



Munich Personal RePEc Archive

# Maximum Likelihood and Bayesian Estimation of Detection Probabilities

Merino Troncoso, Carlos

UNED

20 January 2020

Online at <https://mpra.ub.uni-muenchen.de/110264/>  
MPRA Paper No. 110264, posted 18 Oct 2021 08:03 UTC

## Maximum Likelihood and Bayesian Estimation of Detection Probabilities

### *Abstract*

*This article estimates the probability of detection of cartels using a sample of cartels discovered between 2011 and 2016. This type of study has been questioned for using a sample considered biased of all cartels. This article uses two alternative methodologies to conclude, like (Harrington and Wei 2017), that there is no such bias, and that the estimation of the probability of detection around 15% using the traditional Bryant-Eckard method is acceptable, although it must be interpreted as a probability of disappearance of the cartels, which only equals the probability of detection when we assume that all the cartels disappear by detection. We must therefore consider that the result is an upper limit below which the real probability of detection should be.*

### **1.- Introducción**

Los cárteles suponen un importante daño para la sociedad en términos de mayores precios y menor oferta, mayor gasto público e ineficiencia económica, motivos entre otros que justifican su consideración de ilegales y su persecución en la mayoría de países.

Para calcular el efecto que los cárteles causan en la economía y diseñar sanciones disuasorias contra estas infracciones, es necesario disponer de una estimación de varios parámetros: la magnitud del beneficio ilícito, el overcharge y la probabilidad de detección.

El objetivo de este artículo es estimar el tercer parámetro: la probabilidad de detección, y la aplicación del mismo al caso particular de España. Sorprendentemente, no existen estimaciones sobre la probabilidad de detección aplicada al caso de España. Este artículo trata de cubrir esta laguna.

Efectivamente, para estimar sanciones disuasorias frente a cárteles, es decir, aquellas capaces de evitar la formación o continuidad de un cártel el beneficio ilícito derivado del acuerdo debe ser inferior a la multa esperada multiplicada por la probabilidad de detección.

Según (García-Verdugo, Merino Troncoso y Gómez Cruz 2016) la sanción óptima es *“aquella que desincentiva a una empresa a participar en un cártel, y esto se logra cuando no hay ganancia neta esperada por participar en una conducta prohibida, es*

decir, cuando la ganancia ilícita esperada de entrar en el cartel es menor que la pérdida esperada. Por tanto, para obtener un valor de referencia para las sanciones es necesario comenzar en cualquier caso con la estimación del beneficio que proporciona la pertenencia al cártel”.

La expresión del beneficio ilícito, o diferencia entre el beneficio obtenido por los miembros del cártel y el beneficio que habrían obtenido en condiciones de competencia, según (Buccirossi y Spagnolo 2005) sería:

$$\Delta\pi = f(S; m, k, \varepsilon) = k \frac{(1 + m)(1 - \varepsilon k) - \varepsilon m}{(1 + m)(1 + k)(1 - \varepsilon k)} S$$

En ella se observa que el beneficio ilícito es una proporción del volumen total de ventas en el mercado afectado  $S$ , y ese porcentaje depende de tres parámetros:  $m$  es el margen de beneficio en el escenario competitivo (“*competitive mark-up*”),  $k$  es el porcentaje de aumento del precio como consecuencia de la cartelización (“*cartel overcharge*”), y  $\varepsilon$  es la elasticidad-precio de la demanda (en valor absoluto)<sup>1</sup>.

En (García Verdugo, Merino Troncoso y Gómez Cruz 2016), también se analiza la formación y la estabilidad de un cártel en un entorno dinámico en el que la probabilidad anual de detección es  $\lambda$ ; si el cártel es detectado será disuelto y cada empresa pagará una multa  $F^2$ . El beneficio de no cumplir los acuerdos del cártel es al menos ligeramente superior que el beneficio de cumplirlos, ya que la empresa que los incumple fijaría un precio por debajo del que ha fijado el cártel, aumentando así sus ventas a costa de sus competidores cartelizados; a su vez, suponen que el beneficio de pertenecer al cártel es mayor que el beneficio en una situación de competencia.

La expresión de la multa disuasoria óptima desde una perspectiva dinámica, es decir, la condición suficiente para que la multa tenga efecto disuasorio es:

$$F \geq \frac{\Delta\pi}{\lambda}$$

Una multa igual o superior a  $\frac{\Delta\pi}{\lambda}$  sería suficiente para que al menos alguna de las empresas del sector incumpliera el acuerdo de colusión, mientras que las demás responderían escogiendo la estrategia competitiva y el cártel se dismantelaría. Según la expresión anterior, una multa sería disuasoria si fuera al menos igual al beneficio ilícito

---

<sup>1</sup> Si se utiliza el valor de la elasticidad con signo negativo, la expresión del beneficio ilícito sería análoga pero con un signo positivo en los términos donde aparece la elasticidad.

<sup>2</sup> Tanto la probabilidad de detección como la sanción se asumen constantes a lo largo del tiempo.

anual multiplicado por un factor equivalente a la inversa de la probabilidad anual de detección.

En definitiva, se comprueba que la probabilidad de detección de la infracción debe jugar necesariamente un papel fundamental en la determinación de la multa para que ésta tenga un efecto disuasorio. En un informe de la OCDE<sup>3</sup> sobre la lucha contra los cárteles se subraya con claridad:

*“Effective sanctions against cartels should take into account not only the amount of gain realized by the cartel but also the probability that any cartel will be detected and prosecuted. Because not all cartels are detected, the financial sanction against one that is detected should exceed the gain actually realized by the cartel”.*

Como hemos dicho la probabilidad de detección es difícil de estimar, por un lado, porque la muestra de cárteles disponible está afectada por un sesgo de selección (Heckman 1979), ya que sólo se dispone de datos sobre el número de cárteles que se han detectado, pero se desconoce cuántos siguen operando en el mercado.

Por otro, en general, sólo podemos estimar un límite inferior de la duración de un cártel porque no tenemos la certeza de cuando ha sido con exactitud el nacimiento del cartel, sólo podemos inferirla a partir de las pruebas en el expediente mientras que sí conocemos con exactitud la fecha del descubrimiento y muerte del cartel. Las duraciones de cárteles descubiertos son por tanto infraestimaciones de la duración real de los cárteles que a su vez son sobreestimaciones de la probabilidad de detección.

En este trabajo, primero realizamos una revisión de la literatura existente sobre probabilidades de detección de cárteles para después analizar los datos disponibles de cárteles en España, sobre los que aplicamos, en primer lugar, la metodología de Bryant y Eckard a los datos de cárteles en España para obtener un resultado de probabilidad de detección. En segundo lugar, aplicamos a los mismos datos de España la metodología más reciente de Harrington que se basa en el teorema de Bayes, para acabar con las principales conclusiones y posibles líneas de investigación futuras.

## **2.- Revisión de la literatura sobre la probabilidad de detección de cárteles**

En general, los estudios previos tratan de explicar la probabilidad de detección de forma exógena utilizando modelos matemáticos teóricos.

---

<sup>3</sup> Development. Competition Committee. *Fighting hard-core cartels: Harm, effective sanctions and leniency programmes*. OECD PUBLISHING, (2002), Report on the nature and impact of hard core cartels and sanctions against cartels under national competition laws.

(Bryant et al. 1991) fueron los primeros en estimar la probabilidad anual de detección de acuerdos de precios en EEUU y concluyeron que oscilaban entre un 13% y un 17%, a partir de una base de datos de 184 cárteles detectados en EEUU entre 1961 y 1988. En su artículo consideran que los cárteles nacen y desaparecen al ser detectados siguiendo un proceso estocástico homogéneo de nacimiento y muerte, metodología frecuentemente utilizada en biología y ecología, para describir dinámicas poblacionales. Estos autores utilizan la siguiente intuición: si las duraciones de las infracciones son en la mayoría de los casos de corta duración frente a algunas pocas de larga duración, entonces la probabilidad de detección debe ser elevada, y el número total de cárteles activos debe ser reducido. Otros autores han seguido esta metodología para estimar el número de cárteles activos y la probabilidad de detección en diferentes regiones y países.

(Combe, Monnier, and Legal 2008) siguen la misma línea que el modelo de detección de cárteles de Bryant y Eckard, utilizando en este caso datos de la UE sobre duraciones de 86 cárteles entre 1969 a 2008. Llegan a unos resultados similares: la probabilidad anual de detección se encuentra entre un 12,9 y 13,3%.

Otros autores utilizan estos porcentajes directamente en sus artículos para estimar la sanción óptima contra cárteles, como, por ejemplo, (Katsoulacos and Ulph 2013) o (Harrington 2014). Como hemos visto, en esta línea de investigación, el cálculo de la probabilidad de detección no tiene en cuenta el comportamiento de las empresas ni de la autoridad de competencia, sino que la creación y destrucción de los cárteles sigue un proceso matemático de creación y destrucción.

Si bien lo habitual es no contar con los datos sobre cárteles no descubiertos, la excepción es (Hyytinen et al. 2010) que sí han dispuesto de la muestra completa de cárteles durante el periodo en los cuales eran legales en Finlandia y han podido estudiar así el proceso de creación y destrucción de todos los cárteles en ese país hasta su prohibición, sin verse afectados por el sesgo de selección.

Lo habitual, sin embargo, es contar con una muestra similar a la que contaban los primeros autores en estudiar la probabilidad de detección, (Bryant et al. 1991) que reconocían que sólo se puede concluir que la probabilidad estimada (la probabilidad de detección sería la inversa de la duración del cártel) de un 13 y un 17%, constituye un

límite superior de la probabilidad de desaparición de un cártel, ya que de no ser detectado habría tenido una duración mayor<sup>4</sup>.

Y es que uno de los supuestos restrictivos de este modelo es que los cárteles desaparecen sólo por detecciones, pero no entran a analizar la posible desaparición del cártel por la inestabilidad natural de los mismos.

(Ormosi et al. 2011) utiliza otro modelo de captura-recaptura, frecuentemente utilizado en biología para estimar datos poblacionales (tamaño, capturas, tasa de supervivencia). Los autores concluyen que en la UE la probabilidad de detección es inferior a un quinto entre 1985 y 2005.

(Miller et al. n.d.) también han utilizado un modelo estocástico pero en este caso para mostrar el cambio en el número de detecciones antes y después de la instauración del programa de clemencia en EEUU en 1993, concluyendo que el programa aumentó la capacidad de detección y disuasión de las autoridades de competencia de EEUU.

En la línea de la investigación de Miller, otros autores han construido un modelo más complejo de carácter endógeno que explica el comportamiento de empresas y autoridad (Harrington y Chang 2009a), en especial la creación y disolución de cárteles. Sin embargo, hasta donde tenemos conocimiento su aplicación práctica queda limitada al caso ya comentado de Finlandia (Hyytinen et al. 2010), obteniendo un resultado sustancialmente inferior a los anteriores estudios (la probabilidad obtenida es de tan sólo un 4%).

El estudio realizado por (Harrington y Wei 2017) sobre cárteles detectados en EEUU llega a conclusiones sorprendentes. El sesgo derivado de calcular la probabilidad de detección a partir de la muestra de cárteles detectados es inexistente en un modelo donde todos los cárteles siguen el mismo proceso de nacimiento y muerte, mientras que en un modelo que admite heterogeneidad de los cárteles, sí aparece este error, aunque consideran que es más reducido de lo que se pensaba anteriormente.

Se puede objetar a este estudio que sólo tiene en cuenta la detección de cárteles durante el periodo entre 1960 y 1985 antes de la implantación del programa de clemencia en EEUU. Los autores, a diferencia de los modelos anteriores, consideran que un cartel puede desaparecer no sólo por ser detectado sino también por inestabilidad interna.

---

<sup>4</sup> Otros autores sostienen que el sesgo va en dirección contraria ya que sólo se detectan cárteles de larga duración, los de corta duración pasan inadvertidos a las autoridades de competencia.

Finalmente, (Park, Lee, and Ahn 2018) utilizan una metodología bayesiana para calcular la probabilidad de detección en los EEUU, con resultados coherentes a los anteriores autores, entre 11,4% y 17,4%. No obstante, esta probabilidad está calculada a partir de una muestra de carteles investigados de los que diferencia entre sancionados y no sancionados.

En este artículo vamos a analizar, en primer lugar, los datos disponibles de cárteles en España, datos que son similares a los que tiene cualquier jurisdicción y que son los utilizados en los principales estudios. En segundo lugar, analizamos el método de Bryant y Eckard para estimar la probabilidad de detección y lo aplicamos a los datos de España, y finalmente, hacemos lo mismo con el modelo de Harrington y Wei, llegando a conclusiones muy similares.

### **3. – Datos disponibles de cárteles detectados en España**

Disponemos de datos sobre la duración de 114 cárteles detectados en España entre los años 2011 y 2016. Los datos analizados son los de duraciones de cárteles, aunque estos datos hay que tomarlos con cierta cautela. Suponemos que el cartel se da por finalizado en el momento de la incoación del expediente mientras que de la información contenida en el expediente puede deducirse la fecha de constitución del cartel. Sobre esta fecha de inicio puede existir cierta ambigüedad ya que en ocasiones se conoce sólo aproximadamente como fecha mínima de inicio del cartel. Es posible que como argumentan (Bryant et al. 1991) la fecha de inicio se refiera a las pruebas más antiguas existentes. En todo caso hemos excluido de la muestra aquellos expedientes de los que carecemos de la información sobre la duración de la infracción.

Bryant et al. utilizan dos métodos para calcular las duraciones de las infracciones, el primero es la duración máxima calculada con la información disponible, es decir, se utiliza la fecha de inicio más lejana posible y la más reciente posible para estimar la mayor duración del cartel posible. El segundo método es la duración mínima posible calculada con los datos disponibles (pruebas más recientes de inicio de cartel y fecha más antigua de finalización del cartel).

En la tabla siguiente se muestran los datos estadísticos de duraciones de la muestra de cárteles en España utilizando los dos métodos anteriores:

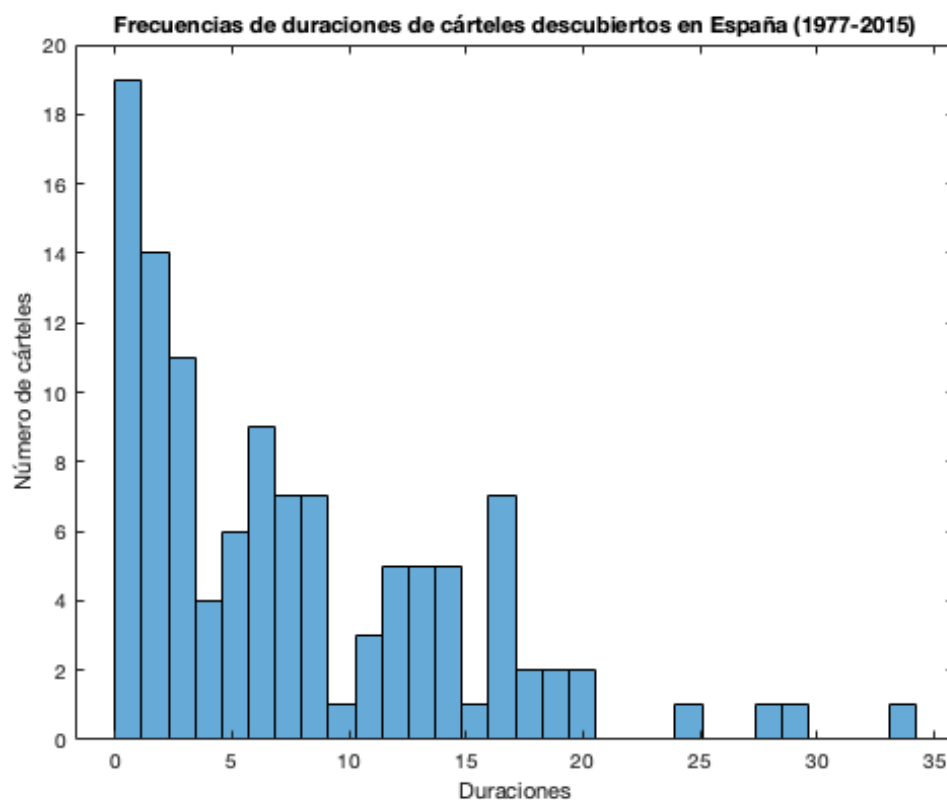
Duración 1	Diferencia entre la fecha máxima final y fecha mínima inicial del cartel
Duración 2	Diferencia entre la fecha mínima final y la fecha máxima inicial

Tabla 1: Frecuencias de duraciones de carteles

Intervalo – Años de duración	Frecuencias	
	Duración 1	Duración 2
0 a 2	20	24
2 a 4	11	15
4 a 6	14	12
6 a 8	14	20
8 a 10	12	15
10 a 12	7	12
12 a 14	5	5
14 a 16	10	3
16 a 18	5	4
18 a 20	9	1
> 20 años	8	4
Tamaño muestra	115	115
Media	7,65	5,55
Mediana	5,83	4,92
Mín	0,04	0,04
Máx	34,01	29,26



Figura 1: Histograma duraciones cárteles descubiertos España



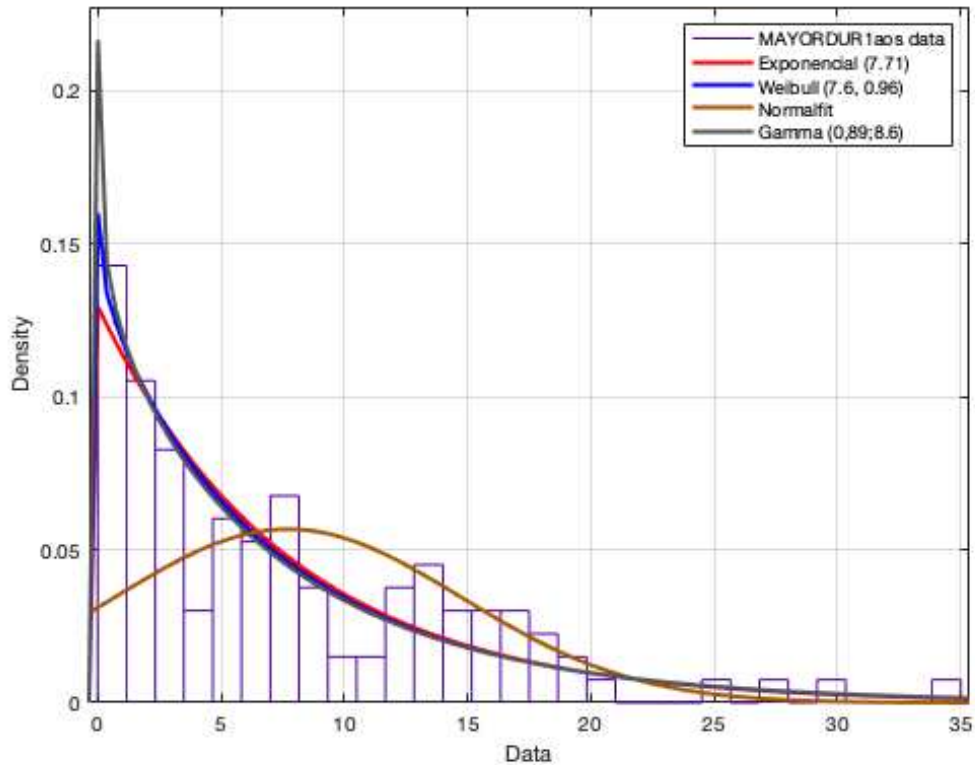
La figura 1 recoge las frecuencias de duraciones de 114 cárteles descubiertos. Para la duración calculada utilizando el método Duración 1 la media es de unos 7,6 años con un mínimo de 0,04 años y un máximo de 34 años. La mediana (5,83 años) es inferior a la media (7,65 años) lo que refleja el sesgo de la distribución a la derecha, por lo que hay más duraciones por debajo de la media que por encima de ella. Hay mucha variabilidad en los datos, lo que se ve reflejado en una desviación típica mayor que la media. En el caso de Duración 2, la media es 5,55 y la mediana 4,92 y los datos varían entre un mínimo de 0,04 y 29,26 años.

Estos resultados se asemejan a los de otros trabajos, por ejemplo, (Levenstein and Suslow 2006) con una duración media de 5 años y una distribución de carácter bimodal o de dos máximos, es decir, numerosos cárteles con una reducida duración (menos de un año) numerosos con duraciones entre 5 y 10 y, en menor medida, duraciones de décadas.

En la figura 2, se muestra el mismo histograma de las duraciones en años, pero ahora con varias curvas que representan el mejor ajuste de los datos a distribuciones de probabilidad típicas como la Normal, Gamma, Exponencial o Weibull. Como se ve en el

gráfico, estas dos últimas son las que se ajustan mejor a los datos. Por el contrario, no se obtiene un buen ajuste utilizando la distribución normal o gaussiana:

Figura 2: Ajuste de distribuciones de probabilidad estándar al histograma de frecuencias:



La mayoría de estudios sobre probabilidades de detección<sup>5</sup> han considerado que los datos de duraciones de cárteles siguen distribuciones de tipo exponencial, puesto que son las que mejor se ajustan a procesos estocásticos estrictamente positivos y *memoryless*, donde el resultado en un momento sólo depende del momento inmediatamente anterior.

Aunque tan sólo disponemos de datos de una muestra de cárteles descubiertos podemos tratar de ajustarlos a varias distribuciones teóricas utilizando la técnica de la máxima probabilidad o verosimilitud (MLE):

---

<sup>5</sup> Bryant-Eckard utilizaron una función exponencial. (Hellwig and Hüschelrath 2018) utilizan una distribución Weibull para el caso de cárteles en la UE.

Tabla 2: Resultados del ajuste a diferentes distribuciones de probabilidad

Tipo de Distribución	Normal	Gamma	Weibull <sup>6</sup>	Exponencial
MLE	-383.619	-346,49	-346.84	-346,9
Media	7.71723	7,72	7.72885	7,72
Varianza	49.4546	66,43	64.2586	59.56
Parámetro A (Forma)	7.71723 (Mu)	0.896525	7.60567	0.522418 (Mu)
Parámetro B (Escala)	7.0324 (Sigma)	8.60794	0.964349	

El mejor resultado se obtiene con cualquiera de las distribuciones de la familia exponencial frente a la distribución normal, por ejemplo, cuyo resultado en términos de verosimilitud es inferior. Bryant y Eckard concluyeron también que los datos obedecían a una función exponencial.

A continuación, dejamos a un lado el análisis de estos datos en particular y sus conclusiones, para pasar a construir un modelo teórico general que explique el proceso por el que pasa un cartel hasta su desaparición y con el cual podemos explicar estos datos de duraciones en particular de cárteles detectados.

---

<sup>6</sup> La distribución Gamma, Weibull y Exponencial están relacionadas puesto que la distribución Gamma con parámetro de forma igual a 1 resulta ser equivalente a una Exponencial de parámetro  $1/\text{Beta}$ , mientras que una distribución Weibull con parámetro forma igual a 1 resulta ser una Exponencial de media igual a Beta.

### **3.- Estimación de la probabilidad de detección**

A la hora de estimar la probabilidad de detección a partir de los datos que disponemos, las duraciones de los cárteles detectados, necesitamos un modelo teórico que explique el proceso de creación y destrucción de cárteles durante un periodo de tiempo del que poder sacar conclusiones.

Analizamos, en primer lugar, el modelo de los primeros autores, Bryant y Eckard, que utilizaron un modelo matemático continuo de nacimiento y muerte, modelo conocido como proceso de Poisson, si bien el modelo se encuentra limitado en la medida en que sus resultados no tienen en cuenta la naturaleza inestable de los cárteles que pueden desaparecer de forma espontánea, sin que sean detectados.

En segundo lugar, analizamos el modelo desarrollado por Harrington y Wei, que evita analizar el proceso de creación de cárteles y se concentra en el proceso de desaparición teniendo en cuenta no sólo la muerte por detección del cártel sino también por inestabilidad interna del mismo. Parece, por tanto, que el proceso de desaparición o muerte de los cárteles quedaría fundamentado en este modelo, pero queda pendiente un análisis teórico del proceso de nacimiento de los cárteles para completar el estudio teórico. Habría que introducir en el modelo el comportamiento de los agentes a la hora de decidir entre crear un cartel o competir (véase por ejemplo el estudio realizado por Motta y Polo). Esto queda pendiente para un estudio futuro.

A continuación, analizamos los dos métodos aplicándolos a los datos observados para el caso de España:

#### **3.1.- Metodología de Bryant y Eckard: Modelo Continuo**

Comenzamos por analizar el método utilizado por Bryant y Eckard, que utilizaron un proceso continuo aleatorio de nacimiento – muerte para explicar la evolución de los cárteles en el tiempo.

Los autores parten de unos supuestos restrictivos: los cárteles se crean y destruyen de acuerdo a un proceso totalmente aleatorio e independiente, donde al final del periodo todos los cárteles desaparecen por detección, no hay otras opciones. Por otro lado, se supone que ya ha transcurrido un largo periodo de tiempo desde que comenzó el proceso de creación de cárteles por lo que nos encontraríamos en un estado estacionario, es decir, las tasas de creación y destrucción de cárteles son constantes.

En este contexto, teniendo en cuenta estos supuestos los autores construyen a una función de verosimilitud o de probabilidad de observar  $n$  cárteles de duraciones  $L_1, \dots, L_n$ , en un intervalo de tiempo  $T_2 - T_1$ :

$$V(L_1, \dots, L_n) = \theta^n \lambda^n e^{-\theta(T_2-T_1)} e^{-\lambda \sum L_i}$$

Donde los parámetros  $\theta$  y  $\lambda$  son la probabilidad nacimiento y probabilidad de detección respectivamente. Tomando logaritmos y derivando se llega a los estimadores máximo verosímiles (MLE) de los parámetros de interés:

$$\frac{1}{\theta} = (T_2 - T_1)/n \quad \text{y} \quad \frac{1}{\lambda} = \sum(L_i)/n$$

Por tanto, en estado estacionario, los parámetros,  $1/\theta$  y  $1/\lambda$  son, sencillamente la media de los tiempos entre cárteles y la media de duraciones de cárteles. (Keiding 1975, pág. 1) llega a la misma conclusión.(en el anexo 1 hay una explicación del proceso).

Los parámetros estimados  $\lambda$  y  $\theta$  de acuerdo con la metodología de estos autores aplicados a los datos de duraciones en España serían:

Tabla 3: Resultados estimación Bryant-Eckard aplicado a datos de cárteles en España

	<i>Medición Duración 1<sup>7</sup></i>	<i>Medición Duración 2</i>
<i>Probabilidad de Detección (<math>\lambda</math>)</i>	0,1295	0,18018
<i>Duración media</i>	7,71	5,5

Los resultados de la estimación aparecen en la tabla 3. La duración media hasta la detección es de 7 años, utilizando el primer método de cálculo de las duraciones y de 5,5 para el segundo método. La probabilidad anual de detección se encuentra entre 13% y 18%. Estos resultados son similares a los obtenidos por Bryant y Eckard, que afirmaban que la duración media de un cartel en EEUU era de 5 a 7 años de media mientras que la probabilidad anual estimada era de entre un 12,8% y un 17,4%.

Para concluir, los autores elaboran una función para explicar los cárteles que sobreviven a este proceso de detección en cada periodo, función decreciente a medida que se acerca al final del periodo lo cual concuerda con los supuestos iniciales que establecieron. Según los autores, dado que observamos el proceso sólo entre  $T_1$  y  $T_2$

7

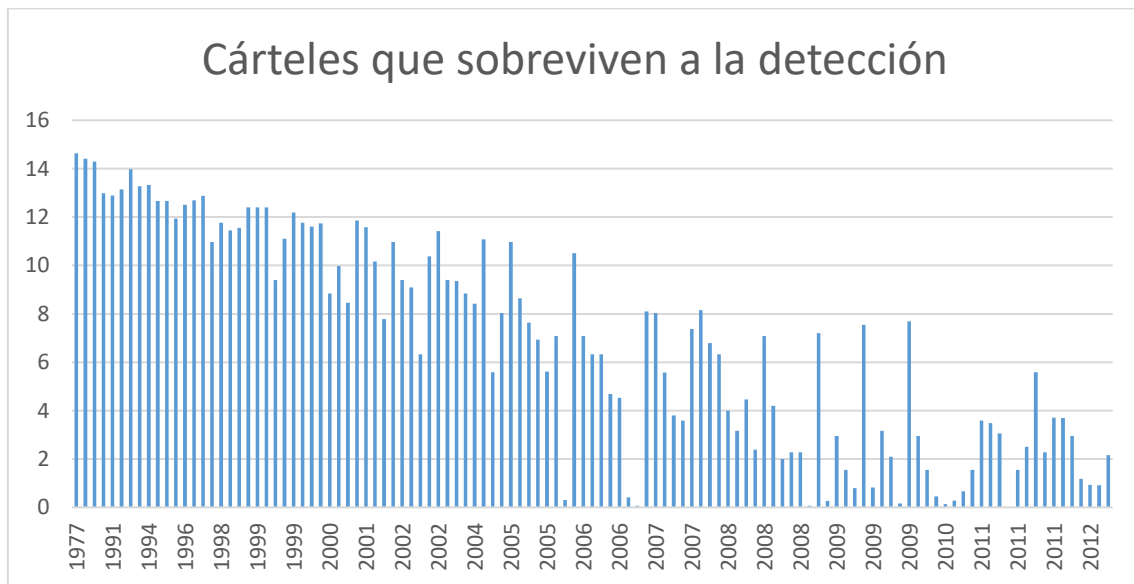
<i>DUR1</i>	<i>Diferencia entre la fecha máxima de la final y fecha mínima del inicio del cartel</i>
<i>DUR2</i>	<i>Diferencia entre el promedio de la fecha fin y el promedio de la fecha inicio</i>

esperamos que los cárteles vivos en  $t$  desaparezcan a medida que nos aproximamos a  $T_2$ , contabilizando los carteles vivos en  $t$ , y todavía vivos en  $T_2$ , llegan a que los cárteles supervivientes en cada periodo  $t$  son:

$$\left(\frac{\theta}{\lambda}\right)(1 - e^{\{\lambda(T_2-t)\}}),$$

Si utilizamos la misma fórmula en nuestra muestra de cárteles detectados en España también se obtiene que, a medida que  $t$  se acerca a  $T_2$ ,  $N(t)$  se reduce. Hemos construido el mismo gráfico con datos de España para obtener el siguiente gráfico:

Figura 3: Cárteles que sobreviven a la detección en España utilizando la metodología de Bryant-Eckard.



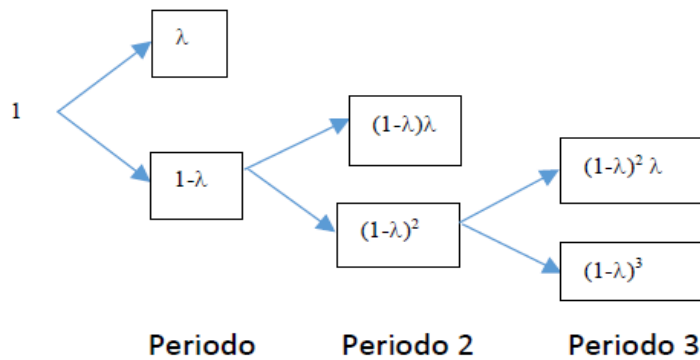
### **3.2.- Metodología basada en un Modelo Discreto. Cárteles homogéneos.**

El análisis anterior adolece de varias limitaciones, la fundamental es que en ese proceso de creación y destrucción de los cárteles no tiene en cuenta la naturaleza inestable de los mismos y por tanto habría que tener en cuenta que una parte de los cárteles desaparecen por causas internas y no por detección. Parece más realista también considerar el proceso de nacimiento y muerte como un proceso discreto, en lugar de continuo como lo plantearon los primeros autores.

Para realizar un modelo teórico sobre este proceso debemos explicar los factores que dan lugar a la creación y a la destrucción de los cárteles. Como hemos dicho, para incluir la creación de cárteles en este modelo deberíamos tener un modelo que explique este proceso que por el momento no disponemos. Sin embargo, si podemos explicar cómo

a partir de una población de cárteles que suponemos estable en el tiempo va decreciendo a causa de dos factores: detección y desestabilización. Partiendo de un número de cárteles inicial estandarizado a 1 (Harrington and Wei 2017), en cada periodo se descubre una proporción  $\lambda$  mientras que  $1-\lambda$  no son descubiertos (suponemos que  $\lambda$  es igual en todos los periodos):

Figura 4: Proporción de cárteles en cada periodo (sólo detección)



En  $n$  periodos, los cárteles descubiertos serían  $\lambda (1-\lambda)^{n-1}$  mientras que aquellos que no son descubiertos y, por tanto, sobreviven serían  $\sum_{n=1}^{\infty} (1-\lambda)^{n-1}$  como suma de las columnas 3 y 2, respectivamente:

Tabla 4: Proporción de cárteles en cada periodo (sólo detección)

Periodo de tiempo (n)	Carteles vivos	Carteles detectados
1	1	$\lambda$
2	$(1-\lambda)$	$(1-\lambda)\lambda$
3	$(1-\lambda)^2$	$(1-\lambda)^2\lambda$
4	$(1-\lambda)^3$	$(1-\lambda)^3\lambda$
...	...	....
$\Sigma$	$\sum_{n=1}^{\infty} (1-\lambda)^{n-1}$	$\lambda \sum_{n=1}^{\infty} (1-\lambda)^{n-1}$

La suma de la columna de cárteles detectados sería igual a:

$$\lambda \sum_{n=1}^{\infty} (1 - \lambda)^{n-1} = \frac{(1-\lambda)\lambda}{1-(1-\lambda)\lambda} = \frac{(1-\lambda)\lambda}{1-\lambda+\lambda^2} \quad (1)$$

Este sería un resultado discreto equivalente al modelo de Bryant y Eckard, aunque en este caso sólo tenemos en cuenta la desaparición del cártel por detección, tal y como hace Harrington.

Ahora introducimos mayor realismo a esta secuencia si tenemos en cuenta no sólo la detección sino también la disolución interna del cártel para explicar la desaparición de los cárteles.

En este caso, la secuencia consiste en partir del número de cárteles inicial normalizado a 1, pero ahora en cada periodo se descubre una proporción  $\lambda$  mientras que  $1-\lambda$  no son descubiertos, (suponemos que  $\lambda$  es igual en todos los periodos). Ahora de estos cárteles no descubiertos hay una proporción  $q$  que desaparecen por desestabilización interna. La secuencia sería entonces:

Figura 5: Proporción de cárteles en cada periodo (detección y desestabilización)

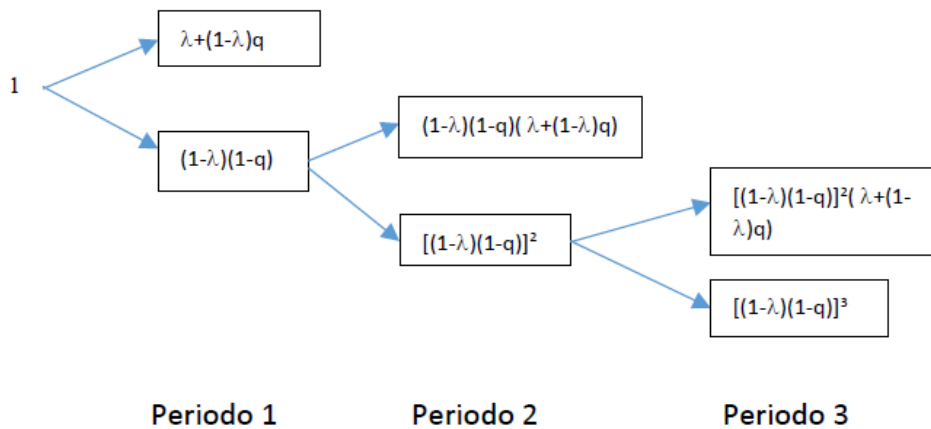




Tabla 5: Proporción de cárteles en cada periodo (detección y desestabilización)

Periodo de tiempo (n)	Cantidad de cárteles vivos en cada periodo t	Cantidad de cárteles muertos (detección o disolución interna) en cada periodo t
1	$1$	$\lambda + (1-\lambda)q$
2	$(1-\lambda)(1-q)$	$[(1-\lambda)(1-q)](\lambda + (1-\lambda)q)$
3	$[(1-\lambda)(1-q)]^2$	$[(1-\lambda)(1-q)]^2(\lambda + (1-\lambda)q)$
4	$[(1-\lambda)(1-q)]^3$	$[(1-\lambda)(1-q)]^3(\lambda + (1-\lambda)q)$
...	...	....
$\Sigma$	$\sum_{n=1}^{\infty} [(1-\lambda)(1-q)]^{n-1}$	$[(\lambda + (1-\lambda)q)] \sum_{n=1}^{\infty} [(1-\lambda)(1-q)]^{n-1}$

La suma de la columna de cárteles desaparecidos teniendo en cuenta la detección y la desestabilización, ponderada por la duración de cada cartel es igual a la duración media de los cárteles (véase Harrington):

$$\begin{aligned}
 [(1-\lambda)(1-q)] \sum_{n=1}^{\infty} n [(1-\lambda)(1-q)]^{n-1} &= [(\lambda + (1-\lambda)q)] \frac{1}{1 - (1-\lambda)(1-q)} \\
 &= \frac{1}{q + \lambda - \lambda q} \quad (2)
 \end{aligned}$$

Este sería el resultado en el caso de que consideremos que  $\lambda$  y  $q$  son iguales para todos los cárteles, es decir, son homogéneos. No obstante, es más realista considerar que son heterogéneos, por lo que tienen parámetros distintos ya que su estabilidad puede verse afectada por características diferentes en función del sector en el que se encuentran. También el parámetro de detección varía en función de la facilidad o dificultad para ser detectado lo cual depende también del sector y del número de empresas que puede facilitar la detección.

Para poder obtener resultados empíricos es necesario realizar una hipótesis sobre qué valores adoptan los parámetros  $\lambda$  y  $q$  y la relación que tienen entre ellos. Esto lo veremos en el apartado siguiente.

### 3.2.1. Modelo discreto y heterogéneo. Estimación de la probabilidad de detección.

En este caso consideramos (Harrington y Wei 2017) la hipótesis más realista: los carteles no son homogéneos y, por tanto, cada cartel tiene parámetros  $(\lambda, p)$  diferentes. Antes de seguir, tenemos que hacer un supuesto sobre cómo se comportan  $p$  y  $\lambda$ , en este sentido consideramos que cada parámetro sigue una distribución de probabilidad de tipo Beta definida de forma que la distribución conjunta de los dos parámetros está definida por cinco parámetros  $\theta = \{\mu_\lambda, c_p, \mu_p, c_\lambda, \psi\}$ , donde  $\mu$  es la media y  $c$  el coeficiente de variación para cada parámetro, mientras que  $\psi$  sería el coeficiente de correlación entre ambos parámetros. Las diferentes combinaciones de los parámetros quedan recogidas en la función de densidad conjunta  $k(\lambda, p)$ .

Ahora se trata de seleccionar valores de los parámetros con los que obtenemos los mayores valores de verosimilitud utilizando la siguiente fórmula basada en el modelo teórico del apartado anterior, con la que obtenemos la probabilidad de que el cartel tenga la duración  $t$  condicionado a que sea descubierto:

$$\iint \frac{p[(1-\lambda)(1-p)]^{t-1}}{\iint \left(\frac{p}{\lambda+p-\lambda p}\right) k(\lambda, p) d\lambda dp} k(\lambda, p) d\lambda dp \quad (3)$$

Para estimar los parámetros utilizamos una aproximación bayesiana que requiere, en primer lugar, especificar unos valores a priori de los parámetros, y que dado el desconocimiento que tenemos sobre los parámetros asumimos que siguen una distribución uniforme, que asigna la misma probabilidad a cada valor. El siguiente paso es calcular los valores a posteriori que se obtienen multiplicando los valores a priori por su verosimilitud calculada según (3).

Sólo utilizaremos aquellos valores cuya verosimilitud (o probabilidad) es mayor que la que se obtendría si simplemente utilizamos valores obtenidos de una distribución uniforme, rechazando los valores con menor verosimilitud. Este proceso lo repetimos con 200.000 valores de los parámetros, aceptando únicamente un 1% de los valores.

Ese conjunto de valores de parámetros resultante los insertamos (3) que se deriva del modelo teórico y que refleja la probabilidad incondicional de desaparición de los carteles:

$$\iint [1 - (1 - \lambda)(1 - q)] k(\lambda, q; \theta) d\lambda dq \quad (4)$$

### 3.3.1. Resultado empírico

Como vemos en la siguiente tabla hemos obtenido unos valores para los parámetros  $p$  (probabilidad de que el cartel desaparezca por desestabilización) y  $\lambda$  (probabilidad de que el cártel desaparezca por detección), una vez aplicado el modelo, menores y con una desviación estándar inferior que los valores apriorísticos, que como recordamos provienen de una distribución totalmente aleatoria.

Los valores medios de  $\lambda$  y de  $q$  a posteriori son 0,0854 y 0,0831, inferiores a los valores a priori de 0,110, en los dos casos, y una reducción también de los valores de los coeficientes de variación desde (0,449;0,446) hasta (0,4028;0,3581), lo que significa que es más probable que los cárteles tengan valores bajos de probabilidad de detección y de desestabilización, y que su variabilidad medida a través de los coeficientes sea más reducida.

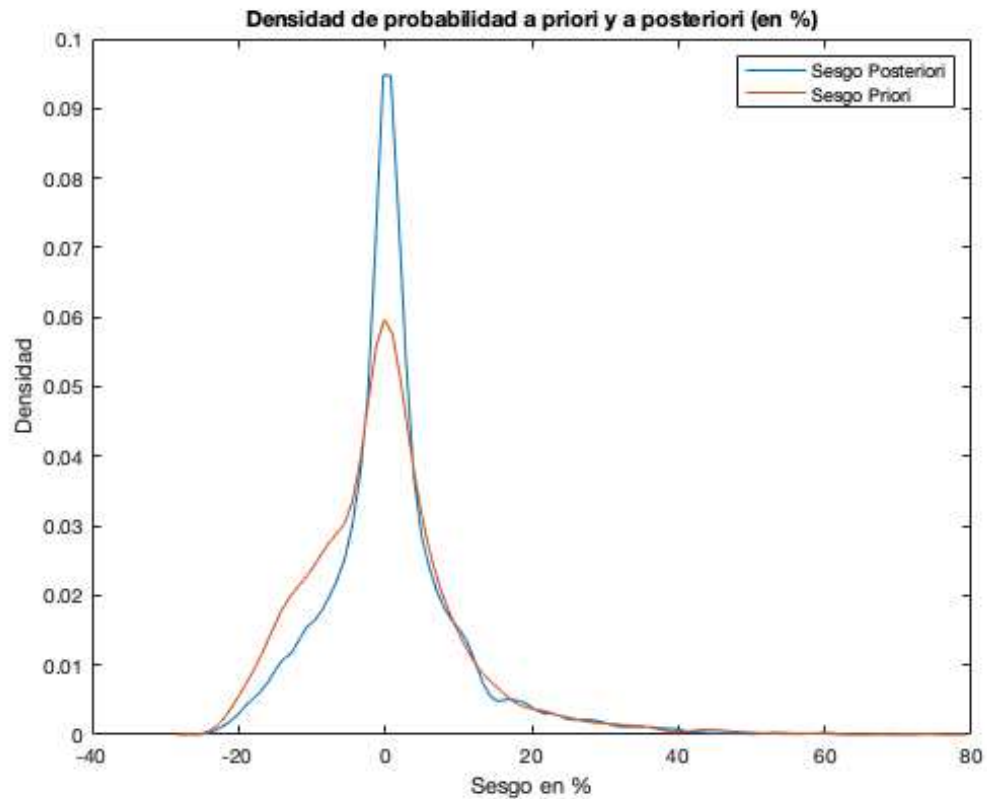
Tabla 6: Resultados estadísticos

	<b>Media Prior</b>	<b>Desv. Est.</b>	<b>Media Posterior</b>	<b>Desv. Est.</b>
<b>Valor medio Lambda (<math>\lambda</math>)</b>	0,109	0,0506	0,0854	0,0384
<b>Valor Medio Rho (<math>q</math>)</b>	0,110	0,0505	0,0831	0,035
<b>Coef. Lambda</b>	0,449	0,261	0,4028	0,2594
<b>Coef. Rho</b>	0,446	0,2601	0,3581	0,2515
<b>Correlación</b>	5,46	2,5964	5,4233	2,609
<b>Prob. muerte Cártel</b>	0,2066	0,063	0,162	0,0256

El sesgo o error derivado de utilizar una muestra de carteles detectados frente a la alternativa de utilizar toda la población de cárteles se acerca a 0 como se ve en el gráfico inferior, con una desviación típica a posteriori de 9,5% inferior a la obtenida a priori de

casi 11%. Por tanto, la distribución a posteriori es menos dispersa que la a priori y más concentrada en valores en torno a un error de 0.

Figura 6: Densidad probabilidad priori y posterior



Con dos desviaciones típicas por encima y debajo de la media en la distribución a priori el sesgo se encontraría entre un 20% y un -18%, mientras que en el caso de la distribución a posteriori se sitúa entre 19% y -18%. Esto significa que la media de duraciones de cárteles descubiertos obtenida en nuestra muestra de cárteles descubiertos de 7,6 años, podría estar como máximo sobreestimando en un 19%, o como mínimo infraestimando en un 18% la duración de los cárteles, es decir, la duración real de los cárteles se encontraría en un intervalo entre 6,2 y 9,1 años. Si utilizamos la distribución a priori la dispersión sería mayor y el intervalo sería entre 5,96 y 9,21 años.

Con ello, podemos concluir que el error de utilizar una muestra de cárteles descubiertos como proxy de todos los cárteles es reducido.

En segundo lugar, podemos hallar la probabilidad de muerte de un cartel utilizando (4).

En la figura 7 se observa que la mayor densidad se encuentra en torno a 0,16, que es la media. En un intervalo de dos desviaciones estándar respecto a la media, entre 0,11 y 0,21, se encuentran englobados los resultados de estudios anteriores. Por ejemplo, el resultado utilizando el modelo Bryant-Eckard, era de un 13-17%. La figura 8 muestra cómo se concentra la densidad de probabilidad en ese intervalo.

Por tanto, coincidimos con las conclusiones de Harrington, el error por usar una muestra de carteles descubiertos en lugar de toda la población de cárteles es reducido, por lo que son aceptables los resultados utilizando una muestra restringida.

En todo caso, los resultados son siempre sobre probabilidades de desaparición del cártel, por lo que lo correcto sería hablar de probabilidades de desaparición o muerte tanto para el caso de Bryant Eckard como de este último resultado.

Figura 7: Densidad probabilidad incondicional muerte

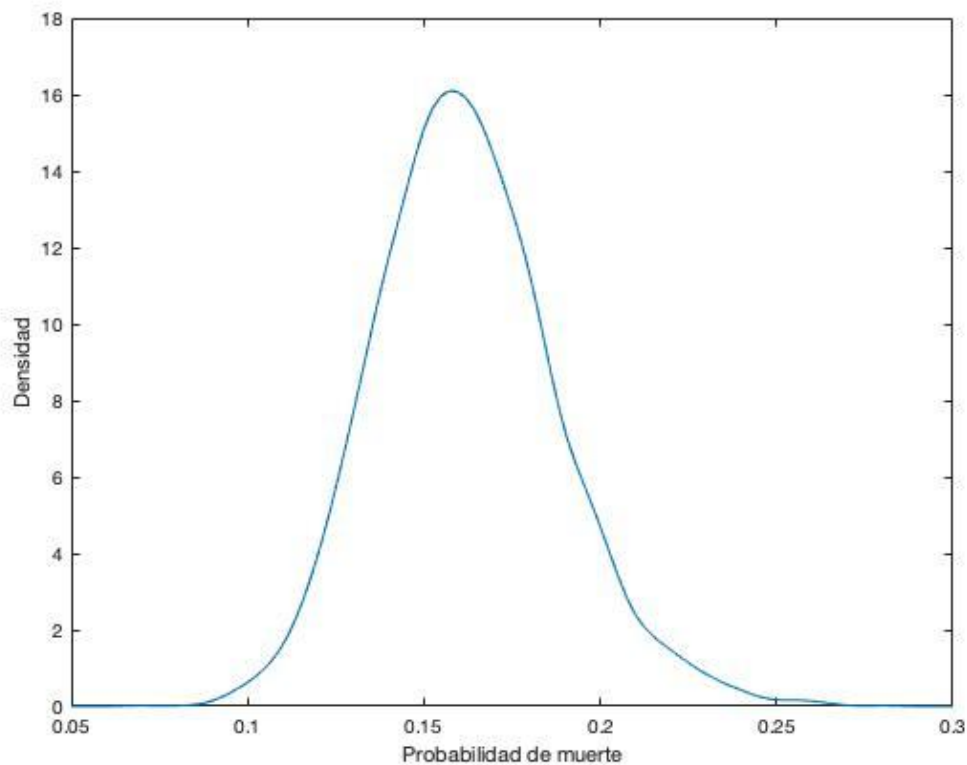
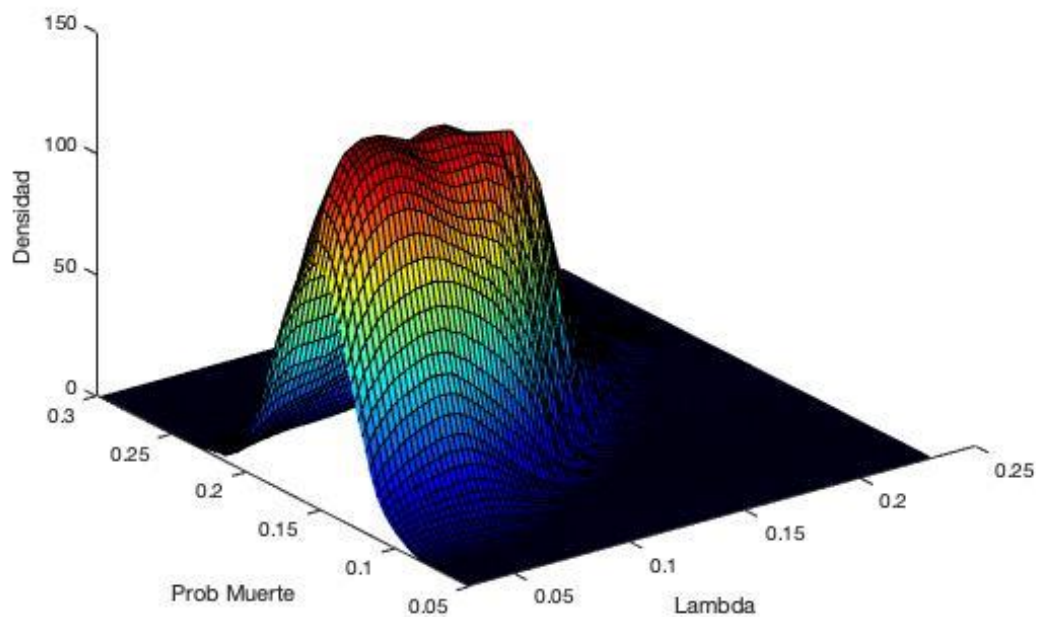


Figura 8: Densidad probabilidad de muerte frente al parámetro  $\lambda$



## 6. Conclusiones

Los autores Bryant y Eckard fueron los primeros en estimar la probabilidad de detección de los cárteles utilizando una muestra de cárteles en EEUU durante los años 1961 y 1988. La probabilidad de detección obtenida fue de entre un 13% y un 17% y la duración media de siete años. Utilizando el mismo procedimiento, Combe y Monnier llegaron a una probabilidad anual de 12,9% y 13,2%, para una muestra de carteles en la UE entre 1969 y 2007.

En este artículo se aplica la misma metodología para calcular la duración hasta la detección y la probabilidad de detección con datos de España, y se obtiene una probabilidad de entre 13% y 18%. El número medio de cárteles detectados de acuerdo con Bryant et al. en un momento dado se encuentra entre 36 y 50. En nuestro modelo el número medio de cárteles sería entre 16 y 22.

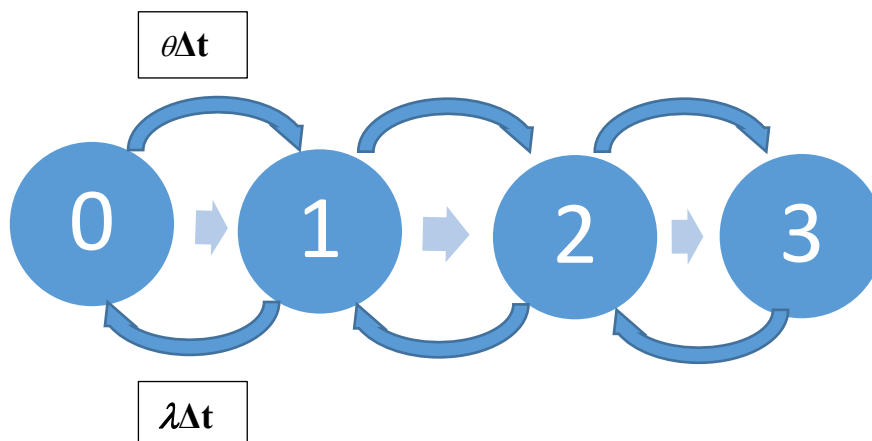
Estos autores se preguntan si se pueden extender estos resultados a aquellos carteles no detectados, si bien aparentemente no es posible estimar su número, consideran que si la duración de un cartel descubierto no es mayor a la de los no descubiertos,

entonces deducen que las estimaciones de duraciones son también aplicables a los no descubiertos (como límite inferior), y las probabilidades estimadas son un límite superior de la probabilidad de que un cártel desaparezca no por detección sino por desavenencias internas.

En la segunda parte, del artículo se aplica un modelo desarrollado por Harrington que combina tanto la muestra de cárteles detectados, el modelo teórico de nacimiento de cárteles y nuestra creencia a priori sobre el comportamiento de los cárteles, obteniendo una probabilidad de desaparición de los cárteles de un 16,2%, dentro del intervalo obtenido utilizando el modelo de Bryant-Eckard, si bien ahora como hemos visto hemos tenido en cuenta toda la población posible de cárteles y no sólo los descubiertos.

### Anexo 1: Bryant-Eckard:

La aproximación que realiza Bryant-Eckard puede explicarse considerando que la creación de cárteles sigue una distribución de Poisson de parámetro  $\theta$ , mientras que la duración del cartel sigue una distribución exponencial negativa  $\lambda e^{-\lambda t}$ . El siguiente diagrama explica el proceso de creación y descubrimiento de cárteles (o muerte porque consideramos que sólo se destruyen con la detección) (Ross, n.d.), los números reflejan el número de cárteles y las flechas las transiciones de un estado a otro, es decir, de un número de cárteles a otro. Sólo hay dos opciones, o se incrementa el número de cárteles en uno o se reduce en uno:



En el estado 0, se puede estar porque se ha descubierto un cartel, por lo que se pasa de 1 a 0 con probabilidad  $\lambda$ , o porque no nazca ningún cartel para un intervalo de tiempo reducido  $\Delta t$ . Esto se traduce en una ecuación de transición:

$$p_0 = p_0 (1 - \theta\Delta t) + p_1 \lambda\Delta t$$

Sucesivamente tendríamos:

$$p_1 = p_0 \theta\Delta t + p_1 (1 - \theta\Delta t - \lambda\Delta t) + p_2 \lambda\Delta t$$

$$p_2 = p_1 \theta\Delta t + p_2 (1 - \theta\Delta t - \lambda\Delta t) + p_3 \lambda\Delta t$$

$$p_3 = p_2 \theta\Delta t + p_3 (1 - \theta\Delta t - \lambda\Delta t) + p_4 \lambda\Delta t$$



...

Simplificando obtenemos:

$$p_0 \theta \Delta t = p_1 \lambda \Delta t$$

$$p_1(\theta \Delta t + \lambda \Delta t) = p_0 \theta \Delta t + p_2 \lambda \Delta t$$

$$p_2(\theta \Delta t + \lambda \Delta t) = p_1 \theta \Delta t + p_3 \lambda \Delta t$$

$$p_3(\theta \Delta t + \lambda \Delta t) = p_2 \theta \Delta t + p_4 \lambda \Delta t$$

Si eliminamos  $\Delta t$  y simplificamos obtenemos:

$$p_0 \theta = p_1 \lambda$$

$$p_1 \theta = p_2 \lambda$$

$$p_2 \theta + p_2 \lambda = p_1 \theta + p_3 \lambda$$

$$p_3 \theta + p_3 \lambda = p_2 \theta + p_4 \lambda$$

...

Lo que nos permite llegar a:

$$p_0 \theta = p_1 \lambda$$

$$p_1 \theta = p_2 \lambda$$

$$p_2 \theta = p_3 \lambda$$

$$p_3 \theta = p_4 \lambda$$

...

O lo que es lo mismo:

$$p_1 = \theta / \lambda \quad p_0$$

$$p_2 = \theta/\lambda p_1$$

$$p_3 = \theta/\lambda p_2$$

$$p_4 = \theta/\lambda p_3$$

...

De forma que se llega a:

$$p_1 = \theta/\lambda p_0$$

$$p_2 = (\theta/\lambda)^2 p_0$$

$$p_3 = (\theta/\lambda)^3 p_0$$

$$p_4 = (\theta/\lambda)^4 p_0$$

...

Dado que  $p_0 + p_1 + p_2 + p_3 + \dots = 1$  y sustituyendo los  $p_i$  anteriores en esta ecuación obtenemos la siguiente sucesión geométrica:

$$p_0 + (\theta/\lambda)^1 p_0 + (\theta/\lambda)^2 p_0 + (\theta/\lambda)^3 p_0 + \dots = 1$$

$$p_0 (1 + (\theta/\lambda)^1 + (\theta/\lambda)^2 + (\theta/\lambda)^3 + \dots) = 1$$

Que se puede simplificar si restamos esta sucesión por la misma multiplicada por  $\theta/\lambda$  a:

$$P_0 = 1 - \left(\frac{\theta}{\lambda}\right)$$

El estado estacionario se obtiene de la siguiente forma:

$$p_0 = (\theta/\lambda)^0 (1 - (\theta/\lambda))$$

$$p_1 = (\theta/\lambda)^1 (1 - (\theta/\lambda))$$

$$p_2 = (\theta/\lambda)^2 (1 - (\theta/\lambda))$$

$$p_3 = (\theta/\lambda)^3 (1 - (\theta/\lambda))$$

....

O lo que es lo mismo:

$$P[n] = (\theta/\lambda)^n (1 - \theta/\lambda)$$

Si acotamos el periodo de tiempo entre  $t_1$  y  $t_2$  ( $t_2 > t_1$ ) (Clarke n.d.), suponiendo que  $\theta/\lambda < 1$ , llega a:

$$\frac{1}{\theta} = (t_2 - t_1)/n \quad \text{y} \quad \frac{1}{\lambda} = \Sigma(L_i)/n$$

## Anexo: Estimación del número de cárteles no descubiertos

Aunque este cálculo no fue realizado por Bryant-Eckard ni Harrington, tratamos de aproximar el número de cárteles existentes pero no descubiertos a partir de la muestra de cárteles descubiertos utilizando también un modelo matemático de cálculo de supervivencia.

Mantenemos los mismos supuestos de independencia de las duraciones de los diferentes cárteles y que éstos se crean como un proceso Poisson. Hay, por tanto, un número de cárteles descubiertos durante el periodo que conocemos  $N_1$  y un número de cárteles creados pero no descubiertos durante el proceso  $N_2$ . Todos los cárteles creados acaban por ser descubiertos.

Suponiendo que conocemos  $\lambda$  de la estimación del apartado anterior, podemos calcular el número de cárteles creados pero no descubiertos durante el proceso (Ross, pg. 316) usando la siguiente ecuación:

$$N_2(t) = \frac{n\lambda(1 - e^{-\frac{T_2-T_1}{\lambda}})}{(T_2 - T_1) - \lambda(1 - e^{-\frac{T_2-T_1}{\lambda}})}$$

Ahora damos valores a los parámetros obtenidos de la estimación del apartado anterior:  $t=14.609$  días, duración media desde creación hasta detección = 2817 días, y número de cárteles detectados durante el periodo  $n= 115$ , obtendríamos 27 cárteles existentes pero no detectados:

$$n = \frac{(115 \times 2.817) \times (1 - \exp(-\frac{14.609}{2.817}))}{(14.609 - 2.817) \times (1 - \exp(-\frac{14.609}{2.817}))} \approx 27$$

Esta estimación tiene la limitación de que la tasa de nacimiento de cárteles  $\lambda$  es constante durante el periodo.

## Bibliografía

- Bryant, Peter G, Edwin Woodrow Eckard, Peter G Bryant, and Edwin Eckard. 1991. "Price Fixing: The Probability of Getting Caught." *The Review of Economics and Statistics* 73 (3): 531–36.  
[https://econpapers.repec.org/article/tprrestat/v\\_3a73\\_3ay\\_3a1991\\_3ai\\_3a3\\_3ap\\_3a531-36.htm](https://econpapers.repec.org/article/tprrestat/v_3a73_3ay_3a1991_3ai_3a3_3ap_3a531-36.htm).
- Buccirossi, Paolo, and Giancarlo Spagnolo. 2005. "Optimal Fines in the Era of Whistleblowers-Should Price Fixers Still Go to Prison?"
- Clarke, A. Bruce. n.d. "Maximum Likelihood Estimates in a Simple Queue." *The Annals of Mathematical Statistics*. Institute of Mathematical Statistics. Accessed December 2, 2018. <https://doi.org/10.2307/2237068>.
- Combe, Emmanuel, Constance Monnier, and Renaud Legal. 2008. "Cartels: The Probability of Getting Caught in the European Union." *Bruges European Economic Research Papers*. <https://ideas.repec.org/p/coe/wpbeer/12.html>.
- García Verdugo, Javier, Carlos Merino, and Lorena Gómez Cruz. 2016. "Valoración Económica de Las Sanciones de La Autoridad Española de Competencia (2011-2015)." *Anuario de La Competencia, Nº 1, 2016, Págs. 345-379*, no. 1: 345–79. <https://dialnet.unirioja.es/servlet/articulo?codigo=6044346>.
- Harrington, Joseph E. 2014. "Penalties and the Deterrence of Unlawful Collusion." *Economics Letters* 124 (1): 33–36.  
<https://doi.org/10.1016/J.ECONLET.2014.04.010>.
- Harrington, Joseph E., and Myong-Hun Chang. 2009. "Modeling the Birth and Death of Cartels with an Application to Evaluating Competition Policy." *Journal of the European Economic Association* 7 (6): 1400–1435.  
<https://doi.org/10.1162/JEEA.2009.7.6.1400>.
- Harrington, Joseph E., and Yanhao Wei. 2017. "What Can the Duration of Discovered Cartels Tell Us About the Duration of All Cartels?" *The Economic Journal* 127 (604): 1977–2005. <https://doi.org/10.1111/eoj.12359>.
- Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153. <https://doi.org/10.2307/1912352>.
- Hellwig, Michael, and Kai Hüschelrath. 2018. "WHEN DO FIRMS LEAVE

CARTELS? DETERMINANTS AND THE IMPACT ON CARTEL SURVIVAL.”

*International Review of Law and Economics* 54 (June): 68–84.

<https://doi.org/10.1016/J.IRLE.2017.11.001>.

Hyytinen, Ari, Frode Steen NHH Bergen, Otto Toivanen HECER, Frode Steen, Otto Toivanen, Susanna Fellman, Joe Harrington, Vesa Kanninen, and Howard Smith. 2010. “Cartels Uncovered\*.”

<https://helda.helsinki.fi/bitstream/handle/10138/16777/cartelsu.pdf?sequence=1>.

Katsoulacos, Yannis, and David Ulph. 2013. “Antitrust Penalties and the Implications of Empirical Evidence on Cartel Overcharges.” *The Economic Journal* 123 (572).

Keiding, Niels. 1975. “Maximum Likelihood Estimation in the Birth-and-Death Process.” *The Annals of Statistics* 3 (2): 363–72.

<https://doi.org/10.1214/aos/1176343062>.

Levenstein, Margaret C, and Valerie Y Suslow. 2006. “What Determines Cartel Success?” *Journal of Economic Literature* 44 (1): 43–95.

<https://doi.org/10.1257/002205106776162681>.

Miller, Nathan H, Severin Borenstein, Joseph Farrell, Russell Pittman, Carl Shapiro, John Sutton, Sofia Villas-Boas, and Gregory Werden. n.d. “Strategic Leniency and Cartel Enforcement.” Accessed September 15, 2018.

<https://doi.org/10.1257/aer.99.3.750>.

Ormosi, Peter L, Stephen Davies, Joseph Harrington, Morten Hviid, Bruce Lyons, Franco Mariuzzo, Matthew Olczak, Greg Shaffer, Iwan Bos, and Greg Shaffer. 2011. “A Tip of the Iceberg? The Probability of Catching Cartels A Tip of the Iceberg? The Probability of Catching Cartels I Am Grateful for the Useful Comments To.”

<http://competitionpolicy.ac.uk/documents/8158338/8253131/CCP+Working+Paper+11-6.pdf/9d9c5c1d-35fe-4139-baa0-06838f7d4fe9>.

Park, Jihyun, Juhyun Lee, and Suneung Ahn. 2018. “Bayesian Approach for Estimating the Probability of Cartel Penalization under the Leniency Program.” *Sustainability* 10 (6): 1938. <https://doi.org/10.3390/su10061938>.

Ross, Sheldon M. n.d. *Introduction to Probability Models*.

Ross, Sheldon M. n.d. “Introduction to Probability Models, Tenth Edition.” Accessed November 1, 2018.

<https://pdfs.semanticscholar.org/13a3/db2415fb38ec77c165a5dfea1913a22655be.pdf>.

