

J. C. GÁZQUEZ ABAD *

M. SÁNCHEZ PÉREZ **

Consideración de la heterogeneidad en el comportamiento de elección del consumidor a través de modelos logit: enfoque paramétrico vs. semiparamétrico

SUMARIO: 1. Introducción. 2. Concepto y clasificación de heterogeneidad. 2.1. Concepto y clasificación de heterogeneidad. 2.2. Métodos orientados al análisis de la heterogeneidad observada. 2.3. Métodos orientados al análisis de la heterogeneidad no observada. 3. Metodología de la investigación. 3.1. Modelos de elección: modelización logit. 3.2. Selección de la categoría de producto y del tipo de datos. 3.3. Conjunto de elección. 3.4. Variables explicativas de la elección. 3.5. Modelos logit utilizados para analizar la heterogeneidad del consumidor. 4. Resultados. 4.1. Modelo básico sin heterogeneidad. 4.2. Modelo paramétrico con heterogeneidad en las preferencias y heterogeneidad estructural. 4.3. Modelo semiparamétrico de clases latentes. 4.4. Contraste estadístico entre los modelos analizados. 5. Conclusiones e implicaciones para la gestión, limitaciones y futuras líneas de investigación. Referencias bibliográficas

RESUMEN: Uno de los grandes retos que el marketing afronta es llegar a comprender la diversidad de preferencias y sensibilidades que existen en un mercado. En este trabajo, analizamos el concepto de heterogeneidad y su tratamiento en los modelos de elección del con-

* Profesor Asociado de Comercialización e Investigación de Mercados. Universidad de Almería. Departamento de Dirección y Gestión de Empresas. Facultad de Ciencias Económicas y Empresariales. Ctra. Sacramento s/n, 04120, La Cañada de San Urbano (Almería). Tel. 950 014035 Fax. 950 015178 e-mail: jcgazque@ual.es

** Catedrático de Comercialización e Investigación de Mercados. Universidad de Almería. Departamento de Dirección y Gestión de Empresas. Facultad de Ciencias Económicas y Empresariales. Ctra. Sacramento s/n, 04120, La Cañada de San Urbano (Almería). Tel. 950 015179 Fax. 950 015178 e-mail: msanchez@ual.es

sumidor. Para ello, y tras realizar una revisión de la literatura que aborda el concepto y la inclusión de la heterogeneidad en los modelos de elección, proponemos una clasificación de las diversas formas de incorporación de dicho comportamiento heterogéneo. Posteriormente, y utilizando datos de escáner contrastamos un modelo logit sin heterogeneidad con algunas formas de tratamiento de este aspecto, contrastando cuál de ellos es estadísticamente más significativo. A partir de estos resultados extraemos las principales conclusiones e implicaciones para la gestión.

ABSTRACT: One of the most important objectives in marketing field is to get a better understanding about consumer. In this paper, we have analysed the concept of heterogeneity and its treatment in consumer choice models. Following a review of the theoretical foundations and formulations about heterogeneity in consumer choice models, we proposed a classification of different kinds of heterogeneity. After this, and using scanner data we describe an empirical application in the context of consumer choice. Firstly, we developed a non heterogeneity MNL model. After that, we comparing it with a random effects model and with a latent segments ones. Results show that models which considered heterogeneity explain consumer choice behavior more efficiently. Finally, main contributions of the study are discussed.

1. Introducción

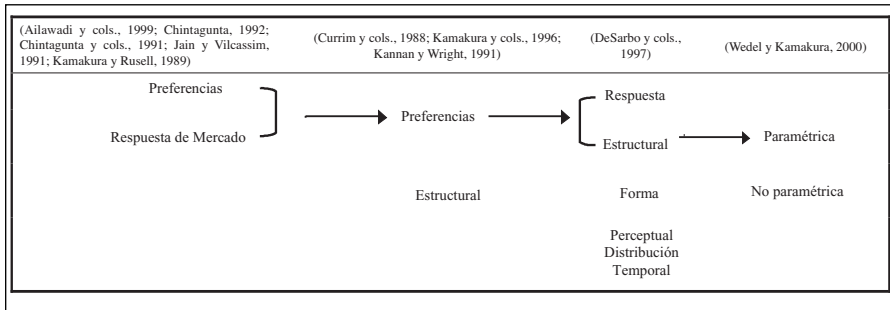
Uno de los grandes retos que el marketing afronta es llegar a comprender la diversidad de preferencias y sensibilidades que existen en un mercado (Allenby y Rossi, 1999). En este sentido, las razones que han llevado a los investigadores de marketing a introducir la heterogeneidad en sus trabajos de marketing han sido de doble índole (Cavero y Cebollada, 2000): en primer lugar, porque la no consideración de diferencias entre el comportamiento de los consumidores llevaría a estimaciones insesgadas y alejadas de la realidad del mismo y, en segundo lugar y desde una perspectiva de gestión, porque el conocimiento de los diferentes grupos existentes en el mercado es fundamental para el diseño de la estrategia comercial para la empresa. El objetivo de este trabajo se centra en analizar el concepto de heterogeneidad y las principales clasificaciones que del mismo se han realizado en la literatura sobre elección del consumidor. A partir de estos trabajos, realizamos una propuesta de clasificación de las fuentes de heterogeneidad del consumidor en su comportamiento de elección, contrastando empíricamente la validez de algunas de estas formas de tratamiento de la heterogeneidad a través de modelos logit. Para ello, hemos estructurado el trabajo en varias partes. Inicialmente, y tras realizar una breve aproximación teórica a los modelos logit, establecemos el concepto de heterogeneidad y su tratamiento en los modelos de elección, haciendo una revisión de las principales clasificaciones que se han realizado en la literatura para, posteriormente, proponer una clasificación que aglutina las diversas fuentes de heterogeneidad tratadas. A continuación, y a partir de datos de escáner, contrastamos un modelo logit multinomial en el que no se considera la heterogeneidad con modelos en los que sí se incorpora, tanto a través de un enfoque paramétrico como de un enfoque semiparamétrico, comparando cuál de ellos es estadísticamente más significativo. A partir de los resultados obtenidos, extraemos las principales conclusiones e implicaciones para la gestión.

2. Concepto y clasificación de heterogeneidad

2.1. CONCEPTO Y CLASIFICACIÓN DE HETEROGENEIDAD

De modo general, la *heterogeneidad* consiste en las diferencias entre individuos en cuanto a su comportamiento de elección debidas a dos tipos de factores: observados y no observados. Las diferencias en el comportamiento de elección consecuencia de factores observados (por ejemplo, características sociodemográficas del individuo) constituyen la heterogeneidad observada, mientras que las diferencias debidas a factores no observables por hábitos del consumidor constituyen la heterogeneidad no observada (Chintagunta, 1993). En la literatura se han elaborado diversas clasificaciones del concepto de heterogeneidad, evolucionando a medida que se ha venido profundizando en su naturaleza. A partir de los criterios de Ailawadi, Gedenk y Neslin (1999), Chintagunta (1992), Chintagunta, Jain y Vilcassim (1991), Currim, Meyer y Le (1988), DeSarbo, Ansari, Chintagunta, Himmelberg, Jedidi, Johnson, Kamakura, Lenk, Srinivasan y Wedel (1997), Jain y Vilcassim (1991), Kamakura, Kim y Lee (1996), Kamakura y Rusell (1989), Kannan y Wright (1991) y Wedel y Kamakura (2000), en la tabla 1 se recogen diversas clasificaciones elaboradas del concepto de heterogeneidad, así como la relación entre ellas.

TABLA 1.—*Principales clasificaciones del concepto de heterogeneidad en la literatura de marketing*



Fuente: Elaboración propia.

Una clasificación tradicional y utilizada con frecuencia es la que distingue entre *heterogeneidad de las preferencias* y *heterogeneidad de respuesta al mercado*, según que las diferencias individuales se refieran a la opción elegida o a los aspectos influyentes sobre dicha opción (Ailawadi *et al.*, 1999; Chintagunta, 1992; Chintagunta *et al.*, 1991; Jain y Vilcassim, 1991; Kamakura y Rusell, 1989):

Otra de las clasificaciones utilizadas en la literatura es la que distingue entre los conceptos *heterogeneidad de las preferencias* y *heterogeneidad estructural*, referida a las diferencias individuales en las opciones disponibles y en el proceso de elección (Currim *et al.*, 1988; Kamakura *et al.*, 1996; Kannan y Wright, 1991):

Wedel y Kamakura (2000) distinguen entre *heterogeneidad paramétrica* y *heterogeneidad no paramétrica*, en referencia a las preferencias del individuo y a su regla de decisión, respectivamente.

DeSarbo *et al.* (1997) realizan un análisis detallado de las posibles fuentes y tipos de heterogeneidad que pueden identificarse en el proceso de elección del consumidor, a partir de la siguiente función de utilidad:

$$V_{ij} = \mathfrak{S}_{it}[X_{ijkt}; \alpha_{ijt}; B_{it}] + \varepsilon_{ij} \quad [1]$$

donde:

$i = 1, \dots, I$ individuos; $j = 1, \dots, J$ alternativas de elección; $k = 1, \dots, K$ atributos de la alternativa;

V_{ij} = Utilidad de la alternativa j en el período t para el individuo i ;

α_{ijt} = constante; B_{it} = parámetros de respuesta para los K atributos;

X_{ijkt} = valor del atributo k para la alternativa j evaluado por el individuo i en el período t ;

\mathfrak{S}_{it} = forma funcional utilizada por el individuo i en el período t ;

ε_{ij} = término de error con distribución $g(\theta)$

Bajo este planteamiento general, DeSarbo *et al.* (1997) distinguen seis posibles fuentes de heterogeneidad:

1.- *Heterogeneidad de respuesta*: Recoge las diferencias en las preferencias hacia las distintas alternativas por parte de los consumidores. Viene recogida por el término constante α_{ijt} .

2.- *Heterogeneidad estructural*. Esta fuente de variación recoge las diferencias en el proceso de decisión. Los parámetros B_{it} pueden variar entre los consumidores, poniendo de manifiesto esta heterogeneidad estructural.

3.- *Heterogeneidad de forma*. Los consumidores difieren en la función de utilidad que utilizan para evaluar y valorar los diferentes atributos de las alternativas. Viene recogida por el término funcional \mathfrak{S}_{it} .

4.- *Heterogeneidad perceptual*. La percepción de los atributos de las diferentes alternativas puede ser muy diferente entre consumidores. Esta heterogeneidad viene recogida por el término X_{ijkt} .

5.- *Heterogeneidad de la distribución*. El término de error puede variar de dos maneras diferentes, bien a través de los parámetros de la distribución del error $g(\theta)$, o bien a través de las distribuciones.

6.- *Heterogeneidad temporal*. Los consumidores difieren en su reacción frente a comportamientos anteriores. La heterogeneidad temporal puede afectar a cualquier aspecto de la función de utilidad planteada en [1].

Esta clasificación de DeSarbo *et al.* (1997) es la más exhaustiva, entendiendo, como se señala en la tabla 1, que la heterogeneidad de las preferencias de Currim *et al.* (1988), Kamakura *et al.* (1996) y Kannan y Wright (1991) equivale a la heterogeneidad de respuesta más la estructural, mientras que la heterogeneidad estructural de aquéllos equivaldría a la heterogeneidad

de forma de DeSarbo *et al.* (1997). Por otra parte, la heterogeneidad paramétrica de Wedel y Kamakura (2000) se corresponde con la heterogeneidad estructural de DeSarbo *et al.* (1997), mientras que la no paramétrica se corresponde con la heterogeneidad de forma.

Un aspecto importante y determinante en el diseño y construcción de los modelos de elección en marketing es el nivel de agregación. Un *modelo desagregado* se define como «aquél que utiliza observaciones individuales para estimar los parámetros de la población», mientras que un *modelo agregado* es «el que utiliza datos que incluyen solamente información sobre grupos de individuos» (Gensch, 1985:462). A partir de esta distinción, pueden considerarse tres niveles diferentes de agregación cuando se estudia el fenómeno de la heterogeneidad (DeSarbo *et al.*, 1997): *nivel agregado*, *nivel de segmento de mercado* y *nivel individual*.

En el primero (por ejemplo, Guadagni y Little, 1983), la función de respuesta del consumidor es estimada agrupando los datos recogidos para toda la muestra. Por tanto, se asume que el proceso de elección es homogéneo para todos los individuos de la población.

Sin embargo, la modelización de la elección a nivel de segmento puede resultar más adecuada que la consideración del mercado a nivel agregado según la extensión de la heterogeneidad (Currim, 1981). Igualmente, y a medida que las preferencias y sensibilidades del consumidor se vuelven más diferentes, resulta menos eficiente considerar el mercado de modo agregado (Allenby y Rossi, 1999) y es más conveniente considerar el nivel individual, si bien para ello es necesario contar con una información sustanciosa y amplia para cada consumidor.

Otro de los criterios fundamentales para el análisis del fenómeno de la heterogeneidad es la estructura del proceso de elección. Así, existen consumidores que siguen un proceso de decisión compensatorio, mientras que existen otros que siguen un proceso de elección jerárquica no compensatorio, en el que la elección de una alternativa pasa por una o varias decisiones intermedias. En la tabla 2 presentamos una clasificación de los diferentes métodos de análisis e incorporación de la heterogeneidad en base a los diferentes criterios analizados.

Como puede observarse, se parte de la dicotomización *heterogeneidad observada vs. no observada* como punto de partida de la clasificación de los métodos de análisis e incorporación de la heterogeneidad de los consumidores en el comportamiento de elección (véase tabla 3 para un resumen de estos métodos).

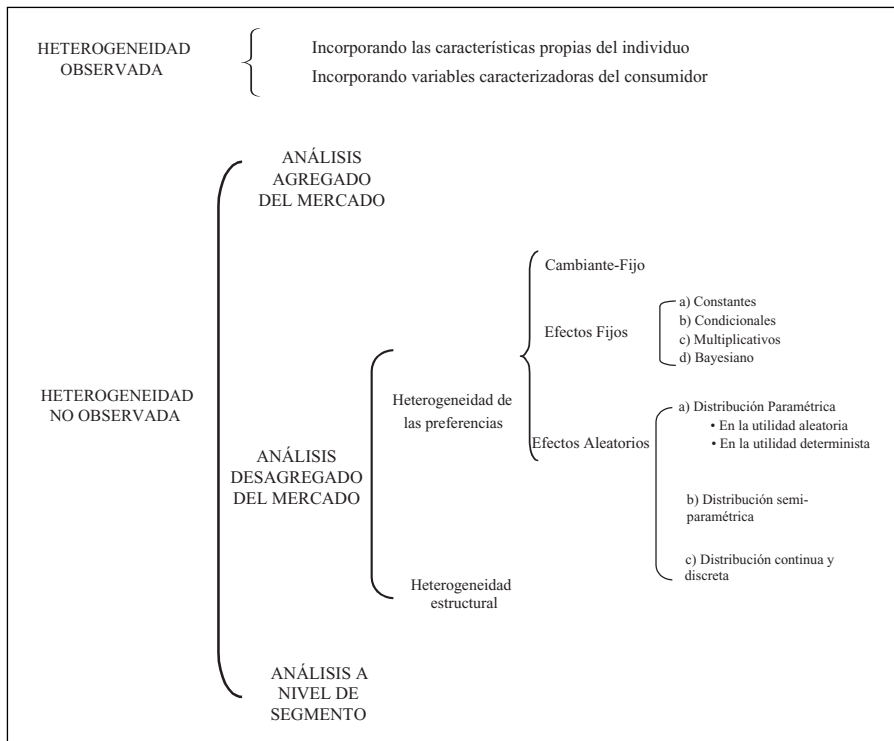
2.2. MÉTODOS ORIENTADOS AL ANÁLISIS DE LA HETEROGENEIDAD OBSERVADA

En la literatura se identifican dos enfoques principales para la medición de la heterogeneidad observada (Ailawadi *et al.*, 1999): un primer enfoque basado en la especificación de la experiencia previa de elección del consumidor, y otro basado en la incorporación de variables específicas y propias del consumidor.

2.2.1. *Métodos basados en la incorporación de la experiencia previa de elección del consumidor*

Este método consiste en la inclusión en la función de utilidad de la proporción de ocasiones de elección en que la alternativa j ha sido elegida, durante un período de inicialización por cada individuo. En la literatura suele denominarse BL (*brand loyalty*) o PREPREF (*previous preference*). Para que no existan sesgos significativos en la inclusión de esta variable se requiere un *período de inicialización* (período que sirve para recoger la experiencia previa del individuo), lo suficientemente largo.

TABLA 2.—*Clasificación de los métodos de análisis e incorporación de la heterogeneidad del consumidor en el comportamiento de elección*



Fuente: Elaboración propia.

2.2.2. *Métodos basados en la especificación de características propias del consumidor*

Otra posibilidad consiste en la utilización de variables que recojan las características propias del individuo (por ejemplo, características sociodemográficas). Las variables demográficas tienen una larga tradición de utilización

TABLA 3.—Principales enfoques de tratamiento de la heterogeneidad en los modelos de elección

		Especificación	Ventaja	Inconveniente
Heterogeneidad observada	1. Incorporando la experiencia previa de elección	Inclusión en la función de utilidad de la proporción de ocasiones de elección de cada alternativa durante un período de inicialización	Es una forma muy parsimoniosa de recoger el comportamiento de lealtad	Es necesario que el período de inicialización sea lo suficientemente largo para que no existan sesgos
	2. Incorporando características del individuo	Inclusión en la función de utilidad de variables que recojan características del individuo (p.e., económicas o sociodemográficas)	Permite a la empresa desarrollar estrategias de marketing y de posicionamiento específicas	No son capaces de capturar, por sí solas, todas las posibles diferencias entre los individuos
Heterogeneidad no observada	1. Efectos fijos	El parámetro de preferencia hacia cada alternativa (α) es constante y específico para cada individuo	Permite que las preferencias sean, realmente, heterogéneas	Es necesario estimar un gran número de parámetros y contar con un elevado número de observaciones
	2. Efectos aleatorios	• Paramétrica	Se asume una forma funcional determinada para la distribución de heterogeneidad entre los consumidores	La posible elección incorrecta de la función de distribución de la heterogeneidad
		• Semiparamétrica	La forma funcional de la distribución de heterogeneidad es estimada empíricamente a partir de los datos disponibles	Permite obtener segmentos de mercado latentes en función del comportamiento de los consumidores

Fuente: Elaboración propia.

en el área de marketing, habiéndose utilizado para la segmentación de mercados o la localización de establecimientos de venta debido, tal y como indican Kalyanam y Putler (1997), a las numerosas ventajas que presentan.

2.3. MÉTODOS ORIENTADOS AL ANÁLISIS DE LA HETEROGENEIDAD NO OBSERVADA

Las diferencias en el comportamiento debidas a factores no observados se deben a aspectos tales como información de interés del consumidor en el producto objeto de elección, factores demográficos como la raza o la edad, nivel de ingresos, etc., los cuales no están siempre disponibles. En este contexto, el concepto de heterogeneidad no observada quedaría representado en el modelo, permitiendo que en la función de utilidad del individuo definida en [1] los parámetros de preferencias intrínsecas (α_{ijt}) y algunos de los elementos del vector de parámetros (B_{it}) sean específicos para cada individuo. Para analizar la heterogeneidad no observada se han desarrollado diversos procedimientos a medida que se han producido avances en las técnicas analíticas. Ailawadi *et al.* (1999) agrupan los diversos trabajos sobre heterogeneidad en la elección según qué aspecto de la heterogeneidad han considerado (preferencias y/o respuesta), si se trata de *heterogeneidad observada o no observada* y, finalmente, según la modelización utilizada. Yamaguchi (1986) distingue entre modelos de *efectos fijos y aleatorios*.

La estimación de un parámetro constante para cada uno de los consumidores que forman el mercado y para cada marca, es denominada en la literatura de marketing como *modelo de efectos fijos* (Chamberlain, 1980):

$$U_{ijt} = \alpha_j + \sum_{k \in K} \beta_k X_{ijkt} + \varepsilon_{ijt} \quad [2]$$

En la formulación de [2], los parámetros de preferencia de las alternativas $\{\alpha_j\}_{j \in J}$ son comunes a todos los individuos y diferentes para cada una de las alternativas, mientras que los parámetros de respuesta $\{\beta_k\}_{k \in K}$ no varían en el tiempo entre individuos ni alternativas, sino que son comunes para cada uno de los atributos o variables explicativas.

Una especificación más general sería proporcionar *preferencias específicas de alternativas* para cada individuo y cada alternativa. El modelo se formularía según se recoge en la siguiente expresión:

$$U_{ijt} = \alpha_{ij} + \sum_{k \in K} \beta_k X_{ijkt} + \varepsilon_{ijt} \quad [3]$$

El principal problema que presenta este modelo es la necesidad de estimar un gran número de parámetros, lo que obliga a tener que contar con un elevado número de observaciones para cada consumidor, lo que ha llevado a una escasa utilización de los modelos de efectos fijos en la investigación de marketing.

Una forma alternativa de incorporar la heterogeneidad entre los consumidores son los llamados modelos de efectos aleatorios, los cuales consisten en

suponer que dicha heterogeneidad está repartida entre la población siguiendo una determinada distribución de probabilidad. Chintagunta *et al.* (1991) distinguen dentro de estos modelos entre estimaciones *paramétricas* y *no paramétricas*. La aproximación *paramétrica* consiste en asumir una determinada forma funcional para la distribución de la heterogeneidad entre los consumidores y estimar los parámetros de dicha función. Considerando el caso más general en que preferencias intrínsecas y parámetros de respuesta de las alternativas varían entre individuos, la utilidad del consumidor toma la siguiente forma :

$$U_{ijt} = \alpha_{ij} + \sum_{k \in K} \beta_{ik} X_{ijkt} + \varepsilon_{ijt} \quad [4]$$

donde $\tilde{\alpha}_{ij}^1, \tilde{\beta}_{ij} \sim \mathfrak{S}(\cdot)$

$\mathfrak{S}(\cdot)$ denota una distribución continua de los parámetros normal (Chintagunta, 2001; Chintagunta *et al.*, 1991; Gönül y Srinivasan, 1993; Papatla y Krishnamurthi, 1996), gamma (Chintagunta, 1993; Chintagunta y Honore, 1996; Chintagunta *et al.*, 1991), o ambas conjuntamente (Arora, Allenby y Ginter, 1998). El enfoque de efectos aleatorios supone que la heterogeneidad entre la población se distribuye según alguna función de distribución predefinida. El principal problema de estos modelos consiste en el error en la selección de la función de distribución de los parámetros, lo que llevaría a estimaciones inconsistentes y sesgadas de la distribución de la heterogeneidad. Para solucionar este problema, se puede utilizar una aproximación *no paramétrica* o también llamada *semiparamétrica*, en la que no se presupone ninguna forma funcional para la distribución de probabilidad de la heterogeneidad, sino que dicha distribución es estimada de forma empírica con los datos de mercado.

Este enfoque consiste en asumir una distribución conjunta discreta de los parámetros de respuesta, suponiendo la existencia de un conjunto de S posibles perfiles de respuesta y de un reparto $\{P(s)\}_{s \in S}$ de la probabilidad. Cada uno de estos perfiles implica la existencia de un segmento latente con tamaño relativo dado por su probabilidad. De acuerdo con el modelo de Kamakura y Russell (1989), la probabilidad de que el consumidor i elija la alternativa j condicionada a la pertenencia a un segmento s vendrá definida por:

$$P_{ij|s} = \frac{\exp(\alpha_{ji} + \sum_{k \in K} \beta_{ks} X_{ijk})}{\sum_{s \in S} \exp(\alpha_{js} + \sum_{k \in K} \beta_{ks} X_{ijk})} \quad [5]$$

donde α_{js} y β_{ks} representan las preferencias intrínsecas y los parámetros de respuesta, respectivamente, de los segmentos. Cada segmento S representa un perfil de respuesta o familia de parámetros ($\{\alpha_{js}\}_{j \in J}, \{\beta_{ks}\}_{k \in K}$). Este conjunto

¹ El símbolo \sim colocado encima de un parámetro, significa que el mismo es aleatorio.

de segmentos configuran la distribución de probabilidad, con un tamaño relativo definido por la probabilidad de ocurrencia $\{P(s)\}_{s \in S}$, de acuerdo con una formulación logística que proporciona unos valores de probabilidad inferiores a la unidad ($0 \leq P_s \leq 1$):

$$P_s = \frac{\exp \lambda_s}{\sum_{s'} \exp \lambda_{s'}} \quad [6]$$

siendo $\sum_{s \in S} \lambda_s = 1$. Como el número de clases latentes (S) es desconocido, la estimación semiparamétrica es llevada a cabo condicionada a un valor previo de S , para lo cual se utilizan heurísticos con diferentes combinaciones en el número de segmentos, minimizando estadísticos como AIC (*Criterio de Información de Akaike*), CAIC (*Criterio de Información de Akaike Condicionado*) o BIC (*Criterio de Información de Bayes*).

A partir de las probabilidades estimadas en [6], se puede calcular la probabilidad de pertenencia de un individuo a cada segmento $P_{is|H_i}$ condicionada a la historia de elección observada-probabilidades a posteriori.

$$P_{is|H_i} = \frac{P_{H_i|s} P_s}{\sum_{s' \in S} P_{H_i|s'} P_{s'}} \quad [7]$$

Chintagunta *et al.* (1991) y Jain y Vilcassim (1991), entre otros, utilizan esta aproximación. Así, Chintagunta *et al.* (1991) comparan la aproximación paramétrica y la semiparamétrica, llegando a la conclusión de que la segunda produce estimaciones más fiables.

En este sentido, la identificación de segmentos dentro del mercado viene siendo considerada como una de las formas fundamentales de tratamiento de la heterogeneidad del mercado, constituyendo una necesidad estratégica de la organización para el diseño de estrategias diferenciadas y la asignación de recursos entre los diferentes grupos detectados (Beane y Ennis, 1987; Wind, 1978). Igualmente, la organización debe decidir qué estrategia de posicionamiento adoptar en los grupos identificados, buscando la diferenciación de su oferta respecto a sus competidores (Corstjens y Corstjens, 1995).

Por todo esto, cabe preguntarse si los resultados derivados del análisis entre la modalidad paramétrica y semiparamétrica de incorporación de la heterogeneidad en el MNL, ponen de manifiesto la superioridad de la identificación de segmentos de mercado con diferentes comportamientos, en consonancia con el desarrollo de estrategias de cobertura del mercado que están desarrollando las empresas en la actualidad.

3. Metodología de la investigación

3.1. MODELOS DE ELECCIÓN: MODELIZACIÓN LOGIT

La formalización cuantitativa de la teoría de la utilidad aleatoria asume una descomposición aditiva de la utilidad V_{ij} de la alternativa j para el individuo i en la ocasión t en una componente determinista o sistemática U_{ij} y una componente aleatoria ε_{ij} , es decir:

$$V_{ij} = U_{ij} + \varepsilon_{ij} \quad [8]$$

El modelo tipo logit más extensamente utilizado es el modelo MNL (*MultiNomial Logit model*), cuya formalización se atribuye a McFadden (1974). La utilidad U_{ij} de una alternativa se representa como una función aditiva de un conjunto K de variables explicativas $\{X_{ijk}\}_{k \in K}$ ponderada por parámetros de respuesta β_k .

$$U_{ij} = \sum_{k \in K} \beta_k X_{ijk} \quad [9]$$

de modo que el modelo se resume en la siguiente expresión:

$$P_{ij} = \frac{\exp(\sum_{k \in K} \beta_k X_{ijk})}{\sum_{m \in M_{it}} \exp(\sum_{k \in K} \beta_k X_{imk})} \quad [10]$$

En econometría es muy habitual una especificación de este modelo que incluya interacciones de los atributos y de variables *dummy* específicas de las alternativas, que en el caso de la probabilidad P_{ij} toma valor 1 para dicha alternativa j y 0 para el resto. Es decir, la utilidad U_{ij} quedaría del siguiente modo:

$$U_{ij} = \alpha_j X_{ij} + \sum_{k \in K} \beta_k X_{ijk} \quad [11]$$

Por tanto, el número de parámetros a estimar (B) es igual al número de atributos más el número de alternativas menos uno, es decir, $B = (K+J-1)$.

Al no ser lineales, los modelos logit requieren procedimientos distintos a los modelos lineales para analizar la bondad del ajuste. Malhotra (1984) recomienda utilizar más de un criterio para evaluar el poder predictivo del modelo, así como evaluar la bondad del ajuste en términos relativos y no absolutos, es decir examinando distintas especificaciones del modelo. Estos criterios de comparación de los modelos deben considerar tanto la capacidad explicativa del modelo como el grado de parsimonia del mismo (Cavero y Cebollada, 1999). Entre los métodos y criterios disponibles, destacamos los siguientes (Guadagni y Little, 1983; Malhotra, 1984):

- Valor de la función de verosimilitud. La estimación de los modelos logit busca maximizar el valor de la función de verosimilitud $LL(\beta)$, por ello, en la elección de especificaciones alternativas, un primer criterio es el de máxima función de verosimilitud del modelo estimado.
- Índice de la razón de verosimilitud (ρ^2). Indica el grado de variabilidad del modelo que es explicado por los coeficientes del mismo, teniendo así un significado similar al coeficiente de determinación R^2 en el modelo lineal. Sin embargo, como indica Novales (1993:547), este coeficiente es difícil de interpretar, porque «ni llega a valer 1 ni es fácil interpretar los valores entre 0 y 1». En cuanto al valor del estadístico que se considera aceptable, McFadden (1978) indica que un coeficiente entre 0,2 y 0,4 es indicador de un ajuste satisfactorio del modelo². La expresión del estadístico es:

$$\rho^2 = 1 - \frac{LL(\beta)}{LL(C)} \quad [12]$$

El valor del estadístico aumenta con la inclusión de parámetros adicionales, por lo que se recomienda, la corrección con el número de parámetros a estimar (B) (Ahn y Ghosh, 1989):

$$\bar{\rho}^2 = 1 - \frac{LL(\beta) - B}{LL(C)} \quad [13]$$

- Contraste de la razón de verosimilitud. Permite comprobar que el ajuste del modelo estimado es mejor que el de un modelo restringido en los parámetros, es decir, que incluya un número menor de variables explicativas. Se hace operativo a través de un estadístico que compara el valor de ambas funciones de verosimilitud, y que se distribuye como una chi-cuadrado cuyos grados de libertad son la diferencia entre el número de parámetros a estimar en el modelo no restringido y restringido.
- Criterios de información teórica. Este tipo de estadísticos consideran que es mejor aquel modelo que combina un bajo número de parámetros con un alto valor de la función de verosimilitud, por lo que, según indica Malhotra (1984), son especialmente indicados para contrastar modelos con diferentes grados de libertad. Entre estos criterios teóricos, destacan el Criterio AIC, y el Criterio BIC, si bien existen otros que, posteriormente, en el análisis empírico utilizaremos.

² En efecto, observando los resultados de investigaciones previas, los valores del coeficiente más habituales se sitúan en un rango entre 0,2 y 0,3, como por ejemplo Ahn y Ghosh (1989), que obtienen un valor de 0,22.

3.2. SELECCIÓN DE LA CATEGORÍA DE PRODUCTO Y DEL TIPO DE DATOS

Los datos utilizados para esta investigación han sido recogidos en un hipermercado de un grupo internacional de distribución con presencia en todo el país. Se trata de datos de escáner de elección de marcas de aceite de oliva 0,4º en el formato de botellas de 1 litro. La elección de la categoría del aceite de oliva se fundamenta en dos aspectos: el primero y principal, viene justificado por el importante protagonismo y tradición que este producto posee en España, tanto desde el punto de vista económico, cultural y social; y el segundo aspecto, se fundamenta en su elevada frecuencia de compra y utilización. Los datos de compra abarcan un período total de 53 semanas (entre el 1 de Enero y el 31 de diciembre de 2002).

3.3. CONJUNTO DE ELECCIÓN

El conjunto de elección está constituido por diez marcas. Estas marcas son, en orden decreciente de cuota de mercado: *Carbonell*, *Marca de distribuidor (MdD)*, *Coosur*, *La Masía*, *La Española*, *Elosúa*, *Giraldal*, *Ybarra*, *Marca de primer precio* y *Mueloliva*. Debido a la necesidad de construir un panel para conocer el comportamiento de cada individuo a lo largo del período analizado, solo se han considerado aquellos consumidores que realizaron su pago con la tarjeta del establecimiento. Por otra parte, el período total de 53 semanas ha sido dividido en dos subperíodos: un primer subperíodo llamado de inicialización, y que hemos utilizado para el cálculo de una variable de agregación que capture el comportamiento del consumidor en el tiempo (lealtad histórica), y un período denominado de calibrado, que ha sido utilizado para la estimación de los parámetros del modelo. El período de inicialización abarca desde la semana 1 a la semana 20³, mientras que el período de calibrado comprende desde la semana 21 a la semana 53. Una vez realizada esta división, y siguiendo a Sivakumar y Raj (1997), hemos considerado, únicamente, aquellos individuos que han realizado, al menos dos compras, en cada uno de los subperíodos. De este modo, el conjunto definitivo de datos de escáner utilizados está constituido por 389 individuos, que realizan un total de 3241 compras⁴ (8,33 ocasiones de compra por individuo).

³ En la línea de otros trabajos de investigación que utilizan como período de inicialización, aproximadamente, el 40% del período total (por ejemplo, Guadagni y Little, 1983 ó Krishnamurthi, Mazumdar y Raj, 1992)

⁴ Se refiere a ocasiones en las que los individuos han acudido al hipermercado y han elegido alguna de las diez marcas analizadas, no al número de unidades adquirido. No obstante, en el software utilizado para la estimación del modelo se ha tenido en cuenta, igualmente, el número de unidades adquiridas en cada ocasión de compra.

3.4. VARIABLES EXPLICATIVAS DE LA ELECCIÓN

Para cada alternativa y en la unidad de tiempo definida (semana), se han considerado dos tipos de variables explicativas de la elección del consumidor: por una parte, variables de marketing relativas a las alternativas de elección, y por otra, variables relativas al consumidor. La elección de las variables utilizadas en este trabajo se justifica desde una doble perspectiva: por un lado, desde la disponibilidad de información suministrada por el hipermercado y, por otro, desde la intención de los autores de analizar el efecto que sobre el comportamiento de elección tienen las variables que el distribuidor maneja en el punto de venta, por lo que variables como la publicidad que realizan las marcas en medios masivos o su imagen, no han sido tenidas en cuenta.

En el primer grupo, se incluyen las siguientes variables: precio de compra, precio de referencia, rebaja absoluta del precio de compra, rebaja relativa del precio de compra, promoción en precio y presencia en folleto promocional. Por otra parte, en el segundo grupo de variables explicativas se definieron dos variables relativas a la lealtad del consumidor hacia las alternativas que configuran el conjunto de elección. Para ello se obtienen las cuotas de mercado⁵ de aquellos individuos que están presentes tanto en el período de inicialización como de calibrado, y que han realizado, al menos, 2 compras en cada uno de esos períodos. Estas cuotas de mercado se indexan al período completo, manteniéndose fijas por individuo y semana a lo largo del período analizado (lealtad histórica); del mismo modo, a esos clientes se les indexa una medición binaria de su grado de lealtad a la marca, demostrada en acciones de compra sucesivas en los períodos analizados (lealtad anterior). En la tabla 4 se refleja la estadística descriptiva de las variables explicativas utilizadas.

3.5. MODELOS LOGIT UTILIZADOS PARA ANALIZAR LA HETEROGENEIDAD DEL CONSUMIDOR

Con el objetivo de contrastar las hipótesis relativas a la heterogeneidad en el comportamiento de elección del consumidor, hemos estimado diferentes modelos multinomiales utilizando las variables explicativas definidas, con objeto de contrastar cuál de ellos permite explicar mejor el comportamiento de elección a partir de diversos criterios estadísticos y de la capacidad de parsimonia del modelo. Inicialmente, desarrollamos un modelo centrado en el

⁵ Hace referencia a la proporción que supone cada marca en relación al total de marcas adquiridas durante el período de inicialización (p.e., si en el período de inicialización el individuo ha adquirido 5 unidades, 3 de la marca Carbonell, 1 de la MdD y 1 de Coosur, Carbonell tendría una cuota de 3/5, mientras que la MdD y Coosur tendrían una cuota de 1/5 cada una. El resto de marcas tendrían una cuota de mercado igual a cero. Esos valores son los que se indexan al período completo.

precio y las promociones relacionadas con el mismo como variables explicativas de la elección. Además, incluimos las variables relativas al nivel de descuento realizado en la promoción del precio.

$$U_{ij} = \beta_0 X_j + \beta_1 \text{PCOM_ACT}_{jt} + \beta_2 \text{PREF}_{jt} + \beta_3 \text{PRO_PREC}_{jt} + \beta_4 \text{RB_PR_AB}_{jt} + \beta_5 \text{RB_PR_RE}_{jt} \quad [14]$$

A continuación, al modelo anterior le añadimos la variable relativa a la presencia de la marca en los folletos publicitarios del establecimiento, para considerar, igualmente, otra herramienta promocional además de las tradicionales promociones en precio.

$$U_{ij} = \beta_0 X_j + \beta_1 \text{PCOM_ACT}_{jt} + \beta_2 \text{PREF}_{jt} + \beta_3 \text{PRO_PREC}_{jt} + \beta_4 \text{RB_PR_AB}_{jt} + \beta_5 \text{RB_PR_RE}_{jt} + \beta_6 \text{FOLLETO}_{jt} \quad [15]$$

El modelo [16] puede ser considerado el «modelo completo», ya que en él vamos a incorporar el efecto que sobre el comportamiento de elección del consumidor tienen todas las variables explicativas.

$$U_{ij} = \beta_0 X_j + \beta_1 \text{PCOM_ACT}_{jt} + \beta_2 \text{PREF}_{jt} + \beta_3 \text{PRO_PREC}_{jt} + \beta_4 \text{RB_PR_AB}_{jt} + \beta_5 \text{RB_PR_RE}_{jt} + \beta_6 \text{FOLLETO}_{jt} + \beta_7 \text{LEAL_HIS}_{ijt} + \beta_8 \text{LEAL_ANT}_{ijt} \quad [16]$$

Una vez que hemos utilizado la totalidad de variables explicativas analizadas, en el último modelo introducimos únicamente, y con objeto de lograr una mayor parsimonia, tres variables explicativas: el precio de compra, por considerar que es fundamental en el comportamiento de elección, la presencia de las marcas en los folletos publicitarios (en detrimento de la promoción del precio, ya que no se trata de una herramienta muy frecuentemente utilizada, tal y como podemos observar en la tabla 4), y una de las variables de lealtad, la lealtad anterior.

$$U_{ij} = \beta_0 X_j + \beta_1 \text{PCOM_ACT}_{jt} + \beta_2 \text{FOLLETO}_{jt} + \beta_3 \text{LEAL_ANT}_{ijt} \quad [17]$$

A partir de este modelo, desarrollamos un enfoque paramétrico con coeficientes aleatorios tanto en las preferencias del individuo como en las variables explicativas para, posteriormente, plantear un modelo semiparamétrico de clases latentes, que nos va a permitir identificar segmentos de mercado en función de su comportamiento de elección. A partir de estas tres modelizaciones, realizamos la comparación entre ellas, con objeto de identificar qué tipo de representación de la heterogeneidad es la que mejor se ajusta al comportamiento de elección observado en el mercado del aceite de oliva analizado. El software utilizado ha sido Limdep 7.0. y su módulo Nlogit 3.0.

TABLA 4.—*Estadísticos descriptivos de las variables objeto de análisis*

	Carbonell	MdD	Coosur	Masía	Española	Elosúa	Giralda	Ybarra	1er precio	Mueloliva
Cuota de mercado (VarY)										
Media	0,33	0,23	0,22	0,04	0,07	0,033	0,032	0,03	0,013	0,002
Máximo	1	1	1	1	1	1	1	1	1	1
Mínimo	0	0	0	0	0	0	0	0	0	0
Desviación	0,42	0,37	0,34	0,14	0,18	0,11	0,1	0,12	0,04	0,001
Precio compra (PCOM_ACT)										
Media	2,69	2,04	2,34	2,42	2,4	2,35	2,26	2,44	2,09	2,55
Máximo	2,69	2,1	2,39	2,49	2,45	2,35	2,59	2,45	2,09	2,55
Mínimo	2,67	1,99	2,29	2,37	2,08	2,35	1,95	1,75	2,09	2,45
Desviación	0,0006	0,023	0,03	0,051	0,031	0	0,32	0,036	0	0,02
Promoción de precio (PROPREC)										
Media	0,0016	0,071	0,012	0,12	0,05	0	0,044	0,05	0	0
Máximo	1	1	1	1	1	0	1	1	0	0
Mínimo	0	0	0	0	0	0	0	0	0	0
Desviación	0,04	0,26	0,11	0,32	0,22	0	0,2	0,21	0	0
Precio referencia (PREF)										
Media	2,69	2,05	2,34	2,42	2,4	2,35	2,26	2,44	2,09	2,55
Máximo	2,69	2,1	2,39	2,49	2,45	2,35	2,59	2,49	2,09	2,55
Mínimo	2,69	1,99	2,29	2,37	2,08	2,35	1,95	2,37	2,09	2,45
Desviación	0	0,02	0,03	0,05	0,03	0	0,31	0,02	0	0,002
Rebaja absoluta precio (RBPRAB)										
Media	0,22	0,42	0,06	0,68	0,33	0	0,97	0,41	0	0
Máximo	2	8	5	12	7	0	50	70	0	0
Mínimo	0	0	0	0	0	0	0	0	0	0
Desviación	0,06	1,65	0,54	2,49	1,43	0	5,52	3,14	0	0
Rebaja relativa precio (RBPRE)										
Media	0,0008	0,2	0,02	0,27	0,13	0	0,42	0,17	0	0
Máximo	0,74	3,86	2,09	4,82	2,86	0	19,3	28,6	0	0
Mínimo	0	0	0	0	0	0	0	0	0	0
Desviación	0,02	0,79	0,23	1	0,59	0	2,24	1,28	0	0

Presencia en folleto(FOLLETO)										
Media	0,37	0,33	0,06	0,19	0,19	0	0	0,07	0	0
Máximo	1	1	1	1	1	0	0	1	0	0
Mínimo	0	0	0	0	0	0	0	0	0	0
Desviación	0,48	0,47	0,24	0,39	0,39	0	0	0,26	0	0
Lealtad histórica (LEAL_HIS)										
Media	0,37	0,21	0,16	0,09	0,07	0,056	0,006	0,02	0,11	0
Máximo	1	1	1	1	1	1	0,5	1	0,67	0
Mínimo	0	0	0	0	0	0	0	0	0	0
Desviación	0,44	0,36	0,3	0,21	0,2	0,15	0,04	0,12	0,07	0
Lealtad actual (LEAL_ANT)										
Media	0,29	0,19	0,15	0,02	0,035	0,016	0,009	0,01	0,002	0
Máximo	1	1	1	1	1	1	1	1	1	0
Mínimo	0	0	0	0	0	0	0	0	0	0
Desviación	0,45	0,39	0,36	0,14	0,18	0,12	0,09	0,12	0,05	0

4. Resultados

4.1. MODELO BÁSICO SIN HETEROGENEIDAD

En la tabla 5 podemos observar los parámetros de los modelos básicos sin heterogeneidad (modelos 14, 15, 16 y 17), así como los coeficientes de bondad de cada uno de estos ajustes.

Como podemos observar en la tabla 5, en los tres primeros modelos, y a medida que introducimos variables explicativas, la preferencia hacia cada marca se reduce. Incluso algunas marcas (por ejemplo, la *MdD*) ven como la preferencia positiva y significativa del modelo [14] se convierte en una preferencia negativa e, incluso, no significativa a medida que introducimos variables explicativas en los modelos. No obstante vemos como la mayor preferencia es, en todos los casos, hacia la marca *Carbonell*, seguida de otras marcas nacionales como *Coosur* y *La Española*. Este hecho confirma la relación observada, entre otros, por Kamakura y Russell (1989) y Krishnamurthi y Raj (1991) entre el valor de una marca y su precio, ya que la alternativa con mejor valor intrínseco es la que presenta el precio más elevado. Igualmente, podemos observar la baja significatividad de la variable relativa a la promoción en precio y al nivel de rebaja (incluso el coeficiente del descuento relativo es negativo), lo que confirma lo apuntado anteriormente acerca de la escasa importancia que en este mercado tienen las promociones del precio debido, fundamentalmente, al conocimiento y a la implicación del consumidor español con el aceite de oliva.

Para determinar cuál de los modelos es superior, analizamos los criterios de bondad del ajuste; como podemos observar en la tabla 5, existe una importante diferencia entre los 2 primeros modelos y los dos últimos en términos del valor de la función de verosimilitud, los estadísticos AIC y BIC y el coeficiente ρ^2 . Dentro de estos dos últimos modelos, observamos como el modelo que hemos denominado «completo» presenta resultados levemente superiores en todos los criterios (AIC, BIC, función de verosimilitud y coeficiente ρ^2). Esto, unido a que existen diferencias estadísticas significativas ($p < 0,01$) entre ambos modelos (estadístico del test de diferencias significativas de 175,37 [5 g.l.]), hace pensar en la conveniencia de utilizar este modelo. Sin embargo, y siguiendo principios de parsimonia en la construcción de modelos, creemos que la consideración de 17 parámetros a estimar en el modelo [16] (8 variables + 9 coeficientes de las alternativas) es excesiva, lo que complicaría las estimaciones paramétricas y semiparamétricas posteriores. Es por ello que nos inclinamos por utilizar el modelo [17] en los posteriores análisis de este trabajo, ya que con tan sólo 3 variables explicativas es capaz de predecir más del 65% del comportamiento de elección del consumidor, siendo este porcentaje únicamente un 1% inferior al que es capaz de explicar el modelo [16] con 5 variables explicativas más; además, también presenta valores de $LL(\beta)$ y de los criterios AIC y BIC muy similares al modelo «completo».

TABLA 5.—Parámetros estimados y criterios de evaluación de bondad de los modelos sin heterogeneidad

	Modelo 14	Modelo 15	Modelo 16	Modelo 17
<i>Preferencia hacia cada alternativa^a</i>				
Carbonell	6,435* (0,522)	6,27* (0,062)	3,961* (0,444)	4,355* (0,519)
MdD	2,978* (0,513)	3,042* (0,552)	-0,405 (0,573)	0,091 (0,641)
Coosur	4,269* (0,536)	4,322* (0,563)	2,217* (0,586)	2,566* (0,532)
La Masía	3,038* (0,517)	3,025* (0,525)	1,489** (0,519)	1,845* (0,524)
La Española	3,514* (0,511)	3,5* (0,521)	2,087* (0,577)	2,344* (0,521)
Elosúa	2,667* (0,504)	2,732* (0,465)	1,258** (0,621)	1,552* (0,535)
Giralda	1,044** (0,572)	1,221** (0,596)	-0,563 (0,741)	-0,348 (0,691)
Ybarra	2,767* (0,534)	2,778* (0,511)	1,513** (0,541)	1,642 (0,529)
Marca de primer precio	0,165 (0,634)	0,314 (0,753)	-1,219*** (0,693)	-0,989*** (0,639)
Mueloliva	—	—	—	—
<i>Parámetros específicos de las variables explicativas^a</i>				
Precio de compra	-2,428* (0,751)	-2,266** (0,745)	-3,639 (0,849)	-6,805* (0,783)
Precio de referencia	-2,425* (0,162)	-2,263* (0,13)	3,633 (2,121)	—
Promoción en precio	0,00658 (0,136)	-0,274* (0,115)	0,00044 (0,658)	—
Reducción absoluta del precio	0,352* (0,137)	0,288 (0,506)	0,596 (0,502)	—
Reducción relativa del precio	-0,766 (0,494)	-0,59** (0,038)	-1,345** (0,533)	—
Presencia en folletos publicitarios	—	0,323* (0,097)	0,715* (0,103)	0,671* (0,082)
Lealtad histórica	—	—	1,3* (0,086)	—
Lealtad anterior	—	—	4,943* (0,158)	5,709* (0,136)
<i>Criterios de evaluación de bondad del ajuste</i>				
Nº parámetros	14	15	17	12
LL (β)	-6632,1	-6612,23	-2240,14	-2327,825
χ ² -LL(C)	189,228	228,977	8973,165	8797,789
ρ ²	0,014	0,017	0,667	0,655
ρ ² ajust.	0,012	0,0148	0,664	0,653
AIC	13292,2	13254,46	4514,28	4679,65
BIC	13369,24	13337,01	4607,83	4745,69
LL(C): -6.726,72				
Nº obs. = 1.814				

*p<0,01 **p<0,05 ***p<0,1

^a entre paréntesis, errores estándar

Centrándonos, por tanto, en el modelo [17], y si realizamos un análisis de los coeficientes que presentan los atributos de las alternativas podemos observar que, por lo general, el signo es el esperado. Así, el precio de compra presenta coeficientes negativos, lo que indica la existencia de una relación inversa entre el precio de una alternativa y su probabilidad de elección; además, y por lo general, los valores son elevados, lo que indica una fuerte e intensa relación inversa. En relación a la variable que recoge la presencia de las alternativas en los folletos publicitarios del establecimiento, en aquellas ocasiones en las que las marcas se encuentran presentes en los folletos publicitarios su probabilidad de ser seleccionadas se ve incrementada (0,671), lo que indica la capacidad de atracción que tiene esta variable promocional. Por último, podemos observar como la elección del consumidor se encuentra fuertemente influida por las elecciones que ha realizado en las ocasiones anteriores de compra, tal y como indica el valor positivo, significativo y tan elevado (5,709) del coeficiente de la variable de lealtad.

4.2. MODELO PARAMÉTRICO CON HETEROGENEIDAD EN LAS PREFERENCIAS Y HETEROGENEIDAD ESTRUCTURAL

Una vez que hemos analizado el modelo básico que asume la homogeneidad del consumidor en la preferencia hacia cada una de las alternativas y en su respuesta hacia las variables explicativas, introducimos la heterogeneidad en el modelo [17]. Inicialmente, hemos estimado un modelo paramétrico en el que se asume que tanto la preferencia hacia cada una de las alternativas como el coeficiente de respuesta a las variables explicativas siguen una distribución uniforme. Sus parámetros estimados los resumimos en la tabla 6.

Como podemos observar en la tabla 6, las estimaciones de las preferencias que proporciona el modelo paramétrico son positivas en todos los casos, con excepción de las marcas propiedad del distribuidor y la marca *Giralda*. No obstante, en el caso de la *MdD* el valor de la desviación típica es elevado y significativo (2,782), lo que indica que esa preferencia difiere mucho entre la población, existiendo individuos para los que la preferencia hacia esta marca es positiva ($-1,39 \pm 2,782$) y otros para los que la preferencia hacia la *MdD* es muy negativa. Para el caso de *Giralda* y la *marca de primer precio*, la desviación típica no es significativa, lo que indica que esa preferencia negativa es compartida por el conjunto de individuos. En relación a las variables explicativas, el efecto sobre la probabilidad de elección es el esperado, es decir, el precio de compra tiene un efecto negativo y significativo (-8,555) sobre la probabilidad de elección, mientras que las otras dos variables, presencia de las marcas en los folletos publicitarios y lealtad anterior, tiene un efecto positivo y significativo (0,829 y 6,999 respectivamente). Por el contrario, las desviaciones típicas de estos parámetros aleatorios presentan comportamientos diferentes, ya que la del precio de compra no es significativa mientras que la de las otras dos variables sí lo es. Esto indica que el efecto del precio de compra sobre la probabilidad de elección es ciertamente similar para el conjunto de individuos, mientras que el efecto de la lealtad y los folletos publicitarios pre-

senta una mayor dispersión (fundamentalmente el caso del parámetro relativo a la presencia de las marcas en los folletos publicitarios, cuya desviación típica es elevada en relación al valor del parámetro medio), si bien en ningún caso se modifica el efecto final (siempre positivo) de las variables.

TABLA 6.—Parámetros estimados del modelo con heterogeneidad de las preferencias y estructural

Modelo con parámetros aleatorios (distribución uniforme)	
<i>Preferencias hacia cada alternativa</i>	
Carbonell - media - desv. típica	4,281* (0,541) 1,241** (0,534)
MdD - media - desv. típica	-1,39*** (0,79) 2,782* (0,357)
Coosur - media - desv. típica	1,801** (0,592) 2,303* (0,476)
La Masía - media - desv. típica	1,469** (0,543) 0,222 (0,558)
La Española - media - desv. típica	1,337** (0,587) 2,671* (0,472)
Elosúa - media - desv. típica	0,617 (0,614) 2,097* (0,518)
Giralda - media - desv. típica	-1,36*** (0,843) 0,541 (0,346)
Ybarra - media - desv. típica	1,013*** (0,575) 1,623* (0,452)
Marca de primer precio - media - desv. típica	-1,837** (0,743) 0,424 (0,421)
Mueloliva	—
<i>Parámetros de las variables explicativas</i>	
Precio de compra - media - desv. típica	-8,555* (1,127) 0,234 (0,386)
Presencia en folletos publicitarios - media - desv. típica	0,829* (0,102) 0,549** (0,22)
Lealtad anterior - media - desv. típica	6,999* (0,322) 0,474*** (0,28)
- LL (β)	2297,458

*p<0,01 **p<0,05 ***p<0,1

4.3. MODELO SEMIPARAMÉTRICO DE CLASES LATENTES

Uno de los aspectos más importantes en el desarrollo de los modelos semiparamétricos es la determinación del número óptimo de segmentos, S . Si S fuese suficientemente grande, podríamos explicar toda la variabilidad en las preferencias y en la sensibilidad a las herramientas de marketing. Sin embargo, perderíamos parsimonia y sencillez, por lo que creemos que es preferible, desde el punto de vista estadístico y de la aplicabilidad en marketing, un número de segmentos bajo. En este sentido, el investigador debe utilizar conjuntamente criterios estadísticos y su propio juicio de valor en relación al contexto en el que se esté realizando la segmentación y a la utilidad de la misma para la empresa. En relación al criterio estadístico, el más utilizado en la literatura es el criterio *BIC* (por ejemplo, Hensher y Greene, 2002) o el *AIC* (por ejemplo, Kamakura y Russell, 1989), cuyas expresiones son:

$$\begin{aligned} AIC &= -2 LL(\beta) + 2B \\ BIC &= -2 \left[LL(\beta) - \frac{\beta}{2} (\ln(N)) \right] = -2 LL(\beta) + B \ln(N) \end{aligned} \quad [18]$$

Igualmente, existen otros criterios estadísticos como el *AIC consistente* (CAIC) o el *AIC modificado* (MAIC):

$$\begin{aligned} CAIC &= -2 LL(\beta) + B \ln(N+1) \\ MAIC &= -2 LL(\beta) + 3B \end{aligned} \quad [19]$$

El procedimiento para determinar el número de segmentos es el siguiente (Kamakura y Russell, 1989): primero se estima el modelo para el mercado sin segmentar ($S = 1$); se vuelve a estimar con dos segmentos ($S = 2$), y así sucesivamente. De este modo, como número óptimo de segmentos se toma aquel en que el criterio estadístico sea menor. En caso de que el criterio estadístico utilizado siga disminuyendo continuamente, se toma como número óptimo de segmentos aquel a partir del cual el valor del estadístico comience a incrementarse. La tabla 7 muestra el valor alcanzado por estos cuatro criterios para el caso de 1, 2, 3, y 4 segmentos.

Como podemos observar la tabla 7, en función del criterio estadístico que utilicemos la solución óptima difiere. Así, para el criterio AIC y MAIC, la solución óptima está en la existencia de 3 segmentos, ya que es el punto en el que el estadístico toma el menor valor (4220,666 y 4256,666, respectivamente), mientras que para el BIC y CAIC, la solución óptima parte de la existencia de 2 segmentos de mercado, ya que a partir de ese punto, el estadístico comienza a subir.

TABLA 7.—Comparación entre diferente número de segmentos (criterios: mínimo valor de AIC, BIC, CAIC y MAIC)

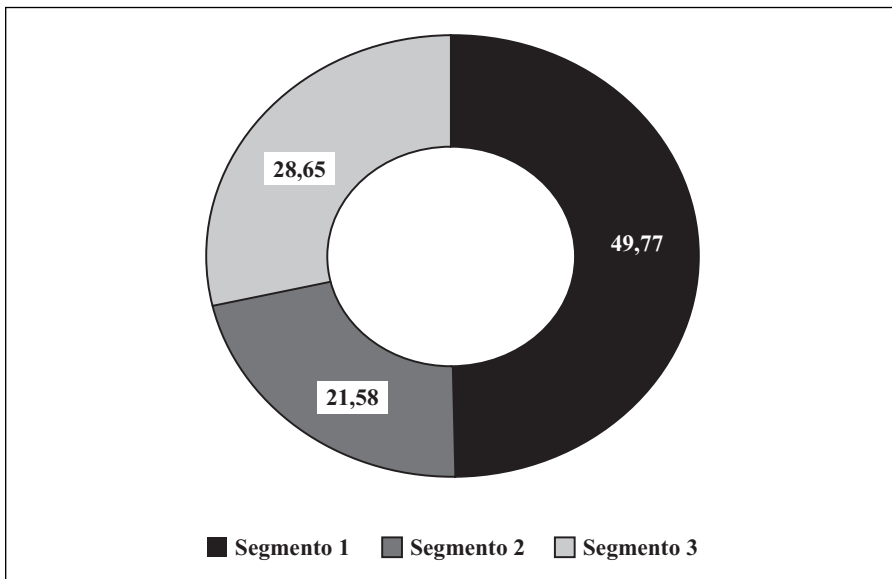
NÚMERO DE SEGMENTOS	AIC	BIC	CAIC	MAIC
1	4679,65	4745,689	4745,696	4691,65
2	4278,512	4410,591	4410,604	4302,512
3	4220,666	4418,784	4418,804	4256,666
4	4227,516	4491,674	5041,696	4825,512
5	—	—	—	—

Ante esta diferencia en el número de segmentos a determinar según cada criterio, hemos optado por realizar un *test de razón de verosimilitud* entre ambos modelos, con objeto de determinar si la consideración de un segmento más (tres), respecto a la solución que proponen BIC y CAIC (dos), mejora, de modo significativo, el poder explicativo del modelo. El valor del estadístico es:

$$RV = -2 [-2074,333 - (-2115,256)] = 81,846 \text{ (13 g.l.)}$$

Este valor es significativo ($p < 0,01$), por lo que aceptamos la hipótesis de existencia de diferencias significativas entre ambos modelos, considerando que el mercado está formado por 3 grupos de consumidores, tal y como indican los criterios AIC y MAIC, con el siguiente tamaño (figura 1).

FIGURA 1.—Tamaño de los segmentos de mercado



El segmento 1 es el de mayor tamaño, significando la mitad del mercado. La otra mitad restante, se la reparten entre el segmento 2 y el segmento 3, si bien este último es algo más grande. En la tabla 8 podemos observar los segmentos resultantes del proceso de modelización semiparamétrica y sus parámetros respectivos.

TABLA 8.—Parámetros estimados para cada uno de los tres segmentos existentes en el mercado

	Segmento 1 (49,77%) ^a	Segmento 2 (21,58%)	Segmento 3 (28,65%)
Preferencia hacia cada alternativa			
Carbonell	38,058* (8,08)	4,242* (0,749)	25,3* (7,123)
MdD	4,302 (4,123)	6,648 (6,236)	2,444* (0,983)
Coosur	2,338 (2,413)	6,276*** (4,451)	4,034* (1,345)
La Masía	20,126*** (13,598)	12,793*** (8,416)	-41,393 (43,412)
La Española	11,998*** (8,332)	6,392*** (4,348)	3,244* (1,065)
Elosúa	6,404 (7,649)	22,649 (22,101)	-83,722 (85,631)
Giralda	-14,998*** (9,991)	-15,709 (14,658)	-64,314 (63,456)
Ybarra	13,476 (13,112)	13,078*** (8,492)	0,632* (0,223)
Marca de primer precio	38,584 (42,134)	-7,745 (8,128)	-6,303* (2,021)
Mueloliva	—	—	—
Parámetros específicos de las variables explicativas			
Precio de compra	-16,494* (2,277)	-56,9 (58,108)	-19,42* (3,556)
Presencia en folletos publicitarios	0,735* (0,113)	-1,332* (0,32)	0,409*** (0,254)
Lealtad anterior	67,48 (56,743)	12,709* (0,883)	85,681 (75,172)
-LL(β)		2074,333	

* $p < 0,01$ ** $p < 0,05$ *** $p < 0,1$ ^a(entre paréntesis el tamaño del segmento)

A partir de los diferentes valores obtenidos en la estimación de cada segmento, hemos desarrollado distintos *test-t para muestras independientes*, con objeto de comparar si las diferencias entre dichos valores es estadísticamente significativa. De este modo, podemos caracterizar a cada uno de los segmentos de mercado en función de su preferencia hacia las marcas y al efecto de las variables explicativas:

- Segmento 1: Se trata del segmento de mayor tamaño, y manifiesta una elevada preferencia por la marca líder *Carbonell* (38,058). Además manifiesta preferencias significativas hacia otras marcas nacionales como *La Masía* o *La Española*. El efecto, tanto del precio como de la variable folleto, es el esperado (negativo para el precio de compra y positivo para la variable folleto); por el contrario, se trata de un segmento que no presenta comportamientos de lealtad significativos.
- Segmento 2: Es el segmento de tamaño más reducido. Su preferencia hacia las marcas nacionales se encuentra mucho más repartida (*Carbonell*, *Coosur*, *La Masía*, *La Española*, *Ybarra*), siendo todos sus valores muy similares (si bien hacia *Ybarra* y *La Masía* esa preferencia es mayor). Al igual que el segmento anterior, no presenta ninguna preferencia significativa hacia la *MdD*. Se trata del único segmento para el que el precio de compra no es una variable de elección significativa, no utilizando los folletos publicitarios (ya que no les influyen las posibles promociones que puedan aparecer en dichos folletos); todo esto lo convierte en un segmento con un comportamiento significativo de lealtad en la elección de marca que realizan sus consumidores.
- Segmento 3: Presenta un tamaño algo mayor que el segmento anterior. Se trata de individuos que, si bien presentan una preferencia muy elevada hacia la marca líder (menor que la de los individuos del primer segmento y más elevada que los del segmento anterior), también manifiestan, de modo significativo, preferencia hacia la *MdD*, siendo el único segmento que muestra esta característica. Igual que el segmento anterior, manifiestan preferencia por marcas nacionales de precio medio, si bien estas preferencias son mucho menores que para el segmento 2. En relación a las variables explicativas presenta un comportamiento similar a los individuos del segmento 1, si bien el efecto del precio es más intenso, siendo menor el de la presencia de las marcas en los folletos publicitarios.

4.4. CONTRASTE ESTADÍSTICO ENTRE LOS MODELOS ANALIZADOS

A continuación vamos a contrastar cuál de estos modelos es estadísticamente superior. Con este objetivo contrastamos la existencia de diferencias significativas entre ellos para, posteriormente, comparar sus valores de AIC, BIC, así como de la función de verosimilitud y el coeficiente ρ^2 . En primer lugar realizamos la comparación entre el modelo de parámetros aleatorios y el modelo básico. El valor del estadístico es:

$$RV = -2 [-2297,458 - (-2327,826)] = 60,736 \text{ (21 g.l.)}$$

Este valor es significativo ($p < 0,01$), por lo que el modelo que incorpora la heterogeneidad en las preferencias y la heterogeneidad estructural a través de parámetros aleatorios mejora, significativamente, el modelo básico con preferencias y coeficientes de las variables explicativas constantes. Una vez deter-

minada la superioridad del modelo con parámetros aleatorios, vamos a compararlo ahora con el modelo semiparamétrico.

$$RV = -2 [-2074,33 - (-2297,458)] = 446,256 \text{ (14 g.l.)}$$

Este valor es significativo ($p < 0,01$), por lo que podemos establecer que existen diferencias significativas entre ambos modelos. Para determinar cuál de ellos es superior, calculamos los estadísticos de bondad del ajuste, así como los coeficientes ρ^2 (tabla 9).

Como podemos observar en la tabla 9, la aproximación semiparamétrica mejora de modo significativo la explicación de la heterogeneidad del consumidor frente al enfoque paramétrico. Esto viene motivado por el menor valor del modelo con segmentos, tanto en la función de verosimilitud como en los estadísticos AIC y BIC, aún estimando 12 parámetros más que el modelo de parámetros aleatorios.

TABLA 9.—*Criterios de evaluación de la bondad del ajuste (modelo sin heterogeneidad, modelo paramétrico y modelo con segmentos)*

	Modelo básico	Modelo paramétrico	Modelo con segmentos
Nº parámetros	12	24	36
LL (β)	-2327,825	-2297,458	-2074,333
$\chi^2 - LL(C)$	8797,789	13052,1	13498,35
ρ^2	0,655	0,7396	0,7649
ρ^2 ajust.	0,653	0,7392	0,7643
AIC	4679,65	4642,916	4220,666
BIC	4745,69	4774,995	4418,784
LL(C): -6.726,72			
Nº obs. = 1.814			

Estos resultados ponen de manifiesto que el profundizar al máximo en el comportamiento heterogéneo del consumidor, llegando al extremo de poder estimar un parámetro para cada individuo, es menos eficiente que considerar grupos de consumidores con comportamientos diferenciados. En este sentido, los resultados avalan la tendencia actual de los directivos de marketing, más interesados, tal y como indican Corstjens y Gautschi, (1983) o Wedel y Kamakura (2000), en el comportamiento de compra y elección de los segmentos de individuos que existen en el mercado, así como en las diferencias que existen entre ellos.

5. Conclusiones e implicaciones para la gestión, limitaciones y futuras líneas de investigación

En el presente trabajo se ha pretendido analizar las consecuencias que tiene la inclusión de la heterogeneidad en el comportamiento de elección del consumidor a través de modelos logit, y evaluar la mejora explicativa que la consideración de este comportamiento supone en dichos modelos. Se pretende, asimismo, contribuir a que las empresas puedan comprender mejor su mercado para optimizar el diseño de su estrategia de diferenciación.

Para el análisis de la heterogeneidad en el comportamiento de elección del consumidor se han utilizado dos enfoques principales: el enfoque paramétrico y el enfoque semiparamétrico. Ambos modelos suponen una mejora sustancial en la capacidad de predicción del comportamiento de elección del consumidor respecto a un modelo en el que no se incluya esa componente heterogénea. En este sentido, se han aportado evidencias empíricas que señalan la necesidad de considerar la heterogeneidad en los modelos de elección del consumidor, ya que la suposición de homogeneidad puede conducir a conclusiones sesgadas por parte de la empresa, al considerar un comportamiento único para el conjunto de mercado. Sin embargo, el descender hasta el nivel individual de heterogeneidad, tal y como defiende el enfoque paramétrico, si bien es idealista, no es práctico desde la perspectiva de marketing, tanto por motivos económicos como por motivos operativos. Es por ello que la empresa prefiere considerar la existencia de grupos de consumidores con un comportamiento muy similar pero a la vez muy diferente del comportamiento del resto de individuos de otros segmentos. A este nivel, tal y como defienden, entre otros, Kamakura y Russell (1989), la heterogeneidad es adecuadamente identificada, pudiendo resultar más beneficioso que la consideración de cada individuo como un segmento, en particular, en los mercados de productos de consumo frecuente.

Los resultados empíricos obtenidos avalan esta hipótesis. Así, en la comparación que hemos realizado de los dos enfoques de tratamiento de la heterogeneidad, el enfoque de clases latentes parece adecuarse mejor al comportamiento de elección observado, permitiendo a la empresa conocer cuántos grupos de consumidores forman su mercado y, sobre todo, conocer cómo reaccionan a las variables de marketing utilizadas. En este sentido, hemos encontrado tres grandes segmentos de mercado en función de su reacción a las variables de marketing analizadas.

Un primer grupo de consumidores, el menos numeroso, para los que el precio del producto no es una variable excesivamente relevante a la hora de considerar qué marca elegir, siendo poco sensibles a cambios en los precios de las marcas. Derivado de este comportamiento, este grupo de individuos es muy leal a las marcas que adquieren con habitualidad, siendo difícil que cambien de marca a través de la realización de promociones o descuentos del precio de venta. Esa menor sensibilidad al precio y su mayor lealtad determinan que estos consumidores, habitualmente, no lean los folletos publicitarios del establecimiento antes de realizar sus compras, ya que no valoran tanto el pre-

cio y, en consecuencia, tampoco valoran las posibles ofertas o promociones que el folleto publicitario les puede proporcionar.

Por todo esto, se trata de un segmento que prefiere marcas nacionales que suelen aparecer con menor intensidad en los folletos publicitarios que, por ejemplo, la marca de distribuidor, y que, por término medio, realizan menos promociones del precio. Desde el punto de vista de la gestión detallista, conocer la existencia de este tipo de consumidor es fundamental para optimizar el presupuesto promocional, ya que puede ayudar a la empresa a seleccionar a qué consumidores es más conveniente hacer llegar sus folletos promocionales e, incluso, a cuáles dirigir sus promociones del precio, lo que a su vez se puede traducir en un incremento de dicho gasto para aquellos consumidores que, efectivamente sí son sensibles a estas herramientas promocionales.

Por el contrario, el segundo de los segmentos de mercado identificados (la mitad del mercado total), sí presenta esa sensibilidad a las herramientas promocionales de la empresa. Así, se trata de los individuos más sensibles al precio de venta de las marcas, lo que hace que, constantemente, busquen el mejor precio y, por tanto, sí acudan a los folletos publicitarios del establecimiento como fuente de información de posibles promociones del precio. Por ello, aquellas marcas que aparecen en estos folletos tienen una mayor probabilidad de ser elegidas. Esta búsqueda del precio como variable importante en las decisiones de elección del individuo, hace que la lealtad hacia la marca sea una dimensión poco significativa en la función de utilidad de este grupo de consumidores. Este es, desde el punto de vista de marketing, el segmento de consumidores al que el establecimiento debe dirigir sus mayores esfuerzos promocionales, ya que es el grupo más receptivo a este tipo de acciones.

El tercer y último grupo (aproximadamente, la tercera parte del mercado), es un segmento cuyo comportamiento se sitúa en un nivel intermedio al de los dos grupos anteriores. En este sentido, se trata de consumidores también sensibles al precio de venta, aunque en menor medida que el segmento anterior. Igualmente gustan de acudir a los folletos publicitarios para informarse de las promociones existentes, lo que hace que manifiesten una preferencia positiva hacia la marca del distribuidor, si bien también presentan preferencias importantes hacia marcas nacionales de precio medio (por ejemplo, *Coosur*, *La Española* o *Ybarra*).

Los distribuidores deben ser conscientes de la existencia de grupos de consumidores que, como en nuestro caso, presentan comportamientos totalmente diferentes, ya que ello va a incrementar la precisión con la que van a poder llegar a ellos, incrementando la satisfacción del consumidor y acentuando el nivel de lealtad hacia su marca.

Ciertas limitaciones deben ser consideradas. En primer lugar, los resultados han sido obtenidos para una categoría y un formato de producto muy concretos: aceite de oliva 0,4º, formato 1 litro. Por este motivo es indispensable que en otras investigaciones se extienda este análisis a otras categorías de producto, con objeto de comprobar posibles diferencias y poder obtener mayores generalizaciones empíricas. Igualmente hay que indicar las limitaciones derivadas de la metodología empleada. En este sentido, la denominada *Propiedad de Independencia de Alternativas Irrelevantes* y la *endogeneidad* pueden limi-

tar la validez de las estimaciones resultantes. Finalmente, queremos dejar constancia de nuestro deseo de seguir profundizando en el análisis de la heterogeneidad en los modelos de elección utilizando, para ello, otras metodologías complementarias a los modelos logit, así como otros enfoques de inclusión de la heterogeneidad, que permitan contrastar diferentes aproximaciones al tratamiento del comportamiento heterogéneo del consumidor en los mercados de bienes de consumo masivo.

Referencias bibliográficas

- AHN, K. y GHOSH, A. (1989). «Hierarchical models of store choice», *International Journal of Retailing*, 4 (5), págs. 39-52.
- AILAWADI, K., GEDENK, K. y NESLIN, S. (1999): «Heterogeneity and purchase event feedback in choice models: an empirical analysis with implications for model building», *International Journal of Research in Marketing*, 16, págs. 177-198.
- ALLENBY, G. y ROSSI, P. (1999). «Marketing models of consumer heterogeneity», *Journal of Econometrics*, 89, págs. 57-78
- ARORA, N., ALLENBY, G y GINTER, J. (1998): «A hierarchical bayes model of primary and secondary demand», *Marketing Science*, 17, (1), págs. 29-44.
- BEANE, T. y ENNIS, D. (1987): «Market segmentation: a review», *European Journal of Marketing*, 21, (5), págs. 20-42.
- CAVERO, S. y CEBOLLADA, J. (1999): «Análisis de la competencia entre tipos de marca. Una aplicación empírica» *Información Comercial Española*, 779, (Julio-Agosto), págs. 55-70.
- (2000): «Buscando segmentos latentes en el mercado. Una aplicación empírica con datos de elección de marca», *Revista Española de Investigación de Marketing*, 6, (Marzo), págs. 95-112.
- CHAMBERLAIN, G. (1980): «Analysis of covariance with qualitative data», *Review of Economic Studies*, 47, págs. 225-238
- CHINTAGUNTA, P. (1992): «Heterogeneity in nested logit models: an estimation approach and empirical results», *International Journal of Research in Marketing*, 9, págs. 161-175.
- (1993): «Investigating purchase incidence, brand choice and purchase quantity decisions of households», *Marketing Science*, 12, (2), págs. 184-204.
- (2001): «Endogeneity and heterogeneity in a probit demanda model: estimation using aggregate data», *Marketing Science*, 20, (4), págs. 442-456.
- CHINTAGUNTA, P. y HONORE, B. (1996): «Investigating the effects of marketing variables and unobserved heterogeneity in a multinomial probit model», *International Journal of Research in Marketing*, 13, págs. 1-15.
- CHINTAGUNTA, P., JAIN, D. y VILCASSIM, N. (1991): «Investigating heterogeneity in brand preferences in logit models for panel data», *Journal of Marketing Research*, XXVIII, (Noviembre), págs. 417-428.
- CORSTJENS, J. y CORSTJENS, M. (1995): *Store wars: the battle for mindspace and shelfspace*, New York, John Wiley & Sons.
- CORSTJENS, M. y GAUTSCHI, D. (1983): «Formal choice models in marketing», *Marketing Science*, 2, (1), págs. 19-56.
- CURRIM, I. (1981): «Using segmentation approaches for better prediction and understanding from consumer choice models», *Journal of Marketing Research*, XVIII, (Agosto), págs. 301-309.

- CURRIM, I., MEYER, R. y LE, N. (1988): «Disaggregate tree-structured modeling of consumer choice data», *Journal of Marketing Research*, XXV, (Agosto), págs. 253-265.
- DESARBO, W., ANSARI, A., CHINTAGUNTA, P., HIMMELBERG, C., JEDIDI, K., JOHNSON, R., KAMAKURA, W., LENK, P., SRINIVASAN, K. y WEDEL, M. (1997): «Representing heterogeneity in consumer response models», *Marketing Letters*, 8, (3), págs. 335-348.
- GENSCH, D. (1985): «Empirically testing a disaggregate choice model for segments», *Journal of Marketing Research*, XXII, (Noviembre), págs. 462-67.
- GÖNÜL, F. y SRINIVASAN, K. (1993): «Modeling multiple sources of heterogeneity in multinomial logit models: methodological and managerial issues», *Marketing Science*, 12, (3), págs. 213-229.
- GUADAGNI, P. y LITTLE, J. (1983): «A logit model of brand choice calibrated on scanner data», *Marketing Science*, 2, (3), págs. 203-238.
- HENSHER, D. y GREENE, W. (2002): «A latent class model for discrete choice analysis: contrasts with mixed logit», *Working Paper*, School of Business, The University of Sydney.
- JAIN, D. y VILCASSIM, N. (1991) : «Investigating household purchase timing decisions: a conditional hazard function approach», *Marketing Science*, 10, (Primavera), págs. 1-15.
- KALYANAM, K. y PUTLER, D. (1997): «Incorporating demographic variables in brand choice models: an indivisible alternatives framework», *Marketing Science*, 16, (2), págs. 166-181.
- KAMAKURA, W., KIM, B. y LEE, J. (1996): «Modeling preference and structural heterogeneity in consumer choice», *Marketing Science*, 15, (2), págs. 152-172.
- KAMAKURA, W. y RUSSELL, G. (1989): «A probabilistic choice model for market segmentation and elasticity structure», *Journal of Marketing Research*, XXVI, (Noviembre), págs. 379-390.
- KANNAN, P. y WRIGHT, G. (1991): «Modeling and testing structured markets: a nested logit approach», *Marketing Science*, 10, (1), págs. 58-82.
- KRISHNAMURTHI, L., MAZUMDAR, T. y RAJ, S. (1992): «Asymmetric response to price in consumer brand choice and purchase quantity decisions», *Journal of Consumer Research*, 19, (Diciembre), págs. 387-400.
- KRISHNAMURTHI, L. y RAJ, S. (1991): «An empirical analysis of the relationship between brand loyalty and consumer price elasticity», *Marketing Science*, 10, (2), págs. 172-183.
- MALHOTRA, N. (1984): «The use of linear logit models in marketing research», *Journal of Marketing Research*, XXI, (Febrero), págs. 20-31.
- MC FADDEN, D. (1974): «Conditional logit analysis of qualitative choice behavior», en Zarembka, P. (ed.), *Frontiers in econometrics*, Academic Press, New York, págs. 105-142.
- (1978): «Modelling the choice of residential location», en Karlquist, A., Lundqvist, L., Snickars, F. y Weibull, J.W. (eds.): *Spatial interaction theory and residential location*, Studies in Regional Science and Urban Economics, 3, North Holland, Amsterdam, págs. 75-96.
- NOVALES, A. (1993). *Econometría*. McGraw-Hill, Madrid.
- PAPATLA, P. y KRISHNAMURTHI, L. (1996): «Measuring the dynamic effects of promotions on brand choice», *Journal of Marketing Research*, XXXIII, (Febrero), págs. 20-35.
- SIVAKUMAR, K. y RAJ, S. (1997): «Quality tier competition: how price change influences brand choice and category choice», *Journal of Marketing*, 61, (Julio), págs. 71-84.

Consideración de la heterogeneidad en el comportamiento de elección...

- WEDEL, M. y KAMAKURA, W. (2000): *Market segmentation. Conceptual and methodological foundations*, 2ªed., Dordrecht: Kluwer.
- WIND, Y. (1978): «Issues and advances in segmentation research», *Journal of Marketing Research*, XV, (Agosto), págs. 317-37.
- YAMAGUCHI, K. (1986):«Alternative approaches to unobserved heterogeneity in the analysis of repeatable events», en N. Tuma (ed.) *Sociological Methodology 1986*, American Sociological Assoc., Washington, págs. 213-249.