UNIVERSIDAD TORCUATO DI TELLA
MAESTRÍA EN ECONOMÍA | MAESTRÍA EN ECONOMETRÍA
2021

**Econometría**
**Problem Set 0**

# 1 Soluciones

1. *Si $A$ es una matriz simétrica y definida positiva, mostrar que puede hallarse una matriz $P$ no singular tal que se verifica que $A = PP'$.*

   $A$ es simétrica $\Longrightarrow A = A'$

   $A$ es definida positiva $\Longrightarrow v'Av > 0$ para todo $v \neq 0$. Todos los autovalores de $A$ son positivos.

   Luego:

$$
\begin{aligned}
A &= C\Lambda C' \\
&= C\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}C'
\end{aligned}
$$

   Donde $\Lambda^{\frac{1}{2}} = \text{diag}(\sqrt{\lambda_1}, ..., \sqrt{\lambda_K})$

Definamos:

$$P = C\Lambda^{\frac{1}{2}}$$

Esta matriz es no singular:

$$
\begin{aligned}
P^{-1} &= \left(C\Lambda^{\frac{1}{2}}\right)^{-1} \\
P^{-1} &= \left(\Lambda^{\frac{1}{2}}\right)^{-1} C^{-1} \\
P^{-1} &= \text{diag}\left(\frac{1}{\sqrt{\lambda_1}}, ..., \frac{1}{\sqrt{\lambda_K}}\right) C^{-1} \\
P^{-1} &= \text{diag}\left(\frac{1}{\sqrt{\lambda_1}}, ..., \frac{1}{\sqrt{\lambda_K}}\right) C'
\end{aligned}
$$

Como los autovalores de $A$ son positivos, sabemos que $P^{-1}$ existe.
Finalmente notar que:

$$
\begin{aligned}
PP' &= C\Lambda^{\frac{1}{2}}\left(C\Lambda^{\frac{1}{2}}\right)' \\
PP' &= C\Lambda^{\frac{1}{2}}\left(\Lambda^{\frac{1}{2}}\right)' C' \\
PP' &= C\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}C' \\
PP' &= C\Lambda C' \\
PP' &= A
\end{aligned}
$$

2. *Sea $x$ un vector aleatorio de $n \times 1$ tal que $x \sim N(\mu, \Sigma)$, donde $\Sigma$ es una matriz simétrica y definida positiva. Mostrar que*

$$(x-\mu)'\,\Sigma^{-1}\,(x-\mu) \sim \chi^2\,(n)$$

   Primero, veamos algunos resultados que serán útiles.

   Al ser $\Sigma$ una matriz definida positiva, como probamos en el punto 1., existe una matriz $P$ no singular tal que

$$\Sigma = PP' \Longrightarrow \Sigma^{-1} = (P')^{-1}P^{-1}$$

Notar que:

$$\begin{aligned}
\det(\Sigma) &= \det(PP') \\
&= \det(P)\det(P') \\
&= \det(P)\det(P) \\
&= \det(P)^2 \\
\sqrt{\det(\Sigma)} &= |\det(P)|
\end{aligned}$$

Vale que:

$$P'\Sigma^{-1}P = P'(P')^{-1}P^{-1}P = I$$

Y que:

$$(P'\Sigma^{-1}P)^{-1} = P^{-1}\Sigma(P^{-1})' = I^{-1} = I$$

Definamos un vector aleatorio $y$ de $nx1$ tal que $y = P^{-1}\left(x-\mu\right)$. Notar que:

$$E\left[y\right] = 0$$

$$\begin{aligned}
V\left[y\right] &= E[yy'] \\
&= E\left[P^{-1}\left(x-\mu\right)\left(x-\mu\right)'\left(P^{-1}\right)'\right] \\
&= P^{-1}E\left[\left(x-\mu\right)\left(x-\mu\right)'\right]\left(P^{-1}\right)' \\
&= P^{-1}\Sigma\left(P^{-1}\right)' \\
&= I
\end{aligned}$$

$y$ es una transformación afín de un vector de variables normales, pero estas no son necesariamente independientes. Sin embargo, su distribución también será normal multivariada, como demostramos a continuación:
Siendo $y = g(x) \Rightarrow g^{-1}(y) = x$, entonces (cambio de variable):

$$f_Y(y) = f_X(g^{-1}(y))|\det(\nabla g^{-1}(y))|$$

Ahora consideremos:

$$\begin{aligned}
y &= P^{-1}\left(x-\mu\right) = P^{-1}x - P^{-1}\mu \equiv g(x) \\
x &= Py + \mu \equiv g^{-1}(y) \\
\nabla g^{-1}(y) &= P
\end{aligned}$$

Y tengamos en cuenta que:

$$f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det(\Sigma))^{\frac{1}{2}}} e^{\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right]}$$

Entonces:

$$\begin{aligned}
f_Y(y) &= f_X(g^{-1}(y))|\det(\nabla g^{-1}(y))| \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}(\det(\Sigma))^{\frac{1}{2}}} e^{\left[-\frac{1}{2}(g^{-1}(y)-\mu)'\Sigma^{-1}(g^{-1}(y)-\mu)\right]}\sqrt{\det(\Sigma)} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{\left[-\frac{1}{2}(Py+\mu-\mu)'\Sigma^{-1}(Py+\mu-\mu)\right]} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{\left[-\frac{1}{2}(Py)'\Sigma^{-1}(Py)\right]} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{\left[-\frac{1}{2}y'P'\Sigma^{-1}Py\right]} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{\left[-\frac{1}{2}y'y\right]}
\end{aligned}$$

En conclusión, $y \sim N\left(0_n, I_n\right)$.

La suma de variables aleatorias normales estándar independientes elevadas al cuadrado tiene una distribución $\chi^2_n$. Luego $y'y = y_1^2 + ... + y_n^2 \sim \chi^2_n$. Entonces:

$$\begin{aligned}
y'y &\sim \chi^2_n \\
\left[P^{-1}\left(x-\mu\right)\right]'\left[P^{-1}\left(x-\mu\right)\right] &\sim \chi^2_n \\
\left(x-\mu\right)'\left(P'\right)^{-1}P^{-1}\left(x-\mu\right) &\sim \chi^2_n \\
\left(x-\mu\right)'\left(PP'\right)^{-1}\left(x-\mu\right) &\sim \chi^2_n \\
\left(x-\mu\right)'\Sigma^{-1}\left(x-\mu\right) &\sim \chi^2_n
\end{aligned}$$

3. *Sea $x$ un vector aleatorio de $n \times 1$ $x \sim N(0_n, I_n)$ y $A$ y $B$ son matrices de $n \times n$ simétricas e idempotentes. Mostrar que las formas cuadráticas $x'Ax$ y $x'Bx$ son independientes si y sólo si $AB = 0$.*

   Dado que $A$ y $B$ son matrices simétricas e idempotentes:

   $$\begin{aligned}
   x'Ax &= x'A'Ax = \left(Ax\right)'\left(Ax\right) = g(Ax) \\
   x'Bx &= x'B'Bx = \left(Bx\right)'\left(Bx\right) = g(Bx)
   \end{aligned}$$

   Luego, si $Ax$ es idependiente de $Bx$ entonces $x'Ax$ es independiente de $x'Bx$.

   Como $Ax$ y $Bx$ son transformaciones lineales de un vector normal estándar:

   $$\begin{aligned}
   Ax &\sim N(0_n, AA') = N(0_n, A) \\
   Bx &\sim N(0_n, BB') = N(0_n, B)
   \end{aligned}$$

   Como se trata de vectores normales, para chequear que sean independientes basta con cjequear que la correlación es 0 :

   $$\begin{aligned}
   Cov(Ax, Bx) &= E\left[\left(Ax\right)\left(Bx\right)'\right] \\
   &= E\left[Axx'B\right] \\
   &= AE\left[xx'\right]B \\
   &= AI_nB \\
   &= AB
   \end{aligned}$$

   Entonces, si $AB = 0 \Longrightarrow xB$ y $xA$ son independientes $\Rightarrow x'Ax$ y $x'Bx$ son independientes.

4. *Sean $x$ e $y$ variables aleatorias con función de densidad conjunta*

   $$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{C}{2(1-\rho^2)}\right], \qquad x, y \in \mathbb{R}$$

   *donde*

   $$C = \left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right),$$

   $$\rho = \frac{\operatorname{cov}(X,Y)}{\sigma_x\sigma_y}, \qquad \sigma_x^2 = \operatorname{var}(X), \qquad \sigma_y^2 = \operatorname{var}(Y).$$

   *Mostrar que las variables aleatorias $u$ y $v$,*

   $$u = \frac{x-\mu_x}{\sigma_x} + \frac{y-\mu_y}{\sigma_y},$$

   $$v = \frac{x-\mu_x}{\sigma_x} - \frac{y-\mu_y}{\sigma_y},$$

   *son independientes y tienen distribución normal con media cero y varianza $2(1+\rho)$ y $2(1-\rho)$, respectivamente.*

Trabajemos con C:

$C = \left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)$

$C = \frac{1}{\sigma_x^2}\left(x-\mu_x\right)^2 + \frac{1}{\sigma_y^2}\left(y-\mu_y\right)^2 - 2\frac{\rho}{\sigma_x\sigma_y}\left(x-\mu_x\right)\left(y-\mu_y\right)$

$C = [x-\mu_x \quad y-\mu_y]\begin{bmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix}\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}$

Ahora definamos:

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

Entonces:

$\Sigma^{-1} = \frac{1}{\det(\Sigma)} adj(\Sigma)$

$\Sigma^{-1} = \frac{1}{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2}\begin{bmatrix} \sigma_y^2 & -\sigma_{xy} \\ -\sigma_{xy} & \sigma_x^2 \end{bmatrix}$

$\Sigma^{-1} = \frac{1}{\sigma_x^2\sigma_y^2 - \rho^2\sigma_x^2\sigma_y^2}\begin{bmatrix} \sigma_y^2 & -\sigma_{xy} \\ -\sigma_{xy} & \sigma_x^2 \end{bmatrix}$

$\Sigma^{-1} = \frac{1}{(1-\rho^2)\sigma_x^2\sigma_y^2}\begin{bmatrix} \sigma_y^2 & -\sigma_{xy} \\ -\sigma_{xy} & \sigma_x^2 \end{bmatrix}$

$\Sigma^{-1} = \frac{1}{(1-\rho^2)}\begin{bmatrix} \frac{\sigma_y^2}{\sigma_x^2\sigma_y^2} & \frac{-\sigma_{xy}}{\sigma_x^2\sigma_y^2} \\ \frac{-\sigma_{xy}}{\sigma_x^2\sigma_y^2} & \frac{\sigma_x^2}{\sigma_x^2\sigma_y^2} \end{bmatrix}$

$\Sigma^{-1} = \frac{1}{(1-\rho^2)}\begin{bmatrix} \frac{1}{\sigma_x^2} & \frac{-\rho}{\sigma_x\sigma_y} \\ \frac{-\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix}$

$(1-\rho^2)\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix}$

Entonces, relancionándolo con el resultado anterior:

$C = [x-\mu_x \quad y-\mu_y]\,(1-\rho^2)\Sigma^{-1}\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}$

$\frac{C}{(1-\rho^2)} = [x-\mu_x \quad y-\mu_y]\,\Sigma^{-1}\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}$

Definamos el vector aleatorio: $z = \begin{bmatrix} x \\ y \end{bmatrix}$ y $\mu_z = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$, de modo que:

$\frac{C}{(1-\rho^2)} = (z-\mu_z)'\Sigma^{-1}(z-\mu_z)$

Luego:

$f_Z(z) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{C}{2(1-\rho^2)}\right]$

$f_Z(z) = \left(\frac{1}{2\pi}\right)^{\frac{2}{2}}\frac{1}{\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}(z-\mu_z)'\Sigma^{-1}(z-\mu_z)\right]$

Recordar: $\det(\Sigma) = (1-\rho^2)\sigma_x^2\sigma_y^2 \Rightarrow \sqrt{\det(\Sigma)} = \sigma_x\sigma_y\sqrt{1-\rho^2}$. De modo que:

$f_Z(z) = \frac{1}{2\pi}\frac{1}{\sqrt{\det(\Sigma)}}\exp\left[-\frac{1}{2}(z-\mu_z)'\Sigma^{-1}(z-\mu_z)\right]$

Entonces: $z \sim N(\mu_z, \Sigma)$

Ahora consideremos el siguiente vector aleatorio:

$w = \begin{bmatrix} u \\ v \end{bmatrix}$

$w = \begin{bmatrix} \frac{x-\mu_x}{\sigma_x} + \frac{y-\mu_y}{\sigma_y} \\ \frac{x-\mu_x}{\sigma_x} - \frac{y-\mu_y}{\sigma_y} \end{bmatrix}$

$w = \begin{bmatrix} \frac{1}{\sigma_x} & \frac{1}{\sigma_y} \\ \frac{1}{\sigma_x} & -\frac{1}{\sigma_y} \end{bmatrix}\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}$

$w = \begin{bmatrix} \frac{1}{\sigma_x} & \frac{1}{\sigma_y} \\ \frac{1}{\sigma_x} & -\frac{1}{\sigma_y} \end{bmatrix}(z-\mu_z) \equiv h(z)$

Y vale que:

$$z = \begin{bmatrix} \frac{1}{\sigma_x} & \frac{1}{\sigma_y} \\ \frac{1}{\sigma_x} & -\frac{1}{\sigma_y} \end{bmatrix}^{-1} w + \mu_z \equiv \tfrac{1}{2} \begin{bmatrix} \sigma_x & \sigma_x \\ \sigma_y & -\sigma_y \end{bmatrix} w + \mu_z = h^{-1}(w)$$

$$\nabla h^{-1}(w) = \tfrac{1}{2} \begin{bmatrix} \sigma_x & \sigma_x \\ \sigma_y & -\sigma_y \end{bmatrix}$$

Nuevamente usamos que:

$$f_W(w) = f_Z(h^{-1}(w))|\det(\nabla h^{-1}(w))|$$

$$f_W(w) = \tfrac{1}{2\pi} \tfrac{1}{\sqrt{\det(\Sigma)}} \exp\left[-\tfrac{1}{2}(h^{-1}(w)-\mu_z)'\Sigma^{-1}(h^{-1}(w)-\mu_z)\right] \tfrac{\sigma_y\sigma_x}{2}$$

$$f_W(w) = \tfrac{1}{2\pi} \tfrac{1}{\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\tfrac{1}{2}\left(\tfrac{1}{2}\begin{bmatrix} \sigma_x & \sigma_x \\ \sigma_y & -\sigma_y \end{bmatrix} w\right)' \Sigma^{-1} \left(\tfrac{1}{2}\begin{bmatrix} \sigma_x & \sigma_x \\ \sigma_y & -\sigma_y \end{bmatrix} w\right)\right] \tfrac{\sigma_y\sigma_x}{2}$$

$$f_W(w) = \tfrac{1}{2\pi} \tfrac{1}{2\sqrt{1-\rho^2}} \exp\left[-\tfrac{1}{2}\left(\tfrac{1}{2}\begin{bmatrix} \sigma_x & \sigma_x \\ \sigma_y & -\sigma_y \end{bmatrix} w\right)' \Sigma^{-1} \left(\tfrac{1}{2}\begin{bmatrix} \sigma_x & \sigma_x \\ \sigma_y & -\sigma_y \end{bmatrix} w\right)\right]$$

$$f_W(w) = \tfrac{1}{2\pi} \tfrac{1}{2\sqrt{1-\rho^2}} \exp\left[-\tfrac{1}{8}w' \begin{bmatrix} \sigma_x & \sigma_x \\ \sigma_y & -\sigma_y \end{bmatrix}' \Sigma^{-1} \begin{bmatrix} \sigma_x & \sigma_x \\ \sigma_y & -\sigma_y \end{bmatrix} w\right]$$

$$f_W(w) = \tfrac{1}{2\pi} \tfrac{1}{2\sqrt{1-\rho^2}} \exp\left[-\tfrac{1}{8}w' \begin{bmatrix} \sigma_x & \sigma_x \\ \sigma_y & -\sigma_y \end{bmatrix}' \tfrac{1}{(1-\rho^2)} \begin{bmatrix} \frac{1}{\sigma_x^2} & \frac{-\rho}{\sigma_x\sigma_y} \\ \frac{-\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix} \begin{bmatrix} \sigma_x & \sigma_x \\ \sigma_y & -\sigma_y \end{bmatrix} w\right]$$

$$f_W(w) = \tfrac{1}{2\pi} \tfrac{1}{2\sqrt{1-\rho^2}} \exp\left[-\tfrac{1}{8(1-\rho^2)}w' \begin{bmatrix} \sigma_x & \sigma_y \\ \sigma_x & -\sigma_y \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_x} - \frac{\rho}{\sigma_x} & \frac{1}{\sigma_x} + \frac{\rho}{\sigma_x} \\ \frac{1}{\sigma_y} - \frac{\rho}{\sigma_y} & -\frac{1}{\sigma_y} - \frac{\rho}{\sigma_y} \end{bmatrix} w\right]$$

$$f_W(w) = \tfrac{1}{2\pi} \tfrac{1}{2\sqrt{1-\rho^2}} \exp\left[-\tfrac{1}{8(1-\rho^2)}w' \begin{bmatrix} 2(1-\rho) & 0 \\ 0 & 2(1+\rho) \end{bmatrix} w\right]$$

$$f_W(w) = \tfrac{1}{2\pi} \tfrac{1}{2\sqrt{1-\rho^2}} \exp\left[-\tfrac{1}{4}w' \begin{bmatrix} \frac{1}{1+\rho} & 0 \\ 0 & \frac{1}{1-\rho} \end{bmatrix} w\right]$$

$$f_W(w) = \tfrac{1}{2\pi} \tfrac{1}{2\sqrt{1-\rho^2}} \exp\left[-\tfrac{1}{4}[u \quad v] \begin{bmatrix} \frac{1}{1+\rho} & 0 \\ 0 & \frac{1}{1-\rho} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}\right]$$

$$f_W(w) = \tfrac{1}{2\pi} \tfrac{1}{2\sqrt{1-\rho^2}} \exp\left[-\tfrac{1}{4}\left[\tfrac{u^2}{1+\rho} + \tfrac{v^2}{1-\rho}\right]\right]$$

$$f_W(w) = \tfrac{1}{2\pi} \tfrac{1}{2\sqrt{1-\rho^2}} \exp\left[-\tfrac{1}{4}\tfrac{u^2}{1+\rho}\right] \exp\left[-\tfrac{1}{4}\tfrac{v^2}{1-\rho}\right]$$

$$f_W(w) = \tfrac{1}{\sqrt{2\pi}} \left(\tfrac{1}{4(1-\rho^2)}\right)^{\frac{1}{2}} \exp\left[-\tfrac{1}{4}\tfrac{u^2}{1+\rho}\right] \tfrac{1}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{4}\tfrac{v^2}{1-\rho}\right]$$

$$f_W(w) = \tfrac{1}{\sqrt{2\pi}} \left(\tfrac{1}{2(1-\rho)2(1+\rho)}\right)^{\frac{1}{2}} \exp\left[-\tfrac{1}{4}\tfrac{u^2}{1+\rho}\right] \tfrac{1}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{4}\tfrac{v^2}{1-\rho}\right]$$

$$f_W(w) = \tfrac{1}{\sqrt{2\pi 2(1+\rho)}} \exp\left[-\tfrac{1}{2}\tfrac{u^2}{2(1+\rho)}\right] \tfrac{1}{\sqrt{2\pi 2(1-\rho)}} \exp\left[-\tfrac{1}{2}\tfrac{v^2}{2(1-\rho)}\right]$$

La densidad conjunta de $u$ y $v$ es el producto de las densidades marginales de $u$ y $v$ :

$$\begin{aligned} u &\sim N(0, 2(1+\rho)) \\ u &\sim N(0, 2(1-\rho)) \end{aligned}$$

y ambas son independientes entre sí.

# Ejercicio 1

Sea el modelo de regresión lineal

$$y = X\beta + u$$

donde $y$ es un vector de $n \times 1$, $X$ es una matriz de $n \times k$ observaciones de variables con $ran(X) = k < n$ con probabilidad 1, $\beta$ es un vector de $k \times 1$ de parámetros desconocidos, y $u$ es un vector de perturbaciones estocásticas de $n \times 1$ con $\mathbf{E}[u/X] = 0$ y $\mathbf{E}[uu'] = \sigma^2 I_n$. El estimador OLS de $\beta$ es

$$\hat{\beta} = (X'X)^{-1} X'y$$

De aquí sale directamente la propiedad de linealidad. Para que dicho estimador exista es necesario que exista la inversa de $(X'X)$, lo cual se asegura con el supuesto de rango completo. Considerando que $y = \beta X + u$ se puede reescribir al estimador como

$$\hat{\beta} = \beta + (X'X)^{-1} X'u$$

Por lo tanto

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

por el supuesto de exogeneidad estricta. Entonces el estimador de OLS es insesgado. El Teorema de Gauss-Markov afirma que el estimador OLS es el de mínima varianza entre todos los estimadores linealmente insesgados. Con este fin debemos comparar su matriz de varianzas y covarianzas con la del resto de los estimadores linealmente insesgados. Sabemos que

$$\hat{\beta} - \beta = (X'X)^{-1} X'u$$

Entonces, sea $\Sigma_{\hat{\beta}}$ la matriz de varianzas y covarianzas del estimador OLS condicional en $X$

$$\Sigma_{\hat{\beta}} = \mathbf{E}\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)'/X\right] = \mathbf{E}\left[(X'X)^{-1} X'uu'X (X'X)^{-1}/X\right] = (X'X)^{-1} X' \underbrace{\mathbf{E}[uu'/X]}_{\sigma^2 I_n} X (X'X)^{-1} = \sigma^2 (X'X)^{-1}$$

Sea $\tilde{\beta}$ otro estimador linealmente insesgado. Por linealidad $\tilde{\beta} = \tilde{D}y$, con $\tilde{D}$ siendo una matriz de $k \times n$. Se puede reescribir como $\tilde{\beta} = \tilde{D}X\beta + \tilde{D}u$, para que sea insesgado necesitamos que

$$\mathbf{E}\left[\tilde{\beta}/X\right] = \mathbf{E}\left[\tilde{D}X\beta + \tilde{D}u/X\right] = \tilde{D}X\beta = \beta \Longleftrightarrow \tilde{D}X = I_k$$

La matriz de varianzas y covarianzas de $\tilde{\beta}$ condicional en $X$, $\Sigma_{\tilde{\beta}}$ es

$$\Sigma_{\tilde{\beta}} = \mathbf{E}\left[\left(\tilde{\beta} - \beta\right)\left(\tilde{\beta} - \beta\right)'/X\right] = \mathbf{E}\left[\tilde{D}uu'\tilde{D}'/X\right] = \sigma^2 \tilde{D}\tilde{D}'$$

Para probar que $\hat{\beta}$ es de varianza mínima basta con probar que $\Sigma_{\tilde{\beta}} - \Sigma_{\hat{\beta}}$ es una matriz semidefinida positiva. Con este fin definamos $D = \tilde{D} - (X'X)^{-1} X'$. Notar que $DX = 0$ (esto implica también que $D'X' = 0$. $\Sigma_{\tilde{\beta}}$ se puede reescribir, entonces

$$\begin{aligned}\Sigma_{\tilde{\beta}} &= \sigma^2 \left(D + (X'X)^{-1} X'\right)\left(D + (X'X)^{-1} X'\right)' \\ &= \sigma^2 \left(DD' + (X'X)^{-1} X'D' + DX (X'X)^{-1} + (X'X)^{-1}\right) = \sigma^2 DD' + \sigma^2 (X'X)^{-1}\end{aligned}$$

Como $DD'$ es una matriz semidefinida positiva entonces $\Sigma_{\tilde{\beta}} - \Sigma_{\hat{\beta}} = DD'\sigma^2$ es semidefinida positiva. Esto prueba que el estimador OLS de $\beta$ es el de mínimza varianza, y completa la prueba que son BLUE. Sea $\theta = c'\beta$, odnde $c$ es un vector de $k \times 1$ variables, se desea construir un estimador BLUE De $\theta$. Queremos probar que el estimador BLUE es $c'\hat{\beta}$, con $\hat{\beta}$ como el estiamdor OLS de $\beta$. Como es lineal considere el estiamdor $\hat{\theta} = a'y$, donde $a$ es un vector de $n \times 1$ variables. Como $\hat{\theta}$ debe ser insesgado

$$\mathbf{E}\left[\hat{\theta}/X\right] = \mathbf{E}[a'(X\beta + u)] = a'X\beta + a'\mathbf{E}[u/X] = a'X\beta = \beta \Longleftrightarrow a'X = c' \implies X'a = c$$

De aquí se desprende que

$$\hat{\theta} = a'X\beta + a'u \implies \hat{\theta} - \theta = a'u$$

La matriz de varianzas y covarianzas condicional en $X$, $\Sigma_{\hat{\theta}}$ en este caso es un escalar, y tiene la forma

$$\Sigma_{\hat{\theta}} = \mathbf{E}\left[\left(\hat{\theta} - \theta\right)\left(\hat{\theta} - \theta\right)'/X\right] = \mathbf{E}[a'uu'a/X] = \sigma^2 a'a$$

Encontrar el estimador BLUE implica minimizar la varianza condicional $\Sigma_{\hat{\theta}}$ sujeto a que sea insesgado. Como $\sigma^2$ es un escalar positivo este problema equivale a

$$\max_a a'a$$
$$s.a. X'a = c$$

El lagrangiano de este problema y las condiciones de primer orden (necesarias y suficientes) son

$$\mathcal{L} = a'a - 2\lambda' (X'a - c)$$

$$a' = \lambda' X' \implies a = X\lambda$$
$$X'a = c$$

Utilizanndo ambas ecuaciones llegamos a que $\lambda = (X'X)^{-1} c$. Esto implica que $a = X (X'X)^{-1} c$. Utilizando esto en $\hat{\theta}$ llegamos a que $\hat{\theta} = c' (X'X)^{-1} X'y = c'\hat{\beta}$, que es lo que queríamos probar.

# Ejercicio 2

Sea $\hat{\beta}$ el estimador OLS de $\beta$, la condición de primer orden del problema es la siguiente:

$$\left(\nabla_{\hat{\beta}} \left(y - X\hat{\beta}\right) \left(y - X\hat{\beta}\right)'\right)' = 0 \implies -2X'y + 2 (X'X) \hat{\beta} = 0$$

La condición de segundo orden del problema implica que la matriz Hessiana sea semidefinida positiva. En este caso

$$H\left(\left(y - X\hat{\beta}\right) \left(y - X\hat{\beta}\right)'\right) = D_{\hat{\beta}} \left(\nabla_{\hat{\beta}} \left(y - \hat{\beta}X\right) \left(y - \hat{\beta}X\right)'\right) = 2 (X'X)$$

Como $X$ es una matriz de rango completo entonces $(X'X)$ es una matriz semidefinida positiva, lo cual verifica la condición de segundo orden del problema.

# Ejercicio 3

Sea el modelo $y = X\beta + u$. El estimador OLS de $\beta$ es $\hat{\beta}$. Consideremos primero un cambio de escala. Un cambio de escala puede representarse como una transformación lineal que afecta a las observaciones de la misma variable por igual. Sean $\tilde{y}$ y $\tilde{X}$ las variables transformadas esto implica que, con $\lambda, \lambda_j \neq 0$

$$\begin{cases} \tilde{y} = \lambda y \\ \tilde{x}_j = \lambda_j x_j \end{cases}$$

En notación matricial esto implica que

$$\tilde{y} = \lambda y$$
$$\tilde{X} = X \underbrace{\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_k \end{bmatrix}}_{\Lambda} = X\Lambda$$
$$\Lambda = diag (\lambda_1, \dots, \lambda_k) \; det(\Lambda) \neq 0 \; \Lambda' = \Lambda$$

El nuevo modelo es, entonces

$$\lambda y = X\Lambda\beta^* + w$$

El estimador OLS de $\beta^*$ en el nuevo modelo será

$$\hat{\beta}^* = \left(\tilde{X}'\tilde{X}\right)^{-1} \tilde{X}'\tilde{y} = \left((X\Lambda)' (X\Lambda)\right)^{-1} (X\Lambda)' \lambda y = \lambda (\Lambda' X' X \Lambda)^{-1} \Lambda' X' y = \lambda\Lambda^{-1} (X'X)^{-1} \Lambda^{-1}\Lambda X' y = \lambda\Lambda^{-1} (X'X)^{-1} X'y$$

$$\hat{\beta}^* = \lambda\Lambda^{-1}\hat{\beta}$$

$$\begin{bmatrix} \hat{\beta}_1^* \\ \hat{\beta}_2^* \\ \vdots \\ \vdots \\ \hat{\beta}_k^* \end{bmatrix} = \begin{bmatrix} \frac{\lambda}{\lambda_1}\hat{\beta}_1 \\ \frac{\lambda}{\lambda_2}\hat{\beta}_2 \\ \vdots \\ \frac{\lambda}{\lambda_k}\hat{\beta}_k \end{bmatrix}$$

Se puede probar que el ajuste del modelo es el mismo

$$\hat{\tilde{y}} = \tilde{X}\hat{\beta}^* \Longleftrightarrow \hat{\tilde{y}} = X\Lambda\lambda\Lambda^{-1}\hat{\beta} = \lambda\hat{y} \Longleftrightarrow \tilde{y} = \lambda\hat{y} + \hat{w} \Longleftrightarrow \lambda\left(y - \hat{y}\right) = \hat{w} \implies \lambda\hat{u} = \lambda\hat{w}$$

Entonces para calcular el ajuste del modelo

$$\tilde{R}^2 = 1 - \frac{\hat{w}'\hat{w}}{\tilde{y}'\tilde{y}} = 1 - \frac{\lambda^2\hat{u}'\hat{u}}{\lambda^2 y'y} = 1 - \frac{\hat{u}'\hat{u}}{y'y} = R^2$$

Consideremos ahora una traslación en los regresores. Sea $d$ un vector columna de $k$ constantes, la traslación consiste en sumar a todas las observaciones de un mismo regresor una misma constante. Es decir

$$\tilde{x}_{ji} = x_{ji} + d_j$$

El nuevo modelo es, particionano el vector de coeficientes e intercepto y pendientes

$$y_i = \left[1 \vdots x_i' + d'\right]\beta + v_i = \beta_1 + \left(x_i' + d'\right)\beta_{-1} + v_i$$

Donde $\beta_1$ es el intercepto y $\beta_{-1}$ es el vector de los $k-1$ coeficientes pendientes. En notación matricial, sean $0_n$ y $1_n$ vectores columna de $n \times 1$ variables y $X_{-1}$ la matrix de observaciones de los regresores de la variable

$$\tilde{X} = X + \left[0 \vdots 1_n \cdot d'\right]$$

$$y = \left\{X + \left[0 \vdots 1_n \cdot d'\right]\right\}\beta + v = \left\{\left[1_n \vdots X_1\right] + \left[0_n \vdots 1_n \cdot d'\right]\right\}\begin{bmatrix}\beta_1 \\ \beta_{-1}\end{bmatrix} + v = 1_n\beta_1 + \left[X_{-1} + 1_n d'\right]\beta_{-1} + v$$

$$= \left[1_n + 1_n \cdot d'\beta_{-1}\right] + X_{-1}\beta_{-1} + v = 1_n\left[\beta_1 + d'\beta_{-1}\right] + X_{-1}\beta_{-1} = \left[1_n \vdots X_{-1}\right]\begin{bmatrix}\beta_1 + d'\beta_{-1} \\ \beta_1\end{bmatrix} + v$$

Sea $\beta_1^* = \beta_1 + d'\beta_{-1}$, el nuevo modelo resulta

$$y = X\beta^* + v \quad \beta^* = \left[\beta_1 + d'\beta_{-1}\right]$$

Este último cambio afecta sólo al intercepto del modelo. El estimador OLS para $\beta_{-1}^*$ es el mismo que en el modelo original. Por lo hallado en el ejercicio 1A sabemos que el estimador OLS de una combinación lineal de los parámetros originales es la combinación lineal de los estimadores OLS de los parámetros. De este modo

$$\hat{\beta}^* = \begin{bmatrix}\hat{\beta}_1 + d'\hat{\beta}_{-1} \\ \hat{\beta}_{-1}\end{bmatrix}$$

# Ejercicio 4

Sea $\hat{\beta}$ el estimador OLS de $\beta$, entonces

$$\hat{\beta} \in arg\min_{\tilde{\beta}}\left\{SSR\left(\tilde{\beta}\right)\right\}$$

$$\hat{\beta} \in arg\min_{\tilde{\beta}}\left\{\left(y - X\tilde{\beta}\right)\left(y - X\tilde{\beta}\right)'\right\}$$

$$y'y - n\bar{y}^2 \geq 0 \implies \hat{\beta} \in arg\min_{\tilde{\beta}}\left\{\frac{\left(y - X\tilde{\beta}\right)\left(y - X\tilde{\beta}\right)'}{y'y - n\bar{y}^2}\right\}$$

$$\hat{\beta} \in arg\max_{\tilde{\beta}}\left\{-\frac{\left(y - X\tilde{\beta}\right)\left(y - X\tilde{\beta}\right)'}{y'y - n\bar{y}^2}\right\}$$

$$\hat{\beta} \in arg\max_{\tilde{\beta}}\left\{1 - \frac{\left(y - X\tilde{\beta}\right)\left(y - X\tilde{\beta}\right)'}{y'y - n\bar{y}^2}\right\} \therefore \hat{\beta} \in arg\max_{\tilde{\beta}}\left\{R^2\right\}$$

# Ejercicio 5

Considerando el modelo en forma matricial con $n = 33$ observaciones

$$y = X\beta + u$$

donde

$$Y_{n\times 1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \; X_{n\times 3} = \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & x_{3n} \end{bmatrix} \; \beta_{3\times 1} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \; u_{n\times 1} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Estimando por OLS llegamos a los siguientes resultados

$$\hat{\beta} = (X'X)^{-1} X'y$$
$$\Sigma_{\hat{\beta}} = \hat{\sigma}^2 (X'X)^{-1}$$
$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n-3}$$

Sabemos que

$$(X'X) = \begin{bmatrix} \sum_{i=1}^{33} x_{1i}^2 & \sum_{i=1}^{33} x_{1i}x_{2i} & \sum_{i=1}^{33} x_{1i}x_{3i} \\ \sum_{i=1}^{33} x_{1i}x_{2i} & \sum_{i=1}^{33} x_{2i}^2 & \sum_{i=1}^{33} x_{2i}x_{3i} \\ \sum_{i=1}^{33} x_{1i}x_{3i} & \sum_{i=1}^{n} x_{2i}x_{3i} & \sum_{i=1}^{33} x_{3i}^2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 10 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Luego

$$(X'X)^{-1} = \frac{1}{9} \begin{bmatrix} 10 & -1 & -10 \\ -1 & 1 & 1 \\ -10 & 1 & 19 \end{bmatrix}$$

También sabemos que

$$X'y = \begin{bmatrix} \sum_{i=1}^{n} x_{1i}y_i \\ \sum_{i=1}^{n} x_{2i}y_i \\ \sum_{i=1}^{n} x_{3i}y_i \end{bmatrix} = \begin{bmatrix} 5 \\ 10 \\ 4 \end{bmatrix}$$

Entonces podemos computar los estimadores OLS

$$\hat{\beta} = \frac{1}{9} \begin{bmatrix} 10 & -1 & -10 \\ -1 & 1 & 1 \\ -10 & 1 & 19 \end{bmatrix} \begin{bmatrix} 5 \\ 10 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}$$

Ahora para computar la varianza de los errores sabemos que

$$\hat{u} = y - X\hat{\beta} = y - X (X'X)^{-1} X'y = \underbrace{\left(I - X (X'X)^{-1} X'\right)}_{M} y$$

M es una matriz idempotente, entonces

$$\hat{u}'\hat{u} = y'MMy = y'My = y' \left(I - X (X'X)^{-1} X'\right) y = y'y - y'X (X'X)^{-1} Xy$$

Sabemos que $y'y = \sum_{i=1}^{33} y_i^2 = 35$. De pasos anteriores sabemos que

$$y'X = (X'y)' = \begin{bmatrix} \sum_{i=1}^{33} x_{1i}y_i & \sum_{i=1}^{33} x_{2i}y_i & \sum_{i=1}^{33} x_{3i}y_i \end{bmatrix} = \begin{bmatrix} 5 & 10 & 4 \end{bmatrix}$$

Entonces

$$\hat{u}'\hat{u} = 35 - \begin{bmatrix} 5 & 10 & 4 \end{bmatrix} \frac{1}{9} \begin{bmatrix} 10 & -1 & -10 \\ -1 & 1 & 1 \\ -10 & 1 & 19 \end{bmatrix} \begin{bmatrix} 5 \\ 10 \\ 4 \end{bmatrix} = 9 \therefore \hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n-3} = \frac{9}{30}$$

Finalmente

$$\Sigma_{\hat{\beta}} = \hat{\sigma}^2 \left(X'X\right)^{-1} = \frac{9}{30}\frac{1}{9} \begin{bmatrix} 10 & -1 & -10 \\ -1 & 1 & 1 \\ -10 & 1 & 19 \end{bmatrix} = \frac{1}{30} \begin{bmatrix} 10 & -1 & -10 \\ -1 & 1 & 1 \\ -10 & 1 & 19 \end{bmatrix}$$

UNIVERSIDAD TORCUATO DI TELLA
MAESTRÍA EN ECONOMÍA — MAESTRÍA EN ECONOMETRÍA
2021

---

**Econometría**
**Problem Set 2**
**Tópicos en el Modelo Lineal**

---

# 1 Un breve repaso...

La traza de una matriz cuadrada es la suma de los elementos de su diagonal principal. Algunas propiedades:

1. $tr(A+B) = tr(A) + tr(B)$
2. $tr(A') = tr(A)$
3. $tr(AB) = tr(BA)$
4. El rango de una matriz idempotente es igual a su traza.

Sea $A$ una matriz cuadrada con autovalores $\lambda_1, \lambda_2, ..., \lambda_n$. Entonces:

1. $\det A = \lambda_1 \lambda_2, ... \lambda_n$. Si $A$ es una matriz no singular entonces todos sus autovalores son no nulos.
2. $trA = \lambda_1 + \lambda_2 + ... + \lambda_n$.
3. Si $A$ es una matriz idempotente entonces sus autovalores son 0 o 1.

Algunos Teoremas (ver Apéndice del Greene) que nos serán útiles:

**Theorem 1 (B.8)** *Si $x \sim N(0, I)$ y $A$ es idempotente, entonces $x'Ax \sim \chi^2_k$ donde $k = rank(A) = \#$ autovalores iguales a 1.*

**Theorem 2 (B.12)** *Sean $x \sim N(0, I)$, $Lx$ una transformación lineal de $x$, y $x'Ax$ con $A$ idempotente y simétrica. Entonces $Lx$ y $x'Ax$ son independientes si $LA = 0$.*

Recordar del Problem Set 0 el siguiente resultado:
*Sea $x$ un vector aleatorio de $n \times 1$ tal que $x \sim N(\mu, \Sigma)$, donde $\Sigma$ es una matriz simétrica y definida positiva. Entonces:*

$$(x-\mu)' \Sigma^{-1} (x-\mu) \sim \chi^2_n$$

Recordar también:
*Sea $x$ un vector aleatorio con $E(x) = \mu$ y $Var(x) = E[(x-\mu)(x-\mu)'] = \Sigma$ ("matriz de varianzas y covarianzas"), entonces, dada una matriz $A: E[Ax] = A\mu$ y $Var[Ax] = A\Sigma A'$.*

---

# 2 Problemas

---

1. Por el supuesto de normalidad: $u|X \sim N(0, \sigma^2_u I_n)$.

   Sea $y$ :

   $$y = X\beta + u$$

   Puede verse que $y$ es una transformación afín de $u$, condicional a los regresores. Luego, $y$ tiene distribución normal. A partir del supuesto de exogeneidad estricta,

   $$E(y|X) = X\beta + E(u|X) = X\beta$$

Además, por el supuesto de errores esféricos,

$$\begin{aligned}\Sigma_y &= E\left\{[y-E(y|X)][y-E(y|X)]'/X\right\}\\ &= E\left\{[y-X\beta][y-X\beta]'/X\right\}\\ &= E\left\{uu'/X\right\}\\ &= \sigma_u^2 I_n\end{aligned}$$

Luego,

$$\boxed{y|X \sim \mathcal{N}\left(X\beta, \sigma_u^2 I_n\right)}$$

Sea $\widehat{\beta}$ :

$$\widehat{\beta} = (X'X)^{-1} X'y = \beta + (X'X)^{-1} X'u$$

El estimador de MCC se escribe como una transformación afín de $u$. Entonces es normal. Sus momentos ya han sido calculados en el práctico pasado. Luego,

$$\boxed{\widehat{\beta}/X \sim \mathcal{N}\left(\beta, \sigma_u^2 (X'X)^{-1}\right)}$$

Sea ahora $\widehat{u}$ :

$$\begin{aligned}\widehat{u} &= y - X\widehat{\beta}\\ &= y - X(X'X)^{-1} X'y\\ &= My\\ &= MX\beta + Mu\\ &= Mu\end{aligned}$$

En el último paso se usó que $MX = (I - X(X'X)^{-1} X')X = X - X = 0$. El vector de residuos de MCC es entonces una transformación lineal del vector $u$. Entonces, es normal. Sus momentos relevantes son, bajo los supuestos de exogeneidad estricta y errores esféricos:

$$\begin{aligned}E(\widehat{u}/X) &= ME(u|X) = 0_n\\ \Sigma_{\widehat{u}} &= E[\widehat{u}\widehat{u}'/X]\\ &= E[Muu'M'/X]\\ &= E[Muu'M/X]\\ &= ME[uu'/X]M\\ &= \sigma_u^2 MM\\ &= \sigma_u^2 M\end{aligned}$$

en virtud de que $M$ es simétrica e idempotente. Luego,

$$\boxed{\widehat{u}/X \sim \mathcal{N}\left(0_n, \sigma_u^2 M\right)}$$

Sea $\frac{\widehat{\sigma}_u^2(n-k)}{\sigma_u^2}$ :

$$\frac{\widehat{\sigma}_u^2(n-k)}{\sigma_u^2} = \frac{\frac{RSS}{n-k}(n-k)}{\sigma_u^2} = \frac{RSS}{\sigma_u^2}$$

$$\begin{aligned}\text{Además :}&\quad RSS = \widehat{u}'\widehat{u} = u'M'Mu = u'Mu\\ \text{Entonces} \quad :&\quad \frac{\widehat{\sigma}_u^2(n-k)}{\sigma_u^2} = \frac{u'Mu}{\sigma_u^2}\end{aligned}$$

Notar que $M$ es idempotente y que $\frac{u}{\sigma_u} \sim N(0, I)$, entonces podemos aprovechar el Teorema B.8 para afirmar que:

$$\frac{\widehat{\sigma}_u^2(n-k)}{\sigma_u^2} = \frac{u'Mu}{\sigma_u^2}\tilde{}$$

$\frac{\widehat{\sigma}_u^2(n-k)}{\sigma_u^2}$ y $\widehat{\beta}$ son transformaciones de $u$:

$$\begin{aligned}\widehat{\beta} &= \beta + (X'X)^{-1} X'u\\ \frac{\widehat{\sigma}_u^2(n-k)}{\sigma_u^2} &= \frac{\widehat{u}'\widehat{u}}{\sigma_u^2} = \frac{u'Mu}{\sigma_u^2}\tilde{}\chi_\nu^2\end{aligned}$$

Donde $\upsilon = rank(M) = tr(M)$ dado que $M$ es idempotente. La traza de $M$ es $n-k$ :

$$\begin{aligned}
tr(M) &= tr(I_n - X\left(X'X\right)^{-1}X') \\
&= tr(I_n) - tr\left[(X)((X'X)^{-1}X')\right] \\
&= tr(I_n) - tr\left[((X'X)^{-1}X')(X))\right] \\
&= tr(I_n) - tr(I_k) \\
&= n-k
\end{aligned}$$

Entonces:

$$\boxed{\frac{\widehat{\sigma}_u^2(n-k)}{\sigma_u^2} \sim \chi^2_{n-k}}$$

$\frac{\widehat{\sigma}_u^2(n-k)}{\sigma_u^2}$ y $\widehat{\beta}$ son transformaciones de $u$:

$$\begin{aligned}
\widehat{\beta} &= \beta + \left(X'X\right)^{-1}X'u \\
\frac{\widehat{\sigma}_u^2\left(n-k\right)}{\sigma_u^2} &= \frac{\widehat{u}'\widehat{u}}{\sigma_u^2} = \frac{u'Mu}{\sigma_u^2}
\end{aligned}$$

Aprovechando el Teorema 8.12, sabemos que estos dos vectores aleatorios son independientes si y sólo si:

$$\begin{aligned}
\left(X'X\right)^{-1}X'M &= O_{k\times n} \\
\left(X'X\right)^{-1}X'M' &= O_{k\times n} \\
\left(X'X\right)^{-1}\left(MX\right)' &= O_{k\times n} \\
\left(X'X\right)^{-1}O_{k\times n} &= O_{k\times n} \checkmark
\end{aligned}$$

2. Para este ejercicio se utilizarán las distribuciones derivadas en el punto 1. Sea $\widehat{\beta}$ :

$$\begin{aligned}
\widehat{\beta}/X &\sim \mathcal{N}\left(\beta, \sigma_u^2\left(X'X\right)^{-1}\right) \\
\left(\widehat{\beta}-\beta\right)'\left[\sigma_u^2\left(X'X\right)^{-1}\right]^{-1}\left(\widehat{\beta}-\beta\right) &\sim \chi^2_k \\
\frac{\left(\widehat{\beta}-\beta\right)'\left(X'X\right)\left(\widehat{\beta}-\beta\right)}{\sigma_u^2} &\sim \chi^2_k \\
\frac{\widehat{\sigma}_u^2\left(n-k\right)}{\sigma_u^2} &\sim \chi^2_{n-k} \\
&\Rightarrow \frac{\left\{\frac{\left[\frac{(\widehat{\beta}-\beta)'(X'X)(\widehat{\beta}-\beta)}{\sigma_u^2}\right]}{k}\right\}}{\left\{\frac{\left[\frac{\widehat{\sigma}_u^2(n-k)}{\sigma_u^2}\right]}{n-k}\right\}} \sim \mathcal{F}\left(k, n-k\right) \\
\frac{\left[\frac{(\widehat{\beta}-\beta)'(X'X)(\widehat{\beta}-\beta)}{k}\right]}{[\widehat{\sigma}_u^2]} &\sim \mathcal{F}\left(k, n-k\right) \\
\frac{\left(\widehat{\beta}-\beta\right)'\left(X'X\right)\left(\widehat{\beta}-\beta\right)}{k\widehat{\sigma}_u^2} &\sim \mathcal{F}\left(k, n-k\right)
\end{aligned}$$

La región de confianza es conceptualmente análoga al intervalo de confianza, para vectores de parámetros. En este caso,

$$P\left(\frac{\left(\widehat{\beta}-\beta\right)'\left(X'X\right)\left(\widehat{\beta}-\beta\right)}{k\widehat{\sigma}_u^2} \leq \mathcal{F}_{1-\alpha}\left(k, n-k\right)\right) = 1-\alpha$$

$$\boxed{C\left(\frac{(\widehat{\beta}-\beta)'(X'X)(\widehat{\beta}-\beta)}{k\widehat{\sigma}_u^2} \leq \mathcal{F}_{1-\alpha}\left(k, n-k\right)\right) = 1-\alpha}$$

La siguiente expresión:

$$\frac{\left(\widehat{\beta}-\beta\right)'(X'X)\left(\widehat{\beta}-\beta\right)}{k\widehat{\sigma}_u^2}=\mathcal{F}_{1-\alpha}(k,n-k)$$

es simplemente una curva de nivel de una forma cuadrática. Para el caso $k=2$, tiene representación gráfica como una elipse (o círculo, según los valores) en el espacio $\mathbb{R}^2$.

Sea ahora $c'\beta$, $c$ no nulo:

$$\begin{aligned}
c'\widehat{\beta}/X &\sim \mathcal{N}\left(c'\beta,\sigma_u^2.c'\left(X'X\right)^{-1}c\right)\\
\frac{c'\left(\widehat{\beta}-\beta\right)}{\sqrt{\sigma_u^2.c'\left(X'X\right)^{-1}c}} &\sim \mathcal{N}(0,1)\\
\frac{c'\left(\widehat{\beta}-\beta\right)}{\sigma_u\sqrt{.c'\left(X'X\right)^{-1}c}} &\sim \mathcal{N}(0,1)\\
\frac{\left[\frac{c'\left(\widehat{\beta}-\beta\right)}{\sigma_u\sqrt{.c'(X'X)^{-1}c}}\right]}{\sqrt{\frac{\left(\frac{\widehat{\sigma}_u^2(n-k)}{\sigma_u^2}\right)}{n-k}}} &\sim \tau_{n-k}\\
\frac{\left[\frac{c'\left(\widehat{\beta}-\beta\right)}{\sqrt{.c'(X'X)^{-1}c}}\right]}{\left[\widehat{\sigma}_u\right]} &\sim \tau_{n-k}\\
\frac{c'\left(\widehat{\beta}-\beta\right)}{\widehat{\sigma}_u\sqrt{.c'\left(X'X\right)^{-1}c}} &\sim \tau_{n-k}\\
\widehat{\sigma}_u\sqrt{.c'\left(X'X\right)^{-1}c} &= \widehat{Se(c'\widehat{\beta})}
\end{aligned}$$

El intervalo de confianza es:

$$\begin{aligned}
P\left(\left|\frac{c'\left(\widehat{\beta}-\beta\right)}{\widehat{Se(c'\widehat{\beta})}}\right|\le\tau_{1-\alpha,n-k}\right) &= 1-\alpha\\
P\left(-\tau_{1-\alpha,n-k}\le\frac{c'\left(\widehat{\beta}-\beta\right)}{\widehat{Se(c'\widehat{\beta})}}\le\tau_{1-\alpha,n-k}\right) &= 1-\alpha\\
P\left(-\tau_{1-\alpha,n-k}\widehat{Se(c'\widehat{\beta})}\le c'\left(\widehat{\beta}-\beta\right)\le\tau_{1-\alpha,n-k}\widehat{Se(c'\widehat{\beta})}\right) &= 1-\alpha\\
C\left(-c'\widehat{\beta}-\tau_{1-\alpha,n-k}\widehat{Se(c'\widehat{\beta})}\le -c'\beta\le -c'\widehat{\beta}+\tau_{1-\alpha,n-k}\widehat{Se(c'\widehat{\beta})}\right) &= 1-\alpha
\end{aligned}$$

$$\boxed{C\left(c'\widehat{\beta}-\tau_{1-\alpha,n-k}\widehat{Se(c'\widehat{\beta})}\le c'\beta\le c'\widehat{\beta}+\tau_{1-\alpha,n-k}\widehat{Se(c'\widehat{\beta})}\right)=1-\alpha}$$

3. $y_t=\beta_1x_{1t}+\beta_2x_{2t}+\beta_3x_{3t}+u_t,\qquad (t=1,\ldots,n)$

El primero es un Test de Significatividad Individual:

$t=\frac{\widehat{\beta}_2}{se(\widehat{\beta}_2)}\sim t_{n-k}$

Según los resultados obtenidos en el Problem Set 1:

$t=\frac{\widehat{\beta}_2}{se(\widehat{\beta}_2)}=\frac{1}{\sqrt{\frac{1}{30}}}=\sqrt{30}=5.4772$

El segundo es un Test de Wald:

$R=[1\quad 1\quad 0]$

$r=2$

$H_0: R\beta=2$

$W=\frac{(R\widehat{\beta}-r)'(Rs^2(X'X)^{-1}R')^{-1}(R\widehat{\beta}-r)}{\#r}\sim F_{\#r,n-k}$

Según los resultados obtenidos en el Problem Set 1:

$W=\frac{10}{3}=3.33.$

4. Algunas consideraciones previas. Considere el modelo:

$$y = X\beta + u = X_1\beta_1 + X_2\beta_2 + u$$

Donde $X_1$ y $X_2$ son las observaciones de dos sets de variables.

De la condición de primer orden de minimización de RSS obteníamos:

$$\begin{aligned}
(X'X)\widehat{\beta} &= X'y \\
\left[(X_1 \vdots X_2)'(X_1 \vdots X_2)\right]\widehat{\beta} &= (X_1 \vdots X_2)'y \\
\left[\begin{bmatrix} X_1' \\ \cdots \\ X_2' \end{bmatrix}(X_1 \vdots X_2)\right]\begin{bmatrix}\widehat{\beta}_1 \\ \widehat{\beta}_2\end{bmatrix} &= \begin{bmatrix} X_1' \\ \cdots \\ X_2' \end{bmatrix} y \\
\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}\begin{bmatrix}\widehat{\beta}_1 \\ \widehat{\beta}_2\end{bmatrix} &= \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}
\end{aligned}$$

Tenemos entonces el siguiente sistema de ecuaciones:

$X_1'X_1\widehat{\beta}_1 + X_1'X_2\widehat{\beta}_2 = X_1'y$

$X_2'X_1\widehat{\beta}_1 + X_2'X_2\widehat{\beta}_2 = X_2'y$

Consideremos la primera parte de este sistema:

$X_1'X_1\widehat{\beta}_1 = X_1'y - X_1'X_2\widehat{\beta}_2 \Rightarrow \widehat{\beta}_1 = (X_1'X_1)^{-1}X_1'y - (X_1'X_1)^{-1}X_1'X_2\widehat{\beta}_2 \Rightarrow \widehat{\beta}_1 = (X_1'X_1)^{-1}X_1'(y - X_2\widehat{\beta}_2)$

Notar que si $X_1'X_2 = 0$, entonces $\widehat{\beta}_1 = (X_1'X_1)^{-1}X_1'y$, que es el resultado de regresar $y$ sobre $X_1$. Esto es un teorema: **en un modelo de regresión lineal múltiple de $y$ sobre dos sets de variables $X_1$ y $X_2$, si los dos sets de variables son ortogonales, entonces el vector de coeficientes estimados asociados a $X_1$ se puede obtener haciendo una regresión de $y$ únicamente sobre $X_1$; idem para el vector de coeficientes estimados asociados a $X_2$.**

Ahora tomemos la segunda parte del sistema e insertemos la expresión obtenida para $\widehat{\beta}_1$ :

$X_2'X_1\widehat{\beta}_1 + X_2'X_2\widehat{\beta}_2 = X_2'y$

$X_2'X_1(X_1'X_1)^{-1}X_1'(y - X_2\widehat{\beta}_2) + X_2'X_2\widehat{\beta}_2 = X_2'y$

$X_2'X_1(X_1'X_1)^{-1}X_1'y - X_2'X_1(X_1'X_1)^{-1}X_1'X_2\widehat{\beta}_2 + X_2'X_2\widehat{\beta}_2 = X_2'y$

$X_2'(X_1(X_1'X_1)^{-1}X_1'y - y) = X_2'(X_1(X_1'X_1)^{-1}X_1'X_2 - X_2)\widehat{\beta}_2$

$X_2'(X_1(X_1'X_1)^{-1}X_1' - I)y = X_2'(X_1(X_1'X_1)^{-1}X_1' - I)X_2\widehat{\beta}_2$

$X_2'(I - X_1(X_1'X_1)^{-1}X_1')y = X_2'(I - X_1(X_1'X_1)^{-1}X_1')X_2\widehat{\beta}_2$

$\left[X_2'(I - X_1(X_1'X_1)^{-1}X_1')X_2\right]^{-1} X_2'(I - X_1(X_1'X_1)^{-1}X_1')y = \widehat{\beta}_2$

$$\widehat{\beta}_2 = \left[X_2'M_1X_2\right]^{-1} X_2'M_1y$$

Finalmente:

$$\widehat{\beta}_1 = \left[X_1'M_2X_1\right]^{-1} X_1'M_2y$$

Ahora bien, en el ejercicio 1 se vio que, dado $y = X\beta + u$, los residuos de la regresión pueden escribirse: $\widehat{u} = My$, donde $M = I - X\left(X'X\right)^{-1}X'$. Análogamente, podemos ver a $M_1X_2$ como una matriz de residuos, donde cada columna de $M_1X_2$ es un vector de residuos provenientes de una regresión de dicha columna de $X_2$ contra $X_1$. Esta observación nos lleva al Teorema de Frisch-Waugh:

**Theorem 3** *Teorema de Frisch-Waugh. En una regresión lineal de $y$ contra dos sets de variables, $X_1$ y $X_2$, el subvector $\widehat{\beta}_2$ se puede obtener regresando los residuos provenientes de regresar $y$ contra $X_1$, contra los residuos provenientes de regresar cada columna de $X_2$ contra $X_1$.*

Demostración:

Se realizan las regresiones de $y$ y de todas las columnas de $X_2$ sobre $X_1$:

$$\begin{aligned}
y &= X_1\gamma + v \\
X_2 &= X_1\Delta + W
\end{aligned}$$

El segundo modelo es simplemente una notación compactada de todas las regresiones de cada columna de $X_2$. Nótese que $\Delta$ y $W$ son matrices de coeficientes y residuos, respectivamente. Los estimadores de MCC y residuos asociados son:

$$\begin{aligned}
\widehat{\gamma} &= \left(X_1'X_1\right)^{-1} X_1'y \\
\widehat{v} &= y - X_1\widehat{\gamma} = M_1 y \\
\widehat{\Delta} &= \left(X_1'X_1\right)^{-1} X_1'X_2 \\
\widehat{W} &= X_2 - X_1\widehat{\Delta} = M_1X_2 \\
\text{Donde } M_1 &= I_n - X_1\left(X_1'X_1\right)^{-1} X_1'
\end{aligned}$$

Sea ahora la regresión de los residuos del primer modelo sobre la matriz de residuos del segundo modelo:

$$\begin{aligned}
\widehat{v} &= \widehat{W}\phi + \varepsilon \\
\widehat{\phi} &= \left(\widehat{W}'\widehat{W}\right)^{-1} \widehat{W}'\widehat{v} \\
\widehat{\phi} &= \left(X_2'M_1'M_1X_2\right)^{-1} X_2'M_1'M_1y \\
\widehat{\phi} &= \left(X_2'M_1X_2\right)^{-1} X_2'M_1y
\end{aligned}$$

Por los resultados de regresión particionada,

$$\widehat{\beta}_2 = \left(X_2'M_1X_2\right)^{-1} X_2'M_1y$$

Luego,

$$\boxed{\widehat{\phi} = \widehat{\beta}_2}$$

que es el resultado de Frisch-Waugh. Intercambiando los subíndices 1 y 2 es claro que también vale para $\widehat{\beta}_1$.

Notar:

$$\begin{aligned}
\widehat{\phi} &= \left(X_2'X_2\right)^{-1} X_2'y \Leftrightarrow M_1X_2 = X_2 \\
&\Leftrightarrow X_2 - X_1\left(X_1'X_1\right)^{-1} X_1'X_2 = X_2 \\
&\Leftrightarrow X_1\left(X_1'X_1\right)^{-1} X_1'X_2 = O \\
&\Leftrightarrow \boxed{X_1'X_2 = O}
\end{aligned}$$

En este caso se dice que los regresores son ortogonales.

Nótese entonces que en un modelo de regresión lineal múltiple con variable dependiente salario y variables explicativas educación y promedio en la carrera de grado, para obtener la estimación de mínimos cuadrados clásicos del coeficiente de educación se puede:

-Empezar regresando salario sobre promedio; y eduación sobre promedio.

-Obtener los residuos de la primera regresión y regresarlos sobre los residuos de la segunda regresión.

Así acabamos haciendo sólamente regresiones simples.

5. Sea el modelo

$$y = X\beta + c.z + u = \begin{bmatrix} X & \vdots & z \end{bmatrix} \begin{bmatrix} \beta \\ \dots \\ c \end{bmatrix} + u$$

donde $z$ es un vector columna con las observaciones de uno de los regresores del modelo. Se asume que la primer columna de $X$ es una columna de unos.

Considere las regresiones parciales respecto de $X$ y los residuos asociados:

$$\begin{aligned}
y^* &= y - X\widehat{\theta} \\
z^* &= z - X\widehat{\eta}
\end{aligned}$$

Como ambas regresiones tienen intercepto,

$$\sum_{i=1}^{n} y_i^* = \sum_{i=1}^{n} z_i^* = 0 \Rightarrow \overline{y^*} = \overline{z^*} = 0$$

Del ejercicio anterior sabemos que podemos obtener $\widehat{c}$ de regresar $y^*$ contra $z^*$: $y^* = cz^* + \varepsilon$.

El coeficiente de correlación (muestral) parcial entre $y, z$ respecto de $X$ es:

$$r_{y,z;X} = r_{y^*,z^*} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(y_i^* - \overline{y^*}\right)\left(z_i^* - \overline{z^*}\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i^* - \overline{y^*}\right)^2}\sqrt{\frac{1}{n}\left(z_i^* - \overline{z^*}\right)^2}} = \frac{y^{*T}z^*}{\sqrt{y^{*T}y^*}\sqrt{z^{*T}z^*}}$$

De acuerdo a los resultados de regresión particionada y de de MCC,

$$\begin{aligned} y^* &= \widehat{c}z^* + \widehat{\varepsilon} : z^{*T}\widehat{\varepsilon} = 0 \\ y^{*T} &= \widehat{c}z^{*T} + \widehat{\varepsilon}' \\ y^{*T}z^* &= \widehat{c}z^{*T}z^* + \widehat{\varepsilon}'z^* = c\left(z^{*T}z^*\right) \end{aligned}$$

Reemplazando esta expresión en el coeficiente de correlación parcial,

$$\boxed{r_{y,z;X} = \widehat{c}\frac{\sqrt{z^{*T}z^*}}{\sqrt{y^{*T}y^*}}}$$

6. Sea $Q_n$ la matriz sample demeaning:

$$\begin{aligned} Q_n x &= \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} \overline{x} \\ \vdots \\ \overline{x} \end{bmatrix} = \widetilde{x} \\ Q_n &= I_n - \frac{1}{n}1_n 1_n' \end{aligned}$$

Sea el modelo con intercepto:

$$y = X\beta + u$$

y la partición:

$$y = \begin{bmatrix} 1_n & \vdots & X_{-1} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \cdots \\ \beta_{-1} \end{bmatrix} + u$$

donde $\beta_1$ es el intercepto del modelo.

Por los resultados de regresión particionada,

$$\begin{aligned} \widehat{\beta}_{-1} &= \left(X_{-1}'M_1X_{-1}\right)^{-1}X_{-1}'M_1y \\ \widehat{\beta}_{-1} &= \left(X_{-1}'M_1'M_1X_{-1}\right)^{-1}X_{-1}'M_1'M_1y \\ \widehat{\beta}_{-1} &= \left[(M_1X_{-1})'(M_1X_{-1})\right]^{-1}(M_1X_{-1})'(M_1y) \\ M_1 &= I_n - 1_n\left(1_n'1_n\right)^{-1}1_n' \\ M_1 &= I_n - 1_n(n)^{-1}1_n' \\ M_1 &= I_n - \frac{1}{n}1_n1_n' \\ &\Rightarrow M_1 = Q_n \\ \widehat{\beta}_{-1} &= \left[(Q_nX_{-1})'(Q_nX_{-1})\right]^{-1}(Q_nX_{-1})'(Q_ny) \\ &\boxed{\widehat{\beta}_{-1} = \left[\widetilde{X}_{-1}'\widetilde{X}_{-1}\right]^{-1}\widetilde{X}_{-1}'\widetilde{y}} \end{aligned}$$

7. El estimador de mínimos cuadrados restringidos es el estimador de $\beta$ que resuelve el siguiente problema:

$$\begin{aligned} &\min_{\widetilde{\beta}}\left(y - X\widetilde{\beta}\right)'\left(y - X\widetilde{\beta}\right) \\ s.a &: R\widetilde{\beta} = r \end{aligned}$$

El lagrangiano del problema es:

$$\begin{aligned} \mathcal{L} &= \left(y - X\widetilde{\beta}\right)'\left(y - X\widetilde{\beta}\right) - 2\lambda'\left(R\widetilde{\beta} - r\right) \\ &= y'y - 2y'X\widetilde{\beta} + \widetilde{\beta}'X'X\widetilde{\beta} - 2\lambda'\left(R\widetilde{\beta} - r\right) \end{aligned}$$

Las condiciones de primer orden son:

$$\begin{aligned} -2X'y + 2X'X\widetilde{\beta} - 2R'\lambda &= 0 \\ R\widetilde{\beta} &= r \end{aligned}$$

Trabajamos un poco las expresiones:

$$\begin{aligned} -X'y + X'X\widetilde{\beta} - R'\lambda &= 0 \\ -R(X'X)^{-1}X'y + R(X'X)^{-1}X'X\widetilde{\beta} &= R(X'X)^{-1}R'\lambda \\ -R\widehat{\beta} + R\widetilde{\beta} &= R(X'X)^{-1}R'\lambda \\ -R\widehat{\beta} + r &= R(X'X)^{-1}R'\lambda \\ \left(R(X'X)^{-1}R'\right)^{-1}\left(r - R\widehat{\beta}\right) &= \lambda \\ -X'y + X'X\widetilde{\beta} - R'\left(R(X'X)^{-1}R'\right)^{-1}\left(r - R\widehat{\beta}\right) &= 0 \end{aligned}$$

$$\begin{aligned} X'X\widetilde{\beta} &= X'y + R'\left(R(X'X)^{-1}R'\right)^{-1}\left(r - R\widehat{\beta}\right) \\ \widetilde{\beta} &= (X'X)^{-1}X'y + (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}\left(r - R\widehat{\beta}\right) \\ \widetilde{\beta} &= \widehat{\beta} + (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}\left(r - R\widehat{\beta}\right) \\ \widetilde{\beta} &= \widehat{\beta} - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}\left(R\widehat{\beta} - r\right) \\ \widetilde{\beta} &= \left\{I_k - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R\right\}\widehat{\beta} \\ &\quad + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}r \\ \text{Reescribimos usando } \widehat{\beta} &= \beta + (X'X)^{-1}X'u \\ \widetilde{\beta} &= \left\{I_k - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R\right\}\beta \\ &\quad + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}r \\ &\quad + \left\{I_k - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R\right\}(X'X)^{-1}X'u \\ &= \beta - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R\beta \\ &\quad + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}r \\ &\quad + \left\{I_k - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R\right\}(X'X)^{-1}X'u \\ &= \beta + \left\{I_k - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R\right\}(X'X)^{-1}X'u \\ &= \beta + Cu \end{aligned}$$

que es una transformación afín de $u$. Luego, es normal. Busquemos la esperanza y varianza:

$$\begin{aligned} E(\widetilde{\beta}/X) &= \beta \\ Var(\widetilde{\beta}/X) &= E\left[(\widetilde{\beta} - \beta)(\widetilde{\beta} - \beta)'/X\right] \\ &= E\left[Cu(Cu)'/X\right] \\ &= CE\left[u\,u'/X\right]C \\ &= \sigma_u^2 CC' \\ &= \sigma_u^2\left\{(X'X)^{-1} - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}\right\} \end{aligned}$$

La distribución es:

$$\boxed{\widehat{\beta}_R/X \sim \mathcal{N}\left(\beta, \sigma_u^2 \left\{(X'X)^{-1} - (X'X)^{-1} R' \left[R(X'X)^{-1} R'\right]^{-1} R(X'X)^{-1}\right\}\right)}$$

8. El vector de residuos restringidos y su suma de cuadrados son:

$$\begin{aligned}
\widehat{u}_R &= y - X\widehat{\beta}_R \\
&= y - X\widehat{\beta}_R + X\widehat{\beta} - X\widehat{\beta} \\
&= \widehat{u} - X\left(\widehat{\beta}_R - \widehat{\beta}\right) \\
\widehat{u}_R'\widehat{u}_R &= \left[\widehat{u} - X\left(\widehat{\beta}_R - \widehat{\beta}\right)\right]'\left[\widehat{u} - X\left(\widehat{\beta}_R - \widehat{\beta}\right)\right] \\
&= \widehat{u}'\widehat{u} + \left(\widehat{\beta}_R - \widehat{\beta}\right)' X'X\left(\widehat{\beta}_R - \widehat{\beta}\right) \\
&\quad - \left(\widehat{\beta}_R - \widehat{\beta}\right)' X'\widehat{u} - \widehat{u}'X\left(\widehat{\beta}_R - \widehat{\beta}\right) \\
X'\widehat{u} &= 0_k \Rightarrow \widehat{u}_R'\widehat{u}_R = \widehat{u}'\widehat{u} + \left(\widehat{\beta}_R - \widehat{\beta}\right)' X'X\left(\widehat{\beta}_R - \widehat{\beta}\right) \\
\widehat{\beta}_R - \widehat{\beta} &= -(X'X)^{-1} R'\left[R(X'X)^{-1} R'\right]^{-1}\left(R\widehat{\beta} - r\right) \\
\left(\widehat{\beta}_R - \widehat{\beta}\right)' X'X\left(\widehat{\beta}_R - \widehat{\beta}\right) &= \left(R\widehat{\beta} - r\right)'\left[R(X'X)^{-1} R'\right]^{-1} R(X'X)^{-1}(X'X) \\
&\quad (X'X)^{-1} R'\left[R(X'X)^{-1} R'\right]^{-1}\left(R\widehat{\beta} - r\right) \\
&= \left(R\widehat{\beta} - r\right)'\left[R(X'X)^{-1} R'\right]^{-1} R(X'X)^{-1} \\
&\quad R'\left[R(X'X)^{-1} R'\right]^{-1}\left(R\widehat{\beta} - r\right) \\
&= \left(R\widehat{\beta} - r\right)'\left[R(X'X)^{-1} R'\right]^{-1}\left(R\widehat{\beta} - r\right) \\
&\Rightarrow \boxed{\widehat{u}_R'\widehat{u}_R - \widehat{u}'\widehat{u} = \left(R\widehat{\beta} - r\right)'\left[R(X'X)^{-1} R'\right]^{-1}\left(R\widehat{\beta} - r\right) \geq 0}
\end{aligned}$$

Este resultado implica que el estadístico F equivale al siguiente estadístico de residuos restringidos:

$$\begin{aligned}
\frac{\left[\frac{\widehat{u}_R'\widehat{u}_R - \widehat{u}'\widehat{u}}{q}\right]}{\widehat{\sigma}_u^2} &\sim {}^{H_0}\mathcal{F}(q, n-k) \\
\frac{\left[\frac{\widehat{u}_R'\widehat{u}_R - \widehat{u}'\widehat{u}}{q}\right]}{\left[\frac{\widehat{u}'\widehat{u}}{n-k}\right]} &\sim {}^{H_0}\mathcal{F}(q, n-k) \\
&\boxed{\frac{\left[\frac{SSR_R - SSR}{q}\right]}{\left[\frac{SSR}{n-k}\right]} \sim^{H_0} \mathcal{F}(q, n-k)}
\end{aligned}$$

9. Para probar este resultado, nótese que la inclusión de nuevos regresores a un modelo puede pensarse como un caso particular de una restricción de exclusión de q variables:

$$y = X\beta + \varepsilon \rightarrow y = X\beta + Z\gamma + u \Leftrightarrow \gamma = 0_q$$

donde Z es la matriz de observaciones de q regresores adicionales. Sean $\mathrm{RSS}_0$ la suma de residuos cuadrados del modelo restringido y $\mathrm{RSS}_1$, del modelo ampliado. Estas sumas pueden pensarse como:

$$\begin{aligned}
RSS_1 &= RSS = \widehat{u}'\widehat{u} \\
RSS_0 &= RSS_R = \widehat{\varepsilon}'\widehat{\varepsilon} = \widehat{u}_R'\widehat{u}_R
\end{aligned}$$

Por el resultado del ejercicio 8:

$$\begin{aligned}
\widehat{u}_R'\widehat{u}_R &\geq \widehat{u}'\widehat{u} \\
\frac{\widehat{u}_R'\widehat{u}_R}{y'y-n\overline{y}^2} &\geq \frac{\widehat{u}'\widehat{u}}{y'y-n\overline{y}^2} \\
-\frac{\widehat{u}_R'\widehat{u}_R}{y'y-n\overline{y}^2} &\leq -\frac{\widehat{u}'\widehat{u}}{y'y-n\overline{y}^2} \\
1-\frac{\widehat{u}_R'\widehat{u}_R}{y'y-n\overline{y}^2} &\leq 1-\frac{\widehat{u}'\widehat{u}}{y'y-n\overline{y}^2} \\
& \boxed{R_0^2 \leq R_1^2}\checkmark
\end{aligned}$$

10. Probamos la convergencia en probabilidad. Sea

$$\begin{aligned}
\widehat{\sigma}_u^2 &= \frac{\widehat{u}'\widehat{u}}{n-k} \\
&= \frac{u'Mu}{n-k} \\
&= \frac{u'u-u'X(X'X)^{-1}X'u}{n-k} \\
&= \frac{n}{n-k}\left(\frac{u'u}{n}-\frac{u'X}{n}(\frac{X'X}{n})^{-1}\frac{X'u}{n}\right) \\
p\lim \widehat{\sigma}_u^2 &= 1.\left(E(u_i^2)-0_k'Q^{-1}0_k\right) \\
p\lim \widehat{\sigma}_u^2 &= E(u_i^2) \\
p\lim \widehat{\sigma}_u^2 &= \sigma_u^2
\end{aligned}$$

11. Sea el modelo simple

$$y_i = \alpha + \beta x_i^* + u_i$$

El investigador estima en su lugar el siguiente modelo

$$\begin{aligned}
y_i &= \alpha + \beta\left(x_i - \varepsilon_i\right) + u_i \\
y_i &= \alpha + \beta x_i + u_i - \beta\varepsilon_i \\
y_i &= \alpha + \beta x_i + v_i
\end{aligned}$$

Por tratarse de un modelo simple:

$$\begin{aligned}
\widehat{\beta} &= \frac{\sum_{i=1}^n (x_i-\overline{x})y_i}{\sum_{i=1}^n (x_i-\overline{x})^2} \\
\widehat{\beta} &= \frac{\sum_{i=1}^n (x_i-\overline{x})(\alpha+\beta x_i+v_i)}{\sum_{i=1}^n (x_i-\overline{x})^2} \\
\widehat{\beta} &= \frac{\sum_{i=1}^n (x_i-\overline{x})\alpha}{\sum_{i=1}^n (x_i-\overline{x})^2} + \beta\frac{\sum_{i=1}^n (x_i-\overline{x})x_i}{\sum_{i=1}^n (x_i-\overline{x})^2} + \frac{\sum_{i=1}^n (x_i-\overline{x})v_i}{\sum_{i=1}^n (x_i-\overline{x})^2} \\
\widehat{\beta} &= \beta + \frac{\sum_{i=1}^n (x_i-\overline{x})v_i}{\sum_{i=1}^n (x_i-\overline{x})^2} \\
&= \beta + \frac{\left(\frac{\sum_{i=1}^n (x_i-\overline{x})v_i}{n}\right)}{\left(\frac{\sum_{i=1}^n (x_i-\overline{x})^2}{n}\right)} \\
p\lim \widehat{\beta} &= \beta + \frac{cov(x_{i,}v_i/X)}{var(x_i/X)}
\end{aligned}$$

En este caso,

$$\begin{aligned}
cov(x_i^*, \varepsilon_i/X) &= 0 \\
var(x_i/X) &= \sigma_{x^*}^2 + \sigma_\varepsilon^2 \\
cov(x_i, v_i/X) &= E(x_i.v_i/X) \\
&= E(x_i.u_i/X) - \beta E(x_i.\varepsilon_i/X) \\
&= E(x_i.u_i/X) - \beta E((x_i^* + \varepsilon_i).\varepsilon_i/X) \\
&= 0 - \beta 0 - \beta\sigma_\varepsilon^2 \\
&= -\beta\sigma_\varepsilon^2 \\
&\Rightarrow p\lim\widehat{\beta} = \beta - \frac{\beta\sigma_\varepsilon^2}{\sigma_{x^*}^2 + \sigma_\varepsilon^2} \\
p\lim\widehat{\beta} &= \beta\left[1 - \frac{\sigma_\varepsilon^2}{\sigma_{x^*}^2 + \sigma_\varepsilon^2}\right] \\
&= \beta\left[\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_\varepsilon^2}\right] \\
&= \beta\left[\frac{1}{1 + \left(\frac{\sigma_\varepsilon^2}{\sigma_{x^*}^2}\right)}\right]
\end{aligned}$$

El estimador de MCC del coeficiente pendiente es inconsistente. El parámetro aparece multiplicado por un factor menos a uno, que se conoce como "sesgo de atenuación". El tamaño de este sesgo depende de la varianza relativa de la variable y el error de medición. Si este error fuera sistemático, $\sigma_\varepsilon^2 = 0$, el estimador preservaría la consistencia.

Universidad Torcuato Di Tella
Master en Economía y Econometría
2019

**Econometría**
**Problem Set 4: Heteroscedasticidad y autocorrelación**

1. Sea el modelo:

$$\begin{aligned} y_t &= \alpha x_t + u_t, \qquad t = 1, 2, ..., n \\ y &= \alpha x + u \\ E(u/x) &= 0_n \\ E\left[uu^T/x\right] &= \beta.diag(xx^T) \end{aligned}$$

donde $\{u_t\}$ son variables aleatorias independientes condicionalmente Normales con media 0 y varianza $\beta x_t^2$. Nótese que para que efectivamente sean normales los errores se esta asumiendo que todas las observaciones del regresor son distintas de 0. Se desea estimar $\alpha$ con un estimador lineal insesgado. Sea un miembro de la clase de estimadores lineales en $y$, identificados por un vector real $c$ :

$$\begin{aligned} \widetilde{\alpha} &= c^T y \\ &= \alpha c^T x + c^T u \end{aligned}$$

Para restringir la clase de estimadores lineales a los lineales insesgados:

$$\begin{aligned} E(\widetilde{\alpha}/x) &= \alpha c^T x + c^T E(u/x) \\ &= \alpha c^T x \\ &= \alpha \Leftrightarrow c^T x = 1 \end{aligned}$$

La varianza de este estimador es:

$$\begin{aligned} var(\widetilde{\alpha}/x) &= E\left[(\widetilde{\alpha} - E(\widetilde{\alpha}/x))\left(\widetilde{\alpha} - E(\widetilde{\alpha}/x)\right)^T/x\right] \\ &= E\left[\left(c^T u\right)\left(c^T u\right)^T/x\right] \\ &= c^T E\left[uu^T/x\right] c \\ &= \beta.c^T diag(xx^T)c \end{aligned}$$

El problema a resolver es hallar:

$$c^* \in \arg\min_c \left\{\beta.c^T diag(xx^T)c \;\; s.a : c^T x = 1\right\}$$

$$\mathcal{L} = \beta . c^T diag(xx^T)c - 2\lambda \left(c^T x - 1\right)$$

Las C.P.O. son

$$\begin{aligned}
\nabla_c \mathcal{L} &= \beta 2 diag(xx^T)c - 2\lambda x \\
D_\lambda \mathcal{L} &= -2\left(c^T x - 1\right) \\
& \left\{ \begin{array}{c} \beta 2 diag(xx^T)c^* - 2\lambda^* x = 0 \\ -2\left(c^{*T} x - 1\right) = 0 \end{array} \right. \\
& \left\{ \begin{array}{c} \beta diag(xx^T)c^* = \lambda^* x \\ c^{*T} x = 1 \end{array} \right.
\end{aligned}$$

La matriz diagonal es no singular. Luego, de la primer condición se obtiene:

$$c^* = \lambda^* \left(\beta diag(xx^T)\right)^{-1} x$$

De acuerdo a la segunda condición:

$$\begin{aligned}
c^{*T} x &= 1 \\
\left[\lambda^* \left(\beta diag(xx^T)\right)^{-1} x\right]^T x &= 1 \\
\lambda^* x^T \left(\beta diag(xx^T)\right)^{-1} x &= 1 \\
\lambda^* &= \left[x^T \left(\beta diag(xx^T)\right)^{-1} x\right]^{-1}
\end{aligned}$$

siempre y cuando este último escalar sea no nulo. Volviendo a la anterior ecuación,

$$\begin{aligned}
c^* &= \left[x^T \left(\beta diag(xx^T)\right)^{-1} x\right]^{-1} \left(\beta diag(xx^T)\right)^{-1} x \\
&= \left[x^T diag(xx^T)^{-1} x\right]^{-1} diag(xx^T)^{-1} x
\end{aligned}$$

Esta expresión se reduce a:

$$\begin{aligned}
diag(xx^T)^{-1} x &= \begin{bmatrix} \frac{1}{x_l^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{x_n^2} \end{bmatrix} x \\
&= \begin{bmatrix} \frac{1}{x_l} \\ \vdots \\ \frac{1}{x_n} \end{bmatrix} \\
x^T diag(xx^T)^{-1} x &= x^T \begin{bmatrix} \frac{1}{x_l} \\ \vdots \\ \frac{1}{x_n} \end{bmatrix} \\
&= n
\end{aligned}$$

$$c^* = \frac{1}{n}\begin{bmatrix} \frac{1}{x_l} \\ \vdots \\ \frac{1}{x_n} \end{bmatrix}$$

Volviendo al problema,

$$\begin{aligned}
\widehat{\alpha} &= \widetilde{\alpha}^* = c^{*T} y \\
\widehat{\alpha} &= \frac{1}{n}\begin{bmatrix} \frac{1}{x_l} \\ \vdots \\ \frac{1}{x_n} \end{bmatrix}^T y \\
& \boxed{\widehat{\alpha} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{x_i}\right)}
\end{aligned}$$

Se puede ver claramente que el estimador hallado es el estimador gls. La transformación relevante consiste en dividir todas las unidades muestrales por la observación correspondiente del regresor. El modelo transformado es un modelo con sólo constante y variable dependiente

$$\frac{y_i}{x_i}$$

El estimador ols del nuevo modelo, que es el estimador gls, es para este modelo constante la media muestral de la variable dependiente, el resultado indicado.

2. Sea el modelo en los datos observados:

$$y_j = \beta^T x_j + u_j$$

Las propiedades del error en este modelo son:

$$\begin{aligned}
E(u_j/x) &= E\left(\sum_{i=1}^{n_j} u_i/x\right) \\
&= \sum_{i=1}^{n_j} E(u_i/x) \\
&= 0 \\
var(u_j/x) &= var\left(\sum_{i=1}^{n_j} u_i/x\right) \\
&= n_j \sigma_u^2 \\
&\Rightarrow \quad E(u.u^T/x) = \sigma_u^2 \begin{bmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & n_J \end{bmatrix}
\end{aligned}$$

La agregación ha generado heteroscedasticidad.

Si se toman promedios:

$$\begin{aligned}
\overline{u}_j &= \frac{1}{n_j}\sum_{i=1}^{n_j} u_i \\
E(\overline{u}_j/x) &= 0 \\
E(\overline{u}.\overline{u}^T/x) &= \sigma_u^2 \begin{bmatrix} n_1^{-1} & 0 & \cdots & 0 \\ 0 & n_2^{-1} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & n_J^{-1} \end{bmatrix}
\end{aligned}$$

cambia el patrón pero persiste la heteroscedasticidad.

3. Se tiene el modelo

$$\begin{aligned}
y_t &= \beta y_{t-1} + u_t, \qquad |\beta| < 1 \\
u_t &= \varepsilon_t + \theta\varepsilon_{t-1}, \qquad |\theta| < 1
\end{aligned}$$

($a$) El estimador ols de $\beta$ es:

$$\begin{aligned}
\widehat{\beta} &= \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2} \\
&= \frac{\sum (\beta y_{t-1} + u_t)\, y_{t-1}}{\sum y_{t-1}^2} \\
&= \frac{\sum \beta y_{t-1}^2 + \sum u_t y_{t-1}}{\sum {y_{t-1}}^2} \\
&= \beta + \frac{\sum u_t y_{t-1}}{\sum {y_{t-1}}^2} \\
&= \beta + \frac{\left(\frac{\sum u_t y_{t-1}}{n}\right)}{\left(\frac{\sum {y_{t-1}}^2}{n}\right)}
\end{aligned}$$

$$\begin{aligned}
p\lim \widehat{\beta} &= \beta + \frac{p\lim\left(\frac{\sum u_t y_{t-1}}{n}\right)}{p\lim\left(\frac{\sum {y_{t-1}}^2}{n}\right)} \\
&= \beta + \frac{Cov\,(y_{t-1}, u_t)}{Var\,(y_{t-1})}
\end{aligned}$$

Los momentos poblacionales son:

$$\begin{aligned}
Cov\left(y_{t-1}, u_t\right) &= E\left(y_{t-1} u_t\right) \\
&= E\left[y_{t-1}\left(\varepsilon_t+\theta \varepsilon_{t-1}\right)\right] \\
&= E\left[y_{t-1} \varepsilon_t\right]+\theta E\left[y_{t-1} \varepsilon_{t-1}\right] \\
&= 0+\theta E\left[\left(\beta y_{t-2}+u_{t-1}\right) \varepsilon_{t-1}\right] \\
&= \theta E\left[\beta y_{t-2} \varepsilon_{t-1}\right]+\theta E\left[u_{t-1} \varepsilon_{t-1}\right] \\
&= \theta E\left[\beta y_{t-2} \varepsilon_{t-1}\right]+\theta E\left[\left(\varepsilon_{t-1}+\theta \varepsilon_{t-2}\right) \varepsilon_{t-1}\right] \\
&= 0+\theta E\left[\varepsilon_{t-1}^2\right]+\theta^2 E\left[\varepsilon_{t-2} \varepsilon_{t-1}\right] \\
&= 0+\theta E\left[\varepsilon_{t-1}^2\right]+0 \\
&= \theta \sigma_{\varepsilon}^2
\end{aligned}$$

Dado que $|\beta|<1$, se puede probar que la varianza de $y_{t-1}$ es constante e igual a:

$$Var\left(y_{t-1}\right)=\gamma_0=\frac{1+2 \beta \theta+\theta^2}{1-\beta^2} . \sigma_{\varepsilon}^2$$

Luego,

$$\begin{aligned}
p \lim \widehat{\beta} &= \beta+\frac{\theta \sigma_{\varepsilon}^2}{\gamma_0} \\
&= \beta+\frac{\theta}{\left(\frac{1+2 \beta \theta+\theta}{1-\beta^2}\right)}
\end{aligned}$$

Como se ve,

$$p \lim \widehat{\beta} \neq \beta, \text{ cuando } \theta \neq 0$$

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:

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

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:

$$\begin{aligned}
D^2_\beta l(\beta,\sigma_u^2;y,X) &= D_\beta\nabla_\beta l(\beta,\sigma_u^2;y,X) = \frac{\partial^2 \ln L}{\partial\beta\partial\beta'} = -\frac{1}{\sigma_u^2}X^TX \\
D^2_{\sigma_u^2} l(\beta,\sigma_u^2;y,X) &= \frac{\partial^2 \ln L}{\partial\sigma_u^2\partial\sigma_u^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}u^Tu \\
D_{\sigma_u^2}\nabla_\beta l(\beta,\sigma_u^2;y,X) &= \frac{\partial^2 \ln L}{\partial\beta\partial\sigma_u^2} = -\frac{1}{\sigma_u^4}X^Tu \\
H(\beta,\sigma_u^2;y,X) &= \begin{bmatrix} \frac{\partial^2 \ln L}{\partial\beta\partial\beta'} & \frac{\partial^2 \ln L}{\partial\beta\partial\sigma_u^2} \\ \frac{\partial^2 \ln L}{\partial\sigma_u^2\partial\beta'} & \frac{\partial^2 \ln L}{\partial\sigma^2\partial\sigma_u^2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sigma_u^2}X^TX & -\frac{1}{\sigma_u^4}X^Tu \\ -\frac{1}{\sigma_u^4}u^TX & \frac{n}{2\sigma^4}-\frac{1}{\sigma^6}u^Tu \end{bmatrix}
\end{aligned}$$

Evaluada en el estimador de máxima verosimilitud,

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

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:

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

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 $\gamma$ se obtiene igualando el score a un vector nulo:

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

Los estimadores máximo verosímiles de $\theta$ y de $\tau$ pueden obtener fácilmente a partir del principio de invarianza. Para $\theta$ :

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

Para $\tau$ :

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

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,

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

Para los parámetros $\theta$ y $\tau$ la distribución puede obtenerse de forma sencilla como funciones de $\gamma$.Siendo estas funciones diferenciables, ambas distribuciones límite pueden obtenerse por el método delta.

Para $\theta$ :

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

Para $\tau$ :

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

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 $\beta$ y $\sigma_u^2$ son asintóticamente incorrelacionados. Pero además, bajo normalidad, la incorrelación equivale a independencia.

   El estimador de $\beta$ sin conocer el verdadero valor de $\sigma_u^2$ 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 $\sigma_u^2$ el problema sería maximizar la log-likelihood sólo respecto de $\beta$ :

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

   El score respecto de $\beta$ 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:

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

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

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

Claramente,

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

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 $\phi$ 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:

$$S(y, X) = \left[\begin{array}{c} S_\theta(y, X) \\ S_\phi(y, X) \end{array}\right]$$

El estadístico $S(y, X)$ es ancilliar para $\theta$ 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

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

Supóngase que la densidad de la muestra de los regresores no depende de $\theta$. 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 $\theta$. Así se establece lo deseado.

El principio de Condicionalidad de la inferencia clásica sugiere entonces que la inferencia sobre $\theta$ 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 $\theta$, 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 $\theta$ 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 $\beta$ es:

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

El estadístico de LM es:

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

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 $\delta$ 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.