Universidad Torcuato Di Tella
Master en Economía y en Econometría
2019

Econometría
Problem Set 5: Máxima Verosimilitud

1. En el problema de maximización de la log-likelihood, la condición de segundo orden se expresa como una restricción sobre el "signo" de la matriz hessiana en el óptimo (se pide que sea definida negativa):

$$H(\widehat{\beta}_{ml}, \widehat{\sigma}^2_{ml}; y, X) = D^2 l(\widehat{\beta}_{ml}, \widehat{\sigma}^2_{ml}; y, X)$$

La matriz hessiana se vincula con la matriz de información de la siguiente forma:

$$I_n(\beta, \sigma^2_u; y, X) = E\left[-H(\beta, \sigma^2_u; y, X)\right]$$

Consideremos:

$$y = X\beta + u$$

$$u\tilde{\ }N(0, \sigma^2_u I)$$

1. La función likelihood para una muestra de n errores independientes, i.i.d. es:

$$L\left(\beta, \sigma^2_u; u\right) = \left(\frac{1}{\sqrt{\sigma^2_u 2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum(u_i)^2} = \left(\frac{1}{\sqrt{\sigma^2_u 2\pi}}\right)^n e^{-\frac{1}{2\sigma^2_u}u'u}$$

Teniendo en cuenta que $u = y - X\beta$:

$$\begin{aligned}
L\left(\beta, \sigma^2_u; y, X\right) &= \left(\frac{1}{\sqrt{2\pi\sigma^2_u}}\right)^n \exp\left\{-\frac{1}{2\sigma^2_u}\left(y - X\beta\right)^T\left(y - X\beta\right)\right\} \\
l(\beta, \sigma^2_u; y, X) &= \ln\left[L\left(\beta, \sigma_u; y, X\right)\right] \\
&= -\frac{n}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\left(\sigma^2_u\right) - \frac{1}{2\sigma^2_u}\left(y - X\beta\right)^T\left(y - X\beta\right) \\
&= -\frac{n}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\left(\sigma^2_u\right) - \frac{1}{2\sigma^2_u}(y^T y + \beta^T X^T X\beta - 2\beta^T X^T y)
\end{aligned}$$

El score es:

$$\begin{aligned}
S(y,X;\beta,\sigma_u) &= \nabla l(\beta,\sigma_u;y,X) \\
&= \begin{bmatrix} \nabla_\beta l(\beta,\sigma_u;y,X) \\ D_{\sigma_u} l(\beta,\sigma_u;y,X) \end{bmatrix} \\
&= \begin{bmatrix} -\frac{1}{2\sigma_u^2}\nabla SSR(\beta) \\ D_{\sigma_u} l(\beta,\sigma_u;y,X) \end{bmatrix} \\
&= \begin{bmatrix} -\frac{1}{2\sigma_u^2}\left[-2X^T y + 2X^T X\beta\right] \\ -\frac{n}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}(y-X\beta)^T(y-X\beta) \end{bmatrix} \\
&= \begin{bmatrix} -\frac{1}{\sigma_u^2}\left[-X^T y + X^T X\beta\right] \\ -\frac{n}{\sigma_u^2} + \frac{1}{\sigma_u^4}(y-X\beta)^T(y-X\beta) \end{bmatrix}
\end{aligned}$$

De igualar a 0 se obtenían los estimadores ml:

$$\widehat{\beta}_{ml} = (X'X)^{-1}X'y$$

$$\widehat{\sigma}^2_{ml} = \frac{\widehat{u}'\widehat{u}}{n}$$

1. La matriz hessiana es entonces:

$$\begin{aligned}
D^2_\beta l(\beta,\sigma_u^2;y,X) &= D_\beta \nabla_\beta l(\beta,\sigma_u^2;y,X) = \frac{\partial^2 \ln L}{\partial\beta\partial\beta'} = -\frac{1}{\sigma_u^2}X^T X \\
D^2_{\sigma_u^2} l(\beta,\sigma_u^2;y,X) &= \frac{\partial^2 \ln L}{\partial\sigma_u^2\partial\sigma_u^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}u^T u \\
D_{\sigma_u^2}\nabla_\beta l(\beta,\sigma_u^2;y,X) &= \frac{\partial^2 \ln L}{\partial\beta\partial\sigma_u^2} = -\frac{1}{\sigma_u^4}X^T u \\
H(\beta,\sigma_u^2;y,X) &= \begin{bmatrix} \frac{\partial^2 \ln L}{\partial\beta\partial\beta'} & \frac{\partial^2 \ln L}{\partial\beta\partial\sigma_u^2} \\ \frac{\partial^2 \ln L}{\partial\sigma_u^2\partial\beta'} & \frac{\partial^2 \ln L}{\partial\sigma^2\partial\sigma_u^2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sigma_u^2}X^T X & -\frac{1}{\sigma_u^4}X^T u \\ -\frac{1}{\sigma_u^4}u^T X & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}u^T u \end{bmatrix}
\end{aligned}$$

Evaluada en el estimador de máxima verosimilitud,

$$\begin{aligned}
H(\widehat{\beta}_{ml},\widehat{\sigma}^2_{ml};y,X) &= \begin{bmatrix} -\frac{1}{\widehat{\sigma}^2_{ml}}X^T X & -\frac{1}{\widehat{\sigma}^4_{ml}}X^T u \\ -\frac{1}{\widehat{\sigma}^4_{ml}}u^T X & \frac{n}{2\widehat{\sigma}^4_{ml}} - \frac{1}{\widehat{\sigma}^6_{ml}}\widehat{u}^T_{ml}\widehat{u}_{ml} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{1}{\widehat{\sigma}^2_{ml}}X^T X & 0_k \\ 0_k^T & \frac{n}{2\widehat{\sigma}^4_{ml}} - \frac{1}{\widehat{\sigma}^6_{ml}}\widehat{u}^T_{ml}\widehat{u}_{ml} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{1}{\widehat{\sigma}^2_{ml}}X^T X & 0_k \\ 0_k^T & \frac{n}{2\widehat{\sigma}^4_{ml}} - \frac{1}{\widehat{\sigma}^6_{ml}}n\widehat{\sigma}^2_{ml} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{1}{\widehat{\sigma}^2_{ml}}X^T X & 0_k \\ 0_k^T & -\frac{n}{2\widehat{\sigma}^4_{ml}} \end{bmatrix}
\end{aligned}$$

La condición suficiente de segundo orden exige que esta matriz sea definida negativa. Sea $v^T = \left(v_1^T, v_2\right)$ un vector no nulo de dimensiones tales que:

$$
\begin{aligned}
v^T H(\widehat{\beta}_{ml}, \widehat{\sigma}^2_{ml}; y, X)v &= \left(v_1^T, v_2\right) \left[\begin{array}{cc} -\frac{1}{\widehat{\sigma}^2_{ml}} X^T X & 0_k \\ 0_k^T & -\frac{n}{2\widehat{\sigma}^4_{ml}} \end{array} \right] \left(v_1^T, v_2\right)^T \\
&= \left[\begin{array}{cc} -v_1^T \frac{1}{\widehat{\sigma}^2_{ml}} X^T X & -v_2 \frac{n}{2\widehat{\sigma}^4_{ml}} \end{array} \right] \left(\begin{array}{c} v_1 \\ v_2 \end{array} \right) \\
&= -v_1^T \frac{1}{\widehat{\sigma}^2_{ml}} X^T X v_1 - v_2 \frac{n}{2\widehat{\sigma}^4_{ml}} v_2 < 0
\end{aligned}
$$

Como $X^T X$ es definida positiva bajo el supuesto de rango completo,

$$
-v_1^T \left[\frac{1}{\widehat{\sigma}^2_{ml}} X^T X \right] v_1 = -\frac{1}{\widehat{\sigma}^2_{ml}} v_1^T \left[X^T X \right] v_1 < 0
$$

Además,

$$
-v_2^2 \frac{n}{2\widehat{\sigma}^4_{ml}} < 0
$$

Así se concluye que la matriz hessiana en el óptimo es de hecho definida negativa.

2. El estimador máximo verosímil de γ se obtiene igualando el score a un vector nulo:

$$
\begin{aligned}
S(y, X; \widehat{\beta}_{ml}, \widehat{\sigma}^2_{ml}) &= 0_{k+1} \\
\left[\begin{array}{c} -\frac{1}{\widehat{\sigma}^2_{ml}} \left[-X^T y + X^T X \widehat{\beta}_{ml} \right] \\ -\frac{n}{\widehat{\sigma}^2_{ml}} + \frac{1}{\widehat{\sigma}^4_{ml}} \left(y - X\widehat{\beta}_{ml} \right)^T \left(y - X\widehat{\beta}_{ml} \right) \end{array} \right] &= 0_{k+1} \\
\widehat{\beta}_{ml} &= \widehat{\beta} = \left(X^T X\right)^{-1} X^T y \\
\widehat{\sigma}^2_{ml} &= \frac{\widehat{u}_{ml}^T \widehat{u}_{ml}}{n} = \frac{\widehat{u}^T \widehat{u}}{n} \\
&\Rightarrow \widehat{\gamma}_{ml} = \left[\begin{array}{c} \left(X^T X\right)^{-1} X^T y \\ \frac{\widehat{u}^T \widehat{u}}{n} \end{array} \right]
\end{aligned}
$$

Los estimadores máximo verosímiles de θ y de τ pueden obtener fácilmente a partir del principio de invarianza. Para θ :

$$
\begin{aligned}
\theta &= \left[\begin{array}{c} \beta \\ \sigma_u \end{array} \right] = g\left(\gamma\right) \\
&\Rightarrow \widehat{\theta}_{ml} = g\left(\widehat{\gamma}_{ml}\right) = \left[\begin{array}{c} \left(X^T X\right)^{-1} X^T y \\ \sqrt{\frac{\widehat{u}^T \widehat{u}}{n}} \end{array} \right]
\end{aligned}
$$

Para τ :

$$
\begin{aligned}
\tau &= \frac{\beta}{\sigma_u} = h(\gamma) \\
&\Rightarrow \ \widehat{\tau}_{ml} = h(\widehat{\gamma}_{ml}) = \sqrt{\frac{n}{\widehat{u}^T\widehat{u}}}.\left(X^TX\right)^{-1}X^Ty
\end{aligned}
$$

Para obtener el límite en distribución, recuérdese que todo estimador máximo verosímil, bajo ciertas condiciones de regularidad, converge a una normal multivariada con media cero y matriz de varianzas y covarianzas igual a la inversa de la matriz de información por unidad muestral.

Para $\widehat{\gamma}_{ml}$:

$$
\begin{aligned}
I_n(\gamma; y, X) &= I_n(\beta, \sigma_u; y, X) \\
&= E\left[-H(\beta, \sigma_u; y, X)\right] \\
&= E\begin{bmatrix} \frac{1}{\sigma_u^2}X^TX & \frac{1}{\sigma_u^4}X^Tu \\ \frac{1}{\sigma_u^4}u^TX & -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}u^Tu \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sigma_u^2}X^TX & \frac{1}{\sigma_u^4}X^TE(u) \\ \frac{1}{\sigma_u^4}E(u)^TX & -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}E(u^Tu) \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sigma_u^2}X^TX & 0_k \\ 0_k^T & -\frac{n}{2\sigma_u^4} + \frac{1}{\sigma_u^6}n\sigma_u^2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sigma_u^2}X^TX & 0_k \\ 0_k^T & \frac{n}{2\sigma_u^4} \end{bmatrix}
\end{aligned}
$$

Luego,

$$
\begin{aligned}
I_1(\gamma; y, X) &= \frac{1}{n}I_n(\beta, \sigma_u^2; y, X) \\
&= \begin{bmatrix} \frac{1}{\sigma_u^2}\left(\frac{X^TX}{n}\right) & 0_k \\ 0_k^T & \frac{1}{2\sigma_u^4} \end{bmatrix} \\
I_1(\gamma; y, X) &\rightarrow \begin{bmatrix} \frac{1}{\sigma_u^2}Q & 0_k \\ 0_k^T & \frac{1}{2\sigma_u^4} \end{bmatrix} \\
I_1^{-1}(\gamma; y, X) &\rightarrow \begin{bmatrix} \sigma_u^2Q^{-1} & 0_k \\ 0_k^T & 2\sigma_u^4 \end{bmatrix} \\
\sqrt{n}\left(\widehat{\gamma}_{ml} - \gamma\right) &\rightarrow {}^D\mathcal{N}\left(0_{k+1}, \begin{bmatrix} \sigma_u^2Q^{-1} & 0_k \\ 0_k^T & 2\sigma_u^4 \end{bmatrix}\right)
\end{aligned}
$$

Para los parámetros θ y τ la distribución puede obtenerse de forma sencilla como funciones de γ.Siendo estas funciones diferenciables, ambas distribuciones límite pueden obtenerse por el método delta.

Para θ :

$$\begin{aligned}
\theta &= g(\gamma) = \begin{bmatrix} \beta \\ (\sigma_u^2)^{\frac{1}{2}} \end{bmatrix} \\
Dg(\gamma) &= \begin{bmatrix} I_k & 0_k \\ 0_k^T & \frac{1}{2}(\sigma_u^2)^{-\frac{1}{2}} \end{bmatrix} \\
\sqrt{n}\left(\widehat{\theta}_{ml} - \theta\right) &\to \mathcal{N}\left(0_{k+1}, \begin{bmatrix} I_k & 0_k \\ 0_k^T & \frac{1}{2}(\sigma_u^2)^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \sigma_u^2 Q^{-1} & 0_k \\ 0_k^T & 2\sigma_u^4 \end{bmatrix} \begin{bmatrix} I_k & 0_k \\ 0_k^T & \frac{1}{2}(\sigma_u^2)^{-\frac{1}{2}} \end{bmatrix}^T\right) \\
\sqrt{n}\left(\widehat{\theta}_{ml} - \theta\right) &\to \mathcal{N}\left(0_{k+1}, \begin{bmatrix} \sigma_u^2 Q^{-1} & 0_k \\ 0_k^T & \frac{\sigma_u^2}{2} \end{bmatrix}\right)
\end{aligned}$$

Para τ :

$$\begin{aligned}
\tau &= h(\theta) = \frac{1}{\sigma_u}\beta \\
Dh(\theta) &= \begin{bmatrix} \frac{1}{\sigma_u} I_k & -\frac{1}{\sigma_u^2}\beta \end{bmatrix} \\
\sqrt{n}\left(\widehat{\tau}_{ml} - \tau\right) &\to \mathcal{N}\left(0_{k+1}, \begin{bmatrix} \frac{1}{\sigma_u} I_k & -\frac{1}{\sigma_u^2}\beta \end{bmatrix} \begin{bmatrix} \sigma_u^2 Q^{-1} & 0_k \\ 0_k^T & \frac{\sigma_u^2}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_u} I_k & -\frac{1}{\sigma_u^2}\beta \end{bmatrix}^T\right) \\
\sqrt{n}\left(\widehat{\tau}_{ml} - \tau\right) &\to \mathcal{N}\left(0_{k+1}, Q^{-1} + \frac{\beta\beta^T}{2\sigma_u^2}\right)
\end{aligned}$$

3. Este resultado se deriva del hecho de que la matriz de varianzas y covarianzas asintótica es diagonal. Por ello se dice que los estimadores de β y σ_u^2 son asintóticamente incorrelacionados. Pero además, bajo normalidad, la incorrelación equivale a independencia.

   El estimador de β sin conocer el verdadero valor de σ_u^2 ya fue encontrado antes, con varianza asintótica:

$$Avar\left(\widehat{\beta}_{ml}\right) = \sigma_u^2\left(X^T X\right)^{-1}$$

   Si se conociera σ_u^2 el problema sería maximizar la log-likelihood sólo respecto de β :

$$l(\beta, \sigma_u^2; y, X) = -n\ln\left(\sqrt{2\pi}\right) - n\ln\left(\sigma_u\right) - \frac{1}{2\sigma_u^2}\left(y - X\beta\right)^T\left(y - X\beta\right)$$

   El score respecto de β es:

$$\begin{aligned}
S_\beta(y, X; \beta, \sigma_u^2) &= \nabla_\beta l(\beta, \sigma_u^2; y, X) \\
&= -\frac{1}{\sigma_u^2}\left[-X^T y + X^T X\beta\right]
\end{aligned}$$

con lo que el estimador es el mismo que sin conocer a la varianza del error. La matriz de información es:

$$I_n(\beta;\sigma_u,y,X)=\frac{1}{\sigma_u^2}X^TX$$

La distribución asintótica de este estimador, $\widehat{\beta}_{ml}^{\sigma}$, es:

$$\sqrt{n}\left(\widehat{\beta}_{ml}^{\sigma}-\beta\right)\rightarrow\mathcal{N}\left(0_k,\sigma_u^2Q^{-1}\right)$$

Claramente,

$$Avar\left(\widehat{\beta}_{ml}^{\sigma}\right)=\sigma_u^2\left(X^TX\right)^{-1}=Avar\left(\widehat{\beta}_{ml}\right)$$

por lo que no hay ganancia en eficiencia.

4. Sea la verosimilitud condicional con regresores estocásticos:

$$L\left(\beta,\sigma_u^2;y/X\right)=\left(\frac{1}{\sqrt{2\pi}\sigma_u}\right)^n\exp\left\{-\frac{1}{2\sigma_u^2}\left(y-X\beta\right)^T\left(y-X\beta\right)\right\}$$

Claramente es igual a al verosimilitud incondicional con regresores no estocáticos, sólo cambia la interpretación. Luego,

$$\left(\widehat{\beta}_{ml},\widehat{\sigma}_{ml}^2\right)\in\arg\max_{(\beta,\sigma_u^2)}L\left(\beta,\sigma_u^2;y/X\right)$$

5. Considerando a los regresores como estocásticos, la función de verosimilitud es

$$\begin{aligned}L\left(\beta,\sigma_u^2,\phi;y,X\right)&=f_Y\left(y/X;\beta,\sigma_u^2\right).f(X;\beta,\sigma_u^2,\phi)\\L\left(\beta,\sigma_u^2,\phi;y,X\right)&\equiv L\left(\theta,\phi;y,X\right)\end{aligned}$$

donde ϕ es el vector de parámetros que participa en la identificación de la distribución de los regresores y $\theta=(\beta,\sigma_u^2)$. Sea un estadístico $S(y,X)$ para ambos vectores de parámetros:

$$S(y,X)=\begin{bmatrix}S_\theta(y,X)\\S_\phi(y,X)\end{bmatrix}$$

El estadístico $S(y,X)$ es ancilliar para θ si:

a) $f_{S_\phi}(s_\phi;\theta,\phi)=f_{S_\phi}^*(s_\phi;\phi)$

b) $f_{S_\theta/S_\phi}(s_\theta;\theta,\phi/S_\phi)=f_{S_\theta/S_\phi}^*(s_\theta;\theta/S_\phi)$

Sea el estadístico

$$S(y,X)=\begin{bmatrix}S_\theta(y,X)\\S_\phi(y,X)\end{bmatrix}=\begin{bmatrix}\widehat{\beta}_{ml}\\\widehat{\sigma}_{ml}\\vec\left(X\right)\end{bmatrix}$$

Supóngase que la densidad de la muestra de los regresores no depende de θ. En este caso,

$$\begin{aligned} f_{S_\phi}(s_\phi;\theta,\phi) &= f(X;\theta,\phi) \\ &= f^*(X;\phi) \end{aligned}$$

con lo que se cumple la primera condición. Para ver la segunda condición,

$$\begin{aligned} f_{S_\theta/S_\phi}(s_\theta;\theta,\phi/S_\phi) &= f_{\widehat{\theta}_{ml}/X}(\widehat{\theta}_{ml};\theta,\phi/X) \\ &= f_{\widehat{\theta}_{ml}/X}(\widehat{\theta}_{ml};\theta/X) \end{aligned}$$

bajo el supuesto de normalidad condicional, que sólo involucra a los parámetros en θ. Así se establece lo deseado.

El principio de Condicionalidad de la inferencia clásica sugiere entonces que la inferencia sobre θ debe basarse en el valor obtenido en la muestra para el estadístico ancilliar como un valor fijo. En este caso esto quiere decir considerar al valor de los regresores obtenidos en la muestra como fijos, ignorando su aleatoriedad. Bajo este principio, si la densidad de la muestra de los regresores no depende de θ, la inferencia en el modelo clásico debe ignorar la aleatoriedad de los regresores.

El estimador ml cumple con este principio, ya que el estimador que surge de reconocer la aleatoriedad de los regresores (es decir, el basado en la función de verosimilitud) es equivalente al que surge de la verosimilitud condicional:

$$\begin{aligned} L\left(\beta,\sigma_u^2,\phi;y,X\right) &= L\left(\beta,\sigma_u^2;y/X\right).f(X;\beta,\sigma_u^2,\phi) \\ &= L\left(\beta,\sigma_u^2;y/X\right).f^*(X;\phi) \\ L\left(\beta,\sigma_u,\phi;y,X\right) &\propto L\left(\beta,\sigma_u;y/X\right) \end{aligned}$$

Esta relación de proporcionalidad que se deriva bajo la hipótesis de que la distribución de los regresores no depende de θ garantiza que el estimador máximo verosímil marginal y condicional son iguales.

2. Sea el lagrangiano del problema:

$$\mathcal{L} = l(\beta,\sigma_u^2;y,X) - \lambda^T\left(R\beta - r\right)$$

Las condiciones de primer orden son:

$$\begin{aligned}
\nabla_{(\beta,\sigma_u^2)}\mathcal{L} &= S(y,X;\beta,\sigma_u^2) - \begin{bmatrix} R^T\lambda \\ 0 \end{bmatrix} \\
\nabla_{\lambda}\mathcal{L} &= R\beta - r \\
&\Rightarrow \begin{bmatrix} S(y,X;\beta^*,\sigma_u^{2*}) \\ R\beta^* \end{bmatrix} = \begin{bmatrix} R^T\lambda^* \\ 0 \\ r \end{bmatrix} \\
-\frac{1}{\sigma_u^{2*}}\left[-X^Ty + X^TX\beta^*\right] &= R^T\lambda^* \\
R(X^TX)^{-1}\left[-\frac{1}{\sigma_u^{2*}}\left[-X^Ty + X^TX\beta^*\right]\right] &= R(X^TX)^{-1}R^T\lambda^* \\
\frac{1}{\sigma_u^{2*}}R(X^TX)^{-1}X^Ty - \frac{1}{\sigma_u^{2*}}R(X^TX)^{-1}X^TX\beta^* &= R(X^TX)^{-1}R^T\lambda^* \\
\frac{1}{\sigma_u^{2*}}R\widehat{\beta}_{ml} - \frac{1}{\sigma_u^{2*}}R\beta^* &= R(X^TX)^{-1}R^T\lambda^* \\
\frac{1}{\sigma_u^{2*}}R\left(\widehat{\beta}_{ml} - \beta^*\right) &= R(X^TX)^{-1}R^T\lambda^* \\
\frac{1}{\sigma_u^{2*}}\left(R\widehat{\beta}_{ml} - r\right) &= R(X^TX)^{-1}R^T\lambda^* \\
&\Rightarrow \lambda^* = \frac{1}{\sigma_u^{2*}}\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta}_{ml} - r\right)
\end{aligned}$$

Asi,

$$\begin{aligned}
-\frac{1}{\sigma_u^{2*}}\left[-X^Ty + X^TX\beta^*\right] &= R^T\frac{1}{\sigma_u^{2*}}\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta}_{ml} - r\right) \\
X^Ty - X^TX\beta^* &= R^T\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta}_{ml} - r\right) \\
X^TX\beta^* &= X^Ty - R^T\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta}_{ml} - r\right) \\
\beta^* &= \widehat{\beta}_{ml} - (X^TX)^{-1}R^T\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta}_{ml} - r\right) \\
\widehat{\beta}_{mlR} &= \widehat{\beta}_{ml} - (X^TX)^{-1}R^T\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta}_{ml} - r\right) = \widehat{\beta}_R
\end{aligned}$$

3. El estadístico de Wald para hipótesis lineales sobre β es:

$$
\begin{aligned}
W &= \left(R\widehat{\beta}_{ml} - r\right)^T \left[R.I_n^{-1}(\widehat{\beta}_{ml}, \widehat{\sigma}_{ml}; y, X).R^T\right]^{-1} \left(R\widehat{\beta}_{ml} - r\right) \\
&= \left(R\widehat{\beta}_{ml} - r\right)^T \left[\widehat{\sigma}^2_{ml} R\left(X^T X\right)^{-1} R^T\right]^{-1} \left(R\widehat{\beta}_{ml} - r\right) \\
&= \frac{1}{\widehat{\sigma}^2_{ml}} \left(R\widehat{\beta}_{ml} - r\right)^T \left[R\left(X^T X\right)^{-1} R^T\right]^{-1} \left(R\widehat{\beta}_{ml} - r\right) \\
&= \frac{1}{\widehat{\sigma}^2_{ml}} \left(R\widehat{\beta} - r\right)^T \left[R\left(X^T X\right)^{-1} R^T\right]^{-1} \left(R\widehat{\beta} - r\right) \\
&= \frac{1}{\widehat{\sigma}^2_{ml}} \left[\widehat{u}_R^T \widehat{u}_R - \widehat{u}^T \widehat{u}\right] \\
&= \frac{1}{\widehat{\sigma}^2_{ml}} \left[\widehat{u}_{mlR}^T \widehat{u}_{mlR} - \widehat{u}_{ml}^T \widehat{u}_{ml}\right] \\
&= n\frac{\left(\widehat{\sigma}^2_{mlR} - \widehat{\sigma}^2_{ml}\right)}{\widehat{\sigma}^2_{ml}}
\end{aligned}
$$

El estadístico de LM es:

$$
\begin{aligned}
LM &= S(y, X; \widehat{\beta}_{mlR}, \widehat{\sigma}^2_{mlR})^T I_n(\widehat{\beta}_{mlR}, \widehat{\sigma}^2_{mlR}; y, X)^{-1} S(y, X; \widehat{\beta}_{mlR}, \widehat{\sigma}^2_{mlR}) \\
&= \lambda^{*T} R.I_n^{-1}(\widehat{\beta}_{mlR}, \widehat{\sigma}_{mlR}; y, X).R^T \lambda^* \\
&= \lambda^{*T} R.\left[\widehat{\sigma}^2_{mlR}(X^T X)^{-1}\right] R^T \lambda^* \\
&= \left[\frac{1}{\widehat{\sigma}^2_{mlR}} \left[R(X^T X)^{-1} R^T\right]^{-1} \left(R\widehat{\beta}_{ml} - r\right)\right]^T R.\left[\widehat{\sigma}^2_{mlR}(X^T X)^{-1}\right] R^T \\
&\quad \left[\frac{1}{\widehat{\sigma}^2_{mlR}} \left[R(X^T X)^{-1} R^T\right]^{-1} \left(R\widehat{\beta}_{ml} - r\right)\right] \\
&= \frac{1}{\widehat{\sigma}^2_{mlR}} \left(R\widehat{\beta}_{ml} - r\right)^T \left[R(X^T X)^{-1} R^T\right]^{-1} R.\left[(X^T X)^{-1}\right] R^T \\
&\quad \left[R(X^T X)^{-1} R^T\right]^{-1} \left(R\widehat{\beta}_{ml} - r\right) \\
&= \frac{1}{\widehat{\sigma}^2_{mlR}} \left(R\widehat{\beta}_{ml} - r\right)^T \left[R(X^T X)^{-1} R^T\right]^{-1} \left(R\widehat{\beta}_{ml} - r\right) \\
&= n\frac{\left(\widehat{\sigma}^2_{mlR} - \widehat{\sigma}^2_{ml}\right)}{\widehat{\sigma}^2_{mlR}}
\end{aligned}
$$

Finalmente, el estadístico LR es:

$$\begin{aligned}
LR &= -2\ln\left(\frac{L(\widehat{\beta}_{ml},\widehat{\sigma}^2_{ml};y,X)}{L(\widehat{\beta}_{mlR},\widehat{\sigma}^2_{mlR};y,X)}\right)\\
&= -2\ln\left(\frac{\left(\frac{1}{\sqrt{\widehat{\sigma}^2_{mlR}2\pi}}\right)^n e^{-\frac{1}{2\widehat{\sigma}^2_{mlR}}\widehat{u}'_R\widehat{u}_R}}{\left(\frac{1}{\sqrt{\widehat{\sigma}^2_{ml}2\pi}}\right)^n e^{-\frac{1}{2\widehat{\sigma}^2_{ml}}\widehat{u}'\widehat{u}}}\right)\\
&= -2\ln\left(\left(\frac{\widehat{\sigma}^2_{ml}}{\widehat{\sigma}^2_{mlR}}\right)^{\frac{n}{2}}\frac{e^{-\frac{1}{2\widehat{\sigma}^2_{mlR}}n\widehat{\sigma}^2_{mlR}}}{e^{-\frac{1}{2\widehat{\sigma}^2_{ml}}n\widehat{\sigma}^2_{ml}}}\right)\\
&= -2\left[\frac{n}{2}\ln\left(\widehat{\sigma}^2_{ml}\right)-\frac{n}{2}\ln\left(\widehat{\sigma}^2_{mlR}\right)\right]\\
&= n\left[\ln\left(\widehat{\sigma}^2_{mlR}\right)-\ln\left(\widehat{\sigma}^2_{ml}\right)\right]\\
&= n\ln\left(\frac{\widehat{\sigma}^2_{mlR}}{\widehat{\sigma}^2_{ml}}\right)
\end{aligned}$$

Nótese que:

$$\begin{aligned}
LR &= n\ln\left(1+\frac{W}{n}\right)\\
LM &= \frac{W}{1+\frac{W}{n}}\\
z &> 0 \Rightarrow z > \ln(1+z) > \frac{z}{1+z}\\
z &= \frac{W}{n} \Rightarrow \frac{W}{n} > \ln\left(1+\frac{W}{n}\right) > \frac{\frac{W}{n}}{1+\frac{W}{n}}\\
\frac{W}{n} &> \ln\left(1+\frac{W}{n}\right) > \frac{1}{n}\frac{W}{1+\frac{W}{n}}\\
\frac{W}{n} &> \frac{LR}{n} > \frac{LM}{n}\\
&\Rightarrow W > LR > LM
\end{aligned}$$

4. La reparametrización lineal puede representarse como una transformación lineal no singular:

$$\begin{aligned}
F\beta &= \delta\\
\beta &= F^{-1}\delta
\end{aligned}$$

La hipótesis es:

$$\begin{aligned}
R\beta &= r\\
RF^{-1}\delta &= r\\
R^*\delta &= r
\end{aligned}$$

La distribución asintótica del estimador de máxima verosimilitud de δ es:

$$\sqrt{n}\left(\widehat{\delta}_{ml}-\delta\right)\rightarrow\mathcal{N}\left(0_{k+1},F.I_1^{-1}(\beta;\sigma_u^2,y,X).F^T\right)$$

El estadístico de Wald es:

$$\begin{aligned}
W &= \left(R^*\widehat{\delta}_{ml}-r\right)^T\left[R^*.I_n^{-1}(\widehat{\delta}_{ml};y,X).R^{*T}\right]^{-1}\left(R^*\widehat{\delta}_{ml}-r\right)\\
&= \left(RF^{-1}\widehat{\delta}_{ml}-r\right)^T\left[RF^{-1}.I_n^{-1}(\widehat{\delta}_{ml};y,X).\left(RF^{-1}\right)^T\right]^{-1}\left(RF^{-1}\widehat{\delta}_{ml}-r\right)\\
&= \left(RF^{-1}\widehat{\delta}_{ml}-r\right)^T\left[RF^{-1}.I_n^{-1}(\widehat{\delta}_{ml};y,X).\left(F^{-1}\right)^T R^T\right]^{-1}\left(RF^{-1}\widehat{\delta}_{ml}-r\right)\\
&= \left(RF^{-1}\widehat{\delta}_{ml}-r\right)^T\left[R.I_n^{-1}(\widehat{\beta}_{ml};\sigma_{ml}^2,y,X).R^T\right]^{-1}\left(RF^{-1}\widehat{\delta}_{ml}-r\right)
\end{aligned}$$

Por el principio de invarianza,

$$\begin{aligned}
\widehat{\delta}_{ml} &= F\widehat{\beta}_{ml}\\
F^{-1}\widehat{\delta}_{ml} &= \widehat{\beta}_{ml}
\end{aligned}$$

Luego,

$$W=\left(R\widehat{\beta}_{ml}-r\right)^T\left[R.I_n^{-1}(\widehat{\beta}_{ml};\widehat{\sigma}_{ml}^2,y,X).R^T\right]^{-1}\left(R\widehat{\beta}_{ml}-r\right)$$

que es el estadístico de Wald para la hipótesis original.