Test basados en Máxima Verosimilitud

Fundamentos de Econometría

Juan Palomino[1]

[1]Magister en Economía Aplicada con Mención Estudios Regionales
juan.palominoh@pucp.pe

Departamento de Economía

Índice

The holy trinity

- Hemos cubierto las propiedades teóricas más importantes del MLE: es consistente, asintóticamente normal y eficiente.
- Ahora dirigiremos nuestra atención a un problema diferente: la inferencia basada en la probabilidad o verosimilitud.
- Específicamente, iremos más allá de la verosimilitud como mecanismo para producir estimaciones puntuales y veremos cómo podemos usar la función de verosimilitud para realizar pruebas de hipótesis.

The holy trinity

- Hay tres enfoques ampliamente utilizados para llevar a cabo la inferencia basada en verosimilitud:
  - **Wald** [Abraham Wald]
  - **Ratio de verosimilitud** (LR) [C.R. Rao]
  - **Multiplicador de Lagrange** (LM) [Jerzy Neyman, Egon Pearson, Samuel Wilks]
- En muestras grandes, estos tests tienden a coincidir.
- Estas pruebas pueden considerarse como una comparación entre las estimaciones obtenidas después de que las restricciones implícitas por la hipótesis se hayan impuesto a las estimaciones obtenidas sin las restricciones.

Índice

Test Ratio de Verosimilitud (Explicación Conceptual)

- Esta es una prueba de hipótesis que ayuda a elegir el "mejor" modelo entre dos modelos anidados (nested models) o jerarquizados.
- Modelos anidados son aquellos que tienen las mismas variables, sin embargo, se puede construir otro modelo con un número menor de las mismas variables.

Test Ratio de Verosimilitud (Explicación Conceptual)

Ejemplo 1

Definir cuál de los siguientes modelos es el que mejor se ajusta a los datos reales:

- Modelo 1 - Modelo irrestricto:

$$anemia_i = \beta_0 + \beta_1 edad_i + \beta_2 sexo_i + \beta_3 altura_i + \beta_4 peso_i + \varepsilon_i$$

- Modelo 2 - Modelo restringido:

$$anemia_i = \beta_0 + \beta_1 edad_i + \beta_2 sexo_i + \varepsilon_i$$

LR compara los dos modelos; si el modelo 1 es mejor que el modelo 2, los parámetros adicionales del modelo más complejo se usan en los análisis posteriores.

Test Ratio de Verosimilitud (Explicación Conceptual)

- Dado que LR solo se usa para comparar modelos anidados, es necesario tener en cuenta, que el agregar variables adicionales siempre incrementará la verosimilitud, sin embargo, esto es marginalmente decreciente, es decir llegará un punto en el que agregar más variables no mejora significativamente el ajuste.
- El mejor modelo es el que hace que los datos sean más probables, o maximice la función de verosimilitud: $L(\theta, y \mid X)$.
- En la H_0 se evalúa el modelo más pequeño:
  - Aceptar la H_0 indica que el modelo restringido es el mejor
  - Rechazar la H_0 significa que el modelo irrestricto es mejor que que el modelo restringido.

Test Ratio de Verosimilitud (Explicación Conceptual)

El LR comienza con una comparación de las puntuaciones de verosimilitud de los dos modelos:

$$LR = 2 * (lnLNR - lnLR) \sim \chi^2_{1-\alpha}(q)$$

el cual se suele presentar en forma de diferencia logarítmica:

$$LR = 2 * (\frac{lnLNR}{lnLR}) \sim \chi^2_{1-\alpha}(q)$$

y cambiando los signos:

$$LR = -2 * (lnLR - lnLNR) \sim \chi^2_{1-\alpha}(q)$$

Para probar si la diferencia entre los dos modelos es estadísticamente significativa, se debe considerar los grados de libertad que son iguales al número de variables adicionales en el modelo más grande o irrestricto.

Test Ratio de Verosimilitud (Explicación Matemática)

- Retomando la función de verosimilitud: cuando no hay restricciones en el espacio de parámetros, se sabe que:

$$\mathscr{L} = lnL = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

- Supongamos que queremos probar la hipótesis de coeficientes restringidos: $H_0 : R\beta_R = r$ donde R es $(q \times k)$
- Para incoporar las restricciones en la función de verosimilitud se construye:

$$lnL_R = lnL + \lambda'(r - R\beta)$$

donde puede haber cualquier número de restricciones sobre el vector de coeficientes.

Test Ratio de Verosimilitud (Explicación Matemática)

- La estimación con restricciones genera estimadores "restringidos" $(\tilde{\beta}_R, \tilde{\sigma}^2_R)$, que al ser sustituidos en la función L, producen la $lnL_R(\tilde{\beta}_R, \tilde{\sigma}^2_R)$.
- Obteniendo las derivadas parciales con respecto a los parámetros desconocidos e igualándolas a cero, se llega a:

$$\tilde{\beta}_R = \hat{\beta} + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\hat{\beta})$$

$$\tilde{\sigma}^2_R = \frac{(y - X\tilde{\beta}_R)'(y - X\tilde{\beta}_R)}{n}$$

Test Ratio de Verosimilitud (Explicación Matemática)

Test Ratio de Verosimilitud (LR)

Defínase λ como la razón de verosimilitud:

$$\lambda = \frac{lnL_R(\tilde{\beta}_R, \tilde{\sigma}_R^2)}{lnL(\hat{\beta}, \hat{\sigma}^2)}$$

El valor de λ nunca podrá ser mayor que 1 porque el valor de los restringidos es menor que el de los no restringidos, entonces $lnL_R(\tilde{\beta}_R, \tilde{\sigma}_R^2) < lnL(\hat{\beta}, \hat{\sigma}^2)$.
Al ser λ muy pequeña , se aumenta la probabilidad de rechazar H_0. Asimismo, la razón de verosimilitud para muestras grandes se define como:

$$LR = -2ln\lambda = 2[lnL(\hat{\beta}, \hat{\sigma}^2) - lnL_R(\tilde{\beta}_R, \tilde{\sigma}_R^2)] \sim \chi^2_{1-\alpha}(q)$$

la cual converge asintóticamente a una distribución chi-cuadrada con q grados de libertad.

Test Ratio de Verosimilitud (Explicación Matemática)

- Entonces evaluando $(\hat{\beta}, \hat{\sigma}^2)$ en:

$$
\begin{aligned}
lnL(\hat{\beta}, \hat{\sigma}^2) &= -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta) \\
&= -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n}) - \frac{1}{2(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n})}(\hat{\varepsilon}'\hat{\varepsilon}) \\
&= -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n}) - \frac{n}{2}
\end{aligned}
$$

- Similarmente, el logaritmo de la verosimilitud evaluada en $(\tilde{\beta}_R, \tilde{\sigma}_R^2)$ es:

$$
lnL(\tilde{\beta}_R, \tilde{\sigma}_R^2) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{n}) - \frac{n}{2}
$$

Test Ratio de Verosimilitud (Explicación Matemática)

- El estadístico del test de razón de verosimilitud es:

$$
\begin{aligned}
-2ln\lambda &= -2[lnL_R(\tilde{\beta}_R, \tilde{\sigma}_R^2) - lnL(\hat{\beta}, \hat{\sigma}^2)] \\
&= n[log(\frac{\tilde{\varepsilon}^{'}\tilde{\varepsilon}}{n}) - log(\frac{\hat{\varepsilon}^{'}\hat{\varepsilon}}{n})] \\
&= n[ln(\tilde{\varepsilon}^{'}\tilde{\varepsilon}) - ln(\hat{\varepsilon}^{'}\hat{\varepsilon})]
\end{aligned}
$$

- Este se distribuye asintóticamente como un chi-cuadrado con q grados de libertad si la H_0 es cierta.

Índice

Test Wald

- Es un test alternativo al anterior y que prueba la misma hipótesis $H_0 : R\hat{\beta} - r = 0$ basándose en estimadores de máxima verosimilitud.
- El vector $R\hat{\beta} - r = 0$ con un valor cercano a cero, acepta H_0; de lo contrario se rechaza.
- A diferencia del test de razón de verosimilitud, esta prueba solo requiere el cálculo de los estimadores del modelo irrestricto $\hat{\beta}$.
- Asintóticamente debe dar los mismos resultados que el test de razón de verosimilitud.

Test Wald

Test Wald

El estadístico W de Wald es:

$$W = (R\hat{\beta} - r)^{'}[Var(R\hat{\beta} - r)]^{-1}(R\hat{\beta} - r)$$

La expresión $Var(R\hat{\beta} - r)$ puede deducirse por el método delta o directamente, siendo igual a $Var(R\hat{\beta} - r) = R\sigma^2(X^{'}X)^{-1}R^{'}$.
Esto resulta en:

$$W = \frac{(R\hat{\beta} - r)^{'}[R(X^{'}X)^{-1}R^{'}]^{-1}(R\hat{\beta} - r)}{\sigma^2}$$

Test Wald

- Nótese que el estadístico W de Wald es muy parecido al estadístico

$$F = \frac{(R\hat{\beta}-r)^{'}[R(X^{'}X)^{-1}R^{'}]^{-1}(R\hat{\beta}-r)/q}{\hat{\varepsilon}'\hat{\varepsilon}/(n-k)}$$

- El estadístico W solo requiere el cómputo de los estimadores sin restricciones, al igual que el estadístico F.
- Adicionalmente, si n es muy grande, es cierto que:

$$\frac{1}{q}W \approx F$$

- Se puede expresar en términos de las sumas de cuadrados de los residuos de los modelos restringido y del irrestricto

$$W = \frac{\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}/n} = \frac{n(\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon})}{\hat{\varepsilon}'\hat{\varepsilon}}$$

Índice

Test Multiplicador Lagrange (ML)

- Este se basa en los estimadores restringidos de máxima verosimilitud $\tilde{\theta} = \tilde{\beta}_R, \tilde{\sigma}^2_R$.
- Se prueba también la misma hipótesis $H_0 : C(\theta) - r = 0$ basándose en estimadores de máxima verosimilitud.

Test LM

El estadístico LM es:

$$LM = \left(\frac{\partial lnL(\tilde{\theta})}{\partial \tilde{\theta}}\right)' [I(\tilde{\theta})]^{-1} \left(\frac{\partial lnL(\tilde{\theta})}{\partial \tilde{\theta}}\right) \sim \chi^2_{1-\alpha}(q)$$

Si la hipótesis es cierta, lo más probable es que LM tome valores pequeños cercanos a cero y, si la hipótesis es falsa, LM tomará valores grandes.
Si el valor LM calculado es mayor que el percentil $\chi^2_{1-\alpha}(q)$, se rechaza la H_0 con $\alpha\%$ de significancia.

- Por ejemplo, si queremos probar la hipótesis $H_0 : R\beta = r$ con el test LM.
- Partiendo de las derivadas parciales evaluadas en $\tilde{\beta}_R, \tilde{\sigma}_R^2$ es:

$$\frac{\partial lnL}{\partial \beta}(\tilde{\beta}_R,\tilde{\sigma}_R^2)=\frac{1}{\tilde{\sigma}^2}\text{X'}(\text{y-X}\tilde{\beta})=\frac{1}{\tilde{\sigma}^2}\text{X'}\tilde{\varepsilon}$$

$$\frac{\partial lnL}{\partial \sigma^2}(\tilde{\beta}_R,\tilde{\sigma}_R^2)=-\frac{n}{2\tilde{\sigma}^2}+\frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{2\tilde{\sigma}^4}=-\frac{n}{2\tilde{\sigma}^2}+\frac{n\tilde{\sigma}^2}{2\tilde{\sigma}^4}=0$$

Test Multiplicador Lagrange (ML)

- Utilizando la inversa de la matriz de información $I^{-1}(\beta,\sigma^2) = \begin{bmatrix} \tilde{\sigma}^2(X'X)^{-1} & 0 \\ 0 & \frac{2\tilde{\sigma}^4}{n} \end{bmatrix}$ se obtiene:

$$LM = \begin{bmatrix} \frac{1}{\tilde{\sigma}^2}\tilde{\varepsilon}'X & 0 \end{bmatrix} \begin{bmatrix} \tilde{\sigma}^2(X'X)^{-1} & 0 \\ 0 & \frac{2\tilde{\sigma}^4}{n} \end{bmatrix} \begin{bmatrix} \frac{1}{\tilde{\sigma}^2}X'\tilde{\varepsilon} \\ 0 \end{bmatrix}$$

$$= \frac{\tilde{\varepsilon}'X(X'X)^{-1}X'\tilde{\varepsilon}}{\tilde{\sigma}^2} = \frac{n\tilde{\varepsilon}'X(X'X)^{-1}X'\tilde{\varepsilon}}{\tilde{\varepsilon}'\tilde{\varepsilon}}$$

- LM solo requiere de las estimación de los parámetros del modelo sujeto a las restricciones.

Test Multiplicador Lagrange (ML)

- Este estadístico también puede expresarse en términos de las sumatorias de cuadrados de los residuos restringidos e irrestrictos:

$$LM = \frac{n\tilde{\varepsilon}'X(X^{'}X)^{-1}X'\tilde{\varepsilon}}{\tilde{\varepsilon}'\tilde{\varepsilon}} = \frac{n\tilde{\varepsilon}'[I-M]\tilde{\varepsilon}'}{\tilde{\varepsilon}'\tilde{\varepsilon}} = \frac{n[\tilde{\varepsilon}'\tilde{\varepsilon}-\tilde{\varepsilon}'M\tilde{\varepsilon}]}{\tilde{\varepsilon}'\tilde{\varepsilon}}$$

- Dado que $\tilde{\varepsilon}' = \hat{\varepsilon} - X(\tilde{\beta} - \hat{\beta})$, multiplicando esta expresión por M y dado que $MX = 0$ y que $X'\hat{\varepsilon} = 0$:

$$M\tilde{\varepsilon} = M\hat{\varepsilon} - MX(\tilde{\beta} - \hat{\beta}) = M\hat{\varepsilon} = (I - X(X'X)^{-1}X')\hat{\varepsilon} = \hat{\varepsilon}$$

- Luego $\tilde{\varepsilon}'M\tilde{\varepsilon} = \hat{\varepsilon}'\hat{\varepsilon}$, entonces:

$$LM = \frac{n[\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon}]}{\tilde{\varepsilon}'\tilde{\varepsilon}} = n(1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{\tilde{\varepsilon}'\tilde{\varepsilon}}) = nR^2_{aux}$$

- donde R^2_{aux} es el R-cuadrado resultante de una regresión del tipo $\tilde{\varepsilon} = X\delta \hat{+} \varepsilon$.

Índice

Relación entre las pruebas

- Comparando los tres test mencionados:

$$LR = n[ln(\tilde{\varepsilon}^{'}\tilde{\varepsilon}) - ln(\hat{\varepsilon}^{'}\hat{\varepsilon})]$$

$$W = \frac{\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}/n} = \frac{n(\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon})}{\hat{\varepsilon}'\hat{\varepsilon}}$$

$$LM = \frac{n[\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon}]}{\tilde{\varepsilon}'\tilde{\varepsilon}} = n(1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{\tilde{\varepsilon}'\tilde{\varepsilon}})$$

- No arrojan los mismos valores.

- De hecho:

$$W \geq LR \geq LM$$

- Wald tiende a rechazar con más frecuencia que LM y LR, considerando que los tres estadísticos se comparan con el mismo percentil de la distribución chi-cuadrado.
- Estas diferencias se reducen a cero cuando $n \to \infty$, con lo cual los tres test son asintóticamente equivalentes:

$$W = LR = LM$$

- Por ello, es recomendable usar estos tests en muestras grandes.

Índice

Gráfico: Test LR, Wald y LM

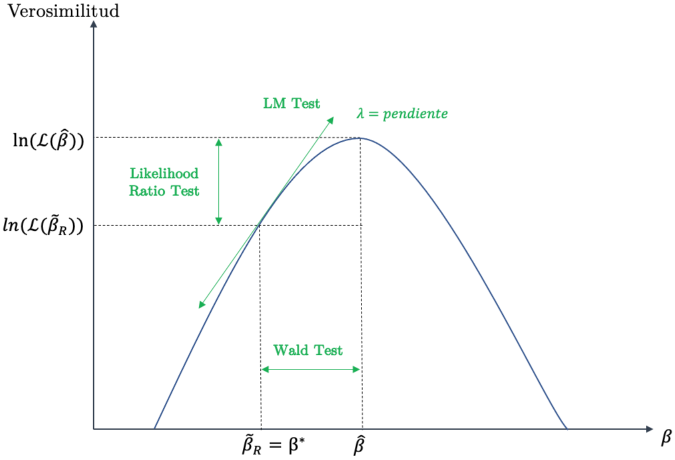

Figura: Comparación de Test LR, Wald y LM

Gráfico: Test Ratio de Verosimilitud

- El test LR evalúa la restricción mediante la comparación de $lnL(\hat{\beta})$ al $lnL(\tilde{\beta}_R)$
- La $H_0 : \beta = \beta^*$ impone el estimador restringido $\tilde{\beta}_R = \beta^*$
- Si $\hat{\beta} = \beta^*$, entonces $lnL(\tilde{\beta}_R) < lnL(\hat{\beta})$
- Si la restricción reduce significativamente la verosimilitud, entonces la hipótesis nula es rechazada.
- La función log-verosimilitud es una línea sólida.

Gráfico: Test Wald

El Test Wald estima el modelo sin restricciones, y evalua la restricción considerando:

1. La distancia $\hat{\beta} - \tilde{\beta}_R = \hat{\beta} - \beta^*$: a mayor distancia, menos probable que la restricción sea cierta.
2. La distancia es ponderada por la curvatura de la función de verosimilitud $\frac{\partial^2 lnL}{\partial \beta^2}$
   - Cuanto más grande es la segunda derivada, más rápido cambia la curva.
   - La función log verosimilitud (line punteada) en casi plana, por lo que la segunda derivada evaluada en $\hat{\beta}$ es relativamente pequeña.
   - Cuando la segunda derivada es pequeña, la distancia $\hat{\beta}$ y $\tilde{\beta}_R$ es relativamente menor a la variación muestral.
   - La segunda función log-verosimilitud tiene una segunda derivada más grande.
   - Con una segunda derivada más grande, la misma distancia entre $\hat{\beta}$ y $\tilde{\beta}_R$ podría ser significativa.

Gráfico: Test Multiplicador Lagrange (ML)

- Este solo estima el modelo restringido.
- Este evalúa la pendiente de la función log verosimilitud en la restringida.
- Si H_0 es cierta, la pendiente (score) en la restringida debe ser cercana a cero.
- Al igual que con la prueba de Wald, la curvatura de la función de probabilidad logarítmica en la restricción se usa para evaluar la importancia de una pendiente distinta de cero.

Relación entre las pruebas

- Cuando H_0 es cierto, los tests de Wald, LR y LM son asintóticamente equivalentes.
- A medida que $n \longrightarrow \infty$, la distribuciones muestrales de los tres tests $\xrightarrow{d} \chi_r^2$, donde r es el número de restricciones a testear.

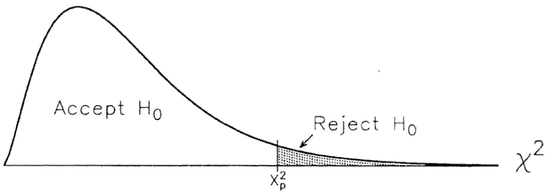

Figura: Distribución muestral del χ_p^2

- Son similares solo cuando $n \longrightarrow \infty$. En pequeñas muestras esto no es necesariamente cierto.

Referencias

- **Chapter 4.1 Hypothesis Testing and Goodness of Fit.** - Scott Long, J. (1997). Regression models for categorical and limited dependent variables. Advanced quantitative techniques in the social sciences, 7.
- **Apéndice 8A2** - Gujarati, D., & Porter, D. (2010). Econometría (Quinta edición ed.). & P. Carril Villareal, Trad.) México: Mc Graw Hill educación.