MAT-206: Test asintóticos para extremum estimation

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Test en el contexto de extremum estimation

Problema 1:

Considere la hipótesis simple $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ con $\boldsymbol{\theta}_0$ fijado. En el contexto de extreme estimation, deseamos resolver

$$\max_{\theta \in \Theta} Q_n(\boldsymbol{\theta}) \qquad \text{sujeto a:} \quad \boldsymbol{\theta} = \boldsymbol{\theta}_0.$$

Problema 2:

Deseamos probar hipótesis no lineales de la forma $H_0 : \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}$, 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, deseamos resolver el problema restringido:

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

Supuestos

B1: $Q_n(\boldsymbol{\theta})$ es dos veces diferenciable en $\boldsymbol{\theta}$.

B2: Sea $\boldsymbol{A}_n(\boldsymbol{\theta}) = \partial^2 Q_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ y asuma que

$$\boldsymbol{A}_n(\boldsymbol{\theta}) \xrightarrow{\text{a.s}} \boldsymbol{A},$$

uniformemente.

B3: La matriz $\boldsymbol{A}$ es no singular.

B4: La secuencia $\{\sqrt{n}(\partial Q_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta})\}$ converge en distribución a una normal estándar con vector de medias cero y matriz de covarianza $\boldsymbol{B}$. Es decir,

$$\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \xrightarrow{\text{D}} \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{B}).$$

Resultado 1 (test de Wald):

El test de Wald (asintóticamente) de tamaño α para probar la hipótesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, adopta la forma:

Rechazar H_0 si,

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

donde

$$W_n = n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{A}\boldsymbol{B}^{-1}\boldsymbol{A}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

con $\chi^2_{1-\alpha}(p)$ el valor cuantil $(1-\alpha)$ de la distribución chi-cuadrado con p grados de libertad.

Demostración:

Sabemos que

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

luego

$$\boldsymbol{Z}_n = \sqrt{n}(\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1})^{-1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathsf{D}} \mathsf{N}_p(\mathbf{0}, \boldsymbol{I}),$$

como $(\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1})^{-1/2}(\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1})^{-1/2} = \boldsymbol{A}\boldsymbol{B}^{-1}\boldsymbol{A}$, sigue que

$$\boldsymbol{Z}_n^\top \boldsymbol{Z}_n = n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \boldsymbol{A}\boldsymbol{B}^{-1}\boldsymbol{A}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathsf{D}} \chi^2(p).$$

Resultado 2 (test tipo-score):[1]

El test tipo-score para probar $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, es definido por la región crítica,

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

donde

$$R_n = n\Big(\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big)^\top \boldsymbol{B}^{-1} \, \frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}},$$

y bajo H_0 es asintóticamente de tamaño α.

[1] Boos (1992). The American Statistician **46**, 327-333.

Demostración:

Por el supuesto B4, tenemos que

$$\sqrt{n}\,\boldsymbol{B}^{-1/2}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{\mathsf{D}} \mathsf{N}_p(\mathbf{0}, \boldsymbol{I}).$$

luego

$$R_n = n\Big(\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big)^\top \boldsymbol{B}^{-1/2}\boldsymbol{B}^{-1/2}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{\mathsf{D}} \chi^2(p).$$

Resultado 3 (test de forma bilineal):[2]

El estadístico de forma bilineal para probar la hipotesis nula $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, asume la forma:

$$BF_n = n\Big(\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big)^\top \boldsymbol{B}^{-1}\boldsymbol{A}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

y bajo H_0, BF_n tiene distribución asintótica $\chi^2(p)$.

Observación:

Asumiendo que $\boldsymbol{B} = -\boldsymbol{A}$, sigue que

$$BF = n\Big(\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big)^\top (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

cuando $Q_n(\boldsymbol{\theta}) = \overline{\ell}_n(\boldsymbol{\theta})$, obtenemos la estadística gradiente dada por Terrell (2002).[3]

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

[3] Computing Sciences and Statistics **34**, 206-215.

Demostración:

Considere $\boldsymbol{B} = \boldsymbol{R}\boldsymbol{R}^\top$. De este modo,

$$\sqrt{n}\,\boldsymbol{R}^{-1}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{\mathsf{D}} \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{I}),$$

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

Basado en estas versiones estandarizadas, tenemos

$$BF_n = \Big\{\sqrt{n}\,\boldsymbol{R}^{-1}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big\}^\top \{\sqrt{n}\,\boldsymbol{R}^{-1}\boldsymbol{A}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\},$$

y de este manera, sigue que $BF_n \xrightarrow{\mathsf{D}} \chi^2(p)$.

Ejemplo (Una función objetivo cuadrática):

Sea

$$Q_n(\boldsymbol{\theta}) = -\tfrac{1}{2}\boldsymbol{f}_n^\top(\boldsymbol{\theta})\boldsymbol{W}^{-1}\boldsymbol{f}_n(\boldsymbol{\theta}),$$

donde $\boldsymbol{f}_n(\boldsymbol{\theta})$ es un conjunto de condiciones de momentos y $\boldsymbol{W}$ es una matriz definida positiva con órdenes apropiados, entonces

$$\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\boldsymbol{F}_n^\top(\boldsymbol{\theta})\boldsymbol{W}^{-1}\boldsymbol{f}_n(\boldsymbol{\theta}),$$

con $\boldsymbol{F}_n(\boldsymbol{\theta}) = \partial \boldsymbol{f}_n(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^\top$, y

$$\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} = -\Big[\frac{\partial \boldsymbol{F}_n^\top(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big]\big[\boldsymbol{W}^{-1}\boldsymbol{f}_n(\boldsymbol{\theta})\big] - \boldsymbol{F}_n^\top(\boldsymbol{\theta})\boldsymbol{W}^{-1}\boldsymbol{F}_n(\boldsymbol{\theta}),$$

donde $[\cdot][\cdot]$ denota multiplicación entre arreglos. Si $\sqrt{n}\,\boldsymbol{f}_n(\boldsymbol{\theta}_0) \xrightarrow{\mathsf{D}} \mathsf{N}(\boldsymbol{0}, \boldsymbol{W})$, entonces:

$$\boldsymbol{B} = \boldsymbol{F}^\top\boldsymbol{W}^{-1}\boldsymbol{F}, \qquad \boldsymbol{A} = -\boldsymbol{F}^\top\boldsymbol{W}^{-1}\boldsymbol{F}.$$

De ahí que, $\boldsymbol{B} = -\boldsymbol{A}$ se satisface.

Propiedad 1:

Tenemos que

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \overset{\mathsf{a}}{=} \boldsymbol{A}^{-1}\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}.$$

De ahí que, bajo $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ podemos apreciar

$$\begin{aligned} BF_n &= \sqrt{n}\Big(\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big)^\top \boldsymbol{B}^{-1}\boldsymbol{A}\{\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\} \\ &\overset{\mathsf{a}}{=} \sqrt{n}\Big(\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big)^\top \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{A}^{-1}\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \\ &\overset{\mathsf{a}}{=} n\Big(\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big)^\top \boldsymbol{B}^{-1}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \equiv R_n \end{aligned}$$

donde R_n corresponde al estadistico tipo-score para probar $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ en extreme estimation (ver Rotnitzky and Jewell, 1990[4] y Boos, 1992).

[4] Biometrika **77**, 485-497.

Sea $\{\widetilde{\boldsymbol{\theta}}_n\}$ una secuencia de estimadores que son solución del problema

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

El estimador restringido $\widetilde{\boldsymbol{\theta}}_n$ satisface las condiciones de primer-orden:

$$\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} - \boldsymbol{G}(\widetilde{\boldsymbol{\theta}}_n)^\top \widetilde{\boldsymbol{\lambda}}_n = \mathbf{0}$$

$$\boldsymbol{g}(\widetilde{\boldsymbol{\theta}}_n) = \mathbf{0},$$

donde $\widetilde{\boldsymbol{\lambda}}_n$ es una secuencia de multiplicadores de Lagrange.

Objetivo:

Deseamos obtener estadísticos de prueba para test de hipótesis no lineales, de la forma:

$$H_0 : \boldsymbol{g}(\boldsymbol{\theta}_0) = \mathbf{0} \qquad \text{against} \qquad H_1 : \boldsymbol{g}(\boldsymbol{\theta}_0) \neq \mathbf{0}.$$

Sabemos que, bajo H_0,

$$\sqrt{n}\, \boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \xrightarrow{\mathsf{D}} \mathsf{N}_q(\mathbf{0}, \boldsymbol{\Omega}), \tag{A.1}$$

donde $\boldsymbol{\Omega} = \boldsymbol{G}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{G}^\top$, y

$$\boldsymbol{G} = \frac{\partial \boldsymbol{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\Big|_{\theta=\theta_0}.$$

Notando que

$$\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \overset{\mathsf{a}}{=} \boldsymbol{G}(-\boldsymbol{A})^{-1}\sqrt{n}\,\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}$$

y desde la condición de primer orden

$$\sqrt{n}\,\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \overset{\mathsf{a}}{=} \boldsymbol{G}^\top(\widetilde{\boldsymbol{\theta}}_n)\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n.$$

Obtenemos que

$$\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n \overset{\mathsf{a}}{=} [\boldsymbol{G}(-\boldsymbol{A})^{-1}\boldsymbol{G}^\top]^{-1}\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n). \qquad \text{(A.2)}$$

Entonces, usando la distribución asintótica dada en (A.1), hallamos

$$\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n \xrightarrow{\mathsf{D}} \mathsf{N}_q(\mathbf{0}, \boldsymbol{S}^{-1}\boldsymbol{\Omega}\boldsymbol{S}^{-1}),$$

donde $\boldsymbol{S} = \boldsymbol{G}(-\boldsymbol{A})^{-1}\boldsymbol{G}^\top$.

Resultado 4 (test de Wald):

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

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

donde

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

y bajo H_0 es asintóticamente de tamaño α.

Demostración:

Basta notar que por Ecuación (A.1),

$$W_n = \{\sqrt{n}\,\widehat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{g}^\top(\widehat{\boldsymbol{\theta}}_n)\}^\top\sqrt{n}\,\widehat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{g}^\top(\widehat{\boldsymbol{\theta}}_n) \xrightarrow{\mathsf{D}} \chi^2(q).$$

Resultado 5 (test tipo-score o de multiplicadores de Lagrange):

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

$$R_n = n\widetilde{\boldsymbol{\lambda}}_n^\top \widetilde{\boldsymbol{S}}\widetilde{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{S}}\widetilde{\boldsymbol{\lambda}}_n,$$

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

Demostración:

Por la Ecuación (A.2), tenemos

$$R_n = \{\sqrt{n}\,(\widetilde{\boldsymbol{S}}^{-1}\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{S}}^{-1})^{-1/2}\widetilde{\boldsymbol{\lambda}}_n\}^\top \sqrt{n}\,(\widetilde{\boldsymbol{S}}^{-1}\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{S}}^{-1})^{-1/2}\widetilde{\boldsymbol{\lambda}}_n \xrightarrow{\mathsf{D}} \chi^2(q).$$

Resultado 6 (test de forma bilineal)[5]

El estadístico de forma bilineal para probar la hipótesis nula $H_0 : \boldsymbol{g}(\boldsymbol{\theta}_0) = \mathbf{0}$ es dada por

$$BF_{n,1} = n\, \widetilde{\boldsymbol{\lambda}}_n^\top \boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n),$$

y bajo H_0, BF_1 tiene una distribución chi-cuadrado con q grados de libertad.

Demostración:

Sea $\boldsymbol{\Omega} = \boldsymbol{R}\boldsymbol{R}^\top$ donde $\boldsymbol{R}$ es una matriz no singular $q \times q$. De este modo,

$$\sqrt{n}\,\boldsymbol{R}^{-1}\boldsymbol{S}\widetilde{\boldsymbol{\lambda}}_n \xrightarrow{\mathsf{D}} \mathsf{N}_q(\mathbf{0}, \boldsymbol{I}), \qquad \sqrt{n}\,\boldsymbol{R}^{-1}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \xrightarrow{\mathsf{D}} \mathsf{N}_q(\mathbf{0}, \boldsymbol{I}),$$

lo que permite establecer el resultado.

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

Resultado 7

El estadístico de forma bilineal tiene las siguientes formulaciones alternativas. Sea $\boldsymbol{G}^{+} = \boldsymbol{G}^{\top}(\boldsymbol{G}\boldsymbol{G}^{\top})^{-1}$ la inversa Moore-Penrose de $\boldsymbol{G}$. Entonces,

$$BF_{n,2} = n\Big(\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}\Big)^{\top}\boldsymbol{G}^{+}\boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n),$$

$$BF_{n,3} = n\Big(\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}\Big)^{\top}\boldsymbol{G}^{+}\boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{G}(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n),$$

y bajo H_0, $BF_k \xrightarrow{\mathsf{D}} \chi^2(q)$, $k = 2, 3$.

Observación:

Defina $\boldsymbol{P} = \boldsymbol{G}^{+}\boldsymbol{G}$ y asuma que $\boldsymbol{B} = -\boldsymbol{A}$, lo que lleva a $\boldsymbol{S} = \boldsymbol{\Omega}$. Obtenemos las siguientes especificaciones

$$
\begin{aligned}
BF_{n,4} &= n\,\widetilde{\boldsymbol{\lambda}}_n^{\top}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n),\\
BF_{n,5} &= n\Big(\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}\Big)^{\top}\boldsymbol{G}^{+}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n),\\
BF_{n,6} &= n\Big(\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}\Big)^{\top}\boldsymbol{P}(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n),\\
BF_{n,7} &= n\Big(\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}\Big)^{\top}(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n),
\end{aligned}
$$

y $BF_k \xrightarrow{\mathsf{D}} \chi^2(q)$, $k = 4, 5, 6, 7$, bajo $H_0 : \boldsymbol{g}(\boldsymbol{\theta}_0) = \boldsymbol{0}$.

Observación:

Cuando $Q_n(\boldsymbol{\theta}) = \bar{\ell}_n(\boldsymbol{\theta})$ es la función de log-verosimilitud obtenemos que la estadística BF_n es dada por

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

donde $\boldsymbol{U}_n(\widetilde{\boldsymbol{\theta}}) = \partial\bar{\ell}_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ es la función score.

Observación:

Sea

$$D_n = n\big(Q_n(\widetilde{\boldsymbol{\theta}}) - Q_n(\widehat{\boldsymbol{\theta}})\big),$$

la estadística de distancia metrica. Debemos resaltar que para el modelo lineal $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$, D_n y BF_n son iguales. Más aún, el estadístico BF_n asume la forma

$$BF_n = (\boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\theta}}_n)^\top \boldsymbol{X}\boldsymbol{B}^{-1}\boldsymbol{X}^\top\boldsymbol{X}(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_n).$$

Propiedad 2:

El estadístico $BF_{n,1}$ y el estadístico de multiplicadores de Lagrange son asintóticamente equivalentes y bajo $H_0 : \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}$ su distribución asintótica común es $\chi^2(q)$.

Demostración:

Desde (A.2), tenemos que

$$\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \stackrel{\text{a}}{=} \boldsymbol{G}(-\boldsymbol{A})^{-1}\boldsymbol{G}^\top \sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n = \boldsymbol{S}\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n,$$

y esto lleva a,

$$\begin{aligned} BF_n &= \sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n \boldsymbol{S}\boldsymbol{\Omega}^{-1}\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \\ &\stackrel{\text{a}}{=} \sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n \boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{S}\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n \equiv R_n. \end{aligned}$$

Basado en el estudio de simulación introducido por Gregory y Veall (1985)[6].

Se generó $M = 5\,000$ conjuntos de datos de tamaños $n = 20, 50, 100, 500$ considerando la siguiente especificación de modelo:

$$\boldsymbol{Y} = \boldsymbol{1}_n\theta_1 + \boldsymbol{x}_2\theta_2 + \exp(\boldsymbol{x}_3\theta_3) + \boldsymbol{\epsilon},$$

donde $\boldsymbol{x}_j \sim \mathsf{N}_n(\boldsymbol{0}, 0.16\boldsymbol{I})$, $j = 2, 3$ y $\boldsymbol{\epsilon} \sim \mathsf{N}_n(\boldsymbol{0}, 0.16\boldsymbol{I})$.

Deseamos probar dos hipótesis nulas equivalentes

$$H_0^A : \theta_2 - \frac{1}{\theta_3} = 0, \qquad H_0^B : \theta_2\theta_3 - 1 = 0.$$

usandos las estadísticas de forma bilineal (BF), Wald (W), multiplicadores de Langrange (R) y distancia metrica (D).

Corrimos nuestras simulaciones en un servidor HP Proliant DL360, procesador Intel Xeon E5-2630 con 192 GB de RAM (tiempo total de simulación: 23 hrs, 40 min, 3 seg).[7]

---

[6] Econometrica **53**, 1465-1468.

[7] código R diponible en: `https://github.com/faosorios/BF_EE`

Resultados: Tamaños empíricos para un test del 5%[8]

Escenario	$(\theta_1, \theta_2, \theta_3)$	n	W^A	W^B	BF^A	BF^B	R	D
I	(1,10,0.1)	20	0.420	0.176	0.067	0.064	0.084	0.087
		50	0.282	0.106	0.065	0.059	0.068	0.074
		100	0.197	0.077	0.059	0.059	0.061	0.061
		500	0.104	0.052	0.048	0.048	0.049	0.049
II	(1,5,0.2)	20	0.277	0.178	0.068	0.065	0.083	0.086
		50	0.171	0.108	0.058	0.057	0.067	0.068
		100	0.127	0.078	0.058	0.058	0.061	0.062
		500	0.070	0.052	0.047	0.047	0.048	0.048
III	(1,2,0.5)	20	0.145	0.175	0.066	0.067	0.082	0.082
		50	0.096	0.113	0.062	0.057	0.070	0.075
		100	0.078	0.082	0.056	0.056	0.062	0.062
		500	0.049	0.055	0.045	0.045	0.050	0.050
IV	(1,1,1)	20	0.140	0.170	0.084	0.070	0.086	0.101
		50	0.095	0.108	0.062	0.062	0.070	0.070
		100	0.074	0.080	0.066	0.066	0.067	0.067
		500	0.055	0.055	0.056	0.056	0.061	0.061

[8] Los superíndices A y B refieren al hecho de que W y BF son calculados usando las hipótesis nulas H_0^A y H_0^B

Se condujo en estudio de simulación adicional basado en $M = 1\,000$ conjuntos de datos con tamaños muestrales $n = 20, 50, 100$, calculando la potencia empírica bajo las hipótesis alternativas:

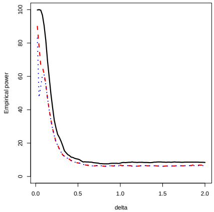
$$H_1^A : \beta_2 - \delta/\beta_3 = 0,$$

y

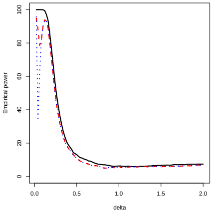
$$H_1^B : \beta_2\beta_3 - \delta = 0,$$

considerando $\delta \in [0, 2]$ para cada uno de los escenarios.[9]

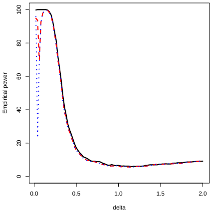
[9] Sólo las potencias empíricas de R, BF^A y BF^B son reportadas.



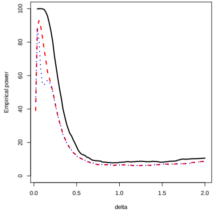
(a) $n = 20$



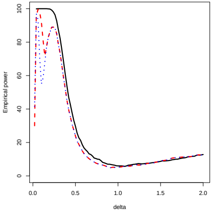
(b) $n = 50$



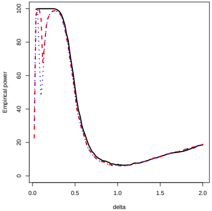
(c) $n = 100$



(d) $n = 20$

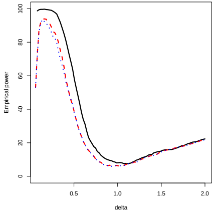


(e) $n = 50$

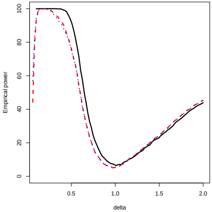


(f) $n = 100$

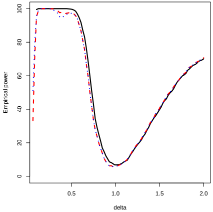
Resultados: Potencias empíricas para R, BF^A y BF^B, escenarios III y IV



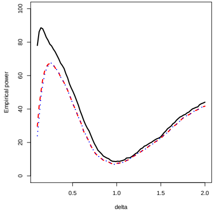
(a) $n = 20$



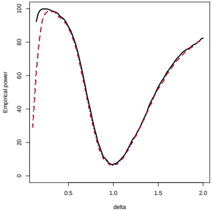
(b) $n = 50$



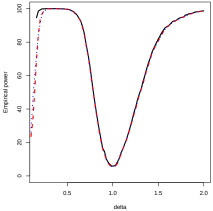
(c) $n = 100$



(d) $n = 20$



(e) $n = 50$



(f) $n = 100$

Probando la igualdad de razones de Sharpe

Considere $\boldsymbol{Y} = (Y_1, \dots, Y_p)^\top$ un vector aleatorio p-dimensional con vector de medias $\boldsymbol{\mu} \in \mathbb{R}^p$ y matriz de covarianza $\boldsymbol{\Sigma}$. Para evaluar una estrategia de inversión, Sharpe (1966)[10] propuso la razón:

$$\lambda_j = \frac{\mu_j}{\sqrt{\sigma_{jj}}}, \qquad j = 1, \dots, p.$$

Sea $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)^\top$ el vector de razones de Sharpe para un portfolio con p inversiones. Deseamos probar la hipótesis:[11]

$$H_0 : \lambda_1 = \lambda_2 = \cdots = \lambda_p \qquad (= \lambda)$$

[10] Journal of Business **39**, 119-138.

[11] Trabajo en colaboración con Manuel Galea (PUC, Chile).

Considere una muestra aleatoria $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$ from $\mathsf{T}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$ con densidad

$$f(\boldsymbol{y}) = \Big(\frac{c(\eta)}{\pi}\Big)^{p/2} \frac{\Gamma(\frac{1+\eta p}{2\eta})}{\Gamma(\frac{1}{2\eta})} \big(1 + c(\eta) D^2\big)^{-\frac{1}{2}(1+\eta p)}, \qquad 0 < \eta < 1/2,$$

donde $c(\eta) = \eta/(1-2\eta)$ y $D^2 = (\boldsymbol{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu})$ denota la distancia de Mahalanobis.

Note que, bajo H_0 tenemos $\mu_j = \lambda\sqrt{\sigma_{jj}}$, $j = 1, \dots, p$. Esto es, podemos escribir:

$$\boldsymbol{Y}_i | \omega_i \overset{\text{ind}}{\sim} \mathsf{N}_p(\lambda \boldsymbol{\sigma}^{1/2}, \boldsymbol{\Sigma}), \qquad \omega_i \overset{\text{ind}}{\sim} \mathsf{Gamma}\Big(\frac{1}{2\eta}, \frac{1}{2c(\eta)}\Big),$$

para $i = 1, \dots, n$. Además $\boldsymbol{\Sigma} = \boldsymbol{D}^{-1/2} \boldsymbol{\Omega} \boldsymbol{D}^{-1/2}$ con $\boldsymbol{D} = \operatorname{diag}(\boldsymbol{\sigma})$.

Probando la igualdad de razones de Sharpe

- Usamos un algoritmo ECM (Meng and Rubin, 1993)[12] para obtener estimadores ML restringidos de $\boldsymbol{\theta} = (\lambda, \boldsymbol{\sigma}^\top, \boldsymbol{\rho}^\top, \eta)^\top$, con $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}(\boldsymbol{\rho})$.

- Para eficiencia computacional, la siguiente reparametrización fue considerada

$$\begin{aligned} \phi_j &= \log \sigma_{jj}, \qquad j = 1, \dots, p, \\ \boldsymbol{\Gamma}^{-1} &= \boldsymbol{\Delta}^\top \boldsymbol{\Delta}, \end{aligned}$$

  con $\boldsymbol{\Delta}$ matriz triangular inferior.

- De este modo, podemos escribir:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = Q_1(\boldsymbol{\tau}|\boldsymbol{\theta}^{(k)}) + Q_2(\eta|\boldsymbol{\theta}^{(k)}),$$

  donde $\boldsymbol{\tau} = (\lambda, \boldsymbol{\phi}^\top, \boldsymbol{\rho}^\top)^\top$.

[12] Biometrika **80**, 267-278.

- Paso E: Para una estimación inicial $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, calcular:

$$\begin{aligned} Q_1(\boldsymbol{\tau}|\boldsymbol{\theta}^{(k)}) &= -\frac{n}{2}\operatorname{tr}\boldsymbol{\Phi} + n\log|\boldsymbol{\Delta}| \\ &\quad - \frac{1}{2}\sum_{i=1}^{n}\omega_i^{(k)}\{\boldsymbol{\Delta}(e^{-\boldsymbol{\Phi}/2}\boldsymbol{y}_i - \lambda\mathbf{1})\}^\top\boldsymbol{\Delta}(e^{-\boldsymbol{\Phi}/2}\boldsymbol{y}_i - \lambda\mathbf{1}) \end{aligned}$$

y

$$\begin{aligned} Q_2(\eta|\boldsymbol{\theta}^{(k)}) = n\Big\{&\frac{1}{2\eta}\log\Big(\frac{1}{2c(\eta)}\Big) - \log\Gamma\Big(\frac{1}{2\eta}\Big) + \frac{1}{2c(\eta)}\Big[\psi\Big(\frac{1/\eta^{(k)}+p}{2}\Big) \\ &- \log\Big(\frac{1/\eta^{(k)}+p}{2}\Big) + \frac{1}{n}\sum_{i=1}^{n}(\log\omega_i^{(k)} - \omega_i^{(k)})\Big]\Big\}, \end{aligned}$$

donde

$$\omega_i^{(k)} = \frac{1/\eta^{(k)} + p}{1/c(\eta^{(k)}) + D_i^2(\lambda^{(k)}, \boldsymbol{\phi}^{(k)}, \boldsymbol{\rho}^{(k)})},$$

con[13]

$$D_i^2(\lambda, \boldsymbol{\phi}, \boldsymbol{\rho}) = \|\boldsymbol{z}_i\|^2, \qquad \boldsymbol{z}_i = \boldsymbol{\Delta}(e^{-\boldsymbol{\Phi}/2}\boldsymbol{y}_i - \lambda\mathbf{1}), \qquad i = 1, \dots, n.$$

[13] Note que $\boldsymbol{z}_i$ es solución de $\boldsymbol{L}\boldsymbol{z}_i = e^{-\boldsymbol{\Phi}/2}\boldsymbol{y}_i - \lambda\mathbf{1}$, con $\boldsymbol{\Gamma} = \boldsymbol{L}\boldsymbol{L}^\top$.

Probando la igualdad de razones de Sharpe

- Paso CM-1: Para $\boldsymbol{\phi} = \boldsymbol{\phi}^{(k)}$ y $\boldsymbol{\rho} = \boldsymbol{\rho}^{(k)}$, actualizar:[14]

$$\lambda^{(k+1)} = \frac{1}{\mathbf{1}^\top \mathbf{\Delta} \mathbf{1}} \mathbf{1}^\top \mathbf{\Delta} e^{-\mathbf{\Phi}/2} \boldsymbol{\mu}^{(k+1)}, \qquad \boldsymbol{\mu}^{(k+1)} = \frac{1}{\sum_{i=1}^n \omega_i^{(k)}} \sum_{i=1}^n \omega_i^{(k)} \boldsymbol{y}_i.$$

- Paso CM-2: Calcular:

$$\mathbf{\Sigma}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \omega_i^{(k)} (\boldsymbol{y}_i - \boldsymbol{\mu}^{(k+1)})(\boldsymbol{y}_i - \boldsymbol{\mu}^{(k+1)})^\top = \frac{1}{n} \sum_{i=1}^n \omega_i^{(k)} \boldsymbol{z}_i \boldsymbol{z}_i^\top.$$

  Para $\boldsymbol{\phi} = \boldsymbol{\phi}^{(k)}$ y $\lambda = \lambda^{(k+1)}$, sea $\boldsymbol{D}^{-1/2} \mathbf{\Sigma}^{(k+1)} \boldsymbol{D}^{-1/2} = \mathbf{\Gamma}^{(k+1)}$. Usando la condición de primer orden, tenemos

$$\mathbf{\Delta}^{-1} - \mathbf{\Gamma} \mathbf{\Delta}^\top = \mathbf{0},$$

  esto es, $\mathbf{\Gamma}^{(k+1)}$ es la solución de la ecuación:

$$\mathbf{\Gamma}^{-1} - \mathbf{\Delta}^\top \mathbf{\Delta} = \mathbf{0}.$$

[14] Note que, $\boldsymbol{r}^{(k+1)} = e^{-\mathbf{\Phi}/2} \boldsymbol{\mu}^{(k+1)}$ es el vector de razones de Sharpe estimadas.

- Paso CM-3: Para actualizar $\boldsymbol{\phi}^{(k+1)}$ consideramos el método secante multivariado (Nash, 1990), usando BFGS maximizamos

$$f(\boldsymbol{\phi}) = -Q_1(\lambda^{(k+1)}, \boldsymbol{\phi}, \boldsymbol{\rho}^{(k+1)}|\boldsymbol{\theta}^{(k)}),$$

donde

$$f(\boldsymbol{\phi}) = \frac{n}{2}\operatorname{tr}\boldsymbol{\Phi} - n\log|\boldsymbol{\Delta}| + \frac{1}{2}\sum_{i=1}^{n}\omega_i^{(k)}\|\boldsymbol{z}_i\|^2$$

$$\dot{f}(\boldsymbol{\phi}) = \mathbf{1} + \frac{1}{n}\sum_{i=1}^{n}\omega_i^{(k)}\operatorname{diag}(\boldsymbol{y}_i\boldsymbol{z}_i^\top\boldsymbol{\Delta}^\top\boldsymbol{\Delta}e^{-\boldsymbol{\Phi}/2})$$

- Paso CM-4: actualizar el parámetro de forma maximizando $Q_2(\eta|\boldsymbol{\theta}^{(k)})$ con relación a η, esto es,

$$\eta^{(k+1)} = \arg\max_{\eta} Q_2(\eta|\boldsymbol{\theta}^{(k)}),$$

usando un procedimiento de optimización unidimensional (Brent, 1973).

Rentabilidades de administradoras de fondos de pensiones (AFP) chilenas

Aplicación:

El sistema de pensiones Chileno (o sistema de capitalización individual) ha estado en efecto desde 1980.

AFPs son compañías privadas que invierten los fondos de los cotizantes en un portfolio de títulos de inversión.

Existe 5 tipos de fondos (A, B, C, D, E), diferenciados por la proporción que es invertida en instrumentos de renta variable.

Fondo A representa un portfolio de alto riesgo. El riesgo disminuye progresivamente para los fondos B, C, D y E.

Datos:

Retornos mensuales de AFPs: Cuprum, Habitat, PlanVital y ProVida entre Agosto 2005 y Diciembre 2013.

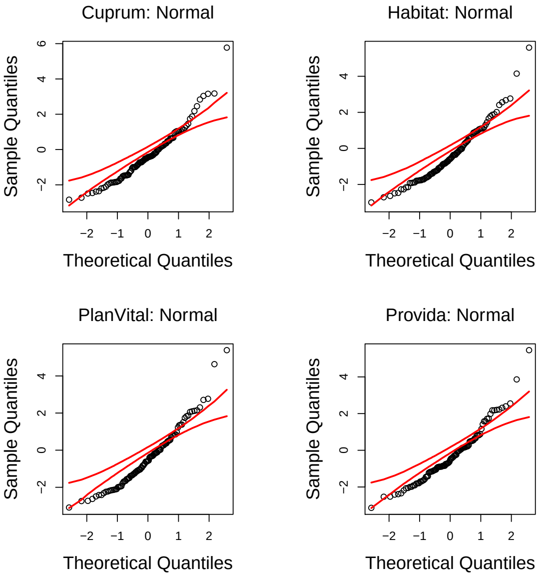
Datos fueron obtenidos desde el sitio web de la superintendencia de pensiones (www.spensiones.cl)

Conjunto de datos con 101 observaciones y 4 variables.

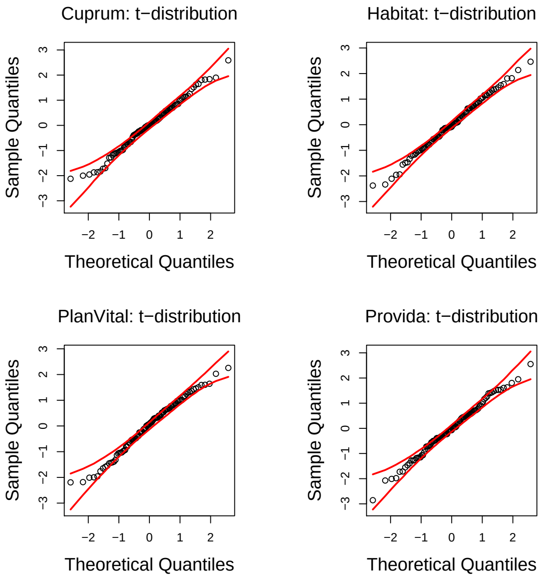
Obs. 28, 30, 35, 38, 39, 42, 66, 73 y 75 son identificadas como outliers.

QQ-plot de distancias transformadas resalta la presencia de colas más pesadas que la normal.

QQ-plots bajo el modelo normal: datos de AFP



QQ-plots bajo el modelo t multivariado: datos de AFP



Estimaciones de las razones de Sharpe: datos de AFP

Estimaciones de las razones de Sharpe $\widehat{\lambda}_i = \widehat{\mu}_i / \widehat{\sigma}_{ii}^{1/2}$, $i = 1, \dots, p$, asumiendo modelos normal y t multivariado.

- Modelo normal:

AFP	Fondo				
	A	B	C	D	E
Cuprum	0.083	0.106	0.168	0.253	0.309
Habitat	0.095	0.122	0.181	0.291	0.355
Planvital	0.093	0.118	0.173	0.260	0.325
Provida	0.085	0.102	0.150	0.253	0.300

- Modelo t multivariado:

AFP	Fondo				
	A	B	C	D	E
Cuprum	0.164	0.182	0.233	0.324	0.387
Habitat	0.176	0.195	0.240	0.336	0.385
Planvital	0.177	0.194	0.231	0.302	0.334
Provida	0.187	0.199	0.244	0.347	0.367

Probando $H_0 : \lambda_1 = \cdots = \lambda_p$, asumiendo datos normales.

	Cuprum		Habitat	
Test	estadistica	valor-p	estadistica	valor-p
LRT	21.123	0.000	19.598	0.000
Wald	11.934	0.018	10.448	0.033
BF	23.332	0.000	21.499	0.000

	PlanVital		Provida	
Test	estadistica	valor-p	estadistica	valor-p
LRT	13.896	0.008	19.666	0.001
Wald	8.063	0.089	10.485	0.033
BF	14.852	0.005	21.581	0.000

Probando $H_0 : \lambda_1 = \cdots = \lambda_p$, asumiendo datos t multivariados.

	Cuprum		Habitat	
Test	estadistica	valor-p	estadistica	valor-p
LRT	51.578	0.000	54.781	0.000
Wald	13.427	0.009	13.071	0.011
BF	52.625	0.000	57.899	0.000

	PlanVital		Provida	
Test	estadistica	valor-p	estadistica	valor-p
LRT	37.799	0.000	48.544	0.000
Wald	8.614	0.072	14.312	0.006
BF	39.302	0.000	50.659	0.000

Otros tópicos de interés

- Los resultados vistos en esta sesión pueden ser usado para probar hipótesis en modelos mal especificados.

- Existe muy poca literatura sobre test de hipótesis para funciones de inferencia (FI). (i.e., cuando una función objetivo $Q_n(\boldsymbol{\theta})$ no es disponible). Estamos trabajando en extender el estadístico de forma bilineal para FI[15]

- Test alternativos están basados en bootstrap, rangos y permutaciones.

- Test de hipótesis también son usados como ingredientes para realizar análisis de diagnóstico así como definir residuos en modelos más generales (usando, por ejemplo, el modelo de salto en la media).

[15] Trabajo conjunto con Cibele Russo (ICMC-USP, Brazil).

Referencias

Boos, D.D. (1992).
On generalized score tests.
The American Statistician **46**, 327-333.

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

Dagenais, M.G., Dufour, J.M. (1991).
Invariance, nonlinear models, and asymptotic tests.
Econometrica **59**, 1601-1615.

Gourieroux, C., and Monfort, A. (1995).
Statistics and Econometric Models: Testing, confidence regions, model selection, and asymptotic theory.
Cambridge University Press, Cambridge.

Gregory, A.W., Veall, M.R. (1985).
Formulating Wald tests of nonlinear restrictions.
Econometrica **53**, 1465-1468.

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