PUBLICACIONES DE 4er CURSO

Grado: Economía

Asignatura: ECONOMETRÍA III

Tema 3: Evaluación
Apartado 3.2: Contrastes LR, LM y Wald

Grupos:

Profesores: Antonio Aznar y M. Teresa Aparicio

Departamento de ANÁLISIS ECONÓMICO

Curso Académico 2014/15

1542
Facultad de
Economía y Empresa
Universidad Zaragoza

Apartado 3.2

CONTRASTES LR, W y LM

Hemos visto en la Sección anterior que la Región Crítica de un contraste se determinaba a partir del cociente de los valores tomados por la función de verosimilitud sustituyendo los parámetros por sus estimaciones con y sin restricción. En particular, se definía una distancia entre los valores que toma la función de verosimilitud, y si esta distancia era grande se rechazaba la hipótesis nula.

Pero no hay un procedimiento único para dar cuenta de esta distancia. En la literatura se han desarrollado otras formas de tratar esa distancia y nosotros vamos a prestar atención a tres de ellas asociadas con los llamados contrastes de la Razón de Verosimilitud (LR), Wald (W) y Multiplicadores de Lagrange (LM).

En principio, la argumentación la llevaremos a cabo utilizando un caso muy simple en el que la muestra de tamaño T se ha obtenido a partir de una población con media μ y varianza σ^2 conocida.

En aquellos casos en que sea de interés para el contenido de otras secciones posteriores, los resultados se extenderán a un marco más general en el que la función de verosimilitud depende de un vector de k parámetros , θ , y se pretende contrastar las r restricciones lineales que escribimos como:

$$R\,\theta = q \qquad (3.2.1)$$

En donde R es una matriz r.k y q es un vector de r elementos.

La función de verosimilitud para el caso simple viene dada por:

$$L\left(\mu;\sigma^2,y\right)= (2\pi)^{-T/2}(\sigma^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}\Sigma\left(y_t-\mu\right)^2\right\}$$

Su logaritmo puede escribirse como:

$$\ell(\mu) = \log L(\ ) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\Sigma\,(y_t-\mu)^2$$

<u>Estimación MV sin restricciones</u> ($\tilde{\mu}$)

El estimador MV sin restricciones es aquél que cumple que:

$$\ell\left(\tilde{\mu}\right) = \sup_{\forall\mu} \ell\left(\mu\right)$$

La condición necesaria para alcanzar el máximo es que el gradiente se iguale a cero. El gradiente se define como:

$$d(\mu) = \frac{\partial\ell(\mu)}{\partial\mu} = \frac{\Sigma\ (y_t-\mu)}{\sigma^2}$$

Por lo tanto, la condición necesaria puede escribirse:

$$d(\tilde{\mu}) = 0 \Rightarrow \tilde{\mu} = \bar{y}$$

El valor que toma el logaritmo de la función de verosimilitud sustituyendo los parámetros por los estimadores MV sin restricciones es:

$$\ell(\tilde{\mu}) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\Sigma\,(y_t - \tilde{\mu})^2$$

<u>Estimación MV con restricciones</u> ($\tilde{\mu}_R$)

Supongamos que queremos estimar μ cumpliéndose la restricción: $\mu=\mu_0$, siendo μ_0 un valor conocido.

El estimador MV restringido de μ es aquel que cumple que:

$$\ell(\tilde{\mu}_R) = \sup_{\mu=\mu_0} \ell(\mu)$$

Para obtener la condición necesaria se define la función lagrangiana como:

$$L = \ell(\mu) + \lambda(\mu - \mu_0)$$

Las primeras derivadas de esta función respecto a μ y λ vienen dadas por:

$$\frac{\partial L}{\partial \mu} = d(\mu) + \lambda$$

$$\frac{\partial L}{\partial \lambda} = \mu - \mu_0$$

Las condiciones necesarias vienen dadas, por tanto, por:

$$d(\tilde{\mu}_R) + \tilde{\lambda} = 0 \Rightarrow d(\tilde{\mu}_R) = -\tilde{\lambda}$$

$$\tilde{\mu}_R = \mu_0$$

Podemos resumir los resultados obtenidos en el siguiente Cuadro:

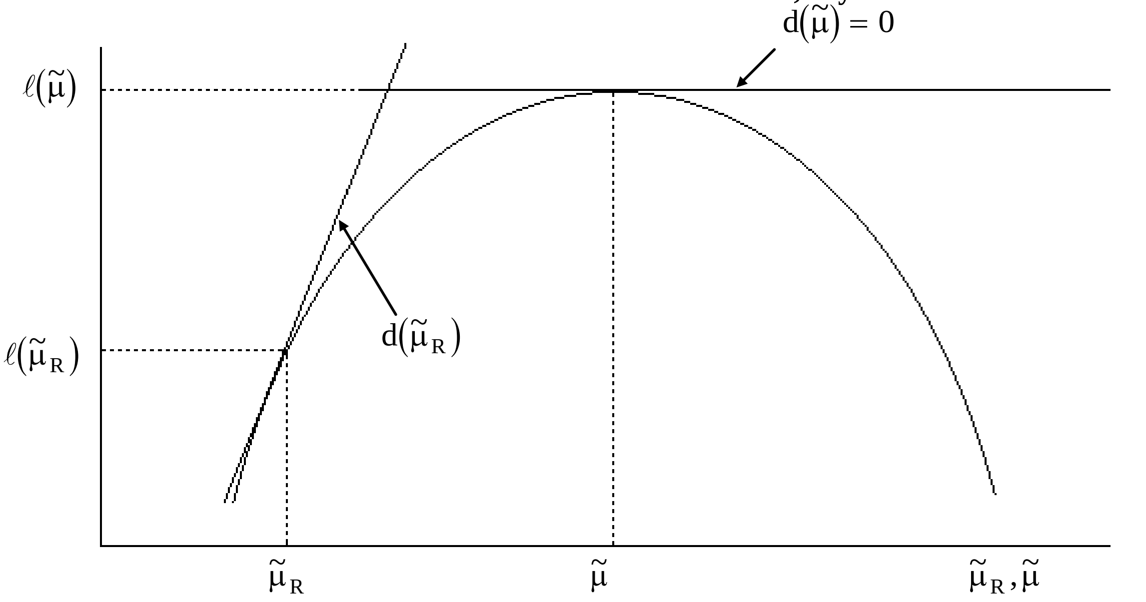
	SIN RESTRICCIONES	CON RESTRICCIONES
Logaritmo de la Func. de Verosimilitud	$\ell(\tilde{\mu})$	$\ell(\tilde{\mu}_R)$
Estimador MV	$\tilde{\mu}$	$\tilde{\mu}_R$
Gradiente	$d(\tilde{\mu}) = 0$	$d(\tilde{\mu}_R)$

Hemos comentado en la sección anterior que la Región Crítica se determina prestando atención a la distancia entre los valores tomados por la función de verosimilitud o, equivalentemente, por los tomados por sus logaritmos, es decir, $\ell(\tilde{\mu}) - \ell(\tilde{\mu}_R)$.

Pero a similares resultados puede llegarse si prestamos atención a la distancia entre los dos estimadores o entre los valores tomados por el gradiente.

El siguiente gráfico puede ayudarnos a entender lo que subyace a estos tres planteamientos:

FIGURA 3.2. Ilustración de los Contrastes LR,W y LM.



Puede apreciarse cómo existe una relación monótona entre las tres medidas alternativas de la distancia. Conforme $\ell(\tilde{\mu}_R)$ se aproxima a $\ell(\tilde{\mu}), \tilde{\mu}_R$ se aproxima a $\tilde{\mu}$ y $d(\tilde{\mu}_R)$ se aproxima a $d(\tilde{\mu})$.

El contraste de cualquier hipótesis está basado en el siguiente principio: **"sólo se puede aceptar una restricción**

cuando su toma en consideración no distorsiona de forma relevante la evidencia contenida en los datos".

El que una distorsión sea o no relevante puede ponerse en relación con las tres medidas de distancia comentadas.

Podemos decir que la distorsión es relevante cuando:

1) La distancia entre $\ell(\tilde{\mu})$ y $\ell(\tilde{\mu}_R)$ es grande.

2) La distancia entre $\tilde{\mu}$ y $\tilde{\mu}_R$ es grande.

3) La distancia entre $d(\tilde{\mu})$ y $d(\tilde{\mu}_R)$ es grande o, equivalentemente, si la distancia

entre $d(\tilde{\mu}_R)$ y cero es grande.

Ahora bien, la distancia entre estadísticos hay que establecerla prestando atención a la distribución estocástica que siguen dichos estadísticos.

Los dos resultados que se precisan para calibrar esas distancias son:

$$\sqrt{T}\,(\tilde{\mu}-\mu_0)\xrightarrow[H_0\ cierta]{d}\mathrm{N}[0,\ \lim\ \mathrm{T\ I}(\mu)^{-1}\ ]$$

$$\frac{1}{\sqrt{T}}\,d\,(\tilde{\mu}_R)\xrightarrow[H_0\ cierta]{d}\mathrm{N}[0,\ \lim\ \frac{\mathrm{I}(\mu)}{\mathrm{T}}\ ]$$

(la demostración de estos resultados puede verse en Godfrey (1988))

en donde I(μ) es la matriz de información -en este caso en que μ es un escalar, elemento de información- definida como:

$$I(\mu) = - E \frac{\partial^2 \ell(\mu)}{\partial \mu^2} = \frac{T}{\sigma^2}$$

Para llegar a los estadísticos que sirven de base para el contraste necesitamos un resultado adicional que podemos formular como:

<u>Resultado 3.2.1</u>: Sea X un vector de p variables aleatorias distribuidas conjuntamente como: N(0, Σ_p) siendo Σ_p una matriz no singular. Entonces se tiene que $X' \Sigma_p^{-1} X \sim \chi^2(p)$.

Veamos ahora la forma que toman los tres contrastes comentados.

<u>Contraste de Wald</u>: Está basado en la distancia $(\tilde{\mu} - \tilde{\mu}_R)$ o bien en la distancia $(\tilde{\mu} - \mu_0)$. Por los resultados comentados, podemos escribir:

$$W = \sqrt{T}(\tilde{\mu} - \mu_0)\, [\widehat{Var}\, \sqrt{T}(\tilde{\mu} - \mu_0)]^{-1}\, \sqrt{T}(\tilde{\mu} - \mu_0) =$$

$$= \sqrt{T}(\tilde{\mu} - \mu_0)\, [\, I(\tilde{\mu}) / T\,]\, \sqrt{T}(\tilde{\mu} - \mu_0) = (\tilde{\mu} - \mu_0)\, I(\tilde{\mu})(\tilde{\mu} - \mu_0)$$

Asintóticamente y bajo la hipótesis nula este estadístico se distribuye como una χ^2 con un grado de libertad.

La forma de operar con este estadístico es la siguiente: se fija un tamaño de error tipo 1, $\varepsilon = \varepsilon_0$, se busca en las tablas el correspondiente punto crítico $\chi^2_{\varepsilon_0}(1)$ y la Región Crítica del contraste se determina como:

$$W > \chi^2_{\varepsilon_0}(1)$$

Para el caso general en el que lo que se pretende es contrastar (3.2.1) entonces el contraste toma la forma siguiente:

$$W = (R\tilde{\theta} - q)'[RI(\tilde{\theta})^{-1}R']^{-1}(R\tilde{\theta} - q) \qquad (3.2.2)$$

y bajo H_0, W sigue una distribución χ^2 con r grados de libertad.

<u>Contraste de Multiplicadores de Lagrange (LM)</u>

Este contraste se basa en la distancia $d(\tilde{\mu}_R) - d(\tilde{\mu})$, y teniendo en cuenta que $d(\tilde{\mu}) = 0$, simplemente en lo distante que $d(\tilde{\mu}_R)$ está de cero. Por los resultados comentados se tiene que:

$$LM = T^{-1/2} d(\tilde{\mu}_R) \left[V\hat{a}r\left(T^{-1/2} d(\tilde{\mu}_R)\right)\right]^{-1} T^{-1/2} d(\tilde{\mu}_R) = d(\tilde{\mu}_R)\, I(\tilde{\mu}_R)^{-1} d(\tilde{\mu}_R)$$

Asintóticamente y bajo la hipótesis nula este estadístico se distribuye como una χ^2 con un grado de libertad.

La Región Crítica del contraste es:

$$\mathrm{LM} > \chi^2_{\varepsilon_0}(1)$$

Para el caso general asociado con el contraste de (3.2.1), el contraste de los Multiplicadores de Lagrange es:

$$\mathrm{LM} = d(\tilde{\theta}_R)' I(\tilde{\theta}_R)^{-1} d(\tilde{\theta}_R) \qquad (3.2.3)$$

que, bajo H_0, sigue una distribución χ^2 con r grados de libertad.

Contraste de la Razón de Verosimilitud

El contraste está basado en la distancia $(\ell(\tilde{\mu}) - \ell(\tilde{\mu}_R))$. La distribución de este estadístico puede obtenerse fácilmente tras tomar una expansión de Taylor a partir de la de $\sqrt{T}(\tilde{\mu} - \mu_0)$ pudiéndose escribir:

$$\mathrm{LR} = 2[\ell(\tilde{\mu}) - \ell(\tilde{\mu}_R)] \sim \chi^2(1) \qquad (3.2.4)$$

la Región Crítica del contraste se determina como:

$$\mathrm{LR} > \chi^2_{\varepsilon_0}(1)$$

Para el caso general, el contraste de la Razón de Verosimilitud es:

$$\mathrm{LR} = 2[\ell(\tilde{\theta}) - \ell(\tilde{\theta}_R)] \qquad (3.2.5)$$

que, bajo H_0, sigue una distribución χ^2 con r grados de libertad.

Hemos presentado una justificación intuitiva de los tres contrastes. Pero ahora la pregunta relevante es: ¿son admisibles estos tres contrastes?. La admisibilidad con carácter general se ha demostrado asintóticamente; como puede verse en Godfrey (1988), los tres son consistentes garantizando el mejor uso de la evidencia disponible cuando ésta es grande.

3.4. CONTRASTE DE HIPOTESIS EN EL MARCO DEL MODELO LINEAL GENERAL

Partiremos del modelo escrito en (1.3), suponiendo que se cumplen las ocho hipótesis introducidas en la Sección 1.2.

Nuestro objetivo es contrastar las r restricciones lineales $R\beta = q$. Para tomar la decisión de aceptar o rechazar esta hipótesis nula utilizaremos el siguiente estadístico:

$$F = \frac{(R\hat{\beta} - q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)/r}{\hat{\sigma}^2} \qquad (3.2.6)$$

Resultado 3.3: Cumplidas las hipótesis enunciadas en el Apartado 2.3, el estadístico escrito en (3.2.6), bajo la hipótesis nula, sigue una distribución de probabilidad F central con r grados de libertad en el numerador y T-k en el denominador.

Prueba: Una forma equivalente de escribir (3.2.6) es:

$$F = \frac{(R\hat{\beta}-q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-q)/r\sigma^2}{\frac{1}{T-k} x \left(\frac{(T-k)\hat{\sigma}^2}{\sigma^2}\right)}$$

El numerador sigue una distribución χ^2 con r grados de libertad utilizando resultados vistos en el Apartado 2.3. Tambien se ve que el denominador es una distribución χ^2 con T-K grados de libertad dividida por esos grados de libertad. Por último, numerador y denominador son independientes y el resultado sigue.

Resultado 3.2.2: Una forma equivalente de escribir el estadístico F escrito en (3.2.6) es la siguiente:

$$F = \frac{(\hat{u}_R^{'}\hat{u}_R - \hat{u}'\hat{u})/r}{\hat{\sigma}^2} \qquad (3.2.7)$$

en donde $\hat{u}$ es el vector de residuos MCO, sin restringir y $\hat{u}_R$ es el vector de residuos restringidos definido en (2.3.58).

Prueba: Por coincidir los denominadores, basta probar la equivalencia de los numeradores.

A partir de (2.3.51) se tiene que:

$$\hat{\beta}-\hat{\beta}_R=(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-q) \qquad (3.2.8)$$

Por lo tanto, podemos escribir:

$$\left(\hat{\beta}-\hat{\beta}_R\right)'X'X(\hat{\beta}-\hat{\beta}_R)=(R\hat{\beta}-q)'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'X$$

$$(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-q)=$$

$$=(R\hat{\beta}-q)'[R(X'X)^{-1}R]^{-1}(R\hat{\beta}-q)$$

que es precisamente el numerador de (3.2.6). Por otra parte, utilizando lo que hemos llamado la Identidad Fundamental en el Resultado 2.3.3 podemos escribir:

$$(\hat{\beta}-\hat{\beta}_R)'X'X(\hat{\beta}-\hat{\beta}_R)=\hat{u}_R'\hat{u}_R-\hat{u}'\hat{u} \qquad (3.2.9)$$

llegándose así al resultado buscado tras corregir con los grados de libertad.

La forma de determinar la región crítica utilizando el estadístico F se hace siguiendo lo que hemos llamado un enfoque verificacionista. Se especifica a priori un nivel del tamaño del Error Tipo 1, $\varepsilon=\varepsilon_0$, y la hipótesis nula se rechaza cuando:

$$F>F_{\varepsilon_0}\ (r, T\text{-}k) \qquad (3.2.10)$$

Para estudiar la forma que toman los estadísticos LR, W y LM cuando se contrastan las r restricciones vamos a recopilar los resultados relevantes reescribiendo algunos de ellos. Primero, hay que tener en cuenta que:

$$\theta' = (\beta', \sigma^2)$$

El logaritmo de la función de verosimilitud, los estimadores MV de $\beta \ y \ \sigma^2$, el gradiente y la matriz de información pueden verse , respectivamente, en (2.3.37), (2.3.41), (2.3.42), (2.3.38) y (2.3.48).

Para obtener los estimadores MV restringidos la función Lagrange la escribimos como:

$$L = \ell(\theta) + \lambda'(R\beta - q) \qquad (3.2.11)$$

en donde λ es el vector de r multiplicadores de Lagrange.

Las condiciones necesarias de máximo son:

$$\frac{\partial L}{\partial \beta} = 0; \qquad \frac{\partial L}{\partial \sigma^2} = 0; \qquad y \quad \frac{\partial L}{\partial \lambda} = 0$$

Teniendo en cuenta los k primeros elementos del gradiente en (2.3.38) se tiene:

$$\frac{1}{\tilde{\sigma}_R^2} X'(y - X\hat{\beta}_R) + R'\tilde{\lambda} = 0 \qquad (3.2.12)$$

y teniendo en cuenta el correspondiente a σ^2 :

$$-\frac{T}{\tilde{\sigma}_R^2}+\frac{\Sigma\tilde{u}_{Rt}^2}{\tilde{\sigma}_R^4}=0 \qquad (3.2.13)$$

A partir de (3.2.12) y (3.2.13) se obtienen los estimadores MV restringidos. El correspondiente a β coincide con el MC0 y el correspondiente a σ^2 ,a partir de (3.2.13), es:

$$\tilde{\sigma}_R^2=\frac{\tilde{u}_R^{'}\tilde{u}_R}{T} \qquad (3.2.14)$$

Hay que destacar que los elementos del gradiente correspondientes a β evaluados con los estimadores restringidos son diferentes a cero tal como se ve a partir de (3.2.12). El elemento del gradiente correspondiente a σ^2 evaluado con estimadores restringidos si que es cero como se deriva de (3.2.13).

<u>Resultado 3.2.3</u>: En el marco del Modelo Lineal General en el que se cumplen las hipótesis habituales el estadístico que se utiliza para el contraste de la Razón de Verosimilitud puede escribirse como:

$$LR = T\log\left(\frac{\tilde{\sigma}_R^2}{\tilde{\sigma}^2}\right) \qquad (3.2.15)$$

<u>Prueba</u>: Los valores que toma el logaritmo de la función de verosimilitud sustituyendo los parámetros de la misma escrito en

(2.3.37) por sus estimadores sin restringir y restringidos son, respectivamente:

$$\ell(\tilde{\theta}) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\tilde{\sigma}^2 - \frac{\tilde{u}'\tilde{u}}{2\tilde{\sigma}^2} = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\tilde{\sigma}^2 - \frac{T}{2}$$

$$\ell(\tilde{\theta}_R) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\tilde{\sigma}_R^2 - \frac{\tilde{u}_R'\tilde{u}_R}{2\tilde{\sigma}_R^2} = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\tilde{\sigma}_R^2 - \frac{T}{2}$$

sustituyendo ahora en (3.2.4) se tiene:

$$LR = 2\left[-\frac{T}{2}\log\tilde{\sigma}^2 + \frac{T}{2}\log\tilde{\sigma}_R^2\right] = T\log\frac{\tilde{\sigma}_R^2}{\tilde{\sigma}^2}$$

<u>Resultado 3.2.4</u>: En el mismo marco supuesto en el Resultado 3.2.3, el estadístico que se utiliza para el contraste de Wald puede escribirse como:

$$W = T\,\frac{(\tilde{\sigma}_R^2 - \tilde{\sigma}^2)}{\tilde{\sigma}^2} \qquad (3.2.16)$$

<u>Prueba</u>: Teniendo en cuenta que $R^*\tilde{\theta} = R\tilde{\beta}$ y recordando la forma que toma la matriz de información escrita en (2.3.48), podemos escribir (3.2.2) como:

$$W = \frac{1}{\tilde{\sigma}^2}(R\tilde{\beta} - q)'[R(X'X)^{-1}R']^{-1}(R\tilde{\beta} - q) \qquad (3.2.17)$$

Los desarrollos contenidos en el Resultado 3.2.2 nos permiten escribir:

$$(R\tilde{\beta} - q)'[R(X'X)^{-1}R']^{-1}(R\tilde{\beta} - q) = \tilde{u}'_R \tilde{u}_R - \tilde{u}'\tilde{u}$$

Utilizando esta identidad en (3.2.17) y multiplicando y dividiendo por T nos permite demostrar el resultado.

<u>Resultado 3.2.5</u>: En el mismo marco asumido en el Resultado 3.2.3, el estadístico que se utiliza para el contraste de los Multiplicadores de Lagrange puede escribirse como:

$$LM = T\,\frac{(\tilde{\sigma}^2_R - \tilde{\sigma}^2)}{\tilde{\sigma}^2_R} \qquad (3.2.18)$$

<u>Prueba</u>: Si evaluamos el gradiente con los estimadores restringidos se tiene que:

$$d(\tilde{\theta}_R) = \begin{bmatrix} \dfrac{1}{\tilde{\sigma}^2_R} X'\tilde{u}_R \\ 0 \end{bmatrix}$$

Por lo tanto, (3.2.3) es ahora:

$$LM = d(\tilde{\theta}_R)' I(\tilde{\theta}_R)^{-1} d(\tilde{\theta}_R) =$$

$$= \begin{bmatrix} \dfrac{1}{\tilde{\sigma}^2_R} \tilde{u}'_R X & 0 \end{bmatrix} \tilde{\sigma}^2_R \begin{bmatrix} (X'X)^{-1} & 0 \\ 0 & \dfrac{2\tilde{\sigma}^2_R}{T} \end{bmatrix} \begin{bmatrix} \dfrac{1}{\tilde{\sigma}^2_R} X'\tilde{u}_R \\ 0 \end{bmatrix} =$$

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

y como:

$$X'\tilde{u}_R = X'(y - X\tilde{\beta}_R) = X'y - X'X\tilde{\beta}_R = X'X(\tilde{\beta} - \tilde{\beta}_R)$$

podemos escribir:

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

y teniendo en cuenta (3.2.9) basta dividir y multiplicar por T para llegar al resultado.

Un último resultado que tiene interés es el que relaciona estos tres estadísticos con el que corresponde al contraste de la F.

<u>Resultado 3.2.6</u>: Los contrastes LR, W y LM pueden escribirse como:

$$LR = T \log \left(1 + \frac{rF}{T-k} \right) \tag{3.2.19}$$

$$W = \frac{TrF}{T-k} \tag{3.2.20}$$

$$LM = \frac{Tr}{(T-k)F^{-1} + r} \tag{3.2.21}$$

<u>Prueba</u>: La demostración es inmediata teniendo en cuenta que a partir de (3.2.7) podemos escribir:

$$F = \frac{T}{r\hat{\sigma}^2}(\tilde{\sigma}_R^2 - \tilde{\sigma}^2) = \frac{T-k}{r} \text{ x } \frac{\tilde{\sigma}_R^2 - \tilde{\sigma}^2}{\tilde{\sigma}^2}$$

de donde:

$$\frac{\tilde{\sigma}_R^2 - \tilde{\sigma}^2}{\tilde{\sigma}^2} = \frac{Fr}{T-k}$$

o, equivalentemente:

$$\frac{\tilde{\sigma}_R^2}{\tilde{\sigma}^2} = 1 + \frac{Fr}{T-k}$$

Utilizando estas expresiones y sustituyéndolas en (3.2.15), (3.2.16) y (3.22.18) y haciendo ligeras transformaciones se llega al resultado.

Referencias

Aznar, A. (2012): "Curso de Econometría" Copy Center Digital. Zaragoza.