

¿INEFICIENCIA O DIFERENCIAS TECNOLÓGICAS EN EL SECTOR LECHERO?*

ANTONIO ÁLVAREZ PINILLA
JULIO DEL CORRAL CUERVO
Universidad de Oviedo

La estimación de funciones de producción suele basarse en el supuesto de que la tecnología es común para todos los productores. Sin embargo, en la muestra puede haber empresas que usen tecnologías diferentes. En este trabajo se emplea un modelo de clases latentes para estimar la tecnología de varios sistemas de producción de explotaciones lecheras según su grado de intensificación. Los resultados de esta estimación se comparan con los obtenidos usando otras dos aproximaciones. Una, el modelo clásico de frontera estocástica. La otra consiste en dividir la muestra mediante un análisis *cluster* y estimar una frontera de producción distinta para cada uno de los grupos. El análisis empírico se realiza con un panel completo compuesto por 169 explotaciones lecheras asturianas para el período comprendido entre 1999 y 2004.

Palabras clave: Explotaciones lecheras, función de producción, frontera estocástica, heterogeneidad inobservable, modelo de clases latentes.

Clasificación JEL: Q12, C23.

La estimación de funciones de producción, costes o beneficios suele basarse en el supuesto de que la tecnología subyacente es común para todos los productores. Sin embargo, en la muestra puede haber empresas que usen tecnologías diferentes. En este caso, la estimación de una función común para todas las observaciones no es apropiada ya que se pueden obtener estimaciones sesgadas de las características de la tecnología. Para evitar este problema de especificación, a veces se realiza un proceso en dos etapas. En la primera, las observaciones de la muestra se clasifican en varios grupos. Esta separación se puede hacer de dos formas: a partir de información *a priori* sobre las empresas, como la localización [véase, por ejemplo, Hoch (1962)] o mediante un análisis *cluster* [véase, por ejemplo, Maudos *et al.* (2002), Álvarez *et al.* (2007)]. En la se-

(*) Los autores agradecen las sugerencias de Carlos Arias, Ángel Gavilán, Manuel Fernández Bagüés, Luis Orea, Juan Prieto y David Roibás, así como los comentarios de dos revisores anónimos. El trabajo ha contado con la financiación del proyecto MEC-SEJ-2004-07997 y del proyecto ALFA – NERUDA de la Unión Europea. El segundo autor agradece el apoyo financiero del Ministerio de Educación y Ciencia (programa FPU).

gunda etapa se estima una función distinta para cada grupo. Una alternativa es el uso de modelos que permiten, en una sola etapa, determinar los grupos y estimar una función distinta para cada grupo. A esta categoría pertenecen los denominados modelos de clases latentes¹.

En este trabajo se investiga la existencia de diferentes tecnologías en el sector lechero asturiano. En las últimas décadas se ha producido una importante reducción en el número de explotaciones, lo que ha llevado, dado que la cuota regional ha permanecido más o menos constante, a que las explotaciones que han continuado en el sector hayan aumentado de tamaño. Paralelamente a este proceso de crecimiento está teniendo lugar una transformación en el sistema de producción empleado por las explotaciones, observándose una tendencia hacia sistemas de producción más intensivos. En general, se considera que un sistema es más intensivo cuando el ganado se alimenta en mayor medida con pienso comprado y en menor medida con forraje producido en la explotación, lo que se traduce en mayores ratios de vacas por hectárea y kilogramos de pienso por vaca. Sin embargo, las diferencias entre explotaciones intensivas y extensivas no han sido estudiadas adecuadamente dado que el concepto de intensificación no viene definido por una sola variable, por lo que no resulta fácil asignar las explotaciones a un sistema de producción determinado.

Una posible consecuencia de este proceso de intensificación es que aumente el coste medio de producción debido a que el sistema de alimentación de las ganaderías intensivas se basa en un mayor uso de piensos y en un menor uso de forrajes producidos en la propia explotación. Sin embargo, un aspecto positivo del sistema de producción intensivo es que es más sencillo de gestionar, por lo que es posible que los ganaderos que lo usan sean técnicamente más eficientes que los extensivos. Efectivamente, los ganaderos que basan la alimentación del ganado en los forrajes producidos en la explotación tienen que realizar muchas más tareas (sembrar forrajes, cosechar, ensilar, combinar los alimentos, etc.), por lo que la probabilidad de tomar una decisión equivocada es mayor. Por tanto, aunque la eficiencia de las explotaciones se ha analizado en varios trabajos [véase, por ejemplo, Álvarez y Arias (2003 y 2004)] es interesante conocer si hay diferencias en la eficiencia técnica de las explotaciones extensivas e intensivas.

En este trabajo se emplea un modelo de clases latentes para estimar la tecnología de varios sistemas de producción según su grado de intensificación. En concreto, se identifican las tecnologías de tres sistemas de producción que hemos denominado como extensivo, semi-intensivo e intensivo. Como existe interés en calcular la eficiencia de los distintos sistemas, el modelo de clases latentes se ha aplicado en el contexto de una frontera estocástica² [Greene (2005)]. Los resulta-

(1) En inglés, *Latent-class models* o *Finite mixture models*. Véase Cameron y Trivedi (2005, pág. 621) para una exposición de estos modelos.

(2) Recientemente han aparecido algunos trabajos que combinan el enfoque de frontera estocástica con la estructura de clases latentes usando datos de panel. Caudill (2003) estima una función de costes estocástica para una muestra de bancos estadounidenses con dos grupos. Orea y Kumbhakar (2004) estiman una frontera de costes estocástica para el sector bancario español con cuatro grupos. El-Gamal e Inanoglu (2005) estiman una función de costes estocástica para bancos turcos con dos grupos. Previamente, Beard *et al.* (1991) combinaron estos dos enfoques usando datos de corte transversal.

dos se comparan con los obtenidos a partir de una frontera estocástica que considera que la tecnología es común para todas las explotaciones. Por otra parte, también se comparan los resultados del modelo de clases latentes con una aproximación habitual para modelizar las diferencias tecnológicas: dividir la muestra mediante un análisis *cluster* y estimar una frontera de producción distinta para cada uno de los grupos. El análisis empírico se realiza con un panel de datos compuesto por 169 explotaciones lecheras asturianas para el período 1999-2004.

Los objetivos del trabajo son los siguientes. En primer lugar, queremos averiguar si los métodos mencionados en el párrafo anterior permiten identificar grupos de explotaciones cuyas características puedan asociarse con sistemas productivos conocidos. En segundo lugar, queremos conocer si las diferencias entre las características tecnológicas estimadas para estos sistemas son importantes. En concreto, estamos interesados en investigar si las economías de escala varían con el nivel de intensificación del sistema productivo. Anteriores trabajos habían encontrado que las deseconomías de tamaño aparecían a partir de tamaños relativamente pequeños [Álvarez y Arias (2003)]. De alguna manera, este resultado no permitía explicar el gran aumento de tamaño de las explotaciones en un período en el que el precio real de la leche ha disminuido ligeramente. Por último, queremos analizar la importancia de suponer la existencia de varias tecnologías de cara a la estimación de la ineficiencia productiva. El problema es que si sólo se estima una tecnología las unidades que no utilicen dicha tecnología tendrán una ineficiencia mayor que si fuesen comparadas con su tecnología. En cambio, si se estiman varias tecnologías cada unidad productiva será comparada con su tecnología. De este modo no se etiqueta como ineficiencia lo que realmente son diferencias en la tecnología empleada por las explotaciones [Stigler (1976)].

El resto del trabajo se estructura de la forma siguiente. La sección 1 aborda la modelización de diferencias tecnológicas entre empresas. En la sección 2 se describe el modelo de clases latentes de frontera estocástica. En la sección 3 se describen los datos y el modelo empírico, mientras que en la sección 4 se presentan y discuten los resultados. En la sección 5 se analiza la eficiencia técnica de las explotaciones. Por último, en la sección 6 se extraen algunas conclusiones.

1. DIFERENCIAS TECNOLÓGICAS ENTRE EMPRESAS

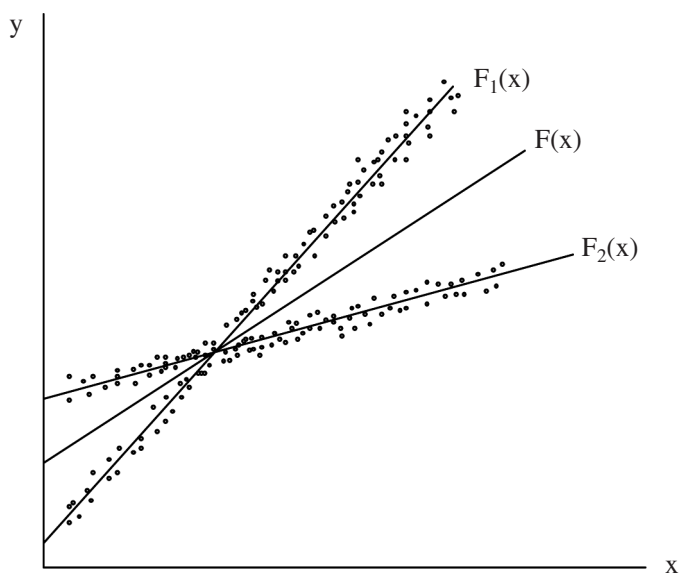
En todos los trabajos empíricos existe siempre alguna diferencia entre las unidades muestrales que no viene reflejada en los datos (la calidad de algunos inputs es un ejemplo típico). Esta información no observada se conoce como heterogeneidad inobservable. Cuando esta información es poco importante para explicar el fenómeno que se analiza, se puede subsumir en la perturbación aleatoria del modelo. La cuestión de interés es cómo modelizar las situaciones en las que estas diferencias son importantes.

Esta heterogeneidad inobservable entre individuos puede ser una fuente importante de problemas econométricos si no se modeliza correctamente, ya que su efecto es equivalente al de una variable omitida. El problema surge cuando las variables relevantes omitidas del análisis están correlacionadas con alguna de las variables explicativas, ya que la estimación será sesgada [Griliches (1957)]. El caso

más estudiado es cuando la información no observada puede considerarse invariante en el tiempo. Por ejemplo, la habilidad en los modelos salariales o la calidad de la tierra en una función de producción agraria. Si se dispone de datos de panel, la solución al problema consiste en modelizar la heterogeneidad como un efecto individual [Mundlak (1961)].

Otro tipo de heterogeneidad inobservable se refiere a la relación económica subyacente; es decir, las empresas de la muestra pueden provenir de distintos procesos generadores de datos. Este caso puede véase gráficamente en el gráfico 1, donde hay dos grupos de empresas que tienen distintas tecnologías de producción. Si se estima una única función de producción (por ejemplo mediante mínimos cuadrados), se obtendrá la función $F(x)$, por lo que las características tecnológicas estimadas (elasticidades de producción, economías de escala, elasticidades de sustitución,...) estarán sesgadas.

Gráfico 1: FUNCIONES DE PRODUCCIÓN CON DIFERENTES TECNOLOGÍAS



Fuente: Elaboración propia.

Para superar este problema es necesario usar modelos que estimen parámetros distintos para cada uno de los grupos. Algunos de estos modelos son continuos en el sentido de que permiten la existencia de una tecnología distinta para cada empresa. Los modelos de parámetros aleatorios [Hildreth y Houck (1968)] y la estimación por máxima verosimilitud local [Kumbhakar *et al.* (2007)] son dos ejemplos. Otros métodos son discretos, ya que forman varios grupos, estimando una referencia tecnológica para cada grupo. A esta categoría pertenecen el análisis *cluster* y los modelos de clases latentes.

El análisis *cluster* permite clasificar la muestra en varios grupos. A continuación se estima una representación de la tecnología para cada grupo. Sin embargo, este procedimiento tiene dos limitaciones. Una está asociada con la primera etapa, ya que como el algoritmo no tiene en cuenta la relación de las variables separadoras con la variable dependiente, las empresas pueden separarse por un criterio distinto al deseado (el de la tecnología). El segundo problema está asociado a la eficiencia en la estimación, ya que este procedimiento no usa la información contenida en una clase para estimar la tecnología de las empresas pertenecientes a otras clases. Sin embargo, esta información puede ser importante porque las tecnologías, aunque sean diferentes, pueden tener rasgos comunes. Por lo tanto, la estimación de distintas referencias tecnológicas usando los grupos obtenidos mediante un algoritmo *cluster* es un proceso estadísticamente ineficiente puesto que no se aprovecha toda la información disponible.

Por otra parte, los modelos de clases latentes suponen que hay un número finito de estructuras subyacentes (clases) en los datos, de forma que cada una de las empresas pertenece a una de las clases aunque la clase a la que pertenece cada empresa es desconocida para el analista [Greene (2004)]. Estos modelos permiten estimar para cada empresa la probabilidad de pertenecer a cada grupo.

2. MODELO DE CLASES LATENTES DE FRONTERA ESTOCÁSTICA

En este trabajo se va a estimar una frontera de producción estocástica usando un modelo de clases latentes. Los modelos de frontera estocástica se empezaron a desarrollar a finales de la década de los setenta [Aigner, Lovell y Schmidt (1977), Meeusen y van den Broeck (1977)]³. Una función de producción frontera estocástica puede escribirse como:

$$\ln y = f(x) + \varepsilon ; \quad \varepsilon = v - u \quad [1]$$

donde y representa el output de cada empresa, x es un vector de inputs, $f(x)$ es la representación de la tecnología, y ε es una perturbación aleatoria compuesta por dos términos de error: uno simétrico, v , que refleja la aleatoriedad del proceso productivo y que se supone que se distribuye como una normal con esperanza cero y desviación típica σ_v , y otro asimétrico, u , que recoge la ineficiencia de las empresas, y que se suele modelizar mediante una distribución seminormal⁴, por lo que $u \geq 0$. Además, se supone que los componentes u y v son independientes. Para estimar estos modelos hay que utilizar determinados supuestos que no están exentos de problemas, como la elección de la forma funcional y las distribuciones de las perturbaciones aleatorias⁵.

(3) Una revisión de esta literatura puede véase en Álvarez (2001).

(4) Otras distribuciones utilizadas en la literatura son la exponencial, la normal truncada y la gamma.

(5) Alternativamente, existen métodos no paramétricos que no necesitan ninguno de estos supuestos para ser estimados [véase Charnes *et al.* (1995)]. Sin embargo, tampoco están exentos de limitaciones puesto que atribuyen todo el error a la ineficiencia de las unidades productivas y se ven más afectados por valores atípicos.

La frontera estocástica de la ecuación [1] se puede expresar como un modelo de clases latentes de la siguiente forma [Greene (2001)]:

$$\ln y_{it} |_j = f(x_{it}) |_j + v_{it} |_j - u_{it} |_j \quad [2]$$

donde el subíndice i indica empresas, t denota tiempo y j representa las distintas clases (grupos). La barra vertical indica que existe un modelo distinto para cada clase j .

Modelizando los términos de ineficiencia con distribuciones seminormales, la función de verosimilitud (LF) de la empresa i en el momento t para el grupo j es:

$$LF_{ijt} = f(y_{it} | x_{it}, \beta_j, \sigma_j, \lambda_j) = \frac{\Phi(\lambda_j \cdot \varepsilon_{it|j} / \sigma_j)}{\Phi(0)} \cdot \frac{1}{\sigma_j} \cdot \phi\left(\frac{\varepsilon_{it|j}}{\sigma_j}\right) \quad [3]$$

donde $\varepsilon_{it|j} = y_{it|j} - \beta'_j x_{it}$, $\sigma_j = [\sigma_{uj}^2 + \sigma_{vj}^2]^{1/2}$, $\lambda_j = \sigma_{uj} / \sigma_{vj}$, Φ y ϕ son respectivamente la función de distribución y la función de densidad de una normal estándar.

La función de verosimilitud de la empresa i en el grupo j se obtiene multiplicando las funciones de verosimilitud de cada uno de los períodos considerados:

$$LF_{ij} = \prod_{t=1}^T LF_{ijt} \quad [4]$$

La función de verosimilitud de la empresa i se obtiene como una media ponderada de su función de verosimilitud en cada uno de los grupos j , siendo los pesos las probabilidades *a priori* de pertenecer a cada grupo:

$$LF_i = \sum_{j=1}^J LF_{ij} P_{ij} \quad [5]$$

La probabilidad *a priori* de pertenecer a un grupo tiene que estar comprendida entre cero y uno: $0 \leq P_{ij} \leq 1$. Asimismo, la suma de estas probabilidades para cada empresa debe ser uno: $\sum_j P_{ij} = 1$. La forma habitual de modelizar estas probabilidades es la especificación logit multinomial puesto que asegura que se verifiquen las dos condiciones. La expresión correspondiente es:

$$P_{ij} = \frac{\exp(\delta_j q_i)}{\sum_{j=1}^J \exp(\delta_j q_i)} \quad [6]$$

donde q_i es un vector de variables separadoras y δ_j son parámetros a estimar. Un supuesto del modelo es que las probabilidades *a priori* de pertenecer a un grupo son constantes en el tiempo, por lo que se emplean variables separadoras invariantes en el tiempo. En concreto, se han utilizado los valores del primer año de la muestra⁶.

(6) Los resultados son muy similares si se usa la media del período.

El logaritmo de la función de verosimilitud conjunta se obtiene como el sumatorio de los logaritmos de las funciones de verosimilitud individuales:

$$\log LF = \sum_{i=1}^N \log LF_i \quad [7]$$

Por tanto, la función de verosimilitud del modelo se puede escribir como:

$$\begin{aligned} \log LF &= \sum_{i=1}^N \log \sum_{j=1}^J P_{ij} \prod_{t=1}^T LF_{ijt} = \\ &= \sum_{i=1}^N \log \sum_{j=1}^J \frac{\exp(\delta_j \cdot q_i)}{\sum_{j=1}^J \exp(\delta_j \cdot q_i)} \prod_{t=1}^T \frac{\Phi(\lambda_j \cdot \varepsilon_{itj} / \sigma_j)}{\Phi(0)} \cdot \frac{1}{\sigma_j} \cdot \phi\left(\frac{\varepsilon_{itj}}{\sigma_j}\right) \end{aligned} \quad [8]$$

El logaritmo de la función de verosimilitud puede maximizarse respecto al conjunto de parámetros $\theta_j = (\beta_j, \sigma_j, \lambda_j, \delta_j)$ usando métodos convencionales [Greene (2002)]. Además, los parámetros estimados pueden utilizarse para calcular mediante el Teorema de Bayes la probabilidad *a posteriori* de pertenecer a cada uno de los grupos:

$$P(j/i) = \frac{P_{ij} LF_{ij}}{\sum_{j=1}^J P_{ij} LF_{ij}} \quad [9]$$

Es decir, la probabilidad *a posteriori* de pertenecer a un grupo aumenta cuanto mayor sea la probabilidad *a priori* (variables separadoras) y cuanto mayor sea la función de verosimilitud de la explotación *i* en el grupo *j* (bondad de ajuste).

Una dificultad de los modelos de clases latentes es que el investigador debe seleccionar el número de clases. El procedimiento habitual es estimar el modelo con distinto número de clases y usar algún contraste estadístico para seleccionar el modelo preferido (lo normal suele estar entre 2 y 4 clases). Greene (2005) propuso el *testing down*, en el que comenzando por un J^* al menos tan grande como el verdadero J se puede contrastar hacia abajo, dado que un modelo de $J-1$ clases está anidado en otro de J clases, basándose en el contraste de la ratio de verosimilitud⁷. Otra alternativa es utilizar criterios de información, como el Criterio de Información de Akaike (AIC) o el Criterio de Información Bayesiana de Schwarz (SBIC):

$$SBIC = -2 \cdot \log LF(J) + m \cdot \log(n) \quad [10]$$

$$AIC = -2 \cdot \log LF(J) + 2 \cdot m \quad [11]$$

(7) El estadístico se construye como $-2 \cdot (\log LF_r - \log LF_u)$, donde LF_r es el valor de la función de verosimilitud del modelo restringido y LF_u es el valor de la función de verosimilitud del modelo libre. El estadístico sigue una distribución chi-cuadrado de tantos grados de libertad como restricciones se impongan.

donde $LF(J)$ es el valor de la función de verosimilitud para J grupos, m es el número de parámetros del modelo y n el número de observaciones ($n = N \cdot T$, siendo N el número de empresas y T el número de años). El modelo preferido será el que tenga un menor valor de estos estadísticos.

3. DATOS Y MODELO EMPÍRICO

Los datos empleados en el análisis empírico son un panel completo compuesto por 169 explotaciones lecheras asturianas. Los datos corresponden al período 1999-2004. En el cuadro 1 se presentan las estadísticas descriptivas de las principales variables usadas en el análisis.

Cuadro 1: ESTADÍSTICAS DESCRIPTIVAS DE LOS DATOS

	Media	Coefficiente de variación	Mínimo	Máximo
Leche (litros)	300.835	0,49	90.484	908.612
Pienso (kg.)	144.582	0,53	29.777	458.475
Vacas (unidades)	40	0,36	20	100
Trabajo (ud. trabajo anual)	1,96	0,30	0,71	4
Gasto en forraje (euros)	21.737	0,64	3.658	106.526
Leche / Vacas	7.245	0,18	3.586	10.973
Leche / Tierra (l./ha.)	16.289	0,37	4.292	43.216
Pienso / Vacas	3.463	0,25	1.240	6.182
Vacas / Tierra	2,22	0,28	0,90	4,43

Fuente: Elaboración propia.

En general, las explotaciones de la muestra son relativamente grandes comparadas con la explotación media asturiana⁸. Las últimas cuatro filas hacen referencia a variables que recogen de alguna manera la mayor o menor intensificación del sistema productivo, de las cuales las dos últimas son las variables separadoras del modelo de clases latentes y las variables que se emplean en el algoritmo *cluster*. Las medias de estas cuatro variables indican que la intensificación es bastante elevada, aunque se observan grandes diferencias entre los mínimos y los máximos.

La forma funcional empleada es la translogarítmica. La variable dependiente es la cantidad de leche producida (litros)⁹. Los datos originales de los inputs se

(8) La cuota media de leche por explotación en el año 2004 fue de 127.705 kilogramos (SADEI, 2005).

(9) Sólo se ha considerado un output dado que estas explotaciones están altamente especializadas (más del 90% de sus ingresos provienen de las ventas de leche).

han dividido por su media geométrica, por lo que los coeficientes de primer orden se pueden interpretar como las elasticidades de producción evaluadas en la media geométrica de la muestra. Se han considerado cuatro inputs: número de vacas (media anual de vacas en producción), trabajo (unidades de trabajo anual), alimentación externa (kilogramos de pienso consumidos anualmente) y gasto en forraje (incluye los gastos realizados en la producción de forraje: semillas, tratamientos, abono, carburante, alquiler de maquinaria y materiales diversos, así como la amortización de la maquinaria y una imputación del coste de oportunidad de la tierra). Adicionalmente, se introducen efectos temporales, para controlar los factores que afectan del mismo modo a todas las explotaciones pero varían a lo largo del tiempo (la categoría excluida es el año 1999).

Por tanto, la ecuación a estimar es:

$$\ln y_{it} | _j = \beta_0 | _j + \sum_{l=1}^L \beta_l | _j \ln x_{it} + \frac{1}{2} \sum_{l=1}^L \sum_{k=1}^L \beta_{lk} | _j \ln x_{it} \ln x_{ikt} + \sum_{t=2000}^{t=2004} \beta_t D_t | _j + v_{it} | _j - u_{it} | _j \quad [12]$$

donde se ha supuesto que v sigue una distribución normal centrada en cero y u una distribución semi-normal, es decir, $v_{it} | _j \sim N(0, \sigma_{vj}^2)$ y $u_{it} | _j \sim |N(0, \sigma_{uj}^2)|$.

4. ESTIMACIÓN Y RESULTADOS

En esta sección se estima el modelo de clases latentes de la ecuación [12] y se compara con otros dos modelos: una frontera estocástica, que hemos denominado frontera *pooled* puesto que agrupa todas las observaciones sin considerar la existencia de heterogeneidad, y un modelo donde se separa la muestra en tres grupos mediante un análisis *cluster*, estimando posteriormente una frontera estocástica para cada grupo.

El modelo de clases latentes se estimó por máxima verosimilitud¹⁰. El modelo con tres grupos fue el preferido según el Criterio de Información de Akaike y el *testing down* propuesto por Greene (2005)¹¹ por lo que fue el modelo elegido. Las variables “separadoras” para identificar los grupos, que se miden en ratios para evitar que las explotaciones se agrupen principalmente en función del tamaño, son los kilogramos de pienso por vaca y el número de vacas por hectárea¹². Los tres grupos obtenidos se corresponden con tres sistemas de producción diferentes que se han etiquetado como extensivo, semi-intensivo e intensivo en función de las siguientes ratios: litros de leche por vaca, litros de leche por hectárea, vacas por hectárea y pienso por vaca.

Para permitir la comparación, en el análisis *cluster* se dividió la muestra también en tres grupos usando las mismas variables separadoras que en el modelo de clases latentes. El análisis se realizó usando la media de estas variables durante

(10) Las estimaciones de todos los modelos se han realizado utilizando el programa LIMDEP 8.0.

(11) Sin embargo, el modelo preferido según el SBIC fue el de dos grupos.

(12) No se consideran las ratios litros de leche por vaca y litros de leche por hectárea para evitar posibles problemas de endogeneidad.

los seis años¹³. De este modo, cada explotación está asignada al mismo grupo en todos los periodos, por lo que los grupos están formados por las mismas explotaciones durante todos los años. Los grupos obtenidos se han denominado de la misma forma que los del modelo de clases latentes (extensivo, semi-intensivo e intensivo) en función de las mismas variables.

En el cuadro 2 se muestran las estadísticas descriptivas de los grupos formados por ambos métodos¹⁴. En general, las características de los grupos son similares en ambos métodos: las explotaciones más intensivas producen más leche, usan más inputs, tienen vacas de mayor producción, consumen más pienso por vaca, tienen más vacas por unidad de superficie y, como consecuencia, producen más litros de leche por hectárea.

Una cuestión interesante es si el modelo de clases latentes y el algoritmo *cluster* realizan la misma clasificación de las explotaciones. Para ello, en el cuadro 3 se compara la clasificación de ambos métodos. Las celdas en la diagonal principal del cuadro indican el número de explotaciones que se clasifican en el mismo sistema por los dos métodos. Como se puede ver ambos métodos clasifican a las explotaciones de forma diferente. La mayor coincidencia se produce en el grupo extensivo (el 68% de las explotaciones extensivas según el modelo de clases latentes también son consideradas como extensivas en el *cluster*). Por lo tanto, a pesar que en los dos métodos se han utilizado las mismas variables separadoras, la clasificación difiere bastante.

A continuación se procede a comparar los resultados de la estimación de los tres modelos. La estimación del modelo de clases latentes se puede véase en el cuadro A1 del Anexo¹⁵ mientras que la estimación del modelo de frontera estocástica para el *pool* de datos y para los tres grupos del *cluster* están recogidas en el cuadro A2¹⁶. En primer lugar se analizan las elasticidades-output (coeficientes de primer orden en la translog) y la elasticidad de escala (la suma de las elasticidades de producción).

El cuadro 4 contiene las elasticidades-output estimadas para los dos inputs principales (vacas y pienso) así como la elasticidad de escala. Para que la comparación de la frontera estocástica *pooled* con los otros dos modelos sea más informativa, se evalúan las elasticidades de la frontera *pooled* en los grupos obtenidos en el modelo de clases latentes. Se analizan dos tipos de diferencias: diferencias

(13) El algoritmo *cluster* empleado ha sido el “conglomerado de k-medias” del programa SPSS 12.0.

(14) En el modelo de clases latentes se considera que cada explotación pertenece al grupo en el que tiene una mayor probabilidad *a posteriori*. También se calcularon las estadísticas como una media ponderada, en donde las ponderaciones son las probabilidades *a posteriori* de pertenecer a cada uno de los grupos y los resultados son prácticamente idénticos.

(15) De las variables separadoras empleadas sólo resultó significativa el pienso por vaca. Esta variable tiene signo positivo para el grupo intensivo y negativo para el extensivo, lo que indica que un valor mayor de la variable aumenta la probabilidad de clasificar una explotación en un grupo más intensivo.

(16) Los efectos temporales tienen la misma estructura en todos los modelos y grupos. Estas variables recogen básicamente dos aspectos: el cambio técnico neutral y el efecto de variables no medidas variantes en el tiempo y comunes entre explotaciones (por ejemplo, la climatología). El año 2003 fue un año muy seco, especialmente desde febrero hasta mayo. Dado que los expertos consideran que la precipitación primaveral es la más importante para determinar la cantidad de forraje producida en los pastos, la caída en el efecto temporal del 2003 es posible que recoja este efecto.

Cuadro 2: ESTADÍSTICAS DE LOS GRUPOS SEGÚN EL MÉTODO EMPLEADO PARA CLASIFICAR

	Modelo de clases latentes				Clasificación <i>cluster</i>			
	Grupo Extensivo	Grupo Semi-intensivo	Grupo Intensivo	Grupo Extensivo	Grupo Semi-intensivo	Grupo Intensivo	Grupo Extensivo	Grupo Intensivo
Explotaciones	38	71	60	66	67	36		
Litros de leche	210.739	283.490	378.421	216.976	313.906	430.248		
Vacas	36	39	45	34	41	50		
Tierra	19	19	20	17	19	23		
Trabajo	1,84	1,92	2,08	1,76	2,02	2,22		
Leche / Vacas	5.821	7.140	8.271	6.321	7.468	8.524		
Leche / Tierra	11.784	15.995	19.489	13.319	17.201	20.035		
Leche / Pienso	2,04	2,16	2,21	2,32	2,11	1,91		
Pienso / Vacas	2.959	3.409	3.846	2.770	3.585	4.505		
Vacas / Tierra	2,02	2,24	2,34	2,10	2,28	2,34		

Fuente: Elaboración propia.

Cuadro 3: COMPARACIÓN DE LA CLASIFICACIÓN DEL MODELO DE CLASES LATENTES Y EL ANÁLISIS *CLUSTER*

		Análisis <i>cluster</i>			
		Extensivo	Semi-intensivo	Intensivo	Total
Modelo	Extensivo	26	9	3	38
Clases	Semi-intensivo	28	31	12	71
Latentes	Intensivo	12	27	21	60
	Total	66	67	36	169

Fuente: Elaboración propia.

Cuadro 4: ELASTICIDADES-OUTPUT Y DE ESCALA

	Elasticidad-output del input vacas			Elasticidad-output del input pienso			Elasticidad de escala		
	Ext.	Sem.	Int.	Ext.	Sem.	Int.	Ext.	Sem.	Int.
<i>Pooled</i>	0,37	0,45	0,51	0,50	0,46	0,42	1,02	1,05	1,07
MCL*	0,36	0,48	0,73	0,39	0,35	0,34	0,95	0,94	1,17
<i>Cluster</i>	0,47	0,54	0,69	0,37	0,42	0,33	0,95	1,13	1,07

* Modelo de clases latentes.

Fuente: Elaboración propia.

entre los modelos (frontera estocástica *pooled*, modelo de clases latentes y *cluster*) y diferencias entre los tres grupos de cada modelo (extensivo, semi-intensivo e intensivo).

La comparación de las elasticidades-output de los tres modelos indica que existen importantes diferencias en las elasticidades estimadas. Con respecto a la comparación entre el modelo de clases latentes y el modelo *pooled*, dado que las explotaciones que forman los grupos son las mismas, estas diferencias se deben sólo a las características tecnológicas estimadas, por lo que se verifica que hay diferencias importantes entre estos modelos. Con respecto a las diferencias entre grupos, a medida que aumenta la intensificación en todos los modelos se observa una tendencia a aumentar la elasticidad-output de las vacas y a disminuir la elasticidad-output del pienso.

En la elasticidad de escala también se observan diferencias importantes entre los modelos. En el modelo *pooled* todos los grupos gozan de economías de escala, mientras que en el *cluster* hay dos grupos con economías de tamaño, y en el modelo

de clases latentes sólo el grupo intensivo tiene rendimientos crecientes a escala¹⁷. En general, la elasticidad de escala tiende a aumentar con la intensificación. Este resultado es en cierta manera sorprendente dado que, en general, las explotaciones más intensivas son más grandes y se supone que a medida que las explotaciones crecen agotan las economías de tamaño. Por esta razón, se esperaba que la elasticidad de escala de las explotaciones extensivas fuese mayor que la de las explotaciones intensivas. Sin embargo, los diferentes grupos de explotaciones emplean distintas tecnologías, por lo que es posible que las explotaciones intensivas hayan agotado menos las economías de tamaño que las extensivas. De hecho, los resultados del modelo *cluster* indican que el 90% de las explotaciones intensivas y semi-intensivas presentan rendimientos crecientes a escala, mientras que este porcentaje sólo es del 27% para las explotaciones extensivas.

Este resultado permite matizar el resultado encontrado en anteriores trabajos [Álvarez y Arias (2003)] en los que, al no considerar la posible existencia de diferencias tecnológicas, se había obtenido la conclusión de que existían deseconomías de tamaño generalizadas en el sector lechero asturiano. En ese mismo trabajo se sugería que la existencia de las deseconomías de tamaño podía deberse a la presencia de algún factor fijo, como la capacidad de gestión de los ganaderos, lo que implicaba que una pieza clave para el desarrollo del sector era la potenciación del asesoramiento técnico de los ganaderos. Sin contradecir esa idea, el presente trabajo sugiere que la adopción de sistemas de producción más intensivos proporciona una fuente de economías de tamaño en el sector y, por tanto, una forma de superar el tradicional problema del minifundismo sin caer en las deseconomías de tamaño.

Por otra parte, nuestro resultado de que muchas de las explotaciones intensivas no han agotado sus economías de tamaño permite dar explicación al hecho de que durante los últimos años las explotaciones más grandes han comprado cuota a las más pequeñas. Hasta el momento este era un resultado que no podía ser explicado a la luz de anteriores caracterizaciones de la tecnología. Dado que ambos tipos de explotaciones se enfrentan a precios de la leche similares, la presencia de deseconomías de escala hacía que las explotaciones más pequeñas tuviesen costes marginales menores que las más grandes, por lo que el valor marginal de un litro de cuota debería ser mayor para las explotaciones pequeñas que para las grandes. Sin embargo, el descubrimiento en nuestro análisis de que las explotaciones más grandes son casi todas intensivas y, por tanto, con mayores economías de tamaño, permite entender el anterior fenómeno.

5. EFICIENCIA TÉCNICA

La eficiencia técnica (ET) refleja la habilidad de una explotación para producir la máxima cantidad de output dada la cantidad utilizada de todos los inputs. En esta sección se estiman índices de eficiencia técnica en los tres modelos. En una frontera estocástica (como las que se estiman en el modelo *pooled* y en los grupos

(17) Las economías de escala del *cluster* calculadas según los grupos del modelo de clases latentes son 0,98, 1,05 y 1,08 para el grupo extensivo, semi-intensivo e intensivo, respectivamente.

del *cluster*) el índice de eficiencia técnica orientado al output se calcula mediante la siguiente expresión:

$$ET_{it} = \exp(-\hat{u}_{it}) \quad [13]$$

donde la ineficiencia u_{it} se separa del otro componente del error (v_{it}) aplicando la fórmula de Jondrow *et al.* (1982).

El cálculo del índice de eficiencia técnica en el modelo de clases latentes no es inmediato pues cada explotación tiene tres fronteras de referencia, cada una de ellas con una probabilidad asociada. En la literatura se han propuesto dos alternativas [Greene (2002)]: se puede calcular la eficiencia técnica respecto a la referencia tecnológica más probable (la de mayor probabilidad *a posteriori*) o se puede calcular una media ponderada de la eficiencia técnica para todas las posibles referencias tecnológicas usando como ponderación las probabilidades *a posteriori*. Es decir:

$$\ln ET_{it} = \sum_{j=1}^J P(j/i) \cdot \ln ET_{it}|_j \quad [14]$$

Se ha optado por la segunda alternativa al considerar que refleja mejor toda la información que contiene el modelo de clases latentes, mientras que la primera omite parte de las probabilidades *a posteriori* del modelo de clases latentes, información que puede ser relevante. También utilizan esta alternativa Orea y Kumbhakar (2004) y Moutinho *et al.* (2003).

Cuadro 5: MEDIA DE LOS ÍNDICES DE EFICIENCIA TÉCNICA

	Extensivo	Semi-intensivo	Intensivo
<i>Pooled</i>	0,84	0,90	0,93
MCL	0,90	0,97	0,98
<i>Cluster</i>	0,86	0,90	0,89

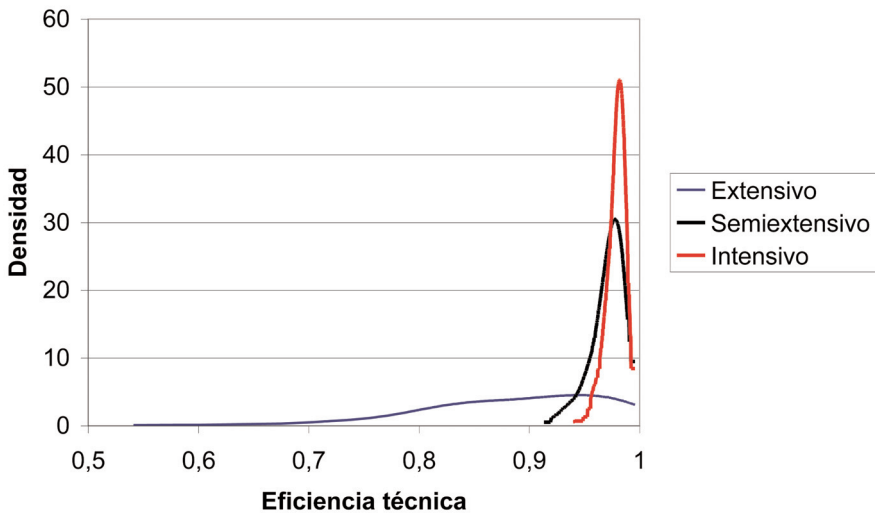
Fuente: Elaboración propia.

En el cuadro 5 se pueden ver las medias de los índices de eficiencia técnica obtenidos en cada modelo. La eficiencia media más alta para cada grupo se obtiene siempre en el modelo de clases latentes. Asimismo, la media de la eficiencia del modelo de clases latentes (0,96) es mayor que la media de la eficiencia en el modelo *pooled* (0,90). Estos resultados eran esperables ya que el modelo de clases latentes utiliza como criterio adicional para identificar grupos la bondad de ajuste de cada frontera estimada, por lo que es capaz de separar mejor las explotaciones que usan tecnologías distintas. De este modo, no se etiqueta como ineficiencia lo que realmente son diferencias en la tecnología empleada por las explotaciones [Stigler (1976)]. En los otros dos modelos la eficiencia media se sitúa por encima de 0,84. Estos niveles de eficiencia son superiores a los obtenidos en otros estudios anterior-

res del mismo sector¹⁸. Adicionalmente, se observa que el grupo intensivo es el más eficiente en los tres modelos, lo que parece confirmar nuestra idea inicial de que, al ser una tecnología más sencilla, el sistema de producción intensivo puede permitir lograr mejoras en la eficiencia con respecto al sistema extensivo.

A continuación se presentan en el gráfico 2 las funciones de densidad de la eficiencia técnica de los tres grupos en el modelo de clases latentes calculadas mediante un *suavizado kernel*. Se aprecia claramente que la mayor dispersión se da en las explotaciones extensivas mientras que la menor dispersión se produce en las explotaciones intensivas. Este resultado también entra dentro de lo esperado ya que el manejo de las explotaciones intensivas es mucho más homogéneo que el de las extensivas, donde se observan grandes diferencias en decisiones importantes (época de siembra, altura de corte del pasto, tipos de forrajes sembrados,...).

Gráfico 2: *KERNELS* DE LA EFICIENCIA TÉCNICA EN EL MODELO DE CLASES LATENTES



Fuente: Elaboración propia.

(18) Álvarez y Arias (2004) con una base de datos similar pero con datos del período 1993-1998 obtuvieron un índice de eficiencia media de 0,7. Álvarez, Arias y Roibás (2005) obtuvieron un índice de eficiencia técnica media de 0,71 utilizando índices genéticos como inputs en la función de producción.

6. CONCLUSIONES

En este trabajo se ha estimado una función de producción usando tres modelos: 1) una frontera estocástica *pooled*; 2) un modelo de clases latentes en el contexto de una frontera estocástica; y 3) separando la muestra en grupos mediante un algoritmo *cluster* para posteriormente estimar una frontera estocástica para cada uno de los grupos. Estos métodos conducen a resultados diferentes. En concreto, las elasticidades output y las economías de escala son diferentes no sólo entre grupos sino también entre modelos. De hecho, a diferencia de estudios anteriores que han encontrado que existen deseconomías de tamaño a partir de tamaños de producción relativamente pequeños, en este trabajo se ha encontrado que esas deseconomías existen fundamentalmente para explotaciones extensivas, mientras que las explotaciones intensivas, a pesar de ser mucho mayores, todavía no han agotado sus economías de escala. Por otra parte, se han obtenido diferencias importantes en los índices de eficiencia técnica entre los tres modelos. Los resultados sugieren que el modelo de clases latentes parece una técnica adecuada para no confundir ineficiencia con tecnologías diferentes.

Un resultado importante de este trabajo es que las explotaciones intensivas están situadas más cerca de su frontera que las extensivas de la suya. Esta mayor eficiencia técnica de las explotaciones intensivas parece confirmar que el sistema intensivo es más fácil de manejar que el extensivo. A pesar de que las causas de las diferencias entre ambos sistemas no quedan identificadas en nuestro análisis, no parece aventurado concluir que una buena parte de la ineficiencia de las explotaciones extensivas puede deberse al mal manejo de la producción de forrajes.

Una pregunta que tiene gran interés para la política agraria es la de qué sistema se debería promocionar. Esta pregunta no puede contestarse con la información de la que se dispone. Esta decisión tiene que basarse en criterios de rentabilidad y no sólo en el nivel de eficiencia técnica de los distintos grupos.

ANEXO

Cuadro A1: ESTIMACIÓN DEL MODELO DE CLASES LATENTES CON TRES GRUPOS

	Grupo Extensivo	Grupo Semi-intensivo	Grupo Intensivo
Constante	12,55***	12,51***	12,62***
Vacas	0,359***	0,478***	0,729***
Pienso	0,390***	0,348***	0,344***
Trabajo	0,028	-0,001	0,005
Gforraje	0,172***	0,111***	0,095***
0,5·Vacas·Vacas	-0,007	0,528	-0,324
0,5·Pienso·Pienso	0,147	-0,135	0,006
0,5·Trabajo·Trabajo	-0,444**	0,137**	-0,069
0,5·Gforraje·Gforraje	0,097	0,059	0,066
Vacas·Pienso	-0,231	0,029	0,073
Vacas·Trabajo	0,705***	-0,058	-0,122
Vacas·Gforraje	-0,119	-0,128	0,107
Pienso·Trabajo	-0,340*	0,089	0,023
Pienso·Gforraje	-0,009	0,019	-0,183***
Trabajo·Gforraje	-0,036	0,022	0,099
D2000	-0,065	0,012	0,027
D2001	-0,040	0,029*	0,040**
D2002	0,005	0,065***	0,036**
D2003	-0,018	0,030**	0,031
D2004	0,005	0,045***	0,060***
Constante	2,60		-3,09**
Vacas/Tierra	-0,535		0,131
Pienso/Vacas	-0,001		0,001**
$\lambda = \sigma_u / \sigma_v$	14,25	1,34**	0,97***
$\sigma = [\sigma_v^2 + \sigma_u^2]^{1/2}$	0,15***	0,09***	0,08***
Log F. Verosimilitud		1.119	
Observaciones		1.014	
Elasticidad de escala	0,95	0,94	1,17

*, **, *** Niveles de significación al 10%, 5% y 1%, respectivamente.

Fuente: Elaboración propia.

Cuadro A2: ESTIMACIONES DE LAS FRONTERAS ESTOCÁSTICAS
PARA LOS MODELOS ALTERNATIVOS

	Frontera Estocástica <i>Pooled</i>	Fronteras estocásticas para los grupos <i>cluster</i>		
		Grupo Extensivo	Grupo Semi-intensivo	Grupo Intensivo
Constante	12,62***	12,41***	12,67***	12,99***
Vacas	0,454***	0,474***	0,536***	0,691***
Pienso	0,454***	0,373***	0,418***	0,334***
Trabajo	0,020	-0,028	0,054*	-0,026
Gforraje	0,123***	0,127***	0,118***	0,070**
0,5·Vacas·Vacas	-0,914***	-1,297***	-0,565	1,883**
0,5·Pienso·Pienso	-0,223**	-0,979***	-0,192	1,326***
0,5·Trabajo·Trabajo	-0,078	-0,277***	0,221*	-0,068
0,5·Gforraje·Gforraje	-0,012	0,037	0,064	-0,112
Vacas·Pienso	0,441***	1,062***	0,222	-1,575***
Vacas·Trabajo	0,209***	-0,065	-0,017	-0,099
Vacas·Gforraje	0,174**	-0,051	0,255	0,044
Pienso·Trabajo	-0,086	0,071	0,138	-0,015
Pienso·Gforraje	-0,116**	0,010	-0,235*	-0,039
Trabajo·Gforraje	0,059*	0,065*	0,010	0,117
D2000	0,008	-0,017	0,013	0,018
D2001	0,018	0,018	0,013	0,035
D2002	0,035**	0,045**	0,018	0,040
D2003	0,009	0,008	0,001	0,022
D2004	0,031**	0,044**	0,010	0,053*
$\lambda = \sigma_u / \sigma_v$	1,91***	5,89***	1,95***	3,43***
$\sigma = [\sigma_v^2 + \sigma_u^2]^{1/2}$	0,16***	0,19**	0,15***	0,16***
Log F. Verosimilitud	795	325	329	192
N.º observaciones	1.014	396	402	216
Elasticidad de escala	1,05	0,95	1,13	1,07

*, **, *** Niveles de significación al 10%, 5% y 1%, respectivamente.

Fuente: Elaboración propia.



REFERENCIAS BIBLIOGRÁFICAS

- Aigner, D., K. Lovell y P. Schmidt (1977): "Formulation and Estimation of Stochastic Frontier Production Function Models", *Journal of Econometrics*, vol. 6, págs. 21-37.
- Álvarez, A. (2001): *La medición de la eficiencia y la productividad*, Pirámide, Madrid.
- Álvarez, A. y C. Arias (2003): "Diseconomies of Size with Fixed Managerial Ability in Dairy Farms", *American Journal of Agricultural Economics*, vol. 85, 1, págs. 134-142.
- Álvarez, A. y C. Arias (2004): "Technical Efficiency and Farm Size. A Conditional Analysis", *Agricultural Economics*, vol. 30, 3, págs. 241-250.
- Álvarez, A., C. Arias y D. Roibás (2005): "Análisis de la calidad de la leche en un modelo microeconómico multi-output: el papel de la genética", *Economía Agraria y Recursos Naturales*, vol. 5, págs. 3-17.
- Álvarez, A., J. del Corral, J.A. Pérez y D. Solís (2007): "Efecto de la intensificación sobre la eficiencia de las explotaciones lecheras", *Economía Agraria y Recursos Naturales*, vol. 7, 2, págs. 91-106.
- Beard T., S. Caudill y D. Gropper (1991): "Finite Mixture Estimation of Multiproduct Cost Functions", *Review of Economics and Statistics*, vol. 73, págs. 654-664.
- Cameron, C. y P.K. Trivedi (2005): *Microeconometrics: Methods and Applications*, Cambridge University Press, New York.
- Caudill, S. (2003): "Estimating a Mixture of Stochastic Frontier Regression Models via the EM: A Multiproduct Cost Function Application", *Empirical Economics*, vol. 28, 3, págs. 581-598.
- Charnes, A., W. Cooper, A. Lewin y L. Seiford (1995): *Data Envelopment Analysis: Theory, Methodology and Applications*, Kluwer Academic Publishers, Boston.
- El-Gamal, M.A. y H. Inanoglu (2005): "Inefficiency and Heterogeneity in Turkish Banking: 1990-2000", *Journal of Applied Econometrics*, vol. 20, págs. 641-664.
- Greene, W. (2001): "New Developments in the Estimation of Stochastic Frontier Models with Panel Data", *Efficiency Series Paper 6/2001*, Departamento de Economía, Univ. de Oviedo.
- Greene, W. (2002): "Alternative Panel Data Estimators for Stochastic Frontier Models", Working Paper, Department of Economics, Stern School of Business, NYU.
- Greene, W. (2004): "Distinguishing between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organization's Panel Data on National Health Care Systems", *Health Economics*, vol. 13, 10, págs. 959-980.
- Greene, W. (2005): "Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model", *Journal of Econometrics*, vol. 126, págs. 269-303.
- Griliches, Z. (1957): "Specification Bias in Estimates of Production Function", *Journal of Farm Economics*, vol. 39, 1, págs. 8-20.
- Hildreth, C. y J. Houck (1968): "Some Estimators for a Linear Model with Random Coefficients", *Journal of the American Statistical Association*, vol. 63, págs. 584-595.
- Hoch, I. (1962): "Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data", *Econometrica*, vol. 30, 1, págs. 34-53.
- Jondrow, J., K. Lovell, I. Materov y P. Schmidt (1982): "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model", *Journal of Econometrics*, vol. 19, págs. 233-238.
- Kumbhakar, S., E. Tsionas, B. Park y L. Simar (2007): "Nonparametric Stochastic Frontiers: A Local Maximum Likelihood Approach", *Journal of Econometrics*, vol. 137, págs. 1-27.
- Maudos, J., J. Pastor y F. Pérez (2002): "Competition and Efficiency in the Spanish Banking Sector: the Importance of Specialization", *Applied Financial Economics*, vol. 12, págs. 505-516.

- Meeusen, W. y J. van den Broeck (1977): "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error", *International Economic Review*, vol. 18, págs. 435-444.
- Moutinho, N., F. Machado y E. Silva (2003): "An International Comparison of Productivity Change in Agriculture and the Economy as a Whole", Discussion Papers 9-2003, Faculdade de Economia, Universidade de Porto.
- Mundlak, Y. (1961): "Empirical Production Function Free of Management Bias", *Journal of Farm Economics*, vol. 43, págs. 44-56.
- Orea, L. y S. Kumbhakar (2004): "Efficiency Measurement using a Stochastic Frontier Latent Class Model", *Empirical Economics*, vol. 29, págs. 169-183.
- SADEI (2005): *Las explotaciones ganaderas en Asturias, 2004*, Servicio de Publicaciones del Principado de Asturias.
- Stigler, G. (1976): "The Xistence of X-Efficiency", *American Economic Review*, vol. 66, 213-216.

Fecha de recepción del original: enero, 2006
Versión final: enero, 2007

ABSTRACT

The estimation of production functions usually relies on the assumption that all producers use the same technology. However, it is possible that some firms use different technologies. In this paper, we use a stochastic frontier latent class model to estimate the technology of several dairy farms' production systems which present different levels of intensification. The results of this model are compared with two other methods: the classical stochastic frontier model; and the estimation of separate production functions based on the classification of farms using a *cluster* algorithm. The empirical analysis uses a balanced panel of 169 dairy farms over the period 1999 to 2004.

Key words: Dairy farms, production function, stochastic frontier, unobserved heterogeneity, latent class model.

JEL classification: Q12, C23.