

Modelo de producción trans-log con frontera estocástica: ¿Estimación máximo verosímil o bayesiana?

Francisco Javier Ortega Irizo

e-mail: fjortega@us.es

Tfno: +34 954556970

José Antonio Camúñez Ruiz

José Manuel Gavilán Ruiz

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

RESUMEN

En este artículo se analizan y comparan las propiedades muestrales de los estimadores máximo verosímil y bayesiano en un modelo de producción translog con frontera estocástica, a través de un estudio de tipo Monte Carlo. Los resultados indican que el estimador Bayes es preferible al máximo verosímil, dado que el comportamiento del error cuadrático medio es significativamente mejor en el caso bayesiano.

Palabras Claves: Frontera estocástica, estimación máximo verosímil, estimación bayesiana, Monte Carlo.

Área temática: Economía y Metodología. Métodos Cuantitativos. Didáctica de la Economía.

Stochastic frontier trans-log production function: maximum likelihood or Bayesian estimation?

ABSTRACT

In this paper we investigate and compare the finite sample properties of the maximum likelihood and Bayesian estimators of the stochastic frontier trans-log production function, through a Monte Carlo study. The results show that Bayesian estimator should be used in preference to maximum likelihood owing to the fact that the mean square error performance is substantially better in the Bayesian framework.

Key words: stochastic frontier, maximum likelihood, Bayesian estimator, Monte Carlo.

Topic: Economy and Methodology. Quantitative Methods.

Modelo de producción trans-log con frontera estocástica: ¿Estimación máximo verosímil o bayesiana?

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 pueden ser identificadas con perturbaciones aleatorias en un modelo de regresión.

Aigner y Chu (1968), siguiendo la iniciativa de Farrel (1957), 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.

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 (MV) (Ortega y Basulto, 2009; Ortega *et al.* 2009). 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 y Corra (1977) y Meusen y 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 variaciones aleatorias.

En estos modelos, se verifican las condiciones de regularidad habituales, por lo que podemos aplicar las propiedades asintóticas del estimador MV 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.

Alternativamente, podemos usar el enfoque bayesiano para estimar los parámetros del modelo, así como la ineficiencia (o eficiencia) de cada una de las firmas analizadas. El uso de la metodología bayesiana presenta algunas ventajas, como la facilidad para incorporar restricciones sobre los parámetros, pero sobre todo el uso de métodos tipo Monte Carlo con Cadenas de Markov a través del algoritmo de Gibbs (Gelfand y Smith, 1990) es una herramienta muy adecuada en este contexto debido a la complejidad de los modelos de frontera estocástica. De hecho, gran parte de la literatura reciente sobre el tema utiliza la metodología Bayesiana (Griffin y Steel 2007, Dorfman y Koop 2005).

Uno de los problemas principales del enfoque bayesiano, sobre todo de cara a las aplicaciones, ha sido la falta de un software fácil de utilizar y lo suficientemente versátil como para adaptarse a distintas situaciones y conjuntos de datos. En Griffin y Steel (2007), se utiliza el software libre WinBUGS para estimar diversas variantes de modelos de producción con frontera estocástica, comprobándose que este software es una herramienta potente y flexible en este contexto, que puede ser manejado con facilidad por parte del investigador y que consigue estimar los modelos en un tiempo razonable. Obviamente, una implementación específica para un modelo concreto puede ser más eficiente que WinBUGS, que es una herramienta genérica que puede ser usada en una gran variedad de situaciones, aunque la facilidad de uso y la posibilidad de aplicar análisis bayesiano sin necesidad de programar son ventajas evidentes.

El objetivo principal de este artículo es llevar a cabo un análisis de simulación para comparar los resultados del método MV (usando el software FRONTIER) con los obtenidos a través del enfoque bayesiano (usando el software WinBUGS). Para ello, aprovechamos que tanto uno como otro pueden integrarse en el entorno del software R, lo que facilita la aplicación de ambas metodologías al mismo conjunto de muestras simuladas para que, de esta forma, la comparación de resultados sea más fiable. Nuestro diseño del experimento Monte Carlo será similar al usado en Coelli (1995), donde se compara el estimador MV con el de mínimos cuadrados corregidos. Es importante señalar que en Zhang (1999) también se compara la estimación bayesiana con la MV, aunque ésta

se hace para unos valores únicos del espacio paramétrico y poniendo especial interés en analizar el efecto de usar diferentes distribuciones a priori. En Ortega *et al.* (2009) se lleva a cabo una comparación entre ambas metodologías aplicadas sobre el modelo de frontera determinista.

Aunque analizaremos también el sesgo de los estimadores, como criterio de comparación usaremos el error cuadrático medio (ECM), haciendo especial énfasis en el parámetro que indica qué proporción de varianza del error compuesto se debe a ineficiencia y en la estimación de las ineficiencias (o eficiencias) individuales.

A partir de aquí, en la sección 2 presentamos el modelo de producción sobre el que se realiza el análisis; en la sección 3 explicamos cómo obtener los estimadores MV y bayesianos usando FRONTIER y WinBUGS, respectivamente, todo ello en el entorno del software R; en la sección 4 describimos el diseño del experimento Monte Carlo; en la sección 5 ofrecemos los resultados obtenidos, así como la interpretación de los aspectos más interesantes. Por último, en la sección 6 exponemos las principales 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 $f(\bullet)$ es la función de producción.

La perturbación aleatoria ε_i se compone de dos partes (de ahí que también hablemos 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.

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

representan el efecto de la ineficiencia en la producción, suponemos la hipótesis más habitual, es decir, que $u_i = |u_i^*|$, 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 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 ha considerado con frecuencia otro tipo de densidad de probabilidad para las perturbaciones u_i , entre las que cabe destacar las distribuciones Exponencial y Gamma, que han sido ampliamente utilizadas sobre todo cuando se adopta el enfoque bayesiano (Koop et al. 1995, Osiewalski and Steel 1998, Koop and Stell 2003)

Con respecto a la función de producción, consideraremos un modelo de tipo translog que ha sido ampliamente utilizado debido a su versatilidad. Concretamente, siendo k el número de inputs considerados la función de producción es:

$$f(x_i, \beta) = \beta_0 + \sum_{r=1}^k \beta_r x_{ir} + \frac{1}{2} \sum_{r,s=1}^k \beta_{rs} x_{ir} x_{is} \quad i = 1, \dots, n.$$

Impondremos la condición de simetría, es decir, $\beta_{rs} = \beta_{sr} \quad \forall s, r$ y además tanto los inputs como los outputs se consideran en escala logarítmica. Por ello, en lugar de interesarnos directamente en u_i como medida de ineficiencia, se suele estimar $\exp(-u_i)$ que es una medida de eficiencia acotada entre 0 y 1.

Para obtener los estimadores MV, en el paquete FRONTIER se considera la parametrización $\sigma^2 = \sigma_v^2 + \sigma_u^2$ y $\gamma = \sigma_u^2 / \sigma^2$. Observemos que γ es un parámetro que toma valores entre 0 y 1 y es un indicador de la proporción de varianza debida a ineficiencia. Resaltemos que γ no es exactamente la proporción de varianzas, ya que $\text{var}[u] = p \sigma_u^2$, donde $p = (\pi - 2) / \pi$. Concretamente, si llamamos γ^* a la proporción de la varianza total debida a ineficiencia (es decir, $\gamma^* = \text{var}[u] / (\text{var}[u] + \text{var}[v])$) es fácil ver que se verifica $\gamma^* = \gamma / (\gamma + (1 - \gamma)p^{-1})$.

3. MÉTODOS DE ESTIMACIÓN UTILIZADOS

Como se ha indicado en la introducción, la estimación MV la hemos llevado a cabo usando el software FRONTIER y la estimación bayesiana a través de WinBUGS, ambos integrados en el entorno del software R.

Concretando más, para el estimador MV usamos el paquete *frontier* en su versión 0.996-6 el cual utiliza el código fuente *Fortran* del software FRONTIER 4.1 (Coelli, 1996). Para la inferencia bayesiana se ha usado el paquete *R2WinBUGS*, el cual enlaza y abre el programa WinBUGS v1.4, le transfiere los datos y recoge los resultados hacia el entorno R. El software WinBUGS puede descargarse desde la dirección web <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml> y el software R puede obtenerse en <http://www.r-project.org>, siendo ambos de distribución libre. Una vez ejecutado R, debemos instalar y cargar los paquetes *frontier* y *R2WinBUGS*. Para los detalles y opciones de ambos paquetes, puede consultarse la ayuda incluida en los mismos. Una detallada exposición del funcionamiento del paquete *R2WinBUGS* junto con algunos ejemplos de aplicación puede encontrarse en Sturtz *et al.* (2005).

En el diseño del experimento de simulación se ha considerado, por simplicidad, un modelo con dos variables explicativas más una ordenada en el origen, de forma que las simulaciones se hacen en base al modelo:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \frac{1}{2}(\beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2) + \beta_{12} x_{i1} x_{i2} + v_i - u_i, \quad i = 1, \dots, n.$$

Para obtener las estimaciones bayesianas (tanto de los parámetros como de las eficiencias), consideramos las distribuciones a priori que se utilizan en Griffin y Steel (2007), salvo el hiperparámetro correspondiente a la varianza de la distribución Half-Normal. Concretamente, utilizamos distribuciones a priori no informativas para la ordenada en el origen y para todas las pendientes (se supone que siguen una $N(0, 10^6)$) y para la varianza de la perturbación aleatoria (se supone $\sigma_v^{-2} \sim Ga(10^{-3}, 10^{-3})$), donde $N(\mu, \delta^2)$ representa una distribución Normal de media μ y varianza δ^2 mientras que $Ga(a_0, a_1)$ representa una distribución Gamma con media a_0/a_1 y parámetro de forma a_0 .

La distribución a priori para el parámetro σ_u^{-2} se suele elegir dentro de la familia Gamma, fijando los hiperparámetros de forma que la mediana a priori recoja nuestras

creencias acerca de la eficiencia mediana del sector productivo en el que se está llevando a cabo el análisis (Van den Broeck *et al.* 1994, Zhang, X. 1999). Por ejemplo, Griffin y Steel (2007), en su análisis aplicado al mercado de la electricidad en Estados Unidos, usan la distribución a priori $\sigma_u^{-2} \sim Ga(1, 0.0267)$, lo que supone asumir que la eficiencia mediana a priori del sector es 0.875. En nuestro caso, dado que se hace un estudio de simulación, hemos elegido como distribución de referencia a priori $\sigma_u^{-2} \sim Ga(1, 1)$, que se corresponde con una eficiencia mediana a priori igual a 0.44 (Van de Broeck *et al.* 1994).

En cuanto a la implementación del modelo en WinBUGS, se han generado tres cadenas de Markov de 1500 iteraciones cada una, de las cuales se desechan las 500 primeras, lo que conlleva un total de 3000 muestras simuladas. Como puntos de partida se toman los estimadores de Mínimos Cuadrados Corregidos para la ordenada en el origen y las pendientes de la función de producción (Green, 1980), mientras que para el resto de parámetros se generan valores aleatorios.

4. DISEÑO DEL EXPERIMENTO MONTE CARLO.

Como hemos indicado en la sección anterior, las simulaciones se hacen sobre un modelo trans-log con dos variables explicativas. El espacio muestral del experimento viene dado inicialmente por los distintos valores de los parámetros $\vec{\beta}^t = (\beta_0, \beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12})$, σ^2 y γ , del tamaño muestral n así como de los valores del vector de observaciones de las covariables x_1 y x_2 . El objetivo fundamental es analizar el comportamiento de las estimaciones del parámetro γ y de las eficiencias individuales.

Sin pérdida de generalidad, podemos fijar los valores de $\vec{\beta}$ y σ^2 (Zhang 1999, Coelli, 1995). Nosotros consideraremos $\vec{\beta}^t = (1, 1, 1, 1, 1, 1)$ y $\sigma^2 = 1$. Con respecto a los valores de las covariables x_1 y x_2 , se han generado de acuerdo a modelos uniformes independientes en el intervalo $[0, 10]$. Pretendemos analizar el comportamiento de ambos métodos tanto en muestras pequeñas como en muestras de gran tamaño, por lo que hemos seleccionado $n \in \{30, 50, 100, 500\}$. Con respecto al parámetro γ , siguiendo a Coelli (1995), se fijarán valores del parámetro γ^* (proporción de la varianza total debida a ineficiencia) desde 0 a

1 (ambos inclusive) en saltos de 0.2. Teniendo en cuenta la relación $\gamma^* = \gamma / (\gamma + (1-\gamma)p^{-1})$, obtenemos la siguiente correspondencia entre γ y γ^* :

γ^*	0.0	0.2	0.4	0.6	0.8	1.0
γ	0.00	0.41	0.65	0.80	0.92	1.00

Se simulan, pues, un total de 6 valores de γ y 4 valores de n , lo que conlleva 24 combinaciones. En cada una de ellas, hacemos un total de $m = 1000$ replicaciones del modelo. Para obtener los números pseudoaleatorios, utilizamos los generadores implementados por defecto en el software R. La estimación MV y bayesiana en cada caso, se efectúa según el procedimiento descrito en el epígrafe anterior.

Para cada parámetro, se calcula el sesgo medio (SM) y el error cuadrático medio (ECM) observado en las m replicaciones con cada uno de los métodos. Para el análisis de las eficiencias, no nos limitaremos a considerar la eficiencia media, sino que calcularemos tanto el SM como el ECM para cada una de las eficiencias individualmente. Posteriormente, como indicador conjunto ofrecemos el promedio de los SM y los ECM de cada una de las firmas. Es decir, si llamamos E_i a la eficiencia de la firma i -ésima y \hat{E}_{ij} al correspondiente valor estimado (por cualquiera de los dos métodos) en la replicación j -ésima, tenemos que $SM_i = m^{-1} \sum_{j=1}^m (\hat{E}_{ij} - E_i)$ y $ECM_i = m^{-1} \sum_{j=1}^m (\hat{E}_{ij} - E_i)^2$, para $i = 1, \dots, n$.

Posteriormente, el SM y el ECM de las eficiencias se obtienen como $SM = n^{-1} \sum_{i=1}^n SM_i$ y

$$ECM = n^{-1} \sum_{i=1}^n ECM_i, \text{ respectivamente.}$$

5. RESULTADOS DEL EXPERIMENTO MONTE CARLO.

Los resultados completos del experimento Monte Carlo, es decir, sesgo y ECM de todos los parámetros y de las eficiencias individuales en las 24 situaciones consideradas se ofrecen en el Anexo. Aquí vamos a analizar en mayor profundidad los resultados más relevantes, centrandó nuestra atención en el criterio del ECM, que es al fin y al cabo el que indica qué metodología consigue reproducir mejor los verdaderos valores del modelo considerado. Nos centraremos en el comportamiento de las estimaciones del parámetro γ

y de las eficiencias individuales. Como ya hemos indicado, el parámetro γ es el que recoge la estructura del error compuesto y, como es conocido, uno de los inconvenientes de este tipo de modelos es la dificultad para identificar qué proporción del error total se debe a ineficiencia y qué parte se debe a los efectos aleatorios. Por otra parte, la correcta estimación de γ es muy importante para establecer las eficiencias individuales de cada firma, que es uno de los objetivos fundamentales al usar este tipo de modelos.

En la Tabla I recogemos el ECM correspondiente al parámetro γ y a las eficiencias en función de los valores de γ^* para los diferentes tamaños muestrales y para ambos métodos de estimación. Estos valores se han representado gráficamente en las figuras 1 y 2. Como se indicó en el epígrafe anterior, en todos los casos se ha obtenido el ECM de las eficiencias individuales y posteriormente se ha calculado la media de todos ellos, siendo este promedio el que se ofrece en la tabla I y se representa en la figura 2.

Comparando por pares los resultados de la estimación MV y la bayesiana, hemos resaltado en fondo oscuro aquella que presenta un mejor comportamiento en cada caso, es decir, la que tiene un menor valor de ECM.

Tabla I. ECM correspondiente al parámetro γ y a las eficiencias.

γ^*	n=30				n=50				n=100				n=500			
	MV		BAYES		MV		BAYES		MV		BAYES		MV		BAYES	
	γ	EFI														
0	0.4403	0.1634	0.5833	0.3103	0.2832	0.1444	0.3884	0.2529	0.1833	0.1123	0.2923	0.2150	0.0816	0.0785	0.1830	0.1712
0.2	0.2353	0.1281	0.1567	0.0738	0.1634	0.1099	0.0834	0.0553	0.1144	0.0884	0.0489	0.0467	0.0514	0.0596	0.0150	0.0388
0.4	0.2453	0.1359	0.0404	0.0504	0.1681	0.0990	0.0227	0.0428	0.1135	0.0731	0.0111	0.0377	0.0266	0.0408	0.0071	0.0352
0.6	0.2527	0.1219	0.0084	0.0394	0.1390	0.0752	0.0098	0.0336	0.0621	0.0452	0.0095	0.0303	0.0040	0.0276	0.0040	0.0275
0.8	0.2233	0.0981	0.0028	0.0298	0.0569	0.0399	0.0051	0.0222	0.0145	0.0232	0.0053	0.0193	0.0006	0.0156	0.0007	0.0156
1	0.1601	0.0724	0.0069	0.0208	0.0016	0.0066	0.0002	0.0033	0.0002	0.0033	0.0003	0.0024	0.0000	0.0006	0.0000	0.0002

En la figura 1 representamos el ECM del parámetro γ . En el caso del método MV el gráfico se corresponde con la figura 2 de Coelli (1995). Como puede verse, los resultados son prácticamente coincidentes para los tamaños $n=50$ y $n=100$, mientras que nuestra gráfica para $n=500$ es muy similar a la que presenta Coelli (1995) para $n=400$.

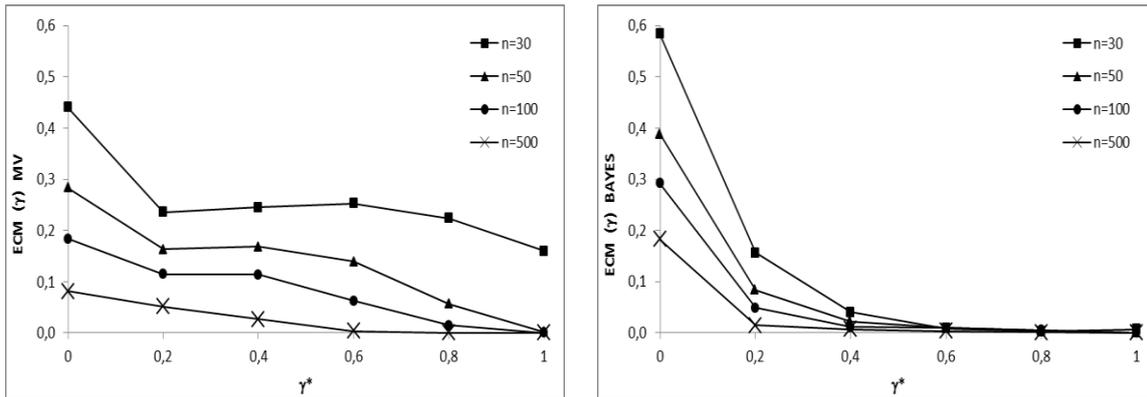


Figura 1: ECM del parámetro γ^* para ambos métodos de estimación

Como puede apreciarse claramente, el patrón de comportamiento del ECM del estimador MV es muy diferente al de la metodología bayesiana, siendo favorable para esta última. En efecto, aunque para el caso extremo $\gamma^* = 0$ el ECM bayesiano es superior al MV, a partir de $\gamma^* = 0.2$ el ECM bayesiano desciende rápidamente, situándose por debajo del correspondiente MV (en el caso $n = 500$ los valores de ECM tienden a igualarse a partir de $\gamma^* = 0.6$). Por tanto, la metodología bayesiana muestra un mejor comportamiento en cuanto a ECM en la mayor parte del espacio paramétrico de γ^* . Es también interesante destacar que el ECM en general es decreciente como función de γ^* , circunstancia que se aprecia con mucha mayor claridad en el caso bayesiano. Entonces, tanto una como otra metodología estiman con mayor precisión el modelo cuando la proporción de varianza debida a ineficiencia es elevada.

Indiquemos también que, como es de esperar, los ECM disminuyen con ambos métodos de estimación a medida que aumentamos el tamaño muestral. Este comportamiento puede observarse para todos los parámetros analizados.

En la figura 2, representamos el ECM correspondiente a las eficiencias individuales. Los comentarios hechos para la figura 1 son igualmente válidos para la figura 2, ya que el patrón de comportamiento es básicamente el mismo, si bien en el caso de la eficiencia la escala del ECM es algo menor en los dos enfoques considerados.

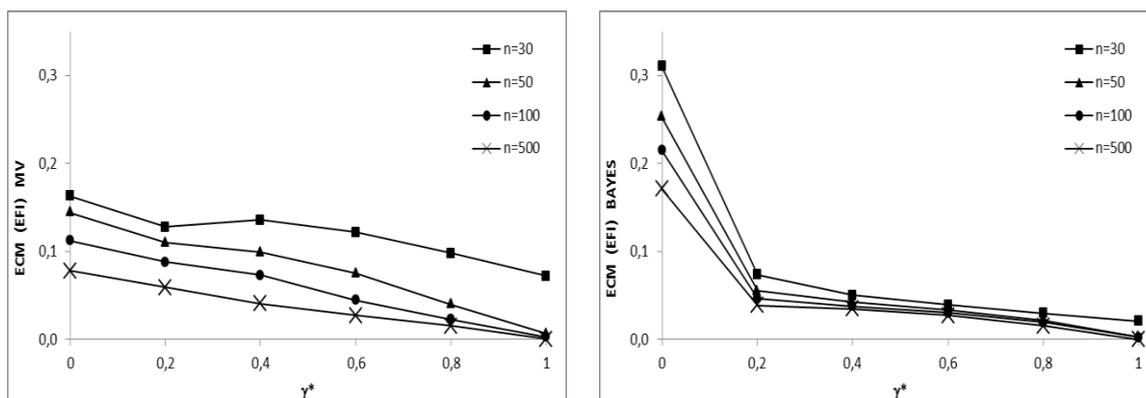


Figura 2: ECM medio de las eficiencias individuales para ambos métodos de estimación

La similitud entre las figuras 1 y 2 pone de manifiesto la incidencia directa que tiene la estimación del parámetro γ sobre la determinación de las eficiencias de las firmas.

Para cuantificar la diferencia de comportamiento entre ambos métodos de estimación, en la Tabla II hemos calculado los ratios entre el ECM máximo verosímil y el ECM bayesiano en cada caso. Así, valores superiores a 1 (resaltados en la tabla con fondo oscuro) se identifican con un mejor comportamiento del estimador bayesiano, mientras que valores inferiores a 1 indican ventaja para el estimador MV. Estos ratios se representan gráficamente en la figura 3.

Tabla II. Ratios de ECM (MV/Bayes).

γ^*	n=30		n=50		n=100		n=500	
	γ	EFI	γ	EFI	γ	EFI	γ	EFI
0	0.7548	0.5264	0.7291	0.5709	0.6273	0.5225	0.4457	0.4585
0.2	1.5021	1.7357	1.9591	1.9864	2.3404	1.8953	3.4236	1.5358
0.4	6.0785	2.6969	7.4213	2.3137	10.1841	1.9379	3.7506	1.1579
0.6	30.0566	3.0946	14.1293	2.2367	6.5683	1.4916	1.0040	1.0024
0.8	80.3456	3.2934	11.1347	1.7964	2.7300	1.1993	0.8238	0.9991
1	23.0864	3.4774	9.6686	1.9922	0.5678	1.3716	22.3813	3.2049

Como ya habíamos observado en las figuras 1 y 2, el comportamiento del estimador bayesiano es mejor salvo en el caso extremo $\gamma^* = 0$. No obstante, los ratios ponen de manifiesto que además las diferencias entre ambos métodos son realmente importantes. Por ejemplo, si analizamos el caso del parámetro γ con tamaño muestral $n = 50$, para el valor $\gamma^* = 0.6$ el ECM del estimador MV es del orden de 14 veces superior al que presenta el estimador bayesiano. Las diferencias siguen siendo importantes también cuando consideramos $n = 100$ e incluso con $n = 500$, ya que en este último caso cuando

γ^* varía entre 0.2 y 0.4, los ratios observados son del orden de 3.5. Si analizamos las eficiencias, observamos un comportamiento parecido al caso del parámetro γ , aunque a una escala bastante menor. En este caso los mayores ratios se sitúan en torno a 3.5, si bien todos ellos son favorables al estimador bayesiano (salvo $\gamma^* = 0$).

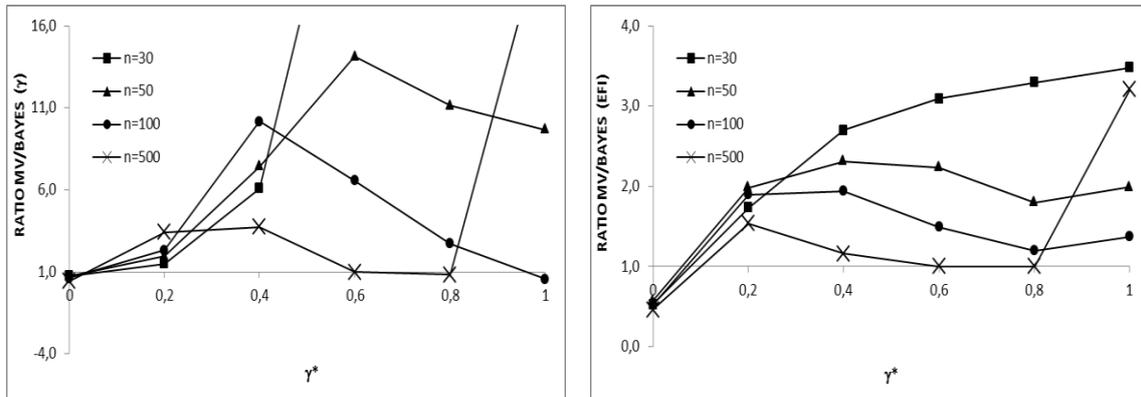


Fig. 3: Ratios de ECM para γ y para las eficiencias

En la figura 3, al representar los ratios de ECM correspondientes al parámetro γ , hemos preferido elegir un valor máximo de 16 en el eje OY para poder apreciar con claridad las diferentes gráficas, aunque ello provoque que algunos ratios especialmente elevados queden fuera de la escala elegida. Indicar también que el ratio de ECM para el parámetro γ correspondiente a $\gamma^* = 1$ y $n = 500$, puede ser poco representativo debido a que ambos valores ECM son muy cercanos a 0 (concretamente, con una precisión de 8 decimales los ECM que se obtienen son 0.00003499 para el estimador MV y 0.00000156 para el bayesiano).

En base al criterio de comparación utilizado, los resultados indican una clara diferencia a favor de la metodología bayesiana, por lo que sugerimos su uso práctico para la estimación del modelo de producción con frontera estocástica analizado en este trabajo.

Para finalizar el análisis, vamos a efectuar algunos comentarios generales acerca del ECM del resto de parámetros, así como de todos los sesgos.

Con respecto a las pendientes la conclusión más relevante es que ambos métodos consiguen estimarlas correctamente y con la misma precisión, ya que tanto los sesgos como los ECM son cercanos a 0 en todos los casos (algo mayores cuando $n = 30$) y

similares entre sí. Cabe destacar que las pendientes correspondientes a los términos cuadráticos (es decir, β_{11} , β_{22} y β_{12}) se estiman con bastante más precisión con ambas metodologías. Por su parte, la ordenada en el origen β_0 presenta el mismo comportamiento que el parámetro γ y las eficiencias, pues el ECM es siempre menor con la metodología bayesiana salvo en el caso $\gamma^* = 0$. El ECM del parámetro σ^2 muestra un comportamiento más diverso, pues para el tamaño muestral $n=30$ es casi siempre inferior en el estimador MV, mientras que para los tamaños $n=50$, $n=100$ y $n=500$ el ECM bayesiano es inferior al MV a partir de $\gamma^* = 0.4$ en adelante.

Lo más destacable del comportamiento de los sesgos (analizados en valor absoluto) es que las diferencias entre los métodos dependen en mayor medida de los valores de γ^* (salvo en el caso de las pendientes en que los sesgos son siempre similares). En los parámetros β_0 y γ , así como en la eficiencia, el sesgo del estimador MV tiende a ser menor cuando $\gamma^* \leq 0.4$, dándose la desigualdad contraria a partir de $\gamma^* = 0.6$. Sin embargo, el sesgo MV del parámetro σ^2 es en general menor que el sesgo bayesiano para $\gamma^* \leq 0.8$. Por tanto, con respecto al criterio del sesgo en valor absoluto, no podríamos concluir que un método es claramente superior al otro.

Con respecto a los signos de los sesgos, lo más destacable es que en general siguen un mismo patrón en ambos métodos, si bien el cambio de signo no se da para los mismos valores de γ^* . Por ejemplo, los sesgos del parámetro γ comienzan siendo positivos con ambos métodos para los primeros valores de γ^* y terminan siendo negativos; sin embargo, el cambio de signo en el estimador MV se da en torno a $\gamma^* = 0.2$ mientras que en el caso bayesiano aparece en torno a $\gamma^* = 0.6$ (el cambio de signo también depende del tamaño muestral). Resaltar también que los signos del sesgo de la ordenada en el origen y del parámetro γ son casi siempre coincidentes y a su vez distintos del signo del sesgo en las eficiencias (es decir, cuando subestimamos la ordenada en el origen, también subestimamos la proporción de varianza debida a ineficiencia y como consecuencia sobrestimamos las eficiencias individuales).

6. CONCLUSIONES.

La principal conclusión es que la estimación bayesiana en general es preferible a la estimación MV en el modelo de producción con frontera estocástica analizado. Esta conclusión se ha obtenido en base a que el ECM del parámetro γ y de las eficiencias estimadas es inferior en la inferencia bayesiana salvo en el caso extremo en que la muestra provenga de un modelo en el que todo el error se debe a efectos aleatorios (es decir, en el caso de ausencia de ineficiencia).

Queremos destacar también que la estimación bayesiana se ha llevado a cabo usando el software bayesiano de propósito general WinBUGS, usado desde el entorno R, lo que hace que su uso sea viable y atractivo para investigadores aplicados con unos conocimientos básicos de estadística bayesiana.

Los resultados de la simulación se han obtenido usando la misma distribución a priori en todos los casos; para los parámetros $\bar{\beta}$ y σ_v^{-2} se han usado distribuciones no informativas y para el parámetro σ_u^{-2} se ha elegido una distribución de referencia que supone una eficiencia mediana a priori igual a 0.44. En las aplicaciones concretas, puede incorporarse el conocimiento previo del sector por parte del investigador, sobre todo cambiando la distribución a priori de σ_u^{-2} para ajustarla a la creencia acerca de la eficiencia mediana. Esta información adicional, en principio, conseguiría mejorar aún más la precisión de la estimación bayesiana; si el investigador carece de esta información, hemos comprobado que con la distribución utilizada la metodología bayesiana presenta un mejor comportamiento que la MV en la práctica totalidad del espacio paramétrico.

Para finalizar, resaltemos que nuestro análisis se ha centrado en un modelo con datos de sección cruzada y en el que se ha supuesto que las perturbaciones u_i siguen un modelo Half-Normal. Ampliaciones naturales de este trabajo consistirían en suponer otro tipo de modelos para las perturbaciones u_i y en la consideración de datos de panel.

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. (1995), "Estimators and hypothesis test for a stochastic frontier function: A Monte Carlo analysis", *The Journal of Productivity Analysis*, 6, pp. 247-268.
- 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>
- FARRELL, M.J. (1957), "The measurement of productive efficiency", *Journal of the Royal Statistical Society (A)*, 120, pp. 253-281.
- DORFMAN, J.H.; KOOP, G. (2005), "Current developments in productivity and efficiency measurement", *Special issue of Journal of Econometrics*, 126, pp. 233-570.
- GELFAND, A.E.; SMITH, A.F. (1990): "Sampling-based approaches to calculating marginal densities", *Journal of the American Statistical Association*, 85, pp. 398-409
- GREEN, W.H. (1980): "Maximun Likelihood Estimations of econometric frontier functions", *Journal of Econometrics*, 13, pp. 27-56.
- GRIFFIN, J.E.; STEEL, M.F.J. (2007), "Bayesian stochastic frontier analysis using WinBUGS", *Journal of Productivity Analysis*, 27, pp. 163-176.
- KOOP, G.; STEEL, F.J. (2003), "Bayesian analysis of stochastic frontier models". En: Baltagi, B.H. (editor), *A companion to theoretical econometrics*, Blackwell.
- KOOP, G.; STEEL, M.F.J.; OSIEWALSKI, J. (1995), "Posterior analysis of stochastic frontier models using Gibbs sampling", *Computational Statistics*, 10, pp. 353-373.
- 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.
- OSIEWALSKI, J.; STEEL, M.F.J. (1998), "Numerical tools for the Bayesian analysis of stochastic frontier models", *Journal of Productivity Analysis*, 10, pp. 103-117.
- STURZ, S.; LIGGES, U.; GELMAN, A. (2005), "R2WinBUGS: A package for running WinBUGS from R", *Journal os Statistical Software*, 12 (3), pp. 1-16.
- VAN DEN BROEK, J.; KOOP, G., OSIEWALSKI, J.; STEEL, M.F.J., (1994), "Stochastic frontier models. A Bayesian perspective", *Journal of Econometrics*, 61, pp. 273-303.
- ZHANG, X. (1999), "A Monte Carlo study on the finite sample properties of the Gibbs sampling method for a stochastic frontier model", *Journal of Productivity Analysis*, 14, pp. 71-83.

ANEXO

En este anexo ofrecemos los resultados completos del experimento Monte Carlo realizado.

$\gamma^*=0.0;$ $\gamma=0.00$		SES(β_0)	SES(β_1)	SES(β_2)	SES(β_{11})	SES(β_{22})	SES(β_{12})	SES(σ^2)	SES(γ)	SES(EF)	ECM(β_0)	ECM(β_1)	ECM(β_2)	ECM(β_{11})	ECM(β_{22})	ECM(β_{12})	ECM(σ^2)	ECM(γ)	ECM(EF)
n=30	MV	0,292981	0,059984	0,044901	-0,012792	-0,009473	0,000814	0,158407	0,461395	-0,255017	3,186585	0,222726	0,243231	0,006990	0,007709	0,001877	0,543015	0,440258	0,163353
	BAYES	0,953158	0,027523	0,012604	-0,004938	-0,001842	-0,000298	1,434642	0,755246	-0,532831	3,995965	0,222944	0,239360	0,006850	0,007221	0,001778	2,628467	0,583256	0,310340
n=50	MV	0,419476	0,017440	0,002040	-0,003629	-0,001185	0,000242	0,347540	0,349568	-0,248591	1,105840	0,060630	0,062858	0,001964	0,001874	0,000422	0,622887	0,283211	0,144354
	BAYES	0,855816	0,005832	-0,008361	-0,001445	0,000925	0,000235	0,918132	0,610414	-0,480626	1,456223	0,058869	0,058255	0,001909	0,001736	0,000398	1,020468	0,388439	0,252869
n=100	MV	0,359678	0,003122	0,003816	-0,000187	-0,000525	-0,000379	0,242295	0,277794	-0,226637	0,585718	0,023484	0,025033	0,000737	0,000778	0,000166	0,287349	0,183349	0,112347
	BAYES	0,728309	0,000939	0,001764	0,000172	-0,000114	-0,000370	0,645274	0,530892	-0,444986	0,845978	0,023461	0,024647	0,000734	0,000770	0,000165	0,499189	0,292287	0,215012
n=500	MV	0,286966	0,003203	0,002033	-0,000229	0,000059	-0,000437	0,153313	0,189908	-0,207213	0,215994	0,004664	0,004631	0,000154	0,000147	0,000031	0,072119	0,081584	0,078480
	BAYES	0,598194	0,003309	0,002030	-0,000243	0,000066	-0,000447	0,407207	0,422561	-0,400109	0,416622	0,004669	0,004650	0,000154	0,000148	0,000032	0,180045	0,183026	0,171161

$\gamma^*=0.2;$ $\gamma=0.41$		SES(β_0)	SES(β_1)	SES(β_2)	SES(β_{11})	SES(β_{22})	SES(β_{12})	SES(σ^2)	SES(γ)	SES(EF)	ECM(β_0)	ECM(β_1)	ECM(β_2)	ECM(β_{11})	ECM(β_{22})	ECM(β_{12})	ECM(σ^2)	ECM(γ)	ECM(EF)
n=30	MV	-0,205141	0,047870	0,029234	-0,010902	-0,006891	0,001527	-0,081687	0,070498	0,107161	2,613634	0,186968	0,179853	0,005696	0,005634	0,001325	0,326203	0,235323	0,128063
	BAYES	0,407539	0,019935	-0,001619	-0,003954	0,000131	0,000518	0,996573	0,383548	-0,159510	2,610703	0,180645	0,173052	0,005341	0,005197	0,001264	1,343828	0,156658	0,073783
n=50	MV	-0,112260	0,022973	0,006135	-0,004634	-0,000635	-0,000078	0,056397	-0,002435	0,092780	0,687979	0,045801	0,047652	0,001453	0,001535	0,000295	0,313708	0,163446	0,109863
	BAYES	0,273188	0,011739	-0,004625	-0,002510	0,001486	-0,000057	0,536415	0,264831	-0,108433	0,571311	0,041807	0,043592	0,001323	0,001428	0,000270	0,410753	0,083430	0,055306
n=100	MV	-0,074322	0,007810	0,001351	-0,000722	-0,000037	-0,000526	0,040699	-0,009648	0,070635	0,338623	0,018582	0,018838	0,000611	0,000603	0,000115	0,153796	0,114442	0,088443
	BAYES	0,188026	0,004761	-0,002155	-0,000137	0,000648	-0,000522	0,336260	0,197085	-0,079032	0,260646	0,018300	0,018685	0,000607	0,000597	0,000114	0,169122	0,048898	0,046665
n=500	MV	-0,040297	-0,004740	-0,003814	0,000896	0,000803	-0,000062	-0,009321	-0,040481	0,048120	0,096553	0,003334	0,003456	0,000108	0,000111	0,000021	0,037783	0,051399	0,059578
	BAYES	0,108570	-0,005098	-0,004082	0,000974	0,000851	-0,000068	0,131934	0,098263	-0,038280	0,056829	0,003335	0,003451	0,000108	0,000111	0,000021	0,029054	0,015013	0,038794

$\gamma^*=0.4; \gamma=0.65$		SES(β_0)	SES(β_1)	SES(β_2)	SES(β_{11})	SES(β_{22})	SES(β_{12})	SES(σ^2)	SES(γ)	SES(EF)	ECM(β_0)	ECM(β_1)	ECM(β_2)	ECM(β_{11})	ECM(β_{22})	ECM(β_{12})	ECM(σ^2)	ECM(γ)	ECM(EF)
n=30	MV	-0,296317	0,015899	0,025777	-0,002802	-0,005302	0,000114	-0,261735	-0,122898	0,167131	1,921227	0,131787	0,138386	0,004132	0,004095	0,000993	0,293275	0,245348	0,135906
	BAYES	0,259766	-0,015210	-0,005025	0,004105	0,001980	-0,000747	0,617078	0,184782	-0,075899	1,702375	0,126990	0,125596	0,003892	0,003666	0,000889	0,605621	0,040363	0,050394
n=50	MV	-0,123924	0,015852	0,011742	-0,003655	-0,003215	0,000666	-0,013584	-0,083654	0,085180	0,545084	0,034441	0,036085	0,001081	0,001183	0,000240	0,238733	0,168120	0,099045
	BAYES	0,149286	0,001986	-0,002595	-0,000525	0,000075	0,000291	0,334594	0,101377	-0,047752	0,399332	0,030049	0,031654	0,000955	0,001035	0,000212	0,212832	0,022654	0,042809
n=100	MV	-0,093917	0,007080	0,001074	-0,001326	-0,000266	-0,000137	-0,040318	-0,093968	0,067361	0,264300	0,014956	0,013806	0,000454	0,000427	0,000098	0,120489	0,113521	0,073109
	BAYES	0,067258	0,002566	-0,002372	-0,000453	0,000428	-0,000116	0,159072	0,036019	-0,018674	0,180524	0,014519	0,013578	0,000442	0,000419	0,000095	0,071902	0,011147	0,037726
n=500	MV	-0,029135	0,001043	-0,000137	-0,000343	0,000104	0,000041	-0,023722	-0,034792	0,017766	0,048747	0,002664	0,002729	0,000088	0,000086	0,000016	0,026731	0,026578	0,040771
	BAYES	0,000127	0,000478	-0,000632	-0,000224	0,000209	0,000034	0,020368	-0,012428	0,003276	0,035117	0,002646	0,002727	0,000087	0,000086	0,000016	0,014659	0,007086	0,035211

$\gamma^*=0.6; \gamma=0.80$		SES(β_0)	SES(β_1)	SES(β_2)	SES(β_{11})	SES(β_{22})	SES(β_{12})	SES(σ^2)	SES(γ)	SES(EF)	ECM(β_0)	ECM(β_1)	ECM(β_2)	ECM(β_{11})	ECM(β_{22})	ECM(β_{12})	ECM(σ^2)	ECM(γ)	ECM(EF)
n=30	MV	-0,406530	0,040448	0,042413	-0,008383	-0,008781	0,000153	-0,332367	-0,195654	0,171946	1,572095	0,119837	0,116959	0,003881	0,003565	0,000841	0,270285	0,252658	0,121883
	BAYES	0,096778	0,002886	0,011112	0,000326	-0,001355	-0,000986	0,423523	0,064408	-0,037487	1,250039	0,104796	0,103200	0,003294	0,003143	0,000755	0,358941	0,008406	0,039386
n=50	MV	-0,158005	0,020675	0,021415	-0,003851	-0,004550	0,000061	-0,054901	-0,098337	0,066039	0,446336	0,030633	0,031338	0,000974	0,001027	0,000198	0,194641	0,138998	0,075220
	BAYES	0,060390	-0,002060	0,002571	0,000515	-0,000801	0,000144	0,198181	0,015614	-0,018493	0,324385	0,026016	0,026711	0,000828	0,000879	0,000175	0,126998	0,009838	0,033630
n=100	MV	-0,077149	0,013707	0,004960	-0,002542	-0,000949	-0,000192	-0,039566	-0,054263	0,030929	0,188944	0,011890	0,011637	0,000368	0,000368	0,000081	0,091681	0,062114	0,045155
	BAYES	-0,002668	0,007534	-0,001025	-0,001353	0,000226	-0,000162	0,069607	-0,022886	0,000967	0,144386	0,011284	0,011215	0,000347	0,000356	0,000078	0,049484	0,009457	0,030274
n=500	MV	-0,003788	-0,001408	-0,000836	0,000269	0,000090	0,000012	-0,013477	-0,006530	0,004509	0,026964	0,001947	0,002084	0,000061	0,000064	0,000014	0,013552	0,004047	0,027567
	BAYES	-0,003739	-0,002279	-0,001787	0,000443	0,000278	0,000008	-0,002812	-0,019664	0,006163	0,026711	0,001945	0,002087	0,000061	0,000064	0,000014	0,012046	0,004031	0,027502

$\gamma^*=0.8; \gamma=0.92$		SES(β_0)	SES(β_1)	SES(β_2)	SES(β_{11})	SES(β_{22})	SES(β_{12})	SES(σ^2)	SES(γ)	SES(EF)	ECM(β_0)	ECM(β_1)	ECM(β_2)	ECM(β_{11})	ECM(β_{22})	ECM(β_{12})	ECM(σ^2)	ECM(γ)	ECM(EF)
n=30	MV	-0,370849	0,025926	0,032656	-0,005991	-0,007626	0,000623	-0,394851	-0,203892	0,162197	1,342800	0,087296	0,086208	0,002652	0,002753	0,000690	0,280093	0,223310	0,098089
	BAYES	0,039394	-0,007040	0,000734	0,001778	0,000236	-0,000588	0,223798	-0,021288	-0,002793	1,028130	0,077542	0,073083	0,002285	0,002315	0,000601	0,200523	0,002779	0,029784

n=50	MV	-0,133682	0,025871	0,022613	-0,004729	-0,004661	-0,000126	-0,067987	-0,031846	0,031727	0,341391	0,024320	0,024120	0,000767	0,000758	0,000163	0,112773	0,056865	0,039908
	BAYES	0,001152	0,001478	-0,000180	0,000584	0,000344	-0,000491	0,098401	-0,020855	-0,002303	0,246242	0,019054	0,020075	0,000607	0,000622	0,000133	0,078132	0,005107	0,022215
n=100	MV	-0,033197	0,007624	0,012694	-0,001881	-0,002915	0,000424	-0,000738	-0,002367	0,003800	0,117004	0,009171	0,009239	0,000292	0,000299	0,000056	0,059564	0,014491	0,023154
	BAYES	0,009466	-0,003644	0,000928	0,000379	-0,000589	0,000412	0,045143	-0,023043	0,000827	0,098918	0,007881	0,007977	0,000247	0,000259	0,000050	0,043031	0,005308	0,019307
n=500	MV	-0,007361	0,000425	0,001318	-0,000169	-0,000404	0,000192	-0,005050	0,000556	0,001414	0,016656	0,001351	0,001467	0,000045	0,000046	0,000009	0,008003	0,000573	0,015590
	BAYES	-0,003464	-0,001177	-0,000216	0,000151	-0,000098	0,000186	0,002255	-0,006967	0,001915	0,016525	0,001347	0,001455	0,000044	0,000046	0,000009	0,007876	0,000696	0,015604

$\gamma^*=1.0;$	$\gamma=1.00$	SES(β_0)	SES(β_1)	SES(β_2)	SES(β_{11})	SES(β_{22})	SES(β_{12})	SES(σ^2)	SES(γ)	SES(EF)	ECM(β_0)	ECM(β_1)	ECM(β_2)	ECM(β_{11})	ECM(β_{22})	ECM(β_{12})	ECM(σ^2)	ECM(γ)	ECM(EF)
n=30	MV	-0,352162	0,022095	0,029638	-0,007790	-0,009545	0,002746	-0,439738	-0,173517	0,159084	1,089515	0,069940	0,068982	0,002043	0,002036	0,000515	0,293667	0,160085	0,072449
	BAYES	-0,012383	-0,011468	-0,003302	0,001206	-0,000311	0,000557	0,080650	-0,076377	0,027827	0,779390	0,056321	0,058063	0,001645	0,001694	0,000451	0,142313	0,006934	0,020834
n=50	MV	-0,188920	0,025217	0,027608	-0,004820	-0,005438	-0,000220	-0,202837	-0,017888	0,048819	0,199848	0,013405	0,012555	0,000392	0,000400	0,000080	0,082745	0,009875	0,011742
	BAYES	-0,047146	0,000489	0,000572	0,000425	0,000173	-0,000323	-0,018798	-0,030783	0,017715	0,129475	0,010342	0,009721	0,000301	0,000309	0,000069	0,045643	0,001647	0,006640
n=100	MV	-0,101495	0,014490	0,013939	-0,002841	-0,002740	-0,000015	-0,105766	-0,003551	0,025389	0,059141	0,003915	0,004051	0,000114	0,000125	0,000025	0,036920	0,000170	0,003333
	BAYES	-0,027750	-0,000181	-0,000734	0,000070	0,000234	-0,000015	-0,017274	-0,011153	0,012422	0,037692	0,002816	0,002979	0,000086	0,000090	0,000020	0,025492	0,000300	0,002430
n=500	MV	-0,022950	0,000359	0,001673	-0,000288	-0,000503	0,000265	-0,041457	-0,002768	0,009280	0,005877	0,000453	0,000478	0,000015	0,000015	0,000005	0,009545	0,000035	0,000585
	BAYES	-0,009215	-0,000305	-0,000111	0,000028	0,000008	-0,000011	-0,012064	-0,001150	0,005229	0,002037	0,000164	0,000169	0,000005	0,000005	0,000001	0,004280	0,000002	0,000183