

Análisis del comportamiento variado de compra de los hogares españoles: una aproximación bayesiana(*)

por

CARMEN BERNÉ MANERO
NOEMÍ MARTÍNEZ CARABALLO

Departamento de Economía y Dirección de Empresas. Universidad de Zaragoza

y

PILAR GARGALLO VALERO
MANUEL SALVADOR FIGUERAS

Departamento de Estructura e Historia Económica y Economía Pública.
Universidad de Zaragoza

RESUMEN

Este trabajo analiza la composición del escenario de compra de los hogares españoles utilizando datos de un panel de consumo proporcionado por la empresa AC Nielsen. Para ello realiza una comparación bayesiana de un modelo Tobit y un modelo de Decisiones Separadas. Los resultados obtenidos confirman la influencia de indicadores económicos, demográficos y de comportamiento de la unidad de

(*) Este trabajo se ha realizado en el marco del Grupo de Excelencia "CREVALOR". Los autores desean agradecer la ayuda recibida a través de los Proyectos de Investigación del Ministerio de Educación y Ciencia SEJ2005-06603/ECON y SEJ2006-09961 y del Proyecto PM 090/2006 del Departamento de Ciencia, Tecnología y Universidad del Gobierno de Aragón. Así mismo, los autores agradecen los comentarios y sugerencias realizados por dos evaluadores anónimos los cuales contribuyeron a mejorar la exposición de los resultados obtenidos.

consumo sobre el grado de variación del escenario de compra. La comparación de los dos modelos resuelve que el mejor ajuste lo proporciona la aplicación del modelo Tobit, probándose así que no es necesario un tratamiento analítico diferente para aquellos hogares con un escenario de variación nula.

Palabras clave: Comportamiento del Consumidor, Inferencia Bayesiana, MCMC, Regresión Tobit, Decisiones Separadas, Comparación de Modelos.

Clasificación AMS: 62F15, 62J12, 91B42.

1. INTRODUCCION

Desde hace tres décadas, el estudio del comportamiento variado del consumidor ha atraído la atención de los investigadores sobre consumo en Economía y en Marketing. Esta variación en el comportamiento está dirigida a la obtención de estímulos en la compra mediante la alternancia entre objetos de elección, con la intención de obtener un cambio de ritmo en las actuaciones.

Centrando la cuestión en las decisiones de compra de productos de gran consumo de las familias, la situación actual es que, siempre y cuando existan alternativas disponibles, los hogares alternan y/o complementan la compra en su establecimiento habitual o principal con compras en otros establecimientos (Kahn y McAlister, 1997; Rhee y Bell, 2002). En los mercados muy competitivos, como son precisamente los mercados de bienes de gran consumo, este comportamiento variado ha devenido en una compra cruzada entre ofertas alternativas que es cada vez más acusada (Kahn y McAlister, 1997; McGoldrick y André, 1997).

De este modo, aceptamos que cada unidad de consumo resuelve sus necesidades de compra entre un conjunto o relación de establecimientos seleccionados por el responsable de la compra, quien los utiliza de manera habitual durante un periodo de tiempo en un entorno comercial determinado. Esta relación la denominamos "escenario de compra", la cual estará sujeta a una mayor o menor variación que será un reflejo directo de un mayor o menor comportamiento variado de consumo.

Hasta el momento, la literatura especializada ha analizado el comportamiento variado relacionándolo con conceptos tratados en Psicología, como el nivel óptimo de estimulación (e.g., Berlyne, 1960, Steenkamp y Baumgartner, 1992) y ha teorizado acerca de sus motivaciones intrínsecas y extrínsecas (McAlister y Pessemier,

1982; Kahn y Raju, 1991) trasladando estas ideas a aplicaciones en contextos económicos de consumo de productos (e.g., Kahn, Kalwani y Morrison, 1988; Steenkamp y Baumgartner, 1992; Van Trijp, Hoyer e Inmann, 1996; Kahn, 1998; Kim, Allenby y Rossi, 2002) en un establecimiento habitual de compra. Sin embargo, no ha habido aportaciones concluyentes sobre las influencias de variables demográficas y tampoco nos constan estudios que consideren la traslación de la situación de variación en el comportamiento a un escenario de compra habitual.

Este trabajo, que se desarrolla en un contexto de compra de bienes de consumo frecuente, pretende cubrir la falta señalada desde la observación de la composición del escenario de compra de los hogares españoles obtenida del panel de consumo de AC Nielsen del año 2003. Los objetivos teóricos parten de la formulación de hipótesis de trabajo que, tomando como base los antecedentes encontrados en la literatura, plantean relaciones entre indicadores sociales, económicos y demográficos de las unidades de consumo, y el grado de variación del escenario de compra del hogar. Los objetivos analíticos consideran la oportunidad que proporciona la aproximación bayesiana para la estimación y posterior comparación de modelos que estudian el tratamiento de la variable dependiente como continua (modelos Tobit) o como mixta (modelos de Decisiones Separadas). Esto es, resolver la compra del hogar en un único establecimiento o resolverla en más de uno, puede ser tratado como una decisión incluida en un continuo o como una decisión separada del comportamiento variado de compra, que requiere un análisis de antecedentes diferente. En última instancia, lo que se persigue es encontrar la mejor aproximación estadística de entre los modelos propuestos en este trabajo, para el tratamiento y análisis de los datos revelados en el panel de consumidores.

Los resultados obtenidos muestran que los patrones de compra de la unidad familiar (volumen de gasto realizado y frecuencia de compra), el número de personas que la conforman, su poder adquisitivo y el área geográfica en la que realiza sus compras ejercen una influencia significativa sobre el grado de variación de su escenario de compra.

El trabajo se organiza como sigue: en la sección 2 se formulan las hipótesis de trabajo y se describen los datos y las variables del estudio; en la sección 3 se describe la metodología estadística utilizada; en la sección 4 se muestran los resultados obtenidos y, finalmente, la sección 5 recoge las conclusiones más relevantes y plantea las futuras líneas de investigación.

2. PLANTEAMIENTO DEL PROBLEMA

2.1 Hipótesis de trabajo

Desde la literatura, distintas relaciones se han venido defendiendo entre motivaciones intrínsecas y extrínsecas del cambio y el comportamiento variado en el consumo (Berné, Múgica y Yagüe, 2001; Berné, Múgica y Rivera, 2005). Sin embargo, son escasos los trabajos dedicados al análisis de las características demográficas y su ulterior efecto en el comportamiento variado (Popkowski-Leszczyc y Timmermans, 1997) y, en algunos casos, los resultados no han sido concluyentes. Por ejemplo, Chintagunta y Gupta (1994) y Fox, Montgomery y Lodish (2004) no encontraron relación en un contexto de elección de marca en distintas categorías de productos.

En este trabajo nos centramos en la variación observada en un escenario de compra en un momento dado. Las variables que consideramos como potencialmente determinantes de dicha variación son (1) características propias de la unidad de consumo (nivel de ingresos, tamaño y ciclo de vida del hogar, área geográfica en la que reside y tamaño de la población del lugar de residencia), (2) características propias del responsable de la compra (edad y condición laboral) y (3) variables de comportamiento de compra (volumen y frecuencia de la compra).

La relación entre las características del hogar y la variación en el escenario de compra se ha venido analizando desde indicadores socio-económicos y demográficos. Así, McGoldrick y André (1997) encontraron que aquellos hogares con un mayor nivel de ingresos manifiestan una mayor lealtad al establecimiento principal de compra. Por otra parte, Seetharaman y Chintagunta (1998) constataron que el nivel de ingresos de la familia tiene un efecto negativo sobre el comportamiento de variedad. Considerando estas aportaciones y la composición del escenario de compra, cabe esperar que un mayor nivel de ingresos dé lugar a una menor variación de dicho escenario.

Seetharaman y Chintagunta (1998) encuentran que el tamaño de la unidad familiar tiene un efecto positivo en el comportamiento variado de los hogares. Una posible explicación es que los diferentes componentes del hogar revelen unos gustos muy diferentes, lo cual puede manifestarse en una mayor propensión a implementar variación en la elección. Por otro lado Mägi (2003) sugiere que los hogares de mayor tamaño pueden tener más restricciones de tiempo y una mayor tendencia a concentrar sus compras en un único establecimiento -establecimiento principal-. Pero, también cabe pensar que el establecimiento principal sea cambiado por otro que ocupe su lugar en un determinado momento.

Otro factor que potencialmente puede influir en el comportamiento variado de los hogares es el ciclo de vida del hogar. Este factor ha recibido un especial interés en la literatura como determinante del patrón de consumo de las familias (Wells y Gubar 1966; Gilly y Enis 1982; Deaton, Ruiz-Castillo y Thomas, 1989). Su posible influencia sobre el comportamiento de compra se debe a que el ciclo de vida del hogar refleja de manera conjunta el efecto de factores tales como la edad, el estatus socio-económico de la familia y su tamaño. Consiguientemente, influye en el reparto del gasto realizado en diferentes bienes y servicios, como es el caso de los productos alimenticios (Burnett, 1977).

En este trabajo, siguiendo a Wilkes (1995), definimos el ciclo de vida del hogar en función de la composición del mismo (tamaño de la unidad familiar y presencia de niños en la familia). Cabe pensar que la etapa del ciclo de vida en que se encuentre el hogar afecta al reparto del gasto realizado en la compra de productos de gran consumo y, por ende, al grado de variación en su escenario de compra. En este sentido, la literatura especializada observa que una de las mayores restricciones que soportan los responsables de la compra es la escasez de tiempo (Nickols y Fox, 1983; Soberon-Ferrer y Dardis, 1991). En particular, las mayores restricciones de tiempo que presentan los hogares con niños, esto es, que se encuentran en una etapa intermedia de su ciclo de vida, es muy probable que concentren el gasto en alimentación y productos para el hogar en un número menor de establecimientos. Se espera, por tanto, que su escenario de compra esté sujeto a una menor variación.

Por consiguiente, se plantean las siguientes hipótesis:

H1. *Cuanto mayor sea el nivel de ingresos del hogar, menor será la variación del escenario de compra.*

H2. *Cuanto mayor sea el tamaño del hogar, mayor será la variación del escenario de compra.*

H3. *La etapa del ciclo de vida del hogar influye en la variación del escenario de compra.*

Otros aspectos potencialmente relevantes en la variación del escenario de compra son el área geográfica y el tamaño de la población en las que residen las familias. En este sentido, se ha comprobado que la localización geográfica ejerce, junto al ciclo de vida de la familia y el nivel de educación, un efecto significativo sobre el reparto del gasto en servicios de los hogares (Soberon-Ferrer y Dardis, 1991). En concreto, desde un punto de vista comercial, las áreas geográficas son consideradas por las empresas para planificar sus rutas comerciales (decisiones en ventas y distribución de productos). Esta partición del mercado puede recoger parte de la heterogeneidad existente desde el lado de la oferta (número de estableci-

mientos, entorno competitivo, etc.) así como parte de la heterogeneidad de la muestra (diferencias sociológicas, demográficas o económicas). Por estas razones incluimos este factor en el modelo a contrastar como una variable de control de los efectos de esta heterogeneidad. Además, la oferta comercial de la zona geográfica difiere en función de cuál sea el tamaño de la población. Un tamaño mayor de población atrae a una mayor oferta minorista que busca aprovechar las ventajas del mayor gasto comercializable del área y de la atracción de la oferta conjunta para el consumidor. Consecuentemente, se plantean las siguientes hipótesis de trabajo:

H4. *El área geográfica en la que reside la familia influye en la variación del escenario de compra.*

H5. *Cuanto mayor sea el tamaño de la población en la que reside la familia, mayor será la variación del escenario de compra.*

Respecto de las relaciones entre rasgos personales del responsable de la compra y la variación del escenario, diversos estudios han demostrado que la edad del responsable de la compra se encuentra positivamente relacionada con el reparto del gasto en diversos establecimientos (East, Harris, Willson y Lomax, 1995; Mägi, 2003). Una posible explicación de este fenómeno puede ser que los consumidores de mayor edad –jubilados en su caso más extremo– disponen de más tiempo libre, por lo que son capaces de invertir más de este recurso, generalmente escaso, en la compra cotidiana y, al mismo tiempo, de utilizar diversos establecimientos simultáneamente para resolver su proceso de compra (East, Hammond, Harris y Lomax, 2000). Por tanto, esperamos que la edad del responsable de la compra ejerza una influencia positiva sobre la variación del escenario de compra.

Por otro lado, aquellos hogares en los que la persona encargada de realizar la compra de productos de uso frecuente trabaja a jornada completa, tenderán a realizar un mayor gasto en el denominado establecimiento principal y acudirán a realizar la compra a un número reducido de establecimientos a lo largo del tiempo (Fox, Montgomery y Lodish, 2004; McGoldrick y André, 1997). Consecuentemente, se espera una menor variación del escenario de compra del hogar cuando el responsable de la compra trabaje fuera del mismo.

Las hipótesis ahora se formulan como sigue:

H6. *Cuanto mayor sea la edad del responsable de la compra de productos para el hogar, mayor será la variación del escenario de compra.*

H7. *Si el responsable de la compra de productos para el hogar trabaja fuera del mismo, la variación del escenario de compra será menor.*

Por otro lado, Mägi (2003) señala que cuanto mayor sea el gasto total realizado por el hogar, mayores serán los beneficios derivados de comprar a mejores precios. Por consiguiente, se esperará que cuanto mayor sea el gasto del hogar mayor será la variación que presente su escenario de compra. Sin embargo, la misma autora admite que un mayor volumen de compra puede estar relacionado con un mayor tamaño de la unidad familiar. En consecuencia, aquellos hogares más grandes dispondrán de menos tiempo y tenderán a concentrar sus compras en un menor número de establecimientos, incluso en uno único, manifestando así un mayor grado de lealtad.

Para tratar de capturar el efecto ejercido por el volumen de gasto en el grado de variación que presenta en el escenario de compra de la familia, descontando el efecto de la variable tamaño de la unidad familiar, en este trabajo se considerará el gasto *per cápita*. Así, se plantea la siguiente hipótesis:

H8. *Cuanto mayor sea el volumen total de la compra per cápita, mayor será la variación del escenario de compra.*

También la frecuencia de compra ha sido considerada en varios trabajos como un factor determinante del comportamiento variado. En este sentido, cuanto más frecuentemente se realizan viajes de compra, más rápidamente se sentirá saciado y cobrará más importancia la necesidad de variedad provocada por aburrimiento o saciedad (Park, Moorman y Kohli, 1991). Así mismo, en un contexto de elección de variedad en la selección de productos, Berné, Mollá y Múgica (2004) encuentran que la frecuencia de compra tiene un efecto significativo y positivo sobre el comportamiento variado de consumo. En consecuencia, se plantea la siguiente hipótesis:

H9. *Cuanto mayor sea la frecuencia de compra, mayor será la variación del escenario de compra.*

2.2 Base de datos y medición de variables

La investigación empírica se abordará teniendo en cuenta la información contenida en una base de datos construida a partir de un panel de consumidores de AC Nielsen. En dicho panel de consumidores se recoge información sobre las compras de cinco categorías de productos de consumo frecuente y con alta tasa de penetración, realizadas por 2.017 hogares españoles (desde el 28 de abril de 2003 hasta el 25 de abril de 2004), concretamente: lugar y fecha de compra, volumen total de la cesta de compra, área Nielsen, tamaño poblacional, clase social a la que pertenece

el hogar, edad y situación laboral de la persona responsable de la compra en el hogar, número de miembros del hogar y presencia de niños en el mismo(1).

Por lo que respecta a la composición de dicho escenario de compra, únicamente el 6% de los hogares panelistas resuelve la compra familiar en un único establecimiento, un 13% lo hace en dos y el resto de los hogares, 81%, compra en tres o más establecimientos (concretamente, el 20,7% lo hace en 4 establecimientos; el 15,9% en 5; el 12,3% en 6; el 6,7% compra en 7 establecimientos y el 7,3% lo hace en 8 o más establecimientos). Estas observaciones ponen de manifiesto el marcado carácter de alternancia en las compras como una tónica actual de comportamiento para resolver una misma necesidad básica.

La medición del comportamiento variado en el escenario se abordará teniendo en cuenta el porcentaje de gasto relativo realizado en cada uno de los establecimientos que forman parte del escenario de compra.

Explicamos, a continuación, la medición de la que es nuestra variable principal de estudio, el comportamiento variado, y después la de las variables independientes.

Para medir el comportamiento variado en el escenario de compra se utiliza el Índice de Herfindahl-Hirschman (H), propuesto originalmente como una medida de concentración industrial (Theil, 1967) y utilizado después por Van Trijp (1995), para cuantificar la variación en el consumo. Dicho índice responde a la siguiente expresión:

$$H = - \sum_{j=1}^m p_j^2$$

donde p_j es el porcentaje de gasto realizado en el establecimiento j , y m es el número total de establecimientos que forman parte del escenario de compra. En nuestro caso, m vale 11, por lo que dicho índice se encontrará comprendido entre -1, cuando la familia no tenga un comportamiento variado y compre únicamente en un establecimiento ($p_1 = 1, p_2 = \dots = p_m = 0$), y $\frac{1}{m} = 0,09$, cuando el porcentaje de compra en

cada uno de los establecimientos sea el mismo ($p_1 = \dots = p_m = \frac{1}{m}$) y, por lo tanto, el grado de variación sea máximo.

(1) Aunque podría ser interesante estudiar la influencia de la variable sexo del responsable de la compra, la base de datos no permite su análisis puesto que no hay en la misma representación masculina. Todos los panelistas son mujeres.

Con respecto a las variables independientes del modelo:

1. Siguiendo a Martos (2004), la medición de la variable nivel de ingresos se ha abordado considerando como *proxy* la **clase_social**. Esta variable tiene tres posibles categorías: clase social alta, media o baja.
2. El tamaño de la unidad familiar (**tamaño_hogar**) toma el valor 1 si se trata de un hogar unipersonal, 2 si es bipersonal y así sucesivamente.
3. La covariable **ciclo_de_vida** cuenta con las categorías: Hogar unipersonal, Hogar pluripersonal con niños menores de 12 años y Hogar pluripersonal con niños mayores de 12 años, tomándose esta última categoría como valor de referencia.
4. El área geográfica se delimita mediante el **área_Nielsen** a la que pertenece el hogar. La empresa AC Nielsen (de la que proceden los datos con los que contamos) trabaja desde hace unos cuarenta años con una asignación de áreas que reparten la península española en zonas geográficas. El criterio original de distribución que la empresa siguió en el momento de su especificación atendió a un reparto en cuatro cuadrantes del mapa peninsular y a los circuitos comerciales utilizados por las empresas. Hoy en día, esta asignación es conocida como las "áreas Nielsen"⁽²⁾. Su reconocimiento es generalizado así como su utilización, habiéndose invertido la relación entre su aparición y su razón de origen; de modo que las áreas Nielsen se consideran más bien como un dato que condiciona sus decisiones comerciales. Por todo ello, decidimos la inclusión de esta variable en el conjunto de variables explicativas.
5. El tamaño poblacional (**tamaño_población**) toma los valores 0, 1, 2 y 3 dependiendo de si el hábitat en el que reside la familia es menor o igual a 10.000 habitantes, entre 10.000 y 50.000 habitantes, entre 50.000 y 200.000 y superior a 200.000, respectivamente.

(2) AC Nielsen clasifica los establecimientos teniendo en cuenta su situación geográfica, dividiendo la zona peninsular de España en 8 zonas conocidas como "áreas Nielsen", según el producto o tipo de venta. El área Nielsen 1 se encuentra integrada por: Barcelona (excluida su área metropolitana), Gerona, Lérida, Palma de Mallorca, Huesca y Zaragoza; el área Nielsen 2 está compuesta por Castellón, Valencia, Alicante, Murcia y Albacete; Badajoz, Córdoba, Huelva, Sevilla, Cádiz, Málaga, Granada y Almería forman parte del área Nielsen 3; el área Nielsen 4 se encuentra integrada por Zamora, Valladolid, Segovia, Soria, Guadalajara, Teruel, Cuenca, Toledo, Ávila, Cáceres, Ciudad Real y Salamanca; La Coruña, Lugo, Orense, Pontevedra y Oviedo forman parte del área 5, mientras que el área Nielsen 6 está constituida por las provincias españolas de Cantabria, Vizcaya, Guipúzcoa, Navarra, Burgos, Álava, La Rioja y Palencia; el Área Metropolitana de Barcelona constituye el área Nielsen 7 y, finalmente, Madrid se corresponde con el área Nielsen 8.

6. La edad del responsable de la compra (**edad**) tiene tres categorías que representan tres segmentos de edad. Siguiendo los criterios de segmentación de AC Nielsen, los puntos de corte son las edades de 35 y 54 años.

7. La **situación_laboral** del responsable de la compra hace alusión a si el comprador trabaja también fuera del hogar.

8. El volumen del gasto total *per cápita* de la unidad de consumo se mide mediante el gasto anual *per cápita* realizado por el hogar tomando logaritmos (**log[gastos]**).

9. La variable frecuencia de compra (**frecuencia**) es el número de veces -número de viajes de compra- que ha acudido la familia a realizar la compra a lo largo del año (período de estudio).

En la tabla 1 se muestran las estadísticas descriptivas de la muestra considerada en la investigación así como las categorías de referencia de cada variable cualitativa.

Tabla 1

ANÁLISIS DESCRIPTIVO DE LAS VARIABLES INDEPENDIENTES(1)

VARIABLES	Media	Desviación Típica
Gasto <i>per cápita</i> (€)	606,96	382,15
Frecuencia	51,39	29,47

	1*	2	3	4	5	6	7	8	Total
Área_Nielsen	12,1%	15,4%	17,6%	9%	11,9%	10,1%	11,2%	12,7%	100%

	Joven (< 35 años)	Media (Entre 35 y 54 años)*	Anciana (> 55 años)	Total
Edad	11,6%	51,9%	36,5%	100%

	1	2	3	4	5 o más*	Total
Tamaño_hogar	6,0%	21,8%	24,1%	30,9%	17,2%	100%

	Baja	Media*	Alta	Total
Clase_social	25,9%	52,6%	21,5%	100%

	0	1	2	3*	Total
Tamaño_población	15,7%	22,5%	27,6%	34,3%	100%

	No*	Sí	Total
¿Hijos < 12 años?	64,8%	35,2%	100%
¿El responsable de la compra trabaja fuera del hogar?	81,0%	19,0%	100%

(1) Para cada variable categórica la clase de referencia se señala con *

3. METODOLOGÍA ESTADÍSTICA

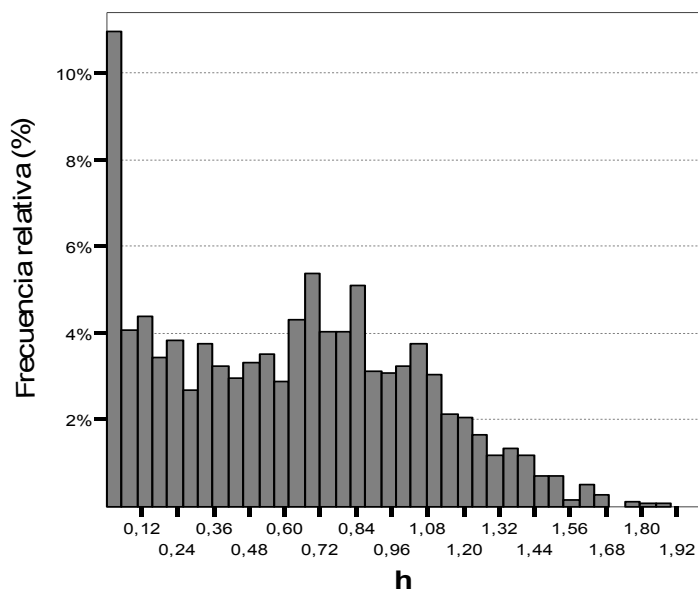
En esta sección se presenta la descripción de los modelos propuestos, la aproximación bayesiana que se ha seguido en el análisis, y los criterios que se han utilizado en el proceso de comparación de los modelos.

3.1 Descripción de los modelos

Sea N el número de hogares de la muestra considerada en el estudio. Sea H el índice Herfindahl-Hirschman de comportamiento variado para todo el año y sean $\mathbf{X} = (X_1, \dots, X_p)$ las covariables medidas en cada hogar donde supondremos, en particular, que $X_1 = 1$.

Los datos del problema vienen dados por $\mathbf{D} = \{(\mathbf{x}_i', H_i)\}_{i=1, \dots, N}$ con $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$, el vector de covariables y H_i el valor del índice Herfindahl-Hirschman de comportamiento variado para todo el año correspondientes al i -ésimo hogar de la muestra. Con el fin de conseguir un mayor grado de normalidad de la variable dependiente se realizó la transformación $h_i = -\log(-H_i)$. En la figura 1 se muestra el histograma correspondiente a dicha variable, en el que se aprecia el carácter mixto de la misma debido a la presencia de un porcentaje de familias (6%) que muestran un comportamiento no variado del escenario de compra ($h = 0$) respecto del resto.

Figura 1

HISTOGRAMA DE LA VARIABLE TRANSFORMADA $h = -\log(-H)$ 

Para contrastar todas y cada una de las hipótesis planteadas y dado el carácter mixto de la variable a explicar, se han utilizado dos tipos de modelos: por un lado, un modelo de regresión tipo Tobit dado por la expresión:

$$h_i = -\log(-H_i) = \begin{cases} 0 & \text{con probabilidad } 1 - \Phi(\tau^{1/2}(\beta' x_i)) \\ \beta' x_i + \varepsilon_i & \text{con probabilidad } \Phi(\tau^{1/2}(\beta' x_i)) \end{cases} \quad [3.1]$$

con $\varepsilon_i \approx N(0, \tau^{-1})$; $i = 1, \dots, N$, $\beta = (\beta_1, \dots, \beta_p)$, y donde Φ denota la función de distribución de una normal estándar. Por lo tanto, el modelo [3.1] supone que es el mismo el efecto ejercido por las covariables X sobre la decisión que toma un hogar acerca de si comprar o no en más de un establecimiento y su efecto sobre su grado de comportamiento variado. Con el fin de analizar hasta qué punto dicha hipótesis es cierta consideramos como alternativo un modelo de Decisiones Separadas (Lancaster, 2004, página 243) dado por las expresiones:

$$h_i = -\log(-H_i) = \begin{cases} 0 & \text{con probabilidad } 1 - \Phi(\gamma' x_i) \\ \beta' x_i + \varepsilon_i & \text{con probabilidad } \Phi(\gamma' x_i) \end{cases} \quad [3.2]$$

con $\gamma = (\gamma_1, \dots, \gamma_p)'$ y $\varepsilon_i \approx N(0, \tau^{-1})$; $i = 1, \dots, N$.

El modelo dado por [3.2] plantea un marco más general para explicar el proceso generador de los datos observados, distinguiendo la decisión de comprar en más de un establecimiento de la de comprar únicamente en uno y, en caso que se haya decidido comprar en más de un establecimiento, midiendo el grado de variabilidad del escenario de compra del hogar. En particular, si $\tau^{-1/2} \gamma = \beta$ la influencia ejercida por las covariables, X , en dichas decisiones es la misma y vendría descrita por el modelo Tobit [3.1].

Para comparar y estimar los parámetros de cada uno de los modelos considerados en el trabajo se ha adoptado una metodología bayesiana (ver, por ejemplo, Rossi, Allenby y McCulloch, 2005, para una reciente descripción de dicha perspectiva y sus ventajas en el tratamiento de datos de Marketing).

3.2 Análisis bayesiano

La aproximación bayesiana considera a los parámetros del modelo como variables aleatorias, cuyas distribuciones de probabilidad a priori cuantifican la incertidumbre asociada al desconocimiento del valor de dichos parámetros, previa a la observación de los datos. En nuestro caso, se han utilizado las siguientes *distribuciones a priori*, tradicionalmente adoptadas en la literatura (i.e., Lancaster, 2004, apéndice 3, página 377):

$$\gamma \approx N_{p+1}(0, \Sigma_\gamma) \quad [3.3]$$

$$\beta \approx N_{p+1}(0, \Sigma_\beta) \quad [3.4]$$

$$\tau \approx \text{Gamma}\left(\frac{n}{2}, \frac{\alpha}{2}\right) \quad [3.5]$$

donde las matrices $\Sigma_\gamma, \Sigma_\beta$ de dimensión $(p + 1) \times (p + 1)$ y las constantes n y α se suponen conocidas y [3.3]-[3.5] se suponen independientes entre sí. Con el fin de

que dichas distribuciones no sean muy informativas y no ejerzan excesiva influencia sobre los resultados obtenidos se han tomado $\sum_{\gamma} = \sum_{\beta} = 10000I_{p+1}$ y $n = \alpha = 0.01$.

Una vez especificadas las ecuaciones del modelo y la distribución a priori sobre los parámetros del mismo, las inferencias se realizan utilizando su *distribución a posteriori* conjunta que, aplicando el Teorema de Bayes, viene dada por:

$$\begin{aligned} \pi(\beta, \tau | D) \propto \prod_{i=1}^N \left(\tau^{1/2} \exp \left[-\frac{\tau}{2} (h_i - \beta' x_i)^2 \right] \right)^{(1-y_i)} \left(1 - \Phi(\tau^{1/2} (\beta' x_i)) \right)^{y_i} \\ \times \exp \left[-\frac{1}{2} \beta' \Sigma_{\beta}^{-1} \beta \right] \tau^{\frac{n}{2}-1} \exp \left[-\frac{a}{2} \tau \right] I_{(0, \infty)}(\tau) \end{aligned} \quad [3.6]$$

en el caso del modelo Tobit [3.1] y por:

$$\begin{aligned} \pi(\gamma, \beta, \tau | D) \propto \prod_{i=1}^N \left(\frac{\tau^{1/2} \exp \left[-\frac{\tau}{2} (h_i - \beta' x_i)^2 \right]}{\Phi(\tau^{1/2} (\beta' x_i))} \right)^{(1-y_i)} \prod_{i=1}^N \left[\Phi(\gamma' x_i) \right]^{(1-y_i)} \left[1 - \Phi(\gamma' x_i) \right]^{y_i} \times \\ \times \exp \left[-\frac{1}{2} \gamma' \Sigma_{\gamma}^{-1} \gamma \right] \exp \left[-\frac{1}{2} \beta' \Sigma_{\beta}^{-1} \beta \right] \tau^{\frac{n}{2}-1} \exp \left[-\frac{a}{2} \tau \right] I_{(0, \infty)}(\tau) \end{aligned} \quad [3.7]$$

en el caso del modelo de Decisiones Separadas [3.2], donde $I_A(x) = 1$ si $x \in A$ y 0 en otro caso, es la función indicador e $y_i = I_{(-\infty, 0]}(h_i)$; $i=1, \dots, N$.

Ya que estas dos distribuciones no son tratables analíticamente utilizaremos métodos MCMC (*Monte Carlo Markov Chain*) para su cálculo. Dichos métodos se basan en construir sendas cadenas de Markov cuyas distribuciones estacionarias sean aquellas que se quiere calcular (en nuestro caso, las distribuciones [3.6] y [3.7]). Simulando dichas cadenas un número suficientemente elevado de veces se obtendrá una muestra de cada una de dichas distribuciones a partir de las cuales se podrán calcular momentos, distribuciones marginales, cuantiles, etc. que nos permitirán estimar, los parámetros de los modelos [3.1] y [3.2]. En nuestro caso hemos utilizado el *Gibbs sampling* (Robert y Casella, 2004), cuya implementación

se realiza mediante cálculos estándar similares a los realizados en Lancaster (2004) o Rossi, Allenby y McCulloch (2005)(3).

3.3 Bondad de ajuste y comparación de modelos

Dado que en el estudio realizado se analizan y comparan diversos modelos, en este apartado se describen, brevemente, los criterios utilizados para analizar su bondad de ajuste a los datos y para llevar a cabo dicha comparación.

3.3.1 Bondad de ajuste

Debido a que en los modelos considerados en el trabajo se supone que los términos de error son normales, la evaluación de la bondad de los mismos hace referencia a la normalidad de sus residuos. Es bien conocido que, en modelos de variable dependiente limitada, el estimador máximo-verosímil puede ser inconsistente si falla dicha hipótesis (Arabmazar y Schmidt, 1982) y de ahí el interés de analizar este aspecto. Para ello utilizamos el estadístico LM propuesto en Bera, Jarque y Lee (1984) que, como el reciente trabajo de Holden (2004) demuestra, tiene un buen comportamiento en muestras finitas. Dado que la aproximación adoptada en este trabajo es bayesiana, utilizamos como medio de evaluación de la bondad de ajuste el pvalor predictivo a posteriori dado por:

$$\text{pvalor-Normalidad} = P[S_{B.J.L.}(h^{rep}, \theta) \geq S(h, \theta) | D] \quad [3.8]$$

donde $S_{B.J.L.}(h, \theta)$ denota el estadístico LM propuesto en Bera, Jarque y Lee (1984), $\mathbf{h} = (h_1, \dots, h_N)'$, θ es el vector de parámetros del modelo y \mathbf{h}^{rep} es una réplica de \mathbf{h} construida a partir de la distribución predictiva a posteriori:

$$f(\mathbf{h} | \mathbf{D}, \mathbf{X}) = \int f(\mathbf{h} | \theta, \mathbf{X}) \pi(\theta | \mathbf{D}) d\theta \quad [3.9]$$

Como Gelman, Carlin, Stern y Rubin (2004, Capítulo 6) argumentan, un pvalor muy bajo (<0.01) o muy alto (>0.99), indica un mal ajuste del modelo a los datos en términos del estadístico de contraste, $S_{B.J.L.}(h, \theta)$, y pone en duda la idoneidad de la hipótesis de normalidad utilizada en el modelo correspondiente.

(3) Todos los detalles matemáticos de los algoritmos utilizados en el trabajo están a disposición del lector interesado previa petición a los autores.

3.3.2 Comparación de modelos

Todos los criterios de comparación utilizados evalúan el comportamiento predictivo de los modelos en términos de cubrimientos de intervalos, errores medios y densidades.

Así, los cubrimientos empíricos de los intervalos de credibilidad bayesianos del 95% y del 99% vienen dados por la expresión:

$$\text{CUB}(100(1-\alpha)) = 100 \left(\frac{1}{N} \sum_{i=1}^N I_{\left[q_i\left(\frac{\alpha}{2}\right), q_i\left(1-\frac{\alpha}{2}\right) \right]}(h_i) \right) \quad [3.10]$$

donde $q_i(\alpha)$ denota el cuantil de orden α de la distribución predictiva $h_i | D, x_i$ $i = 1, \dots, N$.

Para evaluar el comportamiento predictivo en términos del error de predicción utilizamos la raíz cuadrada del error cuadrático medio (RMSE) y el error absoluto medio (MAD), que vienen dados por las siguientes expresiones:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (h_i - \hat{h}_i)^2} \quad [3.11]$$

$$\text{MAD} = \frac{1}{N} \sum_{i=1}^N |h_i - \hat{h}_i| \quad [3.12]$$

donde $\hat{h}_i = E[h_i | D, x_i]$; $i = 1, \dots, N$.

El comportamiento predictivo en términos de distribuciones predictivas se evalúa utilizando dos criterios: el primero es el logaritmo de la densidad predictiva a posteriori que viene dado por:

$$\text{LPRED} = \log(f(h|D, X)) \quad [3.13]$$

donde $f(h|D, X)$ es la función de densidad predictiva a posteriori dada por [3.9].

Este criterio analiza el grado de bondad de ajuste del modelo a los datos desde un punto de vista predictivo intramuestral pero tiene en cuenta la incertidumbre asociada a la estimación de los parámetros del modelo y penaliza a los modelos menos parsimoniosos. Su comportamiento asintótico es similar al del criterio AIC y

se utiliza para calcular el factor Bayes a posteriori (Aitkin, 1991). Tiene la ventaja, sobre el factor Bayes tradicional, de ser poco sensible a la distribución a priori utilizada, evitando la llamada *paradoja de Lindley*.

El segundo criterio es el logaritmo de la densidad predictiva univariante, propuesto por Draper, Krnjajić y Kottas (2006), que viene dado por:

$$LS_{FS} = \sum_{i=1}^N \log(f(h_i | D, x_i)) \quad [3.14]$$

Esta expresión es una aproximación a la densidad predictiva extramuestral, dejando un caso fuera, la cual viene dada por

$$LS_{CV} = \sum_{i=1}^N \log(f(h_i | D_{-i}, x_i)) \text{ donde } D_{-i} = D - \{(h_i, x_i)\}$$

Como demuestran Draper, Krnjajić y Kottas (2006), el comportamiento de LS_{FS} es mejor que el de LS_{CV} tanto asintóticamente como en muestras finitas. Su cálculo es mucho más rápido desde un punto de vista computacional.

Estos criterios se obtienen mediante cálculos estándar utilizando las muestras obtenidas de la distribución a posteriori de θ mediante el *Gibbs sampling*(4).

4. RESULTADOS

En esta sección se presentan los resultados obtenidos tanto en la estimación de los parámetros de los modelos como en la de los efectos marginales. Finalmente, se presenta una comparación de los modelos.

4.1 Estimación de los parámetros

En las tablas 2.1 y 2.2 se muestran las estimaciones de los parámetros de los modelos Tobit y de Decisiones Separadas, respectivamente. Más concretamente, para cada parámetro se presenta una estimación puntual dada por la mediana de su distribución a posteriori así como un intervalo de credibilidad bayesiano del 95%, cuyos límites vienen dados por los cuantiles 2.5 y 97.5 de dicha distribución.

(4) Todos los detalles matemáticos de dicho cálculo están a disposición del lector interesado previa petición a los autores.

Tabla 2.1
RESULTADOS DE LA ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO
TOBIT (3.1)

	<i>Efecto Tobit (β)</i>		
	Q2.5	Mediana	Q97.5
constante	1,5638	1,8142	2,0525
log[gastos]	-0,2497	-0,2118	-0,1736
frecuencia	0,0011	0,0018	0,0024
área_nielsen2	-0,1958	-0,1189	-0,0459
área_nielsen3	-0,1582	-0,0823	-0,0095
área_nielsen4	-0,0629	0,0244	0,1086
área_nielsen5	-0,0139	0,0668	0,1452
área_nielsen6	-0,1286	-0,0494	0,0341
área_nielsen7	-0,0739	0,0033	0,0819
área_nielsen8	-0,0459	0,0310	0,1109
edad_joven	-0,0796	-0,0134	0,0529
edad_anciana	-0,0819	-0,0329	0,0185
tamaño_hogar1	0,0380	0,1350	0,2250
tamaño_hogar2	0,0174	0,0881	0,1586
tamaño_hogar3	0,0006	0,0630	0,1203
tamaño_hogar4	-0,0407	0,0154	0,0723
clase_social_alta	-0,1017	-0,0576	-0,0151
clase_social_baja	-0,0513	-0,0004	0,0484
tamaño_población0	-0,0441	0,0200	0,0789
tamaño_población1	-0,0024	0,0451	0,1027
tamaño_población2	-0,0126	0,0374	0,0875
hijos<12 años	-0,0656	-0,0178	0,0331
situación_laboral	-0,0132	0,0333	0,0841
τ	5,1819	5,5456	5,8958

Nota: En negrita los coeficientes significativos al 95%.

Tabla 2.2
RESULTADOS DE LA ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO
DE DECISIONES SEPARADAS (3.2)

(Continúa)

	<i>Efecto Probit $(\tau^{-1/2} \gamma)$**</i>		
	Q2.5	Mediana	Q97.5
constante	1,2217	1,6895	2,1730
log[gastos]	-0,2640	-0,1929	-0,1167
frecuencia	0,0003	0,0018	0,0033
área_nielsen2	-0,2188	-0,0935	0,0192
área_nielsen3	-0,0688	0,0502	0,1688
área_nielsen4	0,0251	0,1890	0,3762
área_nielsen5	0,1771	0,3686	0,6433
área_nielsen6	-0,0207	0,1322	0,2757
área_nielsen7	-0,0348	0,1235	0,2886
área_nielsen8	-0,0592	0,0845	0,2299
edad_joven	-0,0733	0,0562	0,1991
edad_anciana	-0,1249	-0,0404	0,0536
tamaño_hogar1	-0,0464	0,1358	0,3226
tamaño_hogar2	-0,0654	0,0669	0,2010
tamaño_hogar3	-0,0701	0,0494	0,1727
tamaño_hogar4	-0,1196	-0,0134	0,0977
clase_social_alta	-0,1769	-0,0950	-0,0031
clase_social_baja	-0,1061	-0,0036	0,0962
tamaño_población0	-0,1781	-0,0624	0,0563
tamaño_población1	-0,0888	0,0126	0,1149
tamaño_población2	-0,0970	-0,0060	0,0902
hijos<12 años	-0,1038	-0,0130	0,0844
situación_laboral	-0,0515	0,0538	0,1609
τ			

Nota: En negrita los coeficientes significativos al 95%.

** Con el fin de hacer comparables los coeficientes de regresión de la parte probit de los modelos (3.1) y (3.2) presentamos las estimaciones a posteriori de $(\tau^{-1/2} \gamma)$

Tabla 2.2
RESULTADOS DE LA ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO
DE DECISIONES SEPARADAS (3.2)
(Conclusión)

	<i>Efecto Regresión (β)</i>		
	Q2.5	Mediana	Q97.5
constante	1,4616	1,6731	1,9043
log[gastos]	-0,2128	-0,1772	-0,1412
frecuencia	0,0008	0,0015	0,0021
área_nielsen2	-0,1629	-0,0953	-0,0181
área_nielsen3	-0,1704	-0,1068	-0,0372
área_nielsen4	-0,0947	-0,0146	0,0687
área_nielsen5	-0,0613	0,0090	0,0781
área_nielsen6	-0,1584	-0,0853	-0,0109
área_nielsen7	-0,1030	-0,0268	0,0512
área_nielsen8	-0,0631	0,0085	0,0785
edad_joven	-0,0836	-0,0236	0,0388
edad_anciana	-0,0775	-0,0267	0,0167
tamaño_hogar1	0,0133	0,1055	0,1981
tamaño_hogar2	0,0166	0,0766	0,1418
tamaño_hogar3	0,0012	0,0542	0,1161
tamaño_hogar4	-0,0373	0,0203	0,0773
clase_social_alta	-0,0813	-0,0373	0,0040
clase_social_baja	-0,0472	-0,0021	0,0423
tamaño_población0	-0,0134	0,0415	0,0978
tamaño_población1	-0,0021	0,0481	0,0976
tamaño_población2	-0,0070	0,0402	0,0845
hijos<12 años	-0,0669	-0,0211	0,0274
situación_laboral	-0,0195	0,0260	0,0750
τ	6,1722	6,5434	6,9896

Nota: En negrita los coeficientes significativos al 95%.

** Con el fin de hacer comparables los coeficientes de regresión de la parte probit de los modelos (3.1) y (3.2) presentamos las estimaciones a posteriori de $(\tau^{-1/2} \gamma)$

Con el fin de comprobar la validez de la hipótesis sobre la relación entre el comportamiento variado y la etapa del ciclo de vida a la que pertenece la familia, se realizó una estimación de dichos modelos utilizando covariables que miden la etapa del ciclo de vida en la que se encuentran los hogares. Los resultados para estos últimos modelos se muestran en las tablas 3.1 y 3.2.

Tabla 3.1

RESULTADOS DE LA ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO TOBIT (3.1) QUE INCORPORA LA HIPÓTESIS DEL CICLO DE VIDA

	<i>Efecto Tobit (β)</i>		
	<i>Q2.5</i>	<i>Mediana</i>	<i>Q97.5</i>
constante	1,5597	1,7935	2,0254
log[gastos]	-0,2376	-0,2003	-0,1653
frecuencia	0,0009	0,0016	0,0022
área_nielsen2	-0,1879	-0,1213	-0,0408
área_nielsen3	-0,1535	-0,0833	-0,0112
área_nielsen4	-0,0483	0,0323	0,1084
área_nielsen5	-0,0073	0,0687	0,1497
área_nielsen6	-0,1297	-0,0505	0,0377
área_nielsen7	-0,0751	0,0057	0,0762
área_nielsen8	-0,0463	0,0321	0,1068
edad_joven	-0,0521	0,0119	0,0698
edad_anciana	-0,0710	-0,0287	0,0102
clase_social_alta	-0,1026	-0,0572	-0,0111
clase_social_baja	-0,0529	-0,0042	0,0416
tamaño_población0	-0,0508	0,0127	0,0775
tamaño_población1	-0,0125	0,0409	0,0940
tamaño_población2	-0,0156	0,0323	0,0787
situación_laboral	-0,0011	0,0445	0,0914
hogar_unipersonal	-0,0209	0,0625	0,1405
hogar pluripersonal con niños < 12 años	-0,0673	-0,0240	0,0166
τ	5,1930	5,5245	5,8644

Nota: En negrita los coeficientes significativos al 95%.

Tabla 3.2

RESULTADOS DE LA ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO DECISIONES SEPARADAS (3.2) QUE INCORPORA LA HIPÓTESIS DEL CICLO DE VIDA

(Continúa)

	<i>Efecto Probit</i> $(\tau^{-1/2} \gamma)^{**}$		
	Q2.5	Mediana	Q97.5
constante	1,2476	1,6754	2,1410
log[gastos]	-0,2559	-0,1847	-0,1152
frecuencia	0,0003	0,0017	0,0032
área_nielsen2	-0,2146	-0,0962	0,0258
área_nielsen3	-0,0777	0,0535	0,1723
área_nielsen4	0,0188	0,1917	0,3773
área_nielsen5	0,1658	0,3817	0,6366
área_nielsen6	-0,0120	0,1252	0,2946
área_nielsen7	-0,0156	0,1281	0,2763
área_nielsen8	-0,0572	0,0834	0,2359
edad_joven	-0,0602	0,0709	0,2180
edad_anciana	-0,1303	-0,0477	0,0359
clase_social_alta	-0,1737	-0,0936	-0,0152
clase_social_baja	-0,0981	-0,0079	0,0917
tamaño_población0	-0,1857	-0,0719	0,0486
tamaño_población1	-0,0925	0,0075	0,1086
tamaño_población2	-0,1020	-0,0139	0,0849
situación_laboral	-0,0309	0,0624	0,1708
hogar_unipersonal	-0,0582	0,0880	0,2483
hogar pluripersonal con niños < 12 años	-0,0934	-0,0091	0,0726
τ			

Nota: En negrita los coeficientes significativos al 95%.

** Con el fin de hacer comparables los coeficientes de regresión de la parte probit de los modelos (3.1) y (3.2) presentamos las estimaciones a posteriori de $(\tau^{-1/2} \gamma)$

Tabla 3.2

RESULTADOS DE LA ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO
DECISIONES SEPARADAS (3.2) QUE INCORPORA LA HIPÓTESIS DEL
CICLO DE VIDA

(Conclusión)

	<i>Efecto Regresión</i> (β)		
	Q2.5	Mediana	Q97.5
constante	1,4519	1,6709	1,8942
log[gastos]	-0,2048	-0,1684	-0,1344
frecuencia	0,0007	0,0013	0,0019
área_nielsen2	-0,1649	-0,0929	-0,0279
área_nielsen3	-0,1731	-0,1051	-0,0399
área_nielsen4	-0,0918	-0,0103	0,0662
área_nielsen5	-0,0639	0,0132	0,0816
área_nielsen6	-0,1599	-0,0869	-0,0086
área_nielsen7	-0,1099	-0,0296	0,0468
área_nielsen8	-0,0696	0,0083	0,0815
edad_joven	-0,0592	-0,0041	0,0546
edad_anciana	-0,0597	-0,0202	0,0221
clase_social_alta	-0,0778	-0,0369	-0,0046
clase_social_baja	-0,0504	-0,0029	0,0428
tamaño_población0	-0,0172	0,0339	0,0917
tamaño_población1	-0,0058	0,0433	0,0919
tamaño_población2	-0,0066	0,0383	0,0834
situación_laboral	-0,0066	0,0361	0,0830
hogar_unipersonal	-0,0322	0,0459	0,1271
hogar pluripersonal con niños < 12 años	-0,0682	-0,0256	0,0185
τ	6,1343	6,5436	6,9854

Nota: En negrita los coeficientes significativos al 95%.

** Con el fin de hacer comparables los coeficientes de regresión de la parte probit de los modelos (3.1) y (3.2) presentamos las estimaciones a posteriori de $(\tau^{-1/2} \gamma)$

Los parámetros estimados para ambos modelos coinciden en su mayoría. Se observa, en primer lugar, que las familias de clase social alta tienden a tener un escenario de compra menos variado que aquellas de clase social inferior, lo cual viene a corroborar la primera de las hipótesis planteadas (H1).

Por otra parte, cuanto mayor es el tamaño de la unidad familiar, menor es la variación del escenario de compra, por lo que no podemos aceptar la hipótesis H2. Es decir, los hogares uni, bi y tri-personales presentan un escenario de compra más variado que los hogares de mayor tamaño.

No se observa que la variable ciclo de vida ejerza una influencia significativa sobre el comportamiento variado de las familias para ninguno de los modelos considerados en el trabajo (ver tablas 3.1 y 3.2). Consecuentemente, no podemos aceptar la hipótesis H3.

Por otro lado, existe una relación entre el área geográfica en la que vive la familia y el grado de variación del escenario de compra. Particularmente, se observa que las familias que viven en las zonas del Levante y del Sur (áreas Nielsen 2 y 3) tienen un comportamiento menos variado que las familias del resto de la Península, mientras que el tamaño de la población no ejerce una influencia significativa sobre la variación en el escenario de compra. Consecuentemente, podemos aceptar la hipótesis H4 pero no la hipótesis H5.

La influencia de la edad del responsable de la compra sobre la variación en el escenario de compra al igual que su situación laboral no resulta significativa, por lo que no podemos aceptar las hipótesis H6 y H7.

Por lo que respecta al volumen total de la compra *per cápita*, la estimación comprueba que, cuanto mayor es este, menor es la variación del escenario de compra, de manera que no podemos aceptar la hipótesis de trabajo H8.

Finalmente, se observa que cuanto mayor es el número de veces con que la unidad de consumo realiza la compra, esto es, cuanto mayor es la frecuencia de compra, mayor es la variación del escenario de compra. Por tanto, podemos aceptar la hipótesis H9.

4.2 Estimación de los efectos marginales

Con el fin de evaluar las influencias de las variables explicativas sobre la probabilidad de que una familia muestre un comportamiento variado y sobre el valor esperado de este, en las tablas 4.1, 4.2, 5.1 y 5.2 se muestran, para cada uno de los modelos, los efectos marginales de las variaciones unitarias de cada una de dichas variables sobre la probabilidad de que una familia tenga un comportamiento

variado (denominado efecto probit) y sobre el valor esperado del logaritmo del índice de Herfindahl-Hirschman, h_i (denominado efecto regresión). Dichos efectos vienen dados por las expresiones (ver Lancaster, 2004 y Greene, 2004):

$$\text{Efecto Probit} = \frac{\partial P(h_i > 0 | x_i, \theta)}{\partial x_{ik}} = \theta_k \phi(x_i; \theta) \quad [4.1]$$

donde $\theta = \beta$ si el modelo es un Tobit y $\theta = \gamma$ si el modelo es uno de Decisiones Separadas y

$$\text{Efecto Regresión} = \frac{\partial E(h_i | h_i > 0, x_i, \theta)}{\partial x_j} = \beta_j \left[1 - \frac{\phi(z_i)(\phi(z_i) + z_i \Phi(z_i))}{(\Phi(z_i))^2} \right]$$

$$\text{con } z_i = \tau^{1/2}(\beta' x_i) \quad [4.2]$$

donde ϕ y Φ denotan las funciones de densidad y distribución de una normal estándar, respectivamente.

Tabla 4.1
EFECTOS MARGINALES DE LAS COVARIABLES DEL MODELO TOBIT (3.1)

(Continúa)

	<i>Efecto Probit (4.1)</i>		
	Q2.5	Mediana	Q97.5
log[gastos]	-0,0884	-0,0746	-0,0615
frecuencia	0,0004	0,0006	0,0009
área_nielsen2	-0,0691	-0,0419	-0,0162
área_nielsen3	-0,0565	-0,0291	-0,0033
área_nielsen4	-0,0219	0,0086	0,0382
área_nielsen5	-0,0050	0,0235	0,0510
área_nielsen6	-0,0454	-0,0173	0,0120
área_nielsen7	-0,0265	0,0012	0,0287
área_nielsen8	-0,0160	0,0110	0,0387
edad_joven	-0,0282	-0,0047	0,0186
edad_anciana	-0,0285	-0,0117	0,0066
tamaño_hogar1	0,0133	0,0476	0,0811
tamaño_hogar2	0,0061	0,0310	0,0563
tamaño_hogar3	0,0002	0,0221	0,0434
tamaño_hogar4	-0,0138	0,0054	0,0261
clase_social_alta	-0,0360	-0,0204	-0,0052
clase_social_baja	-0,0180	-0,0001	0,0178
tamaño_población0	-0,0155	0,0070	0,0279
tamaño_población1	-0,0008	0,0160	0,0371
tamaño_población2	-0,0045	0,0132	0,0314
hijos<12 años	-0,0228	-0,0063	0,0118
situación_laboral	-0,0045	0,0118	0,0296

Nota: En negrita los coeficientes significativos al 95%.

Tabla 4.1
EFFECTOS MARGINALES DE LAS COVARIABLES DEL MODELO TOBIT (3.1)
 (Conclusión)

	<i>Efecto Regresión (4.2)</i>		
	Q2.5	Mediana	Q97.5
log[gastos]	-0,0467	-0,0381	-0,0307
frecuencia	0,0002	0,0003	0,0005
área_nielsen2	-0,0359	-0,0213	-0,0080
área_nielsen3	-0,0297	-0,0148	-0,0016
área_nielsen4	-0,0113	0,0044	0,0202
área_nielsen5	-0,0026	0,0120	0,0264
área_nielsen6	-0,0234	-0,0090	0,0062
área_nielsen7	-0,0138	0,0006	0,0143
área_nielsen8	-0,0081	0,0056	0,0201
edad_joven	-0,0148	-0,0023	0,0094
edad_anciana	-0,0144	-0,0059	0,0035
tamaño_hogar1	0,0068	0,0241	0,0433
tamaño_hogar2	0,0030	0,0157	0,0307
tamaño_hogar3	0,0001	0,0113	0,0232
tamaño_hogar4	-0,0070	0,0027	0,0134
clase_social_alta	-0,0181	-0,0104	-0,0025
clase_social_baja	-0,0090	-0,0001	0,0089
tamaño_población0	-0,0075	0,0036	0,0144
tamaño_población1	-0,0004	0,0083	0,0191
tamaño_población2	-0,0024	0,0068	0,0166
hijos<12 años	-0,0116	-0,0032	0,0061
situación_laboral	-0,0023	0,0061	0,0154

Nota: En negrita los coeficientes significativos al 95%.

Tabla 4.2
EFFECTOS MARGINALES DE LAS COVARIABLES DEL MODELO DE
DECISIONES SEPARADAS (3.2)

(Continúa)

	<i>Efecto Probit (4.1)</i>		
	Q2.5	Mediana	Q97.5
log[gastos]	-0,0982	-0,0510	-0,0215
frecuencia	0,0001	0,0005	0,0011
área_nielsen2	-0,0680	-0,0241	0,0055
área_nielsen3	-0,0214	0,0126	0,0445
área_nielsen4	0,0055	0,0483	0,1228
área_nielsen5	0,0372	0,0971	0,2200
área_nielsen6	-0,0044	0,0339	0,0920
área_nielsen7	-0,0079	0,0320	0,0946
área_nielsen8	-0,0158	0,0221	0,0753
edad_joven	-0,0205	0,0135	0,0584
edad_anciana	-0,0374	-0,0100	0,0168
tamaño_hogar1	-0,0100	0,0356	0,1169
tamaño_hogar2	-0,0126	0,0181	0,0773
tamaño_hogar3	-0,0162	0,0124	0,0647
tamaño_hogar4	-0,0272	-0,0036	0,0365
clase_social_alta	-0,0531	-0,0241	-0,0008
clase_social_baja	-0,0270	-0,0009	0,0320
tamaño_población0	-0,0499	-0,0154	0,0178
tamaño_población1	-0,0209	0,0033	0,0421
tamaño_población2	-0,0248	-0,0015	0,0326
hijos<12 años	-0,0266	-0,0032	0,0312
situación_laboral	-0,0133	0,0135	0,0534

Nota: En negrita los coeficientes significativos al 95%.

Tabla 4.2
EFFECTOS MARGINALES DE LAS COVARIABLES DEL MODELO DE
DECISIONES SEPARADAS (3.2)

(Conclusión)

	<i>Efecto Regresión (4.2)</i>		
	Q2.5	Mediana	Q97.5
log[gastos]	-0,0355	-0,0282	-0,0219
frecuencia	0,0001	0,0002	0,0004
área_nielsen2	-0,0263	-0,0152	-0,0027
área_nielsen3	-0,0290	-0,0170	-0,0061
área_nielsen4	-0,0149	-0,0024	0,0113
área_nielsen5	-0,0098	0,0014	0,0123
área_nielsen6	-0,0253	-0,0134	-0,0017
área_nielsen7	-0,0165	-0,0042	0,0081
área_nielsen8	-0,0098	0,0013	0,0128
edad_joven	-0,0134	-0,0038	0,0063
edad_anciana	-0,0122	-0,0042	0,0027
tamaño_hogar1	0,0022	0,0168	0,0331
tamaño_hogar2	0,0025	0,0121	0,0241
tamaño_hogar3	0,0002	0,0085	0,0199
tamaño_hogar4	-0,0058	0,0032	0,0135
clase_social_alta	-0,0127	-0,0060	-0,0006
clase_social_baja	-0,0074	-0,0003	0,0071
tamaño_población0	-0,0021	0,0065	0,0159
tamaño_población1	-0,0004	0,0077	0,0161
tamaño_población2	-0,0010	0,0064	0,0143
hijos<12 años	-0,0104	-0,0034	0,0046
situación_laboral	-0,0030	0,0041	0,0122

Nota: En negrita los coeficientes significativos al 95%.

Tabla 5.1
EFFECTOS MARGINALES DE LAS COVARIABLES DEL MODELO TOBIT (3.1)
QUE INCORPORA LA HIPÓTESIS DEL CICLO DE VIDA

(Continúa)

	<i>Efecto Probit (4.1)</i>		
	<i>Q2.5</i>	<i>Mediana</i>	<i>Q97.5</i>
log[gastos]	-0,0818	-0,0691	-0,0570
frecuencia	0,0003	0,0005	0,0008
área_nielsen2	-0,0645	-0,0418	-0,0140
área_nielsen3	-0,0538	-0,0289	-0,0037
área_nielsen4	-0,0167	0,0111	0,0370
área_nielsen5	-0,0026	0,0237	0,0516
área_nielsen6	-0,0450	-0,0173	0,0133
área_nielsen7	-0,0253	0,0019	0,0267
área_nielsen8	-0,0157	0,0111	0,0372
edad_joven	-0,0180	0,0041	0,0243
edad_anciana	-0,0242	-0,0098	0,0036
clase_social_alta	-0,0349	-0,0198	-0,0039
clase_social_baja	-0,0182	-0,0014	0,0147
tamaño_población0	-0,0172	0,0044	0,0273
tamaño_población1	-0,0043	0,0140	0,0328
tamaño_población2	-0,0053	0,0111	0,0277
situación_laboral	-0,0004	0,0154	0,0320
hogar_unipersonal	-0,0070	0,0214	0,0492
hogar pluripersonal con niños < 12 años	-0,0234	-0,0083	0,0057

Nota: En negrita los coeficientes significativos al 95%.

Tabla 5.1

EFFECTOS MARGINALES DE LAS COVARIABLES DEL MODELO TOBIT (3.1)
QUE INCORPORA LA HIPÓTESIS DEL CICLO DE VIDA

(Conclusión)

	<i>Efecto Regresión (4.2)</i>		
	Q2.5	Mediana	Q97.5
log[gastos]	-0,0413	-0,0343	-0,0276
frecuencia	0,0001	0,0003	0,0004
área_nielsen2	-0,0323	-0,0207	-0,0070
área_nielsen3	-0,0276	-0,0143	-0,0018
área_nielsen4	-0,0086	0,0055	0,0184
área_nielsen5	-0,0013	0,0118	0,0258
área_nielsen6	-0,0226	-0,0085	0,0068
área_nielsen7	-0,0124	0,0009	0,0133
área_nielsen8	-0,0079	0,0055	0,0186
edad_joven	-0,0087	0,0020	0,0122
edad_anciana	-0,0118	-0,0048	0,0018
clase_social_alta	-0,0173	-0,0098	-0,0020
clase_social_baja	-0,0090	-0,0007	0,0075
tamaño_población0	-0,0085	0,0022	0,0138
tamaño_población1	-0,0022	0,0069	0,0165
tamaño_población2	-0,0027	0,0054	0,0141
situación_laboral	-0,0002	0,0077	0,0159
hogar_unipersonal	-0,0034	0,0106	0,0247
hogar pluripersonal con niños < 12 años	-0,0117	-0,0041	0,0028

Nota: En negrita los coeficientes significativo

Tabla 5.2
EFFECTOS MARGINALES DE LAS COVARIABLES DEL MODELO
DECISIONES SEPARADAS (3.2) QUE INCORPORA LA HIPÓTESIS DEL
CICLO DE VIDA

(Continúa)

	<i>Efecto Probit (4.1)</i>		
	Q2.5	Mediana	Q97.5
log[gastos]	-0,0767	-0,0444	-0,0230
frecuencia	0,0001	0,0004	0,0009
área_nielsen2	-0,0583	-0,0219	0,0068
área_nielsen3	-0,0232	0,0123	0,0389
área_nielsen4	0,0042	0,0454	0,1058
área_nielsen5	0,0323	0,0899	0,1856
área_nielsen6	-0,0031	0,0289	0,0857
área_nielsen7	-0,0039	0,0298	0,0779
área_nielsen8	-0,0157	0,0195	0,0641
edad_joven	-0,0140	0,0172	0,0608
edad_anciana	-0,0318	-0,0112	0,0088
clase_social_alta	-0,0435	-0,0220	-0,0035
clase_social_baja	-0,0227	-0,0018	0,0248
tamaño_población0	-0,0421	-0,0165	0,0132
tamaño_población1	-0,0191	0,0016	0,0337
tamaño_población2	-0,0221	-0,0033	0,0281
situación_laboral	-0,0079	0,0143	0,0490
hogar_unipersonal	-0,0142	0,0207	0,0759
hogar pluripersonal con niños < 12 años	-0,0229	-0,0022	0,0196

Nota: En negrita los coeficientes significativos al 95%.

Tabla 5.2
EFFECTOS MARGINALES DE LAS COVARIABLES DEL MODELO
DECISIONES SEPARADAS (3.2) QUE INCORPORA LA HIPÓTESIS DEL
CICLO DE VIDA

(Conclusión)

	<i>Efecto Regresión (4.2)</i>		
	Q2.5	Mediana	Q97.5
log[gastos]	-0,0316	-0,0255	-0,0201
frecuencia	0,0001	0,0002	0,0003
área_nielsen2	-0,0254	-0,0140	-0,0042
área_nielsen3	-0,0280	-0,0161	-0,0057
área_nielsen4	-0,0139	-0,0016	0,0100
área_nielsen5	-0,0099	0,0020	0,0124
área_nielsen6	-0,0241	-0,0131	-0,0013
área_nielsen7	-0,0165	-0,0044	0,0073
área_nielsen8	-0,0103	0,0013	0,0121
edad_joven	-0,0088	-0,0006	0,0082
edad_anciana	-0,0088	-0,0030	0,0034
clase_social_alta	-0,0113	-0,0056	-0,0007
clase_social_baja	-0,0076	-0,0005	0,0069
tamaño_población0	-0,0026	0,0051	0,0145
tamaño_población1	-0,0008	0,0065	0,0146
tamaño_población2	-0,0009	0,0058	0,0131
situación_laboral	-0,0010	0,0055	0,0126
hogar_unipersonal	-0,0048	0,0070	0,0195
hogar pluripersonal con niños < 12 años	-0,0101	-0,0038	0,0029

Nota: En negrita los coeficientes significativos al 95%.

Las tablas 4.1, 4.2, 5.1 y 5.2 contienen los valores estimados de los efectos marginales medianos estimados a posteriori para cada una de las variables explicativas y cada uno de los modelos comparados, así como sus correspondientes intervalos de credibilidad del 95%. Esta estimación se realiza para una familia de características medias tomando, en cada una de las variables categóricas, los valores de referencia como punto de comparación (ver tabla 1). Así, por ejemplo, el

modelo Tobit estima que si una familia compra y vive en el área Nielsen 2, tiene una probabilidad un 4,19% menor que una familia del área Nielsen 1 de mostrar un comportamiento variado de compra, y un índice de variación de compra, H_i , un 2,13% menor (ver tabla 4.1).

Con carácter general, los valores de los efectos estimados por los modelos Tobit y de Decisiones Separadas toman valores similares siendo los signos y su significación estadística similar a la de los coeficientes γ y β correspondientes. Todo ello viene a corroborar los resultados obtenidos tras la estimación de los parámetros de los modelos.

4.3 Bondad de ajuste y comparación de modelos

En la tabla 6 se muestran los resultados obtenidos en el proceso de análisis de la bondad de ajuste y comparación de los modelos considerados en el trabajo. Se observa, en primer lugar, que la hipótesis de normalidad de los residuos no es rechazada en ninguno de los modelos para un nivel del 1% de credibilidad. Sin embargo, para los modelos Tobit, existe un rechazo débil de la normalidad para un nivel del 5%. No obstante, dado que el porcentaje de censura es bajo (un 6%), concluimos que el ajuste de los modelos a los datos es adecuado para llevar a cabo inferencias acerca del efecto de las covariables sobre la variable dependiente ya que, según Arabmazar y Schmidt (1982), el sesgo asintótico de los estimadores no es importante para porcentajes de censura inferiores al 10%.

Respecto al proceso de comparación, se observa que los modelos Tobit se ajustan mejor a los datos tanto en términos explicativos como predictivos en todos los criterios considerados. Ello se debe, muy probablemente, a que el porcentaje de familias que no tiene un comportamiento variado es bajo (un 6%) por lo que su influencia en el proceso de estimación de los modelos comparados es muy pequeña y, por lo tanto, puede no resultar suficiente como para capturar comportamientos diferenciados entre las familias que tienen un comportamiento variado y las que no.

Tabla 6
COMPARACIÓN DE LOS MODELOS AJUSTADOS(*)

<i>Criterio</i>	<i>Tobit</i>	<i>Decisiones Separadas</i>	<i>Tobit Ciclo de Vida</i>	<i>Decisiones Separadas Ciclo de Vida</i>		
RMSE	0,4015 (2,3x10 ⁻⁵) ⁺	0,4038 (3,3x10 ⁻⁵)	0,4022 (2,1x10 ⁻⁵)	0,4045 (3,3x10 ⁻⁵)		
MAD	0,3387 (2,5x10 ⁻⁵) ⁺	0,3414 (3,2x10 ⁻⁵)	0,3393 (2,4x10 ⁻⁵)	0,3420 (3,0x10 ⁻⁵)		
LPRED	-1203,9 (0,16) ⁺⁺	-1214,5 (0,50)	-1206,7 (0,12)	-1212,5 (0,98)		
LS_{FS}	-1207,7 (0,12) ⁺	-1231,8 (0,29)	-1209,9 (0,11)	-1233,4 (0,29)		
CUB95	97,72 (0,50) ⁺⁺⁺	96,93 (0,58)	97,62 (0,51)	96,83 (0,58)		
CUB99	99,65 (0,20) ⁺⁺⁺	99,60 (0,21)	99,60 (0,21)	99,60 (0,21)		
Pvalor Normalidad	0,0222	Probit	Regresión Truncada	0,0167	Probit	Regresión Truncada
		0,1100	0,1455		0,1722	0,1478

(*) Entre paréntesis figura el error estándar de Monte Carlo.

⁺ El error estándar ha sido calculado mediante la expresión $\frac{s_T}{\sqrt{S}}$ donde s_T^2 es la varianza muestral de $\{T(\theta^{(i)}); i=1, \dots, S\}$ siendo $\{\theta^{(i)}; i=1, \dots, S\}$ la muestra de la distribución a posteriori y donde $T(\theta^{(i)})$ con $T \in \{RMSE, MAD, LS_{FS}\}$ denota el valor de T en $\theta^{(i)}$

⁺⁺ El error estándar ha sido calculado aplicando el método delta, mediante $\frac{s_f}{\text{PRED}\sqrt{S}}$ donde s_f^2 es la varianza muestral de $\{f(h|X, \theta^{(i)}); i=1, \dots, S\}$ siendo $\{\theta^{(i)}; i=1, \dots, S\}$ la muestra de la distribución a posteriori, $\text{PRED} = \exp(\text{LPRED})$, y donde $f(h|X, \theta^{(i)})$ es el valor de la verosimilitud del modelo correspondiente calculado en $\theta^{(i)}$.

⁺⁺⁺ El error estándar ha sido calculado mediante la expresión $\sqrt{\frac{p(1-p)}{S}}$ donde $p = \frac{\text{CUB}}{100}$

De los cuatro modelos sometidos a comparación, el que mejor comportamiento muestra es el modelo Tobit que no incorpora la variable ciclo de vida en el conjunto de variables explicativas. Por lo tanto, concluimos:

1) la inexistencia de diferencias significativas entre las familias que tienen un comportamiento variado y las que no

2) que el ciclo de vida del hogar no ejerce una influencia digna de mención sobre dicho comportamiento.

CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

Esta investigación quiere ser una contribución original en el campo de estudio del comportamiento del consumidor teniendo en cuenta que, hasta el momento, ningún estudio había tratado el comportamiento variado considerando la composición del escenario de compra. Concretamente, se han estudiado relaciones entre la composición del escenario de compra de los hogares y algunas características demográficas y de comportamiento de compra de los mismos. Dada la naturaleza mixta de la variable dependiente objeto de estudio, se ha hecho uso de modelos de regresión Tobit y de Decisiones separadas (Lancaster, 2004), analizados mediante una metodología bayesiana. Lancaster (2004) y Rossi, Allenby y McCulloch (2005) justifican la adecuación de dicha metodología para el tratamiento y análisis de datos en la investigación desarrollada en el área de Comercialización e Investigación de Mercados.

Esta investigación parte de reconocer que, en el mercado minorista de alimentación, las unidades de consumo tienen un determinado escenario de compra regular que está formado por un conjunto de establecimientos entre los que se reparte el gasto total de la compra destinada a resolver una necesidad. El análisis de la base de datos realizado evidencia que entre las unidades de consumo: hogares españoles, predominan los escenarios de compra cruzada, de tal modo que, tan sólo el 6% de los hogares panelistas acude a resolver la compra a un único establecimiento.

Tras observar el comportamiento de compra realizado durante un año, la generalidad de los hogares españoles realiza la compra en una diversidad de establecimientos que conforman su escenario de compra habitual. Este se encuentra influenciado por diversas variables: la frecuencia de compra, el volumen total de la cesta de la compra (en sentido contrario al esperado), el nivel de ingresos del hogar, su tamaño y el área geográfica en la que residen. Sin embargo, no se ha encontrado relación entre el grado de variación que presenta el escenario de compra del hogar y la edad ni la situación laboral del responsable de la compra.

Tampoco la etapa del ciclo de vida en la que se encuentra la familia ni el tamaño de la población en la que vive son determinantes de la variación observada en su escenario de compra habitual.

Consecuentemente, por lo que respecta a las características demográficas y socio-económicas de los hogares, cabe afirmar que cuanto mayor es el nivel de ingresos de la familia y/o mayor es su tamaño, el gasto en un mismo formato es mayor y, por tanto, menor es la variación observada en el escenario de compra.

Por lo que respecta a las obligaciones familiares derivadas de la situación laboral del responsable de la compra en el hogar, cabe señalar que, en contra de lo esperado, no se ha encontrado una relación entre aquellos compradores que disponen de menos tiempo y tienen más compromisos, y un comportamiento de compra menos variado.

Por otro lado, el hecho de que el escenario de compra en un momento dado sea menos variado cuanto mayor es la asignación de gasto de cada miembro del hogar, viene a confirmar que existe un mayor grado de lealtad al establecimiento principal de compra por parte de aquellas unidades de consumo con un volumen total de compra relativamente elevado. Mägi (2003), consideraba de interés para la gestión averiguar en qué grado los compradores que incurren en un mayor volumen de gasto, difieren de los demás en su propensión a concentrar más sus compras, llegando incluso a resolver la compra en un único establecimiento. En efecto, las empresas de distribución comercial pueden aprovechar esta circunstancia en sus programas de fidelización de clientes, identificando los hogares con un mayor volumen de compra como aquéllos con un mayor potencial para la empresa. La gestión del minorista puede decidir un tratamiento especial para este tipo de clientes que estarán dispuestos a dirigir a su establecimiento una mayor cantidad de gasto.

En cuanto a la frecuencia de compra, se observa que cuanto mayor es ésta, mayor es el comportamiento variado de consumo de la unidad familiar. Además, la pertenencia a áreas Nielsen 2 y 3, que cubren el litoral sur y levantino de la península, se relaciona con un menor comportamiento variado en el escenario de compra, mientras que los hogares que viven en las áreas 4 y 5 -zona centro-norte de la Península y Galicia- tienden a ofrecer una mayor variación, siendo este último efecto mucho más débil. Algunos motivos para justificar este hecho pueden ser la existencia de una menor oferta comercial en las zonas del sur, un menor número de alternativas percibidas de elección por los compradores y/o una diferente estructura de la población, con características demográficas particulares que favorezcan un comportamiento de compra proclive a implementar menos variación en la compra de productos de uso frecuente. Una revisión de la ratio número de establecimientos

comerciales por habitante, que proporciona la Dirección General de Comercio (año 2003) para las zonas en cuestión, no nos permite aceptar la primera posibilidad. Puesto que no existen diferencias claras entre las zonas en este sentido, resulta de interés realizar una investigación más profunda sobre la heterogeneidad que puede estar recogiendo esta variable, en cuanto a alternativas de oferta percibidas por los consumidores en las distintas localizaciones consideradas y en cuanto a rasgos de los hogares que puedan ser discriminatorios en el contexto de compra que contempla este trabajo. Explorar la primera opción necesita un estudio *ad-hoc* que recoja datos primarios. Para explorar la segunda, puede ser de utilidad el uso de una versión jerárquica del modelo Tobit (3.1), en la que los coeficientes de regresión de las variables cuantitativas dependan de los valores tomados por las variables cualitativas de la base.

Así mismo, sería interesante ver hasta qué punto las relaciones que hemos comprobado se mantienen si el problema se plantea de manera dinámica, esto es, contemplando la evolución del escenario de compra del hogar a lo largo del tiempo, para lo cual sería necesario considerar el historial completo de compra del hogar. Todas estas reflexiones e ideas configuran nuestra propuesta de investigación futura en este campo de estudio.

REFERENCIAS

- AITKIN, M. (1991). «Posterior Bayes Factors (with discussion)», *Journal of the Royal Statistical Society Series B*, 53, 111-142.
- ARABMAZAR, A. Y SCHMIDT, P. (1982). «An investigation of the robustness of the Tobit estimator to non-normality». *Econometrica*, 50, 1055-1063.
- BERA, A.K.; JARQUE, C. M. Y LEE, L.F. (1984). «Testing the normality assumption in limited dependent variable models». *International Economic Review*, 25, 563-578.
- BERLYNE, D.E. (1960). «Conflict, Arousal and Curiosity», New York: McGraw-Hill.
- BERNÉ, C.; MOLLÁ, A. Y MÚGICA, J. M. (2004). «The Influence of Purchase Frequency on Varied Consumption», 33rd *European Marketing Academy Conference*, Murcia, Mayo 2004.
- BERNÉ, C.; MÚGICA, J. M. Y YAGÜE, M. J. (2001). «The Effect of Variety-Seeking on Customer Retention in Services», *Journal of Retailing and Consumer Services*, 8, 335-345.

- BERNÉ, C.; MÚGICA, J. M. Y RIVERA, P. (2005). «Managerial Ability to Control the Varied Behavior of Regular Customers in Retailing: Interformat Differences», *Journal of Retailing and Consumer Services*, 12, 151-164.
- BURNETT, P. (1977). «Tests of a Linear Learning Model of Destination Choice: Applications to Shopping Travel by Heterogeneous Population Groups», *Geografiska Annaler. Series B, Human Geography*, 59 (2), 95-108.
- CHINTAGUNTA, P. K. Y GUPTA, S. (1994). «On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models», *Journal of Marketing Research*, 31, February, 128-139.
- DEATON, A.; RUIZ-CASTILLO, J. Y THOMAS, D. (1989). «The Influence of Household Composition on Household Expenditure Patterns: Theory and Spanish Evidence» *Journal of Political Economy*, 97 (1), 179-200
- DRAPER, D.; KRNJAJIĆ, M. Y KOTTAS, T. (2006). «Bayesian Model Specification». Valencia/ISBA Eighth World Meeting on Bayesian Statistics. June 2006.
- EAST, R.; HAMMOND, K.; HARRIS, P. Y LOMAX, W. (2000). «First-Store Loyalty and Retention», *Journal of Marketing Management*, 16 (4), 307-325.
- EAST, R.; HARRIS, P.; WILLSON, G. Y LOMAX, W. (1995). «Loyalty to Supermarkets», *The International Review of Retail, Distribution and Consumer Research*, 5 (1), 99-109.
- FOX, E. J.; MONTGOMERY, A. L. Y LODISH, L. (2004). «Consumer Shopping and Spending across Retail Formats», *Journal of Business*, 77 (2), S25-S60.
- GELMAN, A.; CARLIN, J.B.; STERN, H.S. Y RUBIN, D.B. (2004). «Bayesian Data Analysis (Second Editio)». Texts in Statistical Science. Chapman&Hall.
- GILLY, M.C. Y ENIS, B.M. (1982). «Recycling the Family Life Cycle: A Proposal for Redefinitio». *Advances in Consumer Research*, 9, 271-276.
- GREENE, W. (2004). «Fixed effects and bias due to the Incidental Parameters problem in the Tobit model». *Econometric Reviews*, 23 (2), 125-147.
- HOLDEN, D. (2004). «Testing the Normality Assumption in the Tobit Model». *Journal of Applied Statistics*, 31 (5), 521-532.
- KAHN, B.E. (1998). «Dynamic Relationships with Customers: High-Variety Strategies», *Journal of the Academy of Marketing Science*, 26 (1), 45-53.

- KAHN, B.E., KALWANI, M.U. Y MORRISON, D.G. (1988). «Niching versus Change of Pace Brands: Using Purchase Frequencies and Penetration Rates to Infer Brand Positionings», *Journal of Marketing Research*, 25, November, 384-390.
- KAHN, B. E. Y MCALISTER, L. (1997). *Grocery Revolution: «The New Focus on the Consumer»*, Reading, MA: Addison Wesley, Longman.
- KAHN, B.E. Y RAJU, J.S. (1991). «Effects of Price Promotions on Variety-Seeking and Reinforcement Behavior», *Marketing Science*, 10 (4), 316-337.
- KIM, J., ALLENBY, G. M., Y ROSSI, P. (2002). «Modeling Consumer Demand for Variety», *Marketing Science*, 21 (3), 229-235.
- LANCASTER, T. (2004). «An Introduction to Modern Bayesian Econometrics». Blackwell Publishing.
- MÄGI, A. W. (2003). «Share of Wallet in Retailing: The Effects of Customer Satisfaction, Loyalty Cards and Shopper Characteristics», *Journal of Retailing*, 79 (2), 97-106.
- MARTOS, M. (2004). «Brand-Loyalty Versus Category-Loyalty Using Discrete Choice Models», 33rd *European Marketing Academy Conference*, Murcia, Mayo 2004.
- MCALISTER, L. Y PESSEMIER, E. (1982). «Variety Seeking Behavior: An Interdisciplinary Review», *Journal of Consumer Research*, 9, December, pp. 312-322.
- MCGOLDRICK, P. J. Y ANDRE, E. (1997). «Consumer Misbehaviour: Promiscuity or Loyalty in Grocery Shopping» *Journal of Retailing and Consumer Services*, 4 (2), 73-81.
- NICKOLS, S. Y. Y FOX, K. D. (1983). «Buying Time and Saving Time: Strategies for Managing Household Production», *Journal of Consumer Research*, 10 (September), 197-208.
- PARK, C. W.; MOORMAN, C. Y KOHLI, R. (1991). «A Strategic Framework for Predicting and Managing Consumer Loyalty to Need-Satisfying Alternatives», *Working Paper*, University of Wisconsin.
- POPKOWSKI-LESZCZYC, P. T. L. Y TIMMERMANS, H. J. P. (1997). «Store Switching Behavior», *Marketing Letters*, 8 (2), 193-204.
- RHEE, H. Y BELL, D. R. (2002). «The Inter-store Mobility of Supermarket Shoppers», *Journal of Retailing*, 78 (4), 225-237.
- ROBERT, C. Y CASELLA, G. (2004). «Monte Carlo Statistical Methods» (2nd Ed.). *Springer-Verlag*.

- ROSSI, P.E.; ALLENBY, G.M. Y MCCULLOCH, R. (2005). «Bayesian Statistics and Marketing». *Wiley Series in Probability and Statistics*.
- SEETHARAMAN, P. B. Y CHINTAGUNTA, P. (1998). «A Model of Inertia and Variety-Seeking with Marketing Variables», *International Journal of Research in Marketing*, 15, 1-17.
- SOBERON-FERRER, H. Y DARDIS, R. (1991). «Determinants of Household Expenditures for Services», *Journal of Consumer Research*, 17 (4), 385-397.
- STEENKAMP, J.E.M. Y BAUMGARTNER, H. (1992). «The Role of Optimum Stimulation Level in Exploratory Consumer Behavior», *Journal of Consumer Research*, 19, 434-448.
- THEIL, H. (1967). «Economics and Information Theory», Amsterdam: North-Holland.
- VAN TRIJP, H.C.M. (1995). «Variety Seeking in Product Choice Behaviour: Theory with Applications in the Food Domain», *Mansholt Series, N° 1, Wageningen*, The Netherlands: Wageningen University.
- VAN TRIJP, H.C.M.; HOYER, W.D. E INMANN, J.J. (1996). «Why Switch?. Product Category-Level Explanations for True Variety Seeking Behavior», *Journal of Marketing Research*, 33, August, 281-292.
- WELLS, W. C. Y GUBAR, G. (1966). «The Life Cycle Concept in Marketing Research», *Journal of Marketing Research*, 3, November, 355-363.
- WILKES, R. E. (1995). «Household Life-Cycle Stages, Transitions and Product Expenditures», *Journal of Consumer Research*, 22, 27-42.

ANALYSIS OF SHOPPING VARIED BEHAVIOR OF SPANISH HOUSEHOLDS: A BAYESIAN APPROACH

ABSTRACT

This paper analyses the composition of the scenario of purchase of Spanish households using an AC Nielsen panel database on consumption data. To that end a Bayesian comparison of a Tobit model and a model of Separate Decisions is made. The obtained results confirm the influence of economic, demographic and behavioural variables of the consumption unit on the quantity of variation of the purchase scenario. The comparison of the two models solves that the best fit is provided by the application of the Tobit model, being proved that is not necessary a different analytic treatment for those households with a scenario of null variation.

Key words: Consumer Behaviour, Bayesian Inference, MCMC Methods, Tobit Regression, Separate Decisions, Models Comparison.

AMS Classification: 62F15, 62J12, 91B42.