

Título de la comunicación:

**ANÁLISIS SOBRE LA FORMA DE LA DISTRIBUCIÓN DE ERROR COMPUESTO DEL
MODELO DE FRONTERA ESTOCÁSTICA. UN ESTUDIO MONTE CARLO.**

Autores:

**Rafaela Dios Palomares.
Antonio Ramos Millán.**

Universidad de Córdoba
Dpto. de Estadística e Investigación Operativa
Apartado de Correos 3048 14080 Córdoba
Email : ma1dipar@uco.es

RESUMEN

El presente trabajo analiza los parámetros de asimetría y curtosis de las distribuciones compuestas que forman el error de la función frontera de eficiencia. Dichas distribuciones son la normal- exponencial, y la normal-seminormal. Se realiza un estudio teórico de dichos parámetros así como de sus relaciones con el parámetro γ que define la proporción de la variable de ineficiencia con la del error fuera de control. Se establece el estadístico de Jarque-Bera como discriminante entre las dos distribuciones compuestas estudiadas, en base a los residuos minimocuadráticos modificados. A través de un experimento Monte Carlo se calculan los valores empíricos de los parámetros de interés. Por último, se realiza otro experimento Monte Carlo que deja patente la buena potencia del estadístico de Jarque Bera como contraste discriminante.

Palabras Clave: Eficiencia. Frontera estocástica. Monte Carlo. Jarque-Bera.

ANÁLISIS SOBRE LA FORMA DE LA DISTRIBUCIÓN DE ERROR COMPUESTO DEL MODELO DE FRONTERA ESTOCÁSTICA. UN ESTUDIO MONTE CARLO.

Rafaela Dios Palomares.

Antonio Ramos Millán.

Dpto. de Estadística e Investigación Operativa
Universidad de Córdoba

RESUMEN

El presente trabajo analiza los parámetros de asimetría y curtosis de las distribuciones compuestas que forman el error de la función frontera de eficiencia. Dichas distribuciones son la normal- exponencial, y la normal-seminormal. Se realiza un estudio teórico de dichos parámetros así como de sus relaciones con el parámetro γ que define la proporción de la variable de ineficiencia con la del error fuera de control. Se establece el estadístico de Jarque-Bera como discriminante entre las dos distribuciones compuestas estudiadas, en base a los residuos minimocuadráticos modificados. A través de un experimento Monte Carlo se calculan los valores empíricos de los parámetros de interés. Por último, se realiza otro experimento Monte Carlo que deja patente la buena potencia del estadístico de Jarque Bera como contraste discriminante.

Palabras Clave: Eficiencia. Frontera estocástica. Monte Carlo. Jarque-Bera.

1. Introducción.

La productividad empresarial ha sido objeto de numerosos estudios a lo largo de la historia. Para un correcto análisis de la misma, hemos de aislar los efectos que las diferentes fuentes de variación ejercen sobre ella, es decir los derivados de diferencias en la eficiencia del proceso productivo, de la tecnología, y finalmente los provenientes de las variaciones debidas al medio concreto en el que se halle el proceso de producción. En este trabajo nos centraremos exclusivamente en el estudio de la primera fuente mencionada, es decir de la eficiencia técnica.

Las dos metodologías más usadas en la actualidad para estimar eficiencias mediante función frontera son la programación matemática aplicando DEA, y la que se denomina frontera econométrica, que será el objeto del presente trabajo, y a la que nos referiremos a continuación

En el análisis econométrico de eficiencia a través de la estimación de función frontera, se han establecido distintos enfoques tanto desde el punto de vista de la especificación del modelo como de la estimación del mismo y de la propia eficiencia. Así, el planteamiento general de una función de producción frontera se especifica por medio del modelo:

$$y = X * \mathbf{b} + \mathbf{e}$$

Donde la variable de error \mathbf{e} , recoge la diferencia entre la parte sistemática del modelo y los valores observados. Este enfoque es paramétrico, y por ello es propenso a confundir los efectos de una especificación errónea de la forma funcional (tecnología) o de las distribución elegida para representar a la ineficiencia.

El enfoque de Error Compuesto supone que la variable de error no recoge únicamente el efecto de la ineficiencia, sino que también existe otra fuente de error incluida en la misma, que no es controlable por el individuo. Se admite por tanto, que la variable \mathbf{e} se genera como diferencia entre una variable estocástica v (no controlable, simétrica, y definida entre $-\infty$ e ∞) y la variable de ineficiencia u , que será siempre positiva y asimétrica, dado el fenómeno que recoge. La variable v se especifica con distribución Normal con media cero y varianza σ_v^2 , mientras que las distribuciones que han sido consideradas con más frecuencia en la literatura para especificar la variable u son la Exponencial, Seminormal, Normal Truncada y Gamma.

Para realizar el análisis de la eficiencia con este planteamiento, en base a la especificación que se establezca para u , se procede a estimar el modelo de función frontera, suponiendo para el error compuesto la distribución que le corresponda a \mathbf{e} como diferencia de v y u . Como resultado inmediato de dicha estimación se obtienen los residuos que recogen dicho error compuesto, del que habrá que extraer la

parte que se deba realmente a ineficiencia, para lo cual se hace uso de los algoritmos propuestos por Jondrow et al.(1981)

La aplicación de la metodología hasta aquí expuesta para el estudio de la eficiencia mediante función frontera presenta, sin embargo, junto a las limitaciones propias de todo trabajo empírico de modelización de una realidad desconocida, y de suponer la homogeneidad del proceso productivo para todo el sistema, la de encontramos con la incertidumbre en la especificación del error de la función frontera, y confundir el efecto de una especificación errónea con la de ineficiencia.

En la práctica es frecuente que esta incertidumbre se resuelva estimando la eficiencia bajo varios supuestos para dicha especificación, como puede verse en los trabajos de Alauddin et al. (1993) y de Taylor y Shonkwiler (1984). Sin embargo, consideramos que la estimación del modelo y de la eficiencia bajo varias especificaciones no aporta información válida sobre cual ha sido el verdadero Proceso Generador de Datos (P.G.D.). La obtención de resultados dispares ocasiona con frecuencia conclusiones ambiguas en cuanto a la eficiencia del sistema estudiado. En estas condiciones, los posibles errores que se producen al estudiar la eficiencia bajo una especificación distinta de la que generó los datos, ocasionan falta de fiabilidad. Creemos que dicha fiabilidad puede mejorar con el acierto en la especificación, ya que es importante que el procedimiento de estimación sea congruente con la verdadera estructura que generó la ineficiencia del sistema.

Así, resulta interesante establecer un método que permita detectar a priori la especificación del P.G.D. que generó la muestra, cuya eficiencia técnica queremos estudiar. Dado que las especificaciones alternativas que con más frecuencia han sido utilizadas, Normal-Seminormal y Normal-Exponencial, difieren en la forma de la distribución, consideramos que los estadísticos de forma de ϵ pueden ser determinantes a la hora de decidir entre ellas. Para plantear por tanto una metodología que ayude a decidir sobre cual será la especificación del P.G.D., se impone la realización de un estudio teórico y otro empírico sobre dichas distribuciones, haciendo especial hincapié en las medidas de forma de las mismas y en la distribución de estas medidas para cada especificación.

El presente trabajo tiene como objetivo general el análisis de los estadísticos de forma, asimetría y curtosis, el cual se realizará con el propósito de estudiar las diferencias de los mismos, en ambas especificaciones. Una vez constatada claramente dicha diferencia, investigaremos sobre alguna medida que aglutine a estos dos estadísticos. Consideramos que el estadístico de Jarque-Bera puede constituir una buena vía de decisión discriminatoria entre las dos especificaciones estudiadas.

La investigación se llevado a cabo con los siguientes objetivos específicos:

- Análisis teórico de las medidas de forma.
- Estudiar las medidas de forma de la distribución de ϵ desde el punto de vista empírico para muestras de tamaño 100, y para las especificaciones Normal-Exponencial y Normal-Seminormal.
- Desarrollar un método que permita discriminar entre las especificaciones planteadas.
- Una vez sondeada la validez de este método como posible vía de decisión entre las especificaciones de error consideradas, se pretende llevar a cabo una optimización del método auxiliándonos de curvas percentiles de esas dos distribuciones para tal propósito.

2. Estudios Teóricos.

2.1. Repercusión de la hipótesis de especificación.

Se exponen a continuación los resultados del estudio teórico para las especificaciones Normal-Seminormal y Normal-Exponencial y que nos servirán de base para los desarrollos de los estadísticos de forma de las mismas y demás resultados.

Las especificaciones de la variable de errores minimocuadráticos que estudiaremos serán la Normal-Seminormal y la Normal-Exponencial, cuyos momentos se expresan a continuación:

- Caso Normal-Seminormal:

$$\mathbf{e} = v - u; \quad v \equiv N(0, \mathbf{s}_v^2); \quad u \equiv N(0, \mathbf{s}_u^2) \quad \text{donde} \quad f(u) = \sqrt{\frac{2}{p}} e^{-\frac{\left(\frac{u}{\mathbf{s}_u}\right)^2}{2}} \quad \text{Para } u > 0$$

$$m_1 = -\sqrt{\frac{2}{p}} \mathbf{s}_u \quad m_2 = \mathbf{s}_v^2 + \frac{p-2}{p} \mathbf{s}_u^2$$

$$m_3 = \sqrt{\frac{2}{p}} \left(\frac{p-4}{p} \right) \mathbf{s}_u^3 \quad m_4 = 3\mathbf{s}_v^4 + 6\frac{p-2}{p} \mathbf{s}_v^2 \mathbf{s}_u^2 + \left(3 - \frac{4}{p} - \frac{12}{p^2} \right) \mathbf{s}_u^4$$

Despejando de las anteriores expresiones obtenemos que:

$$\mathbf{s}_{u(N-SN)}^2 = \left(\sqrt{\frac{p}{2}} \left(\frac{p}{p-4} \right) m_3 \right)^{\frac{2}{3}} = 2.760647293 |m_3|^{\frac{2}{3}}$$

$$\mathbf{s}_{v(N-SN)}^2 = m_2 - \left(\frac{p-2}{p} \hat{\mathbf{s}}_{u(N-SN)}^2 \right) = m_{2(N-SN)} - 1.003164642 |m_{3(n-sn)}|^{\frac{2}{3}}$$

- Caso Normal-Exponencial:

$$\mathbf{e} = v - u; \quad v \equiv N(0, \mathbf{s}_v^2); \quad u \equiv \text{Exp}(\mathbf{q}) \quad \text{donde} \quad f(u) = \mathbf{q} e^{-\mathbf{q}u} \quad \text{Para } u > 0$$

$$m_1 = -\frac{1}{\mathbf{q}} \quad m_2 = \mathbf{s}_v^2 + \frac{1}{\mathbf{q}^2}$$

$$m_3 = -\frac{2}{\mathbf{q}^3} \quad m_4 = \frac{9}{\mathbf{q}^4} + \frac{6\mathbf{s}_v^2}{\mathbf{q}^2} + 3\mathbf{s}_v^4$$

De donde se deduce que:

$$\hat{\mathbf{q}}_{(N-EXP)} = \left(-\frac{2}{m_3} \right)^{\frac{1}{3}}$$

$$\mathbf{s}_{v(N-EXP)}^2 = m_2 - \left(\frac{-m_3}{2} \right)^{\frac{2}{3}} \hat{\mathbf{s}}_{u(N-EXP)}^2 = \frac{1}{\hat{\mathbf{q}}^2} = \left(-\frac{m_3}{2} \right)^{\frac{2}{3}}$$

A lo largo de este estudio utilizaremos los parámetros \mathbf{s}^2 y \mathbf{g}_B (acotado entre 0 y 1) propuestos por Aigner et al. (1977) y Battese-Corra (1977) respectivamente que vienen definidos por la siguientes expresiones:

$$\mathbf{s}^2 = \mathbf{s}_u^2 + \mathbf{s}_v^2 \quad \mathbf{g}_B = \frac{\mathbf{s}_u^2}{\mathbf{s}_u^2 + \mathbf{s}_v^2}$$

De ambas fórmulas se deducen las siguientes: $\mathbf{s}_u^2 = \mathbf{g}_B \mathbf{s}^2$ y $\mathbf{s}_v^2 = (1 - \mathbf{g}_B) \mathbf{s}^2$

La definición del parámetro \mathbf{I}_A dada por Aigner et al. es: $\mathbf{I}_A = \frac{\mathbf{s}_u}{\mathbf{s}_v}$;

expresión que puesta en función de \mathbf{g}_B resulta ser: $\mathbf{I}_A = \sqrt{\frac{\mathbf{g}_B}{(1 - \mathbf{g}_B)}}$

El parámetro γ ($\mathbf{g} = \frac{V(u)}{V(u) + \mathbf{s}_v^2}$) recoge la proporción de la varianza residual que se debe a la varianza de la ineficiencia. Además definimos el parámetro \mathbf{I} como la relación entre las desviaciones típicas de u y v respectivamente.

Llegados a este punto es importante puntualizar la distinción en la especificación de errores Normal-Seminormal, entre \mathbf{s}_u^2 y $V(u)$ (Dios,1999). El parámetro \mathbf{s}_u^2 es la varianza de la $N(0, \mathbf{s}_u^2)$ de la cual procede la distribución Seminormal correspondiente a la componente $u_{i..}$. Obviamente, la varianza de la distribución Seminormal será diferente a la de la Normal de la que procede, estando relacionadas mediante la siguiente fórmula:

$$V(u) = \frac{p-2}{p} \mathbf{s}_u^2$$

En este caso \mathbf{s}_e^2 , que definimos como la varianza de ϵ será igual a la suma de varianzas de v (\mathbf{s}_v^2) y de u ($V(u)$) de modo que: $\mathbf{s}_e^2 = \mathbf{s}_v^2 + V(u)$

Al haber diferenciado entre \mathbf{s}_u^2 y $V(u)$, cabe también diferenciar entre los parámetros: \mathbf{g} y \mathbf{I} (hallados mediante el uso de $V(u)$) y sus correspondientes valores homólogos \mathbf{g}_B y \mathbf{I}_A , definidos anteriormente.

Así podemos expresar:

$$\mathbf{g} = \frac{V(u)}{V(u) + \mathbf{s}_v^2}; \quad y \quad \mathbf{I} = \frac{\sqrt{V(u)}}{\mathbf{s}_v}$$

En el caso de la especificación de errores Normal-Exponencial no cabe hacer la distinción entre \mathbf{s}_u^2 y $V(u)$, puesto que ambas son iguales. Ni tampoco entre γ y \mathbf{I} con respecto a \mathbf{g}_B y \mathbf{I}_A .

Así pues, tendremos que: $V(u) = \mathbf{s}_u^2$ donde $\mathbf{s}_u^2 = \frac{1}{q^2}$

Realizando las pertinentes operaciones encontramos las siguientes relaciones en función de \mathbf{g} y \mathbf{s}_e^2 :

$$q^2 = \frac{1}{s_e^2 g} \quad s_v^2 = (1-g)s_e^2 = \frac{1-g}{q^2 g}$$

Para valorar la repercusión que tiene la especificación supuesta para e en la estimación de los parámetros, nos basaremos en el análisis de eficiencia mediante el método de los mínimos cuadrados modificados. La estimación de los parámetros s_u^2 , s_v^2 , l y g , se realiza a partir de los momentos de los residuos mínimocuadráticos. Para tal fin partimos de las expresiones de los momentos de las especificaciones de los errores Normal-Seminormal y Normal-Exponencial dadas anteriormente.

Dada la serie de residuos mínimocuadráticos de una muestra estudiada, denominaremos m_2^* y m_3^* a sus momentos de segundo y tercer orden.

La varianza estimada de u para la Normal-Seminormal es:

$$\hat{V}(u) = 1.003 \left(-m_3^* \right)^{\frac{2}{3}}$$

mientras que para la Normal-Exponencial la varianza de u tomará el valor :

$$\hat{s}_{u(N-EXP)}^2 = \left(-\frac{m_3^*}{2} \right)^{\frac{2}{3}} = 0.623 \left(-m_3^* \right)^{\frac{2}{3}}$$

Con lo cual, $\hat{V}(u)$ siempre será mayor pudiendo expresar: $\hat{V}(u) > \hat{s}_{u(N-EXP)}^2$

Pasemos ahora a cuantificar el grado de la desigualdad anterior dividiendo las expresiones que nos proporcionan $\hat{V}(u)$ y $\hat{s}_{u(N-EXP)}^2$:

$$\hat{V}(u)_{(N-SN)} = \frac{1.003}{0.630} \hat{s}_{u(N-EXP)}^2 = 1.592 \hat{s}_{u(N-EXP)}^2$$

Comparemos ahora las estimaciones de \hat{s}_v^2 correspondientes a ambas especificaciones de errores:

La varianza estimada de v para las especificaciones Normal-Seminormal y Normal-Exponencial son respectivamente:

$$\hat{s}_{v(N-SN)}^2 = m_2^* - 1.003 \left| m_3^* \right|^{\frac{2}{3}}; \quad \text{y} \quad \hat{s}_{v(N-EXP)}^2 = m_2^* - 0.63 \left| m_3^* \right|^{\frac{2}{3}}$$

Con lo cual, siempre ocurrirá que $\hat{s}_{v(N-SN)}^2 < \hat{s}_{v(N-EXP)}^2$

Siendo:

$$\hat{s}_{v(N-SN)}^2 = m_2^* + 1.592 \left(\hat{s}_{v(N-EXP)}^2 - m_2^* \right)$$

Una vez hecho, esto pasaremos a relacionar las fronteras medias estimadas mediante el método de Mínimos Cuadrados Ordinarios Modificados (MCOM). Siguiendo este método, la frontera global se calcula añadiendo a la estimación de las Y_i un término positivo, que será el correspondiente a la media de

la distribución u_i , concretamente: $E(u) = \sqrt{\frac{2}{p}} \mathbf{s}_{u(N-SN)}$ en el caso de la especificación de errores Normal-Seminormal y $E(u) = \frac{1}{q(N-EXP)}$ en el caso de la especificación de errores Normal-Exponencial.

Las fronteras \hat{Y}_i modificadas (\hat{Y}_{im}) serán:

$$\hat{Y}_{im(N-SN)} = \hat{Y}_{i(M.C.O.)} + \left(\sqrt{\frac{p}{2}} \left(\frac{p}{p-4} \right) m_3^* \right)^{\frac{1}{3}} \quad \hat{Y}_{im(N-EXP)} = \hat{Y}_{i(M.C.O.)} + \left(-\frac{m_3^*}{2} \right)^{\frac{1}{3}}$$

donde $\hat{Y}_{(M.C.O.)}$ es la Y estimada por el método de M.C.O y $\hat{Y}_{im(N-SN)}$, $\hat{Y}_{im(N-EXP)}$ son las fronteras globales estimadas para la especificación Normal-Seminormal y Normal-Exponencial respectivamente. El valor numérico de dichas expresiones es:

$$\hat{Y}_{im(N-SN)} = \hat{Y}_{i(M.C.O.)} + 1.326(-m_3^*)^{\frac{1}{3}} \quad \hat{Y}_{im(N-EXP)} = \hat{Y}_{i(M.C.O.)} + 0.794(-m_3^*)^{\frac{1}{3}}$$

Con lo cual, para una misma muestra de valores Y_i , y para una misma serie de residuos resultados de la estimación minimocuadrática, tendremos que la corrección sobre las Y_i en la especificación Normal-Seminormal será 1.67 veces mayor que la efectuada sobre las Y_i en la especificación Normal-Exponencial.

Al cumplirse que $\hat{\mathbf{e}}_i = Y_i - \hat{Y}_{im}$; para ambas especificaciones obtenemos que:

$\hat{\mathbf{e}}_{i(N-SN)} = \hat{\mathbf{e}}_{i(N-EXP)} - 0.532(-m_3^*)^{\frac{1}{3}}$, con lo cual siempre ocurrirá que $\hat{\mathbf{e}}_{i(N-SN)} < \hat{\mathbf{e}}_{i(N-EXP)}$ dado que el momento de tercer orden debe ser negativo.

2.2 Estudio Teórico De Los Estadísticos De Forma.

A continuación, deduciremos las expresiones de la asimetría y curtosis de las distribuciones Seminormal y Exponencial, para hacernos una idea inicial de cómo variarán ambos coeficientes en las especificaciones de error Normal-Seminormal y Normal-Exponencial. Obviamente la intensidad de la influencia de las distribuciones Seminormal y Exponencial aumentará conforme lo haga la proporción de la componente u_i en el error ε_i , es decir, conforme aumente el valor del parámetro \mathbf{g} en las especificaciones. Estudiaremos los casos en que las componentes u_i sigan una distribución Seminormal y Exponencial. Las funciones de densidad de dichas distribuciones son respectivamente:

Los momentos de tercer y cuarto orden de dichas distribuciones son:

$$m_{3(SN)} = -\sqrt{\frac{2}{p}} \left(\frac{p-4}{p} \right) \mathbf{s}_u^3 \quad m_{3(EXP)} = \frac{2}{q^3}$$

$$m_{4(SN)} = \left(3 - \frac{4}{p} - \frac{12}{p^2} \right) \mathbf{s}_u^4 \quad m_{4(EXP)} = \frac{9}{q^4}$$

Recordando que las fórmulas para el cálculo de los coeficientes de asimetría (S) y curtosis (K) son respectivamente:

$$S = \frac{m_3}{\mathbf{s}^3} \quad K = \frac{m_4}{\mathbf{s}^4}$$

procederemos al cálculo de los mencionados coeficientes de ambas distribuciones, junto con el valor del estadístico Jarque Bera que vendrá dado por la expresión:

$$JB = \frac{T-k}{6} \left(S^2 + \frac{1}{4}(K-3)^2 \right)$$

Donde S es la asimetría, K la curtosis, y k representa el número de coeficientes estimados.

Los coeficientes de asimetría y curtosis de la distribución Seminormal resultan ser:

$$S_{(SN)} = -\sqrt{\frac{2}{p-2}} \left(\frac{p-4}{p-2} \right) = 0.995; \quad y \quad K_{(SN)} = \frac{3p^2 - 4p - 12}{p^2 - 2p + 4} = 0.665$$

Sustituyendo las expresiones de la asimetría y curtosis halladas anteriormente, en la expresión del estadístico Jarque-Bera, resulta que el valor de dicho estadístico para la distribución Seminormal es:

$$JB_{(SN)} = T \left[\frac{\frac{2}{p-2} \left(\frac{p-4}{p-2} \right)^2}{6} + \frac{\left(\frac{3p^2 - 4p - 12}{p^2 - 2p + 4} - 3 \right)^2}{24} \right] = 0.392 * T$$

Donde T es el número de valores de la muestra considerada.

A continuación procederemos a realizar un análisis homólogo al realizado anteriormente con la distribución Exponencial.

Así pues, sustituyendo la varianza de u en las fórmulas de la asimetría y curtosis, tenemos que el valor de dichos estadísticos para una distribución exponencial es:

$$S_{(EXP)} = \frac{\frac{2}{q^3}}{\frac{1}{q^3}} = \frac{2}{1} = 2 \quad K_{(EXP)} = \frac{\frac{9}{q^4}}{\frac{1}{q^4}} = 9$$

Sustituyendo las expresiones anteriores en la expresión del estadístico Jarque Bera, resulta:

$$JB_{(EXP)} = T \left(\frac{2^2}{6} + \frac{(9-3)^2}{24} \right) = T \left(\frac{4}{6} + \frac{36}{24} \right) = 2.167 * T$$

Donde T representa el número de valores de la muestra considerada.

Pasaremos a continuación a estudiar las fórmulas de los coeficientes de asimetría y curtosis de las especificaciones de errores ε , en los casos Normal-Seminormal y Normal-Exponencial. Para ello utilizaremos las expresiones de los momentos de tercer y cuarto orden de dichas especificaciones.

a) *Especificación Normal-Seminormal.*

Para el cálculo de la asimetría correspondiente a esta especificación sustituiremos el valor de la desviación típica de u elevada al cubo en la fórmula del momento de tercer orden, con lo que obtenemos:

$$m_{3(N-SN)} = \sqrt{\frac{2}{p}} \left(\frac{p-4}{p} \right) \left(\frac{pg}{p-2} \right)^{\frac{3}{2}} s_e^2$$

Dicha expresión será sustituida en la que nos define la asimetría, obteniendo:

$$S_{N-SN} = \sqrt{2} (p-4) \left(\frac{g}{p-2} \right)^{\frac{3}{2}}$$

Por otro lado, la curtosis de esta especificación resulta ser:

$$K_{N-SN} = 3(1-g^2) + \frac{(3p^2 - 4p - 12)g^2}{(p-2)^2}$$

El cálculo de la expresión que nos proporciona el valor de Jarque-Bera para esta especificación, lo llevaremos a cabo sustituyendo las expresiones de la asimetría y curtosis en la expresión que nos da el valor del estadístico Jarque-Bera resultando que:

$$JB_{(N-SN)} = T(0.1650943082g^3 + (3.147788274 \cdot 10^{-2})g^4)$$

b) *Especificación Normal-Exponencial.*

Para el cálculo de la asimetría de esta especificación, sustituiremos la expresión del momento de tercer orden de esta especificación y el valor de s_e^3 en la expresión que nos da la asimetría, con lo cual obtenemos que:

$$S_{(N-EXP)} = -2g^{\frac{3}{2}}$$

Seguidamente hallaremos la expresión de la curtosis para esta misma especificación. Para lo cual sustituiremos la expresión del momento de cuarto orden en función de s_e

$$m_{4(N-EXP)} = 9g^2 s_e^4 + 6g(1-g)s_e^4 + 3(1-g)^2 s_e^4 \text{ en la fórmula definitoria de la curtosis:}$$

$$K_{(N-EXP)} = \frac{9g^2 s_e^4 + 6g(1-g)s_e^4 + 3(1-g)^2 s_e^4}{s_e^4} = 6g^2 + 3$$

El cálculo de la expresión del estadístico Jarque-Bera se hallará sustituyendo las expresiones de la asimetría y curtosis para esta especificación en la expresión que nos define al estadístico, resultando que:

$$JB_{(N-EXP)} = T \left(\frac{\left(-2g^{\frac{3}{2}} \right)^2}{6} + \frac{(6g^2 + 3 - 3)^2}{24} \right) = \frac{T}{6} g^3 (4 + 9g)$$

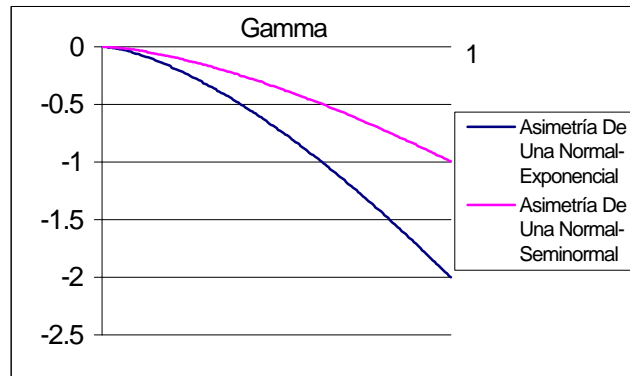
Seguidamente, avanzaremos en nuestro estudio relacionando las expresiones obtenidas de Asimetrías, Curtosis y Jarque-Bera Entre Ambas Especificaciones.

Relacionaremos las expresiones de asimetría dividiéndolas, resultando:

$$\frac{S_{(N-EXP)}}{S_{(N-SN)}} = 2.01$$

Con lo cual se constata que la asimetría de la especificación Normal-Seminormal será para todos los valores de g , 2.01 veces mayor que la asimetría de la especificación Normal-Seminormal.

Gráfico: 1. Representación del estadístico Asimetría correspondiente a ambas especificaciones.



Fuente: Elaboración propia.

En el gráfico 1 podemos observar las representaciones de las funciones encontradas para las asimetrías de las dos especificaciones en función de γ .

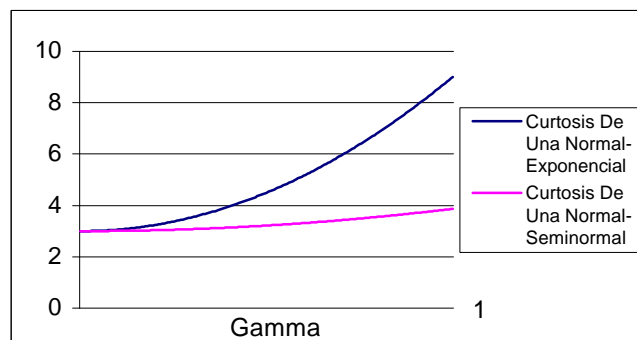
A continuación procederemos a la relación de las expresiones de la curtosis de ambas especificaciones, que una vez operado resulta:

$$\frac{K_{(N-EXP)}}{K_{(N-SN)}} = \frac{2g^2 + 1}{0.289725768g^2 + 1}$$

En este caso la relación no es fija, puesto que como vemos depende del valor de g considerado en cada ocasión. Así, para $g = 0$ resulta que los valores de la curtosis de ambas especificaciones son iguales, y para $g = 1$ la curtosis de la especificación Normal-Exponencial será aproximadamente 2.326 veces mayor que la de la Normal-Seminormal.

En el gráfico 2. se representan las curvas correspondientes a las expresiones halladas de la curtosis para cada especificación considerada.

Gráfico: 2. Representación del estadístico Curtosis correspondiente a ambas especificaciones.

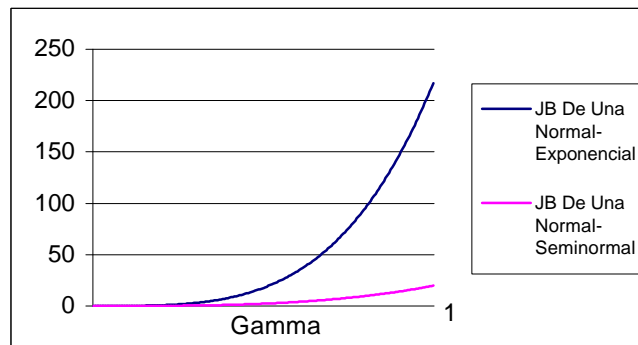


Fuente: Elaboración propia.

A continuación hallaremos la expresión que nos relaciona los valores de los estadísticos Jarque-Bera de ambas especificaciones. Para ello dividiremos las expresiones que nos dan los valores del estadístico en cuestión de ambas especificaciones, con lo que operando obtenemos:

$$\frac{JB_{(N-EXP)}}{JB_{(N-SN)}} = \frac{4 + 9g}{0.991 + 0.189g}$$

Gráfico: 3. Representación del estadístico Jarque-Bera (*JB*) correspondiente a ambas especificaciones.



Fuente: Elaboración propia.

Como podemos apreciar en el gráfico 3, (en el que se representan las expresiones del estadístico Jarque-Bera para muestras de 100) el valor del estadístico JB de la especificación Normal-Exponencial es mayor que el correspondiente a la especificación de errores Normal-Seminormal y esta diferencia aumenta conforme lo hace γ . Obviamente una excepción a lo anterior ocurre en valor $\gamma=0$ en el cual ambos estadísticos valen cero, ya que se trataría de una especificación simple y no compuesta, en otras palabras la distribución sería una Normal.

3. Estudio Empírico De Los Estadísticos De Forma.

Los estudios teóricos se han realizado sobre la base de que las dos especificaciones que con más frecuencia se aplican en el análisis de eficiencia mediante función frontera estocástica, podían tener estadísticos de forma muy distintos. Así, la relación de dichos estadísticos con el parámetro γ y las demás relaciones encontradas, proporcionan una base teórica importante. No obstante, debido a que en la práctica se trabaja con muestras de tamaño limitado, esto nos ha conducido a realizar el estudio empírico que aporta una información más próxima a la realidad.

Con el fin de estudiar la problemática planteada, se ha diseñado un experimento Monte Carlo tomando como fuente de variación el grado de ineficiencia del sistema y las distribuciones Normal-Exponencial y Normal-Seminormal como especificaciones de ϵ .

El experimento se plantea de la siguiente forma:

Se definen las especificaciones para ϵ , que son la Normal-Seminormal y Normal-Exponencial.

Se generan 5000 muestras para cada especificación y distintos valores de γ

El experimento se desarrolla con 2 fuentes de variación:

1) Especificación de los errores.

Se establece un proceso generador de datos bajo los supuestos:

a) Especificación de errores $\mathbf{e}_i = v_i - u_i$, donde la componente $v_i \in N(0, \mathbf{s}_v^2)$ y la $u_i \in N(0, \mathbf{s}_u^2)$

b) Especificación de errores $\mathbf{e}_i = v_i - u_i$, donde la componente $v_i \in N(0, \mathbf{s}_v^2)$ y la $u_i \in \text{Exp}(\mathbf{q})$

2) Proporción de las componentes v_i y u_i en el error total, \mathbf{e}_i .

Dicha proporción variará en función del valor del parámetro γ , acotado entre 0 y 1. Este parámetro mide la proporción de varianza de u contenida en la varianza total residual y se define del modo

$$\mathbf{g} = \frac{V(u)}{\mathbf{s}_e^2} = \frac{V(u)}{\mathbf{s}_v^2 + V(u)}$$

Para $\gamma=0$, \mathbf{e}_i estaría compuesto únicamente por la componente v_i y no existiría u_i . Esto supondría un valor de $\mathbf{s}_u^2=0$ con lo que el modelo no sería de frontera sino un Modelo Lineal General con error $N(0, \mathbf{s}_v^2)$. Sin embargo para el valor de $\gamma=1$ el error \mathbf{e}_i , sería únicamente u_i , y no existiría v_i . En este caso la frontera de producción que tendríamos sería determinística.

El experimento Monte Carlo se ha llevado a cabo para valores de γ que van de ($\gamma=0, 0.025, \dots, 1$).

Se ha tomado la determinación de establecer varios procesos de generación de distribuciones que se han plasmado en programas informáticos que son:

❖ Un programa generador de muestras de la componente v que sigue una distribución Normal: $v_i \in N(0, \mathbf{s}_v^2)$

❖ Un programa generador de muestras de la componente u que seguirá una distribución Seminormal (en el caso de la especificación de errores Normal-Seminormal) o bien, una Exponencial (en el caso de la especificación de errores Normal-Exponencial)

La generación de cada muestra de \mathbf{e} , se obtiene mediante la diferencia: $\mathbf{e}_i = v_i - u_i$; ($i=1, 2, \dots, 100$)

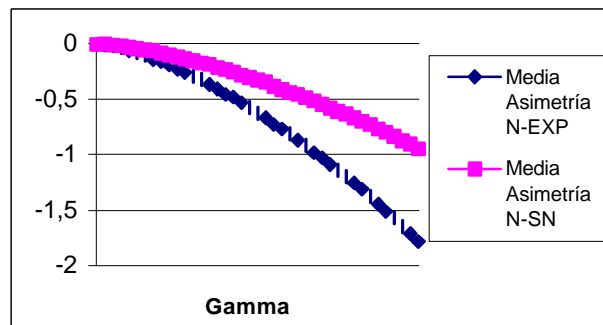
Después de llevar a cabo numerosos tests sobre series muestrales generadas, hemos convenido en que en todos los casos, la generación es lo suficientemente buena como para asegurar que las series de valores de \mathbf{e} seguirán efectivamente las distribuciones deseadas.

Para este estudio, se han generado 5000 series de 100 valores de \mathbf{e}_i . Para cada especificación se han considerado 41 valores de γ ($\gamma=0, 0.025, \dots, 1$) con el fin de aproximarnos lo más posible a cualquier valor de γ estimado de una muestra. A estas series se les ha calculado la asimetría, curtosis y Jarque-Bera. Los valores de estos estadísticos han sido recogidos en series, sobre las que se han calculado distribuciones de frecuencias y se les ha determinado la media y los cuantiles que dejan a la derecha e izquierda un 5% de los datos respectivamente. Este análisis descriptivo permite conocer la variabilidad de dichos estadísticos para muestras de tamaño 100.

Asimetría.

A continuación estudiaremos los resultados que sobre la asimetría hemos obtenido. Para ello representaremos los valores medios de las distribuciones empíricas de ambas especificaciones de errores, para así ver gráficamente como la curva correspondiente a la distribución Normal-Exponencial (N-EXP) se hace cada vez más negativa que la curva media de las distribuciones empíricas de la especificación Normal-Seminormal (N-SN), a medida que aumenta gamma que se representa en el eje de abscisas.

Gráfico: 4. Medias de las distribuciones de frecuencias de la asimetría



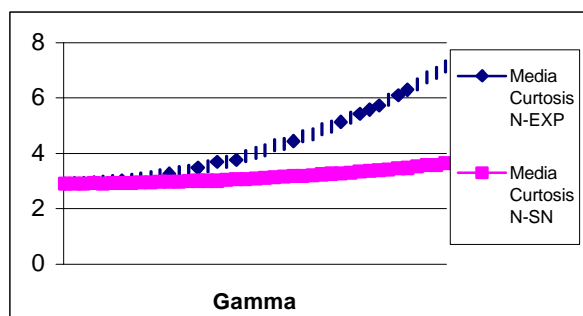
Fuente: Elaboración propia.

Los resultados empíricos obtenidos confirman las relaciones teóricas encontradas para la asimetría de ambas especificaciones. Esto nos da pie a concluir que este estadístico sería una buena medida para distinguir si una muestra de residuos $\hat{\epsilon}$ se parece más a una especificación que a la otra.

Curtosis.

En la gráfica 5 podemos ver con claridad, como la curva media de las distribuciones empíricas del estadístico curtosis correspondiente a la especificación de errores Normal-Exponencial (N-EXP) se distancia cada vez más de la curva media de las distribuciones empíricas de frecuencias de la curtosis correspondiente a la distribución Normal-Seminormal (N-SN).

Gráfico: 5. Medias de las distribuciones de frecuencias de la curtosis.



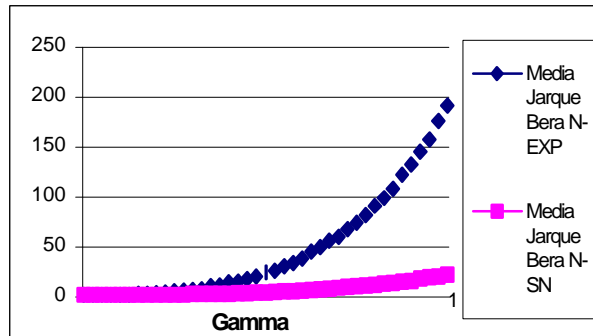
Fuente: Elaboración propia.

Del análisis anterior se deduce que al igual que la asimetría, la curtosis de una muestra de residuos $\hat{\epsilon}$ también se puede considerar un estadístico capaz de discriminar entre las dos especificaciones

Jarque Bera.

A continuación se representan las curvas correspondientes a la media de las distribuciones empíricas de cada observación, comprobándose como la curva de la especificación Normal-Seminormal (N-SN) se sitúa por encima de la correspondiente a la especificación Normal-Exponencial (N-EXP)

Gráfico: 6. Jarque-Bera medios de las distribuciones de frecuencias.

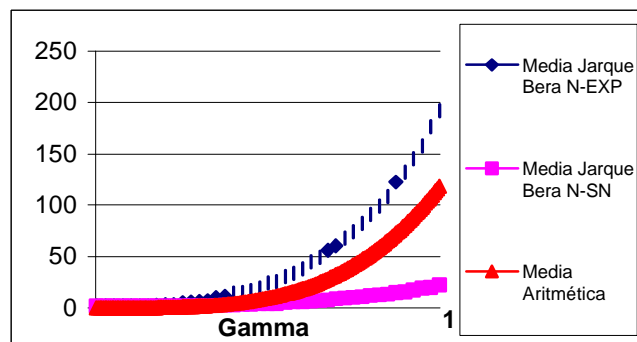


Fuente: Elaboración propia.

El estudio del estadístico Jarque-Bera se emprendió con objeto de aglutinar las medidas de asimetría y curtosis en una sola capaz de distinguir entre las dos especificaciones estudiadas. Se deduce de los resultados empíricos anteriores la confirmación de los teóricos, de modo que, a partir de un valor aproximado de $\gamma=0.4$ los valores de este estadístico para las dos especificaciones son tan distintos que pueden proporcionar una buena medida discriminante.

En la gráfica siguiente se representan las curvas medias de las distribuciones empíricas de frecuencias de los estadísticos Jarque-Bera obtenidos, así como la media aritmética de los valores de dichas curvas.

Gráfico: 7. Jarque-Bera medios de las dos distribuciones, Jarque-Bera media aritmética de las medias de las dos distribuciones.



Fuente: Elaboración propia

El cálculo de las medias aritméticas se efectúa en la búsqueda del punto discriminante para un γ concreto de una muestra. Estos resultados nos servirán para establecer un método que permita deducir cual fue el P.G.D.

4. Medida de discriminación.

Vista la gráfica 7, consideramos en un principio, que la curva que representa al Jarque Bera media aritmética de las dos especificaciones, podría constituir la referencia deseada. Así pues, si el valor del Jarque-Bera de la muestra es mayor que el del Jarque Bera de referencia significará que la muestra se ajusta mejor a una especificación de errores Normal-Exponencial, y si es menor a una Normal-Seminormal.

Con el fin de tomar una referencia de comparación que ayudara a discernir qué especificación de errores se ajusta mejor a una serie de residuos y que fallara lo menos posible, se representaron los valores medios de los estadísticos Jarque-Bera de cada especificación, los valores de los cuantiles anteriormente citados y los valores medios del estadístico de ambas especificaciones de errores. A la vista de lo anterior se llegó a la conclusión de que debido al solapamiento de regiones, la medida de referencia que constituye el Jarque Bera de las medias de las distribuciones empíricas de frecuencias del estadístico de ambas especificaciones podía ser objeto de un estudio de optimización .

Observada gráficamente la región de solapamiento de las distribuciones empíricas del estadístico Jarque-Bera y dada la multitud de intersecciones de curvas percentiles concluimos realizar una optimización de dichos puntos intersección.

Para lo cual se ha seguido el siguiente proceso:

- Se generan 10000 series de 100 datos para 101 valores propuestos de g ($g = 0, 0.01, 0.02, \dots, 1$) de las especificaciones Normal-Seminormal y Normal-Exponencial.
- A cada una de las 10000 series citadas anteriormente se les calcula el valor del estadístico de forma Jarque-Bera. Con lo cual tendremos 2 series de 10000 valores de Jarque-Bera para cada uno de los 101 valores de g considerados.
- A continuación obtenemos las curvas percentiles de la distribución del estadístico Jarque-Bera para las dos especificaciones consideradas.
- Seguidamente se procede al cálculo de las intersecciones de cada una de las curvas percentiles correspondientes a la especificación Normal-Seminormal con todas y cada una de las curvas percentiles de la especificación Normal-Exponencial.

En cada punto de intersección sabremos la probabilidad de acierto o fallo para cada una de las 2 especificaciones. Así, si contamos con una intersección de las curvas percentiles número i de la especificación Normal-Seminormal y j de la especificación Normal-Exponencial, nos indicará que la probabilidad de acierto en caso de que la verdadera especificación sea la Normal-Seminormal será $i/100$ mientras que la de cometer error sería de $(100-i)/100$. En el caso de que la verdadera especificación sea la Normal-Exponencial, la probabilidad de acierto será $(100-j)/100$ y la de error de $j/100$.

Dada la gran cantidad de puntos intersección obtenidos y teniendo en cuenta las características del método, seleccionaremos un punto intersección en cada intervalo (aproximadamente centesimal) de g siguiendo el criterio de tomar aquel punto que se derive de la intersección de curvas percentiles cuyo número corresponda conjuntamente a la menor curva de la Normal-Exponencial y mayor de Normal-Seminormal existentes en cada intersección (de forma que la posibilidad de acierto sea la mayor para cada intervalo). En caso de resultar varios puntos que nos proporcionen igual error, optaremos por aquel punto en el que el error se halle más repartido entre las dos especificaciones.

Una vez llevado a cabo la optimización se han obtenido los valores del JB correspondientes a las intersecciones óptimas. Así pues, para cada intervalo de g sabemos aproximadamente el error que cometeremos al considerar como más apropiada a la distribución de errores Normal-Seminormal (Normal-Exponencial) cuando en realidad la que más se ajusta a nuestro conjunto particular de datos es la especificación Normal-Exponencial (Normal-Seminormal).

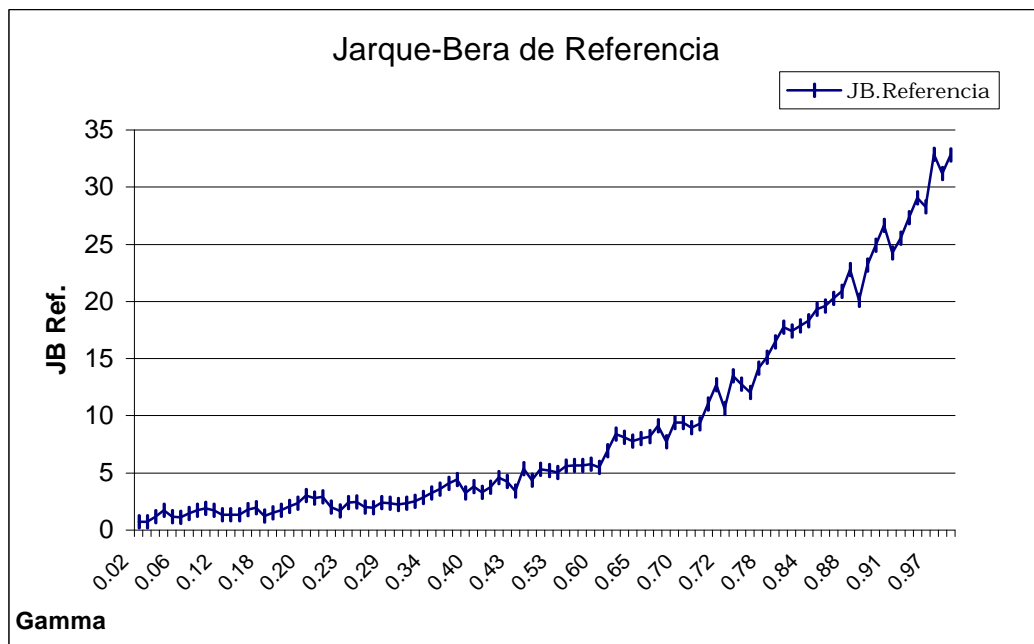
La curva de Jarque-Bera que se tomará como referencia en el método será la representada en la gráfica 8.

En la *tabla 1* (situada en el anexo) se expone la probabilidad estimada de acierto en los valores de gamma dados para las dos especificaciones consideradas, así como el valor numérico de cada uno de los puntos correspondientes al estadístico Jarque-Bera de referencia.

A la vista de dicha tabla se pone de manifiesto como al tomar esta medida de referencia, las probabilidades de acierto aumentan considerablemente conforme crece el valor de γ .

Como vemos se pasa de una probabilidad de acierto aproximada del 50.5% para $\gamma=0.024$ hasta el 84% para $\gamma=0.985$.

Gráfico: 8. Jarque-Bera de referencia obtenido para la comparación.



Fuente: Elaboración propia.

5. Conclusiones.

A continuación se describen las conclusiones generales más relevantes sobre los diferentes estudios llevados a cabo y los derivados de los resultados del experimento Monte Carlo.

- Tanto la asimetría como la curtosis y el estadístico de Jarque-Bera se pueden expresar en función de γ para las dos especificaciones.
- La asimetría, la curtosis y el estadístico Jarque-Bera para la especificación Normal-Exponencial es siempre mayor que para la especificación Normal-Seminormal.
- Cuando estimamos S_u^2 por el método de los momentos del vector de residuos minimocuadráticos, existe una relación teórica entre dichas estimaciones para las dos especificaciones siendo la de la Normal-Exponencial menor la de la Normal-Seminormal.

- Los resultados empíricos obtenidos confirman las relaciones teóricas encontradas para la asimetría de ambas especificaciones. Esto nos da pie a concluir que este estadístico sería una buena medida para distinguir si una muestra de residuos \hat{e} se parece más a una especificación que a la otra.
- A la vista de los resultados se deduce que el método puede ser válido para el diferenciamiento de ambas especificaciones para valores altos de γ .

6. Referencias Bibliográficas.

AIGNER, D., KNOX LOVELL, C.A. y SCHMIDT, P. (1977). "Formulation and estimation of stochastic frontier production models," *Journal of Econometrics* 6 : 21-37.

ALLAUDIN, M, SQUIRES.D. y TISDELL .C. (1993) "Divergency between average and frontier production technologies: an empirical investigation for Bangladesh" *Applied Economics* nº25: 379-388.

BATESSE, G. E. (1992). "Frontier production functions and technical efficiency: a survey of empirical applications in agricultural economics," *Agricultural Economics*, 7:185-208.

BATTESE, G.E. y CORRA, G.S. (1977). "Estimation of a production frontier model: with application to the pastoral zone of eastern Australia" *Aust. J. Prod. Anal.*,3: 153-169.

BYRON J.T. MORGAN.(1984) "Elements of simulation" Science Paperbacks Chapman and Hall. London.

DIOS, R. (1998). "El modelo lineal sin término independiente y el coeficiente de determinación. Un Estudio Monte Carlo," *Qüestió*, vol. 22, 1: 3-37.

DIOS, R., (1999), " Análisis de interpretación de los parámetros de relación de varianzas en el modelo de frontera estocástica" Documento de Trabajo 99.02 Dto de Estadística. Universidad de Córdoba.Córdoba.

GREENE, W.H. (1992). "Limdep Version 6.0. User´s Manual and Reference Guide," Ed. Econometric Software, Inc. Bellport, New York.

JONDROW, J., KNOX LOVELL, C.A., MATEROV, I. S. y SCMIDT, P. (1982). "On the estimation of technical inefficiency in the stochastic frontier production function model," *Journal of Econometrics*, 19: 233-238.

MEEUSEN, W. y VAN DEN BROECK, J. (1977). "Efficiency Estimation from Cobb-Douglas Production functions with composed error," *International Economic Review*. Vol. 18 Nº. 2, Junio: 435-455.

OLSON, J. A., SCHMIDT, P. y WALDMAN, D. M. (1980). "A Monte Carlo study of estimators of stochastic frontier production function," *Journal of Econometrics* , 13: 67-82.

ROSCOE, J. T. y BYARS, J.A. (1971). "An investigation of the restraints with respect to sample size commonly imposed on the use of the chi-square statistic". *Journal of the American Statistical Association* 66: 755-759.

TAYLOR, T.G. y SHONKWILER, J.S. (1984) "Alternative stochastic specifications of the frontier production function in the analysis of agricultural credit programs and technical efficiency". *Journal of Development Economics*, Nº21: 149-160.

Anejo:

Tabla1. Valores del Jarque Bera de referencia resultante y de sus correspondientes del acierto medio y gamma asociados

J.B. Ref.	Gamma	acierto medio	J.B. Ref.	Gamma	acierto medio
0.739229865	0.023644185	50.5	5.223665155	0.525534852	69
0.740360307	0.024544885	50.5	5.024305122	0.534718186	68.5
1.202313757	0.037089496	50.5	5.572180387	0.540991653	70.5
1.763550599	0.043634706	50.5	5.629765445	0.55130675	73.5
1.201431571	0.050769488	50.5	5.669610083	0.581506203	71
1.145325485	0.059346475	50.5	5.769812188	0.586921321	71
1.430148845	0.070164022	50.75	5.504185377	0.596221674	70.5
1.714972206	0.080981569	51	6.950919966	0.608309614	73
1.88128455	0.096653634	52.5	8.397654555	0.620397554	76
1.710890711	0.116632455	51	8.103337123	0.620956833	74.5
1.331478062	0.120138754	51.5	7.809019691	0.632985973	72.5
1.345007955	0.126668816	51	7.981366773	0.654502697	73.5
1.341818623	0.127668816	51.5	8.153713855	0.67601942	74
1.790005424	0.166581668	51.5	9.129565784	0.677940068	73.5
1.969804828	0.168988087	52.5	7.746213602	0.694078233	74
1.23449588	0.178173602	51.5	9.386454678	0.701603462	72.5
1.516182567	0.180360236	53	9.428976234	0.702011462	75
1.758724466	0.189182857	52.5	8.943875213	0.707328197	75
2.0548791	0.190889196	53.75	9.322885984	0.708344695	75.5
2.351033734	0.192595536	55	11.02891341	0.716346363	76.5
3.019204491	0.204726239	55.5	12.73494084	0.724348032	77
2.824534737	0.219761964	54	10.6263378	0.724949139	76
2.909879603	0.220235975	60.5	13.49096933	0.728996499	77.5
2.000532484	0.221783957	55	12.76675492	0.745654636	78
1.667902616	0.224726239	55.5	12.0425405	0.762312772	76.5
2.401491226	0.226287945	60	14.18126049	0.775348562	78
2.475389827	0.241548211	58	15.12417194	0.783214573	78.5
2.037259122	0.245526751	56.5	16.44039037	0.787196679	79
1.975330709	0.252221229	56	17.75660879	0.791178785	79.5
2.380262934	0.272462767	57.5	17.39536678	0.793901067	79.5
2.353975071	0.286699382	58	17.86190882	0.817707888	79.5
2.217538922	0.290788105	61.5	18.32845085	0.841514709	80
2.365316998	0.291892911	61	19.31256788	0.845452624	79
2.513095073	0.292997718	60	19.62105259	0.860384291	81
2.882748979	0.317151583	61	20.26735346	0.869819082	81.5
3.252402884	0.341305448	62	20.91365433	0.879253874	81.5
3.583220726	0.347329597	63	22.79845071	0.88058723	81.5
4.102251368	0.369772318	64	20.10702279	0.891843652	81
4.441586865	0.387741124	66.5	23.18194419	0.892618755	82
3.236425744	0.397661223	64	24.92493421	0.903405436	82.5
3.809561181	0.398490266	65.5	26.66792424	0.914192118	83
3.326760966	0.404676747	63	24.25704917	0.914265749	82.5
3.748628088	0.408168175	65	25.53520005	0.926274849	83
4.606580761	0.41304875	69.5	27.30014619	0.939527715	83
4.251406456	0.422294579	67.5	29.06509233	0.95278058	83
3.431590701	0.426862373	65.5	28.26630577	0.954932803	83
5.360269362	0.459799151	70	32.88382556	0.971207303	84
4.34139012	0.468717992	68.5	31.19019914	0.976752526	84
5.346676196	0.482675585	68.5	32.79346814	0.985013597	84