

DIFICULTADES DEL ESTIMADOR MÁXIMO VEROSÍMIL EN EL MODELO DE PRODUCCIÓN HALF-NORMAL CON FRONTERA ESTOCÁSTICA

FCO. JAVIER ORTEGA IRIZO

e-mail: fjortega@us.es

JOSÉ MANUEL GAVILÁN RUIZ

e-mail: gavi@us.es

JOSÉ ANTONIO CAMÚÑEZ RUÍZ

e-mail: camunez@us.es

Departamento de Economía Aplicada I. UNIVERSIDAD DE SEVILLA
Avenida Ramón y Cajal, nº1. 41018-Sevilla.

RESUMEN

El uso del método de máxima verosimilitud para estimar modelos de producción Half-Normal con frontera estocástica, conlleva algunas dificultades prácticas que tal vez no han sido suficientemente enfatizadas. En consecuencia, los resultados obtenidos por los métodos numéricos implementados en el software desarrollado en este área, se deberían tomar con cautela. Además, tales problemas pueden aparecer en una proporción importante de muestras, por lo que no es extraño encontrarse con ellos al usar este tipo de modelos.

Palabras clave: Modelos de producción, Frontera estocástica, Estimador máximo verosímil.

Área temática: Métodos cuantitativos

ABSTRACT

Using the method of maximum likelihood in order to estimate Half-Normal stochastic frontier productions models entails several practical difficulties that perhaps have not been emphasised enough. As a consequence, the results obtained from the numerical methods implemented in the developed software in this subject, should be taken with caution. Moreover, such problems can appear in an important proportion of samples, and then it is not surprising to come across them when using this kind of models.

Key words: Production models, Stochastic Frontier, Maximum likelihood estimator.

Thematic Area: Quantitative Methods

DIFICULTADES DEL ESTIMADOR MÁXIMO VEROSÍMIL EN EL MODELO DE PRODUCCIÓN HALF-NORMAL CON FRONTERA ESTOCÁSTICA

1. INTRODUCCIÓN

El origen del análisis de modelos econométricos de producción con frontera y el cálculo de medidas de eficiencia puede situarse en el trabajo de Farrell (1957), en el que se estableció la novedosa idea de analizar la eficiencia de un proceso productivo en términos de las desviaciones observadas entre la producción real y una frontera ideal de output máximo. En términos econométricos, dichas desviaciones podrían ser identificadas con perturbaciones aleatorias en un modelo de regresión.

Siguiendo la iniciativa de Farrel, Aigner and Chu (1968) proponen un modelo en el que se expresa un determinado output en función de una serie de inputs y parámetros desconocidos más una perturbación aleatoria negativa; la parte determinista del modelo representa la frontera de producción o valor máximo alcanzable de outputs para unos inputs dados, mientras que la perturbación aleatoria (diferencia entre la producción real y la máxima posible) representaría el grado de ineficiencia en el proceso productivo. Alternativamente, podemos utilizar la formulación dual de este problema, en la que la variable dependiente es el coste de producción y la perturbación aleatoria sería positiva, representando la parte determinista del modelo el coste mínimo de producción para unos inputs dados.

Esta formulación, denominada habitualmente modelo de producción con frontera determinista, presenta dificultades de estimación, debido al hecho de que la modelización de la perturbación unilateral rompe las hipótesis habituales de regularidad asumidas para la obtención de las propiedades asintóticas del estimador máximo verosímil (EMV) (Ortega y Basulto, 2009; Ortega et al. 2009). Otro inconveniente que se ha destacado acerca de este modelo es su gran sensibilidad a la presencia de outliers y/o valores extremos (Simar, 2007). Para evitar esta cuestión, se han propuesto métodos de estimación que permiten dejar un cierto porcentaje de observaciones “por encima” de la frontera estimada (Cazals et al., 2002 y Daouia and Simar, 2004). Pero sobre todo se ha argumentado que esta modelización no tiene en cuenta ninguna fuente de errores de

medida ni ningún otro tipo de variaciones aleatorias que serían achacadas a ineficiencia en el proceso productivo.

Tratando de dar respuesta a este último problema, Aigner et al. (1977), Battese and Corra (1977) y Meusen and van den Broeck (1977) formularon modelos econométricos con término de error compuesto, que actualmente conocemos por modelos de producción con frontera estocástica. En ellos, se introducen dos perturbaciones; una es una medida de la ineficiencia, mientras que la otra refleja todas las posibles fuentes de variación aleatorias.

En estos modelos, se verifican las condiciones de regularidad habituales, por lo que podemos aplicar las propiedades asintóticas del estimador máximo-verosímil para llevar a cabo nuestras inferencias. Además, el desarrollo de software específico (como FRONTIER y LIMDEP) ha hecho que el uso de los modelos de frontera estocástica esté en la actualidad ampliamente difundido.

La obtención del estimador máximo verosímil (EMV) se basa en el uso de algoritmos numéricos de maximización de funciones (tipo gradiente, Newton-Rapson, etc.), ya que no es posible obtener expresiones explícitas para calcular el máximo de la función de verosimilitud. Es decir, en definitiva se trata de resolver numéricamente un problema de maximización.

El objetivo de este trabajo es poner de manifiesto que en este caso se presentan algunas dificultades prácticas que tal vez no han sido enfatizadas hasta ahora suficientemente y que hacen que debamos analizar con cautela los resultados arrojados por los métodos de cálculo numérico implementados en el software desarrollado al efecto. Además, dichos problemas pueden aparecer en una proporción importante de muestras, por lo que no es extraño que nos encontremos con ellos al trabajar con este tipo de modelos.

La primera dificultad, que ha recibido mayor atención en la literatura especializada, aparece cuando los residuos obtenidos por el método de Mínimos Cuadrados muestran asimetría contraria a la asimetría teórica de la distribución del término de perturbación compuesto. En este caso, decimos que los residuos muestran “wrongskewness” o “asimetría errónea” (es decir asimetría positiva en el caso de modelos de producción o asimetría negativa en el caso de modelos de coste). La práctica más común en esta situación ha consistido en desechar el modelo, bajo el argumento de que este hecho es un

síntoma de que está mal especificado o se está usando de una manera inapropiada. Sin embargo, estudios simulados han puesto de manifiesto que en el modelo con frontera estocástica la asimetría “errónea” puede aparecer con probabilidades no despreciables y que en ciertos casos se necesita un tamaño muestral superior a 1000 para que esta probabilidad descienda al 5% (Simar, 2007).

Al aplicar el método de máxima verosimilitud sobre una muestra con asimetría errónea obtenemos que nos encontramos en una situación extrema en la que toda la variabilidad del término de error compuesto se debe a motivos aleatorios y por tanto hay ausencia de ineficiencia. Es decir, este método de estimación nos dará una respuesta incorrecta en las muestras que, aún procediendo de un modelo de frontera estocástica, presenten asimetría errónea.

La otra dificultad aparece cuando el máximo de la función se alcanza en el otro caso extremo, es decir, cuando el resultado de la estimación indica que la totalidad del término de error compuesto se debe a ineficiencia y por tanto hay ausencia de factores aleatorios. En este caso incluso hay discrepancia de resultados entre LIMDEP 9.0 (que detiene el proceso de estimación con un mensaje de error en el que se indica que la matriz de covarianzas es singular) y FRONTIER 4.1 (que ofrece las estimaciones de los parámetros y su matriz de covarianzas, aunque podemos comprobar que dicha matriz es “casi singular”). Como veremos en nuestro estudio simulado, la proporción de muestras procedentes de un modelo con error compuesto en las que el método de máxima verosimilitud estima ausencia de factores aleatorios tampoco es despreciable.

Señalar también que, aparte de LIMDEP 9.0 y FRONTIER 4.1, hemos utilizado el software Mathematica 5.2 para analizar y maximizar la verosimilitud del modelo. Otro problema añadido que hemos encontrado es que con relativa frecuencia nos han aparecido mensajes de error de “overflow”, “underflow” o “indeterminate” al tratar de evaluar la función de verosimilitud en diferentes zonas del espacio paramétrico, aún trabajando con niveles de precisión bastante elevados. Usando la instrucción *FindMaximun* de dicho software en estos casos no es posible obtener el máximo de la función. En definitiva, tenemos varios indicios de que nos encontramos ante una función difícil de manejar no solo desde el punto de vista algebraico, sino también desde el punto de vista numérico.

A partir de aquí, en la sección 2 presentamos el modelo de producción, la función de verosimilitud bajo las hipótesis asumidas para el término de perturbación y las parametrizaciones utilizadas por LIMDEP 9.0 y FRONTIER 4.1 para llevar a cabo el proceso de estimación; en la sección 3 consideramos en detalle el problema de “asimetría errónea”; en la sección 4 analizamos el problema de “ausencia de factores aleatorios”; en la sección 5 presentamos los resultados del proceso de simulación llevado a cabo para analizar en qué proporción de casos podemos encontrarnos frente a ambos problemas. Por último, en la sección 6 exponemos las conclusiones de nuestro trabajo.

2. PLANTEAMIENTO DEL MODELO.

La formulación básica del modelo de producción con frontera estocástica es:

$$y_i = f(x_i, \beta) + v_i - u_i, \quad i = 1, \dots, n,$$

ε_i

donde y_i es la producción de la firma i -ésima, x_i es el vector de todos sus inputs, β es un vector de parámetros a estimar y $f(\cdot)$ es la función de producción.

La perturbación aleatoria ε_i se compone de dos partes (de ahí que algunos autores hablen de modelo de error compuesto), $v_i \in \mathbb{R}$ que recoge las fuentes de variación aleatorias y $u_i > 0$ que recoge la ineficiencia en el proceso productivo. Comúnmente se supone una distribución Normal para v_i , mientras que para u_i hemos de elegir una distribución de probabilidad positiva. Señalemos que la formulación del modelo de producción con frontera determinista, propuesta anteriormente en Aigner and Chu (1968) es $y_i = f(x_i, \beta) + (-\varepsilon_i)$, donde $\varepsilon_i > 0$ recoge la ineficiencia en el proceso productivo y no hay cabida a las fuentes de variación aleatorias.

Las perturbaciones v_i que recogen los efectos aleatorios siguen una distribución Normal, concretamente, $v_i \sim N(0, \sigma_v^2)$. Con respecto a las perturbaciones u_i que recogen el efecto de la ineficiencia en la producción, supondremos la hipótesis más habitual, es

decir, que $u_i = |u_i^*|$, donde $u_i^* \sim N(0, \sigma_u^2)$. Por definición, diremos que las perturbaciones u_i siguen una distribución Half-Normal, lo cual representamos por $u_i \sim HN(0, \sigma_u^2)$. Adicionalmente, se supone que todas las perturbaciones (tanto v_i como u_i) son independientes. Es preciso señalar que se han considerado con frecuencia otro tipo de densidad de probabilidad para las perturbaciones u_i , entre las que cabe destacar las distribuciones Exponencial y Gamma.

Con respecto a la función de producción, consideraremos el caso más usual, es decir, $f(x_i, \beta) = x_i' \beta$, donde β contendrá un término constante. Inicialmente, tanto los inputs como los outputs pueden ser medidos en unidades absolutas o logarítmicas, ya que este aspecto no influirá en el proceso de estimación de los parámetros del modelo, aunque sí habría que tenerlo en cuenta, en su caso, para estimar la eficiencia individual de cada firma.

Los programas LIMDEP 9.0 y FRONTIER 4.1 utilizan reparametrizaciones alternativas para calcular los EMVs y presentar los resultados del modelo estimado.

En LIMDEP 9.0 se consideran los parámetros $\sigma^2 = \sigma_v^2 + \sigma_u^2$ y $\lambda = \sigma_u / \sigma_v$. De este modo, $\sigma^2 > 0$ es la varianza total del término de error compuesto, ya que se supone independencia entre v y u , mientras que $\lambda^2 \geq 0$ se interpreta como una proporción entre la varianza debida a ineficiencia y la varianza debida a factores aleatorios. Con estos parámetros, el logaritmo de la verosimilitud para una muestra individual viene dado por:

$$\log(L_i) = \frac{1}{2} \log\left(\frac{2}{\pi}\right) - \log(\sigma) - \frac{1}{2} \left(\frac{y_i - x_i' \beta}{\sigma}\right)^2 + \log\left(\Phi\left(\frac{-\lambda(y_i - x_i' \beta)}{\sigma}\right)\right).$$

Por su parte, FRONTIER 4.1 considera la parametrización $\sigma^2 = \sigma_v^2 + \sigma_u^2$ y $\gamma = \sigma_u^2 / \sigma^2$. Así, $0 \leq \gamma < 1$ se interpreta como una proporción entre la varianza debida a ineficiencia y la varianza total. En este caso, para obtener la expresión del logaritmo de la verosimilitud bastaría sustituir en la fórmula anterior λ por $\sqrt{\gamma/(1-\gamma)}$.

Resaltemos que en realidad λ^2 y γ no son exactamente proporciones de varianzas, ya que $\text{var}[u] = [(\pi - 2)/\pi] \sigma_u^2$ y así para obtener las proporciones deberíamos corregir con la constante $(\pi - 2)/\pi$.

3. EL PROBLEMA DE ASIMETRÍA ERRÓNEA (WRONGSKWENESS).

El término de error compuesto del modelo de producción Half-Normal tiene asimetría negativa (Green, 1993). El problema de asimetría errónea se presenta cuando los residuos estimados a través del método de mínimos cuadrados ordinarios (MCO) presentan un momento muestral de orden 3 positivo. Es decir, llamando e_i a los residuos MCO, el problema aparece cuando $m_3 = (1/n) \sum_i e_i^3 > 0$, siendo Waldman (1982) el primero en destacar este inconveniente. En este caso, los métodos numéricos llegan a la solución $\hat{\lambda} = 0$ (o bien $\hat{\sigma}_u^2 = \hat{\gamma} = 0$) coincidiendo el EMV de β con el estimador MCO. Así, la conclusión es que en el modelo inicialmente planteado “sobra” el término de perturbación que recoge la ineficiencia, que el estimador óptimo es el de MCO y que todas las firmas están en la frontera de eficiencia.

Este hecho ha sido considerado como una prueba de inconsistencia del modelo, aconsejándose cambiar la especificación o aumentar el número de observaciones ante esta situación. Es preciso señalar que la conclusión que obtiene Waldman (1982) es que cuando hay asimetría errónea la solución señalada anteriormente es un máximo local y que la evidencia empírica mostrada en Olsen et al. (1980) “sugiere” que también es el máximo global. Actualmente algunos autores sugieren efectuar diferentes búsquedas del máximo considerando una amplia variedad de puntos de partida para los métodos numéricos.

A continuación, vamos a considerar un ejemplo numérico para analizar las respuestas ofrecidas por LIMDEP 9.0 y FRONTIER 4.1 ante el problema de wrongskweness.

Ejemplo 1: Hemos considerado una muestra de tamaño $n = 20$ simulada a partir de un modelo de frontera estocástica con una única variable explicativa y ordenada en el origen. Los valores que se han tomado para los parámetros han sido $\beta_0 = 1$, $\beta_1 = 2$, $\sigma_u = 1$,

$\sigma_v = 1$. Una vez simulada la muestra, hemos considerado solo la parte entera de los datos obtenidos. Los valores concretos de la muestra se ofrecen en el apéndice.

Las estimaciones MCO son $\hat{\beta}_0 = -1.0372$ y $\hat{\beta}_1 = 2.0827$, mientras que el tercer momento de los residuos MCO resulta ser $m_3 = 0.3035$.

Al utilizar LIMDEP 9.0 para estimar el modelo Half-Normal con frontera estocástica para estos datos, primero aparece un mensaje con el siguiente error:

“Error 315: Stoch. Frontier: OLS residuals have wrong skew. OLS is MLE.”

Tras pulsar “OK”, la salida que ofrece el programa es la que se muestra a continuación.

```
Error 315: Stoch. Frontier: OLS residuals have wrong skew. OLS is MLE.
WARNING! OLS residuals have the wrong skewness for SFM
Other forms of the model models may also behave poorly.
In this case, one MLE for the half normal model is OLS
for beta and sigma and zero for the inefficiency term.
Warning 141: Iterations:current or start estimate of sigma is nonpositiv
Warning 141: Iterations:current or start estimate of sigma is nonpositiv
Normal exit from iterations. Exit status=0.
```

```
+-----+
| Limited Dependent Variable Model - FRONTIER |
| Maximum Likelihood Estimates                |
| Model estimated: Mar 09, 2010 at 04:52:17PM. |
| Dependent variable                          Y |
| Weighting variable                          None |
| Number of observations                       20 |
| Iterations completed                        24 |
| Log likelihood function                      -31.02107 |
| Number of parameters                        4 |
| Info. Criterion: AIC =                      3.50211 |
| Finite Sample: AIC =                       3.63544 |
| Info. Criterion: BIC =                      3.70125 |
| Info. Criterion:HQIC =                     3.54098 |
| Variances: Sigma-squared(v)=                1.30243 |
|           Sigma-squared(u)=                 .00000 |
|           Sigma(v) =                       1.14124 |
|           Sigma(u) =                       .00013 |
| Sigma = Sqr[(s^2(u)+s^2(v))]=              1.14124 |
| Stochastic Production Frontier, e=v-u.      |
+-----+
+-----+-----+-----+-----+-----+-----+
|Variable| Coefficient | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+-----+-----+-----+-----+-----+-----+
-----+Primary Index Equation for Model
Constant| -1.03707273 | 74.1483059 | -.014 | .9888 |
X       | 2.08270107 | .09940750 | 20.951 | .0000 | 7.10000000
-----+Variance parameters for compound error
Lambda  | .00011512 | 81.4258036 | .000 | 1.0000 |
Sigma   | 1.14123972 | .04866113 | 23.453 | .0000 |
```

Lo más destacable es que las estimaciones ofrecidas para el parámetro β son las MCO y que $\hat{\sigma}_u^2$ es cero con cinco dígitos significativos, es decir, que tal como indica el mensaje de error, la solución obtenida es que el estimador MCO es el EMV (“OLS is MLE”).

Por su parte, la salida (resumida) que proporciona el programa FRONTIER 4.1 es:


```

Output from the program FRONTIER (Version 4.1c)

Error Components Frontier (see B&C 1992)
The model is a production function
The dependent variable is not logged

the ols estimates are :
      coefficient      standard-error      t-ratio
beta 0      -0.10371775E+01    0.79110716E+00  -0.13110456E+01
beta 1       0.20827011E+01    0.10478470E+00   0.19876003E+02
sigma-squared 0.14471421E+01
log likelihood function = -0.31021072E+02

the estimates after the grid search were :
beta 0      -0.83024617E+00
beta 1       0.20827011E+01
sigma-squared 0.13452485E+01
gamma       0.50000000E-01
mu is restricted to be zero
eta is restricted to be zero

the final mle estimates are :
      coefficient      standard-error      t-ratio
beta 0      -0.10326477E+01    0.43841732E+01  -0.23553990E+00
beta 1       0.20827011E+01    0.97988916E-01   0.21254456E+02
sigma-squared 0.13024488E+01    0.41190460E+00   0.31620157E+01
gamma       0.24578015E-04    0.50712482E-01   0.48465416E-03

log likelihood function = -0.31021072E+02

the likelihood value is less than that obtained
using ols! - try again using different starting values

```

Como podemos apreciar, la estimación de β también coincide con la estimación MCO y $\hat{\gamma} = 0.000025$ en consonancia con la respuesta ofrecida por LIMDEP 9.0. En este caso el mensaje de aviso que aparece es “the likelihood value is less than that obtained using ols! - try again using different starting values”, es decir, FRONTIER 4.1 indica la posibilidad de mejorar el valor de la verosimilitud probando con valores iniciales distintos a los usados por defecto.

Usando la instrucción *FindMaximun* del software Mathematica 5.2 y limitando el recorrido del parámetro λ a valores no negativos, con la parametrización considerada en LIMDEP 9.0 obtenemos también que la estimación de β coincide con MCO, $\hat{\sigma} = 1.1412$ y $\hat{\lambda} = 0$, siendo el valor del logaritmo de la verosimilitud en este punto igual a -31.0211 . Es interesante observar que si no se restringe el recorrido de λ el máximo se alcanza en $\tilde{\lambda} = -2.7015 \times 10^{-9}$, si bien el valor de la verosimilitud, con cuatro decimales significativos, sigue siendo el mismo.

4. EL PROBLEMA DE AUSENCIA DE EFECTOS ALEATORIOS.

Este problema aparece porque, en ocasiones, la función de verosimilitud es estrictamente creciente en la dirección del parámetro λ , alcanzándose así el máximo en “ $\hat{\lambda} = +\infty$ ”. Si usamos la parametrización de FRONTIER, el máximo se daría en $\hat{\gamma} = 1$, punto en el que la función de verosimilitud no está definida. Así, el comportamiento de la verosimilitud nos estaría indicando que en el término de error compuesto “sobra” el término de perturbación que recoge los factores aleatorios, y que por tanto estamos ante un modelo de frontera determinista.

En principio, esta situación no debería suponer ningún problema, pues podríamos tomar como estimador de λ el valor $\hat{\lambda} = +\infty$ (o bien $\hat{\gamma} = 1$), calculando el valor del logaritmo de la verosimilitud y estimando la matriz de varianzas covarianzas como valores límites cuando $\lambda \rightarrow +\infty$ (o bien cuando $\gamma \rightarrow 1$). La dificultad estriba en que los resultados que se obtienen son muy sensibles a los errores de redondeo, y además la matriz de varianzas-covarianzas estimada es singular. Esto provoca que las respuestas de LIMDEP 9.0 y FRONTIER 4.1 ante esta situación sean diferentes, pues el primero detiene el proceso de estimación mostrando un mensaje de error mientras que el segundo no muestra ninguna advertencia y simplemente ofrece la estimación obtenida, con un valor de $\hat{\gamma}$ muy cercano a 1.

Para ilustrar esta situación, vamos a considerar un ejemplo concreto, para analizar las respuestas ofrecidas por ambos programas.

Ejemplo 2: Vamos a considerar los datos acerca de telecomunicaciones en 21 países que se recogen en el ejemplo ofrecido en Coelli et al. (1998, p.193) y que pueden consultarse en el apéndice.

En este ejemplo, la producción se mide a través de un índice de producción y tenemos una ordenada en el origen más dos factores explicativos que son la longitud total de líneas instaladas (en miles de kilómetros) y el número de trabajadores (en 10^5 personas). Tanto el output como los inputs se consideran en logaritmos.

Al utilizar LIMDEP 9.0 para estimar el modelo Half-Normal con frontera estocástica para estos datos, obtenemos el mensaje de error:

“Error 143: Models – estimated variance matrix of estimated is singular”

Tras pulsar “OK”, la salida que ofrece el programa es la que se muestra a continuación.

```
--> FRONTIER;Lhs=LY;Rhs=ONE,LLINES,LEMPLOY$
Maximum iterations reached. Exit iterations with status=1.
Error 143: Models - estimated variance matrix of estimates is singular
```

Por su parte, la salida (resumida) que proporciona el programa FRONTIER 4.1 calculando el logaritmo de inputs y output con 6 decimales de precisión es:

```
Output from the program FRONTIER (Version 4.1c)

the ols estimates are :
      coefficient      standard-error      t-ratio
beta 0      -0.16368877E+00      0.17167025E+00      -0.95350691E+00
beta 1       0.63271391E+00      0.29752671E+00      0.21265785E+01
beta 2       0.22442804E+00      0.34443999E+00      0.65157369E+00
sigma-squared 0.15973431E+00

log likelihood function = -0.89195705E+01

the estimates after the grid search were :
beta 0       0.29404501E+00
beta 1       0.63271391E+00
beta 2       0.22442804E+00
sigma-squared 0.34643533E+00
gamma        0.95000000E+00
mu is restricted to be zero
eta is restricted to be zero

the final mle estimates are :
      coefficient      standard-error      t-ratio
beta 0       0.15587831E+00      0.34236457E-01      0.45529919E+01
beta 1       0.10041642E+01      0.33700206E+00      0.29796974E+01
beta 2       -0.13155514E+00      0.35497673E+00      -0.37060214E+00
sigma-squared 0.31704973E+00      0.50224852E-01      0.63126067E+01
gamma        0.99999998E+00      0.27599009E-04      0.36233184E+05
mu is restricted to be zero
eta is restricted to be zero
log likelihood function = -0.37778683E+01

covariance matrix :
0.11721350E-02  0.27554165E-01  -0.27409853E-01  0.84838674E-02  -0.20637434E-05
0.27554165E-01  0.11357039E+00  0.67886928E-01  -0.44501547E-01  0.23910297E-04
-0.27409853E-01  0.67886928E-01  0.12600848E+00  0.43067575E-01  -0.12721073E-04
0.84838674E-02  -0.44501547E-01  0.43067575E-01  0.25225357E-02  0.18367222E-05
-0.20637434E-05  0.23910297E-04  -0.12721073E-04  0.18367222E-05  0.76170530E-09
```

Destaquemos de esta salida que no aparece ningún mensaje de error ni de aviso que nos indique que estamos en una situación problemática. Observemos que $\hat{\gamma} = 0.99999998$. Además, al calcular con Excel 2007 el determinante de la matriz de covarianzas tal como es ofrecida en la salida obtenemos -1.5673×10^{-14} .

Para el mismo ejemplo, si calculamos los logaritmos de inputs y output con 7 cifras decimales, la salida (resumida) de FRONTIER 4.1 es:

```
Output from the program FRONTIER (Version 4.1c)

the ols estimates are :
      coefficient      standard-error      t-ratio
beta 0      -0.16368920E+00      0.17166988E+00      -0.95351152E+00
beta 1       0.63271608E+00      0.29752595E+00      0.21265913E+01
beta 2       0.22442639E+00      0.34443920E+00      0.65157042E+00
sigma-squared 0.15973407E+00

log likelihood function = -0.89195549E+01
```

```

the estimates after the grid search were :
beta 0      0.29404424E+00
beta 1      0.63271608E+00
beta 2      0.22442639E+00
sigma-squared 0.34643482E+00
gamma       0.95000000E+00
mu is restricted to be zero
eta is restricted to be zero

the final mle estimates are :
      coefficient      standard-error      t-ratio
beta 0      0.87485630E-01  0.17436621E+00  0.50173499E+00
beta 1      0.11374549E+01  0.24957561E+00  0.45575565E+01
beta 2      -0.26732249E+00  0.25254726E+00 -0.10585048E+01
sigma-squared 0.32030733E+00  0.77150512E-01  0.41517200E+01
gamma       0.99999999E+00  0.16350605E-04  0.61159816E+05
mu is restricted to be zero
eta is restricted to be zero
log likelihood function = -0.35691767E+01

covariance matrix :
0.30403575E-01 -0.43040476E-01  0.46685654E-01  0.30730179E-03  0.35034813E-06
-0.43040476E-01  0.62287985E-01 -0.60718987E-01 -0.19882570E-02  0.15292573E-05
0.46685654E-01 -0.60718987E-01  0.63780117E-01  0.49752553E-03 -0.13464999E-05
0.30730179E-03 -0.19882570E-02  0.49752553E-03  0.59522015E-02 -0.29833728E-06
0.35034813E-06  0.15292573E-05 -0.13464999E-05 -0.29833728E-06  0.26734229E-09

```

Podemos apreciar que la incorporación del séptimo decimal en los datos iniciales produce cambios en los estimadores MCO en el sexto decimal, mientras que los EMVs sufren variaciones bastante más significativas, ya que afectan a su primera cifra decimal. Concretamente, las variaciones porcentuales que experimentan los estimadores de β_0 , β_1 y β_2 al tener en cuenta el séptimo decimal son -43.87% , 13.27% y -103.20% respectivamente. El determinante de la matriz de covarianzas estimada en este caso es -3.9584×10^{-18} .

Señalemos que los resultados ofrecidos en Coelli et al. (1998, pp.195-197) coinciden totalmente con los mostrados en nuestra segunda salida (7 decimales en los datos de partida) para la estimación MCO, pero los resultados ofrecidos para los EMVs tampoco coinciden con los que hemos obtenido nosotros, lo cual puede ser debido a que en el libro se utiliza la versión 4.1 de FRONTIER y nosotros hemos usado la versión 4.1c.

Esta gran sensibilidad en los resultados ante ligeras variaciones en los datos de partida es consecuencia del “mal comportamiento” de la verosimilitud en la frontera $\gamma \approx 1$. Además, el hecho de que la matriz de covarianzas sea singular es también un indicativo de que no podemos asegurar que nos encontremos ante un máximo, sino que tal vez estemos ante un punto de silla de la función.

En definitiva, pensamos que la respuesta ofrecida por LIMDEP 9.0 de detener el proceso está suficientemente motivada, pues nos encontramos en una situación en la que,

en esencia, los métodos numéricos no son capaces de garantizar que hemos encontrado el máximo de la función de verosimilitud para el modelo de frontera estocástica. Pensamos que FRONTIER al menos debería mostrar algún mensaje de aviso, aunque ofrezca una solución. Una estrategia acertada ante este problema pensamos que sería estimar un modelo de frontera determinista; dadas las dificultades que presenta la estimación máximo verosímil en el modelo de frontera determinista y las malas propiedades del estimador de mínimos cuadrados corregidos (MCC), una buena solución es utilizar la estimación bayesiana (Ortega y Basulto, 2009; Ortega et al, 2009).

5. ANÁLISIS DE SIMULACIÓN.

El término de error compuesto en el modelo de producción con frontera estocástica, como ya hemos indicado, es $\varepsilon_i = v_i - u_i$, donde v_i recoge los efectos aleatorios y u_i la ineficiencia. En el caso de estimar a partir de una muestra que presenta el problema de asimetría errónea, obtenemos básicamente $\varepsilon_i = v_i$ (es decir, $u_i = 0$), mientras que al estimar con una muestra que presenta el problema de no aleatoriedad obtenemos $\varepsilon_i = u_i$. Se puede interpretar que ambos casos indican que se ha elegido un modelo inadecuado (de hecho, esta es la interpretación más frecuente, sobre todo en el caso de asimetría errónea).

Sin embargo, algunos estudios simulados ponen de manifiesto que siendo “el modelo correcto”, la proporción de muestras que nos llevan a uno u otro extremo es bastante significativo, aún para tamaños muestrales bastante elevados (Simar, 2007). Este hecho sugiere que el problema es que en un porcentaje importante de casos “es virtualmente imposible” distinguir entre los dos componentes del término de error. Algunos autores sugieren usar la distribución Exponencial en lugar de la Half-Normal para la perturbación u_i debido a que es “más diferente” de la distribución Normal de v_i y así es más fácil evitar este problema de separación del efecto aleatorio de la ineficiencia (Koop and Steel, 2003).

Nosotros hemos llevado a cabo un análisis de simulación en el que se han generado muestras aleatorias siguiendo el modelo de producción Half-Normal con frontera aleatoria

para distintos valores del parámetro λ , y hemos obtenido el porcentaje de casos en que se presentan los problemas de asimetría errónea y ausencia de efectos aleatorios, es decir, el porcentaje de veces en que el modelo generador de los datos no es identificado correctamente a partir de la muestra. Para ello, hemos usado el software estadístico R, más concretamente el paquete *frontier* en su versión 0.996-6 el cual usa el código fuente *Fortran* de FRONTIER 4.1. Esto nos permite obtener las estimaciones de FRONTIER, pero integrado en un entorno mucho más versátil que permite la programación.

Los resultados obtenidos se muestran en la tabla I. Concretamente, hemos considerado un modelo $y_i = \beta_0 + \beta_1 x_i + v_i - u_i$, donde hemos fijado $\beta_0 = 1$, $\beta_1 = 2$, $\sigma = 1$ y hemos hecho variar λ en el conjunto $\{0.25, 0.5, 1, 2, 4\}$ (los valores correspondientes de γ son $\{0.0588, 0.2, 0.5, 0.8, 0.9412\}$). Los tamaños muestrales utilizados han sido $n \in \{10, 20, 30, 40, 50, 100\}$ y para obtener los porcentajes hemos generado en cada una de las 30 combinaciones un total de $m = 1000$ muestras. Los valores de la covariable x han sido obtenidos según un modelo $U(0,10)$.

Tabla I. Porcentaje de casos en que el modelo queda incorrectamente identificado

	<i>Asimetría errónea (%)</i>						<i>Ausencia de efectos aleatorios (%)</i>					
	<i>n=10</i>	<i>n=20</i>	<i>n=30</i>	<i>n=40</i>	<i>n=50</i>	<i>n=100</i>	<i>n=10</i>	<i>n=20</i>	<i>n=30</i>	<i>n=40</i>	<i>n=50</i>	<i>n=100</i>
$\lambda=0.25$	51.1	49.1	49.2	50.0	46.4	51.7	30.4	12.6	7.2	1.7	0.7	0.0
$\lambda=0.5$	51.8	47.8	50.9	47.9	46.1	47.1	29.4	12.9	6.1	1.2	0.9	0.0
$\lambda=1$	46.0	42.6	40.0	38.8	36.7	30.1	33.8	18.0	7.1	2.9	1.0	0.0
$\lambda=2$	38.8	27.4	21.3	14.8	11.4	3.9	40.3	25.7	13.0	6.4	4.2	0.0
$\lambda=4$	27.9	11.9	6.5	2.5	1.4	0.0	52.3	42.3	28.9	23.1	14.9	2.0

El primer aspecto destacable es que en general los porcentajes no son, ni mucho menos, poco importantes, sobre todo en el caso de asimetría errónea. Observemos que, por ejemplo, cuando $\lambda = 1$ (es decir, $\sigma_v = \sigma_u$) y $n = 30$ el 40% de las muestras nos conducirían al error de considerar que las perturbaciones del modelo carecen de término de ineficiencia, y el 7.1% nos haría pensar en ausencia de efectos aleatorios. Además, aún aumentando el tamaño muestral hasta $n = 100$ el porcentaje sigue siendo bastante elevado para asimetría errónea (30.1%).

También podemos observar que, en general, los porcentajes de muestras que llevan a estimar ausencia de efectos aleatorios son, en general, inferiores a los de muestras con problema de asimetría errónea.

Por otra parte y como es lógico, se aprecia que fijado un valor de λ los porcentajes disminuyen al aumentar el tamaño muestral, si bien esa disminución es bastante lenta para algunos valores de λ . Fijado un tamaño muestral n , también apreciamos (como indica la lógica) que al aumentar $\lambda = \sigma_u / \sigma_v$, disminuye el porcentaje de muestras con asimetría errónea mientras que por el contrario aumenta el de muestras con el problema de ausencia de efectos aleatorios.

6. CONCLUSIONES.

La principal objeción que se ha hecho al modelo de producción con frontera determinista es la ausencia de un término que recoja variaciones accidentales. El modelo con frontera estocástica, inicialmente solventa este inconveniente en su formulación, si bien hemos comprobado en el estudio de simulación que en una proporción importante de casos no podemos discriminar correctamente en el término de error compuesto el efecto debido a cada uno de los componentes (al menos con el método de máxima verosimilitud, que es el que hemos analizado). En efecto, con bastante frecuencia concluimos que todo el término de perturbación se debe sólo a efectos aleatorios o solo a ineficiencia, aún en muestras simuladas de un modelo en el que ambos términos de la perturbación compuesta tienen varianzas similares.

El problema de asimetría errónea ha sido considerado con mayor atención en la literatura específica del tema. Inicialmente la opción más extendida fue rechazar este tipo de modelización ante la presencia de este problema (Green, 1993), si bien con posterioridad se han propuesto otras alternativas como “profundizar” en la búsqueda del máximo de la función o usar modelos de error compuesto pero con otro tipo de distribución diferente a la Half-Normal. Otra opción sería ensayar algún método alternativo para estimar el modelo, como los métodos bayesianos (Koop and Steel, 2003).

La ausencia de factores aleatorios, en gran medida no ha sido considerado como un problema; de hecho, un software tan extendido como FRONTIER 4.1 no muestra ningún mensaje de aviso ante esta situación, sino que ofrece una solución como definitiva, en la que la varianza del término de error aleatorio es prácticamente 0. No obstante, como hemos indicado en el epígrafe 4, en esta situación existen dudas razonables de que efectivamente hayamos encontrado el máximo de la función (al menos aproximadamente), los resultados que se obtienen son extremadamente sensibles a ligeras variaciones en los datos iniciales y además la matriz de covarianzas que estimamos es singular (motivo por el que LIMDEP 9.0 detiene el proceso de estimación y no ofrece ningún resultado). Ante esta situación, las opciones de actuación son similares a las descritas para el caso de asimetría errónea. También podríamos estimar un modelo de producción con frontera determinista usando metodología bayesiana (Ortega y Basulto, 2009; Ortega et al., 2009).

Para terminar, indiquemos que, obviamente, la posibilidad de ampliar el tamaño muestral o de usar datos de panel reduce las posibilidades de que se presente algún tipo de problemas a la hora de estimar.

BIBLIOGRAFÍA

Aigner, D.J.; Chu, S.F. (1968), "On estimating the industry production function", *American Economic Review*, 58, pp. 826-839.

Aigner, D.J.; Lovell, C.A.; Schmidt, P. (1977), "Formulation and estimation of stochastic frontier production function models", *Journal of Econometrics*, 6, pp. 21-37.

Battese, G.E.; Corra, G.S. (1977), "Estimation of a production frontier model: With application to the Pastoral Zone of Eastern Australia", *Australian Journal of Agricultural Economics*, 21, pp.169-179.

Cazals, C., Florens J.P. and Simar, L. (2002), "Nonparametric frontier estimation: a robust approach", *Journal of Econometrics*, 106, pp. 1-25.

Coelli, T.J.; Rao, D.S.P.; Battese, G.E. (1998), *An introduction to efficiency and productivity analysis*, Kluwer Academic Publishers, Boston

Daouia, A. and Simar, L. (2004), "Nonparametric efficiency analysis: a multivariate conditional quantile approach", *Journal of Econometrics*, 140, pp. 375-400.

Farrell, M.J. (1957), "The measurement of productive efficiency", *Journal of the Royal Statistical Society (A)*, 120, pp. 253-281.

Greene, W.H. (1993), "The econometric approach to efficiency analysis". En: Fried, H.O.; Lovell, C.A.K.; Schmidt, S.S. (editores), *The measurement of productive efficiency: Techniques and applications*, Oxford University Press, New York.

Koop, G.; Steel, M.F. (2003), "Bayesian analysis of stochastic frontier models". En: Baltagi, B.H. (editor), *A companion to theoretical econometrics*, Blackwell.

Meeusen, W.; van den Broeck, J. (1977), "Efficiency estimation from Cobb-Douglas production functions with composed error", *International Economic Review*, 18, pp. 435-444.

Olson, J.A.; Schmidt, P.; Waldman, D.M. (1980), "A Monte Carlo study of estimators of stochastic frontier production functions", *Journal of Econometrics*, 13, pp. 67-82.

Ortega, F.J.; Basulto, J. (2009), "Estimación Bayesiana en modelos de producción con frontera determinista", *Estudios de Economía Aplicada*, 27 (2), p. 573.
<http://www.revista-eea.net/documentos/27205.pdf>

Ortega, F.J.; Basulto, J.; Camúñez, J.A. (2009), "Comparing bayesian and corrected least-squares estimators in frontier production models", *Boletín de Estadística e Investigación Operativa*, 25 (2), pp. 86-96.

Simar, L. (2007), "How to improve the performances of DEA/FDH estimators in the presence of noise?", *Journal of Productivity Analysis*, 28, pp. 183-201.

Waldman, D.M. (1982), "A stationary point for the stochastic frontier likelihood", *Journal of Econometrics*, 18, pp. 275-279.

APÉNDICE: DATOS USADOS EN LOS EJEMPLOS

Datos usados en el ejemplo 1:

Y	19	15	15	10	17	10	20	1	16	19	9	3	14	17	20	10	16	15	21	8
X	10	8	7	5	9	6	10	1	7	9	5	2	8	9	10	5	9	7	10	5

Datos usados en el ejemplo 2:

País	Índice Prod.	Líneas (10 ³ km)	Empleados (10 ⁵ personas)
Australia	0.74	0.7767	0.85
Austria	0.24	0.3223	0.18
Belgium	0.36	0.3990	0.26
Canada	1.26	1.5296	1.05
Denmark	0.39	0.2911	0.18
Finland	0.29	0.2670	0.20
France	2.06	2.8085	1.56
Germany	1.73	2.9981	2.12
Iceland	0.02	0.0126	0.02
Ireland	0.11	0.0983	0.13
Italy	1.48	2.2350	1.18
Japan	2.73	5.3236	2.77
Netherlands	0.77	0.6940	0.32
New Zealand	0.16	0.1473	0.17
Norway	0.27	0.2132	0.15
Portugal	0.19	0.2379	0.23
Spain	0.59	1.2603	0.75
Sweden	0.71	0.5849	0.42
Switzerland	0.56	0.3943	0.22
Turkey	0.15	0.6893	0.36
UK	2.53	2.5404	2.27