

# FRONTERAS DE PRODUCCIÓN

Ángela Corte González

Universidad de Oviedo

Máster en Modelización e Investigación Matemática, Estadística y Computación

Febrero 2015

# Índice general

<b>1. Introducción</b>	<b>3</b>
<b>2. Datos transversales</b>	<b>6</b>
2.1. Descripción de los datos . . . . .	8
<b>3. Frontera de producción determinística</b>	<b>10</b>
3.1. Programación matemática . . . . .	11
3.1.1. Ejemplo lineal . . . . .	13
3.1.2. Ejemplo cuadrático . . . . .	15
3.2. Mínimos cuadrados ordinarios corregidos (COLS) . . . . .	19
3.2.1. Ejemplo . . . . .	21
3.3. Mínimos cuadrados ordinarios modificados (MOLS) . . . . .	23
3.3.1. Ejemplo . . . . .	24
3.4. Comparación . . . . .	26
<b>4. Frontera de producción estocástica</b>	<b>28</b>
4.1. Método de máxima verosimilitud . . . . .	29
4.1.1. Modelo normal - seminormal . . . . .	29
4.1.2. Ejemplo N - $N^+$ . . . . .	35
4.1.3. Modelo normal - exponencial . . . . .	40
4.1.4. Ejemplo N - Exp . . . . .	46
4.2. Método de los momentos . . . . .	48

4.2.1.	Modelo normal - seminormal . . . . .	49
4.2.2.	Ejemplo N - $N^+$ . . . . .	52
4.2.3.	Modelo normal - exponencial . . . . .	55
4.2.4.	Ejemplo N - Exp . . . . .	57
4.3.	Comparación . . . . .	61
4.3.1.	Resumen de resultados . . . . .	61
<b>5.</b>	<b>Datos de panel</b>	<b>69</b>
5.1.	Descripción de los datos de panel . . . . .	70
5.2.	Eficiencia técnica invariante en el tiempo . . . . .	73
5.2.1.	Modelo de efectos fijos . . . . .	74
5.2.2.	Modelo de efectos aleatorios . . . . .	75
5.2.3.	Ejemplos . . . . .	76
<b>6.</b>	<b>Extensiones del modelo</b>	<b>81</b>
6.1.	Eficiencia técnica invariante en el tiempo . . . . .	81
6.1.1.	Método de máxima verosimilitud . . . . .	81
6.2.	Eficiencia técnica variante en el tiempo . . . . .	83
<b>7.</b>	<b>Conclusiones y puntos abiertos</b>	<b>85</b>

# Capítulo 1

## Introducción

El presente trabajo trata los diferentes modelos de fronteras de producción y en él se desarrollan los cálculos subyacentes a las fórmulas publicadas en las referencias fundamentales. El objetivo de los productores o empresas puede ser tan simple como buscar el modo de evitar el gasto, es decir, obtener las máximas producciones dadas unas entradas, o minimizar los recursos utilizados en la producción dados los productos. Por esto se estudiarán fronteras de producción, porque identifican las combinaciones de los distintos tipos de bienes que se pueden producir cuando todos los recursos disponibles de una empresa (en un momento dado) se utilizan de forma eficiente. Es importante el cambio de técnicas tradicionales a las fronteras para realizar el análisis de la producción y ser capaces de utilizar con eficiencia los recursos mínimos requeridos para obtener los productos que se desean desarrollar. En la práctica, las fronteras de producción no son más que una regresión que se ajusta a los datos con la restricción de que todas las observaciones deben encontrarse debajo de la frontera y al menos una debe estar sobre ella. De esta forma, se puede interpretar la frontera de producción como el límite máximo de producción que sirve de referencia para el cálculo de la ineficiencia del resto de las unidades observadas. Los productores que operan en su frontera de producción se dirá que son técnicamente eficientes, puesto que son los niveles máximos de producción posibles dados unos recursos determinados. Por otro lado, los productores que están por debajo de ella serán técnicamente ineficientes. Los puntos fuera de la frontera de producción representan volúmenes de producción que no son alcanzables a partir de los recursos con los que cuenta.

El primer autor que motivó la utilización de las fronteras de producción para analizar la eficiencia fue Farrell en 1957. Para él, la forma correcta de medir la eficiencia era mediante la comparación de cada observación con la mejor práctica observada. Desde su trabajo, la forma más popular para cuantificar la eficiencia de un grupo de empresas consiste en especificar y estimar una frontera de producción. Por ejemplo, en 1977 Aigner, Lovell y Shmidt propusieron el modelo de frontera estocástica.

Los modelos que se van a estudiar serán de ecuación única, de los cuales se estiman los parámetros que describen la estructura de la frontera y las estimaciones de la eficiencia técnica de cada productor se obtienen como subproducto de lo anterior. El término de error asociado a las fronteras de producción está compuesto de dos términos, el ruido aleatorio tradicional simétrico y una nueva componente de ineficiencia unilateral. Se retiene la componente de error simétrico para capturar los efectos de la variación aleatoria en el entorno operativo y se incluye la componente unilateral para capturar los efectos de ineficiencia.

Las técnicas de estimación empleadas dependerán de la cantidad de datos disponibles. Se trabajará con dos tipos de datos, transversales y de panel. En los primeros se observa cada empresa en un determinado periodo de tiempo, mientras que los datos de panel recogen varias observaciones correspondientes a diferentes periodos de tiempo.

A continuación se detallará la estructura del trabajo. En el capítulo dos se define lo que son los datos transversales y la base de datos que se utilizará para comprobar las diferentes técnicas que se estudiarán a posteriori. Puesto que existen dos tipos de fronteras de producción para datos transversales (determinísticas y estocásticas), en el capítulo tres se describirán las tres técnicas determinísticas que hay: programación matemática, mínimos cuadrados corregidos y mínimos cuadrados modificados. Así como resolver un ejemplo para cada una de ellas y, al final del capítulo, hay una comparación de las técnicas. En el capítulo cuatro están los métodos para resolver fronteras estocásticas, que son máxima verosimilitud o el método de los momentos. Cada uno de ellos se divide a su vez en dos, en función de la suposición de la distribución de la ineficiencia, la cual puede ser seminormal o exponencial, entre otras. También se hace un ejemplo de cada uno de ellos y, al final del capítulo, una comparación de los métodos, resumen de todos los resultados obtenidos por las diferentes técnicas para datos

transversales y varias gráficas comparativas. Los dos últimos capítulos están dedicados a los datos de panel. En función de si se considera la eficiencia técnica invariante o no en el tiempo, hay distintas metodologías. Cuando la eficiencia técnica es invariante en el tiempo existe el método de efectos fijos y efectos aleatorios, los cuales se estudian en el capítulo cinco. Se hará un ejemplo de los mismos con unos datos de panel que también serán descritos en el capítulo. En el último, se describe el método de máxima verosimilitud para eficiencia técnica invariante en el tiempo, así como la definición del modelo cuando se considera que la eficiencia técnica varía en el tiempo.

## Capítulo 2

# Datos transversales

El objetivo de este capítulo es hacer un estudio de datos transversales (cross-sectional). Es un tipo de estudio observacional e inferencial, que permite estimar la magnitud y distribución de la producción en un momento dado.

Se supone que el vector de recursos es no-negativo y se denotada por  $\vec{x} = (x_1, x_2, \dots, x_n) \in (\mathbb{R}^+)^n$ . La producción es un vector no-negativo de salidas  $y$ . El modelo de frontera de producción se escribe como:

$$y_i = f(\vec{x}_i; \vec{\beta}) \cdot \text{TE}_i$$

donde

- $y_i$  es el escalar salida del productor  $i$ ,  $i = 1, \dots, I$ .
- $\vec{x}_i$  es un vector de  $N$  entradas del productor  $i$ .
- $f(\vec{x}_i, \vec{\beta})$  es la frontera de producción, y se supondrá que la función de producción es la siguiente  $f(\vec{x}_i, \vec{\beta}) = \prod_n x_{ni}^{\beta_n}$ .
- $\vec{\beta}$  es el vector de parámetros a estimar.

Como  $\text{TE}_i$  representa la eficiencia técnica del productor  $i$ , despejando de la ecuación anterior, se tiene que:

$$\text{TE}_i = \frac{y_i}{f(\vec{x}_i; \vec{\beta})}$$

lo que define la eficiencia técnica como una proporción entre la salida observada y el máximo valor factible de la producción. De modo que  $y_i$  alcanza el máximo valor factible de  $f(\vec{x}_i; \vec{\beta})$  si y sólo si  $TE_i = 1$ . En otro caso, la eficiencia técnica es menor que uno y proporciona una medida de déficit observada en la producción sobre el máximo de la producción factible.

En la ecuación anterior de  $y_i$ , la **frontera de producción**  $f(\vec{x}_i, \vec{\beta})$  es **determinística**. Como consecuencia, todo el déficit de producción observado,  $y_i$ , sobre el máximo valor alcanzable,  $f(\vec{x}_i; \vec{\beta})$ , se atribuye a la ineficiencia técnica, ya que no se recoge el efecto que pueden ejercer perturbaciones aleatorias sobre la producción no controlables por el productor.

Para atribuir variaciones en la producción no asociadas con variaciones en las entradas, se combinan efectos aleatorios e ineficiencia que permitan desarrollar un modelo más realista. Para ello se requiere especificar una frontera de producción estocástica, la cual puede ser escrita como:

$$y_i = f(\vec{x}_i, \vec{\beta}) \cdot \exp\{v_i\} \cdot TE_i$$

donde  $f(\vec{x}_i, \vec{\beta}) \cdot \exp\{v_i\}$  es la **frontera de producción estocástica**, la cual está formada por una parte determinística,  $f(\vec{x}_i, \vec{\beta})$ , común a todos los productores y una parte específica de cada productor,  $\exp\{v_i\}$ , que recoge el efecto aleatorio. En este caso, la eficiencia técnica se calcula como:

$$TE_i = \frac{y_i}{f(\vec{x}_i, \vec{\beta}) \cdot \exp\{v_i\}}$$

De esta forma, la eficiencia técnica puede ser estimada utilizando el modelo de frontera de producción estocástica, el cual es preferido por incluir el efecto aleatorio en la producción.

Considerando que  $TE_i = \exp\{-u_i\}$ , suponiendo  $f(\vec{x}_i, \vec{\beta}) = \prod_n x_{ni}^{\beta_n}$  y tomando logaritmos, en el modelo de frontera de producción estocástica, la ecuación se escribe como

$$\ln y_i = \beta_0 + \sum_n \beta_n \ln x_{ni} + v_i - u_i$$

donde  $v_i$  es la componente ruido y  $u_i$  es la componente no negativa que mide la ineficiencia técnica. Juntas, es decir,  $\epsilon_i = v_i - u_i$ , forman el término de error. La componente  $v_i \rightsquigarrow N(0, \sigma_v^2)$  que es simétrica, idénticamente distribuida e independiente de  $u_i$ . Como  $u_i \geq 0$ , el término de error,  $\epsilon_i$ , es asimétrico respecto al cero.



Por tanto, el objetivo es obtener estimaciones tanto de los parámetros de producción  $\vec{\beta}$  en  $f(\vec{x}_i, \vec{\beta})$  como de la eficiencia técnica de cada productor. Para conseguir la segunda parte se requiere que las estimaciones separadas del ruido estadístico,  $v_i$ , y de la ineficiencia técnica,  $u_i$ , sean extraídas de la estimación de  $\epsilon_i$  para cada productor, para lo que hay que asumir distribuciones en las dos componentes de error.

## 2.1. Descripción de los datos

En esta sección se van a describir los datos transversales con los que se va a trabajar para probar cada método que se vaya detallando en los dos siguientes capítulos. El software que se utilizará es R y los datos son del paquete `frontier`, y se llaman `front41Data`. Lo primero que hay que hacer es instalar el paquete (`install.packages("frontier")`), cargar la librería y los datos con los que se va a trabajar. Las órdenes son las siguientes:

```
> library(frontier)
> data(front41Data)
```

Los datos son transversales y correspondientes a 60 empresas. Las variables que contiene la base de datos se describen a continuación:

- `firm`: ID de la empresa.
- `output`: producción de la empresa.
- `capital`: capital de la empresa.
- `labour`: trabajo de la empresa.

A continuación, por comodidad, se abrevia el nombre de la base de datos, que pasará a denotarse por 'f'. Con la función `summary` se hace un resumen de las variables de la base, es decir, hallar los cuartiles, media y valores máximo y mínimo de cada variable:

```
> f <- front41Data
> summary(f)
```

firm	output	capital	labour
Min. : 1.00	Min. : 1.66	Min. : 0.258	Min. : 1.073
1st Qu.: 15.75	1st Qu.: 11.82	1st Qu.: 2.622	1st Qu.: 36.518
Median : 30.50	Median : 14.93	Median : 4.757	Median : 60.041
Mean : 30.50	Mean : 16.62	Mean : 4.849	Mean : 57.682
3rd Qu.: 45.25	3rd Qu.: 21.25	3rd Qu.: 7.197	3rd Qu.: 80.647
Max. : 60.00	Max. : 37.12	Max. : 9.561	Max. : 98.904

El objetivo de las fronteras de producción es estimar la eficiencia técnica de los productores, lo que requiere utilizar los recursos eficientemente para maximizar beneficios. La variable beneficios o producción, en esta base de datos, es la llamada 'output'. A continuación, se define el logaritmo de la variable y se denotará por  $y$ . Las variables de entrada son 'capital' y 'labour', a las que también hay que hacerles la transformación a logaritmo y después denotarlas por  $x_1$  y  $x_2$ , ya que se ha descrito anteriormente que en los modelos de fronteras se trabaja con la transformación logarítmica de las variables:

```
> y <- f$y <- log(f$output)
> x1 <- f$x1 <- log(f$capital)
> x2 <- f$x2 <- log(f$labour)
```

## Capítulo 3

# Frontera de producción determinística

Un modelo de frontera determinística se había descrito como:

$$y_i = f(\vec{x}_i; \vec{\beta}) \cdot TE_i$$

Se puede observar que la frontera es determinística puesto que atribuye todo el déficit a la ineficiencia técnica y no recoge el efecto de perturbaciones aleatorias no controlables sobre la producción. Sustituyendo la eficiencia técnica en la ecuación anterior, el modelo quedaría como sigue:

$$y_i = f(\vec{x}_i; \vec{\beta}) \cdot \exp\{-u_i\}$$

donde  $TE_i = \exp\{-u_i\}$ . Se requiere que  $TE_i \leq 1$ , por lo que se tiene que  $u_i \geq 0$ . Tomando logaritmos en la ecuación anterior y sustituyendo  $f(\vec{x}_i; \vec{\beta}) = \prod_n x_{ni}^{\beta_n}$ , se obtiene:

$$\ln y_i = \beta_0 + \sum_n \beta_n \ln x_{ni} - u_i$$

donde  $u_i \geq 0$  garantiza que  $y_i \leq f(\vec{x}_i; \vec{\beta})$ . La ecuación es un modelo de regresión lineal con una perturbación negativa. El objetivo es obtener estimaciones del vector de parámetros  $\vec{\beta}$ , el cual describe la estructura de la frontera de producción, y también obtener estimaciones de los  $u_i$ , los cuales se utilizan para obtener estimaciones de  $TE_i$  de cada productor, ya que  $TE_i = \exp\{-u_i\}$ .

En este capítulo se estudiarán tres métodos para estimar las fronteras de producción y se hará un ejemplo de cada uno de ellos utilizando los datos descritos en el capítulo anterior.

### 3.1. Programación matemática

El objetivo de este método es calcular el vector de parámetros  $\vec{\beta}$  para el cual se minimiza la suma de las desviaciones de la salida observada de cada productor debajo de la máxima producción factible. Las desviaciones resultantes son convertidas en medidas de la eficiencia técnica de cada productor. Una posibilidad sería utilizar directamente las desviaciones, lo que sería un **modelo lineal** de programación que se escribe de la siguiente forma:

$$\begin{aligned} \text{mín} \quad & \sum_i u_i \\ \text{sujeto a} \quad & \beta_0 + \sum_n \beta_n \ln x_{ni} \geq \ln y_i, \quad \forall i = 1 \dots I. \end{aligned}$$

Si se despeja  $u_i$  del modelo de frontera determinístico, se tiene que el problema que hay que minimizar es:

$$\begin{aligned} \text{mín} \quad & \sum_i \left( \beta_0 + \sum_n \beta_n \ln x_{ni} - \ln y_i \right) \\ \text{sujeto a} \quad & \beta_0 + \sum_n \beta_n \ln x_{ni} \geq \ln y_i, \quad \forall i = 1 \dots I. \end{aligned}$$

Dentro de esta sección, se va a estudiar otro modelo que es el **cuadrático**, donde el objetivo es calcular el vector de parámetros  $\vec{\beta}$  para el cual se minimiza la suma de los cuadrados de las desviaciones de la salida observada de cada productor debajo de la máxima producción factible. El modelo se escribe como:

$$\begin{aligned} \text{mín} \quad & \sum_i u_i^2 \\ \text{sujeto a} \quad & \beta_0 + \sum_n \beta_n \ln x_{ni} \geq \ln y_i, \quad \forall i = 1 \dots I. \end{aligned}$$

cuya función objetivo se podría reescribir como se ha hecho en el modelo lineal.

Una vez calculados los valores de los parámetros de cada modelo, la ineficiencia técnica de cada productor se puede sacar de las holguras de las restricciones correspondientes:

$$u_i = \left( \beta_0 + \sum_n \beta_n \ln x_{ni} \right) - \ln y_i$$

Por tanto, la eficiencia técnica se calcularía como:

$$TE_i = \exp\{-u_i\}$$

En principio, no se puede hacer inferencia estadística sobre los parámetros, ya que no se estiman mediante procedimientos estadísticos, sino que se calculan sin tener en cuenta ninguna distribución de probabilidad. Sin embargo, se ha demostrado que estos modelos son equivalentes al método máximo verosímil cuando se impone una distribución de  $u_i$ , de hecho, la 'estimación' del modelo lineal es la estimación máximo verosímil de los parámetros de una frontera de producción determinística si  $u_i \geq 0$  se distribuye como una exponencial, es decir:

$$f(u) = \frac{1}{\sigma_u} \cdot \exp\left\{-\frac{u}{\sigma_u}\right\}$$

en cuyo caso, el logaritmo de la función de verosimilitud es:

$$\ln L = \ln\left(\prod_i \frac{1}{\sigma_u} \cdot \exp\left\{-\frac{u_i}{\sigma_u}\right\}\right) = -I \ln \sigma_u - \frac{1}{\sigma_u} \sum_i |u_i|$$

La 'estimación' del modelo cuadrático es la estimación máximo verosímil de los parámetros de una frontera de producción si  $u_i \geq 0$  se distribuye como una seminormal:

$$f(u) = \frac{2}{\sqrt{2\pi}\sigma_u} \cdot \exp\left\{-\frac{u_i^2}{2\sigma_u^2}\right\}$$

en cuyo caso el logaritmo de la función de verosimilitud es:

$$\ln L = \ln\left(\prod_i \frac{2}{\sqrt{2\pi}\sigma_u} \cdot \exp\left\{-\frac{u_i^2}{2\sigma_u^2}\right\}\right) = I \ln \frac{2}{\sqrt{2\pi}} - I \ln \sigma_u - \frac{1}{2\sigma_u^2} \sum_i u_i^2$$

Las técnicas de programación matemática calculan el vector de parámetros  $\vec{\beta}$  en vez de estimarlo, pero establecida esta relación entre fronteras de producción calculadas por programación y fronteras de producción estimadas por métodos de máxima verosimilitud, podría parecer que la programación tiene una base estadística.

A continuación se hace un ejemplo del modelo lineal y cuadrático con los datos descritos en el capítulo dos.

### 3.1.1. Ejemplo lineal

Para realizar el ejemplo de la programación lineal se necesita en R instalar el paquete `linprog`. La función que se utilizará para resolver el problema se llama `lpSolve` y requiere que el problema esté escrito de la siguiente forma:

$$\begin{array}{ll} \text{mín} & c'x \\ \text{sujeto a} & Ax \leq b \end{array}$$

Por tanto, hay que reescribir el problema que se quiere minimizar. Para simplificar cálculos se aplican al ejercicio a resolver, es decir, teniendo en cuenta que la base de datos solo tiene dos variables de entrada  $x_1$  y  $x_2$  y que las incógnitas son  $\vec{x} = (\beta_0, \beta_1, \beta_2)$ . Sustituyendo  $u_i$  por su expresión, se tiene que hay que minimizar:

$$I \cdot \beta_0 + \beta_1 \sum_i \ln x_{i1} + \beta_2 \sum_i \ln x_{i2} - \sum_i \ln y_i$$

por lo que el vector  $c'$  que acompaña al vector de incógnitas  $x$  en la función objetivo es  $c' = (I, \sum_i \ln x_{i1}, \sum_i \ln x_{i2})$ .

A las restricciones hay que cambiarlas de signo para que sean desigualdades de menor o igual, y quedarían como:

$$-\beta_0 - \beta_1 \ln x_{i1} - \beta_2 \ln x_{i2} \leq -\ln y_i, \quad i = 1, \dots, I$$

cuya notación es  $Ax \leq b$ , por lo que el vector  $b = -y$ , donde  $\vec{y}$  es el vector de logaritmos de las salidas de cada productor. La matriz

$$A = \begin{pmatrix} -1 & -\ln x_{1,1} & -\ln x_{1,2} \\ \vdots & \vdots & \vdots \\ -1 & -\ln x_{60,1} & -\ln x_{60,2} \end{pmatrix}$$

Ahora que ya está escrito el problema como requiere la función que se va utilizar, se puede resolver con R de la siguiente forma:

```
> library(linprog)
> I <- nrow(f)
```

```

> c <- c(I, sum(x1), sum(x2))
> b <- -y
> A <- cbind(-1, -x1, -x2)
> LP <- solveLP(c, b, A)

```

Tener en cuenta, que al describir la base de datos, se abrevió  $y := \ln y$ ,  $x_1 := \ln x_1$  y  $x_2 := \ln x_2$  y que el número de productores es el parámetro  $I$ . Como la salida es muy larga, se mostrarán solamente los valores de interés:

```

> beta <- LP$solution
> beta

```

	1	2	3
	0.2166105	0.4662259	0.6852786

Lo cual es el vector  $\vec{\beta}$  de parámetros estimados.

```

> restricciones1 <- LP$con
> restricciones1[1:10,]

```

	actual	dir	bvec	free	dual	dual.reg
1	-3.701103	<=	-2.547725	1.1533778	0	1.1533778
2	-3.911790	<=	-3.189859	0.7219314	0	0.7219314
3	-4.057838	<=	-3.037594	1.0202445	0	1.0202445
4	-3.410807	<=	-2.581201	0.8296060	0	0.8296060
5	-3.506614	<=	-2.486406	1.0202082	0	1.0202082
6	-3.346388	<=	-2.726807	0.6195817	0	0.6195817
7	-2.726299	<=	-1.903152	0.8231473	0	0.8231473
8	-4.097764	<=	-3.075283	1.0224811	0	1.0224811
9	-3.123339	<=	-2.679239	0.4440999	0	0.4440999
10	-3.268592	<=	-2.620530	0.6480622	0	0.6480622

En 'restricciones1' están los valores de las 60 restricciones del problema, para abreviar, se muestran solamente las 10 primeras. La columna 'free' se corresponde con las holguras de las restricciones, es decir, los valores  $u_i$  que también eran objeto a estimar.

El último paso es hallar la eficiencia técnica de todos los productores con la expresión  $TE_i = \exp\{-u_i\}$ . Se calcula como:

```
> TE1 <- exp(-restricciones1[,4]); TE1

[1] 0.3155691 0.4858130 0.3605068 0.4362211 0.3605199 0.5381695 0.4390477
[8] 0.3597014 0.6414013 0.5230584 0.3680984 1.0000000 0.1814117 0.3240079
[15] 0.9537863 0.2978997 0.3433427 0.4442395 0.6042778 0.4838818 0.4062809
[22] 0.7361606 0.5308606 0.6888414 0.3927837 0.6715827 0.4262823 0.4589890
[29] 0.7849319 0.4507818 0.3885825 1.0000000 0.7686230 0.1685969 0.1501772
[36] 0.6099808 0.4526023 0.3317396 0.3902408 0.5831097 0.4370800 0.3284715
[43] 0.4390552 0.6556938 0.3838842 0.4202454 0.4764655 0.4350162 0.3536106
[50] 0.2284272 0.5712764 0.4894400 0.4682380 0.4676979 0.5497414 0.3940833
[57] 1.0000000 0.2843091 0.5038161 0.3784635
```

En el vector 'TE1' están recogidas las eficiencias técnicas de los 60 productores de la base de datos con la que se está trabajando. A continuación se hace un resumen de la variable:

```
> summary(TE1)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.1502  0.3662  0.4416  0.4853  0.5551  1.0000
```

De esta forma se observa que la media de la eficiencia técnica de los productores es de 0,4853. Puesto que el valor máximo que toma la variable es 1, al menos un productor es 100 % técnicamente eficiente.

### 3.1.2. Ejemplo cuadrático

Se resolverá ahora el ejemplo con el método de programación cuadrática. Para ello es necesario instalar el paquete `quadprog`. La función que se utiliza para resolver el problema es



solve.QP y requiere que éste se escriba de la siguiente forma:

$$\begin{array}{ll} \text{mín} & -d'b + \frac{1}{2}b'Db \\ \text{sujeto a} & A'b \geq b_0 \end{array}$$

Para averiguar qué se corresponde en el problema a resolver con la notación anterior, se reescribirá el problema que hay que minimizar. Tener en cuenta, que en la base de datos solo hay dos variables de entrada  $x_1$  y  $x_2$  y que las incógnitas son  $\vec{b} = (\beta_0, \beta_1, \beta_2)$ . Sustituyendo  $u_i$  por su expresión y desarrollando el cuadrado, se tiene que hay que minimizar:

$$\begin{aligned} \sum_{i=1}^I u_i^2 &= \sum_i (\beta_0 + \beta_1 \ln x_{i1} + \beta_2 \ln x_{i2} - \ln y_i)^2 = \\ &= \sum_i [(\ln y_i)^2 - 2 \ln y_i \beta_0 - 2 \ln y_i \beta_1 \ln x_{i1} - 2 \ln y_i \beta_2 \ln x_{i2} + \beta_0^2 + 2\beta_0\beta_1 \ln x_{i1}] + \\ &\quad + \sum_i [2\beta_0\beta_2 \ln x_{i2} + \beta_1^2 (\ln x_{i1})^2 + 2\beta_1\beta_2 \ln x_{i1} \ln x_{i2} + \beta_2^2 (\ln x_{i2})^2] \\ &= \sum_i (\ln y_i)^2 - 2\beta_0 \sum_i \ln y_i - 2\beta_1 \sum_i \ln y_i \ln x_{i1} - 2\beta_2 \sum_i \ln y_i \ln x_{i2} + I\beta_0^2 + 2\beta_0\beta_1 \sum_i \ln x_{i1} + \\ &\quad + 2\beta_0\beta_2 \sum_i \ln x_{i2} + \beta_1^2 \sum_i (\ln x_{i1})^2 + 2\beta_1\beta_2 \sum_i \ln x_{i1} \ln x_{i2} + \beta_2^2 \sum_i (\ln x_{i2})^2 \end{aligned}$$

Para compararlo con la expresión que necesita la función para resolverlo, se harán explícitamente los cálculos para después ver la correspondencia de los vectores y matrices entre una y otra. Las letras de la  $a$  a la  $l$  son las incógnitas:

$$\begin{aligned} & -d'b + \frac{1}{2}b'Db = \\ &= -(a, b, c) \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \frac{1}{2}(\beta_0, \beta_1, \beta_2) \begin{pmatrix} d & e & f \\ g & h & i \\ j & k & l \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \\ &= -a\beta_0 - b\beta_1 - c\beta_2 + \frac{1}{2}(d\beta_0 + g\beta_1 + j\beta_2, e\beta_0 + h\beta_1 + k\beta_2, f\beta_0 + i\beta_1 + l\beta_2) \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \\ &= -a\beta_0 - b\beta_1 - c\beta_2 + \frac{1}{2}d\beta_0^2 + \frac{1}{2}g\beta_0\beta_1 + \frac{1}{2}j\beta_0\beta_2 + \frac{1}{2}e\beta_0\beta_1 + \frac{1}{2}h\beta_1^2 + \frac{1}{2}k\beta_1\beta_2 + \frac{1}{2}f\beta_0\beta_2 + \frac{1}{2}i\beta_1\beta_2 + \frac{1}{2}l\beta_2^2 \end{aligned}$$

Comparando con el desarrollo anterior, se tiene que:

$$\begin{aligned}
 a &= 2 \sum_i \ln y_i & b &= 2 \sum_i \ln y_i \ln x_{i1} & c &= 2 \sum_i \ln y_i \ln x_{i2} \\
 d &= 2 \cdot I & e &= 2 \sum_i \ln x_{i1} & f &= 2 \sum_i \ln x_{i2} \\
 g &= 2 \sum_i \ln x_{i1} & h &= 2 \sum_i \ln x_{i1}^2 & i &= 2 \sum_i \ln x_{i1} \ln x_{i2} \\
 j &= 2 \sum_i \ln x_{i2} & k &= 2 \sum_i \ln x_{i1} \ln x_{i2} & l &= 2 \sum_i \ln x_{i2}^2
 \end{aligned}$$

De esta forma, se puede concluir que los valores de los parámetros de entrada para la función son los siguientes:

$$\begin{aligned}
 d' &= \left( 2 \sum_i \ln y_i, 2 \sum_i \ln y_i \ln x_{i1}, 2 \sum_i \ln y_i \ln x_{i2} \right) \\
 D &= \begin{pmatrix} 2 \cdot I & 2 \sum_i \ln x_{i1} & 2 \sum_i \ln x_{i2} \\ 2 \sum_i \ln x_{i1} & 2 \sum_i (\ln x_{i1})^2 & 2 \sum_i \ln x_{i1} \ln x_{i2} \\ j = 2 \sum_i \ln x_{i2} & \sum_i \ln x_{i1} \ln x_{i2} & 2 \sum_i (\ln x_{i2})^2 \end{pmatrix}
 \end{aligned}$$

Las restricciones son:

$$\beta_0 + \beta_1 \ln x_{i1} + \beta_2 \ln x_{i2} \geq \ln y_i \quad i = 1, \dots, I$$

que ya están de la forma que se necesita para resolverlas, por lo que el vector  $b_0 = y$ , con  $y$  el vector de logaritmos de las salidas de cada productor. La matriz

$$A' = \begin{pmatrix} 1 & \ln x_{1,1} & \ln x_{1,2} \\ \vdots & \vdots & \vdots \\ 1 & \ln x_{60,1} & \ln x_{60,2} \end{pmatrix}$$

Como ya están todos los vectores y matrices comparados del problema original con la forma que tiene que tener el problema para resolverlo en  $\mathbb{R}$ , ya se puede calcular:

```

> library(quadprog)
> I <- nrow(f)
> d <- c(2, 2*sum(x1), 2*sum(x2))
> D1 <- c(2*I, 2*sum(x1), 2*sum(x2))
> D2 <- c(2*sum(x1), 2*sum(x1^2), 2*sum(x1*x2))

```

```

> D3 <- c(2*sum(x2), 2*sum(x1*x2), 2*sum(x2^2))
> D <- cbind(D1, D2, D3)
> b_0 <- y
> A <- rbind(1, x1, x2)
> QP <- solve.QP(D, d, A, b_0)

```

Recordar que al describir la base de datos, se abrevió  $y := \ln y$ ,  $x_1 := \ln x_1$  y  $x_2 := \ln x_2$ . El número de productores es el parámetro  $I$  y la función para resolver el problema es `solve.QP`. A continuación se muestran los valores de interés, es decir, las componentes estimadas del vector  $\vec{\beta}$ :

```

> beta <- QP$solution
> beta

[1] 0.2166105 0.4662259 0.6852786

```

Las holguras de cada restricción son los  $u_i$ , que la función no los calcula, por lo que se sacaran como:

$$u_i = \beta_0 + \beta_1 \ln x_{i1} + \beta_2 \ln x_{i2} - \ln y_i = A' \cdot \vec{\beta} - \vec{b}_0$$

En R se haría como sigue:

```

> restricciones <- t(A)%*%beta - b_0
> restricciones[1:10,]

[1] 1.1533778 0.7219314 1.0202445 0.8296060 1.0202082 0.6195817 0.8231473
[8] 1.0224811 0.4440999 0.6480622

```

En 'restricciones' están los valores de las 60 restricciones del problema, para abreviar, se muestran solamente las 10 primeras. El último paso es calcular la eficiencia técnica de todos los productores con la expresión  $TE_i = \exp\{-u_i\}$ :

```

> TE2 <- exp(-restricciones[,1]); TE2

```

```

[1] 0.3155691 0.4858130 0.3605068 0.4362211 0.3605199 0.5381695 0.4390477
[8] 0.3597014 0.6414013 0.5230584 0.3680984 1.0000000 0.1814117 0.3240079
[15] 0.9537863 0.2978997 0.3433427 0.4442395 0.6042778 0.4838818 0.4062809
[22] 0.7361606 0.5308606 0.6888414 0.3927837 0.6715827 0.4262823 0.4589890
[29] 0.7849319 0.4507818 0.3885825 1.0000000 0.7686230 0.1685969 0.1501772
[36] 0.6099808 0.4526023 0.3317396 0.3902408 0.5831097 0.4370800 0.3284715
[43] 0.4390552 0.6556938 0.3838842 0.4202454 0.4764655 0.4350162 0.3536106
[50] 0.2284272 0.5712764 0.4894400 0.4682380 0.4676979 0.5497414 0.3940833
[57] 1.0000000 0.2843091 0.5038161 0.3784635

```

El vector 'TE2' está formado por las eficiencias técnicas de los 60 productores de la base de datos. Se finalizará el ejemplo con un resumen de la variable:

```
> summary(TE2)
```

```

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.1502  0.3662  0.4416  0.4853  0.5551  1.0000

```

Se puede observar que la media de la eficiencia técnica de los productores es baja, puesto que su valor es de 0,4853. Sin embargo, algún productor es 100 % técnicamente eficiente porque el valor máximo que toma la variable es 1.

### 3.2. Mínimos cuadrados ordinarios corregidos (COLS)

Este modelo de estimación de frontera determinística es un método en dos etapas para estimar los parámetros. Se describen a continuación:

1. Mínimos cuadrados ordinarios (OLS): con este paso se obtendrán estimaciones consistentes e insesgadas de los parámetros y una estimación consistente, pero sesgada, del intercepto. Para ello hay que minimizar el cuadrado del error, es decir:

$$\vec{u} = \ln \vec{y} - \ln \vec{\hat{y}} = \ln \vec{y} - \left[ \beta_0 + \sum_n \beta_n \ln x_{ni} \right]$$

se denotará  $\vec{y} = (\ln X)\vec{\beta}$  y se tiene que:

$$\|\vec{u}\|^2 = \sum_{i=1}^I u_i^2 = \left( \ln \vec{y} - (\ln X)\vec{\beta} \right)^t \cdot \left( \ln \vec{y} - (\ln X)\vec{\beta} \right) = (\ln \vec{y})^t \ln \vec{y} - 2(\ln \vec{y})^t (\ln X)\vec{\beta} + \vec{\beta}^t (\ln X)^t (\ln X)\vec{\beta}$$

Hay que minimizar:

$$\min_{\vec{\beta}} \|\vec{u}\|^2 = \min_{\vec{\beta}} \left( (\ln \vec{y})^t \ln \vec{y} - 2(\ln \vec{y})^t (\ln X)\vec{\beta} + \vec{\beta}^t (\ln X)^t (\ln X)\vec{\beta} \right)$$

para hallar la estimación de  $\vec{\beta}$  se deriva respecto del vector y se iguala a cero:

$$\begin{aligned} \frac{\partial(\|\vec{u}\|^2)}{\partial \vec{\beta}} = 0 &\iff -2(\ln \vec{y})^t \ln X + 2(\ln X)^t (\ln X)\vec{\beta} = 0 \iff (\ln X)^t (\ln X)\vec{\beta} = (\ln \vec{y})^t \ln X \\ &\iff (\ln X)^t (\ln X)\vec{\beta} = (\ln X)^t \ln \vec{y} \iff \vec{\beta} = ((\ln X)^t \ln X)^{-1} (\ln X)^t \ln \vec{y} \end{aligned}$$

2. El segundo paso consiste en corregir el intercepto  $\beta_0$  para asegurar que la frontera estimada limita los datos desde arriba para la regresión. Se puede estimar consistentemente como:

$$\hat{\beta}_0^* = \hat{\beta}_0 + \max_i \{\hat{u}_i\}$$

donde  $\hat{u}_i$  son los residuos de los mínimos cuadrados ordinarios y pueden ser mayores o menores que cero, es decir, no están restringidos. Estos residuos son corregidos en la dirección contraria como:

$$\hat{u}_i^* = \max_i \{\hat{u}_i\} - \hat{u}_i$$

Los residuos  $\hat{u}_i^*$  de los mínimos cuadrados ordinarios corregidos son no negativos, con al menos uno de ellos siendo cero, por lo que generan una frontera estimada que supone el límite superior de las salidas de los productores analizados. Los residuos se pueden usar para proporcionar estimaciones consistentes de la eficiencia técnica de cada productor como:

$$TE_i = \exp\{-\hat{u}_i^*\}$$

Esta técnica es fácil de implementar, sin embargo la estimación de la frontera de producción es paralela a la regresión mínimo cuadrática porque solo se corrige el intercepto

de ésta. Lo que implica que la estructura de la mejor producción es la misma que la tendencia central de producción. Esto es una propiedad restrictiva del método puesto que la mejor producción debería poder diferenciarse de la producción media de los datos, donde los productores son menos eficientes que los productores de mejores prácticas.

### 3.2.1. Ejemplo

La primera parte de este modelo consiste en realizar una regresión lineal simple para obtener la estimación del vector  $\vec{\beta}$ . Para ello se utiliza la función `lm` que es para resolver el modelo lineal y después se hace un resumen de los resultados obtenidos.

```
> COLS <- lm(y ~ x1+x2, data=f)
> summary(COLS)
```

Call:

```
lm(formula = y ~ x1 + x2, data = f)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.00298	-0.16854	0.00741	0.23992	0.83999

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.24490	0.21360	1.147	0.256
x1	0.28049	0.04807	5.835	2.67e-07 ***
x2	0.53331	0.05150	10.356	1.03e-14 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3376 on 57 degrees of freedom

Multiple R-squared: 0.7079, Adjusted R-squared: 0.6976

F-statistic: 69.06 on 2 and 57 DF, p-value: 5.866e-16

Para comprobar la calidad del modelo, hay que fijarse en el valor 'Adjusted R-squared'. Este valor es aproximadamente 70% y se puede considerar que el modelo es válido porque los p-valores de los contrastes de hipótesis sobre si los coeficientes  $\beta_i$  son cero, son muy altos, excepto en el caso del intercepto, que es justamente el valor que se modifica en el paso dos de este método.

Para obtener los  $u_i$ , hay que extraer los residuos de los mínimos cuadrados y guardarlos en una variable como puede ser 'residuosLM':

```
> residuosLM <- as.vector(COLS$residuals)
```

Para corregir el intercepto, se necesita calcular el máximo de los residuos  $u_i$  y sumarlo al  $\beta_0$  obtenido con los mínimos cuadrados ordinarios:

```
> max <- max(residuosLM); max
```

```
[1] 0.8399876
```

```
> beta0 <- COLS$coefficients[1] + max; beta0
```

```
(Intercept)
```

```
1.084886
```

Ahora falta corregir las estimaciones  $u_i$  como la diferencia entre el máximo valor de los residuos menos cada uno de ellos, es decir:

```
> residuosCOLS <- max - residuosLM
```

A continuación se halla la eficiencia técnica de cada productor con la exponencial de los valores anteriores cambiados de signo, y para entender mejor la variable, se hace un resumen al final de la misma:

```
> TE3 <- exp(-residuosCOLS); TE3
```

```

[1] 0.3449803 0.5250213 0.4055174 0.4239794 0.3750931 0.4545202 0.2875352
[8] 0.4155921 0.5502369 0.4345357 0.2733792 1.0000000 0.2149564 0.3219708
[15] 0.6197043 0.2774961 0.3932397 0.4358795 0.5790018 0.4910512 0.4037138
[22] 0.6145054 0.5387572 0.4276308 0.2649474 0.7023465 0.4967142 0.4599687
[29] 0.5741450 0.3238467 0.2946093 0.4212522 0.6436825 0.1986321 0.1583470
[36] 0.6697366 0.5471171 0.3658213 0.3615738 0.5393054 0.4353157 0.3684546
[43] 0.4888798 0.6588312 0.4206405 0.4040556 0.5713744 0.4633394 0.3608874
[50] 0.2551151 0.5482843 0.5656015 0.5853864 0.4310753 0.5054529 0.4367685
[57] 0.6118764 0.3561896 0.5778697 0.3893028

```

```
> summary(TE3)
```

```

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.1583  0.3648  0.4349  0.4544  0.5488  1.0000

```

En vista de los resultados obtenidos, con este modelo la media de la eficiencia técnica es 0,4544 y al menos uno de los productores es 100 % técnicamente eficiente porque el valor máximo que toma la variable es 1.

### 3.3. Mínimos cuadrados ordinarios modificados (MOLS)

Este método consiste en estimar el modelo de frontera de producción con mínimos cuadrados ordinarios, bajo la suposición de que las perturbaciones siguen una distribución unilateral, por ejemplo, distribución exponencial o seminormal. Estas suposiciones implican que niveles mayores de ineficiencia técnica son cada vez menos probables.

La estimación consiste en dos etapas. La primera será mínimos cuadrados ordinarios, ya descrito en el apartado anterior, para calcular el vector  $\vec{\beta}$  de parámetros. El segundo paso requiere establecer un supuesto de la distribución de  $u_i$ , para modificar el intercepto estimado por la media de la distribución unilateral asumida para la ineficiencia y también las estimaciones de las  $u_i$ , es decir:

$$\hat{\beta}_0^{**} = \hat{\beta}_0 + E(\hat{u}_i)$$



$$\hat{u}_i^{**} = E(\hat{u}_i) - \hat{u}_i$$

Este método también es fácil de implementar, pero no garantiza que la modificación del intercepto estimado por mínimos cuadrados ordinarios lo desplace lo suficiente para garantizar que todos los productores estén acotados desde arriba por la frontera de producción estimada, ya que si un productor tiene un gran residuo de mínimos cuadrados ordinarios, se puede dar que  $(\hat{u}_i - E(\hat{u}_i)) > 0$  para ese productor. En este caso, el productor estaría en una puntuación de eficiencia técnica mayor que la unidad. Por otra parte, también es posible que MOLS desplace el intercepto estimado tanto que ningún productor resulte técnicamente eficiente. Al igual que en el método COLS, la frontera de producción MOLS es paralela a la regresión por mínimos cuadrados ordinarios, ya que solo es modificado el intercepto.

### 3.3.1. Ejemplo

La primera parte de este modelo es análoga a la de COLS, es decir, realizar una regresión lineal simple para obtener la estimación del vector  $\vec{\beta}$ :

```
> MOLS <- lm(y ~ x1+x2, data=f)
> MOLS$coefficients

(Intercept)          x1          x2
  0.2448982   0.2804926   0.5333064
```

Los valores  $u_i$  hay que extraerlos de los residuos, como se hizo anteriormente, y guardarlos en una variable, que se llamará 'residuosLM':

```
> residuosLM <- as.vector(MOLS$residuals)
```

Para corregir el intercepto, se necesita calcular la esperanza de los  $u_i$  y sumarla al valor inicial del intercepto:

El cálculo de la esperanza depende de la distribución que se asume para el término  $u_i$ , se tiene que  $E(u_i) = \sqrt{2/\pi}\sigma_u$  cuando  $u$  sigue una normal, y que  $E(u_i) = \sigma_u$  cuando  $u$  sigue una exponencial. La estimación de  $\sigma_u$  se hace como  $\hat{u}_i = \frac{1}{n} \sum_{i=1}^n (u_i - \bar{u}_i)$ .

```

> residuosLM

[1] -0.224280369  0.195671251 -0.062603958 -0.018082813 -0.140593381
[6]  0.051474821 -0.406422278 -0.038063296  0.242581372  0.006510456
[11] -0.456907694  0.839987649 -0.697332491 -0.293306714  0.361474853
[16] -0.441960605 -0.093348378  0.009598115  0.293537972  0.128780747
[21] -0.067061311  0.353050013  0.221497348 -0.009507403 -0.488236346
[26]  0.486659182  0.140247248  0.063390717  0.285114384 -0.287497433
[31] -0.382117534 -0.024535896  0.399438034 -0.776313066 -1.002978624
[36]  0.439116822  0.236895167 -0.165622575 -0.177301462  0.222514476
[41]  0.008303982 -0.158450160  0.124349118  0.422699651 -0.025989116
[46] -0.066215184  0.280277136  0.070692146 -0.179201662 -0.526052901
[51]  0.239026310  0.270122063  0.304504579 -0.001484936  0.157687253
[56]  0.011635773  0.348762746 -0.192304384  0.291580741 -0.103410152

> media <- mean(residuosLM); media

[1] 1.498006e-17

> sigma <- sd(residuosLM); sigma

[1] 0.3318449

> esperanza1 <- sqrt(2/pi)*sigma; esperanza1

[1] 0.2647739

> esperanza2 <- sigma; esperanza2

[1] 0.3318449

> beta0 <- MOLS$coefficients[1] + esperanza1; beta0

(Intercept)
0.5096721

```

Ahora que ya se han estimado todas las componentes del vector  $\vec{\beta}$ , falta corregir las estimaciones  $u_i$  como se había descrito en la teoría, es decir, como la diferencia entre la esperanza de la variable con cada  $u_i$ . Por último, se halla la eficiencia técnica de cada productor como la exponencial de los residuos corregidos y cambiados de signo.

```
> residuosMOLS <- esperanza1 - residuosLM;
> TE4 <- exp(-residuosMOLS); TE4

[1] 0.6132060 0.9332309 0.7208113 0.7536278 0.6667319 0.8079145 0.5110968
[8] 0.7387193 0.9780519 0.7723917 0.4859344 1.7775104 0.3820872 0.5723065
[15] 1.1015309 0.4932523 0.6989876 0.7747803 1.0291817 0.8728486 0.7176056
[22] 1.0922897 0.9576465 0.7601182 0.4709467 1.2484281 0.8829147 0.8175991
[29] 1.0205487 0.5756408 0.5236711 0.7487802 1.1441524 0.3530707 0.2814635
[36] 1.1904637 0.9725063 0.6502512 0.6427012 0.9586210 0.7737783 0.6549319
[43] 0.8689890 1.1710792 0.7476928 0.7182130 1.0156240 0.8235906 0.6414811
[50] 0.4534697 0.9745810 1.0053625 1.0405305 0.7662408 0.8984478 0.7763606
[57] 1.0876167 0.6331308 1.0271694 0.6919898
```

Para terminar, se hace un resumen de la eficiencia técnica:

```
> summary(TE4)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.2815  0.6484  0.7731  0.8077  0.9754  1.7780
```

En este caso la media es 0,807 y el valor máximo es 1,778, los cuales están por encima de la unidad, como se había descrito que puede ocurrir con este modelo. Este es el problema del método, que no garantiza que todos los productores estén acotados superiormente por la frontera de producción estimada.

### 3.4. Comparación

En este apartado se hará un breve comentario acerca de las desventajas que tienen los métodos. El principal problema de la programación matemática es que los parámetros son

calculados en vez de estimados, aunque se haya citado que estos modelos tienen una equivalencia estadística si se impone una distribución del error. El inconveniente de los mínimos cuadrados corregidos es que la frontera estimada es paralela a la obtenida por la regresión de los mínimos cuadrados ordinarios, puesto que solo se corrige el término independiente. De esta forma, la mejor producción será la misma que la tendencia central de producción y esto restringe el método, ya que se debería poder diferenciar la mejor producción de la producción media de los productores. Por último, el método de mínimos cuadrados ordinarios modificados no garantiza que todas las producciones queden limitadas superiormente por la frontera de producción, es decir, si un productor tiene un residuo de mínimos cuadrados ordinarios grande, es posible que la eficiencia técnica sea mayor que la unidad. Además, al igual que antes, la frontera de producción de los mínimos cuadrados modificados es paralela a la de la regresión con mínimos cuadrados ordinarios y no permite que haya diferencia entre la mejor producción y la producción media de todos los productores.

Las tres técnicas descritas son fáciles de implementar. A parte de lo comentando en el párrafo anterior, tienen el inconveniente común de que cada una de ellas mide la eficiencia técnica con una frontera de producción determinística. Es decir, toda la variación en la producción, no asociada con la variación en las entradas, se atribuye a la ineficiencia técnica, ninguna de estas técnicas permite el efecto de perturbaciones aleatorias, las cuales también podrían contribuir a la variación de las salidas. Por tanto, lo que se necesita es un modelo que atribuya las variaciones en las producciones, no asociadas con variaciones en las entradas, como combinación de perturbaciones aleatorias e ineficiencia técnica. Este tipo de modelo es más complejo, pero también más realista y se estudiará en el capítulo siguiente.

## Capítulo 4

# Frontera de producción estocástica

El modelo de frontera estocástica supone una modificación de la frontera determinística que permite reconocer la influencia de perturbaciones aleatorias en el análisis. Se puede escribir como:

$$y_i = f(\vec{x}_i, \vec{\beta}) \cdot \exp\{v_i\} \cdot TE_i$$

Considerando que  $TE_i = \exp\{-u_i\}$  y  $f(\vec{x}_i, \vec{\beta}) = \Pi_n x_{ni}^{\beta_n}$  se tiene que:

$$y_i = \Pi_n x_{ni}^{\beta_n} \cdot \exp\{v_i\} \cdot \exp\{-u_i\}$$

Tomando logaritmos en la ecuación anterior, el modelo quedaría como:

$$\ln y_i = \beta_0 + \sum_n \beta_n \ln x_{ni} + v_i - u_i$$

donde  $v_i$  es la componente ruido y  $u_i$  es la componente no negativa que mide la ineficiencia técnica.

El modelo de frontera estocástico se conoce como el modelo de error compuesto, ya que presenta un término de error con dos componentes,  $\epsilon_i = v_i - u_i$ , donde una de ellas recoge las deficiencias en producción debidas a la ineficiencia técnica,  $u_i$ , y la otra, perturbaciones aleatorias de cada productor que pueden a afectar la producción,  $v_i$ .

El objetivo de estas fronteras es obtener estimaciones de los parámetros, así como de la eficiencia técnica de cada productor.

## 4.1. Método de máxima verosimilitud

En esta sección se empezará por la discusión del método de máxima verosimilitud que puede ser utilizado para estimar los parámetros de una frontera de producción estocástica y de las eficiencias técnicas. Para conseguir lo segundo, hay que realizar una estimación separada del error aleatorio,  $v_i$ , y la ineficiencia técnica,  $u_i$ , a partir de las estimaciones de  $\epsilon_i$  para cada productor. Esto implica suponer distribuciones de las componentes del error. La distribución de  $v_i$  siempre se asume normal, pero para la distribución de  $u_i$  hay varias propuestas. Por tanto, se estudiarán dos casos, en función de qué distribución se considera que sigue la ineficiencia técnica.

### 4.1.1. Modelo normal - seminormal

En el modelo de frontera estocástica de producción

$$\ln y_i = \beta_0 + \sum_n \beta_n \ln x_{ni} + v_i - u_i$$

se asume que:

- $v_i \rightsquigarrow N(0, \sigma_v^2)$ , independiente e idénticamente distribuida, es decir:

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v} \cdot \exp \left\{ -\frac{v^2}{2\sigma_v^2} \right\}$$

- $u_i \rightsquigarrow N^+(0, \sigma_u^2)$ , distribución normal no negativa, es decir:

$$f(u) = \frac{2}{\sqrt{2\pi}\sigma_u} \cdot \exp \left\{ -\frac{u^2}{2\sigma_u^2} \right\}$$

- $v_i$  y  $u_i$  son distribuidas independientemente entre ellas, y de los regresores.

El primer punto se asume por consenso y se mantiene siempre. El segundo está basado en la proposición de que un valor de la ineficiencia técnica es cero, con valores crecientes de la ineficiencia técnica cada vez menos probables. La primera parte del punto tres parece sencilla, pero la segunda parte es más problemática puesto que si los productores supieran algo sobre su eficiencia técnica, esto influiría en la elección de sus entradas.

Asumida la independencia entre  $v$  y  $u$ , la función de densidad conjunta será el producto de sus funciones de densidad individuales, es decir

$$\begin{aligned} f(u, v) &= f(u) \cdot f(v) = \left( \frac{1}{\sqrt{2\pi}\sigma_u} \cdot \exp \left\{ -\frac{u^2}{2\sigma_u^2} \right\} \right) \cdot \left( \frac{1}{\sqrt{2\pi}\sigma_v} \cdot \exp \left\{ -\frac{v^2}{2\sigma_v^2} \right\} \right) \\ &= \frac{2}{2\pi\sigma_u\sigma_v} \cdot \exp \left\{ -\frac{u^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2} \right\} \end{aligned}$$

Como el término de error es la diferencia  $\epsilon = v - u$ , se tiene que  $v = \epsilon + u$ , por lo que la función de densidad para  $u$  y  $\epsilon$  es:

$$f(u, \epsilon) = \frac{2}{2\pi\sigma_u\sigma_v} \cdot \exp \left\{ -\frac{u^2}{2\sigma_u^2} - \frac{(\epsilon + u)^2}{2\sigma_v^2} \right\}$$

El siguiente paso que interesa es calcular la función de densidad marginal de  $\epsilon$ . Ésta se obtiene integrando respecto de  $u$  la función  $f(u, \epsilon)$ , es decir:

$$\begin{aligned} f(\epsilon) &= \int_0^\infty f(u, \epsilon) du = \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v} \cdot \exp \left\{ -\frac{u^2}{2\sigma_u^2} - \frac{(\epsilon + u)^2}{2\sigma_v^2} \right\} du = \\ &= \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v} \cdot \exp \left\{ -\frac{u^2}{2\sigma_u^2} - \frac{\epsilon^2 + u^2 + 2u\epsilon}{2\sigma_v^2} \right\} du = \\ &= \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v} \cdot \exp \left\{ -\frac{1}{2} \frac{\epsilon^2}{\sigma_v^2} - \frac{1}{2} \left[ \frac{\sigma_v^2 u^2 + \sigma_u^2 u^2 + 2u\epsilon\sigma_u^2}{\sigma_u^2 \sigma_v^2} \right] \right\} du = \\ &= \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v} \cdot \exp \left\{ -\frac{1}{2} \frac{\epsilon^2}{\sigma_v^2} \right\} \cdot \exp \left\{ -\frac{1}{2} \left[ \frac{\sigma_v^2 u^2 + \sigma_u^2 u^2 + 2u\epsilon\sigma_u^2}{\sigma_u^2 \sigma_v^2} \right] \right\} du = \\ &= \exp \left\{ -\frac{1}{2} \frac{\epsilon^2}{\sigma_v^2} \right\} \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v} \cdot \exp \left\{ -\frac{1}{2} \left[ \frac{(\sigma_v^2 + \sigma_u^2)u^2 + 2u\epsilon\sigma_u^2}{\sigma_u^2 \sigma_v^2} \right] \right\} du = \\ &= \exp \left\{ -\frac{1}{2} \frac{\epsilon^2}{\sigma_v^2} \right\} \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v} \cdot \exp \left\{ -\frac{1}{2} \frac{\sigma_v^2 + \sigma_u^2}{\sigma_v^2 \sigma_u^2} \left[ u^2 + 2u\epsilon \frac{\sigma_u^2}{\sigma_v^2 + \sigma_u^2} \right] \right\} du = \\ &= \exp \left\{ -\frac{1}{2} \frac{\epsilon^2}{\sigma_v^2} \right\} \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v} \cdot \exp \left\{ -\frac{1}{2} \frac{\sigma_v^2 + \sigma_u^2}{\sigma_v^2 \sigma_u^2} \left[ \left( u + \epsilon \frac{\sigma_u^2}{\sigma_v^2 + \sigma_u^2} \right)^2 - \left( \frac{\sigma_u^2}{\sigma_v^2 + \sigma_u^2} \right)^2 \epsilon^2 \right] \right\} du = \\ &= \exp \left\{ -\frac{1}{2} \frac{\epsilon^2}{\sigma_v^2} \right\} \int_0^\infty \frac{2}{2\pi\sigma_u\sigma_v} \cdot \exp \left\{ \frac{1}{2} \frac{\sigma_v^2 + \sigma_u^2}{\sigma_v^2 \sigma_u^2} \cdot \frac{(\sigma_u^2)^2}{(\sigma_v^2 + \sigma_u^2)^2} \epsilon^2 \right\} \cdot \\ &\quad \cdot \exp \left\{ -\frac{1}{2} \frac{\sigma_v^2 + \sigma_u^2}{\sigma_v^2 \sigma_u^2} \left[ \left( u + \epsilon \frac{\sigma_u^2}{\sigma_v^2 + \sigma_u^2} \right)^2 \right] \right\} du \end{aligned}$$

Como la primera exponencial que hay en la integral no depende de la variable de integración  $u$ , se puede sacar de la integral también. Simplificando términos, sacando alguna de las constantes fuera de la integral y teniendo en cuenta que  $\sigma^2 = \sigma_u^2 + \sigma_v^2$ , se tiene:

$$\frac{2}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{\epsilon^2}{\sigma_v^2} + \frac{1}{2} \frac{\sigma_u^2}{\sigma_v^2} \cdot \frac{1}{\sigma^2} \epsilon^2 \right\} \int_0^\infty \frac{1}{\sqrt{2\pi} \sigma_u \sigma_v} \cdot \exp \left\{ -\frac{1}{2} \frac{\sigma_v^2 + \sigma_u^2}{\sigma_v^2 \sigma_u^2} \left[ \left( u + \epsilon \frac{\sigma_u^2}{\sigma_v^2 + \sigma_u^2} \right)^2 \right] \right\} du$$

Reorganizando los términos de la exponencial que está en la integral, y dejando a parte todo lo que está fuera de la integral, se llega a lo siguiente:

$$\int_0^\infty \frac{1}{\sqrt{2\pi} \sigma_u \sigma_v} \cdot \exp \left\{ -\frac{1}{2} \left[ \left( \frac{u - \frac{-\sigma_u^2}{\sigma_v^2 + \sigma_u^2} \epsilon}{\sqrt{\frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2}}} \right)^2 \right] \right\} du$$

Para tener la función de densidad de una normal en la integral, y que sea así una función de distribución, es necesario hacer el siguiente cambio de variable:

$$t = \frac{u - \frac{-\sigma_u^2}{\sigma_v^2 + \sigma_u^2} \epsilon}{\sqrt{\frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2}}}$$

Derivando la expresión anterior, se tiene que:

$$dt = \frac{1}{\sqrt{\frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2}}} du \Rightarrow du = \frac{\sigma_u \sigma_v}{\sqrt{\sigma_u^2 + \sigma_v^2}} dt = \frac{\sigma_u \sigma_v}{\sigma} dt$$

El último paso es cambiar los extremos de integración:

■ Extremo  $u = 0$  :

$$t = \frac{-\frac{\sigma_u^2}{\sigma_v^2 + \sigma_u^2} \epsilon}{\sqrt{\frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2}}} = -\frac{\sigma_u}{\sigma_v} \frac{\epsilon}{\sqrt{\sigma_v^2 + \sigma_u^2}} = -\frac{\lambda \epsilon}{\sigma}$$

■ Extremo  $u = \infty \Rightarrow t = \infty$

Por tanto, con el cambio de variable descrito, la integral anterior quedaría como:

$$\int_{-\frac{\lambda \epsilon}{\sigma}}^\infty \frac{1}{\sqrt{2\pi} \sigma_u \sigma_v} \cdot \exp \left\{ -\frac{1}{2} t^2 \right\} \cdot \frac{\sigma_u \sigma_v}{\sigma} dt$$

Simplificando términos y sacando de la integral el término  $\frac{1}{\sigma}$ , que no se utiliza para hallar la integral, se tiene:

$$\int_{-\frac{\lambda \epsilon}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}} \cdot \exp \left\{ -\frac{1}{2} t^2 \right\} dt = 1 - \Phi \left( \frac{\lambda \epsilon}{\sigma} \right)$$



Ahora hay que recuperar todos los términos que no dependían de la variable de integración y se han sacado de la integral:

$$\begin{aligned} & \frac{2}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{1}{2} \frac{\epsilon^2}{\sigma_v^2} + \frac{1}{2} \frac{\sigma_u^2}{\sigma_v^2} \cdot \frac{1}{\sigma^2} \epsilon^2 \right\} = \frac{2}{\sqrt{2\pi\sigma}} \exp \left\{ \frac{\epsilon^2}{2\sigma_v^2} \left( -1 + \frac{\sigma_u^2}{\sigma^2} \right) \right\} = \\ & = \frac{2}{\sqrt{2\pi\sigma}} \exp \left\{ \frac{\epsilon^2}{2\sigma_v^2} \left( \frac{-\sigma^2 + \sigma_u^2}{\sigma^2} \right) \right\} = \frac{2}{\sqrt{2\pi\sigma}} \exp \left\{ \frac{\epsilon^2}{2\sigma_v^2} \left( \frac{-\sigma_u^2 - \sigma_v^2 + \sigma_u^2}{\sigma^2} \right) \right\} = \\ & = \frac{2}{\sqrt{2\pi\sigma}} \exp \left\{ \frac{\epsilon^2}{2\sigma_v^2} \left( \frac{-\sigma_v^2}{\sigma^2} \right) \right\} = \frac{2}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{\epsilon^2}{2\sigma^2} \right\} \end{aligned}$$

Por último, se juntan los resultados obtenidos de ambas partes para obtener el valor de la función marginal:

$$f(\epsilon) = \int_0^\infty f(u, \epsilon) du = \frac{2}{\sqrt{2\pi\sigma}} \cdot \exp \left\{ -\frac{\epsilon^2}{2\sigma^2} \right\} \cdot \left[ 1 - \Phi \left( \frac{\epsilon\lambda}{\sigma} \right) \right] = \frac{2}{\sigma} \cdot \phi \left( \frac{\epsilon}{\sigma} \right) \cdot \Phi \left( -\frac{\epsilon\lambda}{\sigma} \right)$$

donde  $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$ ,  $\lambda = \sigma_u/\sigma_v$  y  $\Phi(\cdot)$  y  $\phi(\cdot)$  son la función de distribución y de densidad de una normal estándar, respectivamente. La reparametrización de  $\sigma_u^2$  y  $\sigma_v^2$  a  $\sigma$  y  $\lambda$  es conveniente, puesto que  $\lambda$  proporciona un indicador de las contribuciones relativas de  $u$  y  $v$  en la determinación de la perturbación estocástica,  $\epsilon$ . Cuando  $\lambda \rightarrow 0$  implica que  $\sigma_v^2 \rightarrow +\infty$  o  $\sigma_u^2 \rightarrow 0$ , y la componente simétrica del error domina la componente unilateral del error en la determinación de  $\epsilon$ . Por otra parte, si  $\lambda \rightarrow +\infty$  implica que  $\sigma_u^2 \rightarrow +\infty$  o  $\sigma_v^2 \rightarrow 0$ , y la componente unilateral domina la componente simétrica del error en la determinación de  $\epsilon$ . En el primer caso, se volvería a un modelo de producción sin ineficiencia técnica, mientras que en el segundo se tendría un modelo de producción sin ruido.

La función de densidad marginal de  $f(\epsilon)$  es asimétricamente distribuida, con media y varianza:

$$\begin{aligned} E(\epsilon) &= -E(u) = -\sigma_u \sqrt{\frac{2}{\pi}} \\ Var(\epsilon) &= \frac{\pi - 2}{\pi} \sigma_u^2 + \sigma_v^2 \end{aligned}$$

Usando la función de densidad marginal de  $\epsilon$ , el logaritmo de la función de verosimilitud para la muestra de  $I$  productores es:

$$\ln L = \ln \left( \prod_i \left[ \frac{2}{\sqrt{2\pi\sigma}} \left[ 1 - \Phi \left( \frac{\epsilon_i \lambda}{\sigma} \right) \right] \cdot \exp \left\{ -\frac{\epsilon_i^2}{2\sigma^2} \right\} \right] \right) =$$

$$\begin{aligned}
&= \sum_i \left( \ln \left[ \frac{2}{\sqrt{2\pi}\sigma} \left[ 1 - \Phi \left( \frac{\epsilon_i \lambda}{\sigma} \right) \right] \cdot \exp \left\{ -\frac{\epsilon_i^2}{2\sigma^2} \right\} \right] \right) \\
&= \sum_i \left( \ln \left[ \frac{2}{\sqrt{2\pi}\sigma} \left[ \Phi \left( -\frac{\epsilon_i \lambda}{\sigma} \right) \right] \cdot \exp \left\{ -\frac{\epsilon_i^2}{2\sigma^2} \right\} \right] \right) \\
&= \sum_i \left( \ln \frac{\sqrt{2}}{\sqrt{\pi}} + \ln \frac{1}{\sigma} + \ln \Phi \left( -\frac{\epsilon_i \lambda}{\sigma} \right) + \ln \exp \left\{ -\frac{\epsilon_i^2}{2\sigma^2} \right\} \right) \\
&= -\frac{I}{2} \ln \frac{\pi}{2} - I \ln \sigma + \sum_{i=1}^I \ln \Phi \left( -\frac{\epsilon_i \lambda}{\sigma} \right) - \frac{1}{2\sigma^2} \sum_i \epsilon_i^2
\end{aligned}$$

La función anterior puede ser maximizada respecto de los parámetros. Para ello hay que derivar como sigue:

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{I}{2\sigma^2} + \frac{\lambda}{2\sigma^3} \sum_i \frac{\phi \left( \frac{\epsilon_i \lambda}{\sigma} \right)}{1 - \Phi \left( \frac{\epsilon_i \lambda}{\sigma} \right)} \cdot \epsilon_i + \frac{1}{2\sigma^4} \sum_i \epsilon_i^2 = 0$$

Teniendo en cuenta que  $\epsilon_i = \ln y_i - \beta_0 - \sum_n \beta_n \ln x_{ni}$  para hacer la derivada de la verosimilitud respecto de  $\beta$ , la expresión quedaría:

$$\frac{\partial \ln L}{\partial \vec{\beta}} = \frac{\lambda}{\sigma} \sum_i \frac{\phi \left( \frac{\epsilon_i \lambda}{\sigma} \right)}{1 - \Phi \left( \frac{\epsilon_i \lambda}{\sigma} \right)} \cdot x_i + \frac{1}{\sigma^2} \sum_i \epsilon_i x_i = 0$$

donde  $x_i$  es un vector formado por el logaritmo de los elementos de la fila  $i$ -ésima de la matriz de datos  $X$ .

En este caso, derivando la expresión e igualando a cero no se puede sacar una expresión analítica de dichos estimadores, luego hay que recurrir a una maximización numérica del logaritmo de la verosimilitud para obtener los estimadores de todos los parámetros.

El siguiente paso es obtener la estimación de la eficiencia técnica de cada productor. Para ello se utilizará la técnica propuesta por James Jondrow, Know Lovell, Ivan Materov y Peter Schmidt. La técnica JLMS consiste en estimar la eficiencia técnica basándose en el cálculo de la media o la moda de la distribución condicionada del término de la ineficiencia técnica por el término de error compuesto de cada unidad productiva  $i$ . La eficiencia técnica es calculada de acuerdo con la expresión  $TE_i = \exp\{-\hat{u}_i\}$ . Por tanto, hay que hallar la distribución condicional de  $u_i$  dado  $\epsilon_i$ :

$$f(u|\epsilon) = \frac{f(u, \epsilon)}{f(\epsilon)} = \frac{\frac{2}{2\pi\sigma_u\sigma_v} \cdot \exp \left\{ -\frac{u^2}{2\sigma_u^2} - \frac{(\epsilon+u)^2}{2\sigma_v^2} \right\}}{\frac{2}{\sqrt{2\pi}\sigma} \left[ 1 - \Phi \left( \frac{\epsilon\lambda}{\sigma} \right) \right] \cdot \exp \left\{ -\frac{\epsilon^2}{2\sigma^2} \right\}} =$$

$$= \frac{1}{\sqrt{2\pi} \frac{\sigma_u \sigma_v}{\sigma}} \cdot \frac{\exp \left\{ -\frac{u^2}{2\sigma_u^2} - \frac{(\epsilon+u)^2}{2\sigma_v^2} + \frac{\epsilon^2}{2\sigma^2} \right\}}{\left[ 1 - \Phi \left( \frac{\epsilon\lambda}{\sigma} \right) \right]}$$

Para trabajar más cómodamente, se harán las cuentas por partes, es decir, primero se 'arregla' el denominador y después se opera en el numerador.

Sabiendo que  $\lambda = \sigma_u/\sigma_v$ , el denominador quedaría como sigue:

$$1 - \Phi \left( \frac{\epsilon\lambda}{\sigma} \right) = 1 - \Phi \left( \frac{\epsilon\sigma_u}{\sigma_v\sigma} \right) = 1 - \Phi \left( \frac{\epsilon\sigma_u\sigma_u\sigma}{\sigma_v\sigma\sigma_u\sigma} \right) = 1 - \Phi \left( \frac{\epsilon\sigma_u^2\sigma}{\sigma^2\sigma_v\sigma_u} \right) = 1 - \Phi \left( \frac{-\mu_*}{\sigma_*} \right)$$

donde  $\mu_* = -\epsilon\sigma_u^2/\sigma^2$  y  $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$ .

La exponencial del numerador quedaría de la siguiente manera:

$$\begin{aligned} \exp \left\{ -\frac{u^2}{2\sigma_u^2} - \frac{(\epsilon+u)^2}{2\sigma_v^2} + \frac{\epsilon^2}{2\sigma^2} \right\} &= \exp \left\{ \frac{-u^2\sigma_v^2\sigma^2 - (\epsilon+u)^2\sigma_u^2\sigma^2 + \epsilon^2\sigma_u^2\sigma_v^2}{2\sigma_u^2\sigma_v^2\sigma^2\sigma^2/\sigma^2} \right\} \\ &= \exp \left\{ \frac{-u^2\sigma_v^2\sigma^2 - (\epsilon+u)^2\sigma_u^2\sigma^2 + \epsilon^2\sigma_u^2\sigma_v^2}{2\sigma_*^2\sigma^4} \right\} \\ &= \exp \left\{ \frac{-u^2\sigma_v^2\sigma^{-2} - (\epsilon+u)^2\sigma_u^2\sigma^{-2} + \epsilon^2\sigma_u^2\sigma_v^2\sigma^{-4}}{2\sigma_*^2} \right\} \end{aligned}$$

Teniendo en cuenta que  $\sigma_v^2 = \sigma^2 - \sigma_u^2$ , se tiene:

$$\begin{aligned} &= \exp \left\{ \frac{-u^2(\sigma^2 - \sigma_u^2)\sigma^{-2} - (\epsilon+u)^2\sigma_u^2\sigma^{-2} + \epsilon^2\sigma_u^2(\sigma^2 - \sigma_u^2)\sigma^{-4}}{2\sigma_*^2} \right\} \\ &= \exp \left\{ \frac{-u^2 + u^2\sigma_u^2\sigma^{-2} - \epsilon^2\sigma_u^2\sigma^{-2} - u^2\sigma_u^2\sigma^{-2} - 2\epsilon u\sigma_u^2\sigma^{-2} + \epsilon^2\sigma_u^2\sigma^{-2} - \epsilon^2\sigma_u^4\sigma^{-4}}{2\sigma_*^2} \right\} \\ &= \exp \left\{ \frac{-u^2 - 2\epsilon u\sigma_u^2\sigma^{-2} - \epsilon^2\sigma_u^4\sigma^{-4}}{2\sigma_*^2} \right\} = \exp \left\{ \frac{-(u + \epsilon\sigma_u^2\sigma^{-2})^2}{2\sigma_*^2} \right\} \\ &= \exp \left\{ \frac{-\left(u + \frac{\epsilon\sigma_u^2}{\sigma^2}\right)^2}{2\sigma_*^2} \right\} = \exp \left\{ \frac{-(u - \mu_*)^2}{2\sigma_*^2} \right\} \end{aligned}$$

Si se juntan las expresiones a las que se ha llegado del numerador y del denominador, se tiene lo siguiente:

$$f(u|\epsilon) = \frac{1}{\sqrt{2\pi}\sigma_*} \cdot \frac{\exp \left\{ -\frac{(u-\mu_*)^2}{2\sigma_*^2} \right\}}{1 - \Phi \left( -\frac{\mu_*}{\sigma_*} \right)}$$

Por tanto, la distribución de  $f(u|\epsilon)$  es  $N^+(\mu_*, \sigma_*^2)$ . Tanto la media como la moda de la distribución pueden servir como estimador puntual de  $u_i$ , aunque es más frecuente utilizar la esperanza, y en los ejemplos que se resolverán será la que se utilice.

$$E(u_i|\epsilon_i) = \mu_{*i} + \sigma_* \left[ \frac{\phi(-\mu_{*i}/\sigma_*)}{1 - \Phi(-\mu_{*i}/\sigma_*)} \right] = \sigma_* \left[ \frac{\phi(\epsilon_i \lambda / (\sigma))}{1 - \Phi(\epsilon_i \lambda / \sigma)} - \left( \frac{\epsilon_i \lambda}{\sigma} \right) \right]$$

$$\text{Moda}(u_i|\epsilon_i) = \begin{cases} \mu_{*i} & \text{si } \epsilon_i \leq 0 \\ 0 & \text{en otro caso} \end{cases}$$

Una vez obtenido el estimador de  $u_i$ , el estimador de la eficiencia técnica de cada productor se obtiene como  $\text{TE}_i = \exp\{-\hat{u}_i\}$ , donde  $\hat{u}_i$  puede ser  $E(u_i|\epsilon_i)$  o  $\text{Moda}(u_i|\epsilon_i)$ . Sea cual sea el estimador utilizado, el estimador de la eficiencia técnica es inconsistente porque la variación asociada en la distribución de  $(u_i, \epsilon_i)$  es independiente de  $i$ . Desafortunadamente esto es lo mejor que se puede conseguir con los datos transversales.

Por último, se describe como obtener intervalos de confianza para los puntos estimados de eficiencia técnica, ya que se conoce la función de densidad de  $(u_i, \epsilon_i)$ . Para ello hace falta hallar el límite inferior y superior,  $(L_i, U_i)$ , de  $(u_i, \epsilon_i)$ , lo que implica el límite inferior y superior de  $(\exp\{-\hat{u}_i\}, \epsilon_i)$ . El intervalo de confianza al  $(1 - \alpha)100\%$  viene dado por:

$$L_i = \exp\{-\hat{u}_i - z_L \cdot \sigma_*\}$$

$$U_i = \exp\{-\hat{u}_i - z_U \cdot \sigma_*\}$$

donde

$$Pr(Z > z_L) = \frac{\alpha}{2} \left[ 1 - \Phi \left( -\frac{\mu_{*i}}{\sigma_*} \right) \right]$$

$$Pr(Z > z_U) = \left( 1 - \frac{\alpha}{2} \right) \left[ 1 - \Phi \left( -\frac{\mu_{*i}}{\sigma_*} \right) \right]$$

siendo  $Z \rightsquigarrow N(0, 1)$ . Como consecuencia, se tiene que los valores  $z_L$  y  $z_U$  se calculan como:

$$z_L = \Phi^{-1} \left\{ \left[ 1 - \frac{\alpha}{2} \right] \left[ 1 - \Phi \left( -\frac{\mu_{*i}}{\sigma_*} \right) \right] \right\}$$

$$z_U = \Phi^{-1} \left\{ \left[ 1 - \left( 1 - \frac{\alpha}{2} \right) \right] \left[ 1 - \Phi \left( -\frac{\mu_{*i}}{\sigma_*} \right) \right] \right\}$$

#### 4.1.2. Ejemplo N - $\mathbf{N}^+$

En este apartado se va a calcular la frontera de producción estocástica de los datos descritos anteriormente con el método de máxima verosimilitud y suponiendo que  $u_i$  sigue un distribución seminormal. Para resolver el problema, hay dos paquetes en R que se podrían utilizar, por lo que el ejercicio se calculará dos veces, una con cada función:

## Paquete sfa

El primer paquete que se utiliza se llama *sfa*. Hay que instalarlo en R y después cargar la librería:

```
> library(sfa)
```

La función que se va utilizar se llama también *sfa*. Recordando que se quiere estimar la producción,  $y$ , en función de las variables capital de la empresa,  $x_1$ , y trabajo de la empresa,  $x_2$ , el problema se resuelve como:

```
> NN_sfa <- sfa( y ~ x1 + x2, data = f, intercept = TRUE, fun = "hnormal",  
+               form = "production")
```

```
> beta <- c(NN_sfa$coef[1], NN_sfa$coef[2], NN_sfa$coef[3]); beta
```

```
Intercept      x1      x2  
0.5614944 0.2811041 0.5364992
```

```
> sigmau2 <- NN_sfa$coef[4]; sigmau2
```

```
sigmau2  
0.1729516
```

```
> sigmav2 <- NN_sfa$coef[5]; sigmav2
```

```
sigmav2  
0.04402588
```

De esta forma se obtiene la estimación máximo verosímil de todos los parámetros. Falta obtener la estimación de la eficiencia técnica de cada productor, para ello hay una función en R que las calcula, llamada *te.eff.sfa*. También se hace un resumen de la variable:

```
> TE5 <- te.eff.sfa(NN_sfa);
```

```
> TE5 <- as.vector(TE5); TE5
```

```

[1] 0.6507508 0.8288934 0.7264613 0.7478879 0.6913911 0.7765662 0.5652339
[8] 0.7377141 0.8438910 0.7578689 0.5456687 0.9373780 0.4481647 0.6183810
[15] 0.8738342 0.5495963 0.7126680 0.7590994 0.8572660 0.8065311 0.7246231
[22] 0.8722295 0.8368121 0.7522943 0.5298157 0.8973007 0.8101419 0.7818156
[29] 0.8561063 0.6210468 0.5794696 0.7494027 0.8819156 0.4208897 0.3513279
[36] 0.8890695 0.8411825 0.6787437 0.6729645 0.8385411 0.7596803 0.6819483
[43] 0.8043969 0.8865194 0.7430228 0.7261458 0.8534083 0.7852160 0.6721263
[50] 0.5143746 0.8423803 0.8509817 0.8596297 0.7551165 0.8165049 0.7599397
[57] 0.8735144 0.6647656 0.8566987 0.7084751

```

```
> summary(TE5)
```

```

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.3513  0.6773  0.7585  0.7406  0.8428  0.9374

```

Con este modelo estocástico, la media de la eficiencia técnica es 0,7585, la cual es bastante elevada. Por otra parte, el valor máximo que toma la variable es 0,9374, lo que indica que ningún productor es 100% técnicamente eficiente.

### Paquete frontier

En este caso el paquete que hay que instalar se llama `frontier`. Sin embargo, antes de cargarlo hay que desconectar el paquete anterior, puesto que la función que se utiliza para resolver el ejercicio se llama también `sfa` y sino da problemas a la hora de ejecutarla.

```

> detach("package:sfa")
> library(frontier)

```

Al igual que antes, se utiliza la función correspondiente de este paquete para resolver el problema, es decir:

```

> NN_frontier <- sfa( y ~ x1 + x2)
> NN_frontier

```

Call:

```
sfa(formula = y ~ x1 + x2)
```

Maximum likelihood estimates

(Intercept)	x1	x2	sigmaSq	gamma
0.5616	0.2811	0.5365	0.2170	0.7972

Así se obtiene la estimación del vector  $\vec{\beta}$ . Las estimaciones de  $\sigma_v^2$  y  $\sigma_u^2$  no las da la función, sino que hay que calcularlas a través de las de  $\sigma^2$  y  $\gamma$ , donde  $\gamma = \sigma_u^2/\sigma^2$  y  $\sigma^2 = \sigma_u^2 + \sigma_v^2$ . Por lo que  $\sigma_u^2 = \sigma^2 \cdot \gamma$ , y posteriormente se calcula  $\sigma_v^2 = \sigma^2 - \sigma_u^2$ :

```
> gamma <- 0.7972
> sigma2 <- 0.2170
> sigmau2 <- sigma2 * gamma; sigmau2

[1] 0.1729924

> sigmav2 <- sigma2 - sigmau2; sigmav2

[1] 0.0440076
```

El último paso consiste en obtener el vector de eficiencias técnicas de cada productor. Para ello se utiliza la función `eficiencias`:

```
> TE6 <- eficiencias(NN_frontier)
> TE6 <- as.vector(TE6); TE6

[1] 0.6506888 0.8288915 0.7264261 0.7478515 0.6913358 0.7765464 0.5651682
[8] 0.7376818 0.8438897 0.7578417 0.5455846 0.9373950 0.4480971 0.6183105
[15] 0.8738435 0.5495280 0.7126254 0.7590723 0.8572720 0.8065194 0.7245865
[22] 0.8722360 0.8368137 0.7522574 0.5297482 0.8973169 0.8101341 0.7817940
[29] 0.8561059 0.6209790 0.5793820 0.7493422 0.8819257 0.4208221 0.3512628
[36] 0.8890839 0.8411862 0.6786893 0.6729107 0.8385341 0.7596459 0.6818962
```

```
[43] 0.8043876 0.8865300 0.7429931 0.7261024 0.8534151 0.7851920 0.6720713
[50] 0.5143029 0.8423813 0.8509858 0.8596385 0.7550831 0.8164983 0.7599126
[57] 0.8735073 0.6647149 0.8567044 0.7084283
```

```
> summary(TE6)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.3513  0.6772  0.7585  0.7406  0.8428  0.9374
```

Como también se ha hecho un resumen de la variable, se observa que la media de las eficiencias técnicas es 0,7406 y que ningún productor es 100 % técnicamente eficiente, puesto que el máximo valor que toma la variable es 0,9374.

### Intervalos de confianza

En teoría se ha descrito la forma de obtener los intervalos de confianza para la eficiencia técnica de cada productor. Esta es una gran ventaja sobre los métodos determinísticos, puesto que con ellos no se pueden calcular.

El intervalo viene dado por:

$$(\exp\{-\hat{u}_i - z_L \cdot \sigma_*\}, \exp\{-\hat{u}_i - z_U \cdot \sigma_*\})$$

donde

$$z_L = \Phi^{-1} \left\{ \left[ 1 - \frac{\alpha}{2} \right] \left[ 1 - \Phi \left( -\frac{\mu_{*i}}{\sigma_*} \right) \right] \right\}$$

$$z_U = \Phi^{-1} \left\{ \left[ 1 - \left( 1 - \frac{\alpha}{2} \right) \right] \left[ 1 - \Phi \left( -\frac{\mu_{*i}}{\sigma_*} \right) \right] \right\}$$

Se calculará en R como sigue:

```
> X <- cbind(1, x1, x2)
> beta <- as.vector(beta)
> epsilon <- y - X%*%beta
> muEstrella <- - epsilon*sigmau2/sigma2
> sigmaEstrella <- sigmau2*sigmav2/sigma2
```



```

> distribucion <- pnorm(-muEstrella/sqrt(sigmaEstrella))
> zL <- qnorm((1 - (0.05/2))*(1-distribucion))
> zU <- qnorm((0.05/2)*(1-distribucion))
> u <- - log(TE6)
> L <- exp(-u - zL*sqrt(sigmaEstrella))
> U <- exp(-u - zU*sqrt(sigmaEstrella))
> intervalos <- cbind(L,U)
> intervalos[1:5,]

          [,1]      [,2]
[1,] 0.4622924 0.9400033
[2,] 0.7512136 1.2278674
[3,] 0.5512830 1.0526336
[4,] 0.5838061 1.0858067
[5,] 0.5055398 0.9998347

```

Para que no sea muy larga la salida, se muestra el intervalo de confianza para la eficiencia técnica de los cinco primeros productores con un nivel de confianza del 95 %.

#### 4.1.3. Modelo normal - exponencial

Recordar el modelo de frontera de producción estocástica

$$\ln y_i = \beta_0 + \sum_n \beta_n \ln x_{ni} + v_i - u_i$$

en el que se asume:

- $v_i \rightsquigarrow N(0, \sigma_v^2)$ , independiente e idénticamente distribuida, es decir:

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v} \cdot \exp\left\{-\frac{v^2}{2\sigma_v^2}\right\}$$

- $u_i \rightsquigarrow \exp\left(\frac{1}{\sigma_u}\right)$ , independiente e idénticamente distribuida, es decir:

$$f(u) = \frac{1}{\sigma_u} \cdot \exp\left\{-\frac{u}{\sigma_u}\right\}$$

- $v_i$  y  $u_i$  son distribuidas independientemente de cada una, y de los regresores.

Como consecuencia de la independencia asumida entre variables, la función de densidad conjunta de  $v$  y  $u$  será el producto de sus funciones de densidad individuales, por tanto, se tiene que:

$$\begin{aligned} f(u, v) &= f(u) \cdot f(v) = \frac{1}{\sigma_u} \cdot \exp\left\{-\frac{u}{\sigma_u}\right\} \cdot \frac{1}{\sqrt{2\pi}\sigma_v} \cdot \exp\left\{-\frac{v^2}{2\sigma_v^2}\right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp\left\{-\frac{u}{\sigma_u} - \frac{v^2}{2\sigma_v^2}\right\} \end{aligned}$$

Como  $\epsilon = v - u$ , despejando  $v$  quedaría  $v = u + \epsilon$ , y sustituyendo en la expresión anterior, se obtiene la función de densidad conjunta de  $u$  y  $\epsilon$ :

$$f(u, \epsilon) = \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp\left\{-\frac{u}{\sigma_u} - \frac{1}{2\sigma_v^2}(u + \epsilon)^2\right\}$$

El siguiente paso es calcular la función de densidad marginal de  $\epsilon$  integrando la función de densidad conjunta respecto de  $u$ :

$$\begin{aligned} f(\epsilon) &= \int_0^\infty f(u, \epsilon) du = \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp\left\{-\frac{u}{\sigma_u} - \frac{1}{2\sigma_v^2}(u + \epsilon)^2\right\} du \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp\left\{-\frac{u}{\sigma_u} - \frac{1}{2\sigma_v^2}(u^2 + \epsilon^2 + 2u\epsilon)\right\} du \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp\left\{-\frac{\epsilon^2}{2\sigma_v^2} - \frac{2u\sigma_v^2 + u^2\sigma_u + 2u\epsilon\sigma_u}{2\sigma_v^2\sigma_u}\right\} du \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp\left\{-\frac{\epsilon^2}{2\sigma_v^2}\right\} \cdot \exp\left\{-\frac{2u\sigma_v^2 + u^2\sigma_u + 2u\epsilon\sigma_u}{2\sigma_v^2\sigma_u}\right\} du \\ &= \exp\left\{-\frac{\epsilon^2}{2\sigma_v^2}\right\} \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp\left\{-\frac{2u\sigma_v^2 + u^2\sigma_u + 2u\epsilon\sigma_u}{2\sigma_v^2\sigma_u}\right\} du \\ &= \exp\left\{-\frac{\epsilon^2}{2\sigma_v^2}\right\} \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp\left\{-\frac{\sigma_u u^2 + 2u(\sigma_v^2 + \epsilon\sigma_u)}{2\sigma_v^2\sigma_u}\right\} du \\ &= \exp\left\{-\frac{\epsilon^2}{2\sigma_v^2}\right\} \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp\left\{-\frac{\sigma_u}{2\sigma_v^2\sigma_u} \left[u^2 + \frac{2u(\sigma_v^2 + \epsilon\sigma_u)}{\sigma_u}\right]\right\} du \\ &= \exp\left\{-\frac{\epsilon^2}{2\sigma_v^2}\right\} \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp\left\{-\frac{1}{2\sigma_v^2} \left[u^2 + \frac{2u(\sigma_v^2 + \epsilon\sigma_u)}{\sigma_u}\right]\right\} du \\ &= \exp\left\{-\frac{\epsilon^2}{2\sigma_v^2}\right\} \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp\left\{-\frac{1}{2\sigma_v^2} \left[\left(u + \frac{\sigma_v^2 + \epsilon\sigma_u}{\sigma_u}\right)^2 - \left(\frac{\sigma_v^2 + \epsilon\sigma_u}{\sigma_u}\right)^2\right]\right\} du \end{aligned}$$

$$\begin{aligned}
&= \exp \left\{ -\frac{\epsilon^2}{2\sigma_v^2} \right\} \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp \left\{ \frac{1}{2\sigma_v^2} \left( \frac{\sigma_v^2 + \epsilon\sigma_u}{\sigma_u} \right)^2 \right\} \cdot \exp \left\{ -\frac{1}{2\sigma_v^2} \left( u + \frac{\sigma_v^2 + \epsilon\sigma_u}{\sigma_u} \right)^2 \right\} du \\
&= \exp \left\{ -\frac{\epsilon^2}{2\sigma_v^2} + \frac{1}{2\sigma_v^2} \left( \frac{\sigma_v^2 + \epsilon\sigma_u}{\sigma_u} \right)^2 \right\} \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp \left\{ -\frac{1}{2\sigma_v^2} \left( u + \frac{\sigma_v^2 + \epsilon\sigma_u}{\sigma_u} \right)^2 \right\} du \\
&= \frac{1}{\sigma_u} \exp \left\{ -\frac{\epsilon^2}{2\sigma_v^2} + \frac{1}{2\sigma_v^2} \left( \frac{\sigma_v^2 + \epsilon\sigma_u}{\sigma_u} \right)^2 \right\} \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_v} \cdot \exp \left\{ -\frac{1}{2\sigma_v^2} \left( u + \frac{\sigma_v^2 + \epsilon\sigma_u}{\sigma_u} \right)^2 \right\} du
\end{aligned}$$

Todo lo que está fuera de la integral, es decir, todo lo que no depende de la variable de integración  $u$ , se deja aparte y se intenta resolver la integral. Para ello es necesario hacer un cambio de variable:

$$\begin{aligned}
&\int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_v} \cdot \exp \left\{ -\frac{1}{2\sigma_v^2} \left( u + \frac{\sigma_v^2 + \epsilon\sigma_u}{\sigma_u} \right)^2 \right\} du = \\
&\int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_v} \cdot \exp \left\{ -\frac{1}{2} \left( \frac{u - \frac{-(\sigma_v^2 + \epsilon\sigma_u)}{\sigma_u}}{\sigma_v} \right)^2 \right\} du
\end{aligned}$$

El cambio de variable será el siguiente:

$$t = \frac{u - \frac{-(\sigma_v^2 + \epsilon\sigma_u)}{\sigma_u}}{\sigma_v}$$

Derivando la expresión anterior se tiene:

$$dt = \frac{1}{\sigma_v} du \Rightarrow du = \sigma_v \cdot dt$$

Por último hay que modificar los extremos de integración:

■ Extremo  $u = 0$  :

$$t = \frac{\frac{-\sigma_v^2 + \epsilon\sigma_u}{\sigma_u}}{\sigma_v} = \frac{-\sigma_v^2 - \epsilon\sigma_u}{\sigma_u\sigma_v} = -\frac{\sigma_v}{\sigma_u} - \frac{\epsilon}{\sigma_v}$$

■ Extremo  $u = \infty \Rightarrow t = \infty$

Ahora ya se puede hacer la integral:

$$\begin{aligned}
&\int_{-\frac{\sigma_v}{\sigma_u} - \frac{\epsilon}{\sigma_v}}^\infty \frac{1}{\sqrt{2\pi}\sigma_v} \cdot \exp \left\{ -\frac{1}{2}t^2 \right\} \cdot \sigma_v dt \\
&= \int_{-\frac{\sigma_v}{\sigma_u} - \frac{\epsilon}{\sigma_v}}^\infty \frac{1}{\sqrt{2\pi}} \cdot \exp \left\{ -\frac{1}{2}t^2 \right\} dt = \Phi \left( -\frac{\sigma_v}{\sigma_u} - \frac{\epsilon}{\sigma_v} \right)
\end{aligned}$$

A continuación se reagrupan todos los términos que estaban fuera de la integral porque no dependían de  $u$ :

$$\begin{aligned} \frac{1}{\sigma_u} \exp \left\{ -\frac{\epsilon^2}{2\sigma_v^2} + \frac{1}{2\sigma_v^2} \left( \frac{\sigma_v^2 + \epsilon\sigma_u}{\sigma_u} \right)^2 \right\} &= \frac{1}{\sigma_u} \exp \left\{ -\frac{\epsilon^2}{2\sigma_v^2} + \frac{1}{2\sigma_v^2\sigma_u^2} (\sigma_v^2 + \epsilon\sigma_u)^2 \right\} = \\ &= \frac{1}{\sigma_u} \exp \left\{ \frac{-\epsilon^2\sigma_u^2 + \sigma_v^4 + \epsilon^2\sigma_u^2 + 2\epsilon\sigma_u\sigma_v^2}{2\sigma_v^2\sigma_u^2} \right\} = \frac{1}{\sigma_u} \exp \left\{ \frac{\sigma_v^4 + 2\epsilon\sigma_u\sigma_v^2}{2\sigma_v^2\sigma_u^2} \right\} = \\ &= \frac{1}{\sigma_u} \exp \left\{ \frac{\sigma_v^2 + 2\epsilon\sigma_u}{2\sigma_u^2} \right\} = \frac{1}{\sigma_u} \exp \left\{ \frac{\epsilon}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2} \right\} \end{aligned}$$

Juntando todo lo anterior, se obtiene la función marginal de  $\epsilon$ :

$$f(\epsilon) = \int_0^\infty f(u, \epsilon) du = \left( \frac{1}{\sigma_u} \right) \cdot \exp \left\{ \frac{\epsilon}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2} \right\} \cdot \Phi \left( -\frac{\epsilon}{\sigma_v} - \frac{\sigma_v}{\sigma_u} \right)$$

donde  $\Phi(\cdot)$  es la función de distribución de una normal estandar. La función de densidad  $f(\epsilon)$  es asimétricamente distribuida con media y varianza:

$$E(\epsilon) = -E(u) = -\sigma_u$$

$$Var(\epsilon) = \sigma_u^2 + \sigma_v^2$$

La forma de la distribución normal - exponencial está determinada por los parámetros de la desviación estándar  $\sigma_u$  y  $\sigma_v$ . Cuando el cociente  $\sigma_u/\sigma_v$  crece, la distribución se parece más a una distribución exponencial negativa, mientras que si  $\sigma_u/\sigma_v$  decrece la distribución tiende a una normal.

Utilizando la función de densidad marginal de  $\epsilon$ , se puede obtener el logaritmo de la función de verosimilitud para una muestra de  $I$  productores:

$$\begin{aligned} \ln L &= \ln \left( \prod_i \left[ \left( \frac{1}{\sigma_u} \right) \cdot \Phi \left( -\frac{\epsilon_i}{\sigma_v} - \frac{\sigma_v}{\sigma_u} \right) \cdot \exp \left\{ \frac{\epsilon_i}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2} \right\} \right] \right) = \\ &= \sum \left( \ln \left[ \left( \frac{1}{\sigma_u} \right) \cdot \Phi \left( -\frac{\epsilon_i}{\sigma_v} - \frac{\sigma_v}{\sigma_u} \right) \cdot \exp \left\{ \frac{\epsilon_i}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2} \right\} \right] \right) = \\ &= \sum_i \ln \frac{1}{\sigma_u} + \sum_i \ln \Phi \left( -\frac{\epsilon_i}{\sigma_v} - \frac{\sigma_v}{\sigma_u} \right) + \sum_i \ln \exp \left\{ \frac{\epsilon_i}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2} \right\} = \\ &= -I \ln \sigma_u + \sum_i \ln \Phi \left( -\frac{\epsilon_i}{\sigma_v} - \frac{\sigma_v}{\sigma_u} \right) + \sum_i \frac{\epsilon_i}{\sigma_u} + I \left( \frac{\sigma_v^2}{2\sigma_u^2} \right) = \end{aligned}$$

$$= -I \ln \sigma_u + \sum_i \ln \Phi(-A) + \sum_i \frac{\epsilon_i}{\sigma_u} + I \left( \frac{\sigma_v^2}{2\sigma_u^2} \right)$$

donde  $A = -\tilde{\mu}/\sigma_v$  y  $\tilde{\mu} = -\epsilon - (\sigma_v^2/\sigma_u)$ . El logaritmo de la verosimilitud puede ser maximizado con respecto de los parámetros para obtener la estimación máximo verosímil de ellos. Recordar que  $\epsilon_i = \ln y_i - \beta_0 - \sum_n \beta_n \ln x_{ni}$  para hacer la derivada respecto de  $\vec{\beta}$ . Las expresiones son las siguientes:

$$\begin{aligned} \frac{\partial \ln L}{\partial \vec{\beta}} &= \frac{1}{\sigma_u} \sum_i \frac{\phi(-A)}{\Phi(-A)} \cdot x_i + \frac{1}{\sigma_u} \sum_i x_i = 0 \\ \frac{\partial \ln L}{\partial \sigma_u} &= -\frac{I}{\sigma_u} + \frac{\sigma_v}{\sigma_u^2} \sum_i \frac{\phi(-A)}{\Phi(-A)} - \frac{1}{\sigma_u^2} \sum_i \epsilon_i - I \left( \frac{\sigma_v^2}{\sigma_u^3} \right) = 0 \\ \frac{\partial \ln L}{\partial \sigma_v} &= \left( \frac{1}{\sigma_v^2} \sum_i \epsilon_i - \frac{1}{\sigma_u} \right) \cdot \sum_i \frac{\phi(-A)}{\Phi(-A)} + I \left( \frac{\sigma_v}{\sigma_u^2} \right) = 0 \end{aligned}$$

donde  $x_i$  es un vector formado por los logaritmos de los elementos de la fila  $i$ -ésima de la matriz de datos  $X$ .

Al igual que en el caso anterior, derivando la expresión e igualando a cero no se puede sacar una expresión analítica de dichos estimadores, luego hay que recurrir a una maximización numérica de la función.

Los puntos estimados de eficiencia técnica pueden ser obtenidos también, con la técnica JLMS, a través de la media y la moda de la distribución condicional de  $u$  dado  $\epsilon$ , la cual se calcula a continuación:

$$\begin{aligned} f(u|\epsilon) &= \frac{f(u, \epsilon)}{f(\epsilon)} = \frac{\frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp \left\{ -\frac{u}{\sigma_u} - \frac{1}{2\sigma_v^2}(u + \epsilon)^2 \right\}}{\left( \frac{1}{\sigma_u} \right) \cdot \Phi \left( -\frac{\epsilon}{\sigma_v} - \frac{\sigma_v}{\sigma_u} \right) \cdot \exp \left\{ \frac{\epsilon}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2} \right\}} = \\ &= \frac{1}{\sqrt{2\pi}\sigma_v\Phi \left( -\frac{\epsilon}{\sigma_v} - \frac{\sigma_v}{\sigma_u} \right)} \cdot \exp \left\{ -\frac{u}{\sigma_u} - \frac{1}{2\sigma_v^2}(u + \epsilon)^2 - \frac{\epsilon}{\sigma_u} - \frac{\sigma_v^2}{2\sigma_u^2} \right\} = \\ &= \frac{1}{\sqrt{2\pi}\sigma_v\Phi \left( -\frac{1}{\sigma_v} \left( \epsilon + \frac{\sigma_v^2}{\sigma_u} \right) \right)} \cdot \exp \left\{ \frac{-2u\sigma_u\sigma_v^2 - \epsilon^2\sigma_u^2 - u^2\sigma_u^2 - 2\epsilon u\sigma_u^2 - 2\epsilon\sigma_u\sigma_v^2 - \sigma_v^4}{2\sigma_u^2\sigma_v^2} \right\} = \\ &= \frac{1}{\sqrt{2\pi}\sigma_v\Phi \left( \frac{\epsilon - \frac{\sigma_v^2}{\sigma_u}}{\sigma_v} \right)} \cdot \exp \left\{ \frac{-u^2\sigma_u^2 - 2u(\sigma_u\sigma_v^2 + \epsilon\sigma_u^2) - \epsilon^2\sigma_u^2 - 2\epsilon\sigma_u\sigma_v^2 - \sigma_v^4}{2\sigma_u^2\sigma_v^2} \right\} = \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}\sigma_v\Phi\left(\frac{\tilde{\mu}}{\sigma_v}\right)} \cdot \exp\left\{(\sigma_u^2) \cdot \frac{-u^2 - 2u\left(\frac{\sigma_v^2}{\sigma_u} + \epsilon\right) - \epsilon^2 - 2\epsilon\frac{\sigma_v^2}{\sigma_u} - \frac{\sigma_v^4}{\sigma_u^2}}{2\sigma_u^2\sigma_v^2}\right\} = \\
&= \frac{1}{\sqrt{2\pi}\sigma_v\Phi\left(\frac{\tilde{\mu}}{\sigma_v}\right)} \cdot \exp\left\{\frac{-\left(u + \left(\frac{\sigma_v^2}{\sigma_u} + \epsilon\right)\right)^2 + \frac{\sigma_v^4}{\sigma_u^2} + 2\epsilon\frac{\sigma_v^2}{\sigma_u} + \epsilon^2 - \epsilon^2 - 2\epsilon\frac{\sigma_v^2}{\sigma_u} - \frac{\sigma_v^4}{\sigma_u^2}}{2\sigma_v^2}\right\} = \\
&= \frac{1}{\sqrt{2\pi}\sigma_v\Phi\left(\frac{\tilde{\mu}}{\sigma_v}\right)} \cdot \exp\left\{-\frac{\left(u + \left(\frac{\sigma_v^2}{\sigma_u} + \epsilon\right)\right)^2}{2\sigma_v^2}\right\} = \\
&= \frac{1}{\sqrt{2\pi}\sigma_v\Phi\left(\frac{\tilde{\mu}}{\sigma_v}\right)} \cdot \exp\left\{-\frac{\left(u - \left(\frac{-\sigma_v^2}{\sigma_u} - \epsilon\right)\right)^2}{2\sigma_v^2}\right\} = \frac{1}{\sqrt{2\pi}\sigma_v\Phi\left(\frac{\tilde{\mu}}{\sigma_v}\right)} \cdot \exp\left\{-\frac{(u - \tilde{\mu})^2}{2\sigma_v^2}\right\}
\end{aligned}$$

lo que indica que  $f(u|\epsilon)$  se distribuye como  $N^+(\tilde{\mu}, \sigma_v^2)$ .

Tanto la esperanza como la moda pueden ser utilizadas para calcular la estimación de la eficiencia técnica específica de cada productor. Se podrían hallar como:

$$E(u_i|\epsilon_i) = \tilde{\mu}_i + \sigma_v \left[ \frac{\phi(-\tilde{\mu}_i/\sigma_v)}{\Phi(\tilde{\mu}_i/\sigma_v)} \right] = \sigma_v \left[ \frac{\phi(A)}{\Phi(-A)} - A \right]$$

$$\text{Moda}(u_i|\epsilon_i) = \begin{cases} \tilde{\mu}_i & \text{si } \tilde{\mu}_i \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

donde  $\phi(\cdot)$  y  $\Phi(\cdot)$  son la función de densidad y función de distribución de la normal estandar. Ambos estimadores son inconsistentes.

Los correspondientes intervalos de confianza para la eficiencia técnica pueden ser construidos, al igual que en el caso normal - seminormal. La única diferencia es que  $(u|\epsilon)$  tiene una función de densidad distinta. Habría que hallar el límite inferior y superior,  $(L_i, U_i)$ , lo que implica el límite inferior y superior de  $(\exp\{-\hat{u}_i\}, \epsilon_i)$ . De esta forma, el intervalo de confianza al  $(1 - \alpha)100\%$  viene dado por:

$$L_i = \exp\{-\hat{u}_i - z_L \cdot \sigma_v\}$$

$$U_i = \exp\{-\hat{u}_i - z_U \cdot \sigma_v\}$$

donde

$$Pr(Z > z_L) = \frac{\alpha}{2} \left[ 1 - \Phi\left(-\frac{\tilde{\mu}_i}{\sigma_v}\right) \right]$$

$$Pr(Z > z_U) = \left(1 - \frac{\alpha}{2}\right) \left[1 - \Phi\left(-\frac{\tilde{\mu}_i}{\sigma_v}\right)\right]$$

siendo  $Z \rightsquigarrow N(0, 1)$ . Como consecuencia, se tiene que los valores  $z_L$  y  $z_U$  se calculan como:

$$z_L = \Phi^{-1} \left\{ \left[1 - \frac{\alpha}{2}\right] \left[1 - \Phi\left(-\frac{\tilde{\mu}_i}{\sigma_v}\right)\right] \right\}$$

$$z_U = \Phi^{-1} \left\{ \left[1 - \left(1 - \frac{\alpha}{2}\right)\right] \left[1 - \Phi\left(-\frac{\tilde{\mu}_i}{\sigma_v}\right)\right] \right\}$$

#### 4.1.4. Ejemplo N - Exp

Este apartado está dedicado a calcular la frontera de producción estocástica de los datos `front41Data` con el método de máxima verosimilitud y suponiendo que  $u_i$  sigue un distribución exponencial. Para resolver el problema, se utilizará el paquete `sfa`, puesto que la función del paquete `frontier` no deja elegir la distribución de  $u$ . Recordar que se quiere calcular la producción,  $y$ , en función de las variables capital y trabajo:

```
> library(sfa)
> Exp_sfa <- sfa( y ~ x1 + x2, data = f, intercept = TRUE, fun = "exp",
+               form = "production")
> estimaciones <- Exp_sfa$coef; estimaciones
```

```
Intercept      x1      x2      sigmau2      sigmav2
0.44044913 0.28434984 0.54235073 0.05537614 0.05430909
```

De esta forma, ya ha sido calculada la estimación de todos los parámetros del modelo, los tres primeros corresponden a las componentes de  $\vec{\beta}$  y las otras a  $\sigma_u^2$  y  $\sigma_v^2$ . Faltaría la estimación de la eficiencia técnica de cada productor. Para ello se utiliza la función `te.eff.sfa` y se hace un resumen de la variable:

```
> TE7 <- te.eff.sfa(Exp_sfa);
> TE7 <- as.vector(TE7); TE7
```

```
[1] 0.7544882 0.8811071 0.8152280 0.8319712 0.7902668 0.8504787 0.6661217
[8] 0.8232132 0.8907715 0.8386457 0.6448146 0.9454181 0.5050571 0.7241563
```

```
[15] 0.9080285 0.6442747 0.8053508 0.8387400 0.8977582 0.8683862 0.8146410
[22] 0.9068815 0.8857689 0.8371846 0.6213508 0.9204166 0.8697962 0.8535048
[29] 0.8983250 0.7289136 0.6865148 0.8396455 0.9122308 0.4673377 0.3733075
[36] 0.9155910 0.8877346 0.7789095 0.7748039 0.8879350 0.8399490 0.7812638
[43] 0.8664604 0.9146119 0.8270521 0.8168631 0.8948464 0.8557774 0.7736198
[50] 0.5975698 0.8893387 0.8937444 0.8982838 0.8369811 0.8744222 0.8388003
[57] 0.9095792 0.7649100 0.8970351 0.8031703
```

```
> summary(TE7)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.3733  0.7779  0.8388  0.8093  0.8897  0.9454
```

En función de los resultados obtenidos, se concluye que la media de la eficiencia técnica es alta, con un valor de 0,8388, y el máximo valor que toma la variable es 0,9454, por lo que ningún productor es 100 % técnicamente eficiente.

### Intervalos de confianza

En este último apartado se calcularán los intervalos de confianza del 95 % para la eficiencia técnica de cada productor. Se había descrito que se calculan como:

$$(\exp\{-\hat{u}_i - z_L \cdot \sigma_v\}, \exp\{-\hat{u}_i - z_U \cdot \sigma_v\})$$

donde

$$z_L = \Phi^{-1} \left\{ \left[ 1 - \frac{\alpha}{2} \right] \left[ 1 - \Phi \left( -\frac{\tilde{\mu}_i}{\sigma_v} \right) \right] \right\}$$

$$z_U = \Phi^{-1} \left\{ \left[ 1 - \left( 1 - \frac{\alpha}{2} \right) \right] \left[ 1 - \Phi \left( -\frac{\tilde{\mu}_i}{\sigma_v} \right) \right] \right\}$$

Se calculará en R como sigue:

```
> X <- cbind(1, x1, x2)
> beta <- as.vector(estimaciones[1:3])
> epsilon <- y - X%*%beta
```



```

> sigmau2 <- as.vector(estimaciones[4])
> sigmav2 <- as.vector(estimaciones[5])
> muTilde <- - epsilon - (sigmav2/sqrt(sigmau2))
> distribucion <- pnorm(-muTilde/sqrt(sigmav2))
> zL <- qnorm((1 - (0.05/2))*(1-distribucion))
> zU <- qnorm((0.05/2)*(1-distribucion))
> u <- - log (TE7)
> L <- exp(-u - zL*sqrt(sigmav2))
> U <- exp(-u - zU*sqrt(sigmav2))
> intervalos <- cbind(L,U)
> intervalos[1:5,]

```

```

      [,1]      [,2]
[1,] 0.6111180 1.212166
[2,] 1.0654396 1.597886
[3,] 0.7640910 1.346679
[4,] 0.8208299 1.393776
[5,] 0.6931696 1.286336

```

Para que no sea muy larga la salida, se muestra el intervalo de confianza para la eficiencia técnica de los cinco primeros productores.

## 4.2. Método de los momentos

La estrategia hasta ahora consistía en dos pasos. El primero una estimación máximo verosímil de todos los parámetros del modelo y el segundo estimar la distribución condicionada del término de la ineficiencia técnica por el término de error compuesto de cada productor. De esta forma se puede estimar la eficiencia técnica para cada productor basándose en el cálculo de la media o la moda de la distribución anterior (técnica JLMS).

El método de los momentos consiste en modificar la primera parte del método de máxima verosimilitud dividiéndola en dos fases, es decir:

1. a) Proceso de estimación mínimos cuadrados ordinarios (OLS) para generar así estimaciones consistentes de los parámetros describiendo la estructura de la frontera de producción, excepto del intercepto. Esta parte es independiente de la distribución.
- b) Se asume una distribución para obtener estimaciones consistentes del intercepto y los parámetros que describen la estructura de las dos componentes del error.

Esta forma de dividir el primer paso es equivalente a la aplicación de los mínimos cuadrados modificados a un modelo de frontera de producción determinística. Después se repite el segundo paso del método de máxima verosimilitud para estimar los  $u_i$  con la técnica JLMS. Esta estrategia puede ser aplicada a cualquiera de los dos modelos descritos en la sección anterior, es decir, suponiendo que la ineficiencia técnica se distribuye como una seminormal o como una exponencial.

#### 4.2.1. Modelo normal - seminormal

El modelo de frontera estocástica de producción es:

$$\ln y_i = [\beta_0 - E(u_i)] + \sum_n \beta_n \ln x_{ni} + v_i - [u_i - E(u_i)]$$

donde se asume que  $v_i$  es distribuida de forma simétrica con media cero y que  $u_i \geq 0$ . Por otra parte, el término de error,  $\epsilon_i^* = v_i - [u_i - E(u_i)]$ , tiene media cero y varianza constante.

- Etapa 1: aplicar mínimos cuadrados ordinarios (OLS) para obtener estimaciones consistentes de  $\vec{\beta}$ . Para ello hay que minimizar el cuadrado del error, es decir,

$$\vec{\epsilon}_i^* = \ln \vec{y} - \ln \vec{\hat{y}} = \ln \vec{y} - \left[ (\beta_0 - E(u_i)) + \sum_n \beta_n \ln x_{ni} \right]$$

se denota  $\vec{\hat{y}} = (\ln X)\vec{\hat{\beta}}$  y se tiene que el cuadrado del error es:

$$\begin{aligned} \|\vec{\epsilon}_i^*\|^2 &= \sum_{i=1}^I (\epsilon_i^*)^2 = (\ln \vec{y} - (\ln X)\vec{\hat{\beta}})^t \cdot (\ln \vec{y} - (\ln X)\vec{\hat{\beta}}) \\ &= (\ln \vec{y})^t \ln \vec{y} - 2(\ln \vec{y})^t (\ln X)\vec{\hat{\beta}} + \vec{\hat{\beta}}^t (\ln X)^t (\ln X)\vec{\hat{\beta}} \end{aligned}$$

Hay que calcular  $\min_{\vec{\beta}} \|\vec{\epsilon}_i^*\|^2$ , que se hace derivando la expresión anterior respecto de  $\vec{\beta}$  e igualando a 0:

$$\begin{aligned} \frac{\partial(\|\vec{\epsilon}_i^*\|^2)}{\partial \vec{\beta}} = 0 &\iff -2(\ln \vec{y})^t \ln X + 2(\ln X)^t (\ln X) \vec{\beta} = 0 \\ &\iff \vec{\beta} = ((\ln X)^t \ln X)^{-1} (\ln X)^t \ln \vec{y} \end{aligned}$$

- Etapa 2: estimar  $\beta_0$ ,  $\sigma_u^2$  y  $\sigma_v^2$ . Para ello es necesario asumir las distribuciones de las componentes de error  $v_i$  y  $u_i$ , es decir,  $v_i \rightsquigarrow N(0, \sigma_v^2)$  y  $u_i \rightsquigarrow N^+(0, \sigma_u^2)$ , entonces:

$$E(u_i) = \sqrt{2/\pi} \sigma_u$$

$$Var(u_i) = E(u_i^2) = [(\pi - 2)/\pi] \sigma_u^2$$

$$E(u_i^3) = -\sqrt{2/\pi} (1 - 4/\pi) \sigma_u^3$$

Los momentos centrales de  $\epsilon_i = v_i - u_i$  son:

$$\mu_1 = E(\epsilon_i) = -E(u_i) = -\sqrt{2/\pi} \sigma_u$$

$$\mu_2 = \frac{\pi - 2}{\pi} \sigma_u^2 + \sigma_v^2$$

$$\mu_3 = \sqrt{\frac{2}{\pi}} \left(1 - \frac{4}{\pi}\right) \sigma_u^3$$

Los momentos centrales de orden dos y tres de  $\epsilon_i$  son los mismos que los de  $\epsilon_i^* = v_i - [u_i - E(u_i)]$ , puesto que  $E(u_i)$  es constante. Luego se tiene que:

$$m_2 = \frac{\pi - 2}{\pi} \sigma_u^2 + \sigma_v^2$$

$$m_3 = \sqrt{\frac{2}{\pi}} \left(1 - \frac{4}{\pi}\right) \sigma_u^3$$

Para hallar las estimaciones de los parámetros, hay que igualar los momentos muestrales con los momentos poblacionales. Los residuos obtenidos por el método de mínimos cuadrados ordinarios se utilizan para calcular los momentos muestrales  $\hat{m}_2$  y  $\hat{m}_3$ :

$$\hat{m}_2 = \frac{1}{n} \sum_i (\epsilon_i^* - \bar{\epsilon}^*)^2$$

$$\hat{m}_3 = \frac{1}{n} \sum_i (\epsilon_i^* - \bar{\epsilon}^*)^3$$

Ahora ya se pueden utilizar para estimar  $\sigma_u^2$  y  $\sigma_v^2$ . Estas estimaciones son:

$$\hat{\sigma}_u^2 = \left( \frac{\hat{m}_3}{\sqrt{2/\pi}(1 - 4/\pi)} \right)^{2/3}$$

$$\hat{\sigma}_v^2 = \hat{m}_2 - \left( 1 - \frac{2}{\pi} \right) \hat{\sigma}_u^2$$

Una vez calculada la estimación de  $\sigma_u^2$ , se puede obtener una estimación consistente para el intercepto  $\beta_0$  como:

$$\hat{\beta}_0 = [\beta_0 - E(\hat{u}_i)] + E(\hat{u}_i) = [\beta_0 - E(\hat{u}_i)] + \sqrt{\frac{2}{\pi}} \hat{\sigma}_u = \text{intercepto de OLS} + \sqrt{\frac{2}{\pi}} \hat{\sigma}_u$$

Ahora ya se tienen estimaciones consistentes de todos los parámetros del modelo y se puede pasar a la siguiente fase.

- Etapa 3: el último paso consiste en aplicar la técnica JLMS para obtener  $E(u_i|\epsilon_i)$  o  $\text{Moda}(u_i|\epsilon_i)$ . Esta técnica ya ha sido descrita anteriormente y consistía en calcular:

$$f(u|\epsilon) = \frac{f(u, \epsilon)}{f(\epsilon)} = \frac{\frac{1}{\sqrt{2\pi}\sigma_*} \cdot \exp\left\{-\frac{(u-\mu_*)^2}{2\sigma_*^2}\right\}}{1 - \Phi\left(-\frac{\mu_*}{\sigma_*}\right)}$$

donde  $\mu_* = -\epsilon\sigma_u^2/\sigma^2$  y  $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$ . La distribución de  $f(u|\epsilon)$  es  $N^+(\mu_*, \sigma_*^2)$ . La media y la moda de la distribución son estimadores puntuales de  $u_i$ :

$$E(u_i|\epsilon_i) = \sigma_* \left[ \frac{\phi(\epsilon_i\lambda/\sigma)}{1 - \Phi(\epsilon_i\lambda/\sigma)} - \left( \frac{\epsilon_i\lambda}{\sigma} \right) \right]$$

$$\text{Moda}(u_i|\epsilon_i) = \begin{cases} -\mu_{*i} & \text{si } \epsilon_i \leq 0 \\ 0 & \text{en otro caso} \end{cases}$$

Para finalizar, se calcula la eficiencia técnica de cada productor:

$$\text{TE}_i = \exp\{-\hat{u}_i\}$$

donde  $\hat{u}_i$  puede ser  $E(u_i|\epsilon_i)$  o  $\text{Moda}(u_i|\epsilon_i)$ .

Los intervalos de confianza para los parámetros se podrían construir de forma análoga al método de máxima verosimilitud.

### 4.2.2. Ejemplo N - N<sup>+</sup>

Para la estimación de  $\vec{\beta}$  se utiliza la regresión lineal. Recordar que las variables  $y$ ,  $x_1$  y  $x_2$  ya son las transformaciones logarítmicas de las originales.

```
> MM <- lm(y~x1+x2, data=f)
> beta <- as.vector(MM$coefficients); beta
```

```
[1] 0.2448982 0.2804926 0.5333064
```

De esta forma se obtiene la estimación de  $\vec{\beta}$ . Faltan las estimaciones de  $\sigma_u^2$  y  $\sigma_v^2$  y corregir el intercepto. Para ello es necesario calcular las estimaciones de los momentos de orden dos y tres:

$$\hat{m}_2 = \frac{1}{n} \sum_{i=1}^{60} (\epsilon_i^* - \bar{\epsilon}^*)^2 \quad \text{y} \quad \hat{m}_3 = \frac{1}{n} \sum_{i=1}^{60} (\epsilon_i^* - \bar{\epsilon}^*)^3$$

donde  $\epsilon_i^*$  son los errores de las estimaciones por el método de los mínimos cuadrados ordinarios, es decir, los residuos de la regresión lineal. En R se calculan como:

```
> epsilonEstrella <- MM$residuals
> media <- mean(epsilonEstrella)
> m2 <- sum((epsilonEstrella-media)^2)/60; m2
```

```
[1] 0.1082857
```

```
> m3 <- sum((epsilonEstrella-media)^3)/60; m3
```

```
[1] -0.01974243
```

Ya se había demostrado anteriormente que  $\sigma_u^2$ ,  $\sigma_v^2$  y  $\beta_0$  se calculan como:

$$\hat{\sigma}_u^2 = \left( \frac{\hat{m}_3}{\sqrt{2/\pi}(1 - 4/\pi)} \right)^{2/3}$$
$$\hat{\sigma}_v^2 = \hat{m}_2 - \left( 1 - \frac{2}{\pi} \right) \hat{\sigma}_u^2$$
$$\hat{\beta}_0 = \text{intercepto de OLS} + \sqrt{\frac{2}{\pi}} \hat{\sigma}_u$$

Los cuales ya se puede calcular en R porque los valores  $m_2$  y  $m_3$  son conocidos:

```
> sigmau2 <- (m3 / ((sqrt(2/pi))*(1 - (4/pi))))^(2/3); sigmau2
```

```
[1] 0.2016561
```

```
> sigmav2 <- m2 - (1 - (2/pi))*sigmau2; sigmav2
```

```
[1] 0.03500786
```

```
> beta0 <- beta[1] + sqrt(2/pi)*sqrt(sigmau2); beta0
```

```
[1] 0.6031973
```

La última parte consiste en calcular la eficiencia técnica como  $TE_i = \exp\{-\hat{u}_i\}$ , utilizando  $E(u_i|\epsilon_i)$  como estimador de la ineficiencia. Se había descrito que es:

$$E(u_i|\epsilon_i) = \sigma_* \left[ \frac{\phi(\epsilon_i \lambda / \sigma)}{1 - \Phi(\epsilon_i \lambda / \sigma)} - \left( \frac{\epsilon_i \lambda}{\sigma} \right) \right]$$

Por tanto, hay que calcular todos los valores necesarios para hallar la expresión anterior. Hay que tener en cuenta que los valores que se conocen son los de  $\epsilon_i^*$  y los que se necesitan son los de  $\epsilon_i$ , los cuales se pueden calcular como la diferencia de los de  $\epsilon_i^* - E(u_i)$ , que ya son datos conocidos:

```
> sigma <- sqrt(sigmau2+sigmav2)
> sigmaEstrella <- sqrt(sigmau2*sigmav2)/sigma
> e_ui <- sqrt(2/pi)*sqrt(sigmau2)
> epsilon <- epsilonEstrella - e_ui
> lambda <- sqrt(sigmau2)/sqrt(sigmav2)
> densidad <- dnorm(epsilon*lambda/sigma)
> distribucion <- pnorm(epsilon*lambda/sigma)
> esperanza <- sigmaEstrella *
+ ((densidad/(1-distribucion))- (epsilon*lambda/sigma))
> esperanza <- as.vector(esperanza)
```

Una vez obtenidos los valores  $u_i$  que están almacenados en la variable 'esperanza', solo falta calcular la eficiencia técnica y hacer un resumen de la misma:

```
> TE8 <- exp(-esperanza); TE8
```

```
[1] 0.6080410 0.8171918 0.6929662 0.7164857 0.6515176 0.7521090 0.5211810  
[8] 0.7059725 0.8349649 0.7292848 0.4992538 0.9439340 0.4067810 0.5737211  
[15] 0.8721224 0.5056493 0.6766034 0.7308782 0.8522422 0.7888158 0.6905967  
[22] 0.8698434 0.8271990 0.7209687 0.4861076 0.9005184 0.7939221 0.7579995  
[29] 0.8495300 0.5765463 0.5320702 0.7130998 0.8817785 0.3803065 0.3135130  
[36] 0.8908622 0.8329063 0.6383463 0.6322432 0.8275819 0.7302108 0.6421088  
[43] 0.7868169 0.8872231 0.7123359 0.6910466 0.8479470 0.7615710 0.6312531  
[50] 0.4706965 0.8336810 0.8445628 0.8556894 0.7251439 0.8015012 0.7319279  
[57] 0.8686640 0.6244499 0.8516170 0.6712469
```

```
> summary(TE8)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.  
0.3135 0.6368 0.7297 0.7161 0.8340 0.9439
```

Con esta técnica el valor medio que se obtiene de la eficiencia es 0,7161, lo cual es alto. Sin embargo, ningún productor es técnicamente eficiente porque el valor máximo que toma la variable es 0,9439.

## Intervalos de confianza

Los intervalos de confianza se puede construir de forma análoga al método de máxima verosimilitud. Recordar que viene dado por:

$$(\exp\{-\hat{u}_i - z_L \cdot \sigma_*\}, \exp\{-\hat{u}_i - z_U \cdot \sigma_*\})$$

donde

$$z_L = \Phi^{-1} \left\{ \left[ 1 - \frac{\alpha}{2} \right] \left[ 1 - \Phi \left( -\frac{\mu_{*i}}{\sigma_*} \right) \right] \right\}$$
$$z_U = \Phi^{-1} \left\{ \left[ 1 - \left( 1 - \frac{\alpha}{2} \right) \right] \left[ 1 - \Phi \left( -\frac{\mu_{*i}}{\sigma_*} \right) \right] \right\}$$

Se calculará en R como sigue:

```

> beta <- c(beta0, beta[2:3])
> muEstrella <- - epsilon*sigmau2/sigma2
> sigmaEstrella2 <- sigmau2*sigmav2/sigma2
> distribucion <- pnorm(-muEstrella/sqrt(sigmaEstrella2))
> zL <- qnorm((1 - (0.05/2))*(1-distribucion))
> zU <- qnorm((0.05/2)*(1-distribucion))
> L <- exp(-esperanza - zL*sqrt(sigmaEstrella2))
> U <- exp(-esperanza - zU*sqrt(sigmaEstrella2))
> intervalos <- cbind(L,U)
> intervalos[1:5,]

```

	L	U
1	0.4286683	0.8659785
2	0.7113862	1.1837008
3	0.5049928	0.9879819
4	0.5330922	1.0224114
5	0.4640273	0.9281665

Los resultados anteriores son los intervalos de confianza para la eficiencia técnica de los cinco primeros productores.

#### 4.2.3. Modelo normal - exponencial

- Paso 1: aplicar mínimos cuadrados ordinarios (OLS) para obtener estimaciones consistentes del vector  $\vec{\beta}$ , excepto de la primera coordenada, que se corresponde con el intercepto,  $\beta_0$  y hay que corregir posteriormente. Para ello hay que minimizar el cuadrado del error, derivar respecto de  $\vec{\beta}$  e igualar a 0. Este paso es análogo al modelo normal - seminormal porque todavía no se han hecho suposiciones sobre las distribuciones. Por tanto, se tiene que:

$$\vec{\hat{\beta}} = ((\ln X)^t \ln X)^{-1} (\ln X)^t \ln \vec{y}$$



- Paso 2: estimar  $\beta_0$ ,  $\sigma_u^2$  y  $\sigma_v^2$ . Para ello es necesario asumir las distribuciones de las componentes de error  $v_i$  y  $u_i$ , es decir,  $v_i \rightsquigarrow N(0, \sigma_v^2)$  y  $u_i \rightsquigarrow \exp\left(\frac{1}{\sigma_u}\right)$ , entonces:

$$E(u_i) = \frac{1}{1/\sigma_u} = \sigma_u$$

$$Var(u_i) = E(u_i^2) = \frac{2}{1/\sigma_u^2} = 2\sigma_u^2$$

$$E(u_i^3) = \frac{6}{1/\sigma_u^3} = 6\sigma_u^3$$

Los momentos centrales de  $\epsilon_i = v_i - u_i$  son:

$$\mu_1 = E(\epsilon_i) = -E(u_i) = -\sigma_u$$

$$\mu_2 = \sigma_v^2 + \frac{1}{1/\sigma_u^2} = \sigma_v^2 + \sigma_u^2$$

$$\mu_3 = -\frac{2}{1/\sigma_u^3} = -2\sigma_u^3$$

Los momentos centrales de orden dos y tres de  $\epsilon_i$  son los mismos que los de  $\epsilon_i^* = v_i - [u_i - E(u_i)]$ , puesto que  $E(u_i)$  es constante. Luego se tiene que:

$$m_2 = \sigma_u^2 + \sigma_v^2$$

$$m_3 = -2\sigma_u^3$$

Al igual que en el método de los momentos para el modelo normal - seminormal, los residuos de los mínimos cuadrados ordinarios se utilizan para estimar los momentos  $\hat{m}_2$  y  $\hat{m}_3$ . Una vez calculados, se estiman  $\sigma_u^2$  y  $\sigma_v^2$  como:

$$\hat{\sigma}_u^2 = \left(-\frac{\hat{m}_3}{2}\right)^{2/3}$$

$$\hat{\sigma}_v^2 = \hat{m}_2 - \hat{\sigma}_u^2$$

Ya se puede obtener una estimación consistente para  $\beta_0$ :

$$\hat{\beta}_0 = [\beta_0 - E(\hat{u}_i)] + E(\hat{u}_i) = [\beta_0 - E(\hat{u}_i)] + \hat{\sigma}_u = \text{intercepto de OLS} + \hat{\sigma}_u$$

Así ya se tienen estimaciones consistentes de los parámetros del modelo.

- Paso 3: aplicar la técnica JLMS para obtener  $E(u_i|\epsilon_i)$  o  $\text{Moda}(u_i|\epsilon_i)$ . Esta técnica ya ha sido descrita y consistía en:

$$f(u|\epsilon) = \frac{1}{\sqrt{2\pi}\sigma_v\Phi(-\tilde{\mu}/\sigma_v)} \cdot \exp\left\{-\frac{(u-\tilde{\mu})^2}{2\sigma_v^2}\right\}$$

Por tanto,  $f(u|\epsilon)$  se distribuye como  $N^+(\tilde{\mu}, \sigma_v^2)$ . La esperanza y la moda de la distribución son estimadores puntuales de  $u_i$ :

$$E(u_i|\epsilon_i) = \tilde{\mu}_i + \sigma_v \left[ \frac{\phi(-\tilde{\mu}_i/\sigma_v)}{\Phi(\tilde{\mu}_i/\sigma_v)} \right] = \sigma_v \left[ \frac{\phi(A)}{\Phi(-A)} - A \right]$$

$$\text{Moda}(u_i|\epsilon_i) = \begin{cases} \tilde{\mu}_i & \text{si } \tilde{\mu}_i \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

Faltaría hallar la eficiencia técnica de cada productor como:

$$\text{TE}_i = \exp\{-\hat{u}_i\}$$

donde  $\hat{u}_i$  puede ser  $E(u_i|\epsilon_i)$  o  $\text{Moda}(u_i|\epsilon_i)$ .

Los intervalos de confianza para los parámetros se podrían construir de forma análoga al método de máxima verosimilitud.

#### 4.2.4. Ejemplo N - Exp

Para la estimación del vector  $\vec{\beta}$  se utiliza la regresión lineal, cuya función en R es `lm`:

```
> MM <- lm(y~x1+x2, data=f)
> beta <- MM$coefficients
> beta <- as.vector(beta); beta

[1] 0.2448982 0.2804926 0.5333064
```

Así se obtienen las estimaciones de  $\beta_1$  y  $\beta_2$ , faltan las de  $\beta_0$ ,  $\sigma_u^2$  y  $\sigma_v^2$ , por tanto, es necesario calcular los momentos de orden dos y tres como se ha descrito:

```
> epsilonEstrella <- MM$residuals
> media <- mean(epsilonEstrella)
> m2 <- sum((epsilonEstrella-media)^2)/60; m2
```

```
[1] 0.1082857
```

```
> m3 <- sum((epsilonEstrella-media)^3)/60; m3
```

```
[1] -0.01974243
```

A continuación se calculan  $\sigma_u^2$ ,  $\sigma_v^2$  y  $\beta_0$  como:

$$\hat{\sigma}_u^2 = \left( -\frac{\hat{m}_3}{2} \right)^{2/3}$$

$$\hat{\sigma}_v^2 = \hat{m}_2 - \hat{\sigma}_u^2$$

$$\hat{\beta}_0 = \text{intercepto de OLS} + \hat{\sigma}_u$$

En R sería:

```
> sigmau2 <- (-m3/2)^(2/3); sigmau2
```

```
[1] 0.04601651
```

```
> sigmav2 <- m2 - sigmau2; sigmav2
```

```
[1] 0.06226917
```

```
> beta0 <- beta[1] + sqrt(sigmau2); beta0
```

```
[1] 0.4594128
```

La última parte consiste en calcular la eficiencia técnica como  $TE_i = \exp\{-\hat{u}_i\}$ , donde  $\hat{u}_i$  se considera que es  $E(u_i|\epsilon_i)$ :

$$E(u_i|\epsilon_i) = \tilde{\mu}_i + \sigma_v \left[ \frac{\phi(-\tilde{\mu}_i/\sigma_v)}{\Phi(\tilde{\mu}_i/\sigma_v)} \right]$$

El siguiente paso es calcular todos los valores necesarios para la esperanza, teniendo en cuenta que los valores que se necesitan son los de  $\epsilon_i$ , que se calcularán como la diferencia de los de  $\epsilon_i^* - E(u_i)$ . Recordar también que  $\tilde{\mu} = -\epsilon - (\sigma_v^2/\sigma_u)$ , que ya son todo datos conocidos y se puede calcular en R:

```

> e_ui <- sqrt(sigma_u2)
> epsilon <- epsilonEstrella - e_ui
> muTilde <- -epsilon -(sigma_v2 / sqrt(sigma_u2))
> densidad <- dnorm(-muTilde/sqrt(sigma_v2))
> distribucion <- pnorm(muTilde/sqrt(sigma_v2))
> esperanza <- muTilde + sqrt(sigma_v2)*(densidad/distribucion)
> esperanza <- as.vector(esperanza)

```

Ya se tiene la estimación de  $u_i$  que están en la variable 'esperanza'. Falta calcular la eficiencia técnica de cada productor:

```

> TE9 <- exp(-esperanza); TE9

[1] 0.7682208 0.8809174 0.8233326 0.8356575 0.7988526 0.8527304 0.6864259
[8] 0.8302679 0.8883542 0.8419861 0.6608383 0.9413867 0.5346884 0.7395130
[15] 0.9042233 0.6685065 0.8141353 0.8427575 0.8956379 0.8689296 0.8220347
[22] 0.9032224 0.8851035 0.8379018 0.6445647 0.9172789 0.8711095 0.8554075
[29] 0.8944869 0.7420452 0.6983836 0.8339415 0.9085287 0.4953514 0.3956150
[36] 0.9126941 0.8874918 0.7901740 0.7859840 0.8852637 0.8424348 0.7927028
[43] 0.8680725 0.9110101 0.8335518 0.8222820 0.8938167 0.8570141 0.7852938
[50] 0.6246578 0.8878162 0.8923875 0.8971064 0.8399648 0.8743231 0.8432638
[57] 0.9027064 0.7804696 0.8953722 0.8109989

```

Se hace, por último, un resumen de la variable:

```

> summary(TE9)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.3956  0.7891  0.8422  0.8150  0.8880  0.9414

```

Con esta técnica el valor medio que se obtiene de la eficiencia es 0,815, lo cual es muy alto. Sin embargo, ningún productor es 100% técnicamente eficiente porque el valor máximo que toma la variable es 0,9414.

## Intervalos de confianza

Los intervalos de confianza vendrán dados por las mismas expresiones que en el caso de máxima verosimilitud, es decir:

$$(\exp\{-\hat{u}_i - z_L \cdot \sigma_v\}, \exp\{-\hat{u}_i - z_U \cdot \sigma_v\})$$

donde

$$z_L = \Phi^{-1} \left\{ \left[ 1 - \frac{\alpha}{2} \right] \left[ 1 - \Phi \left( -\frac{\tilde{\mu}_i}{\sigma_v} \right) \right] \right\}$$
$$z_U = \Phi^{-1} \left\{ \left[ 1 - \left( 1 - \frac{\alpha}{2} \right) \right] \left[ 1 - \Phi \left( -\frac{\tilde{\mu}_i}{\sigma_v} \right) \right] \right\}$$

Se calculará en R como sigue:

```
> beta <- c(beta0, beta[2:3])
> muTilde <- - epsilon-(sigmav2/sqrt(sigmau2))
> distribucion <- pnorm(-muTilde/sqrt(sigmav2))
> zL <- qnorm((1 - (0.05/2))*(1-distribucion))
> zU <- qnorm((0.05/2)*(1-distribucion))
> L <- exp(-esperanza - zL*sqrt(sigmav2))
> U <- exp(-esperanza - zU*sqrt(sigmav2))
> intervalos <- cbind(L,U)
> intervalos[1:5,]
```

	L	U
1	0.6710692	1.295639
2	1.1601945	1.728332
3	0.8405265	1.446360
4	0.8911244	1.490307
5	0.7560035	1.372360

Donde la salida anterior es la estimación por intervalo de la eficiencia técnica para los cinco primeros productores de la base de datos.

### 4.3. Comparación

El método de los momentos tiene dos problemas. El primero es que, aunque el momento central  $\mu_3$  de las perturbaciones debe ser negativo, es posible que el tercer momento central  $m_3$  de los residuos de los mínimos cuadrados ordinarios sea positivo, lo que implica que  $\hat{\sigma}_u$  es negativo. Esto proporciona un diagnóstico útil, ya que sugiere que el modelo está mal especificado. Si el modelo está mal especificado, es natural que  $\hat{\sigma}_u = 0$ , de donde se llega a la conclusión de que no hay ineficiencia técnica en los datos analizados. El segundo problema del método aparece cuando  $m_3$  es negativo, pero suficientemente grande para causar que  $\hat{\sigma}_v^2 < 0$ . Esto implicaría que  $\hat{\sigma}_v^2 = 0$ , lo que significaría que no hay ruido en los datos. El primer problema ocurriría cuando  $\lambda = \sigma_u/\sigma_v$  es pequeño y el segundo cuando  $\lambda$  es grande.

Por otro lado, el impacto del supuesto sobre el término de ineficiencia es una diferencia entre ambos métodos. En el caso del método de los momentos, la estimación de los parámetros no se ve afectada por un error en el supuesto de la distribución. Como el estimador máximo verosímil utiliza esta información desde el principio, cualquier tipo de mala especificación sobre la distribución de  $u$  afectará simultáneamente a la estimación de los parámetros. Aunque el procedimiento de los momentos genera estimadores consistentes de todos los parámetros del modelo, hay ineficiencia comparado con los estimadores máximo - verosímiles que están basados en la suposición de distribuciones.

El supuesto sobre la distribución de  $u$  es uno de los temas más discutidos. En la práctica es frecuente utilizar la distribución seminormal, aunque no existe un motivo teórico que justifique esta elección.

#### 4.3.1. Resumen de resultados

En este apartado se van a comparar los resultados obtenidos de la base de datos analizada mediante todas las técnicas descritas, tanto de las fronteras de producción determinísticas como de las estocásticas. En la siguiente tabla están recogidas las estimaciones del vector  $\vec{\beta}$ , de  $\sigma_u^2$  y  $\sigma_v^2$  calculados por todos los procedimientos.

MODELO	$\beta_0$	$\beta_1$	$\beta_2$	$\sigma_u^2$	$\sigma_v^2$
P.Lineal	0,2166105	0,4662259	0,6852786	—	—
P.Cuadrática	0,2166105	0,4662259	0,6852786	—	—
COLS	1,084886	0,2805	0,5333	—	—
MOLS	0,5096721	0,2805	0,5333	—	—
MV NN <sup>+</sup> sfa	0,56149436	0,28110414	0,53649916	0,1729516	0,04402588
MV NN <sup>+</sup> frontier	0,5616	0,2811	0,5365	0,1729924	0,0440076
MV Exp sfa	0,44044913	0,28434984	0,54235073	0,05537614	0,05430909
Momentos NN <sup>+</sup>	0,6031973	0,2804926	0,5333064	0,2016561	0,03500786
Momentos Exp	0,4594128	0,2804926	0,5333064	0,04601651	0,06226917

Dentro de las metodologías de fronteras estocásticas, se puede observar que los resultados obtenidos son bastante parecidos, sobretodo de la estimación del vector  $\vec{\beta}$ .

A continuación se presenta otra tabla para resumir los valores que toma la eficiencia técnica calculada con cada uno de los métodos estudiados. Se darán cuatro valores de cada una, que son el valor mínimo que toma la variable, la mediana, la media y el valor máximo:

EFICIENCIA TÉCNICA				
MODELO	Mínimo	Mediana	Media	Máximo
P.Lineal	0,1583	0,4416	0,4853	1,000
P.Cuadrática	0,1583	0,4416	0,4853	1,000
COLS	0,1583	0,4349	0,4544	1,000
MOLS	0,2815	0,7731	0,8077	1,778
MV NN <sup>+</sup> sfa	0,3513	0,7585	0,7406	0,9374
MV NN <sup>+</sup> frontier	0,3513	0,7585	0,7406	0,9374
MV Exp sfa	0,3733	0,8388	0,8093	0,9454
Momentos NN <sup>+</sup>	0,3135	0,7297	0,7161	0,9437
Momentos Exp	0,3956	0,8422	0,8150	0,9414

Para seguir comparando la eficiencia técnica de cada método, se guardarán todas ellas con `data.frame`, es decir, como una colección de datos. También se calculará la matriz de correlaciones entre las eficiencia técnicas de cada método, lo cual permite ver si están o no muy relacionadas. Cuanto más próximo sea el valor a 1 quiere decir que más se parecen entre sí esas variables, mientras que los valores próximos a 0 indican que hay poca relación entre ellas. Los elementos de esta matriz serán redondeados a cuatro decimales con la función `round`.

```
> TE <- data.frame(TE1, TE2, TE3, TE4, TE5, TE6, TE7, TE8, TE9)
> correlacion <- round(cor(TE),4); correlacion
```

	TE1	TE2	TE3	TE4	TE5	TE6	TE7	TE8	TE9
TE1	1.0000	1.0000	0.7750	0.7750	0.7212	0.7212	0.6759	0.7276	0.6628
TE2	1.0000	1.0000	0.7750	0.7750	0.7212	0.7212	0.6759	0.7276	0.6628
TE3	0.7750	0.7750	1.0000	1.0000	0.9182	0.9182	0.8479	0.9314	0.8419
TE4	0.7750	0.7750	1.0000	1.0000	0.9182	0.9182	0.8479	0.9314	0.8419
TE5	0.7212	0.7212	0.9182	0.9182	1.0000	1.0000	0.9800	0.9983	0.9759
TE6	0.7212	0.7212	0.9182	0.9182	1.0000	1.0000	0.9800	0.9983	0.9759
TE7	0.6759	0.6759	0.8479	0.8479	0.9800	0.9800	1.0000	0.9669	0.9995
TE8	0.7276	0.7276	0.9314	0.9314	0.9983	0.9983	0.9669	1.0000	0.9620
TE9	0.6628	0.6628	0.8419	0.8419	0.9759	0.9759	0.9995	0.9620	1.0000

En la matriz anterior se puede observar que la correlación del método uno (programación lineal) con el método dos (programación cuadrática) es 1, y por tanto, tienen las mismas correlaciones con el resto de los métodos, en torno a 0,7. El resto de los métodos están altamente correlados entre ellos, ya que todos los valores son todos muy próximos a 1.

La última forma que se utilizará para comparar las eficiencias técnicas es con representaciones gráficas. Se harán cuatro gráficos, en los cuales se comparan las eficiencias de los métodos que más interesantes pueden resultar, como puede ser el método de mínimos cuadrados corregidos con el de mínimos cuadrados modificados, máxima verosimilitud calculada con el paquete `sfa` frente a la calculada con el paquete `frontier`. Las dos últimas gráficas son las que comparan el método de máxima verosimilitud suponiendo que la distribución de



la ineficiencia es seminormal o exponencial y la última que compara lo mismo, pero con el método de los momentos. En todas ellas se dibujará la recta  $x = y$  para tener en cuenta la escala de las eficiencias técnicas.

Observando la figura 4.1, se puede ver que la información que ofrecen los mínimos cuadrados ordinarios corregidos y modificados sobre la eficiencia técnica es la misma, pero la escala de los cuadrados modificados no es adecuada, puesto que la corrección del intercepto puede desplazar tanto la recta que pocos productores sean técnicamente eficientes.

El resultado obtenido en la figura 4.2 es lógico, puesto que son dos funciones diferentes para resolver lo mismo.

La figura 4.3 muestra que para valores intermedios de la eficiencia técnica, 0,6, la estimación máximo verosímil normal - exponencial es más alta que la otra. No hay mucha desviación respecto de la recta, por lo que no hay grandes diferencias salvo que cuando la ineficiencia se distribuye como una exponencial, la eficiencia toma valores más altos.

Por último, en la figura 4.4 se observa que la eficiencia técnica calculada con el método de los momentos considerando una distribución seminormal de la ineficiencia, tiene valores más bajos comparando cuando la distribución de  $u$  es exponencial.

```
> plot(TE3,TE4, xlab='COLS', ylab='MOLS', bty='l', col=3)
> abline(0,1)
```

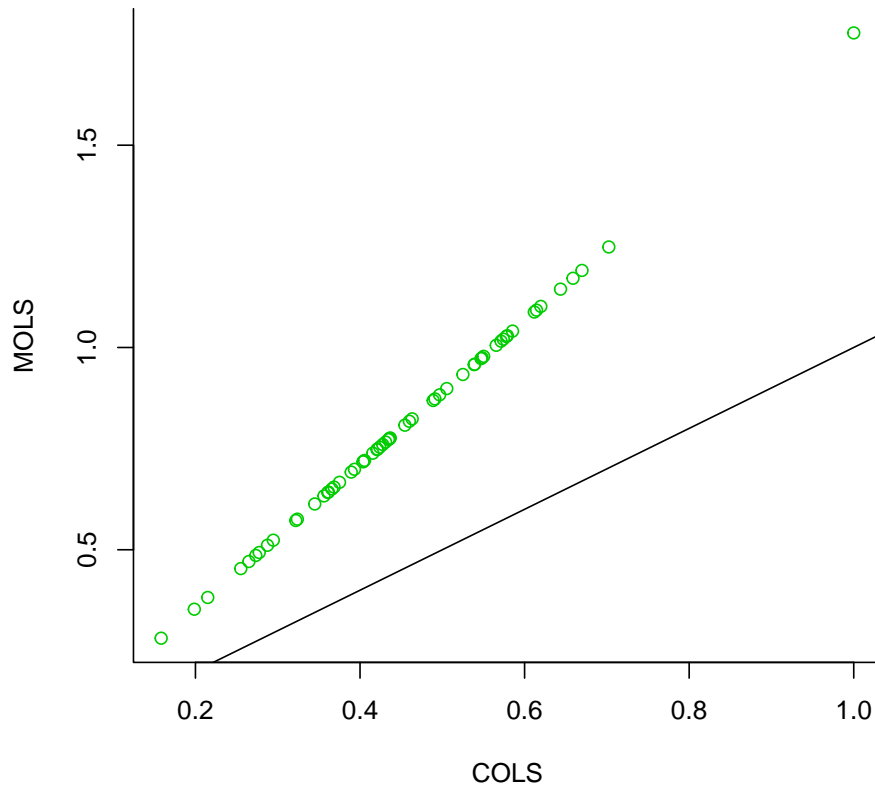


Figura 4.1: Eficiencia Técnica

```
> plot(TE5,TE6, xlab='Máxima Verosimilitud Normal-Seminormal (sfa)',  
+       ylab='Máxima Verosimilitud Normal-Seminormal (frontier)', bty='l', col=4)  
> abline(0,1)
```

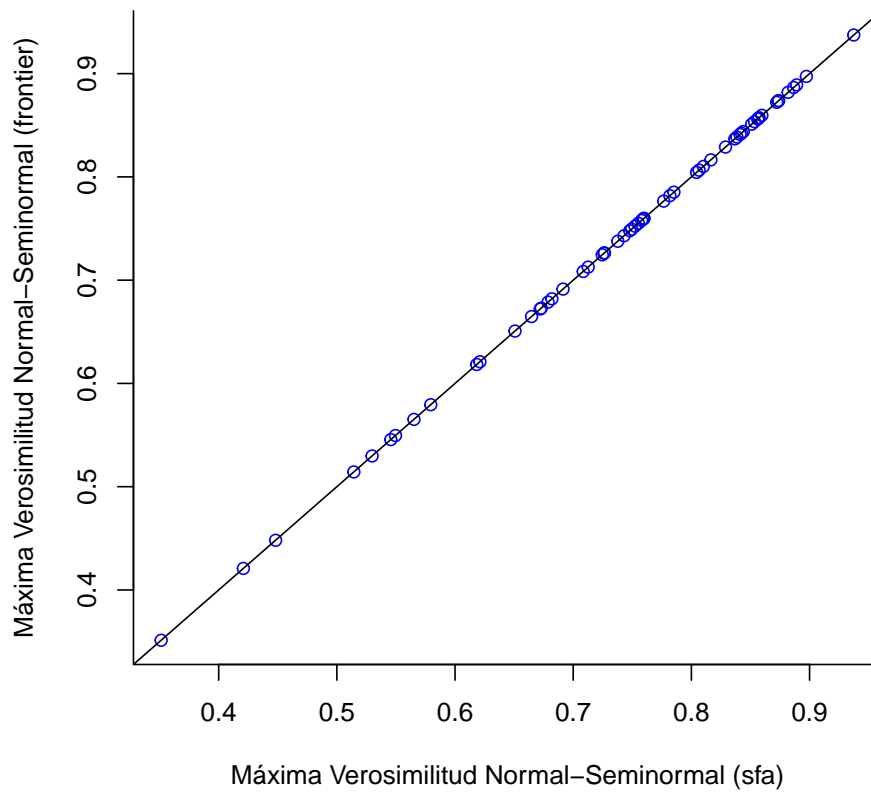


Figura 4.2: Eficiencia Técnica

```
> plot(TE5,TE7, xlab='Máxima Verosimilitud Normal - Seminormal',  
+       ylab='Máxima Verosimilitud Normal- Exponencial', bty='l', col=6)  
> abline(0,1)
```

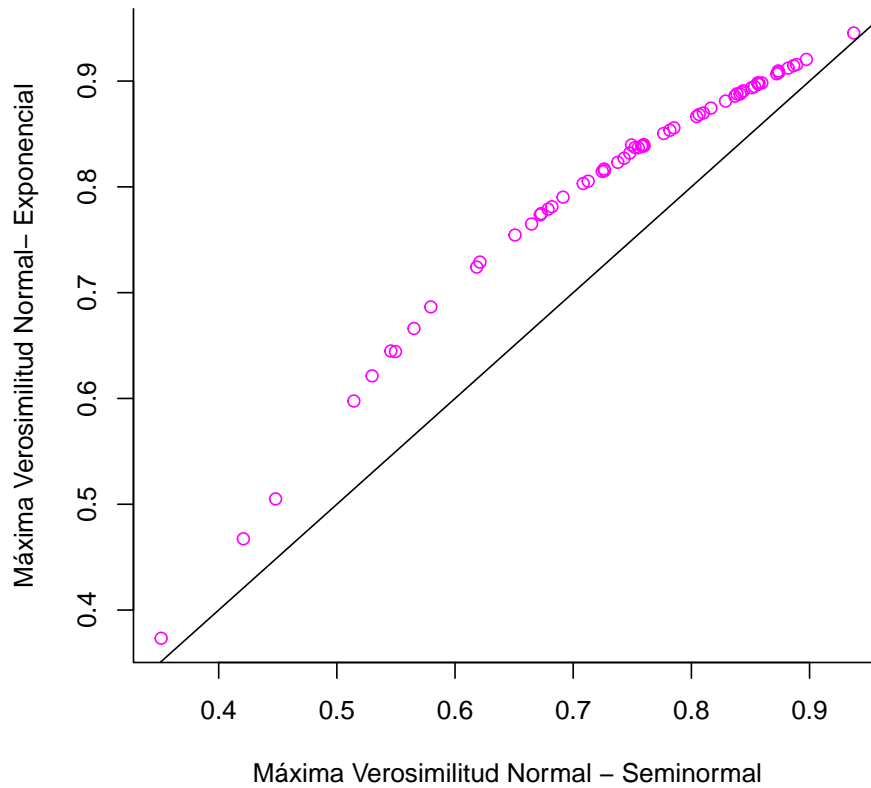


Figura 4.3: Eficiencia Técnica

```
> plot(TE8,TE9, xlab='Método de los momentos Normal - Seminormal',  
+       ylab='Método de los momentos Normal-Exponencial', bty='l', col=5)  
> abline(0,1)
```

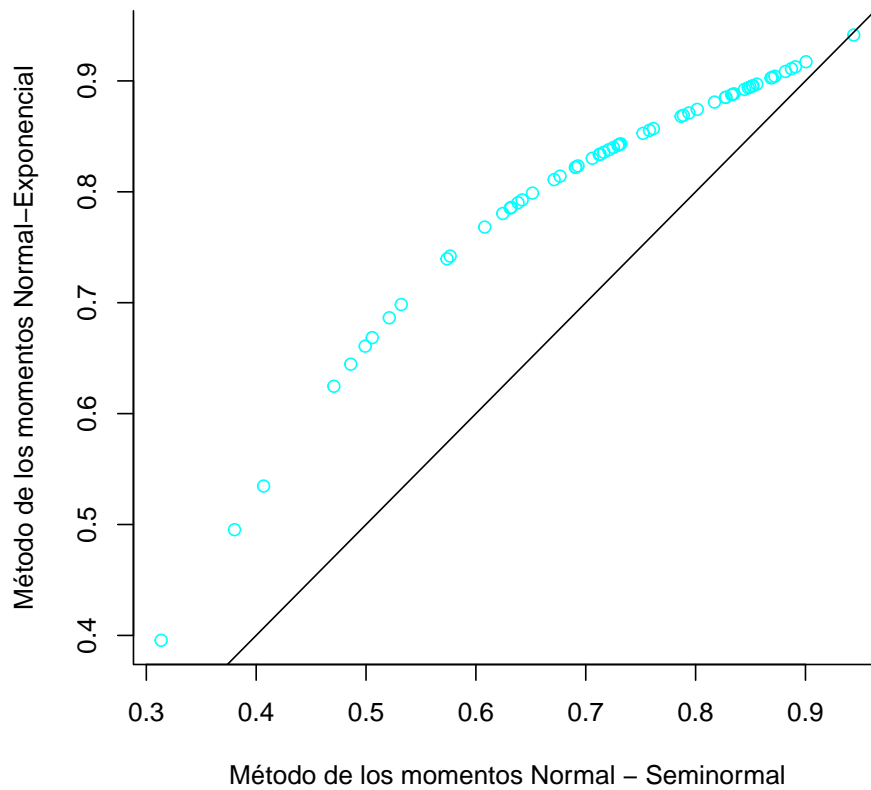


Figura 4.4: Eficiencia Técnica

## Capítulo 5

# Datos de panel

Un panel son observaciones repetidas de cada fenómeno (por ejemplo, de cada productor). Por tanto, los datos de panel recogen observaciones sobre múltiples productores a lo largo de determinados periodos de tiempo. La dimensión temporal enriquece la estructura de los datos y es capaz de aportar información que no aparece en los datos transversales. Este tipo de datos se introducen para evitar las siguientes dificultades que tienen las fronteras estocásticas con datos transversales:

1. La estimación del término de ineficiencia, aunque insesgada es inconsistente, puesto que  $E(u_i|\epsilon_i)$  o  $\text{Moda}(u_i|\epsilon_i)$  no tienden a cero cuando aumenta el tamaño de muestra.
2. Otro problema es que tanto la estimación máximo verosímil de la frontera como la descomposición del error en ruido aleatorio e ineficiencia técnica, requieren de algún supuesto distribucional sobre el término de ineficiencia.
3. Es necesario suponer que la ineficiencia es independiente de los regresores para la estimación máximo - verosímil, lo cual puede no ser un buen supuesto en algunos casos, ya que si un productor conoce su nivel de ineficiencia, es probable que modifique su elección de los recursos.

En teoría, estas tres limitaciones se pueden evitar si se dispone de datos de panel, es decir, si se tienen varias observaciones para cada uno de los productores:

1. La ineficiencia técnica de cada productor puede ser estimada consistentemente a medida que el número de observaciones de cada productor aumenta. Es decir, al añadir más observaciones de cada productor, se genera información no provista con los datos transversales, de esta forma, la eficiencia técnica de cada productor en la muestra puede ser estimada consistentemente cuando  $T \rightarrow +\infty$ , donde  $T$  es el número de observaciones de cada productor.
2. Algunos métodos de estimación con datos de paneles no necesitan realizar fuertes suposiciones sobre la distribución.
3. No hace falta suponer que la ineficiencia es independiente de los regresores, puesto que las observaciones repetidas en la muestra de productores también puede servir como sustituto de la suposición de independencia.

Sin embargo, los modelos de frontera de producción con datos de panel requieren el establecimiento de un supuesto sobre si varía o no la eficiencia técnica a lo largo del tiempo. Es decir, hay modelos en los que varía la eficiencia técnica entre productores, pero es invariante en el periodo de tiempo estudiado. Por otra parte, hay modelos en los que se puede permitir que la eficiencia varíe entre los productores y también a lo largo del tiempo.

En la siguiente sección se van a describir los datos de panel con los que se trabajará para calcular la frontera con los distintos métodos.

## 5.1. Descripción de los datos de panel

Los datos de panel que se utilizarán están en el paquete `frontier` y se llaman `riceProdPhil`. A continuación se carga la librería y el panel:

```
> library(frontier)
> data(riceProdPhil)
```

Este panel contiene datos anuales de 43 pequeños productores de arroz de la región de Tarlac en las Filipinas entre los años 1990 y 1997. El número total de observaciones es 344.

Se van a describir todas las variables de la base de datos. Entre paréntesis están las unidades de medida de cada una de ellas:

- YEARDUM: periodo de tiempo (1 = 1990,  $\dots$ , 8 = 1997).
- FMERCODE: código del agricultor (1,  $\dots$ , 43).
- PROD: producción (toneladas de arroz recién trillado).
- AREA: área plantada (hectáreas).
- LABOR: trabajo empleado (días de hombre de familia y mano de obra contratada).
- NPK: fertilizante utilizado (kg de ingredientes activos).
- OTHER: otros recursos utilizados.
- PRICE: precio del producto (pesos por kg).
- AREAP: precio de alquiler de la renta (pesos por hectárea).
- LABORP: precio del trabajo (pesos por hombre empleado al día).
- NPKP: precio del fertilizante (pesos por kg de ingrediente activo).
- OTHERP: precio de otros recursos (índice de precios implícito).
- AGE: edad del cabeza de familia (años).
- EDYRS: educación del cabeza de familia (años).
- HHSIZE: tamaño de la casa.
- NADULT: número de adultos en la casa.
- BANRT: porcentaje de área clasificada como terreno de Bantog.

Lo primero que hay que hacer es transformar los datos a un formato adecuado para la utilización correcta de los datos de panel. La función que se utiliza se llama `plm.data` que está en el paquete `plm`, que éste a su vez necesita del paquete `formula`, los cuales hay que



instalar previamente. Por tanto, una vez cargadas las librerías se transforma el `riceProdPhil` a datos de panel y se hace un resumen de todas las variables de la misma:

```
> library(plm, formula)
> g <- plm.data( riceProdPhil, c( "FMERCODE", "YEARDUM" ) )
> summary(g)
```

	FMERCODE	YEARDUM	PROD	AREA	LABOR
1	: 8	1 :43	Min. : 0.09	Min. :0.200	Min. : 8.00
2	: 8	2 :43	1st Qu.: 2.79	1st Qu.:1.000	1st Qu.: 49.75
3	: 8	3 :43	Median : 5.11	Median :1.750	Median : 87.50
4	: 8	4 :43	Mean : 6.54	Mean :2.144	Mean :108.34
5	: 8	5 :43	3rd Qu.: 8.92	3rd Qu.:3.400	3rd Qu.:152.00
6	: 8	6 :43	Max. :31.10	Max. :7.000	Max. :437.00
	(Other):296	(Other):86			

	NPK	OTHER	PRICE	AREAP
Min.	: 10.00	Min. : 1.459	Min. :4.500	Min. : 313.9
1st Qu.:	67.72	1st Qu.: 21.352	1st Qu.:5.188	1st Qu.: 2905.2
Median :	142.20	Median : 70.556	Median :6.250	Median : 4666.4
Mean :	189.23	Mean : 125.345	Mean :6.531	Mean : 5289.3
3rd Qu.:	254.35	3rd Qu.: 169.578	3rd Qu.:8.000	3rd Qu.: 6599.5
Max. :	1030.90	Max. :1083.379	Max. :9.000	Max. :27787.8

	LABORP	NPKP	OTHERP	AGE
Min.	: 22.96	Min. : 6.903	Min. : 3.397	Min. :25.00
1st Qu.:	53.27	1st Qu.:12.832	1st Qu.: 13.097	1st Qu.:41.00
Median :	68.86	Median :14.078	Median : 26.758	Median :50.00
Mean :	70.77	Mean :14.885	Mean : 32.836	Mean :49.44
3rd Qu.:	84.72	3rd Qu.:16.190	3rd Qu.: 46.678	3rd Qu.:55.00
Max. :	219.56	Max. :32.396	Max. :105.079	Max. :81.00

EDYRS	HHSIZE	NADULT	BANRAT
Min. : 6.000	Min. : 2.000	Min. : 1.000	Min. : 0.0000
1st Qu.: 6.000	1st Qu.: 4.000	1st Qu.: 2.000	1st Qu.: 0.5200
Median : 6.000	Median : 5.000	Median : 4.000	Median : 0.8200
Mean : 7.244	Mean : 5.026	Mean : 3.849	Mean : 0.7344
3rd Qu.: 10.000	3rd Qu.: 6.000	3rd Qu.: 5.000	3rd Qu.: 1.0000
Max. : 14.000	Max. : 14.000	Max. : 10.000	Max. : 1.0000

El objetivo será estimar la producción a través del área, el trabajo empleado y el fertilizante utilizado (Battese y Coelli, 1992). Por comodidad, se va a definir el logaritmo de las cuatro variables que se van a utilizar posteriormente para estimar la eficiencia técnica:

```
> y <- g$y <- log(g$PROD)
> x1 <- g$x1 <- log(g$AREA)
> x2 <- g$x2 <- log(g$LABOR)
> x3 <- g$x3 <- log(g$NPK)
```

## 5.2. Eficiencia técnica invariante en el tiempo

En esta sección se estudiarán los métodos para calcular fronteras de producción cuando se considera que la eficiencia técnica es invariante a lo largo del tiempo. Suponer que hay  $I$  productores,  $i = 1, \dots, I$ , durante  $T$  periodos de tiempo,  $t = 1, \dots, T$ . Una frontera estocástica de producción con eficiencia técnica invariante en el tiempo, se puede escribir como:

$$\ln y_{it} = \beta_0 + \sum_n \beta_n \ln x_{nit} + v_{it} - u_i$$

donde  $v_{it}$  es el ruido aleatorio y  $u_i \geq 0$  es la ineficiencia técnica.

Notar que la estructura se asume constante a través del tiempo. Este modelo es muy similar al modelo estocástico de datos transversales, la única diferencia es la adición del tiempo en la salida, entrada y ruido. Los parámetros pueden ser estimados y la eficiencia técnica también, por diferentes métodos, como pueden ser modelo de efectos fijos, modelo de efectos aleatorios o máxima verosimilitud.

### 5.2.1. Modelo de efectos fijos

En este modelo se asume que  $v_{it}$  independiente, idénticamente distribuída con media 0 y varianza  $\sigma_v^2$  y está incorrelada con los regresores. No se asume ninguna distribución para  $u_i$ , solo que  $u_i \geq 0$ . Además se permite que esté correlada con los regresores o con los  $v_{it}$ . Como el término  $u_i$  es tratado como efecto fijo, es decir, como un efecto individual de cada empresa, se convierte en parámetro del intercepto específico de cada productor que se estima a lo largo del vector  $\vec{\beta}$ . El modelo se escribe como:

$$\ln y_{it} = \beta_{0i} + \sum_n \beta_n \ln x_{nit} + v_{it}$$

donde  $\beta_{0i} = (\beta_0 - u_i)$  son los interceptos específicos de cada productor. Después de la estimación de  $\vec{\beta}$  por mínimos cuadrados ordinarios, hay que aplicar la normalización  $\hat{\beta}_0 = \max_i \{\hat{\beta}_{0i}\}$  y la estimación de  $u_i$  viene dada por:

$$\hat{u}_i = \hat{\beta}_0 - \hat{\beta}_{0i}$$

lo cual asegura que  $\hat{u}_i \geq 0$ . La estimación de la eficiencia técnica específica de cada productor se calcula como

$$TE_i = \exp\{-\hat{u}_i\}$$

En el modelo de efectos fijos, se asume que al menos un productor es 100% técnicamente eficiente, y que la eficiencia técnica del resto de productores se mide en relación con el productor técnicamente eficiente.

La estimación de los parámetros del vector  $\vec{\beta}$  es consistente cuando  $I \rightarrow \infty$  ó cuando  $T \rightarrow \infty$ , y la propiedad de consistencia no requiere que  $u_i$  estén incorreladas con los regresores.

Este modelo de efectos fijos es simple de implementar y proporciona estimaciones consistentes de la eficiencia técnica de cada productor. Sin embargo, el problema es que intenta capturar la variación de los productores considerando la eficiencia técnica invariante en el tiempo. Esto también captura los efectos de todos los fenómenos que varían a través de los productores, pero que son invariantes en el tiempo para cada productor. Esta confusión de variación en la eficiencia técnica con la variación en otros efectos está incluida en los regresores del modelo. Este problema, motiva el interés por el siguiente modelo.

### 5.2.2. Modelo de efectos aleatorios

En este modelo se considera que  $u_i$  no son fijas, es decir, que están distribuidas aleatoriamente con media y varianza constantes, pero se asume que están incorrelados con los regresores y con los  $v_{it}$ . En este caso tampoco se asume una distribución de  $u_i$ , pero se mantiene que sea positiva. Al igual que antes  $v_{it}$  se asume con media cero y varianza constante. El modelo viene dado por la siguiente ecuación:

$$\begin{aligned}\ln y_{it} &= [\beta_0 - E(u_i)] + \sum_n \beta_n \ln x_{nit} + v_{it} - [u_i - E(u_i)] \\ &= \beta_0^* + \sum_n \beta_n \ln x_{nit} + v_{it} - u_i^*\end{aligned}$$

donde la suposición de que  $u_i$  son aleatorios permite que algún  $x_{nit}$  sea invariante en el tiempo, es decir, este modelo permite regresores invariantes en el tiempo. La estimación de los parámetros de  $\vec{\beta}$  se hace con el método de mínimos cuadrados generalizados. Observar que  $\beta_0$  no depende de  $i$ , ya que  $E(u_i)$  es una constante positiva, por lo que sólo hay un término de intercepto para estimar. Una vez que el parámetro  $\beta_0^*$  y el vector  $\vec{\beta}$  son estimados, el término  $\hat{u}_i^*$  se calcula como la media de los residuos:

$$\hat{u}_i^* = \frac{1}{T} \sum_t \left( \ln y_{it} - \hat{\beta}_0^* - \sum_n \hat{\beta}_n \ln x_{nit} \right)$$

Las estimaciones de  $u_i$  son obtenidas con la siguiente normalización:

$$\hat{u}_i = \max_i \{ \hat{u}_i^* \} - \hat{u}_i^*$$

dichas estimaciones son consistentes cuando  $I \rightarrow \infty$  y  $T \rightarrow \infty$ .

La eficiencia técnica se obtiene como:

$$TE_i = \exp\{-\hat{u}_i\}$$

Al igual que en el modelo de efectos fijos, se asume que al menos un productor es 100% técnicamente eficiente, y la eficiencia técnica del resto de productores se mide en relación con el productor técnicamente eficiente.

Con estos dos métodos se demuestra que los datos de panel permiten evitar las suposiciones de las distribuciones de los datos transversales. Sin embargo, también se podrían asumir

distribuciones en este contexto, y el método máximo verosímil sería factible. Este método se describe en el último capítulo.

### 5.2.3. Ejemplos

En este apartado se calcularán las eficiencias técnicas de los datos de panel `riceProdPhil` suponiendo que ésta es invariante en el tiempo. Se hallarán con el modelo de efectos fijos y con el modelo de efectos aleatorios. Al final se hace una comparación de los resultados obtenidos.

#### Modelo efectos fijos:

Los datos de panel que se utilizarán ya están guardados en el formato correspondiente para trabajar con ellos. El primer paso del método es una estimación del vector  $\vec{\beta}$  por el método de mínimos cuadrados ordinarios. Como es una estimación de datos de panel, se utiliza la función `plm` del paquete `plm`, en vez de la función `lm` que se utilizaba para los datos transversales.

```
> library(plm)
> rice1 <- plm ( y ~ x1 + x2 + x3, data = g )
> rice1$coef
```

```
          x1          x2          x3
0.5407094 0.2353918 0.1962556
```

Como se puede observar en la salida, no se obtiene la estimación del intercepto, entonces la hay que calcular. Se tenía que el modelo es:

$$\ln y_{it} = \beta_{0i} + \sum_n \beta_n \ln x_{nit} + v_{it}$$

donde  $v_{it}$  son los residuos de la estimación anterior. Despejando  $\beta_{0i}$  de la ecuación se tiene que:

$$\beta_{0i} = \ln y_{it} - \sum_n \beta_n \ln x_{nit} - v_{it}$$

lo cual se puede calcular en R como sigue:

```

> v <- rice1$residuals
> X <- cbind(x1, x2, x3)
> beta <- rice1$coef
> beta0 <- y - v - X%%beta
> beta0i <- beta0[(0:42)*8+1]

```

Una vez que se obtiene  $\hat{\beta}_{0i} = \hat{\beta}_0 - \hat{u}_i$ , es necesario separar los términos. Para ello, se ha descrito anteriormente que  $\hat{\beta}_0 = \max_i\{\beta_{0i}\}$ , por lo que  $\hat{u}_i = \hat{\beta}_0 - \hat{\beta}_{0i}$ . Se calculan a continuación:

```

> beta_0 <- max(beta0i); beta_0

[1] -0.3571068

> u <- beta_0 - beta0i

```

El último paso es calcular la eficiencia técnica de cada productor como la exponencial de menos la ineficiencia. Se hará también un resumen de la variable:

```

> te1 <- exp(-u); te1

[1] 0.6233797 0.8917268 0.6084181 0.8119558 0.7410469 0.6868099 0.8768274
[8] 0.7104078 0.7590227 0.7473460 0.4829167 1.0000000 0.6682597 0.7527929
[15] 0.6000349 0.8144714 0.7915291 0.8764250 0.7506852 0.8337416 0.6628714
[22] 0.7244621 0.5888186 0.5859059 0.7139216 0.7482915 0.6790786 0.9399869
[29] 0.5737729 0.4805287 0.8358082 0.7367054 0.7942230 0.3742712 0.8365828
[36] 0.5792508 0.8897544 0.8912732 0.6087871 0.5580206 0.8800406 0.7902515
[43] 0.5957380

> summary(te1)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.3743  0.6086  0.7410  0.7232  0.8241  1.0000

```

La media de la eficiencia técnica es 0,7219, y al menos uno de los productores es 100% técnicamente eficiente, porque el valor máximo que toma la variable es 1.

### Modelo efectos aleatorios:

El primer paso es una estimación de los parámetros por el método de mínimos cuadrados generalizados. Puesto que se está trabajando con datos de panel, se necesita instalar el paquete específico llamado `pglm`. La función se llama igual que el paquete y se utiliza como sigue:

```
> library(pglm)
> rice2 <- pglm ( y ~ x1 + x2 + x3, family = gaussian, data = g )
> coef(rice2)

(Intercept)          x1          x2          x3          sd.mu          sd.eps
-1.1498548  0.4571030  0.2979375  0.2407266  0.1584102  0.2900830
```

Una vez obtenida la estimación de los parámetros, hace falta calcular la estimación de  $\hat{u}_i$  como se había descrito. Los datos de panel están todos en un vector columna, por lo que para calcular la suma de las estimaciones a lo largo de los periodos de tiempo, es más cómodo poner las estimaciones en forma matricial:

```
> X <- cbind(1, x1, x2, x3)
> beta <- c(-1.14985, 0.45710, 0.29794, 0.24073)
> estimacion <- y - X%%beta
> u1 <- NULL
> for (i in 0:7){
+   u1 <- cbind(u1, estimacion[(43*i)+(1:43)])
+ }
```

A continuación se hace la media por filar de la matriz anterior, que es la media a lo largo de los periodos de tiempo. Se calcula el máximo de todos los valores resultantes y se resta a cada una de las estimaciones de  $u_i$  que se tenían:

```
> u2 <- apply(u1, 1, mean)
> m <- max(u2)
> u <- m - u2
```

Como la estimación de  $\hat{\beta}_0$  era la estimación de  $\beta_0$  menos la esperanza de la ineficiencia, es necesario corregir ese término, es decir:

```
> beta00 <- coef(rice2)[1]
> esperanza <- mean(u)
> beta0 <- beta00 + esperanza; beta0
```

```
(Intercept)
-0.9041601
```

El último paso es calcular la eficiencia técnica y hacer un resumen de la misma:

```
> te2 <- exp(-u); te2

[1] 0.6635750 0.8132585 1.0000000 0.8294353 0.7083608 0.8172326 0.7152162
[8] 0.8064857 0.8741356 0.8385329 0.6538496 0.7260750 0.7943092 0.7077068
[15] 0.7360407 0.8258187 0.7767202 0.7343419 0.6351499 0.6615718 0.8226449
[22] 0.8222123 0.7659872 0.7974749 0.6683211 0.8205978 0.8416914 0.9457226
[29] 0.7910720 0.8311001 0.8982443 0.7354619 0.8112710 0.7296581 0.7596416
[36] 0.7879896 0.7144412 0.7629204 0.7276745 0.8376948 0.9897529 0.8278559
[43] 0.8017491
```

```
> summary(te2)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.6351  0.7287  0.7943  0.7863  0.8268  1.0000
```

En vista a los resultados obtenidos, se puede concluir que algún productor es 100 % técnicamente eficiente y que la media de la variable es 0,7843.

## Comparación

En la siguiente tabla - resumen están recogidas las estimaciones del vector  $\vec{\beta}$  calculado por ambos procedimientos:



MODELO	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_2$
Efectos fijos	-0,3571068	0,5407094	0,2353918	0,1962556
Efectos aleatorios	-0,9041601	0,4571030	0,2979375	0,2407266

A continuación se presenta otra tabla para resumir los valores que toma la eficiencia técnica calculada con cada uno de los métodos estudiados.

EFICIENCIA TÉCNICA				
MODELO	Mínimo	Mediana	Media	Máximo
Efectos fijos	0,3743	0,7410	0,7219	1,0000
Efectos aleatorios	0,6351	0,7943	0,7863	1,0000

# Capítulo 6

## Extensiones del modelo

En este capítulo se describen brevemente algunas extensiones de los modelos de datos de panel. Se empieza con la descripción de método de máxima verosimilitud para eficiencia técnica invariante en el tiempo y después se define el modelo de fronteras de producción cuando se considera que la eficiencia técnica varía a través del tiempo.

### 6.1. Eficiencia técnica invariante en el tiempo

#### 6.1.1. Método de máxima verosimilitud

La estimación máximo verosímil de un modelo de frontera de producción con datos de panel es muy similar al procedimiento descrito para los datos transversales. Se comienza asumiendo los siguientes supuestos distribucionales sobre las componentes de error del modelo:

$$\ln y_{it} = \beta_0 + \sum_n \beta_n \ln x_{nit} + v_{it} - u_i$$

donde

- $v_{it} \rightsquigarrow N(0, \sigma_v^2)$  independiente e idénticamente distribuída.
- $u_i \rightsquigarrow N^+(0, \sigma_u^2)$  independiente e idénticamente distribuída, es decir:

$$f(u) = \frac{2}{\sqrt{2\pi}\sigma_u} \cdot \exp\left\{-\frac{u^2}{2\sigma_u^2}\right\}$$

- $u_i$  y  $v_{it}$  independientes entre sí y de los regresores.

La única diferencia con el modelo de datos transversales es que la componente ruido  $v_{it}$  varía a través del tiempo y también de los productores. La función de densidad de  $\vec{v} = (v_1, \dots, v_T)'$  viene dada por:

$$f(\vec{v}) = \frac{1}{(2\pi)^{T/2}\sigma_v^T} \cdot \exp \left\{ -\frac{(\vec{v})'\vec{v}}{2\sigma_v^2} \right\}$$

Dada la independencia, la función de densidad conjunta de  $u$  y  $\vec{v}$  es la siguiente:

$$f(u, \vec{v}) = \frac{2}{(2\pi)^{(T+1)/2}\sigma_u\sigma_v^T} \cdot \exp \left\{ -\frac{u^2}{2\sigma_u^2} - \frac{(\vec{v})'\vec{v}}{2\sigma_v^2} \right\}$$

Por lo que la función de densidad conjunta de  $u$  y  $\vec{\epsilon} = (v_1 - u, \dots, v_T - u)'$  es:

$$f(u, \vec{\epsilon}) = \frac{2}{(2\pi)^{(T+1)/2}\sigma_u\sigma_v^T} \cdot \exp \left\{ -\frac{(u - \mu_*)^2}{2\sigma_*^2} - \frac{(\vec{\epsilon})'\vec{\epsilon}}{2\sigma_v^2} + \frac{\mu_*^2}{2\sigma_*^2} \right\}$$

donde

$$\begin{aligned} \mu_* &= -\frac{T\sigma_u^2\bar{\epsilon}}{\sigma_v^2 + T\sigma_u^2} \\ \sigma_*^2 &= \frac{\sigma_u^2\sigma_v^2}{\sigma_v^2 + T\sigma_u^2} \\ \bar{\epsilon} &= \frac{1}{T} \sum_i \epsilon_{it} \end{aligned}$$

La función de densidad marginal de  $\epsilon$  se calcula como la integral respecto de  $u$  de la expresión anterior, es decir:

$$\begin{aligned} f(\vec{\epsilon}) &= \int_0^\infty f(u, \vec{\epsilon}) du \\ &= \int_0^\infty \frac{2}{(2\pi)^{(T+1)/2}\sigma_u\sigma_v^T} \cdot \exp \left\{ -\frac{(u - \mu_*)^2}{2\sigma_*^2} - \frac{(\vec{\epsilon})'\vec{\epsilon}}{2\sigma_v^2} + \frac{\mu_*^2}{2\sigma_*^2} \right\} du \\ &= \frac{2(1 - \Phi(-\mu_*/\sigma_*))}{(2\pi)^{T/2}\sigma_v^{T-1}(\sigma_v^2 + T\sigma_u^2)^{1/2}} \cdot \exp \left\{ -\frac{(\vec{\epsilon})'\vec{\epsilon}}{2\sigma_v^2} + \frac{\mu_*^2}{2\sigma_*^2} \right\} \end{aligned}$$

La función de verosimilitud para una muestra de  $I$  productores observados en  $T$  periodos de tiempo es:

$$\begin{aligned} \ln L &= \ln(\Pi_{i=1}^I f(\vec{\epsilon})) = \sum_{i=1}^I (\ln f(\vec{\epsilon})) \\ &= I \ln \frac{2}{(2\pi)^{T/2}} - \frac{I(T-1)}{2} \ln \sigma_v^2 - \frac{I}{2} \ln(\sigma_v^2 + T\sigma_u^2) + \end{aligned}$$

$$+ \sum_i \ln([1 - \Phi(-\mu_{*i}/\sigma_*)]) - \frac{1}{2\sigma_v^2} \sum_i \epsilon'_i \epsilon_i + \frac{1}{2} \sum_i \left( \frac{\mu_{*i}}{\sigma_*} \right)^2$$

La función anterior puede ser maximizada respecto de los parámetros  $\vec{\beta}$ ,  $\sigma_v^2$  y  $\sigma_u^2$  para obtener la estimación máximo - verosímil de cada uno de ellos. Como analíticamente no se puede resolver, se hace una maximización numérica para obtener dichas las estimaciones.

El siguiente paso es obtener una estimación de la eficiencia técnica invariante para cada productor. Para ello se empieza calculando la siguiente distribución condicionada:

$$\begin{aligned} f(u|\vec{\epsilon}) &= \frac{f(u, \vec{\epsilon})}{f(\vec{\epsilon})} = \frac{\frac{2}{(2\pi)^{(T+1)/2} \sigma_u \sigma_v^T} \cdot \exp \left\{ -\frac{(u-\mu_*)^2}{2\sigma_*^2} - \frac{(\vec{\epsilon})'\vec{\epsilon}}{2\sigma_v^2} + \frac{\mu_*^2}{2\sigma_*^2} \right\}}{\frac{2(1-\Phi(-\mu_*/\sigma_*))}{(2\pi)^{T/2} \sigma_v^{T-1} (\sigma_v^2 + T\sigma_u^2)^{1/2}} \cdot \exp \left\{ -\frac{(\vec{\epsilon})'\vec{\epsilon}}{2\sigma_v^2} + \frac{\mu_*^2}{2\sigma_*^2} \right\}} \\ &= \frac{1}{(2\pi)^{1/2} \sigma_* (1 - \Phi(-\mu_*/\sigma_*))} \cdot \exp \left\{ -\frac{(u - \mu_*)^2}{2\sigma_*^2} \right\} \end{aligned}$$

Lo cual es la función de densidad de una variable  $N^+(\mu_*, \sigma_*^2)$ . Tanto la media como la moda pueden servir de estimadores puntuales de la ineficiencia técnica, y después calcular la eficiencia como:

$$TE_i = \exp\{-\hat{u}_i\}$$

## 6.2. Eficiencia técnica variante en el tiempo

Suponer que la eficiencia técnica es invariante en el tiempo puede resultar restrictivo en ocasiones. En esta sección se define solamente el modelo para estimar la eficiencia técnica cuando se considera variante en el tiempo, teniendo en cuenta el coste que supone la estimación de parámetros adicionales. El modelo se escribe como:

$$\begin{aligned} \ln y_{it} &= \beta_{0t} + \sum_n \beta_n \ln x_{nit} + v_{it} - u_{it} \\ &= \beta_{it} + \sum_n \beta_n \ln x_{nit} + v_{it} \end{aligned}$$

donde  $\beta_{0t}$  es el intercepto de la frontera de producción de todos los productores en el periodo  $t$ ,  $\beta_{it} = \beta_{0t} - u_{it}$  es el intercepto del productor  $i$  en el periodo  $t$ , y el resto de variables ya estaban previamente definidas.

Es necesario establecer la estructura que indica la evolución de la eficiencia técnica en el tiempo. Los procedimientos de estimación son el método de máxima verosimilitud y el método de los momentos, pero en este trabajo no se profundizará en esta parte.

## Capítulo 7

# Conclusiones y puntos abiertos

En este trabajo se han investigado los diferentes enfoques de la teoría de modelos de las fronteras de producción. En la primera parte se han detallado diferentes técnicas para calcular las fronteras de producción determinísticas. Dadas sus limitaciones, por no tener en cuenta los muchos factores aleatorios que pueden estar presentes, parece imprescindible estudiar modelos más realistas; se introducen así las fronteras estocásticas.

Ninguna de las referencias disponibles sobre este tema muestra explícitamente los cálculos de las distribuciones marginales ni las condicionadas. En este trabajo se presentan dichos cálculos con todo detalle, lo que ha permitido descubrir algunos errores en las fórmulas publicadas en las referencias fundamentales.

Por otra parte, dichas referencias sugieren usar la moda de las distribuciones condicionadas como punto medio de los intervalos de confianza de la eficiencia técnica en las fronteras estocásticas. Sin embargo, para tamaños de muestra habituales los intervalos así estimados ni siquiera cumplen la condición más básica, que contengan a las estimaciones puntuales correspondientes. Para resolver el problema, en este trabajo se propone usar la esperanza de las distribuciones como punto medio del intervalo, lo que asegura la inclusión en el mismo de las estimaciones puntuales, basadas en dicha esperanza.

Como última aportación, cabe destacar la implementación realizada en R para estudiar los modelos de datos de panel, debido a que los paquetes disponibles en el entorno estadístico con el que se trabaja siguen modelos distintos de los de las referencias fundamentales, lo que

impedía hacer un estudio directo de las propiedades de los estimadores de los parámetros de la frontera y de las eficiencias técnicas de los productores.

En lo que se refiere a las posibles extensiones de este trabajo, pueden destacarse las siguientes líneas abiertas:

- Aunque en este trabajo se ha profundizado más en las estimaciones de la eficiencia técnica de cada productor, también se ha comprobado que las estimaciones de todos los parámetros de la frontera para datos transversales son inaceptables cuando el tamaño muestral es pequeño. Se podría estudiar el comportamiento de las estimaciones del vector  $\vec{\beta}$  y de los parámetros  $\sigma_u^2$  y  $\sigma_v^2$  de los métodos de máxima verosimilitud y el de los momentos. Para ello habría que hacer simulaciones con diferentes combinaciones de los valores de los parámetros, realizar las estimaciones pertinentes y buscar las causas de los problemas que presentan.
- Se podría seguir realizando el desarrollo minucioso de las fórmulas presentes descritas en las referencias fundamentales, trabajando con las extensiones descritas para los datos de panel, tanto cuando se considera que la eficiencia técnica es invariante en el tiempo como cuando se considera variante.

# Bibliografía

- [1] Aigner D., Lovell C.A.K, Schmidt P.: *Formulation and estimation of stochastic frontier production function models*, North-Holland Publishing Company (1977).
- [2] Behr A., Tente S.: *Stochastic frontier analysis by means of maximum likelihood and the method of moments*, Banking and Financial Studies (2008).
- [3] Canay I.A.: *Fronteras de eficiencias estocásticas: comparación internacional de distribuidoras eléctricas*, (2001).
- [4] Coelli T., Henningsen A.: *Stochastic frontier analysis - Package 'frontier'*, (2014).
- [5] Croissant Y., Millo G.: *Panel Data Econometrics in R: The plm Package*, (2008).
- [6] Croissant Y.: *Panel generalized linear model - Package 'pglm'*, (2015).
- [7] Farrell M.J.: *The measurement of Productive Efficiency*, (1957).
- [8] Jondrow J., Lovell C.A.K, Materov I.S., Schmidt P.: *On the estimation of technical inefficiency in the stochastic frontier production function model*, North-Holland Publishing Company (1981).
- [9] Kumbhakar S.C, Lovell C.A.K: *Stochastic Frontier Analysis*, Cambridge University Press (2000). Iberoamericana (1988).
- [10] Murillo C.: *Contribuciones al análisis estocástico de la eficiencia técnica mediante métodos no paramétricos*, (2002).
- [11] Pérez M.C.: *Modelos de frontera estocástica. Distribución de la ineficiencia*, (2012).



[12] Straub A.: *Stochastic frontier analysis - Package 'sfa'*, (2014).