
ESTADÍSTICA

Comparing bayesian and corrected least-squares
estimators in frontier production models

Fco. Javier Ortega Irizo, Jesús Basulto Santos and José Antonio
Camúñez Ruíz

Departamento de Economía Aplicada I
Universidad de Sevilla

✉ fjortega@us.es, basulto@us.es, camunez@us.es

Abstract

In the econometric approach to deterministic frontier production models, the use of maximum likelihood estimation has major problems, because the model violates the usual regularity conditions that allow to establish the desirable asymptotic properties of the estimators. To avoid this difficulty, we may use other approaches. One of the methods most widely used is the corrected least-squares estimator. Alternatively, we can perform the Bayesian estimation using Gibbs sampling. In this paper, we make a comparative study of both approaches using simulation methods. We will conclude that the bayesian estimator has better properties in terms of bias and mean squared error, especially for the intercept term; this fact can result very important to estimate the individual efficiencies, which is one of the main objectives of these models.

Keywords: Production models, Bayesian Inference, Gibbs sampling, Corrected least-squares estimator.

AMS Subject classifications: 90B30,62F15,62F30,62J99.

1. Introducción

La estimación de los modelos de producción es un tema de bastante interés dentro del área de economía aplicada, y ha generado en los últimos años gran cantidad de literatura, pudiéndose situar su inicio en el artículo de Aigner y Chu (1968). En este tipo de modelos, 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 una importante dificultad 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. Así, podremos obtener los estimadores máximo-verosímiles resolviendo un problema de optimización con restricciones, pero la determinación de las propiedades estadísticas del estimador es una cuestión complicada.

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 y Simar, 2004).

Alternativamente, se han planteado en la literatura los denominados modelos de producción con frontera estocástica. En ellos, se introducen dos perturbaciones; una es una medida de la ineficiencia, mientras que la otra refleja factores aleatorios, como por ejemplo errores de medida. 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. El problema principal de esta formulación es que, una vez estimado el modelo, no podemos identificar qué parte de cada residuo se debe a ineficiencia y qué parte se debe a errores aleatorios, es decir, no podemos obtener medidas individuales de eficiencia; en este caso, solo es posible estimar niveles medios para el grupo analizado. Una posible salida a este problema se ofrece en Jondrow et al. (1982).

Los modelos con frontera estocástica presentan otro inconveniente importante, que aparece cuando los residuos muestran asimetría contraria a la asimetría teórica de la distribución del término de perturbación compuesto. En este caso, decimos que los residuos muestran 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 (por ejemplo, el programa LIMDEP 7 muestra un mensaje de error y detiene el proceso de estimación), bajo el argumento de que este hecho es un síntoma de que el modelo 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, que pueden llegar a ser hasta del 30 % y

que en ciertos casos se necesita un tamaño muestral superior a 1000 para que esta probabilidad descienda al 5 % (Simar, 2007).

En lo que sigue, nos vamos a centrar en los modelos de producción con frontera determinista. Debido a la dificultad que presenta la estimación máximo-verosímil, se han buscado otras alternativas de estimación. Una de las más utilizadas ha sido el método que llamaremos de mínimos cuadrados corregidos (MCC) (Green, 1980). Alternativamente, podemos utilizar la Inferencia Bayesiana para estimar este modelo con ayuda del algoritmo de Gibbs (Basulto et al., 2006).

El objetivo principal de este trabajo es llevar a cabo una comparación entre ambas alternativas a través de un estudio de simulación. Para ello, en situaciones diversas se van a generar muestras del modelo, se estimarán los parámetros (conocidos) con ambas alternativas y valoraremos los resultados en términos del Sesgo Medio (SM) y del Error Cuadrático Medio (ECM) que se obtiene tras aplicar el procedimiento a un gran número de muestras.

A partir de aquí, en la Sección 2 presentamos el modelo de producción así como su función de verosimilitud bajo las hipótesis asumidas para el término de perturbación; en la Sección 3 exponemos en qué consiste el método de estimación MCC; en la Sección 4 resumimos el proceso de formulación del Algoritmo de Gibbs para el modelo objeto de estudio; en la Sección 5 presentamos la descripción del proceso de simulación llevado a cabo. Por último, en la Sección 6 analizamos los resultados obtenidos y exponemos las conclusiones más importantes que pueden extraerse de los mismos.

2. Planteamiento del modelo

Consideremos el modelo de producción especificado como

$$y_i = x_i' \beta + \varepsilon_i, \quad i = 1, \dots, n$$

donde $y_i \in \mathbb{R}$ es la producción, $\beta \in \mathbb{R}^k$ es un vector de parámetros, $x_i \in \mathbb{R}^n$ es el correspondiente vector de variables exógenas de la observación i -ésima, n es el tamaño de la muestra y $\varepsilon_i < 0$ es una perturbación aleatoria que mide la ineficiencia de la observación i -ésima. Habitualmente, nuestro modelo contendrá un término de ordenada en el origen; sin pérdida de generalidad, supondremos que dicho término es β_1 , es decir, la primera componente del vector β , y por consiguiente la primera componente de todos los vectores x_i , $i = 1, \dots, n$ será igual a la unidad.

Acerca de las perturbaciones, asumiremos la hipótesis de que $-\varepsilon_i$ sigue una distribución *half - Normal*($0, \sigma^2$), $\sigma > 0$, con lo que la densidad inducida para y_i sería

$$f(y_i|x_i, \beta, \sigma) = \frac{2}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (y_i - x_i' \beta)^2 \right\}$$

donde $y_i \leq x'_i \beta \forall i$.

Por tanto, la función de verosimilitud para la muestra de tamaño n del modelo quedaría como

$$L(\beta, \sigma | y, X) \propto \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} (y_i - x'_i \beta)^2 \right\}, \quad y_i \leq x'_i \beta \forall i$$

donde $y = (y_1, \dots, y_n)' \in \mathbb{R}^n$ y $X' = [x_1 | \dots | x_n] \in \mathcal{M}_{k \times n}$. Aclaremos que la columna i -ésima de la matriz X' (o bien la fila i -ésima de X) contiene el vector x_i de valores de las variables exógenas correspondientes a la observación i -ésima. Vamos a llamar B al subconjunto del espacio paramétrico en el que se verifican todas las restricciones y entonces expresaremos la verosimilitud como

$$L(\beta, \sigma | y, X) \propto \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} (y_i - x'_i \beta)^2 \right\}, \beta \in B$$

donde $B = \{\beta \in \mathbb{R}^k : y_i \leq x'_i \beta\}$, es decir el conjunto de valores de los parámetros que son compatibles con los datos observados en el sentido de que posibilitan que todas las producciones observadas queden por debajo de la frontera del modelo.

3. Estimación de Mínimos Cuadrados Corregidos (MCC)

En el modelo de regresión con frontera determinista, puede probarse que los estimadores de Mínimos Cuadrados Ordinarios (MCO) de las pendientes son consistentes y el cálculo usual de los errores estándar correspondientes seguiría siendo válida. Sin embargo, el estimador MCO de la ordenada en el origen es inconsistente para nuestro modelo. Además, es muy posible que el estimador MCO resulte en un vector que no pertenezca al espacio paramétrico, con lo que algunos residuos podrían ser positivos, hecho que no es posible en nuestro modelo pues la producción observada siempre ha de ser inferior a la frontera máxima de producción.

Por ello, se propone una corrección sobre la ordenada en el origen, una vez estimados todos los parámetros por MCO, que conduce a un estimador consistente (pero sesgado) de β_1 y con el que se consigue que el vector de estimaciones pertenezca al espacio paramétrico (y consecuentemente todos los residuos sean negativos).

Concretamente, la corrección que se propone es la siguiente (Green, 1980).

1. Obtener el vector $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)'$ a través del método MCO.
2. Considerar $\tilde{\varepsilon}_i = y_i - \hat{\beta}_2 x_{i2} - \dots - \hat{\beta}_k x_{ik}, \forall i$
3. Sustituir la estimación MCO de la ordenada en el origen por $\hat{\beta}_1^* = \max_{i=1, \dots, n} \{\tilde{\varepsilon}_i\}$.

Con esta corrección, obviamente todos los residuos resultarán ser negativos. Además, como se ha indicado, el estimador así construido resulta ser consistente, si bien como señala el propio Green, es sesgado y de “eficiencia incierta”.

Debemos destacar que uno de los objetivos importantes de estos modelos es la obtención de los residuos, pues van a representar una estimación de la ineficiencia en la producción de cada uno de los ítems que forman parte de la muestra. Para ello, es muy importante contar con buenas estimaciones de las pendientes pero también de la ordenada en el origen, que interviene de manera decisiva en la estimación de la frontera de producción para unos inputs dados y por tanto en la estimación de la ineficiencia. Aún más, es usual plantear este tipo de modelos en términos logarítmicos y así una vez obtenido un residuo $\hat{\varepsilon}_i$ se calcula $\exp\{\hat{\varepsilon}_i\}$ como estimación de la eficiencia/ineficiencia del proceso productivo (1 significa el máximo de eficiencia y 0 el máximo de ineficiencia). Así, los errores que se cometan en las estimaciones de los parámetros de la regresión pueden verse bastante aumentados por el uso de la función exponencial.

4. Estimación Bayesiana. Formulación del algoritmo de Gibbs

En esta sección vamos a exponer los pasos fundamentales que permiten aplicar el algoritmo de Gibbs a nuestro modelo; para una exposición detallada puede consultarse Basulto et al. (2006). Utilizaremos una distribución a priori no informativa para el modelo, cuyo núcleo viene dado por

$$\pi(\beta, \sigma) \propto \sigma^{-1}.$$

El núcleo de la distribución a posteriori conjunta, lo obtendremos multiplicando la verosimilitud por la distribución a priori, con lo que obtendremos

$$\pi(\beta, \sigma | y, X) \propto \sigma^{-(n+1)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_i (y_i - x'_i \beta)^2 \right\}, \quad \beta \in B$$

que, alternativamente, podemos expresar como

$$\pi(\beta, \sigma | y, X) \propto \sigma^{-(n+1)} \exp \left\{ -\frac{1}{2\sigma^2} (y - X\beta)' (y - X\beta) \right\}, \quad \beta \in B$$

La manipulación algebraica de esta distribución es complicada, debido a la restricción $\beta \in B$, que hace que no sea fácil conocer los límites de integración y que lleva también a que las distribuciones marginales de cada componente β_j estén definidas a trozos. Para evitar esta dificultad y dar respuesta al problema de la inferencia, utilizaremos el algoritmo de Gibbs (Gelfand y Smith, 1990), ya que no resulta excesivamente complicado simular muestras de las distribuciones

unidimensionales de cada parámetro condicionadas al resto de parámetros. Por tanto, aplicaremos el algoritmo de Gibbs, obteniendo muestras simuladas de las distribuciones $\sigma|\beta, y, X$ y $\beta_j|\beta_{(j)}, \sigma, y, X$, $j = 1, \dots, k$ donde $\beta_{(j)}$ representa el vector β del cual se elimina la componente j -ésima.

La distribución de σ condicionada al resto de parámetros se obtiene de manera inmediata, pues la distribución a posteriori conjunta puede expresarse como

$$\pi(\beta, \sigma|y, X) \propto (\sigma^{-2})^{\frac{n+1}{2}} \exp\left\{-\frac{1}{2}(\sigma^{-2})(y - X\beta)'(y - X\beta)\right\}, \quad \beta \in B$$

y así obtenemos inmediatamente que

$$\sigma^{-2}|\beta, y, X \sim Ga\left(\frac{n+3}{2}, \frac{1}{2}(y - X\beta)'(y - X\beta)\right)$$

Como podemos observar, el núcleo de la distribución a posteriori es Normal-Gamma, aunque truncado a un cierto subconjunto B , por lo que, siendo $\hat{\beta}$ el estimador MCO de β , la densidad del vector β condicionada al resto de parámetros es

$$\beta|\sigma, X, y \sim N_k\left(\hat{\beta}, \sigma^2(X'X)^{-1}\right), \quad \beta \in B,$$

es decir, una distribución Normal k -dimensional truncada en el subconjunto $B \subseteq \mathbb{R}^k$.

En principio, podrían generarse valores de esta distribución a través del procedimiento:

1. Generar un valor $\beta_0 \sim N_k\left(\hat{\beta}, \sigma^2(X'X)^{-1}\right)$
2. Si $\beta_0 \in B$, aceptar dicho valor; si no, rechazarlo y volver a 1.

Sin embargo, en la práctica, es frecuente que la probabilidad de rechazo de este procedimiento sea tan elevada que el método no resulte viable. Una solución para este problema es analizar las distribuciones condicionadas unidimensionales, que resultan ser Normales truncadas; la generación de valores simulados de esta distribución puede llevarse a cabo fácilmente siguiendo el algoritmo propuesto en Devroye (1986), p. 380.

Para obtener muestras simuladas de $\beta_j|\beta_{(j)}, \sigma, y, X$, reordenando los parámetros, podemos considerar la siguiente partición

$$\beta = \begin{pmatrix} \beta_j \\ \beta_{(j)} \end{pmatrix} \quad y \quad \Sigma = \begin{pmatrix} \sigma_j^2 & \sigma'_{j(j)} \\ \sigma_{j(j)} & \Sigma_{j(j)} \end{pmatrix},$$

donde hemos llamado $\Sigma = \sigma^2(X'X)^{-1}$.

Si suponemos que todas las variables explicativas toman valores positivos, las distribuciones de cada componente del vector β condicionadas al resto de

parámetros son de la forma

$$\beta_j | \beta_{(j)}, \sigma, y, X \sim N \left(\hat{\beta}_j + \Sigma_{j(j)}^{-1} (\beta_{(j)} - \hat{\beta}_{(j)}), \sigma_j^2 - \sigma'_{j(j)} \Sigma_{j(j)}^{-1} \sigma_{j(j)} \right), \\ \beta_j \geq b(\beta_{(j)}, X, y)$$

donde $b(\beta_{(j)}, X, y) = \max_i \left\{ (y_i - x'_{i(j)} \beta_{(j)}) / x_{ij} \right\}$.

En los modelos de producción, es habitual medir las variables en términos de logaritmos, por lo que aparecen con frecuencia valores negativos para las covariables; si no imponemos la hipótesis de que las variables explicativas sean positivas, la distribución condicionada sería la misma, cambiando el recorrido del parámetro, que sería de la forma $b_1(\beta_{(j)}, X, y) \geq \beta_j \geq b_2(\beta_{(j)}, X, y)$, donde

$$b_1(\beta_{(j)}, X, y) = \min_{i / x_{ij} < 0} \left\{ (y_i - x'_{i(j)} \beta_{(j)}) / x_{ij} \right\} \\ b_2(\beta_{(j)}, X, y) = \max_{i / x_{ij} > 0} \left\{ (y_i - x'_{i(j)} \beta_{(j)}) / x_{ij} \right\}$$

Una vez conocidas todas las distribuciones condicionadas unidimensionales, el algoritmo de Gibbs para nuestro modelo puede formularse como se especifica a continuación. Dados los valores β^m y $(\sigma^2)^m$, obtenidos en la etapa m -ésima, los valores generados en la etapa siguiente vendrían dados por:

1. Generar un valor v^{m+1} de la distribución $\sigma^{-2} | \beta^m, y, X$.
Tomar $(\sigma^2)^{m+1} = 1/v^{m+1}$
2. Generar un valor β_j^{m+1} de la distribución $\beta_j | \beta_1^{m+1}, \dots, \beta_{j-1}^{m+1}, \beta_{j+1}^m, \dots, \beta_k^m, (\sigma^2)^{m+1}, y, X, j = 1, \dots, k$
3. $\beta^{m+1} = (\beta_1^{m+1}, \dots, \beta_k^{m+1})'$

Como valor inicial β^0 hemos considerado el estimador obtenido a través del método MCC. Hemos implementado este algoritmo en el entorno Mathematica 5.2, generando muestras de la distribución a posteriori de tamaño 5000. Para estimar los parámetros, hemos utilizado el criterio de la esperanza de las distribuciones marginales.

5. Descripción del procedimiento de simulación llevado a cabo

Para comparar los resultados obtenidos a través del estimador MCC con los que proporciona la Estimación Bayesiana, hemos llevado a cabo un análisis de simulación que hemos configurado de la forma que explicamos a continuación.

Se han considerado, para simplificar, modelos con una única variable explicativa (es de esperar que la presencia de más variables explicativas no influya en

el mejor o peor comportamiento de los estimadores). Así, todas las simulaciones se han llevado a cabo sobre modelos de la forma $y_i = \alpha + \beta x_i + \varepsilon_i$, $i = 1, \dots, n$, donde las perturbaciones ε_i verifican las condiciones expuestas en el epígrafe 2.

Hemos asignado valores distintos a los parámetros α, β, σ y al tamaño muestral n . Concretamente, hemos considerado los casos $(\alpha, \beta) = \{(1, 1), (5, 2), (2, 5)\}$, $\sigma = \{1, 2\}$ y $n = \{5, 10, 20, 50, 100\}$. Además, en todos los casos, se ha duplicado la simulación, considerando primero una muestra de valores de x según un modelo $U(0, 1)$ y posteriormente según un modelo $U(0, 5)$.

En cada situación, una vez generada la muestra de valores de la covariable x se ha procedido a generar 1000 muestras de la variable y del modelo y a obtener los estimadores bayesiano y MCC, comparando los resultados con los parámetros fijados y conocidos de antemano. Concretamente, hemos calculado el sesgo medio y el error cuadrático medio de cada uno de los estimadores en las 1000 muestras simuladas. Por ejemplo, siendo $\hat{\beta}^i$ el estimador de β (por cualquiera de los métodos) en la muestra simulada i -ésima, el SM y el ECM vendrían dados por

$$SM = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\beta}^i - \beta)$$

$$ECM = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\beta}^i - \beta)^2$$

6. Resultados y conclusiones

Los resultados obtenidos en las 60 situaciones resultantes de combinar los distintos valores de los parámetros con los distintos tamaños muestrales y valores de x , se resumen en las 6 tablas que ofrecemos a continuación.

Tabla 1: $\alpha = 1; \beta = 1; x \in U(0, 1)$

		$\sigma = 1$				$\sigma = 2$			
		α		β		α		β	
		SM	ECM	SM	ECM	SM	ECM	SM	ECM
$n = 5$	MCC	-0.3167	0.7929	-0.0219	1.6648	-0.5505	1.7499	-0.0938	3.8169
	BAYES	-0.1895	0.6660	0.0711	1.4849	-0.2858	1.3362	-0.0108	2.8064
$n = 10$	MCC	-0.0487	0.2581	-0.1263	0.7870	-0.2653	0.7349	0.0841	4.1709
	BAYES	-0.0230	0.1568	-0.0333	0.4736	-0.0830	0.3645	-0.0183	2.0755
$n = 20$	MCC	-0.0409	0.0933	0.0256	0.2945	-0.0515	0.3315	-0.0377	1.3468
	BAYES	-0.0232	0.0357	0.0189	0.1035	-0.0102	0.1379	-0.0403	0.5890
$n = 50$	MCC	0.0284	0.0544	-0.0077	0.1377	0.0420	0.1402	-0.0035	0.4259
	BAYES	-0.0040	0.0107	0.0028	0.0273	0.0049	0.0349	-0.0059	0.0912
$n = 100$	MCC	0.0270	0.0222	0.0009	0.0687	0.0923	0.0881	-0.0368	0.2790
	BAYES	0.0054	0.0028	-0.0122	0.0097	0.0046	0.0063	-0.0112	0.0244

Tabla 2: $\alpha = 1; \beta = 1; x \in U(0, 5)$

		$\sigma = 1$				$\sigma = 2$			
		α		β		α		β	
		SM	ECM	SM	ECM	SM	ECM	SM	ECM
$n = 5$	MCC	-0.3543	0.7664	0.0064	0.0810	-0.6764	3.1046	0.0081	0.3473
	BAYES	-0.1964	0.6402	0.0158	0.0738	-0.4571	2.5635	0.0497	0.2907
$n = 10$	MCC	-0.1662	0.5703	0.0156	0.0500	-0.2319	1.5424	-0.0044	0.1460
	BAYES	-0.0867	0.4226	0.0173	0.0360	-0.0737	0.8436	-0.0069	0.0830
$n = 20$	MCC	-0.0338	0.2202	0.0041	0.0202	-0.0489	0.5551	0.0023	0.0664
	BAYES	-0.0185	0.0870	0.0034	0.0070	-0.0179	0.2012	-0.0011	0.0237
$n = 50$	MCC	0.0208	0.0418	0.0025	0.0057	0.0589	0.1930	-0.0064	0.0235
	BAYES	-0.0109	0.0079	0.0042	0.0011	-0.0037	0.0417	-0.0009	0.0054
$n = 100$	MCC	0.0429	0.0260	-0.0050	0.0031	0.0723	0.0883	0.0000	0.0107
	BAYES	0.0023	0.0023	-0.0012	0.0003	-0.0049	0.0082	0.0018	0.0009

Tabla 3: $\alpha = 5; \beta = 2; x \in U(0, 1)$

		$\sigma = 1$				$\sigma = 2$			
		α		β		α		β	
		SM	ECM	SM	ECM	SM	ECM	SM	ECM
$n = 5$	MCC	-0.4230	1.7107	0.1540	3.1552	-0.3289	0.7508	-0.0145	1.3230
	BAYES	-0.2979	1.4458	0.2286	2.7315	-0.1524	0.6587	-0.0002	1.2192
$n = 10$	MCC	-0.1626	0.3329	0.0701	0.7059	-0.1121	0.1418	-0.0062	0.4340
	BAYES	-0.0577	0.2174	0.0280	0.4208	-0.0325	0.0865	-0.0291	0.2441
$n = 20$	MCC	-0.0322	0.1377	0.0098	0.3624	-0.0305	0.1097	0.0057	0.2962
	BAYES	-0.0202	0.0619	0.0129	0.1486	-0.0120	0.0475	-0.0000	0.1170
$n = 50$	MCC	0.0265	0.0390	-0.0070	0.1263	0.0268	0.0422	-0.0013	0.1613
	BAYES	0.0048	0.0057	-0.0137	0.0226	0.0015	0.0094	-0.0072	0.0374
$n = 100$	MCC	0.0377	0.0224	-0.0076	0.0683	0.0238	0.0161	0.0205	0.0573
	BAYES	0.0023	0.0024	-0.0090	0.0076	-0.0042	0.0020	0.0072	0.0054

Tabla 4: $\alpha = 5; \beta = 2; x \in U(0, 5)$

		$\sigma = 1$				$\sigma = 2$			
		α		β		α		β	
		SM	ECM	SM	ECM	SM	ECM	SM	ECM
$n = 5$	MCC	-0.3139	0.7065	0.0027	0.1210	-0.6379	2.1700	-0.0417	0.3939
	BAYES	-0.1835	0.5677	0.0108	0.0927	-0.2248	1.8828	-0.0277	0.3938
$n = 10$	MCC	-0.1282	0.2113	0.0065	0.0333	-0.2499	1.0672	-0.0003	0.0986
	BAYES	-0.0377	0.1284	0.0046	0.0189	-0.0901	0.5964	-0.0060	0.0526
$n = 20$	MCC	-0.0211	0.1575	-0.0041	0.0176	-0.0519	0.7019	0.0016	0.0626
	BAYES	-0.0177	0.0659	0.0011	0.0070	-0.0401	0.3046	0.0052	0.0234
$n = 50$	MCC	0.0123	0.0368	0.0058	0.0062	0.0289	0.1701	0.0021	0.0250
	BAYES	-0.0023	0.0080	-0.0005	0.0012	-0.0118	0.0405	0.0031	0.0054
$n = 100$	MCC	0.0201	0.0188	0.0022	0.0026	0.0576	0.0789	0.0039	0.0108
	BAYES	-0.0028	0.0027	0.0009	0.0004	0.0032	0.0068	-0.0018	0.0011

Tabla 5: $\alpha = 2; \beta = 5; x \in U(0, 1)$

		$\sigma = 1$				$\sigma = 2$			
		α		β		α		β	
		SM	ECM	SM	ECM	SM	ECM	SM	ECM
$n = 5$	MCC	-0.3620	1.0393	0.0581	2.1763	-0.4684	3.8310	-0.2266	9.4462
	BAYES	-0.2103	0.8389	0.0796	1.8089	-0.2942	2.9078	0.0217	7.0655
$n = 10$	MCC	-0.0910	0.1907	-0.0362	0.6790	-0.3458	1.4631	0.2184	4.2833
	BAYES	-0.0223	0.0998	-0.0286	0.3899	-0.1585	0.8396	0.1563	2.5040
$n = 20$	MCC	-0.0621	0.1579	0.0540	0.4101	-0.0207	0.4570	-0.0291	1.0874
	BAYES	-0.0394	0.0864	0.0437	0.1899	-0.0381	0.1970	0.0357	0.4166
$n = 50$	MCC	0.0449	0.0534	-0.0376	0.2597	0.0135	0.1502	0.0573	0.7087
	BAYES	0.0095	0.0103	-0.0244	0.0482	-0.0070	0.0345	0.0016	0.1404
$n = 100$	MCC	0.0353	0.0238	-0.0127	0.0727	0.0613	0.0859	0.0140	0.2546
	BAYES	-0.0005	0.0031	-0.0022	0.0091	-0.0000	0.0087	0.0029	0.0247

Tabla 6: $\alpha = 2; \beta = 5; x \in U(0, 5)$

		$\sigma = 1$				$\sigma = 2$			
		α		β		α		β	
		SM	ECM	SM	ECM	SM	ECM	SM	ECM
$n = 5$	MCC	-0.3119	0.4679	0.0004	0.0504	-0.7007	5.0325	0.0052	0.3086
	BAYES	-0.1444	0.3431	-0.0014	0.0366	-0.5014	4.4014	0.0507	0.2746
$n = 10$	MCC	-0.1103	0.1827	-0.0024	0.0271	-0.2928	1.0243	0.0128	0.1058
	BAYES	-0.0117	0.1019	-0.0111	0.0170	-0.1504	0.5988	0.0153	0.0616
$n = 20$	MCC	0.0120	0.0911	-0.0112	0.0116	-0.0530	0.5086	0.0059	0.0715
	BAYES	0.0034	0.0298	-0.0043	0.0039	-0.0258	0.2238	-0.0006	0.0269
$n = 50$	MCC	0.0170	0.0483	0.0031	0.0061	0.0434	0.1468	-0.0019	0.0195
	BAYES	-0.0045	0.0100	0.0017	0.0012	-0.0012	0.0308	-0.0017	0.0044
$n = 100$	MCC	0.0290	0.0240	0.0015	0.0031	0.0453	0.0896	0.0065	0.0116
	BAYES	-0.0021	0.0029	0.0006	0.0003	-0.0057	0.0132	0.0007	0.0014

En estas tablas hemos resaltado con fondo gris aquellas comparaciones que resultan favorables al estimador MCC frente al estimador Bayes; por contra las celdillas con fondo blanco corresponden a las situaciones en las que el estimador Bayes ha mostrado un mejor comportamiento.

La primera y principal conclusión que podemos obtener es que el estimador Bayes presenta un mejor comportamiento que el MCC en las 60 situaciones analizadas. Este resultado se mantiene para todos los valores de los parámetros, del tamaño muestral y de los valores de las covariables (de su mayor o menor dispersión). Tan sólo en algunos casos el estimador MCC presenta un mejor sesgo medio para la pendiente (y en un único caso para la ordenada en el origen), aunque siempre el ECM es superior al del estimador Bayes.

Otra característica que se repite en todas las situaciones es que las diferencias entre la estimación Bayes y MCC son mucho más acusadas en el caso de la ordenada en el origen que en el caso de la pendiente, en el que los resultados son más similares entre ambas metodologías (aunque favorables al estimador Bayes). Como es conocido y hemos mencionado, el estimador por MCC de la ordenada en el origen, aunque consistente, es sesgado y de "dudosa eficiencia", y según los resultados obtenidos en nuestra simulación el estimador Bayes consigue una notable mejora tanto en sesgo como en ECM, sobre todo para tamaños muestrales pequeños.

Como era de esperar, en ambos casos disminuyen el SM y el ECM a medida que el tamaño de la muestra va aumentando. Observemos también que cuando los datos de x presentan mayor dispersión (es decir, cuando $x \sim U(0, 5)$) manteniendo fija σ , disminuye el ECM de la pendiente para ambos estimadores, como cabía esperar. Asimismo, al aumentar σ manteniendo fija la dispersión de x , aumenta el ECM de la pendiente en ambos casos.

Cabe destacar también que aunque el estimador Bayes consigue una significativa mejora en el sesgo de la ordenada en el origen, sigue mostrando al igual que el estimador MCC persistencia de sesgo negativo, especialmente para pequeñas muestras. Este hecho conllevará que, en general, ambos métodos tenderán a subestimar el grado real de ineficiencia de los procesos productivos (o equivalentemente estimarán un grado de eficiencia superior al real).

Referencias

- [1] Aigner D.J. y Chu S.F. (1968). On estimating the industry production function. *Am. Econ. Rev.*, **58**, 826-839.
- [2] Basulto J., Ortega F.J. y Camúñez J.A. (2006). Estimación Bayesiana del modelo de regresión half-Normal con frontera determinista. *Anales de Economía Aplicada 2006, Delta publicaciones, Madrid*, 375-400.

- [3] Cazals C., Florens J.P. y Simar L. (2002). Nonparametric frontier estimation: a robust approach. *J. Econometrics*, **106**, 1-25.
- [4] Daouia A. y Simar L. (2004). Nonparametric efficiency analysis: a multivariate conditional quantile approach. *J. Econometrics*, **140**, 375-400.
- [5] Devroye L. (1986). *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.
- [6] Gelfand A.E. y Smith A.F. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.*, **85**, 398-409.
- [7] Green W.H. (1980). Maximum Likelihood Estimations of econometric frontier functions. *J. Econometrics*, **13**, 27-56.
- [8] Jondrow J., Novell C.A., Materov I.S. y Schmidt P. (1982). On estimation of technical inefficiency in the stochastic frontier production function model. *J. Econometrics*, **19**, 233-238.
- [9] Simar L. (2007). How to improve the performances of DEA/FDH estimators in the presence of noise?. *J. Prod. Anal.*, **28**, 183-201.

Acerca de los autores

Fco. Javier Ortega es licenciado en Matemáticas, doctor en Economía y profesor de Estadística y Econometría en la Universidad de Sevilla. Sus líneas de investigación principales son métodos de Inferencia Bayesiana (con especial interés en modelos no regulares) e Historia de la Probabilidad y la Estadística.

Jesús Basulto es Catedrático de Estadística en la Universidad de Sevilla. Su línea de investigación principal es Historia de la Probabilidad y la Estadística.

José Antonio Camúñez es licenciado en Matemáticas, doctor en Economía y profesor de Estadística y Econometría en la Universidad de Sevilla. Sus líneas de investigación principales son Análisis de Datos Multivariante e Historia de la Probabilidad y la Estadística.