

Contrastes asintóticos de especificación en funciones frontera

Juan Muro
(Universidad de Alcalá)



Resumen

El contraste de multiplicadores de Lagrange (ML) se utiliza para verificar el diagnóstico realizado en la especificación paramétrica de funciones frontera. En el caso de fronteras estrictas, se construye un contraste ML para la prueba de una especificación $\gamma(P, \lambda)$. En el supuesto de fronteras estocásticas, se diseña un contraste ML en forma de TR^2 para las especificaciones normal-exponencial y normal-gamma.

1 Introducción

Al depender las medidas de eficiencia derivadas de los modelos de funciones frontera, entre otras cosas, de la especificación del término de error de los mismos, el contraste de los supuestos establecidos se convierte en una tarea de gran relevancia empírica. En este artículo, se introduce un contraste de multiplicadores de Lagrange (ML) que permite verificar la forma de la distribución del término de error en las funciones frontera. Mientras que en el apartado 2 se plantea en términos generales la utilización del contraste ML como prueba del diagnóstico efectuado, en los apartados 3 y 4 se aplica dicho contraste a los supuestos generalmente utilizados en la literatura en la especificación de funciones frontera. Finalmente, se sintetizan en el apartado 5 los principales resultados obtenidos.

2 El contraste ML como prueba del diagnóstico en las funciones frontera

El modelo de función frontera que será analizado a lo largo del artículo es de la forma

$$y_t = \beta' X_t + \varepsilon_t, \quad \forall t = 1, 2, \dots, T; \quad [1]$$

donde $\varepsilon_t = v_t + u_t$ en el supuesto de fronteras estocásticas y $\varepsilon_t = u_t$ en el supuesto de fronteras estrictas. En ambas especificaciones $u_t \geq 0$ en el análisis de funciones de coste mientras que $u_t \leq 0$ en el estudio de funciones de producción o beneficio. Los componentes del término estocástico ε_t en (1) presentan el significado habitual: v_t es el error aleatorio debido a factores no sometidos al control de la empresa y u_t es la variable aleatoria representativa de la ineficiencia.

Bajo la hipótesis general de que u_t pertenece a la familia de las distribuciones truncadas de Pearson, la utilización del contraste ML nos permite realizar la prueba de las distintas hipótesis establecidas sobre u_t en cada una de las especificaciones concretas de una función frontera, siempre que estas especificaciones particulares pertenezcan a dicha familia. Este enfoque, que circunscribe la prueba de la distribución de u_t a aquellos casos en que dicha distribución sea un

miembro de la familia de Pearson, resulta de hecho muy general ya que proporciona un contraste de la mayoría de los supuestos que pueden encontrarse en la literatura sobre especificación y estimación de funciones frontera de carácter paramétrico.

La idea primitiva se encuentra en Jarque y Bera(1980) y ha sido aplicada a las fronteras estocásticas con distribución normal-seminormal y normal-normal truncada por Lee(1983).

La función de densidad de probabilidad de un miembro cualquiera de la familia de distribuciones de Pearson puede expresarse como solución de la ecuación diferencial

$$d\ln f(u) = \frac{(P-1) + \gamma\lambda - \lambda u}{-\gamma + u + b_2\lambda u^2} du \quad [2]$$

Llamaremos al vector de parámetros $\varphi = [P \ \lambda \ \gamma \ b_2]'$.

La estimación de (1) bajo el supuesto de que la distribución de u_t es un miembro particular de la familia de las distribuciones truncadas de Pearson, implica la estimación de una ecuación en la que hayamos introducido un conjunto de restricciones en el vector φ en (2). El contraste ML puede ser utilizado para probar la validez del diagnóstico realizado al especificar dicho modelo restringido. Si llamamos $L(\theta)$ al logaritmo de la función de verosimilitud de (1), donde $\theta = [\beta' \ \varphi']'$ es un vector $(k+4) \times 1$ que puede partitionarse en $\theta = [\theta_1' \ \theta_2']'$: θ_2 es un vector de $(m \times 1)$ parámetros restringidos; θ_1 es el correspondiente vector de parámetros sin restricciones, el contraste de la hipótesis nula $H_0: \theta_2 = 0$ puede efectuarse mediante

$$ML = \hat{d}_2' (\hat{l}_{22} - \hat{l}_{21} \hat{l}_{11}^{-1} \hat{l}_{12})^{-1} \hat{d}_2 \approx X_m^2 \quad [3]$$

contraste ML (dada la no singularidad de la matriz de información). Donde \hat{d}_2 : vector de primeras derivadas del logaritmo de la función de verosimilitud calculado para $\theta = \hat{\theta}$, estimación maximoverosímil de los parámetros del modelo restringido; \hat{l}_{ij} : matriz de segundas derivadas calculadas para $\theta = \hat{\theta}$, o cualquier equivalente asintótico.

Por motivos de orden en la exposición desarrollaremos en primer lugar la forma que adopta (3) en los modelos de frontera estricta para estudiar a continuación el contraste ML en la especificación de fronteras estocásticas.

3 Contraste ML en las fronteras estrictas: el caso de la distribución gamma

En los modelos de frontera estricta ($\varepsilon_t = u_t$) el logaritmo de la función de verosimilitud de (1), bajo el supuesto de que el término de ineficiencia se distribuye como un miembro de (3), es

$$L(\theta, u) = L_1(\theta, u) + L_2(\theta, u), \quad u \geq 0. \quad [4]$$

Con

$$L_1(\theta, u) = -T \ln \left\{ \int_0^{\infty} \exp[\Psi(\theta, u)] du \right\}; \quad L_2(\theta, u) = \sum_{t=1}^T \Psi(\theta, u_t);$$

$$\Psi(\theta, u) = \int d \ln f(u);$$

donde hemos introducido en (1) el supuesto de frontera de producción o beneficio.

La forma que adoptan las derivadas parciales en esta situación, con excepción de las derivadas parciales con respecto a β , lleva implícito el cálculo de lo que podríamos denominar *momentos* de distribuciones que pertenecen a la familia de Pearson. En efecto, la distribución de un miembro de la familia de distribuciones truncadas de Pearson puede expresarse como

$$f(u) = \frac{A(\theta, u)}{\int_0^{\infty} A(\theta, u) du}. \quad A(\theta, u) = \exp[\Psi(\theta, u)] \quad [5]$$

Por consiguiente, la palabra *momentos* se utiliza en este caso para representar una amplia gama de integrales de funciones de la variable aleatoria u . Un ejemplo genérico de estos *momentos* podría ser

$$\frac{\int_0^{\infty} z(u) A(\theta, u) du}{\int_0^{\infty} A(\theta, u) du}.$$

La función $z(u)$ que permite identificar estos *momentos* se genera a partir de las derivadas de $A(\theta, u)$ con respecto a los parámetros contenidos en el vector ϕ .

Entre las especificaciones que cumplen las condiciones establecidas por Greene (1980) para los modelos de frontera estricta, la distribución gamma biparamétrica, $G(P, \lambda)$, ofrece ventajas adicionales. Por una parte, su grado de flexibilidad

permite un mejor ajuste a las distintas formas que en la realidad puede presentar la estructura de eficiencias de un sector productivo; por otra, es un miembro de la familia de Pearson.

La distribución gamma corresponde a un miembro de la familia de Pearson si en (2) imponemos la restricción siguiente: $b_2 = \gamma = 0$, $\lambda \neq 0$. En este caso, $\theta_1 = [\beta' \ P \ \lambda]'$, $\theta_2 = [\gamma \ b_2]'$. El contraste del supuesto de que los términos de error en (1) se distribuyen como una $G(P, \lambda)$ es el contraste de $H_0: \theta_2 = 0$, bajo la hipótesis general de que u_t pertenece a (3).

El contraste ML adopta la forma analítica siguiente:

$$d_2 = \left[\lambda T - (P-1) \sum_t \frac{1}{u_t} \quad \frac{TP(P-3)}{2} + \frac{\lambda^2}{2} \sum_t u_t^2 - \lambda(P-1) \sum_t u_t \right]^{-1} \quad [6]$$

$$I_{22} = \begin{bmatrix} \frac{\lambda^2}{P-2} & (P+1)(\lambda-P) \\ \frac{P(P+9)}{2} & \end{bmatrix} \quad [7]$$

$$I_{11} = \begin{bmatrix} \frac{\lambda^2}{(P-2)T} \sum_t X_t X_t' & \frac{\lambda}{(P-1)T} \sum_t X_t & \frac{-1}{T} \sum_t X_t \\ \frac{\partial^2}{\partial P^2} \ln \Gamma(P) & \frac{-1}{\lambda} & \frac{P}{\lambda^2} \end{bmatrix} \quad [8]$$

$$I_{12} = \begin{bmatrix} \frac{\lambda^2}{(P-2)T} \sum_t X_t & \frac{\lambda}{T} \sum_t X_t \\ \frac{\lambda}{P-1} & \frac{3}{2} \\ -1 & -2 \frac{P}{\lambda} \end{bmatrix} \quad [9]$$

Las matrices I_{22} , I_{11} , I_{12} son de dimensión (2×2) , $(k+2) \times (k+2)$, $(k+2) \times 2$, respectivamente, donde k es la dimensión del vector de parámetros explícitos β .

4 Contraste ML en las fronteras estocásticas: distribuciones normal-exponencial y normal-gamma

Si el modelo de función frontera se especifica como una frontera estocástica, en Lee(1983), puede verse la forma que adopta el contraste ML para verificar el supuesto de distribuciones normal-semi-normal y normal-normal truncada en las fronteras estocásticas. Así, el análisis en este artículo queda restringido a los supuestos de distribuciones normal-exponencial y normal-gamma.

El cálculo de las derivadas de $L(\theta, u)$ conlleva en estos modelos, de manera análoga a lo que ocurría en las fronteras estrictas, la obtención de ciertos *momentos* de algunas distribuciones. En las fronteras estocásticas junto a la obtención de los *momentos* del término de ineficiencia u_t , aparecen los *momentos* de la distribución condicional del término de ineficiencia dada la distribución del término aleatorio global del modelo analizado. Esta última distribución, que podríamos denominar $f(u_t | \varepsilon_t)$, nos facilita información sobre la forma del término de ineficiencia dado el término de error compuesto global y nos permite, Jondrow *et al.*(1982), la separación de las componentes del error global, eliminando así una de las mayores dificultades conceptuales de los modelos de frontera estocástica.

4.1

La especificación normal-exponencial en una frontera estocástica da origen a un logaritmo de la función de verosimilitud de la forma

$$L(\theta, u) = T \ln \lambda + \frac{1}{2} (T \lambda^2 \sigma_v^2) + \sum_t \ln(1 - F_{3t}^*) - \lambda \sum_t \varepsilon_t, \quad [10]$$

donde $F_{3t}^* = \phi(\lambda \sigma_v \varepsilon_t / \sigma_v)$; $\phi(\cdot)$ función de densidad de probabilidad de una variable normal estándar. La expresión (10) corresponde a la distribución de una variable aleatoria suma de una variable normal y un miembro de la familia de distribuciones truncadas de Pearson si imponemos sobre esta última las restricciones $P=1$; $\gamma=0$; $b_2=0$. El contraste del supuesto de distribución normal-exponencial en una frontera estocástica es la verificación de $H_0: \theta = 0$. La expresión del contraste ML será en esta oportunidad (3) con $\theta_1 = [\beta' \sigma_v^2 \lambda]'$; $\theta_2 = [(P-1) \gamma b_2]'$.

En esta ocasión la aplicación directa de (3) plantea dificultades. Sustituimos en (3) la matriz de información por una de sus equivalentes asintóticas. Utilizaremos una de las expresiones del contraste denominadas *sin cálculo de esperanzas* en Bera y McKenzie(1983). Sea $\partial/\partial\theta_j [L(\theta, u)]$ un vector $(p \times 1)$ con componentes

$$\partial/\partial\theta_j \left[\sum_{t=1}^T \partial/\partial\theta_j L_t(\theta, u) \right];$$

la matriz de información I se sustituye por la matriz $B = GG'$, donde G' tiene por elementos

$$G'_{ij} = \frac{\partial}{\partial \theta_j} L_t(\theta, u), t = 1, 2, \dots, T; j = 1, 2, \dots, p,$$

con lo que obtenemos una expresión conveniente del contraste ML como

$$ML = \hat{d}_2' (\hat{B}_{22} - \hat{B}_{21} \hat{B}_{11}^{-1} \hat{B}_{12})^{-1} \hat{d}_2, \quad [11]$$

que es asintóticamente equivalente a (3).

La aplicación de (11) al supuesto de frontera estocástica con distribución normal-exponencial implica el cálculo de las primeras derivadas parciales del logaritmo de la función de verosimilitud de una variable normal-miembro de la familia de Pearson, bajo la restricción señalada anteriormente H_0 . Estas son

$$\frac{\partial L(\theta, u)}{\partial \beta} = \sum_t \frac{X_t'}{\sigma_v^2} \{\varepsilon_t + E[u | \varepsilon_t]\}; \quad [12]$$

$$\frac{\partial L(\theta, u)}{\partial \sigma_v^2} = \frac{1}{2\sigma_v^2} \sum_t \left\{ \frac{1}{\sigma_v^2} (\varepsilon_t^2 + 2\varepsilon_t E[u | \varepsilon_t] + E[u^2 | \varepsilon_t]) - 1 \right\}; \quad [13]$$

$$\frac{\partial L(\theta, u)}{\partial \rho} = \sum_t \{E[\ln u | \varepsilon_t] - E[\ln u]\}; \quad [14]$$

$$\frac{\partial L(\theta, u)}{\partial \gamma} = 0; \quad \frac{\partial L(\theta, u)}{\partial \lambda} = -\sum_t \{E[u | \varepsilon_t] - E[u]\}; \quad [15]; [16]$$

$$\frac{\partial L(\theta, u)}{\partial b_2} = \frac{\lambda^2}{2} \sum_t \{E[u^2 | \varepsilon_t] - E[u^2]\}; \quad [17]$$

donde

$$E(u) = \int_0^{\infty} u \lambda \exp(-\lambda u) du; \quad E(u | \varepsilon_t) = \int_0^{\infty} u f(u | \varepsilon_t) du;$$

y $f(u | \varepsilon_t)$ es una distribución normal truncada con moda $-\lambda(\sigma_v^2 + \varepsilon_t / \lambda)$ y varianza σ_v^2 .

En el supuesto normal-exponencial el vector d tendrá por dimensión (3×1) , y por elementos (14), (15) y (17), aunque una de estas componentes es nula. Por otra parte, las matrices B_{22} , B_{21} , B_{11} , tienen por dimensión respectivamente (3×3) , $3 \times (k+3)$, $(k+3) \times (k+3)$, siendo k la dimensión de β . A destacar que

$$B = \sum_t \left[\frac{\partial L_t(\theta, u)}{\partial \theta} \right] \left[\frac{\partial L_t(\theta, u)}{\partial \theta} \right]',$$

donde $\partial/\partial\theta[L_t(\theta, u)]$ es un vector (px1) con componentes $\partial/\partial\theta_j[L_t(\theta, u)]$.

Tanto los vectores d_2, d_2' como la matriz inversa de B contienen una serie de parámetros que no dependen del subíndice i y que pueden ser cancelados en (11).

Para el cálculo de (11) existe un método que permite el cálculo del valor del estadístico ML de forma sencilla. En efecto, si realizamos una regresión de un vector de unos, como variable dependiente, frente a una matriz de datos de variables explicativas igual a G' , o dicha matriz sin los parámetros que no dependen del subíndice muestral como se ha apuntado antes, (11) puede calcularse como TR^2 . Es decir,

$$R^2 = \frac{1}{T} d' B^{-1} d,$$

donde la matriz B se evalúa con las estimaciones maximoverosímiles de la frontera estocástica normal-exponencial.

4.2

En el supuesto de una frontera estocástica con distribución normal-gamma, el contraste se realizará de una forma análoga a la desarrollada en 4.1. El contraste de este supuesto se reducirá a la verificación de $H_0: \gamma=0; b_2=0$. En este caso, la expresión del contraste ML será (3) con $\theta_1 = [\beta' \sigma_v^2 P \lambda]'$ $\theta_2 = [\gamma b_2]'$. Al surgir las mismas dificultades que en 4.1, la expresión conveniente es (11), que puede calcularse asimismo como TR^2 . Aunque algunas de las expresiones que se obtienen sean similares a las obtenidas en el supuesto normal-exponencial, éstas difieren en que $E(u)$, $E(u|\varepsilon_t)$ se calculan mediante

$$f(u) = \frac{e^{-\lambda u} u^{P-1}}{\int_0^{\infty} e^{-\lambda u} u^{P-1} du}; \quad f(u|\varepsilon_t) = \frac{u^{P-1} e^{-\frac{u^2}{2\sigma_v^2} - \left(\lambda + \frac{\varepsilon_t}{\sigma_v^2}\right)u}}{\int_0^{\infty} u^{P-1} e^{-\frac{u^2}{2\sigma_v^2} - \left(\lambda + \frac{\varepsilon_t}{\sigma_v^2}\right)u} du}.$$

5 Conclusiones

La proliferación de trabajos aplicados sobre la estimación de funciones frontera de carácter paramétrico, que incluyen estudios sobre la eficiencia relativa de las empresas de un sector económico, ha puesto cada vez más de manifiesto la necesidad de contrastar los diagnósticos sobre la distribución del término de error

de las fronteras. Este artículo, en la línea de Lee(1983), resalta la utilidad en este contexto de la aplicación del contraste ML.

En el supuesto de fronteras estrictas se facilita la forma analítica del contraste para el caso más flexible de distribución gamma. Esta forma puede programarse con facilidad.

En el supuesto de fronteras estocásticas, la forma analítica del contraste es demasiado compleja en el caso general. Se proporciona como alternativa un método sencillo de cálculo basado en TR^2 , siendo R^2 el coeficiente de determinación no centrado de la regresión de un vector de unos sobre una matriz de datos formada por los elementos del vector de scores eficientes. Los grados de libertad de la distribución del contraste ML dependen del número de parámetros que se restrinjan en cada supuesto analizado y de la singularidad o no de la matriz de información asintótica en cada uno de ellos.

Referencias bibliográficas

BERA, A.K. Y C.R. MCKENZIE (1983). *Alternative Forms and Properties of the Lagrange Multiplier Test*, Discussion Paper, n° 8336, CORE.

GREENE, W.H.(1980). *Maximum Likelihood Estimation of Econometric Frontier Functions*, Journal of Econometrics, 13, págs. 27-56.

JARQUE, C.M. Y A.K. BERA (1980). *Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regressions Residuals*, Economics Letters, 6, págs. 255-259.

JOHNSON, N.L. Y S. KOTZ (1970). *Continuous Univariate Distributions-1*, Houghton Mifflin Company, Boston, MA.

JONDROW, J., C.A.K. LOVELL, I.S. MATEROV Y P. SCHMIDT (1982). *On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model*, Journal of Econometrics, 19,págs. 233-238.

LEE, L.F. (1983). *A Test for Distributional Assumptions for the Stochastic Frontier Functions*, Journal of Econometrics, 22,págs. 245-267.