



Algunas observaciones acerca del uso de software en la estimación del modelo Half-Normal

ORTEGA IRIZO, FRANCISCO JAVIER

Departamento de Economía Aplicada I

Universidad de Sevilla (España)

Correo electrónico: fjortega@us.es

GAVILÁN RUIZ, JOSÉ MANUEL

Departamento de Economía Aplicada I

Universidad de Sevilla (España)

Correo electrónico: gavi@us.es

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. Usando el software FRONTIER, analizamos el caso en que la estimación sugiere la ausencia de factores aleatorios en el término de error compuesto. Hemos comprobado que existen motivos para pensar que las estimaciones de los parámetros y, sobre todo, sus errores estándar son de dudosa validez. El software LIMDEP no obtiene estimaciones en este caso, ofreciendo un mensaje de error.

Palabras clave: frontera estocástica; frontera determinista; estimador máximo verosímil; software FRONTIER.

Clasificación JEL: C52; C63; C87.

MSC2010: 65C60; 91B38.

Some Notes about the Using of Software to Estimate the Half-Normal Model

ABSTRACT

Using the maximum likelihood method, in order to estimate Half-Normal stochastic frontier production models, entails several practical difficulties that, perhaps, have not been sufficiently emphasised. In employing FRONTIER software, we analyse the case in which the estimation obtained suggests the absence of random factors in the composite error term. We have proved that there are reasons to doubt the validity of the parameter estimates and especially of its standard errors. On the other hand, no estimation is obtained in the previous situation, with LIMDEP software, but an error message.

Keywords: stochastic frontier; deterministic frontier; maximum likelihood estimator; software FRONTIER.

JEL classification: C52; C63; C87.

MSC2010: 65C60; 91B38.



1. INTRODUCCIÓN

Los modelos econométricos de producción con frontera permiten estimar el nivel máximo de *output* alcanzable, para unos *inputs* dados, en un proceso productivo. A partir de la diferencia entre la producción máxima (o frontera) y la observada, se obtienen indicadores de ineficiencia en el proceso, siendo esta idea propuesta de manera pionera en Farrell (1957).

Bajo la misma idea, en la literatura sobre el tema, se han propuesto dos tipos de modelos. El primero de ellos, actualmente denominado modelo de producción con frontera determinista, considera que la diferencia entre el *output* máximo y el observado es una variable aleatoria positiva. Este enfoque fue sugerido por primera vez en Aigner y Chu (1968). Esta opción presenta varias dificultades (Ortega *et al.*, 2010, Simar, 2007); una de las más destacables es que viola las condiciones usuales de regularidad que permiten obtener la distribución asintótica del estimador máximo verosímil (EMV). La otra alternativa, actualmente denominada modelo de producción con frontera estocástica (o de error compuesto), consiste en introducir dos perturbaciones: una que recoge la ineficiencia y otra que refleja las posibles fuentes de variaciones aleatorias. Este modelo, que sí verifica las condiciones de regularidad necesarias para utilizar el EMV, fue propuesto de manera independiente y casi simultánea en Aigner *et al.* (1977), Battese y Corra (1977) y Meusen y van den Broeck (1977).

Los modelos de frontera estocástica están ampliamente difundidos en la literatura, en parte debido al desarrollo de software específico (como FRONTIER y LIMDEP), los cuales han implementado los algoritmos necesarios para maximizar numéricamente la función de verosimilitud y obtener estimaciones de la matriz de covarianzas.

El objetivo de este trabajo es poner de manifiesto que, para determinadas muestras, se presentan algunas dificultades prácticas que hacen que debamos analizar con cautela los resultados arrojados por los métodos de cálculo numérico implementados en el software desarrollado al efecto. Concretamente, vamos a centrarnos en el caso en que el EMV concluye que la totalidad del término de error compuesto se debe a ineficiencia y, por tanto, hay ausencia de factores aleatorios o, dicho de otra forma, que en realidad nos encontramos ante un modelo de frontera determinista.

En primer lugar, es destacable que en este caso hay discrepancia de resultados entre LIMDEP (que detiene el proceso de estimación con un mensaje de error) y FRONTIER (que ofrece las estimaciones de los parámetros y su matriz de covarianzas, sin ningún tipo de mensaje de error ni aviso). Hemos analizado la respuesta ofrecida por FRONTIER en esta situación, concluyendo que las estimaciones de los parámetros y, sobre todo, sus errores estándar son de dudosa validez, debido a que estamos usando los métodos tradicionales de estimación en una zona en la que la verosimilitud, aunque verifica las condiciones de regularidad, presenta un comportamiento muy similar al de los modelos no regulares. Aclaremos que las versiones de ambos programas que

hemos utilizado han sido LIMDEP 9.0 y el software estadístico R, más concretamente el paquete *frontier* en su versión 0.996-6, el cual usa el código fuente *Fortran* del software FRONTIER 4.1 (Coelli, 1996). Además, debemos reseñar que la ausencia de factores aleatorios puede aparecer en una proporción no despreciable de muestras, por lo que no es extraño que nos encontremos con esta situación al trabajar con este tipo de modelos de error compuesto (Ortega *et al.* 2010).

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 y FRONTIER para llevar a cabo el proceso de estimación; en la sección 3 presentamos el ejemplo que vamos a analizar así como los resultados que proporcionan LIMDEP y FRONTIER; en la sección 4 analizamos en detalle la solución ofrecida por FRONTIER y exponemos las situaciones conflictivas con las que nos hemos encontrado; en la sección 5 aportamos algunas explicaciones acerca de las situaciones problemáticas observadas. Por último, en la sección 6 exponemos las principales conclusiones de nuestro trabajo.

2. PLANTEAMIENTO DEL MODELO

La formulación del modelo de producción con frontera estocástica que vamos a considerar es:

$$y_i = x_i' \beta + \underbrace{v_i - u_i}_{\varepsilon_i}, \quad i = 1, \dots, n,$$

donde y_i es la producción de la firma i -ésima, x_i es el vector de todos sus *inputs* y β es un vector de parámetros a estimar. Indiquemos que, en este contexto, es frecuente que tanto *inputs* como *outputs* se consideren en unidades logarítmicas, lo cual corresponde a una función de producción de tipo Cobb-Douglas.

En el error compuesto ε_i se recogen las variaciones aleatorias (mediante la perturbación $v_i \in \mathbb{R}$) y la ineficiencia en el proceso productivo (a través de $u_i > 0$). Comúnmente se supone que las perturbaciones v_i siguen una distribución Normal, concretamente, $v_i \sim N(0, \sigma_v^2)$. Sin embargo, con respecto a la distribución de u_i se han considerado distintas opciones (como, por ejemplo, Exponencial y Gamma). En este trabajo se adopta la hipótesis más habitual, que consiste en tomar $u_i = |u_i^*|$, donde $u_i^* \sim N(0, \sigma_u^2)$. Por definición, se dice que las perturbaciones u_i siguen una distribución Half-Normal, lo cual se representa por $u_i \sim HN(0, \sigma_u^2)$. Además, se considera que todas las perturbaciones son independientes entre sí.

Puesto que vamos a utilizar los programas LIMDEP y FRONTIER, resulta conveniente especificar las parametrizaciones que se usan en ellos para obtener el EMV, ya que son diferentes.

En ambos casos se define $\sigma^2 = \sigma_v^2 + \sigma_u^2$, que es la varianza total del término de error compuesto, debido a la independencia entre v y u . Sin embargo, mientras que en LIMDEP se considera $\lambda = \sigma_u/\sigma_v$, en FRONTIER se utiliza $\gamma = \sigma_u^2/\sigma^2$. Así, $\lambda^2 \geq 0$ se interpreta como la proporción entre la varianza debida a ineficiencia y la varianza debida a factores aleatorios, mientras que $0 \leq \gamma < 1$ se interpreta como la proporción entre la varianza debida a ineficiencia y la varianza total. Con la parametrización usada en LIMDEP, el logaritmo de la verosimilitud para una muestra individual viene dado por:

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

Para obtener el logaritmo de la verosimilitud correspondiente a la parametrización usada en FRONTIER, basta con sustituir en la fórmula anterior λ por $\sqrt{\gamma/(1-\gamma)}$.

3. RESULTADOS DE LIMDEP Y FRONTIER EN AUSENCIA DE EFECTOS ALEATORIOS

Para comenzar esta sección, vamos a exponer los resultados ofrecidos por LIMDEP y FRONTIER sobre un conjunto concreto de datos.

Ejemplo. Consideremos 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 Anexo 1.

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 10^3 kilómetros) y el número de trabajadores (en 10^5 personas). Tanto el *output* como los *inputs* se consideran en logaritmos.

Al utilizar LIMDEP 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
```

A continuación, ofrecemos un resumen de los resultados que proporciona FRONTIER, en el que se incluye la matriz de covarianzas estimadas:

```
Error Components Frontier (see Battese & Coelli 1992)
Inefficiency decreases the endogenous variable (as in a production function)
The dependent variable is logged
convergence achieved after 9 iterations

final maximum likelihood estimates
      Estimate Std. Error  z value  Pr(>|z|)
(Intercept)  0.1237296  0.1386500  0.8924  0.3721848
log(x1)      1.0495884  0.2697797  3.8905  0.0001000 ***
```

```

log(x2)      -0.2084949  0.2628753  -0.7931  0.4277007
sigmaSq      0.3456445  0.1021628   3.3833  0.0007163 ***
gamma        1.0000000  0.0030055  332.7271 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
log likelihood value: -4.046282

cross-sectional data
total number of observations = 21
mean efficiency: 0.6863549

```

```

Matriz de Covarianzas:
      (Intercept)      log(x1)      log(x2)      sigmaSq      gamma
(Intercept)  0.0192238129 -0.0266651478  0.0330569866  0.0008946169 -1.598626e-04
log(x1)      -0.0266651478  0.0727811101 -0.0665491785 -0.0018229559  3.333693e-04
log(x2)      0.0330569866 -0.0665491785  0.0691034136  0.0015375604 -3.242120e-04
sigmaSq      0.0008946169 -0.0018229559  0.0015375604  0.0104372378 -1.008058e-04
gamma        -0.0001598626  0.0003333693 -0.0003242120 -0.0001008058  9.032825e-06

```

De este resumen, lo primero que queremos destacar es que no aparece ningún mensaje de error ni de aviso que nos indique que estamos en una situación problemática. Observemos que $\hat{\gamma} = 1$ (aunque este valor realmente es una aproximación, ya que en 1 la función de verosimilitud no es evaluable) y, por tanto, una de las principales conclusiones sería que en el término de error compuesto no está presente el término de perturbación que recoge los factores aleatorios y que, por tanto, en esencia, estamos ante un modelo de frontera determinista. Además, hemos conseguido obtener una matriz estimada de covarianzas, lo cual no es posible formulando inicialmente el modelo determinista, ya que en él se violan las condiciones de regularidad necesarias.

Desafortunadamente, el procedimiento implementado en FRONTIER, aunque ofrece una solución cuando $\hat{\gamma} = 1$, conlleva serias dificultades que nos hacen dudar de la validez de la misma, como veremos en la sección siguiente. Para finalizar esta sección, observemos que el mensaje de error de LIMDEP es “estimated variance matrix of estimates is singular”. Al calcular con R el determinante de la matriz de covarianzas, tal como es ofrecida en los resultados anteriores, obtenemos $1,2110633e-14$, lo que parece confirmar que la matriz de covarianzas es singular, siendo este el primer argumento que nos hace dudar de la validez de las estimaciones obtenidas.

4. ANÁLISIS DE LA SOLUCIÓN OFRECIDA POR FRONTIER

En esta sección, analizamos con más detalle la solución que ofrece FRONTIER a este ejemplo, llevando a cabo dos pruebas encaminadas a verificar la fiabilidad de las estimaciones. Primero, hemos comprobado la solución que ofrece el software tras volver a estimar el modelo considerando como punto de partida la solución obtenida previamente. Segundo, hemos analizado el comportamiento de las estimaciones ante cambios de escala en las variables explicativas.

4.1 Uso de la solución obtenida como punto inicial para una nueva estimación

En esta sección exploramos la posibilidad de mejorar la solución óptima ofrecida por FRONTIER. Para ello, hemos llevado a cabo un proceso de búsqueda en red, evaluando la función de verosimilitud en el entorno del punto óptimo obtenido inicialmente.

Como podemos apreciar en los resultados anteriores, en la solución ofrecida por FRONTIER el máximo de la verosimilitud es $-4,046282$. No obstante, podemos comprobar que en el punto $\hat{\beta}_0 = 0,1115409$, $\hat{\beta}_1 = 1,0324950$, $\hat{\beta}_2 = -0,1780069$, $\hat{\sigma}^2 = 0,3192176$, $\hat{\gamma} \approx 1$, la función de verosimilitud vale $-3,251858$, lo que supone un aumento bastante significativo.

En esta misma línea, también ensayamos la idea de estimar el modelo usando FRONTIER, pero partiendo de un punto inicial cercano a la solución previamente obtenida. Dadas las significativas diferencias que observamos entre los resultados obtenidos, pensamos en volver a estimar usando como punto de partida para la búsqueda del máximo exactamente la solución ofrecida como óptima por el software. Teóricamente, los métodos numéricos de búsqueda de óptimos ante esta situación no deben iterar, proporcionando de nuevo el mismo punto solución. Sin embargo, hemos podido comprobar cómo el programa vuelve a llevar a cabo algunas iteraciones, ofreciendo soluciones diferentes y mejorando la verosimilitud.

En el Anexo 2 ofrecemos algunos de los resultados obtenidos al aplicar al ejemplo considerado este proceso de uso de la solución óptima como punto inicial para una nueva búsqueda. Resultan especialmente significativas, como ya hemos mencionado, las grandes variaciones que se observan en los *z-value* y, en general, en la matriz de covarianzas estimada, resultado que pone en entredicho la validez de dichas estimaciones (comparar, por ejemplo, los valores ofrecidos en los apartados A y B del Anexo 2).

Repitiendo varias veces este procedimiento de usar la solución óptima como punto inicial para una nueva búsqueda, se termina por llegar a una situación en la que el software hace una única etapa (como debería ocurrir teóricamente siempre). Hemos observado que en este caso la matriz de covarianzas estimada es siempre la identidad (Anexo 2.D). Cuando el número de iteraciones es pequeño (del orden de 3 o 4), todos los errores estándar, salvo el del parámetro γ , son cercanos a 1 y todas las covarianzas cercanas a 0 (Anexo 2.C).

4.2. Cambio de escala en las variables explicativas

En este apartado analizamos el comportamiento de las estimaciones ante cambios de escala en las variables explicativas. En principio, si multiplicamos una variable explicativa por una constante k (por ejemplo, como consecuencia de un cambio de unidad de medida), este hecho no debería afectar a la significatividad de dicha variable en el modelo, pues tanto la estimación de la pendiente correspondiente como el error estándar deberían quedar divididos por k (al menos aproximadamente), con lo cual el *z-value* debe permanecer constante.

En el ejemplo considerado, hemos procedido a multiplicar ambas variables explicativas por la constante $k = 100$ (debemos aclarar que en este ejemplo las variables están en logaritmos, por lo que no tendría mucho sentido un cambio de escala de esta naturaleza; no obstante, para nuestro

objetivo de comparación de resultados, este hecho no es relevante y hemos preferido no cambiar de ejemplo). Los resultados que se obtienen ahora son los que se muestran a continuación:

```
Error Components Frontier (see Battese & Coelli 1992)
Inefficiency decreases the endogenous variable (as in a production function)
The dependent variable is logged
convergence achieved after 6 iterations
```

```
final maximum likelihood estimates
      Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.16058720  0.48313437  0.3324  0.7396
100*log(x1)  0.00927335  0.00633806  1.4631  0.1434
100*log(x2) -0.00093949  0.00748136 -0.1256  0.9001
sigmaSq      0.31359422  0.15866606  1.9764  0.0481 *
gamma        0.99999999  0.11651841  8.5823  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
log likelihood value: -3.918338

cross-sectional data
total number of observations = 21

mean efficiency: 0.681226
```

Destaquemos las enormes variaciones que experimentan los z -value, lo que lleva a que la significatividad de las pendientes dependa de las unidades de medida que consideremos para las variables. Esto se debe fundamentalmente a que los errores estándar muestran un comportamiento muy diferente del que deberían mostrar teóricamente. De nuevo, los resultados nos hacen dudar de la validez de las estimaciones, sobre todo en el caso de los errores estándar.

5. ALGUNAS EXPLICACIONES DEL COMPORTAMIENTO OBSERVADO

La explicación básica del mal comportamiento del método de estimación máximo verosímil que hemos observado en el ejemplo considerado está en el hecho de que al tomar límite cuando $\lambda \rightarrow +\infty$ (o, equivalentemente, cuando $\gamma \rightarrow 1$) en la verosimilitud del modelo con frontera estocástica, obtenemos la verosimilitud del modelo de frontera determinista (ya que, en el límite, lo que hacemos es eliminar en el término de error compuesto la perturbación que recoge las variaciones debidas a factores aleatorios).

Como indicamos en la sección 2, la formulación del modelo de frontera determinista es $y_i = f(x_i, \beta) + (-u_i)$, donde $u_i > 0$ recoge la ineficiencia en el proceso productivo. Ahora bien, $u_i > 0 \Rightarrow y_i - f(x_i, \beta) < 0$, por lo que esta verosimilitud está sometida a un conjunto de restricciones que ligan los parámetros y los valores de las variables (siendo esta la causa de que estemos ante un modelo no regular, en el que la distribución asintótica usual del estimador máximo verosímil no es válida). Así, si llamamos B al subconjunto del espacio paramétrico compatible con las restricciones tenemos una verosimilitud discontinua en la frontera de B.

La verosimilitud del modelo de frontera estocástica no está sometida a ningún tipo de restricciones y está definida en todo el espacio paramétrico inicial. Al aplicar el proceso descrito en la sección 4.1 al ejemplo considerado, FRONTIER ofrece la solución $\hat{\gamma} = 1$, $\hat{\beta}_1 = 1,04917$,

siendo el valor de la función objetivo igual a $-3,259794$ (como puede comprobarse en el Anexo 2.D). Pues bien, manteniendo el resto de valores, si evaluamos el logaritmo de la verosimilitud en $\beta_1' = 1,048$ obtenemos $-4,062$; en $\beta_1'' = 1,045$ la función vale $-193,854$; y en $\beta_1''' = 1,040$ desciende hasta $-4867,831$.

Como hemos dicho, en el modelo de frontera determinista la verosimilitud no está definida (o vale 0) fuera del conjunto B y en el modelo de frontera estocástica, cuando $\gamma \simeq 1$, lo que ocurre es que la función está siempre definida, pero en la frontera del conjunto B el logaritmo de la verosimilitud experimenta enormes variaciones al hacer en los parámetros cambios del orden de milésimas. Este sería uno de los motivos de la sensibilidad que presentan las estimaciones.

A grandes rasgos, este sería también el motivo fundamental para dudar de la validez de la matriz de covarianzas estimadas. La estimación de dicha matriz se hace usando la distribución asintótica normal del estimador máximo verosímil que, como sabemos, es válida solo para modelos regulares. En este caso, aunque el modelo es regular, cuando $\gamma \simeq 1$, las enormes variaciones que experimenta el logaritmo de la verosimilitud en la frontera hacen que podamos decir que, a efectos prácticos, nos encontramos ante un modelo no regular, por lo que la estimación obtenida es de dudosa validez.

El motivo de este salto está en el sumando $\log\left(\Phi\left(-\lambda\left(y_i - x_i'\beta\right)/\sigma\right)\right)$ del logaritmo de la verosimilitud. Si usamos la parametrización de FRONTIER, este sumando, como indicamos en la sección 2, es $\log\left(\Phi\left(-\sqrt{\gamma/(1-\gamma)}\left(y_i - x_i'\beta\right)/\sigma\right)\right)$. Cuando fijamos un valor $\gamma \simeq 1$ (o lo que es lo mismo $\lambda \simeq +\infty$), si $\beta \in B$ entonces $y_i - x_i'\beta < 0$ y así tendremos $\Phi\left(-\lambda\left(y_i - x_i'\beta\right)/\sigma\right) \simeq 1$ y, por tanto, $\log\left(\Phi\left(-\lambda\left(y_i - x_i'\beta\right)/\sigma\right)\right) \simeq 0$. Ahora bien, una ligera variación en β que provoque que $y_{i_0} - x_{i_0}'\beta > 0$ para un cierto i_0 hace que $\Phi\left(-\lambda\left(y_{i_0} - x_{i_0}'\beta\right)/\sigma\right) \simeq 0$, por tanto, $\log\left(\Phi\left(-\lambda\left(y_{i_0} - x_{i_0}'\beta\right)/\sigma\right)\right) \simeq -\infty$.

Además, al utilizar algoritmos numéricos para buscar el máximo, también suelen aparecer problemas para evaluar la función de verosimilitud, ya que si el parámetro λ va aumentando en cada etapa y, por otra parte, el parámetro β se va acercando a un valor en la frontera del conjunto B, vamos a tener un cierto i_0 en el que $y_{i_0} - x_{i_0}'\beta$ se va ir acercando a 0. Por tanto, aparece una indeterminación de tipo $0 \times \infty$ cuya evaluación numérica es factible, pero que puede ser extremadamente sensible a las más ligeras variaciones en las condiciones de partida.

En la búsqueda del máximo del logaritmo de la verosimilitud, FRONTIER usa un método de tipo cuasi-Newton (concretamente, el de Davidon-Fletcher-Powell), que en las propias iteraciones,

a la vez que calcula aproximaciones a la solución, va obteniendo también aproximaciones sucesivas a la inversa del hessiano evaluada en el óptimo (Davidon, 1991). Cuando se cumplen los criterios de convergencia, el último punto obtenido es la estimación del parámetro y la última matriz es la estimación de la matriz de covarianzas. Así, en el proceso iterativo no se usa ni el hessiano ni su inversa y es perfectamente posible llegar a una solución en la que la matriz de covarianzas es prácticamente singular, que es lo que, de hecho, hace el software.

El uso de este método es la explicación de que, cuando se encuentra la solución en una única etapa (como explicamos en la sección 4.2), la matriz de covarianzas estimada sea la identidad, que es la matriz que se toma como aproximación inicial a la inversa del hessiano. De hecho, el criterio de parada compara la estimación y la verosimilitud entre una etapa y la siguiente, pero no comprueba si las aproximaciones a la inversa del hessiano en ambas etapas son similares (Coelli, 1996). Esta misma podría ser la base de la explicación de por qué, cuando se alcanza el óptimo en un número pequeño de etapas, la matriz de covarianzas suele ser parecida a la identidad, salvo en el error estándar del parámetro γ , que se aproxima rápidamente a 0.

6. CONCLUSIONES

En el modelo con frontera estocástica se puede utilizar el EMV, puesto que se verifican las condiciones de regularidad necesarias para ello. Sin embargo, cuando el conjunto de datos nos conduce a estimar ausencia de factores aleatorios, nos encontramos ante una situación bastante problemática. El software LIMDEP no proporciona ninguna solución en este caso, mientras que FRONTIER ofrece tanto las estimaciones de los parámetros como de su matriz de covarianzas, aplicando la distribución asintótica usual del EMV. Como hemos puesto de manifiesto a lo largo del trabajo, pensamos que hay motivos suficientes para poner en duda los resultados obtenidos, especialmente en lo referente a la matriz de covarianzas y que, por tanto, el software al menos debería mostrar algún mensaje de aviso (al igual que hace cuando los datos nos conducen a estimar ausencia de ineficiencia). En definitiva, ninguno de los dos programas utilizados ofrece una solución satisfactoria ante esta situación y pensamos que ello es debido a que es el propio modelo el que no es adecuado en este caso extremo.

Aunque en el presente artículo se ha usado un único ejemplo ilustrativo, hemos podido comprobar que las dificultades aparecen siempre que el EMV del parámetro γ es 1. En la Tabla 1 (pág. 10) del análisis de simulación efectuado en Ortega *et al.* (2010), se obtiene que el porcentaje de muestras en esta situación depende del número de observaciones n y del valor del parámetro γ (o λ), pero que, en todo caso, es un porcentaje significativo, especialmente para muestras pequeñas y medianas ($n \leq 50$). Así, el ejemplo considerado no es un caso aislado, pues tales

muestras aparecen con frecuencia al aplicar esta modelización, si bien el problema tiende a desaparecer a medida que aumenta el número de observaciones.

Como alternativa, se puede plantear directamente un modelo Half-Normal de frontera determinista y, para evitar el inconveniente de la no regularidad, utilizar para su estimación la metodología bayesiana (Ortega y Basulto, 2009; Ortega *et al.* 2009). Otras opciones serían asumir para la perturbación alguna distribución de probabilidad que conlleve que el modelo verifique las condiciones de regularidad (Greene, 1993) o usar análisis no paramétrico tipo DEA (Data Envelopment Analysis) junto con técnicas Bootstrap, lo que también permite obtener estimaciones de la matriz de covarianzas (Simar, 2003). En trabajos futuros, sería interesante analizar si en estas alternativas se reproducen o no los problemas detectados al usar el EMV.

AGRADECIMIENTOS

Los autores queremos expresar nuestro agradecimiento a los evaluadores, cuyas sugerencias han contribuido a mejorar la versión inicial del artículo.

REFERENCIAS

- 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.
- COELLI, T. (1996), "A guide to FRONTIER version 4.1: a computer program for frontier production function estimation", CEPA Working Paper 96/07, Department of Econometrics, University of New England, Armidale, Australia:
<http://www.uq.edu.au/economics/cepa/software/FRONT41-xp1.zip>
- COELLI, T.J.; RAO, D.S.P.; BATTESE, G.E. (1998), *An introduction to efficiency and productivity analysis*, Kluwer Academic Publishers, Boston.
- DAVIDON, W.C. (1991), "Variable metric method for minimization", *SIAM Journal on Optimization*, 1, pp. 1–17.
- 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.
- MEEUSEN, W.; VAN DEN BROECK, J. (1977), "Efficiency estimation from Cobb-Douglas production functions with composed error", *International Economic Review*, 18, pp. 435–444.

- 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.
- ORTEGA, F.J.; GAVILÁN, J.M.; CAMÚÑEZ, J.A. (2010), “Dificultades del estimador máximo verosímil en el modelo de producción Half-Normal con frontera estocástica”, en *Anales de Economía Aplicada 2010*, Delta Publicaciones, Madrid.
- SIMAR L.; WILSON, P.W. (2003), “Statistical inference in non-parametric frontier models: the state of the art”, *Journal of Productivity Analysis*, 13, pp. 49–78
- 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.

ANEXO 1. DATOS USADOS EN EL EJEMPLO

| País | Índice Prod. | Líneas (10 ³ km) | Empleados (10 ³ 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 |

ANEXO 2: ESTIMACIONES SUCESIVAS USANDO COMO PUNTO INICIAL EN CADA PASO LA SOLUCIÓN OFRECIDA POR EL SOFTWARE EN EL PASO ANTERIOR

A) Estimación primera (punto de partida por defecto)

Error Components Frontier (see Battese & Coelli 1992)
 Inefficiency decreases the endogenous variable (as in a production function)
 The dependent variable is logged
 convergence achieved after 9 iterations

```
final maximum likelihood estimates
      Estimate Std. Error  z value  Pr(>|z|)
(Intercept)  0.1237296  0.1386500  0.8924  0.3721848
log(x1)      1.0495884  0.2697797  3.8905  0.0001000 ***
log(x2)     -0.2084949  0.2628753 -0.7931  0.4277007
sigmaSq      0.3456445  0.1021628  3.3833  0.0007163 ***
gamma        1.0000000  0.0030055 332.7271 < 2.2e-16 ***
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 log likelihood value: -4.046282

cross-sectional data
 total number of observations = 21

mean efficiency: 0.6863549

Matriz de covarianzas:

| | (Intercept) | log(x1) | log(x2) | sigmaSq | gamma |
|-------------|---------------|---------------|---------------|---------------|---------------|
| (Intercept) | 0.0192238129 | -0.0266651478 | 0.0330569866 | 0.0008946169 | -1.598626e-04 |
| log(x1) | -0.0266651478 | 0.0727811101 | -0.0665491785 | -0.0018229559 | 3.333693e-04 |
| log(x2) | 0.0330569866 | -0.0665491785 | 0.0691034136 | 0.0015375604 | -3.242120e-04 |
| sigmaSq | 0.0008946169 | -0.0018229559 | 0.0015375604 | 0.0104372378 | -1.008058e-04 |
| gamma | -0.0001598626 | 0.0003333693 | -0.0003242120 | -0.0001008058 | 9.032825e-06 |

Determinante de la matriz de covarianzas: 1.210633e-14

B) Segundo paso

Error Components Frontier (see Battese & Coelli 1992)
 Inefficiency decreases the endogenous variable (as in a production function)
 The dependent variable is logged
 convergence achieved after 12 iterations

```
final maximum likelihood estimates
      Estimate Std. Error  z value  Pr(>|z|)
(Intercept)  1.0688e-01  4.9821e-02  2.1453  0.03193 *
log(x1)      1.0530e+00  1.0255e-01  10.2674 < 2e-16 ***
log(x2)     -1.9378e-01  8.0611e-02  -2.4039  0.01622 *
sigmaSq      3.3806e-01  3.4479e-01  0.9805  0.32684
gamma        1.0000e+00  1.8814e-05 53151.4379 < 2e-16 ***
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 log likelihood value: -3.288578

cross-sectional data
 total number of observations = 21

mean efficiency: 0.7097476

Matriz de covarianzas:

| | (Intercept) | log(x1) | log(x2) | sigmaSq | gamma |
|-------------|---------------|---------------|---------------|---------------|---------------|
| (Intercept) | 2.482161e-03 | -1.096311e-01 | 8.235215e-02 | -7.236534e-02 | 2.266578e-06 |
| log(x1) | -1.096311e-01 | 1.051716e-02 | 2.720971e-02 | -2.457205e-01 | 1.803888e-05 |
| log(x2) | 8.235215e-02 | 2.720971e-02 | 6.498124e-03 | 1.496452e-01 | -7.509471e-06 |
| sigmaSq | -7.236534e-02 | -2.457205e-01 | 1.496452e-01 | 1.188780e-01 | 7.897682e-06 |
| gamma | 2.266578e-06 | 1.803888e-05 | -7.509471e-06 | 7.897682e-06 | 3.539729e-10 |

Determinante de la matriz de covarianzas: -6.814318e-14

C) Sexto paso

Error Components Frontier (see Battese & Coelli 1992)
Inefficiency decreases the endogenous variable (as in a production function)
The dependent variable is logged
convergence achieved after 3 iterations

final maximum likelihood estimates

| | Estimate | Std. Error | z value | Pr(> z) |
|-------------|-------------|------------|-----------|------------|
| (Intercept) | 0.10709940 | 0.99971016 | 0.1071 | 0.9147 |
| log(x1) | 1.04917187 | 0.78493713 | 1.3366 | 0.1813 |
| log(x2) | -0.19132009 | 0.62004392 | -0.3086 | 0.7577 |
| sigmaSq | 0.31945882 | 1.00000000 | 0.3195 | 0.7494 |
| gamma | 0.99999999 | 0.00011073 | 9031.1681 | <2e-16 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
log likelihood value: -3.259794

cross-sectional data
total number of observations = 21

mean efficiency: 0.709443

Matriz de covarianzas

| | (Intercept) | log(x1) | log(x2) | sigmaSq | gamma |
|-------------|---------------|---------------|---------------|---------------|---------------|
| (Intercept) | 9.994204e-01 | 1.492682e-02 | 1.890157e-02 | 2.932158e-09 | -1.081708e-04 |
| log(x1) | 1.492682e-02 | 6.161263e-01 | -4.860985e-01 | -1.196774e-07 | -2.006864e-05 |
| log(x2) | 1.890157e-02 | -4.860985e-01 | 3.844545e-01 | -1.511151e-07 | 1.252667e-05 |
| sigmaSq | 2.932158e-09 | -1.196774e-07 | -1.511151e-07 | 1.000000e+00 | 3.633797e-08 |
| gamma | -1.081708e-04 | -2.006864e-05 | 1.252667e-05 | 3.633797e-08 | 1.226061e-08 |

Determinante de la matriz de covarianzas:7.672611e-21

D) Séptimo paso (y último)

Error Components Frontier (see Battese & Coelli 1992)
Inefficiency decreases the endogenous variable (as in a production function)
The dependent variable is logged
convergence achieved after 1 iterations

final maximum likelihood estimates

| | Estimate | Std. Error | z value | Pr(> z) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 0.10710 | 1.00000 | 0.1071 | 0.9147 |
| log(x1) | 1.04917 | 1.00000 | 1.0492 | 0.2941 |
| log(x2) | -0.19132 | 1.00000 | -0.1913 | 0.8483 |
| sigmaSq | 0.31946 | 1.00000 | 0.3195 | 0.7494 |
| gamma | 1.00000 | 1.00000 | 1.0000 | 0.3173 |

log likelihood value: -3.259794

cross-sectional data
total number of observations = 21

mean efficiency: 0.7094435

Matriz de covarianzas

| | (Intercept) | log(x1) | log(x2) | sigmaSq | gamma |
|-------------|-------------|---------|---------|---------|-------|
| (Intercept) | 1 | 0 | 0 | 0 | 0 |
| log(x1) | 0 | 1 | 0 | 0 | 0 |
| log(x2) | 0 | 0 | 1 | 0 | 0 |
| sigmaSq | 0 | 0 | 0 | 1 | 0 |
| gamma | 0 | 0 | 0 | 0 | 1 |

Determinante de la matriz de covarianzas: 1