{n}}\end{aligned}$$

Entonces:

$$\sqrt{n}\left(\hat{\theta}_R-\hat{\theta}\right)\simeq -I^{-1}D'R^{-1}DI^{-1}\frac{s(\theta_0)}{\sqrt{n}}=-I^{-1}D'\frac{\hat{\lambda}}{\sqrt{n}}$$

Reemplazando arriba:

$$LR\simeq \frac{\hat{\lambda}'}{\sqrt{n}}DI^{-1}II^{-1}D'\frac{\hat{\lambda}'}{\sqrt{n}}=\frac{\hat{\lambda}'}{\sqrt{n}}DI^{-1}D'\frac{\hat{\lambda}'}{\sqrt{n}}\simeq LM\xrightarrow{d}\chi^2(r)$$

Consistencia de LR, LM y W

Supongamos $H_A : h(\theta_0) = \delta \neq 0$

$$\begin{aligned} W &= n\, h(\hat{\theta})' \left[D(\hat{\theta}) V_n^{-1} D(\hat{\theta})' \right]^{-1} h(\hat{\theta}) \\ &\to n\, \delta' \left[D(\theta_0) V_n^{-1} D(\theta_0)' \right]^{-1} \delta \to \infty \end{aligned}$$

Notar que bajo H_A, plim $s(\hat{\theta}_R)/n \neq 0$. Entonces:

$$\begin{aligned} LM &= \frac{1}{n}\, s(\hat{\theta}_R)' V(\hat{\theta}_R)^{-1} s(\hat{\theta}_R) \\ &= n\, \frac{s(\hat{\theta}_R)'}{n} V(\hat{\theta}_R)^{-1} \frac{s(\hat{\theta}_R)'}{n} \to \infty \end{aligned}$$

Un argumento similar puede ser utilizado con LR

$$LR = 2\left[l(\hat{\theta};\mathbf{y}) - l(\hat{\theta}_R;\mathbf{y})\right] = n\,2\underbrace{\left[\frac{l(\hat{\theta};\mathbf{y})}{n} - \frac{l(\hat{\theta}_R;\mathbf{y})}{n}\right]}_{\neq 0} \to \infty$$

(recordar la prueba de consistencia...)

Entonces, los tres tests son consistentes: bajo H_0 tienen distribucion asintotica $\chi^2(r)$, de modo que la region de aceptacion es *finita*. Entonces, como bajo H_A tienden a ∞, entonces caen fuera de la region de aceptacion.

Aplicacion: Test de heterocedasticidad

Cuestiones Preliminares

Recordar que un test para la hipotesis nula $h(\theta) = 0$ puede basarse en:

$$LM = n \left[\frac{1}{n} s(\hat{\theta}_R)\right]' I_n^{-1} \left[\frac{1}{n} s(\hat{\theta}_R)\right]$$

en donde I_n^{-1} es un estimador consistente de la matriz de informacion $I(\theta_0, y)$ y $\hat{\theta}_R$ es el estimador maximo verosimil restringido bajo H_0.

1. Test LM para un subvector

Hasta ahora: $H_0 : h(\theta) = 0$. Caso Particular: $\theta = (\theta_1 \quad \theta_2)'$, $H_0 : \theta_2 = \theta_{2o}$, θ_2 es $r \times 1$.
Notar que:

$$H_0 : \underbrace{\begin{bmatrix} 0 & I_r \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} - \theta_{2o}}_{h(\theta)} = 0$$

con $H(\theta) = [0 \quad I_r]$. En forma similar:

$$s(\theta) \begin{bmatrix} s_1(\theta) \\ s_1(\theta) \end{bmatrix}, \qquad I_n(\theta) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}, \qquad I_n^{-1}(\theta) \equiv \begin{bmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{bmatrix}$$

Resultado (Theil, 1971, pp. 18): $I^{22} = \left[I_{22} - I_{21} I_{11}^{-1} I_{12}\right]^{-1}$. En particular, si $I_{21} = 0 \Longrightarrow I^{22} = I_{22}^{-1}$.

CPO del problema de maximizacion restringida: en este caso, $\hat{\theta}_R = (\tilde{\theta}_1, \theta_{20})$ tal que:

$$\begin{cases} s_1(\tilde{\theta}_1, \theta_{20}) &= 0 \\ s_2(\tilde{\theta}_1, \theta_{20}) &= \tilde{\lambda} \\ \tilde{\theta}_2 &= \theta_{20} \end{cases}$$

Reemplazando y simplificando en la formula del test LM:

$$LM = n \left[n^{-1} s_2(\tilde{\theta}_1; \theta_{20}) \right]' I^{22}(\tilde{\theta}) \left[n^{-1} s_2(\tilde{\theta}_1; \theta_{20}) \right]$$

Si ademas $I_{21} = 0$

$$LM = n \left[n^{-1} s_2(\tilde{\theta}_1; \theta_{20}) \right]' I_{22}(\tilde{\theta}) \left[n^{-1} s_2(\tilde{\theta}_1; \theta_{20}) \right]$$

2. Algunas cuestiones algebraicas de MCO

$$SCE = \sum(\hat{Y}_i - \bar{Y})^2 = \sum \hat{Y}_i^2 - N\bar{Y}^2 = \sum \hat{Y}_i^2 - (1/N)(\sum Y_i)^2$$

En terminos matriciales:

$$SCE = \hat{Y}'\hat{Y} - (i'Y)^2/N, \quad \text{con } i = (1, 1, \ldots, 1)'$$

Pero:

$$\hat{Y}'\hat{Y} = \underbrace{Y'X(X'X)^{-1}X'}_{(X\hat{\beta}')}\underbrace{X(X'X)^{-1}X'Y}_{X\hat{\beta}} = Y'X(X'X)^{-1}X'Y$$

Reemplazando: $SCE = Y'X(X'X)^{-1}X'Y - (i'Y)^2/N$

Test de Breusch-Pagan (1979)

$$y_i = x_i'\beta + u_i, \quad i = 1, 2, \ldots, N$$

- x_i vector de K variables explicativas, con constante.
- $u_i = N(0, \sigma_i^2)$
- $V(u_i) = \sigma_i^2 = h(z_i'\alpha)$
- z_i vector de p variables explicativas de la varianza (con constante)
- $h(.) =$ cualquier funcion positiva con dos derivadas.
- Homocedasticidad: $H_0 : \alpha_2 = \alpha_3 = \cdots = \alpha_p = 0$

- e_i, residuos de OLS ignorando heterocedasticidad.
- $\hat{\sigma}^2 = (1/N)\sum e_i^2$. Estimador MV de σ^2 bajo H_0
- $g_t \equiv e_i^2/\hat{\sigma}^2$

Test (Breusch-Pagan): Bajo H_0 :

$$LM = \frac{1}{2} SCE_{g,z} \sim \chi^2(p-1)$$

$SCE_{g,z} = SCE$ de regresar g_i en z_i.

Intuicion: irrelevancia de un modelo para la varianza.

Prueba: $y_i \sim N(x_i'\beta, \sigma_i^2)$, $\sigma_i^2 = h(z_i'\alpha)$.

$$l(\beta,\alpha) = -\frac{1}{2}n\ln(2\pi) - \frac{1}{2}\sum \ln \sigma_i^2 - \frac{1}{2}\sum (1/\sigma_i^2)\left(y_i - x_i'\beta\right)^2$$

Es facil verificar que

$$I_{\alpha,\beta} = -E\left[\frac{\partial^2 l(\beta,\alpha)}{\partial\alpha\partial\beta}\right] = 0$$

Como H_0 es del tipo $\theta_{2o} = 0$ (subvector):

$$LM = \frac{1}{n}\hat{s}_\alpha(\hat{\theta}_R)'\ \hat{I}_{\alpha\alpha}^{-1}(\hat{\theta}_R)\ \hat{s}_\alpha(\hat{\theta}_R)$$

con $\hat{\theta}_R = (\hat{\beta}, \hat{\alpha}_1)$, $\hat{\beta}$ es el estimador MCO, y $\hat{\alpha}_1$ satisface $\hat{\sigma}^2 = h(\hat{\alpha}_1)$.

ASA se puede mostrar

$$\begin{aligned}\hat{s}_\alpha &= \frac{1}{2}\left[\frac{h'(\hat{\alpha}_1)}{\hat{\sigma}^2}\right]\sum z_i\left(\frac{e_i^2}{\hat{\sigma}^2}-1\right)\\ \hat{I}_{\alpha\alpha} &= \frac{1}{2}\left[\frac{h'(\hat{\alpha}_1)}{\hat{\sigma}^2}\right]\sum z_i z_i'\end{aligned}$$

Reemplazando y simplificando:

$$LM = \frac{1}{2}\left(\sum z_i f_i\right)'\left(\sum z_i z_i'\right)^{-1}\left(\sum z_i f_i\right)$$

con $f_i \equiv e_i^2/\hat{\sigma}^2 - 1 = g_i - 1$. en terminos matriciales:

$$LM = \frac{1}{2} f' Z (Z'Z)^{-1} Z' f$$

$$LM = \frac{1}{2} f' Z (Z'Z)^{-1} Z' f$$

En base a :

$$f = g - i; \;\; i'g = N; \;\; i'f = 0; \;\; f' Z (Z'Z)^{-1} Z' i = 0$$

reemplazando $f = g - i$ y simplificando:

$$LM = \frac{1}{2} \left[g' Z (Z'Z)^{-1} Z' g - \frac{(i'g)^2}{N} \right] = \frac{1}{2} \, SCE_{g,z}$$