

OBTENCIÓN DE MAPAS DE POSICIONAMIENTO BASADOS EN MODELOS DE RESPUESTA DEL MERCADO CON HETEROGENEIDAD LATENTE A PARTIR DE DATOS DE ESCANER AGREGADOS

AUTORES:

González-Benito, O.^{1*}, Martínez-Ruiz, M. P.², Mollá-Descals, A.³

AFILIACIONES:

^{1*} *Universidad de Salamanca (España)*

² *Universidad de Valencia (España)*

³ *Universidad de Valencia (España)*

^{1*} **Autor de contacto:**

Dr. Óscar González-Benito Departamento de Administración y Economía de la Empresa. Campus Miguel de Unamuno. 37007 – Salamanca (España). Tel.: +34 (9)23 294 400 (3508). Fax: +34 (9)23 295 715. E-mail: oscargb@usal.es

² **Co-autor:**

Dr. María Pilar Martínez-Ruiz. Departamento de Comercialización e Investigación de Mercados. Facultad de Economía. Avenida de los Naranjos, s/n. 46022 Valencia (España). Tel.: +34 (9)63 828 334. Fax: +34(9)63 828 333. E-mail: M.Pilar.Martinez-Ruiz@uv.es

³ **Co-autor:**

Dr. Alejandro Mollá-Descals. Departamento de Comercialización e Investigación de Mercados. Facultad de Economía. Avenida de los Naranjos, s/n. 46022 Valencia (España). Tel.: +34 (9)63 828 334. Fax: +34(9)63 828 333. E-mail: Alejandro.Molla@uv.es

Agradecimientos

Este trabajo ha sido financiado por la Viceconsejería de Universidades, Investigación e Innovación de la Consejería de Educación y Ciencia de la Junta de Comunidades de Castilla-La Mancha.

OBTENCIÓN DE MAPAS DE POSICIONAMIENTO BASADOS EN MODELOS DE RESPUESTA DEL MERCADO CON HETEROGENEIDAD LATENTE A PARTIR DE DATOS DE ESCANER AGREGADOS

Abstract

Given the increased focus of category management in retailing, this work attempts to contribute to this research line by proposing a methodology based on scanner data aggregated at the store level in order to analyze the competitive positioning of brands within the product categories that compound the retailer's assortment. In particular, this research focuses on the internal analysis of market structure based on market share models with latent heterogeneity estimated with retail scanner data aggregated at the store level. The proposed approach assumes a multidimensional decomposition latent in the preference structure implicit to these types of models. This allows us to simultaneously identify latent dimensions characterizing competing brands and latent segments with different preferences towards these dimensions. The proposed methodology is exemplified and evaluated through an empirical application.

Keywords

Internal analysis of the market structure; positioning maps; latent heterogeneity; market response models; aggregated scanner data

Resumen

Dada la enorme relevancia que durante los últimos años han adquirido los temas de gestión por categorías en el comercio minorista, el presente trabajo pretende contribuir a la investigación realizada en este ámbito proponiendo una metodología basada en datos de escáner agregados a nivel de establecimiento orientada a abordar el posicionamiento competitivo de las marcas que componen las categorías de producto que conforman el surtido del detallista. En concreto, este estudio aborda el análisis interno de la estructura de mercado basado en modelos de cuota de mercado con heterogeneidad latente estimados con datos de escáner agregados a nivel de establecimiento. La propuesta metodológica se centra en asumir una descomposición multidimensional latente en la estructura de preferencias implicada en este tipo de modelos, lo que permite identificar simultáneamente atributos latentes caracterizadores de las marcas competidoras y segmentos latentes con distinto perfil de

preferencias hacia estos atributos. La propuesta metodológica se ejemplifica y evalúa con una aplicación empírica. Los resultados obtenidos permiten constatar el potencial de los modelos propuestos para explicar el posicionamiento competitivo de las marcas sin necesidad de recurrir a datos desagregados a nivel de hogar.

Palabras clave

Análisis interno de la estructura de mercado; mapas de posicionamiento; heterogeneidad latente; modelos de respuesta de mercado; datos de escáner agregados

1. Introducción

La proliferación de nuevas tecnologías y sistemas de información en los canales de distribución ha conllevado la disponibilidad de multitud de datos sobre el comportamiento de compra del consumidor, en particular datos de escáner generados en los puntos de venta. Esta circunstancia ha suscitado un gran interés por desarrollar herramientas analíticas que permitan comprender y pronosticar la respuesta de mercado a los estímulos comerciales a partir de esta información.

El interés de este estudio se centra en dos líneas concretas dentro de esta extensa área de investigación. Por un lado, el análisis interno de la estructura de mercado, que Elrod (1991) define como la inferencia de los atributos de las marcas y las preferencias de los consumidores para estos atributos a partir de información comportamental o actitudinal de los consumidores. Las discusiones de Elrod (1991), DeSarbo et al. (1994), Elrod y Keane (1995) o Elrod et al. (2002) dejan patente la multitud de tradiciones metodológicas orientadas a esta finalidad.

Por otro lado, la incorporación de la heterogeneidad de los consumidores en la modelización explicativa de la respuesta de mercado. Este objetivo resulta fundamental tanto desde un punto de vista explicativo – la segmentación es clave en la configuración de las estrategias comerciales –, como desde un punto de vista predictivo – obviar la diversidad de preferencias y beneficios buscados por los consumidores puede sesgar los modelos de respuesta de mercado. Una metodología ampliamente difundida es el enfoque de heterogeneidad latente mediante efectos aleatorios, consistente en asumir que las preferencias y la sensibilidad de los consumidores a los estímulos de marketing difieren a través de la población conforme a una

distribución de probabilidad. Centrados exclusivamente en los modelos probabilísticos de elección discreta basados en la teoría de utilidad aleatoria, podemos distinguir dos planteamientos: el paramétrico, cuando se asume una distribución continua específica (Gonul y Srinivasan, 1993), y el semi-paramétrico, cuando se asume una distribución discreta (Kamakura y Russell, 1989). Chintagunta et al. (1991) aportan evidencia a favor de este último enfoque, que equivale a asumir la existencia de segmentos latentes caracterizados por un perfil de respuesta.

Algunos trabajos han tratado de combinar ambos objetivos asumiendo una descomposición multidimensional en la estructura de preferencias que parametriza los modelos probabilísticos de elección discreta con efectos aleatorios, tanto con el enfoque paramétrico (Elrod, 1988; Elrod y Keane, 1995) como con el enfoque semiparamétrico (Chintagunta, 1994; Andrews y Manrai, 1999). Esto permite identificar el posicionamiento de las marcas competidoras en un conjunto de atributos latentes y la distribución de las preferencias de los consumidores con respecto a estos atributos.

Sin embargo, todas estas propuestas se centran en modelos de elección estimados con datos desagregados a nivel de hogar, esto es, información sobre el historial de compras de los consumidores. Consecuentemente, su aplicabilidad con datos de escáner generados en los puntos de venta es limitada. La disponibilidad de datos desagregados en un establecimiento se limita a aquellos clientes identificables mediante tarjetas de cliente o tarjetas de pago. Estos criterios de identificación pueden conllevar sesgos, por ejemplo, los derivados de la utilización de distintas tarjetas o la utilización de tarjetas solo en algunas compras. Además, este grupo de consumidores puede no ser representativo de los clientes del establecimiento. Esta circunstancia justifica la necesidad de adaptar este tipo de modelos para que puedan ser estimados con datos agregados a nivel de establecimiento (e.g., Bucklin y Gupta, 1999).

El objetivo de este trabajo es adaptar la metodología de Chintagunta (1994) a los modelos de cuota de mercado estimados con datos agregados a nivel de establecimiento. Esto permite identificar simultáneamente atributos latentes caracterizadores de las marcas competidoras y segmentos latentes con distinto perfil de preferencias hacia estos atributos sin necesidad de recurrir a datos desagregados sobre el historial de compras de los hogares. La modelización simultánea del análisis interno de la estructura de mercado y la segmentación latente, frente a una modelización secuencial, reduce el número de parámetros implicados simplificando la estimación y permite identificar el número adecuado de atributos y segmentos mediante criterios de ajuste estadístico.

Con anterioridad, los trabajos de Zenor y Srivastava (1993) o Bodapati y Gupta (2004) se han centrado en la estimación de modelos de respuesta de mercado con heterogeneidad latente con datos agregados. Este trabajo extiende su planteamiento posibilitando además el análisis interno de la estructura de mercado.

Los contenidos subsiguientes se estructuran en tres secciones adicionales. En primer lugar se describe la propuesta metodológica. En segundo lugar, se describe una aplicación empírica con datos reales que permite ejemplificar y evaluar la propuesta. Finalmente, en una última sección, se resumen las principales conclusiones del estudio.

2. Propuesta metodológica

La propuesta metodológica se desglosa en cinco etapas. Primeramente se propone una interpretación agregada de los modelos de elección mediante modelos de cuota de mercado. En segundo lugar, se incorporan supuestos multidimensionalidad en la estructura de preferencias. Estas dimensiones conforman los ejes de un mapa de posicionamiento competitivo de las marcas competidoras. En tercer lugar, se incorporan supuestos de heterogeneidad latente. Esta condición da sentido a la multidimensionalidad asumida y permite la identificación del modelo. En cuarto lugar, se describe el procedimiento de estimación a partir de datos agregados. Finalmente, en quinto lugar, se proponen indicadores para valorar el ajuste del modelo y comparar las distintas especificaciones.

2.1. Modelo de respuesta de mercado: perspectiva agregada vs. desagregada.

Los modelos probabilísticos de elección discreta basados en la teoría de utilidad aleatoria representan la probabilidad P_{ir} de que un consumidor i en la ocasión r seleccione la marca j como una función parametrizada f_j de un conjunto de variables X_{ir} caracterizadoras de la oferta comercial para el individuo y la ocasión en cuestión. Habitualmente se distinguen dos tipos de parámetros: las constantes de preferencia α_j propias de cada marca, y los parámetros β relacionados con la sensibilidad de los consumidores a los cambios en las variables explicativas. Sujetos a esta notación, el planteamiento general de este tipo de modelos puede resumirse en la siguiente expresión:

$$P_{ir}(j | \alpha_j, \beta) = f_j(X_{ir} | \alpha_j, \beta) \quad [1]$$

La perspectiva agregada implica considerar la respuesta del mercado en conjunto, esto es, el comportamiento de elección agregado de un conjunto de consumidores durante un periodo de tiempo t . Asumamos que:

- (1) El comportamiento de elección de todos los consumidores del mercado se rige por los mismos parámetros (α_j, β) de respuesta;
- (2) Las variables explicativas X_{it} no difieren para los distintos consumidores del mercado en las distintas ocasiones del periodo t . Por ejemplo, este es el caso de la variable precio cuando permanece constante durante un periodo de tiempo independientemente del consumidor y la ocasión considerados. La notación puede entonces simplificarse haciéndose referencia a las variables X_t .
- (3) La elección de marca es independiente de la cantidad de unidades compradas. Para analizar las implicaciones de la cantidad comprada en la relación entre probabilidad de elección y cuota de mercado se puede consultar Cooper y Nakanishi (1988) o Cooper (1993).

Entonces, la cuota de mercado π_t en el periodo t de la marca j coincide exactamente con la probabilidad de elección de la marca j para cualquier consumidor y cualquier ocasión dentro del periodo considerado. Es decir, puede expresarse de manera análoga a la ecuación [1]:

$$\pi_t(j | \alpha_j, \beta) = f_j(X_t | \alpha_j, \beta) \quad [2]$$

Mantenemos esta notación genérica, si bien las formas funcionales más habituales para este tipo de modelos, tanto desde una perspectiva individual como desde una perspectiva agregada, son los modelos tipo logit y los modelos tipo probit.

2.2. Multidimensionalidad en la estructura de preferencias

En línea con el planteamiento de Elrod (1988), Chintagunta (1994) o Elrod y Keane (1995), las constantes de preferencia pueden interpretarse como el resultado de una valoración compensatoria de distintos atributos o dimensiones caracterizadoras de las marcas. Concretamente, se asume:

- (1) La existencia de K dimensiones latentes caracterizadoras de las marcas competidoras. Estos constituyen los ejes de un mapa de posicionamiento.
- (2) El posicionamiento de las distintas marcas en estas dimensiones vendrá dado por una matriz $A = \{a_{jk}\}$ con tantas columnas como dimensiones latentes y tantas filas como marcas competidoras. Cada fila $A_j = [a_{j1}, a_{j2}, \dots, a_{jK}]$ representa la posición de la marca j en el mapa de posicionamiento, es decir, sus coordenadas espaciales.
- (3) La valoración de cada una de estas dimensiones por parte de los consumidores vendrá dada por un vector de ponderaciones $W = [w_1, w_2, \dots, w_K]'$, donde w_k denota la importancia relativa dada a la dimensión k .

Consecuentemente, las constantes de preferencia vienen dadas por la siguiente expresión:

$$\alpha_j = A_j \cdot W \quad [3]$$

y el modelo expresado en [2] queda como sigue:

$$\pi_t(j | A_j \cdot W, \beta) = f_j(X_t | A_j \cdot W, \beta) \quad [4]$$

2.3. Supuestos de heterogeneidad latente

Una limitación importante del planteamiento anterior es que asume unos patrones de comportamiento homogéneos. En otras palabras, se obvia la diversidad de preferencias y beneficios buscados por los consumidores. En particular, esta circunstancia impide la identificación de la descomposición factorial asumida en el modelo. El enfoque de heterogeneidad latente, y más concretamente su planteamiento semiparamétrico (Chintagunta et al., 1991), permite salvar esta situación.

Centrándose exclusivamente en los parámetros de preferencia, se puede asumir la existencia de un conjunto S de segmentos latentes, cada uno caracterizado por una estructura de preferencias y una sensibilidad a las variables explicativas. Así, cada segmento s agrupará consumidores con constantes de preferencia α_{js} y parámetros de respuesta β_s similares. Con un razonamiento idéntico al utilizado para proponer la ecuación [2] a partir de [1], la cuota de mercado de la marca j en el periodo t y en el segmento s puede expresarse de la siguiente manera:

$$\pi_t(j | \alpha_{js}, \beta_s) = f_j(X_t | \alpha_{js}, \beta_s) \quad [5]$$

La cuota de mercado global vendrá dada por una suma ponderada de las cuotas en cada uno de los segmentos, es decir:

$$\pi_t(j) = \sum_{s \in S} \pi_t(j | \alpha_{js}, \beta_s) \cdot \lambda_s \quad [6]$$

donde λ_s denota un parámetro correspondiente al tamaño relativo del segmento s , cumpliéndose que:

$$0 < \lambda_s < 1 \text{ y } \sum_{s \in S} \lambda_s = 1 \quad [7]$$

La combinación de este planteamiento con la descomposición multidimensional de la estructura de preferencias implica asumir que cada segmento s valora de manera distinta las dimensiones que conforman el posicionamiento competitivo, es decir, está caracterizado por un vector de ponderaciones W_s distinto. Así, la ecuación [3] puede aplicarse a cada segmento:

$$\alpha_{js} = A_j \cdot W_s \quad [8]$$

y, por tanto, el modelo propuesto en la expresión [6] queda como sigue:

$$\pi_t(j) = \sum_{s \in S} \pi_t(j | A_j \cdot W_s, \beta_s) \cdot \lambda_s \quad [9]$$

La identificación de este modelo con descomposición factorial está supeditada a que el número de parámetros implicados no sea superior al modelo sin descomposición [6]. A este respecto, se cumple lo siguiente:

- (1) El número de parámetros relacionados con la estructura de preferencias α_{js} en el modelo sin descomposición factorial es $S(J-1)$, donde J es el número de marcas que compiten en el mercado. Puesto que las formas habituales de estos modelos requieren fijar una de las constantes de preferencia para identificar el modelo, deben estimarse $(J-1)$ constantes por cada segmento.
- (2) El número de parámetros relacionados con el posicionamiento A de las marcas competidoras en el modelo con descomposición factorial es $K \cdot J$ menos el número de restricciones necesarias para fijar el mapa con respecto a traslaciones y rotaciones. Fijar el mapa con respecto a traslaciones, es decir, respecto al origen, requiere fijar las K coordenadas de una de las marcas. Fijar el mapa con respecto a rotaciones requiere fijar $(K-1)$ coordenadas de una marca, $(K-2)$ coordenadas de otra marca, así sucesivamente, hasta fijar 1 coordenada de una última marca. En total, el número de restricciones necesarias viene dado por $1/2 \cdot K \cdot (K+1)$.
- (3) El número de parámetros relacionados con los vectores de ponderación W_s en el modelo con descomposición factorial viene dado por $K \cdot S$ menos el número de restricciones necesarias para fijar el mapa con respecto a la escala de los ejes. Esta indeterminación queda resuelta fijando el vector de ponderaciones de uno de los segmentos, es decir, se requieren K restricciones.

Por tanto, la identificación del modelo con descomposición factorial [9] requiere que los números de segmentos, marcas y dimensiones cumplan la siguiente condición:

$$K \cdot J - \frac{K(K+1)}{2} + K \cdot S - K \leq S(J-1) \quad [10]$$

2.4. Estimación

La estimación de los parámetros implicados en el modelo agregado difiere del planteamiento desagregado. La información de partida no consiste en el historial de compras de una muestra de consumidores, sino en las ventas N_{jt} de cada marca j a lo largo de varios periodos t . Estas ventas agregan las elecciones realizadas por distintos individuos pertenecientes a los distintos segmentos latentes.

Zenor y Srivastava (1993) sugieren la estimación mediante el algoritmo EM de máxima verosimilitud. Este procedimiento es aplicado de manera general a modelos con variables

latentes con datos ausentes (Dempster et al., 1977). En este caso, las variables latentes son las ventas en cada segmento latente.

El algoritmo EM se formaliza en dos etapas. Primeramente, se calculan las ventas esperadas N_{jts} por cada marca j en el periodo t y en el segmento s a partir de la información observada N_{jt} . Vendrán dadas por la ecuación:

$$N_{jts} = N_{jt} \left(\frac{\pi_t(j | A_j \cdot W_s, \beta_s) \cdot \lambda_s}{\sum_{s \in S} \pi_t(j | A_j \cdot W_s, \beta_s) \cdot \lambda_s} \right) \quad [11]$$

El calculo inicial requiere fijar unos valores iniciales para los parámetros A , W_s , β_s y λ_s implicados.

En una segunda etapa, se define una función de verosimilitud que representa la probabilidad de las ventas estimadas en la primera etapa. Esta función de verosimilitud viene dada por la siguiente expresión:

$$L = \prod_t \left(k_t \prod_j \prod_s (\pi_t(j | A_j \cdot W_s, \beta_s) \cdot \lambda_s)^{N_{jts}} \right) \quad [12]$$

donde $k_t = \left(\sum_j \sum_s N_{jts} \right)! / \prod_j \prod_s N_{jts}!$

El valor k_t es tan solo un número combinatorio que tiene en cuenta la posible ordenación de la secuencia de ventas en cada periodo t .

La maximización de la función [12] permite identificar los valores de los parámetros A , W_s , β_s y λ_s más adecuados para las ventas estimadas en cada segmento. Estos valores pueden utilizarse para recalcular ventas mediante la expresión [11] y volver a repetir el proceso. Zenor y Srivastava (1990), apoyándose en los resultados de Dempster et al. (1977) demuestran que el valor de la función de verosimilitud crecerá en cada interacción, y convergerá para los valores de los parámetros que mejor representen la mejor segmentación latente del mercado.

Este procedimiento es igualmente aplicable con el modelo sin descomposición factorial de la estructura de preferencias planteado en [6].

2.5. Indicadores de ajuste

El procedimiento de estimación planteado requiere fijar previamente el número de segmentos latentes existentes en el mercado o el número de dimensiones subyacentes a la estructura de preferencias. Consecuentemente, la selección de la solución óptima requiere estimar los modelos para distintas cantidades de segmentos latentes y distinto número de dimensiones, y

compararlos. A este respecto, se sugieren dos tipos de indicadores para formalizar la comparación.

Por un lado, se pueden considerar distintas versiones del coeficiente ρ^2 o pseudo- R^2 sugeridas por Zenor y Srivastava (1993). Para su cálculo, se parte de una función de verosimilitud agregada para todo el mercado. Representa la probabilidad de las cuotas de mercado observadas a partir de unas cuotas estimadas. La transformación logarítmica viene dada por la expresión:

$$\ln(L) = k + \sum_t \sum_j N_{jt} \ln(\pi_t(j))$$

$$\text{donde } k = \sum_t \ln\left(\left(\sum_j N_{jt}\right)!\right) - \sum_t \sum_j \ln(N_{jt}!) \quad [13]$$

El coeficiente pseudo- R^2 viene dado por la expresión:

$$\rho^2 = \frac{\ln(L_e) - \ln(L_o)}{\ln(L_p) - \ln(L_o)} \quad [14]$$

donde $\ln(L_e)$, $\ln(L_o)$ y $\ln(L_p)$ denotan respectivamente el valor de la función [13] para:

- (1) las cuotas estimadas con el modelo agregado con segmentación latente conforme a la expresión [9].
- (2) las cuotas estimadas con un planteamiento trivial, es decir, las cuotas observadas agregando todos periodos:

$$\pi_t(j) = \sum_t N_{jt} / \sum_j \sum_t N_{jt} \quad [15]$$

- (3) las cuotas estimadas con el modelo perfecto, es decir, la cuotas observadas en cada periodo:

$$\pi_t(j) = N_{jt} / \sum_j N_{jt} \quad [16]$$

Por tanto este coeficiente indica en qué medida el modelo con segmentación latente explica la diferencia entre el modelo perfecto y el modelo trivial.

Puesto que este coeficiente crecerá a medida que se incrementa el número de segmentos, conviene ajustarlo para valorar en qué medida su crecimiento compensa el incremento en el número de parámetros. El coeficiente pseudo- R^2 ajustado está dado por la siguiente expresión:

$$\rho^2 \text{ ajustado} = 1 - \left(1 - \rho^2\right) \left(\frac{m-1}{m-p-1}\right) \quad [17]$$

donde p denota el número de parámetros y m denota el número de cuotas observadas, es decir, el producto entre el número de periodos y el número de marcas.

Por otro lado, la comparación se puede basar en Criterios de Información. Elrod y Keane (1993) sugiere cuatro posibles alternativas:

- El Criterio de Información de Akaike (*AIC*):

$$AIC = -2\ln(L_e) + 2p \quad [18]$$

- El Criterio de Información de Hannan y Quinn (*HQ*):

$$HQ = -2\ln(L_e) + 2p \ln(\ln(m)) \quad [19]$$

- El Criterio de Información de Bayes (*BIC*):

$$BIC = -2\ln(L_e) + 2p \ln(m) \quad [20]$$

- El Criterio de Información Consistente de Akaike (*CAIC*):

$$CAIC = -2\ln(L_e) + 2p(\ln(m) + 1) \quad [21]$$

Todos estos indicadores son igualmente aplicables con el modelo sin descomposición factorial de la estructura de preferencias planteado en [6].

3. Aplicación empírica

Con el fin de evaluar y ejemplificar la propuesta metodológica, se describe una aplicación a datos de escáner agregados a nivel de establecimiento minorista. En primer lugar se describe el contexto de estudio y los datos utilizados. A continuación se describen y comentan el análisis realizado y los resultados obtenidos

3.1. Contexto de análisis y datos

Los datos utilizados en el análisis corresponden a los precios y ventas agregadas diarias de un supermercado durante el año 1999. Esto supone un total de 304 observaciones. Las primeras 204 observaciones se utilizaron como muestra de estimación. Las últimas 100 observaciones se utilizaron como muestra de validación. La categoría de producto considerada es el café molido mezcla en envase de 250 gramos. Esta categoría estaba formada por seis marcas, dos marcas de precio reducido: 154 y Bahía, y cuatro marcas de precio elevado: Bonka, Marcilla, Saimaza y Soley. La Tabla 1 contiene descriptivos sobre los precios y las ventas de cada una de las marcas.

3.2. Análisis y resultados

La aplicación de la metodología propuesta se basa en una configuración logit multinomial (MNL) del modelo de cuota de mercado representado en la ecuación [2]. Es decir:

$$\pi_t(j | \alpha_j, \beta) = \frac{\exp(\alpha_j + \beta \cdot P_t(j))}{\sum_j \exp(\alpha_j + \beta \cdot P_t(j))} \quad [22]$$

donde $P_t(j)$ denota la variable de precio de la marca j en el periodo t . Esta variable se ha medido mediante las variaciones de precio con respecto al precio medio del periodo (es decir, precio – precio medio en el año estudiado). Esto implica que los parámetros de atractivo intrínseco también capturan el atractivo derivado del nivel precios medio mantenido por cada marca.

El análisis realizado y los resultados obtenidos se pueden secuenciar en dos fases. En una fase preliminar se aborda la estimación del modelo con heterogeneidad latente pero sin la descomposición multidimensional de la estructura de preferencias. Esto permite identificar el número óptimo de segmentos latentes subyacentes a la estructura de mercado. En una segunda fase donde se incorpora la descomposición factorial de la estructura de preferencias para identificar simultáneamente dimensiones latentes de posicionamiento competitivo y segmentos latentes conforme a las preferencias hacia dichas dimensiones.

3.2.1. Heterogeneidad de mercado: número óptimo de segmentos.

Con el fin de determinar el número óptimo de segmentos latentes, se realizó un primer análisis con el modelo sin descomposición factorial de la estructura de preferencias. El ajuste del modelo desde uno hasta cinco segmentos latentes se resume en la Tabla 2. Los indicadores utilizados son los descritos en el epígrafe 2.5.

En la muestra de estimación el pseudo- R^2 aumenta conforme se incrementa el número de segmentos latentes, puesto que cada modelo es una extensión del anterior. Sin embargo, el pseudo- R^2 ajustado incrementa hasta el modelo realizado con cuatro segmentos, comenzando a decrecer desde este punto. Esto sugiere que el modelo con cuatro segmentos es el más adecuado. La misma conclusión se alcanza observando los Criterios de Información. Todos ellos alcanzan valores mínimos para el modelo con cuatro segmentos.

El ajuste observado sobre la muestra de validación es más modesto. Además, tanto del coeficiente pseudo- R^2 como el Criterio de Información de Akaike, sugieren que el modelo más adecuado es el configurado para dos segmentos. El primero de los indicadores alcanza su valor máximo en este modelo, mientras que el segundo de los indicadores alcanza su valor mínimo. Todo esto nos lleva a considerar que el número óptimo de segmentos es, en todo caso, menor que cinco.

La Tabla 3 resume los resultados de estimación para los modelos realizados con un número de segmentos que varía entre uno y cuatro segmentos latentes respectivamente. Algunos comentarios al respecto son:

- En todos los modelos con segmentación latente, se identifican dos segmentos de mayor tamaño, teniendo el resto un tamaño más reducido.
- Todos los segmentos difieren considerablemente en cuanto a la estructura de preferencias. La magnitud de los parámetros en uno de los segmentos (segmentos 3 y 4 en los modelos para 3 y 4 segmentos latentes, respectivamente) hace pensar en que el conjunto evocado de estos consumidores se limita únicamente a algunas marcas. Precisamente las marcas más desfavorecidas son las que operan en un nivel de precios más bajo (154 y Bahía).
- También los parámetros de precio difieren considerablemente a través de los segmentos. Cabe destacar que el parámetro asociado a uno de los segmentos (segmento 3 en modelo con 4 segmentos latentes) es positivo, aunque muy pequeño. Esto implica una escasa sensibilidad a las variaciones en los precios de este segmento.

3.2.2. Descomposición factorial: mapa de posicionamiento.

Conforme a la ecuación [10], y puesto que el número óptimo de segmentos es menor que cinco, la descomposición factorial de la estructura de preferencias es solamente posible hasta dos dimensiones. Considerar tres dimensiones requeriría al menos cinco segmentos latentes. Puesto que el posicionamiento competitivo basado en una sola dimensión ya queda reflejado en la estructura de preferencias reflejada por el modelo para un solo segmento latente, el interés se centró en la representación bidimensional del posicionamiento competitivo. También conforme a la ecuación [10], la consideración de dos dimensiones requiere al menos tres segmentos latentes. Por ello, se estimó el modelo con descomposición factorial bidimensional para tres y cuatro segmentos latentes.

El ajuste de ambos modelos se resume en la Tabla 4. Puesto que los modelos con descomposición factorial son restricciones de los modelos sin descomposición factorial, el ajuste sobre la muestra de estimación es siempre inferior al obtenido en la Tabla 2. Sin embargo, el ajuste observado sobre la muestra de validación deja patente una capacidad predictiva mucho mayor para los modelos con descomposición factorial, a pesar de tratarse de restricciones. Es decir, asumir una descomposición factorial en la estructura de preferencias y un perfil de valoración de los factores distinto para cada segmento latente parece capturar los patrones de respuesta de mercado de una manera más generalizable y extrapolable a otros

contextos. Esta circunstancia es especialmente notable para el modelo con cuatro segmentos latentes.

Los resultados de estimación de ambos modelos se resumen en la Tabla 5 y la representación de los mapas de posicionamiento obtenidos aparece en el Gráfico 1. Tanto el modelo realizado con tres segmentos como el modelo realizado con cuatro segmentos reflejan posiciones competitivas de las marcas similares. Esto aporta validez interna al análisis realizado, que conduce a una representación gráfica similar independientemente del número de segmentos latentes considerado.

Los mapas obtenidos permiten diferenciar dos grupos de marcas. Por un lado, Marcilla y Bonka, que aparecen con un posicionamiento muy similar. Son las marcas con mayor cuota de mercado. Por otro lado, 154, Saimaza, Bahía y Soley, cuyo posicionamiento parece diferir en una única dimensión (diagonal). Es decir, su posicionamiento en el mapa se ajusta en gran medida a una línea recta. La otra dimensión del mapa de posicionamiento (diagonal) explica la diferencia entre los dos grupos de marcas considerados.

4. Conclusiones

En el presente trabajo hemos llevado a cabo un análisis interno de la estructura de mercado basándonos en modelos de cuota de mercado con heterogeneidad latente, que han sido estimados con datos de escáner agregados de un supermercado. Para ello, la metodología propuesta asume una descomposición multidimensional latente de la estructura de preferencias capturada por un modelo explicativo de cuota de mercado. Adicionalmente, asume la existencia de distintos segmentos latentes con patrones de comportamiento distintos. Esto permite identificar simultáneamente atributos latentes caracterizadores de las marcas competidoras y segmentos latentes con distinto perfil de preferencias hacia dichos atributos. En esencia, la metodología propuesta constituye una combinación de las propuestas de Chintagunta (1994) y Zenor y Srivastava (1993).

Los resultados obtenidos han puesto de manifiesto el potencial de los modelos propuestos para explicar el posicionamiento competitivo de las marcas incluidas en la categoría de producto analizada sin necesidad de recurrir a datos desagregados a nivel de hogar. También se ha detectado que, a pesar de que estos modelos constituyen una versión restringida de los modelos con heterogeneidad latente, pueden presentar una capacidad predictiva superior.

La aplicación de esta propuesta puede resultar de gran utilidad en la planificación y la gestión comerciales. No sólo para orientar e implantar acciones estratégicas de tipo reactivo frente a los movimientos de los competidores, sino también a la hora de llevar acciones comerciales que permitan la anticipación a los movimientos de la competencia. En efecto, la propuesta

implica nuevas posibilidades de explotación de los datos de escáner agregados a nivel de establecimiento. A partir de ellos, es posible identificar distintos perfiles de respuesta en el mercado, valorar el posicionamiento competitivo de las marcas, y estimar el efecto de las variaciones y descuentos promocionales ofrecidos por los distintos competidores. La propuesta igualmente útil para los fabricantes, en la configuración de sus líneas de productos y sus estrategias promocionales, y para los distribuidores, en la gestión de las categorías de productos y la optimización de sus decisiones comerciales al respecto.

A pesar de la utilidad de la presente propuesta y de su notable capacidad predictiva, es importante destacar algunas limitaciones en la aplicación de la misma. Cabe mencionar el elevado coste existente en términos de tiempo para la estimación de los parámetros del modelo, lo cual constituye una dificultad inherente a los procesos de estimación iterativos con algoritmos EM. Los tiempos requeridos en la estimación de los modelos pueden incluso aumentar de forma exponencial al incluir en los mismos un mayor número de marcas y/o segmentos. Adicionalmente, como apuntan Bodapati y Gupta (2004), las estimaciones de modelos con segmentación latente a partir de datos agregados pueden estar sesgadas. La única información disponible sobre la heterogeneidad se basa en las discrepancias entre las cuotas de mercado observadas y las pronosticadas por un modelo sin heterogeneidad latente. Por tanto, el sesgo disminuye en la medida en que existen fuertes discrepancias, el tamaño muestral es mayor, la configuración de los modelos es más completa y flexible, y existen menos segmentos latentes.

Las potencialidades de este modelo se pueden ampliar incorporando información relativa a un mayor número de estímulos de marketing (e.g., publicidad promocional, colocación especial del producto en el lineal, etc.) así como de otras categorías de producto (e.g., categorías de producto percederas; categorías de producto de no alimentación). Otra extensión natural de este trabajo es la caracterización de los segmentos latentes identificados con el fin facilitar el desarrollo de estrategias promocionales enfocadas directamente a cada segmento. Los esfuerzos en esta línea se han centrado en una perspectiva desagregada (Gupta y Chintagunta, 1994; Kamakura et al., 1994).

Referencias

Andrews, R.L. y Manrai, A.K. (1999): "MDS Maps of Product Attributes and Market Response: An Application to Scanner Panel Data", *Marketing Science*, 18 (4), 584-604.

- Bodapati, A.V. y Gupta, S. (2004): "The Recoverability of Segmentation Structure from Store-Level Aggregate Data", *Journal of Marketing Research*, 61 (August), 351-364.
- Bucklin, R. E. y Gupta, S. (1999), "Comercial Use of UPC Scanner Data: Industry and Academic Perspectives", *Marketing Science*, 18 (3), 247-273.
- Chintagunta, P.K. (1994): "Heterogeneous Logit Model Implications for Brand Positioning", *Journal of Marketing Research*, 31 (May), 304-311.
- Chintagunta, P.K.; Jain, D.C. y Vilcasim, N.J. (1991): "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data", *Journal of Marketing Research*, 28 (4), 417-428.
- Cooper, L.G. (1993): "Market-Share Models", en Eliashberg, J. y Lilien, G.L. (Eds.), *Handbooks in Operations Research and Management Science, Vol. 5, Marketing*, North Holland; Amsterdam, 259-314.
- Cooper, L.G. y Nakanishi, M. (1988): *Market-Share Analysis: Evaluating Competitive Marketing Effectiveness*, Kluwer Academic Publishers, Boston.
- Dempster, A.P.; Laird, M.N. y Rubin, D.B. (1977): "Maximum Likelihood Estimation from Incomplete Data Via de EM Algorithm", *Journal of the Royal Statistical Society*, B39, 1-38.
- DeSarbo, W.S.; Manrai, A.K. y Manrai, L.A. (1994): "Latent Class Multidimensional Scaling: A Review of Recent Developments in the Marketing and Psychometric Literature", en Bagozzi, R.P. (ed.), *Advanced Methods of Marketing Research*, Blackwell Publishers, Cambridge, MA, 190-222.
- Elrod, T. (1988): "Choice Map: Inferring a Product-Market Map form Panel Data", *Marketing Science*, 7 (Winter), 21-40.
- Elrod, T. (1991): "Internal Analysis of Market Structure: Recent Developments and Future Prospects", *Marketing Letters*, 2 (August), 253-266.
- Elrod, T. y Keane, M.P. (1995): "A Factor-Analytic Probit Model for Representing the Market Structure in Panel Data", *Journal of Marketing Research*, 32 (February), 1-16.
- Elrod, T.; Russell, G.J.; Andrews, R.L.; Bacon, L.; Bayus, B.L.; Carroll, J.D.; Johnson, R.M.; Kamakura, W.A.; Lenk, P.; Mazanec, J.A.; Rao, V.R. y Shankar, V. (2002): "Inferring Market Structure from Customer Response to Competing and Complementary Products", *Marketing Letters*, 13 (3), 221-232.
- Gonul, F. y Srinivasan, K. (1993): "Modelling Múltiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues", *Marketing Science*, 12 (Summer), 213-229.

Gupta, S. y Chintagunta, P.K. (1994): "On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models", *Journal of Marketing Research*, 31 (February), 128-136.

Kamakura, W.A.; Kim, B.D. y Lee, J. (1996): "Modelling Preference and Structural Heterogeneity in Consumer Choice", *Marketing Science*, 15 (2), 152-172.

Kamakura, W.A. y Russell, G.J. (1989): "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure", *Journal of Marketing Research*, 26 (November), 379-390.

Zenor, M.J. y Srivastava, R.K. (1993): "Inferring Market Structure with Agrégate Data: A Latent Segment Logit Approach", *Journal of Marketing Research*, 30 (August), 369-379.

Tabla 1

Tabla 1. Marcas en la categoría de productos – estadísticos descriptivos.

Ventas (Muestra de estimación)						
	Media ventas	D.T. ventas	Mínimo	Máximo	Ventas totales	Cuota de mercado
154	4,74	6,199	0	44	967	0,095
Bonka	16,52	16,589	0	94	3370	0,330
Marcilla	15,43	17,106	0	91	3147	0,308
Saimaza	5,80	10,446	0	76	1184	0,116
Soley	5,41	8,198	0	57	1104	0,108
Bahia	2,21	3,568	0	26	451	0,044
Ventas (Muestra de validación)						
	Media ventas	D.T. ventas	Mínimo	Máximo	Ventas totales	Cuota de mercado
154	4,64	6,256	0	44	464	0,106
Bonka	17,76	16,631	0	82	1776	0,407
Marcilla	12,52	11,554	0	51	1252	0,287
Saimaza	3,82	6,752	0	37	382	0,088
Soley	2,90	4,218	0	20	290	0,066
Bahia	1,97	3,161	0	16	197	0,045
Ventas (Muestra total)						
	Media ventas	D.T. ventas	Máximo	Mínimo	Ventas totales	Cuota de mercado
154	4,71	6,208	44	0	1431	0,098
Bonka	16,93	16,586	94	0	5146	0,353
Marcilla	14,47	15,541	91	0	4399	0,302
Saimaza	5,15	9,427	76	0	1566	0,107
Soley	4,59	7,228	57	0	1394	0,096
Bahia	2,13	3,436	26	0	648	0,044
Precio (Muestra de estimación) ^a						
	Media	D.T.	Mínimo	Máximo		
154	185,37	5,000	159	189		
Bonka	210,39	13,709	189	235		
Marcilla	223,38	22,624	195	259		
Saimaza	227,92	20,564	189	249		
Soley	210,20	15,468	189	235		

Bahia	190,12	6,399	165	195
Precio (Muestra de validación) ^a				
	Media	D.T.	Mínimo	Máximo
154	182,40	7,839	159	185
Bonka	198,34	9,552	185	215
Marcilla	205,66	10,250	189	215
Saimaza	221,04	11,321	189	225
Soley	193,44	3,069	187	195
Bahia	186,44	8,725	157	189
Precio (Muestra total) ^a				
	Media	D.T.	Máximo	Mínimo
154	184,39	6,228	189	159
Bonka	206,43	13,707	235	185
Marcilla	217,55	21,137	259	189
Saimaza	225,66	18,321	249	189
Soley	204,68	15,019	235	187
Bahia ^a	188,91	7,437	195	157

^a Precio en pesetas; 1 euro = 166,386 pesetas.

Tabla 2

Tabla 2. Comparativa según número de segmentos latentes – modelo sin descomposición factorial

Muestra de estimación						
	Pseudo R ²	Pseudo R ² ajustado	AIC	HQ	BIC	CAIC
1 segmento	0,5413	0,5390	6490,32	6501,86	6520,98	6526,98
2 segmentos	0,6350	0,6311	5689,25	5714,25	5755,68	5768,68
3 segmentos	0,6493	0,6435	5579,39	5617,85	5681,59	5701,59
4 segmentos	0,6567	0,6489	5529,24	5581,16	5667,21	5694,21
5 segmentos	0,6571	0,6473	5539,69	5605,08	5713,53	5747,43
Muestra de validación ¹						
	Pseudo R ²		AIC			
1 segmento	0,3091		3351,52			
2 segmentos	0,3144		3347,86			
3 segmentos	0,2758		3489,26			
4 segmentos	0,2621		3548,43			
5 segmentos	0,2505		3600,52			

¹ Únicamente se han calculado indicadores independientes del tamaño muestral.

Tabla 3

Tabla 3. Estimación hasta 4 segmentos latentes – modelo sin descomposición factorial

	Modelo 1 segmento	Modelo 2 segmentos		Modelo 3 segmentos			Modelo 4 segmentos			
	Segmento 1	Segmento 1	Segmento 2	Segmento 1	Segmento 2	Segmento 3	Segmento 1	Segmento 2	Segmento 3	Segmento 4
<i>Parámetros de preferencia (α_{is})</i>										
154	0,755	0,861	-2,877	1,656	0,185	-19,032	3,189	-0,695	0,387	-30,215
Bonka	2,037	1,609	4,632	0,698	1,822	17,932	0,508	2,569	1,350	27,618
Marcilla	1,974	1,632	4,449	0,503	1,861	19,451	0,715	1,031	1,726	30,154
Saimaza	0,786	0,406	-0,927	-5,708	0,666	7,396	-4,132	-8,155	0,537	36,260
Soley	1,132	0,103	4,681	1,396	-0,044	19,513	3,166	-3,291	-0,013	30,619
Bahia ^a	0	0	0	0	0	0	0	0	0	0
<i>Parámetros variables explicativas (β_s)</i>										
Precio	-0,047	-0,004	-0,306	-0,103	-0,001	-1,325	-0,141	-0,078	0,005	-2,085
<i>Tamaño de los segmentos (λ_s)</i>										
	100%	51,46%	48,54%	19,10%	42,60%	38,30%	11,31%	14,07%	36,19%	38,43%

^a Marca tomada como referencia con parámetro nulo.

Tabla 4

Tabla 4. Comparativa según número de segmentos latentes – modelo con descomposición factorial (2 dimensiones)

Muestra de estimación						
	Pseudo R ²	Pseudo R ² ajustado	AIC	HQ	BIC	CAIC
3 segmentos	0,6353	0,6299	5696,78	5731,40	5788,76	5806,76
4 segmentos	0,6400	0,6334	5664,27	5706,57	5776,68	5798,68
Muestra de validación ¹						
	Pseudo R ²		AIC			
3 segmentos	0,3179		3346,34			
4 segmentos	0,3828		3140,66			

¹ Únicamente se han calculado indicadores independientes del tamaño muestral.

Tabla 5

Tabla 5. Estimación para 3 y 4 segmentos latentes – modelo con descomposición factorial (2 dimensiones)

	Modelo 3 segmentos			Modelo 4 segmentos			
<i>Parámetros de posicionamiento en dimensiones (A)</i>							
	Dimensión 1		Dimensión 2	Dimensión 1		Dimensión 2	
154	0 ^a		0 ^a	0 ^a		0 ^a	
Bonka	1,366		0 ^a	1,793		0 ^a	
Marcilla	1,332		0,027	1,708		0,021	
Saimaza	0,343		-0,408	0,501		-0,235	
Soley	1,308		-1,043	1,920		-0,750	
Bahia ^a	0,497		-0,762	0,871		-0,488	
	Segmento 1	Segmento 2	Segmento 3	Segmento 1	Segmento 2	Segmento 3	Segmento 4
<i>Parámetros de valoración en dimensiones por segmentos (W_s)</i>							
Dimensión 1	1 ^b	6,324	0,302	1 ^b	4,805	0,263	-0,442
Dimensión 2	1 ^b	-0,511	1,859	1 ^b	1,128	3,193	-1,487
<i>Parámetros variables explicativas (β_s)</i>							
Precio	-0,023	-0,351	0,004	-0,032	-0,291	0,004	-0,672
<i>Tamaño de los segmentos (λ_s)</i>							
	35,68%	33,14%	31,18%	21,08%	24,54%	33,63%	20,75%

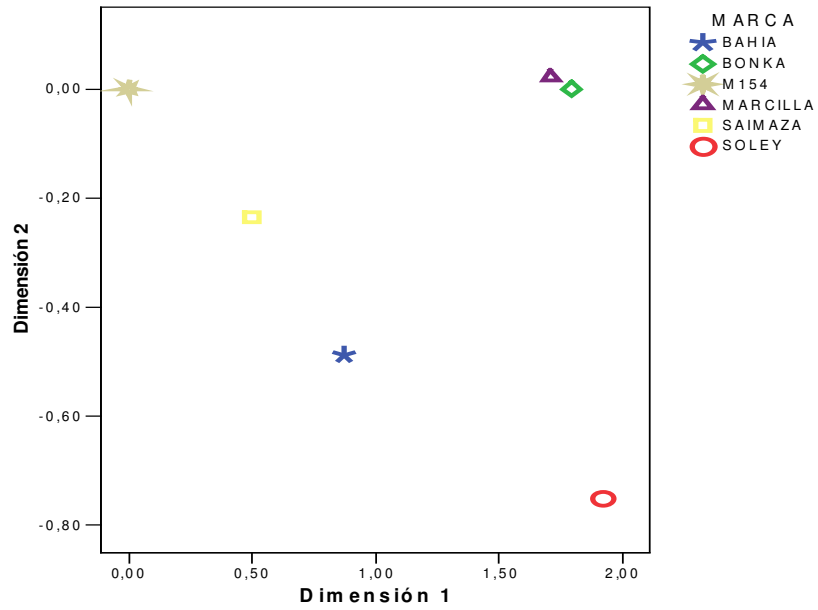
^a Coordenadas tomadas fijadas a 0 para fijar el mapa respecto a traslaciones y rotaciones.

^b Ponderaciones fijadas a 1 para fijar el mapa respecto a la escala de los ejes.

Grafico 1

**Gráfico 1. Mapas de posicionamiento –
modelo con descomposición factorial (2 dimensiones)**

Modelo 4 segmentos



Modelo 3 segmentos

