La trilogía de los contrastes: LR, Wald y LM

Gabriel V. Montes-Rojas

Hipótesis

- Supongamos que estamos interesados en contrastar la hipótesis nula
  $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$
  contra la hipótesis alternativa
  $H_1 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_1$.
- Para el caso univariado tenemos
  - $H_0 : \theta \leq \theta_0$ y $H_1 : \theta > \theta_0$ tal que $\Theta_0 = (-\infty, \theta_0]$ y $\Theta_1 = (\theta_0, \infty)$;
  - $H_0 : \theta \geq \theta_0$ y $H_1 : \theta < \theta_0$ tal que $\Theta_0 = [\theta_0, \infty)$ y $\Theta_1 = (-\infty, \theta_0)$;
  - $H_0 : \theta = \theta_0$ y $H_1 : \theta \neq \theta_0$ tal que $\Theta_0 = \theta_0$ y $\Theta_1 = \mathbb{R}/\theta_0$.
- $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1$.
- Un test o contraste es una función $\boldsymbol{\theta} \mapsto \pi(\boldsymbol{\theta})$ (una regla de decisión) tal que
  - $\pi(\boldsymbol{\theta}) = 0$ si no puedo rechazar H_0;
  - $\pi(\boldsymbol{\theta}) = 1$ si rechazo H_0.
- Dado que no conocemos el verdadero valor del parámetro $\boldsymbol{\theta}_0$, basamos nuestra decisión en el estimador $\hat{\boldsymbol{\theta}}$.

Hipótesis

- En general estamos interesados en contrastes que
  - tienen **tamaño** (size, error tipo I) controlado;
    - Error de tipo I: $P(\pi(\hat{\boldsymbol{\theta}}) = 1)$ cuando H_0 es verdadera.
    - Un contraste tiene **nivel** α si $P(\pi(\hat{\boldsymbol{\theta}}) = 1) \leq \alpha$.
    - Un contraste tiene **nivel** asintótico α si $\lim_{N\rightarrow\infty} P(\pi(\hat{\boldsymbol{\theta}}) = 1) \leq \alpha$.
  - tienen **potencia** (poder), máxima si es posible.
    - Poder: $P(\pi(\hat{\boldsymbol{\theta}}) = 1)$ cuando H_1 es verdadero (1- error tipo II).
    - Un contraste es **consistente** si $lim_{N\rightarrow\infty} P(\pi(\hat{\boldsymbol{\theta}}) = 1) = 1$ cuando H_1 es verdadera.

Hipótesis

- Consideremos una hipótesis nula de Q restricciones continuamente diferenciables y linealmente independientes.

$$H_0 : \boldsymbol{c}(\boldsymbol{\theta}_0) = \mathbf{0} \ \ vs. \ \ H_1 : \boldsymbol{c}(\boldsymbol{\theta}_0) \neq \mathbf{0}$$

- Se asume que $\boldsymbol{\theta}_0 \in int\boldsymbol{\Theta}_0$.
- Definamos
  - $\hat{\boldsymbol{\theta}}$ como el estimador **sin restricciones unrestricted** bajo $\boldsymbol{\Theta}$ y
  - $\tilde{\boldsymbol{\theta}}$ como el estimador **con restricciones** bajo la hipótesis nula $\boldsymbol{\Theta}_0$.

Likelihood ratio (LR)

- Debemos tener que $max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\mathcal{L}(\boldsymbol{\theta}) \geq max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0}\mathcal{L}(\boldsymbol{\theta})$ porque $\boldsymbol{\Theta}_0 \subseteq \boldsymbol{\Theta}$, donde $\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\theta})$ es la función log-likelihood.
- El test LR se basa en la comparación de dos funciones de verosimilitud. Bajo H_0:

$$LR \equiv 2(\mathcal{L}(\hat{\boldsymbol{\theta}}) - \mathcal{L}(\tilde{\boldsymbol{\theta}})) \xrightarrow{d} \chi^2_Q$$

- Esto se puede aplicar a los estimadores M:

$$QLR \equiv 2(\sum_{i=1}^{N} q(\boldsymbol{w}_i, \hat{\boldsymbol{\theta}}) - \sum_{i=1}^{N} q(\boldsymbol{w}_i, \tilde{\boldsymbol{\theta}})) \xrightarrow{d} \chi^2_Q$$

- Este método requiere **ambos** estimadores: $\tilde{\boldsymbol{\theta}}$ y $\hat{\boldsymbol{\theta}}$.
- Ejercicio:Probar que si $y|\boldsymbol{x} \sim N(\boldsymbol{x\beta}, \sigma^2)$, $2(\mathcal{L}(\hat{\boldsymbol{\theta}}) - \mathcal{L}(\tilde{\boldsymbol{\theta}})) = N \ln(\tilde{\sigma}^2/\hat{\sigma}^2)$.

Contrastes de Wald

- El **estadístico de Wald** es

$$W = \boldsymbol{c}(\hat{\boldsymbol{\theta}})'(\hat{\boldsymbol{C}}\hat{\boldsymbol{V}}\hat{\boldsymbol{C}}')^{-1}\boldsymbol{c}(\hat{\boldsymbol{\theta}})$$

donde $\hat{\boldsymbol{V}}$ es un estimador consistente de la varianza de $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{C}} \equiv \boldsymbol{C}(\hat{\boldsymbol{\theta}})$, $\boldsymbol{C}(\boldsymbol{\theta})$ es el $Q \times P$ jacobiano de $\boldsymbol{c}(\boldsymbol{\theta})$. Entonces, bajo la hipótesis nula

$$W \xrightarrow{d} \chi^2_Q.$$

Contrastes de Wald

- La distribución de este contraste depende de $\theta_0 \in int\Theta_0$ y de una aplicación del delta method.
- Usando el teorema del valor medio,

$$\begin{aligned}
\sqrt{N}(\boldsymbol{c}(\hat{\theta}) - \boldsymbol{c}(\theta_0)) &= \ddot{\boldsymbol{C}}_N \cdot \sqrt{N}(\hat{\theta} - \theta_0) \\
&= \boldsymbol{C}(\theta_0) \cdot \sqrt{N}(\hat{\theta} - \theta_0) + [\ddot{\boldsymbol{C}}_N - \boldsymbol{C}(\theta_0)] \cdot \sqrt{N}(\hat{\theta} - \theta_0) \\
&= \boldsymbol{C}(\theta_0) \cdot \sqrt{N}(\hat{\theta} - \theta_0) + o_p(1) \cdot O_p(1) \\
&= \boldsymbol{C}(\theta_0) \cdot \sqrt{N}(\hat{\theta} - \theta_0) + o_p(1)
\end{aligned}$$

  Entonces,

$$\sqrt{N}(\boldsymbol{c}(\hat{\theta}) - \boldsymbol{c}(\theta_0)) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{C}_0 \boldsymbol{A}_0^{-1} \boldsymbol{B}_0 \boldsymbol{A}_0^{-1} \boldsymbol{C}_0'),$$

  where $\boldsymbol{C}_0 \equiv \boldsymbol{C}(\theta_0)$.

- El principal problema del test de Wald es que no son invariantes a como se especifican las restricciones no lineales. Los resultados pueden variar mucho en muestras pequeñas.
- Este método requiere sólo la estimación de $\hat{\theta}$.

Contrastes de Wald para restricciones lineales

- Consideremos una matriz $\boldsymbol{R}$ no estocástica de dimensión $Q \times P$, $Q \leq P$, con $rango(\boldsymbol{R}) = Q$ (hay Q filas linealmente independientes).
- Supongamos

$$H_0 : \boldsymbol{R}\boldsymbol{\theta}_0 = \boldsymbol{r} \ \ contra \ \ \boldsymbol{R}\boldsymbol{\theta}_0 \neq \boldsymbol{r}$$

  donde $\boldsymbol{r}$ es un vector $Q \times 1$ no aleatorio.
- El estadístico de Wald es

$$W = (\boldsymbol{R}\hat{\boldsymbol{\theta}} - \boldsymbol{r})'(\boldsymbol{R}\hat{\boldsymbol{V}}\boldsymbol{R}')^{-1}(\boldsymbol{R}\hat{\boldsymbol{\theta}} - \boldsymbol{r}) \xrightarrow{d} \chi^2_Q$$

Contrastes de Wald: ejercicios

Tomemos el modelo $y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + u_i$. Armar los contrastes de Wald para las siguientes hipótesis nulas:

- $H_0 : \beta_2 = 0$;
- $H_0 : \beta_1 = 0, \beta_2 = 0$;
- $H_0 : \beta_1 = \beta_2$;
- $H_0 : \beta_1 = 1$;
- $H_0 : ln(\beta_1) = 0$.

Rao's score o LM

- Consideremos un contraste basado en las funciones score. Se llama Rao's score o de multiplicadores de Lagrange (Lagrange multiplier, LM). Entonces usando el teorema del valor medio

$$LM = \left(\sqrt{N}\sum_{i=1}^{N}\boldsymbol{s}_i(\tilde{\boldsymbol{\theta}})\right)' \boldsymbol{A}_0^{-1}\boldsymbol{C}_0' \left[\boldsymbol{C}_0\boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}\boldsymbol{C}_0'\right]^{-1} \boldsymbol{C}_0\boldsymbol{A}_0^{-1} \left(\sqrt{N}\sum_{i=1}^{N}\boldsymbol{s}_i(\tilde{\boldsymbol{\theta}})\right) \xrightarrow{d} \chi^2_Q$$

- Este método sólo requiere $\tilde{\boldsymbol{\theta}}$, o sea, la estimación bajo la hipótesis nula. En general, es más simple estimar bajo H_0 que H_1. Esta es la principal ventaja de LM.
- Para máxima verosimilitud

$$LM = \left(\sqrt{N}\sum_{i=1}^{N}\boldsymbol{s}_i(\tilde{\boldsymbol{\theta}})\right)' \boldsymbol{A}_0^{-1} \left(\sqrt{N}\sum_{i=1}^{N}\boldsymbol{s}_i(\tilde{\boldsymbol{\theta}})\right) \xrightarrow{d} \chi^2_Q$$

Rao's score y contrastes de especificación

- Sea $\boldsymbol{g}(\boldsymbol{w},\boldsymbol{\theta})$ un vector $Q \times 1$ que mapea $\mathcal{W} \times \boldsymbol{\Theta} \mapsto \boldsymbol{g}(\boldsymbol{w},\boldsymbol{\theta})$ y asumamos que $\boldsymbol{g}(\boldsymbol{w},\boldsymbol{\theta}_0)$ es diferenciable para $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. Sea $\boldsymbol{g}_i(\boldsymbol{\theta}) \equiv \boldsymbol{g}(\boldsymbol{w}_i,\boldsymbol{\theta})$. Asumamos que $E[\boldsymbol{g}_i(\boldsymbol{\theta}_0)] = \mathbf{0}$.
- Igualdad de la matriz de información: $-E[\nabla_{\theta}\boldsymbol{g}_i(\boldsymbol{\theta}_0)] = E[\boldsymbol{g}_i(\boldsymbol{\theta}_0)\boldsymbol{s}_i(\boldsymbol{\theta}_0)']$.
- Supongamos que $H_0 : E[\boldsymbol{g}_i(\boldsymbol{\theta}_0)] = 0$. Por ejemplo, $\boldsymbol{g}()$ puede ser la función score de los parámetros bajo la hipótesis nula, que deberían ser cero.
- El estadístico Newey-Tauchen-White (NTW) para testeat por esta condición de momento es

$$NTW = \left(\sqrt{N}\sum_{i=1}^{N}\boldsymbol{g}_i(\hat{\boldsymbol{\theta}})\right)' \left[\sqrt{N}\sum_{i=1}^{N}(\boldsymbol{g}_i(\hat{\boldsymbol{\theta}}) - \boldsymbol{s}_i(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\Pi}})(\boldsymbol{g}_i(\hat{\boldsymbol{\theta}}) - \boldsymbol{s}_i(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\Pi}})'\right]^{-1} \left(\sqrt{N}\sum_{i=1}^{N}\boldsymbol{g}_i(\hat{\boldsymbol{\theta}})\right)$$

donde $\hat{\boldsymbol{\Pi}}$ es un estimador consistente de $\boldsymbol{\Pi}_0 \equiv \{E[\boldsymbol{s}_i(\boldsymbol{\theta}_0)\boldsymbol{s}_i(\boldsymbol{\theta}_0)']\}^{-1}$.

Comportamiento bajo alternativas locales

- Para analizar el poder de los contrastes se usa una secuencia de alternativas locales $H_1^N : \boldsymbol{c}(\boldsymbol{\theta}_{0,N}) = \delta_0/\sqrt{N}$ donde δ_0 es una vector $Q \times 1$.
- Dividir por $\sqrt{N}$ asegura que el estadístico esta bien definido en cuanto a su distribución asintótica bajo H_1.
- Under H_1^N,

$$\sqrt{N}(\boldsymbol{c}(\hat{\boldsymbol{\theta}}) - \boldsymbol{c}(\boldsymbol{\theta}_0)) \xrightarrow{d} N(\delta_0, \boldsymbol{C}_0\boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}\boldsymbol{C}_0').$$

- Entonces,

$$LR \overset{a}{\sim} W \overset{a}{\sim} LM \xrightarrow{d} \chi_Q^2(\lambda_0),$$

donde $\chi_Q^2(\lambda_0)$ es una chi cuadrado no centrada con parámetro de no centralidad $\lambda_0 = \delta_0' \left(\boldsymbol{C}_0\boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}\boldsymbol{C}_0'\right)\delta_0$.

Referencias

Estas notas se basan en

- *Capítulos 12, 13 and 14 de Wooldridge.*
- *Bera, A.K., Montes-Rojas, G.V., and Sosa-Escudero, W. (2010), "General specification testing with locally misspecified models," Econometric Theory 26, 1838–1845.*
- *Newey, W.K., y McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing," en Handbook of Econometrics, Volumen 4, ed. R.F. Engle y D. McFadden. Amsterdam: North Holland, 2111–2245.*
- *Van der Vaart, A.W. (1998), Asymptotic Statistics. Cambridge University Press.*