MGE-201: Test basados en la verosimilitud

Felipe Osorio

felipe.osorio@uv.cl

Problema:

Deseamos probar hipótesis (no lineales) de la forma:

$$H_0 : \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}, \qquad \text{versus} \qquad H_1 : \boldsymbol{g}(\boldsymbol{\theta}) \neq \mathbf{0},$$

donde $\boldsymbol{g} : \mathbb{R}^p \to \mathbb{R}^q$, tal que $\boldsymbol{G}(\boldsymbol{\theta}) = \partial \boldsymbol{g}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^\top$ es una matriz $q \times p$ con rango q. En otras palabras, bajo H_0 deseamos resolver el problema restringido:

$$\max_{\theta \in \Theta} \ell_n(\boldsymbol{\theta}), \qquad \text{sujeto a:} \quad \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}.$$

Observación:

Hipótesis del tipo

$$H_0 : \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}, \qquad \text{vs.} \qquad H_1 : \boldsymbol{g}(\boldsymbol{\theta}) \neq \mathbf{0},$$

se dicen en forma implícita.

Mientras que,

$$H_0 : \boldsymbol{\theta} = \boldsymbol{h}(\boldsymbol{\gamma}), \qquad \text{vs.} \qquad H_1 : \boldsymbol{\theta} \neq \boldsymbol{h}(\boldsymbol{\gamma}),$$

con espacio paramétrico nulo,

$$\Theta_0 = \{\boldsymbol{\theta} : \boldsymbol{\theta} = \boldsymbol{h}(\boldsymbol{\gamma}), \boldsymbol{\gamma} \in \Gamma\}, \qquad \boldsymbol{h}(\boldsymbol{\gamma}) = (h_1(\boldsymbol{\gamma}), \dots, h_p(\boldsymbol{\gamma}))^\top,$$

está en forma explícita. Finalmente, la hipótesis nula definida como:

$$H_0 : \{\boldsymbol{\theta} : \exists \boldsymbol{a} \in \mathcal{A} \subset \mathbb{R}^k, \boldsymbol{g}(\boldsymbol{\theta}, \boldsymbol{a}) = \mathbf{0}\},$$

está escrita en forma mixta.[1]

[1] Gourieroux y Monfort (1989). Econometric Theory **5**, 63-82.

Resultado 1 (test de Wald):[2]

El test de Wald para probar $H_0 : \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}$, es definido por la región crítica,

$$\{W_n \geq \chi^2_{1-\alpha}(q)\},$$

donde

$$W_n = n\boldsymbol{g}^\top(\widehat{\boldsymbol{\theta}}_n)[\boldsymbol{G}(\widehat{\boldsymbol{\theta}}_n)\boldsymbol{\mathcal{F}}^{-1}(\widehat{\boldsymbol{\theta}}_n)\boldsymbol{G}^\top(\widehat{\boldsymbol{\theta}}_n)]^{-1}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n),$$

y bajo H_0 es asintóticamente de tamaño α.

[2] Wald (1943). Transactions of the American Mathematical Society **54**, 426-482.

Demostración:

Sabemos que

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathsf{D}} \mathsf{N}_p(\mathbf{0}, \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)).$$

De ahí que

$$\sqrt{n}(\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) - \boldsymbol{g}(\boldsymbol{\theta}_0)) \xrightarrow{\mathsf{D}} \mathsf{N}_q(\mathbf{0}, \boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}^\top(\boldsymbol{\theta}_0)).$$

Bajo H_0, tenemos $\boldsymbol{g}(\boldsymbol{\theta}_0) = \mathbf{0}$. Luego,

$$\sqrt{n}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \xrightarrow{\mathsf{D}} \mathsf{N}_q(\mathbf{0}, \boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}^\top(\boldsymbol{\theta}_0)). \tag{W.1}$$

Como $\boldsymbol{G}(\boldsymbol{\theta}_0)$ es de rango fila completo, sigue que

$$\sqrt{n}[\boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}^\top(\boldsymbol{\theta}_0)]^{-1/2}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \xrightarrow{\mathsf{D}} \mathsf{N}_q(\mathbf{0}, \boldsymbol{I}).$$

Note que $\widehat{\boldsymbol{\theta}}_n$ es un estimador consistente de $\boldsymbol{\theta}_0$. Haciendo,

$$\boldsymbol{Z}_n = \sqrt{n}[\boldsymbol{G}(\widehat{\boldsymbol{\theta}}_n)\boldsymbol{\mathcal{F}}^{-1}(\widehat{\boldsymbol{\theta}}_n)\boldsymbol{G}^\top(\widehat{\boldsymbol{\theta}}_n)]^{-1/2}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \xrightarrow{\mathsf{D}} \mathsf{N}_q(\mathbf{0}, \boldsymbol{I}).$$

De ahí que, bajo H_0

$$\begin{aligned} W_n &= n\boldsymbol{g}^\top(\widehat{\boldsymbol{\theta}}_n)[\boldsymbol{G}(\widehat{\boldsymbol{\theta}}_n)\boldsymbol{\mathcal{F}}^{-1}(\widehat{\boldsymbol{\theta}}_n)\boldsymbol{G}^\top(\widehat{\boldsymbol{\theta}}_n)]^{-1}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \\ &= \boldsymbol{Z}_n^\top \boldsymbol{Z}_n \xrightarrow{\mathsf{D}} \chi^2(q). \end{aligned}$$

Observación:

Para hipótesis de la forma $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$, donde $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$. Es este caso el estadístico de Wald asume la forma:

$$W_n = n(\widehat{\boldsymbol{\theta}}_{1n} - \boldsymbol{\theta}_1)^\top \boldsymbol{K}_{11}^{-1}(\widehat{\boldsymbol{\theta}}_n)(\widehat{\boldsymbol{\theta}}_{1n} - \boldsymbol{\theta}_1),$$

con

$$\boldsymbol{K}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{K}_{11}(\boldsymbol{\theta}) & \boldsymbol{K}_{12}(\boldsymbol{\theta}) \\ \boldsymbol{K}_{21}(\boldsymbol{\theta}) & \boldsymbol{K}_{22}(\boldsymbol{\theta}) \end{pmatrix} = \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}).$$

Note que

$$\boldsymbol{K}_{11}(\boldsymbol{\theta}) = (\boldsymbol{\mathcal{F}}_{11}(\boldsymbol{\theta}) - \boldsymbol{\mathcal{F}}_{12}(\boldsymbol{\theta})\boldsymbol{\mathcal{F}}_{22}^{-1}(\boldsymbol{\theta})\boldsymbol{\mathcal{F}}_{21}(\boldsymbol{\theta}))^{-1},$$

con

$$\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mathcal{F}}_{11}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{F}}_{12}(\boldsymbol{\theta}) \\ \boldsymbol{\mathcal{F}}_{21}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{F}}_{22}(\boldsymbol{\theta}) \end{pmatrix}.$$

Test basados en la verosimilitud

Sea $\widetilde{\boldsymbol{\theta}}_n$ el MLE de $\boldsymbol{\theta}$ sujeto a la restricción $\boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}$. La función Langrangiana asociada con el problema restringido es

$$\ell_n(\boldsymbol{\theta}) - \boldsymbol{g}^\top(\boldsymbol{\theta})\boldsymbol{\lambda},$$

y las condiciones de primer orden son

$$\begin{aligned} \frac{\partial \ell_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} - \boldsymbol{G}^\top(\widetilde{\boldsymbol{\theta}}_n)\widetilde{\boldsymbol{\lambda}}_n &= \mathbf{0}, \\ \boldsymbol{g}(\widetilde{\boldsymbol{\theta}}_n) &= \mathbf{0}, \end{aligned}$$

donde $\widetilde{\boldsymbol{\lambda}}_n$ es un vector de multiplicadores de Lagrange.

Resultado 2 (test score o de multiplicadores de Lagrange):[3]

El estadístico score o de multiplicadores de Lagrange para probar la hipótesis $H_0 : \boldsymbol{g}(\boldsymbol{\theta}_0) = \mathbf{0}$, es dado por

$$\begin{aligned} R_n &= \frac{1}{n}\Big(\frac{\partial \ell_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}\Big)^\top \boldsymbol{\mathcal{F}}^{-1}(\widetilde{\boldsymbol{\theta}}_n)\Big(\frac{\partial \ell_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}\Big) \\ &= \frac{1}{n}\widetilde{\boldsymbol{\lambda}}_n^\top \boldsymbol{G}(\widetilde{\boldsymbol{\theta}}_n)\boldsymbol{\mathcal{F}}^{-1}(\widetilde{\boldsymbol{\theta}}_n)\boldsymbol{G}^\top(\widetilde{\boldsymbol{\theta}}_n)\widetilde{\boldsymbol{\lambda}}_n, \end{aligned}$$

y bajo H_0 tiene distribución asintótica chi-cuadrado con q grados de libertad.

[3] Rao (1948). Proceedings of the Cambridge Philosophical Society **44**, 50-57.

Demostración:

Considere una expansión de Taylor en torno de $\boldsymbol{\theta}_0$,[4] como

$$\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \stackrel{\mathsf{a}}{=} \boldsymbol{g}(\boldsymbol{\theta}_0) + \frac{\partial \boldsymbol{g}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

y análogamente para $\boldsymbol{g}(\widetilde{\boldsymbol{\theta}}_n)$. De ahí que

$$\begin{aligned} \sqrt{n}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) &\stackrel{\mathsf{a}}{=} \boldsymbol{G}(\boldsymbol{\theta}_0)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \\ \sqrt{n}\boldsymbol{g}(\widetilde{\boldsymbol{\theta}}_n) &\stackrel{\mathsf{a}}{=} \boldsymbol{G}(\boldsymbol{\theta}_0)\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \end{aligned}$$

tomando diferencias, obtenemos

$$\begin{aligned} \sqrt{n}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) - \sqrt{n}\boldsymbol{g}(\widetilde{\boldsymbol{\theta}}_n) &\stackrel{\mathsf{a}}{=} \boldsymbol{G}(\boldsymbol{\theta}_0)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \boldsymbol{G}(\boldsymbol{\theta}_0)\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ &= \boldsymbol{G}(\boldsymbol{\theta}_0)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n), \end{aligned}$$

como $\widetilde{\boldsymbol{\theta}}_n$ es el MLE restringido, tenemos $\boldsymbol{g}(\widetilde{\boldsymbol{\theta}}_n) = \boldsymbol{0}$. De este modo,

$$\sqrt{n}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \stackrel{\mathsf{a}}{=} \boldsymbol{G}(\boldsymbol{\theta}_0)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n). \qquad \text{(R.1)}$$

[4] $\boldsymbol{X} \stackrel{\mathsf{a}}{=} \boldsymbol{Y}$ indica que $\boldsymbol{X} - \boldsymbol{Y} = o_{\mathsf{P}}(\boldsymbol{1})$

Por otro lado,

$$\frac{1}{\sqrt{n}}\frac{\partial \ell_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \stackrel{a}{=} \frac{1}{\sqrt{n}}\frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

y

$$\mathbf{0} = \frac{1}{\sqrt{n}}\frac{\partial \ell_n(\widehat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \stackrel{a}{=} \frac{1}{\sqrt{n}}\frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0). \tag{R.2}$$

Tomando diferencias, obtenemos

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial \ell_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} &\stackrel{a}{=} \frac{1}{\sqrt{n}}\frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{1}{\sqrt{n}}\frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \\ &\quad - \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ &\stackrel{a}{=} \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n).\end{aligned}$$

Por tanto,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n) \stackrel{a}{=} \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)\frac{1}{\sqrt{n}}\frac{\partial \ell_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}. \tag{R.3}$$

Test basados en la verosimilitud

Desde Ecuación (R.1), obtenemos

$$\sqrt{n}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \overset{\mathsf{a}}{=} \boldsymbol{G}(\boldsymbol{\theta}_0)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n) \overset{\mathsf{a}}{=} \boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)\frac{1}{\sqrt{n}}\frac{\partial \ell_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}.$$

Por la condición de primer orden, podemos escribir

$$\frac{\partial \ell_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} = \boldsymbol{G}^\top(\widetilde{\boldsymbol{\theta}}_n)\widetilde{\boldsymbol{\lambda}}_n. \tag{R.4}$$

De este modo,

$$\begin{aligned}\sqrt{n}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) &\overset{\mathsf{a}}{=} \boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}^\top(\widetilde{\boldsymbol{\theta}}_n)\frac{\widetilde{\boldsymbol{\lambda}}_n}{\sqrt{n}}\\ &\overset{\mathsf{a}}{=} \boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}^\top(\boldsymbol{\theta}_0)\frac{\widetilde{\boldsymbol{\lambda}}_n}{\sqrt{n}}.\end{aligned}$$

Es decir,

$$\frac{\widetilde{\boldsymbol{\lambda}}_n}{\sqrt{n}} \overset{\mathsf{a}}{=} [\boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}^\top(\boldsymbol{\theta}_0)]^{-1}\sqrt{n}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n).$$

Por (W.1), sigue que

$$\frac{\widetilde{\boldsymbol{\lambda}}_n}{\sqrt{n}} \xrightarrow{\mathsf{D}} \mathsf{N}_q(\mathbf{0}, [\boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}^\top(\boldsymbol{\theta}_0)]^{-1}).$$

De ahí que la forma cuadrática

$$R_n = \frac{\widetilde{\boldsymbol{\lambda}}_n^\top}{\sqrt{n}}\boldsymbol{G}(\widetilde{\boldsymbol{\theta}}_n)\boldsymbol{\mathcal{F}}^{-1}(\widetilde{\boldsymbol{\theta}}_n)\boldsymbol{G}^\top(\widetilde{\boldsymbol{\theta}}_n)\frac{\widetilde{\boldsymbol{\lambda}}_n}{\sqrt{n}} \xrightarrow{\mathsf{D}} \chi^2(q).$$

Nuevamente por la condición en (R.4), podemos escribir:

$$R_n = \frac{1}{n}\Big(\frac{\partial \ell_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}\Big)^\top \boldsymbol{\mathcal{F}}^{-1}(\widetilde{\boldsymbol{\theta}}_n)\frac{\partial \ell_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}.$$

Resultado 3 (test de razón de verosimilitudes):[5]

El test de razón de verosimilitudes es definido por el estadístico

$$LR_n = 2\big(\ell_n(\widehat{\boldsymbol{\theta}}_n) - \ell_n(\widetilde{\boldsymbol{\theta}}_n)\big),$$

y bajo H_0 tiene región crítica asintótica de tamaño α, dada por

$$\{LR_n \geq \chi^2_{1-\alpha}(q)\}.$$

[5] Wilks (1938). The Annals of Mathematical Statistics **9**, 60-62.

Demostración:

Considere las expansiones de Taylor de $\ell_n(\widehat{\boldsymbol{\theta}}_n)$ y $\ell_n(\widetilde{\boldsymbol{\theta}}_n)$ en torno de $\boldsymbol{\theta}_0$. Bajo $H_0 : \boldsymbol{g}(\boldsymbol{\theta}_0) = \mathbf{0}$, estas expresiones son:

$$\begin{aligned}
\ell_n(\widehat{\boldsymbol{\theta}}_n) &\stackrel{\text{a}}{=} \ell_n(\boldsymbol{\theta}_0) + \Big(\frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big)^\top (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \frac{n}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
\ell_n(\widetilde{\boldsymbol{\theta}}_n) &\stackrel{\text{a}}{=} \ell_n(\boldsymbol{\theta}_0) + \Big(\frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big)^\top (\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \frac{n}{2}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).
\end{aligned}$$

Tomando diferencias, obtenemos

$$\begin{aligned}
\ell_n(\widehat{\boldsymbol{\theta}}_n) - \ell_n(\widetilde{\boldsymbol{\theta}}_n) \stackrel{\text{a}}{=} \ell_n(\boldsymbol{\theta}_0) - \ell_n(\boldsymbol{\theta}_0) + \Big(\frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big)^\top (\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n) \\
- \frac{n}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \frac{n}{2}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)
\end{aligned}$$

Es decir,

$$\begin{aligned} LR_n &= 2\big(\ell_n(\widehat{\boldsymbol{\theta}}_n) - \ell_n(\widetilde{\boldsymbol{\theta}}_n)\big) \\ &\overset{\mathsf{a}}{=} 2\Big(\frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big)^\top (\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n) - n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ &+ n(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \end{aligned} \tag{L.1}$$

Usando (R.3), tenemos

$$\frac{1}{\sqrt{n}} \frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \overset{\mathsf{a}}{=} \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).$$

Substituyendo en (L.1), sigue que

$$\begin{aligned} LR_n &\overset{\mathsf{a}}{=} 2n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n) - (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ &+ n(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ &\overset{\mathsf{a}}{=} 2n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n) - (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ &+ n(\widetilde{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n + \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)(\widetilde{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n + \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0). \end{aligned}$$

Después de simple álgebra, obtenemos

$$LR_n \stackrel{\mathsf{a}}{=} n(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n)^\top \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n)$$

Observación:

Por Ecuación (R.3), sigue que

$$LR_n \stackrel{\mathsf{a}}{=} \frac{1}{n}\Big(\frac{\partial \ell_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}\Big)^\top \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)\frac{\partial \ell_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} = R_n.$$

Es decir, LR_n es asintóticamente equivalente a R_n. De ahí que, bajo H_0 tienen la misma distribución asintótica chi-cuadrado.

Definición 1 (test gradiente):[6]

El estadístico gradiente T_n, para probar la hipótesis nula $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ es dado por

$$T_n = \boldsymbol{U}_n^\top(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

y es asintóticamente chi-cuadrado con p grados de libertad.

Demostración:

Sigue de notar que

$$\frac{1}{\sqrt{n}}\boldsymbol{U}_n(\boldsymbol{\theta}_0) \xrightarrow{\mathsf{D}} \mathsf{N}_p(\mathbf{0}, \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)), \qquad \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathsf{D}} \mathsf{N}_p(\mathbf{0}, \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)),$$

de ahí que

$$\Big\{\frac{1}{\sqrt{n}}\boldsymbol{\mathcal{F}}^{-1/2}(\boldsymbol{\theta}_0)\boldsymbol{U}_n(\boldsymbol{\theta}_0)\Big\}^\top \sqrt{n}\boldsymbol{\mathcal{F}}^{1/2}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathsf{D}} \chi^2(p).$$

[6] Terrell (2002). Computing Sciences and Statistics **34**, 206-215.

Resultado 4 (test de forma bilineal):[7]

El estadístico de forma bilineal BF_n, para probar la hipótesis $H_0 : \boldsymbol{g}(\boldsymbol{\theta}_0) = \mathbf{0}$, adopta la forma

$$BF_n = \widetilde{\boldsymbol{\lambda}}_n^\top \boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n),$$

y bajo H_0, BF_n tiene una distribución asintótica chi-cuadrado con q grados de libertad.

Demostración:

Sabemos que

$$\frac{\widetilde{\boldsymbol{\lambda}}_n}{\sqrt{n}} \xrightarrow{\mathsf{D}} \mathsf{N}_q\big(\mathbf{0}, [\boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}^\top(\boldsymbol{\theta}_0)]^{-1}\big),$$

$$\sqrt{n}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \xrightarrow{\mathsf{D}} \mathsf{N}_q\big(\mathbf{0}, \boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}^\top(\boldsymbol{\theta}_0)\big),$$

y el resultado sigue.

[7] Crudu y Osorio (2020). Economics Letters **187**, 108885.

Observación:

Podemos escribir el estadístico BF_n de forma equivalente como

$$BF_n = \boldsymbol{U}_n^\top(\widetilde{\boldsymbol{\lambda}}_n)\boldsymbol{G}^+\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n),$$

donde $\boldsymbol{G}^+ = \boldsymbol{G}^\top(\boldsymbol{G}\boldsymbol{G}^\top)^{-1}$ denota la inversa Moore-Penrose de $\boldsymbol{G} = \boldsymbol{G}(\widetilde{\boldsymbol{\theta}}_n)$. Por (R.1), tenemos que

$$BF_n \stackrel{\text{a}}{=} \boldsymbol{U}_n^\top(\widetilde{\boldsymbol{\theta}}_n)\boldsymbol{G}^+\boldsymbol{G}(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n).$$

Además,

$$BF_n \stackrel{\text{a}}{=} \frac{1}{n}\widetilde{\boldsymbol{\lambda}}_n^\top\boldsymbol{G}(\widetilde{\boldsymbol{\theta}}_n)\boldsymbol{\mathcal{F}}^{-1}(\widetilde{\boldsymbol{\theta}}_n)\boldsymbol{G}^\top(\widetilde{\boldsymbol{\theta}}_n)\widetilde{\boldsymbol{\lambda}}_n = R_n.$$

Referencias bibliográficas

Crudu, F., Osorio, F. (2020).
Bilinear form test statistics for extremum estimation.
Economics Letters **187**, 108885.

Rao, C.R. (1948).
Large sample tests of Statistical hypotheses concerning several parameters with applications to problems of estimation.
Proceedings of the Cambridge Philosophical Society **44**, 50-57.

Terrell, G.R. (2002).
The gradient statistic.
Computing Sciences and Statistics **34**, 206-215.

Wald, A. (1943).
Test of Statistical hypothesis concerning several parameters when the number of observations is large.
Transactions of the American Mathematical Society **54**, 426-482.

Wilks, S.S. (1938).
The large-sample distribution of the likelihood ratio for testing composite hypothesis.
The Annals of Mathematical Statistics **9**, 60-62.