

UN ANÁLISIS BAYESIANO DE LA VARIACIÓN TEMPORAL DEL ESCENARIO DE COMPRA DE LOS HOGARES*

Carmen Berné, Pilar Gargallo, Noemí Martínez y Manuel Salvador**

WP-EC 2008-05

Correspondencia a: N. Martínez, Universidad de Zaragoza, Departamento de Economía y Dirección de Empresas, Facultad de Ciencias Económicas y Empresariales, C/ Gran Vía 2, 50005 Zaragoza, Teléfono: +34 976761000 – Ext. 4668, Fax: +34 976761767, e-mail: noemar@unizar.es

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.

Primera Edición Mayo 2008

Depósito Legal: V-2462-2008

Los documentos de trabajo del Ivie ofrecen un avance de los resultados de las investigaciones económicas en curso, con objeto de generar un proceso de discusión previo a su remisión a las revistas científicas.

* Este trabajo se ha realizado en el marco del Grupo de Excelencia “CREVALOR”. Los autores desean agradecer la ayuda recibida a través del Proyecto de Investigación del Ministerio de Educación y Ciencia SEJ2005-06603/ECON y del Proyecto PM 090/2006 del Departamento de Ciencia, Tecnología y Universidad del Gobierno de Aragón.

** N. Martínez, C. Berné, P. Gargallo, M. Salvador: Universidad de Zaragoza;

UN ANÁLISIS BAYESIANO DE LA VARIACIÓN TEMPORAL DEL ESCENARIO DE COMPRA DE LOS HOGARES

Carmen Berné, Pilar Gargallo, Noemí Martínez y Manuel Salvador

RESUMEN

El objetivo de esta investigación es analizar los determinantes del comportamiento variado de los hogares en la selección de su escenario de compra habitual a lo largo de un período de tiempo (variación temporal). Para ello, se plantea un modelo bayesiano de decisiones separadas que permite, en particular, analizar si existen diferencias entre las familias con y sin comportamiento variado de compra. Utilizando datos de AC Nielsen, los resultados obtenidos muestran que la frecuencia de compra, el volumen de gasto mensual *per cápita* del hogar, la edad del responsable de la compra, la clase social a la que pertenece la familia, el tamaño del hogar, el área geográfica y el tamaño de la población en la que habitan sus miembros, ejercen una influencia significativa en la variación dinámica de dicho escenario. Dichos resultados permiten ofrecer implicaciones directas para la gestión minorista y para la investigación.

Palabras clave: Análisis Bayesiano, Comportamiento del Consumidor, Elección de Establecimiento, Modelo de Decisiones Separadas, Variación Temporal.

ABSTRACT

This paper aims to analyse the underlying factors of varied behaviour shown in the store set of purchase of households during a time period (temporal variation). In order to examine whether there are differences between households with and without shopping varied behaviour, a Bayesian Separate Decisions Model has been used. By using data from AC Nielsen, results show that shopping frequency, monthly expenditure volume per capita, age of the person responsible for purchasing of grocery and household products, social status, household size, geographical area and population size, have a significant influence on the dynamic variation shown by the store set of purchase. The results reported here allow us to provide guidelines for both researchers and practitioners.

Keywords: Bayesian Analysis, Consumer Behaviour, Store Choice, Separate Decisions Model, Temporal Variation

JEL Classification: C11; C33; D12; D91; M31; P36

1. Introducción

Durante las tres últimas décadas, el comportamiento variado de los consumidores ha atraído la atención de los investigadores interesados en el estudio del comportamiento del consumidor. El comportamiento variado está dirigido a la obtención de estímulos en el comportamiento de compra mediante la alternancia entre objetos de elección, y con la intención de obtener un cambio de ritmo en las actuaciones. De ahí que, por ejemplo, una situación de aburrimiento ocasionada por un nivel no óptimo de estimulación derivado de un comportamiento de compra, puede provocar que el individuo complique su proceso de compra con un comportamiento variado (Howard y Sheth, 1969).

Ahora bien, la literatura especializada sobre comportamiento del consumidor en el mercado minorista se encuentra sesgada hacia el estudio del comportamiento de lealtad del consumidor, obviando en cierto modo su comportamiento variado, que no es un comportamiento opuesto sino una realidad que se va imponiendo en los mercados y cuyo conocimiento puede complementar las estrategias de fidelización de los minoristas.

Centrando la cuestión en las decisiones de elección de establecimiento donde resolver la compra de productos de gran consumo, la situación actual es que, cuando existen alternativas disponibles, los hogares alternan y/o complementan la compra en su establecimiento principal con compras en otros establecimientos (Kahn y McAlister, 1997; Rhee y Bell, 2002; Segarra, 2007). Por lo tanto, en mercados muy competitivos como los de bienes de gran consumo, este comportamiento variado o de compra cruzada es cada vez más acusado (Kahn y McAlister, 1997; McGoldrick y André, 1997). En dichos mercados, los hogares seleccionan un conjunto de establecimientos como los habituales para resolver sus necesidades de compra.

Dicho conjunto de establecimientos recibe el nombre de “*escenario de compra*” del consumidor y su selección incluye la elección de un determinado nivel de variación, el cual debe ser analizado y gestionado en la medida en que cuanto mayor sea, más relevante será la parte del presupuesto total no destinada al gasto en el establecimiento principal en el que se realizan la mayor parte de las compras de la unidad de consumo (Rhee y Bell, 2002).

El conocimiento de los determinantes de la variación en el consumo manifestado en el escenario de compra, puede ser de gran utilidad a las empresas minoristas que operan en el sector para diseñar nuevas guías de actuación (multiplicidad de formatos comerciales, programas de fidelización, etc.), destinadas al mantenimiento de su posición en el mercado.

En la literatura se han analizado las relaciones existentes entre las motivaciones intrínsecas y extrínsecas del cambio y el comportamiento variado en el consumo (Berné *et al.*, 2001 y 2005), pero se ha prestado una menor atención al efecto que pueden tener las características de la unidad de consumo sobre su variación en el comportamiento de compra, siendo además los resultados obtenidos no concluyentes.

A la hora de estudiar el comportamiento variado de compra de los consumidores existen dos formas de abordar el problema: una estructural o estática y otra temporal o dinámica (Pessemier, 1985). Esta variación es susceptible de aumentar con el tiempo (Kahn, 1995). En la aproximación estática se analiza el grado de variación de compra en un determinado momento de tiempo (ver, por ejemplo, Berné *et al.*, 2008), mientras que en la dinámica (i.e., Fox *et al.*, 2004) se observa la evolución del comportamiento variado tras una secuencia de elecciones.

El presente trabajo utiliza una aproximación dinámica, analizando el efecto ejercido por características socio-económicas, demográficas y geográficas de una amplia muestra de hogares procedentes de la base de datos AC Nielsen sobre la evolución de su escenario de compra a lo largo de un año. Para ello, se estima un modelo bayesiano de decisiones separadas (Lancaster, 2004) que permite analizar el problema desde un marco más flexible y realista.

El análisis realizado se sitúa en un contexto similar al de Fox *et al.* (2004). Dichos autores analizan el impacto de características demográficas familiares así como diversos aspectos comerciales del establecimiento (políticas promocionales, variedad de oferta, nivel de precios) sobre la decisión de elección de formatos comerciales planteando un modelo Tobit multivariante jerárquico bayesiano que toma como variable dependiente el logaritmo de los gastos mensuales de una familia en un conjunto de establecimientos. Sin embargo, a diferencia de dichos autores, nuestro estudio analiza el grado de comportamiento variado utilizando un modelo de decisiones separadas que permite analizar, en particular, la existencia de comportamientos diferenciados entre las familias que tienen un comportamiento variado de compra y las que no. Nuestro estudio

incorpora, además, información acerca de la localización geográfica de las familias la cual ejerce una influencia relevante en el comportamiento variado de las mismas.

El trabajo se organiza como sigue: en la sección 2 se realiza una revisión de la literatura y se formulan las hipótesis de trabajo; en la sección 3 se describen los datos, las variables y el modelo; en la sección 4 se muestran los resultados obtenidos; en la sección 5 se muestran las conclusiones más relevantes del estudio así las limitaciones del estudio que nos permitirán perfilar las futuras líneas de investigación y, finalmente, en la sección 6 se muestran una serie de implicaciones para la gestión. Se incluye, además, un apéndice en el que se describe con más detalle la metodología estadística utilizada en el trabajo.

2. Planteamientos de hipótesis

En esta sección se realiza una revisión de las variables condicionantes de la variación temporal mensual del escenario de compra de los hogares, y se describen las variables utilizadas en el estudio, formulándose las hipótesis de trabajo a contrastar. Distinguimos entre cuatro tipos de variables según el tipo de características que miden, a saber: características demográficas del hogar, características socio-económicas del responsable de compra, características del comportamiento de compra del hogar y características geográficas de la unidad de consumo.

2.1. *Características demográficas del hogar*

Las variables demográficas del hogar que han sido utilizadas en estudios previos para explicar su comportamiento de compra han sido su clase social y su tamaño.

Así, Seetharaman y Chintagunta (1998) y McGoldrick y André (1997) encontraron que aquellos hogares con un mayor nivel de ingresos manifiestan una mayor lealtad al establecimiento principal de compra, lo cual conduce más probablemente a una disminución del número total de establecimientos en el escenario que a un aumento de los mismos. Esto es, a un escenario menos variado tanto en su estructura como en el tiempo. La idea subyacente es que pertenecer a una clase social más afortunada implica una menor planificación en las decisiones de compra, de manera que cabe esperar que un mayor nivel de ingresos condicione a la baja la variación del escenario.

Respecto al tamaño de la unidad familiar, sí se admite en la literatura que pueda ejercer un efecto sobre el comportamiento variado pero no existe unanimidad sobre el signo de dicho efecto. Así, mientras que, por un lado, Mägi (2003) sostiene que los hogares más grandes pueden tener más restricciones de tiempo y una mayor tendencia a concentrar sus compras en un único establecimiento, por el otro, Fox *et al.* (2004) argumentan que los hogares más grandes tenderán a mayores gastos porque tienen más miembros, así como una mayor diferenciación en los gustos, lo cual puede manifestarse en una mayor propensión a aumentar el grado de variación en la elección del escenario de compra por parte del responsable de la misma, incrementando el número de establecimientos y/o de formatos y/o distribuyendo de manera más desigual el gasto entre los mismos.

En consecuencia, se plantean las siguientes hipótesis:

H1. “Cuanto mayor sea la clase social a la que pertenece el hogar, menor será la variación del escenario de compra”.

H2. “Cuanto mayor sea el tamaño de la unidad de consumo, mayor será la variación temporal del escenario de compra”.

2.2. Variables demográficas del responsable de la compra

En la literatura especializada, los rasgos personales del responsable de la compra que se han considerado como posibles determinantes del grado de variación del escenario, han sido: su edad y su situación laboral.

En este sentido, diversos estudios han demostrado que la edad del responsable de la compra se encuentra positivamente relacionada con un comportamiento de reparto del gasto en diversos establecimientos (East *et al.*, 1995; East *et al.*, 2000; Mägi, 2003). Una explicación de esta relación es que los consumidores de mayor edad, jubilados en su caso más extremo, disponen de más tiempo libre, por lo que son capaces de dedicar más tiempo a la compra cotidiana y a la comparación entre ofertas, y de utilizar diversos establecimientos simultáneamente para resolver su proceso de compra (East *et al.*, 2000). Puesto que un mayor comportamiento de compra cruzada supone una mayor variación del escenario de compra, se espera una relación positiva entre la edad del responsable de la compra y la variación presente en el escenario de compra.

Además, aquellos hogares en los que la persona encargada de realizar la compra de productos de uso frecuente está trabajando a jornada completa (Mason, 1991), el tamaño de la unidad familiar es grande y su edad se encuentra comprendida en el segmento de 25 a 44 años, son más propensos a tener un comportamiento leal, ya que sus compromisos familiares y sus restricciones de horario son mayores (East *et al.*, 1997). Aquellas personas que disponen de menos tiempo libre concentrarán sus compras en un número reducido de establecimientos a lo largo del tiempo, con el fin de simplificar al máximo el escenario de compra e invertir menos tiempo y esfuerzo en realizar la compra de productos de uso frecuente (McGoldrick y André, 1997; East *et al.*, 1997).

Un mayor volumen de necesidades de consumo, menos tiempo y más compromisos pueden favorecer un comportamiento de compra menos variado. En particular, aquellos hogares que presenten mayores restricciones de tiempo derivados de la presencia de niños en los mismos, es muy probable que concentren el gasto en alimentación y productos para el hogar en un número más pequeño de establecimientos, esperando, por tanto, que exista menos variación en su escenario de compra. Así, cuanto mayor sea el coste de oportunidad del tiempo que soporta el responsable de la compra, es previsible que la variación del escenario de compra sea menor.

Consecuentemente, se plantean las siguientes hipótesis de trabajo:

H3. “Cuanto mayor sea la edad del responsable de la compra, mayor será la variación temporal del escenario de compra”.

H4. “Si el responsable de la compra trabaja fuera del hogar, menor será la variación temporal del escenario de compra”.

H5. “Si el responsable de la compra tiene hijos menores de 12 años, menor será la variación temporal del escenario de compra”.

2.3. Variables de comportamiento de compra

El comportamiento de compra propio del hogar se ha abordado en la literatura mediante el volumen total de la compra y la frecuencia de compra.

En lo que respecta a la relación entre el volumen de la compra realizado por la familia y su comportamiento de compra, Mägi (2003) señala que cuanto mayor sea el gasto total realizado por el hogar, mayores serán los beneficios derivados de comprar a

mejores precios. En nuestro contexto, entendemos que este hecho implicaría un mayor esfuerzo en búsqueda de precios y comparación, lo que conduciría, de contar con el tiempo necesario, a un escenario de compra con un mayor número de establecimientos diferentes e incluso con intercambios continuados de posición de los mismos en gasto relativo, y con entradas y salidas del conjunto reflejando los intentos de aprovechar las distintas ofertas. Esto es, un escenario sujeto a una mayor variación estructural y temporal. Sin embargo, por otro lado, este autor admite que un mayor volumen de compra puede estar relacionado con un mayor tamaño de la unidad familiar y, en consecuencia, parece lógico pensar que hogares más grandes dispongan de menos tiempo y tiendan a concentrar sus compras en un único establecimiento, de manera que el escenario de compra habitual elegido resultaría menos variado. El volumen de compra estará, por lo tanto, muy ligado al tamaño del hogar. Por dicha razón y para evitar posibles problemas de multicolinealidad entre ambas variables se ha optado por utilizar como variable explicativa de la variación presente en el escenario de compra de la familia, el gasto *per cápita*.

La frecuencia de compra ha sido otra de las variables consideradas en las investigaciones anteriores como determinante del comportamiento variado de consumo. Así, por ejemplo, Givon (1984) apunta que cuanto mayor sea la frecuencia de compra, más importante es la necesidad de variedad para el individuo. Así mismo, en un contexto de elección de variedad en la selección de productos (categoría de postres), Berné *et al.* (2004) encuentran que la frecuencia de compra tiene un efecto significativo y positivo sobre el comportamiento variado de consumo. Una razón es que cuanto más intensamente encare una familia el consumo, más rápidamente se sentirá saciada y cobrará más importancia la necesidad de variedad provocada por aburrimiento (Park *et al.*, 1991).

En consecuencia, se plantean las siguientes hipótesis sobre la relación entre el comportamiento variado de consumo y el comportamiento de compra propio del hogar:

H6. “Cuanto mayor sea el volumen total de la compra per cápita, menor será la variación temporal del escenario de compra”.

H7. “Cuanto mayor sea la frecuencia de compra, mayor será la variación temporal del escenario de compra”.

2.4. *Características geográficas*

Otros aspectos potencialmente relevantes en la variación del escenario de compra son el área geográfica y el tamaño de la población en los que residen los hogares.

Las áreas geográficas se delimitan desde una unidad de localización determinada (un continente, un país, un estado federal, una comunidad, etc.) siguiendo criterios geográficos, históricos o políticos, entre otros. La localización geográfica de residencia ejerce un efecto significativo sobre el reparto del gasto en servicios de los hogares (Soberon-Ferrer y Dardis, 1991). Desde un punto de vista comercial, las áreas geográficas son consideradas por las empresas para planificar sus rutas comerciales (en las que se toman distintas decisiones en ventas y en distribución de productos).

La partición territorial del mercado puede recoger parte de la heterogeneidad existente en el mismo desde el lado de la oferta (en número de establecimientos, nivel de rivalidad del entorno competitivo, etc.), así como parte de la heterogeneidad de los hogares (en este caso de nuestra muestra) desde el lado de la demanda (diferencias sociológicas, demográficas o económicas). De manera que incluiremos esta variable en el modelo como una variable de control de los efectos de esta heterogeneidad.

La información sobre resultados previos a este respecto es escasa. Sí sabemos que el gasto de los hogares españoles se incrementa en diciembre con la época navideña. Según Resa (2007), este gasto no sólo se reparte según segmentos y canales, sino también por áreas geográficas, de manera que su incremento es mayor en las áreas noroeste y centro de la península (Galicia, Asturias y Cantabria, y Madrid) y menor en el área nordeste (Cataluña y más concretamente, Barcelona).

Con todo, defendemos que vivir en una determinada área geográfica puede condicionar el nivel de variación en el comportamiento de compra y esto, a su vez, implicar diferencias entre áreas y grado de variación dinámica del escenario de compra.

Además, la oferta comercial de la zona geográfica diferirá en función de cuál sea el tamaño de la población. En este sentido, cobra especial relevancia la consideración no sólo de la oferta comercial realmente existente, sino también la percibida –alternativas percibidas de compra– por parte de las personas responsables de la compra en el hogar (Arrondo, 1999; Berné *et al.*, 2005; Berné y Martínez, 2008).

En cualquier caso, el tamaño de la población en la que habite el hogar estará directamente relacionado con el volumen de la oferta disponible y, por lo tanto, *ceteris paribus*, mayor será la variación esperada del escenario de compra. Si no se producen descensos demográficos, la variación de dicho escenario es de esperar se sostendrá o se incrementará en el tiempo. Así mismo, el tamaño de la población en la que habite el hogar estará directamente relacionado con el volumen de la oferta disponible y, por lo tanto, *ceteris paribus*, mayor será la variación esperada del escenario de compra (Berné *et al.*, 2008). En caso de que no hayan de soportarse descensos demográficos, la variación de dicho escenario se sostendrá o se incrementará en el tiempo.

En consecuencia, se plantean las siguientes hipótesis de trabajo:

H8. “El grado de variación temporal del escenario de compra diferirá entre áreas geográficas”.

H9. “Cuanto menor sea el tamaño de la población en la que reside la familia, menor será la variación del escenario de compra en el tiempo”.

3. Metodología

3.1. Base de datos y medición de las variables

La investigación empírica se aborda utilizando la información contenida en una base de datos construida a partir de un panel de consumidores de AC Nielsen España. Se trata de información sobre la compra realizada por 2017 hogares españoles (desde Abril de 2003 hasta Abril de 2004). Concretamente, se dispone de la siguiente información: lugar y fecha de compra, volumen total de la cesta de compra, el área Nielsen ¹ y el tamaño de la población en la que reside, la clase social a la que pertenece

¹ AC Nielsen divide la zona peninsular de España en 8 zonas conocidas como “áreas Nielsen”. El área Nielsen 1 se encuentra integrada por: Barcelona (excluida su área metropolitana), Gerona, Lérida, Tarragona, Huesca y Zaragoza; el área Nielsen 2 está compuesta por Castellón, Valencia, Alicante, Murcia y Albacete; Badajoz, Córdoba, Jaén, 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, Salamanca y Madrid (excluida su área metropolitana); La Coruña, Lugo, Orense, Pontevedra, León y Asturias forman parte del área 5, mientras que el área Nielsen 6 está constituida por las provincias españolas de Cantabria, Vizcaya, Guipúzcoa,

el hogar y el tamaño del mismo, así como, la edad y la situación laboral de la persona responsable de la compra.

Explicamos, a continuación, la medición de la que es nuestra variable principal de estudio, el comportamiento variado, y después la del resto de variables consideradas en el análisis.

La medición del comportamiento variado mensual en el escenario de compra se aborda a través de la observación de la permanencia o variación de la composición del mismo teniendo en cuenta el porcentaje de gasto relativo realizado en cada uno de los establecimientos que forman parte del escenario de compra de los hogares.

Para medir la variación presente en el escenario de compra utilizamos el Índice de Hirschman-Herfindahl (IHH), propuesto originalmente como una medida de concentración industrial en Theil (1967) y utilizado después por Theil y Finke (1983), Meulenbergh, (1989) y Van Trijp (1995) para cuantificar la variación en el consumo. Dicho índice, para nuestro caso concreto –índice de Hirschman-Herfindahl de comportamiento variado– responde a la siguiente expresión:

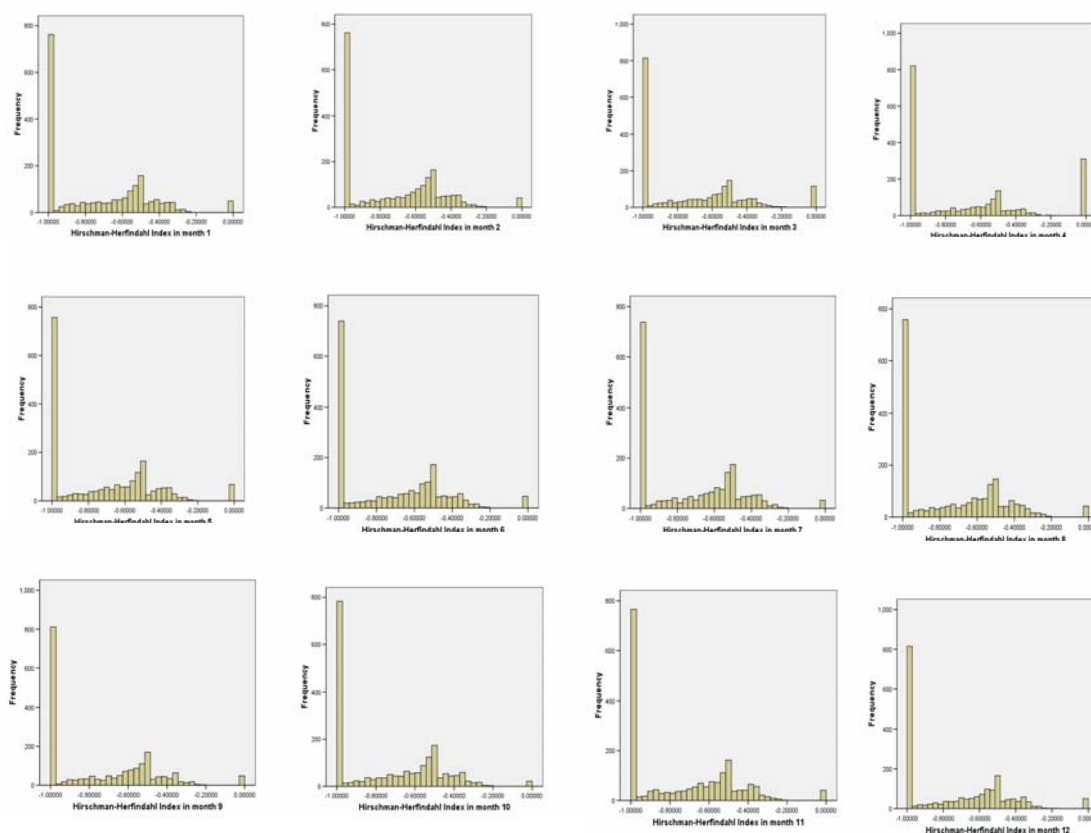
$$IHH_t = - \sum_{j=1}^m p_{j,t}^2$$

donde: $p_{j,t}$ es el porcentaje de gasto realizado en el establecimiento j -ésimo en el mes t ; $j = 1, \dots, m$; $t = 1, \dots, 12$ siendo m el número total de establecimientos que forman parte del escenario de compra.

El número de establecimientos de compra considerados en el trabajo es $m = 11$ por lo que dicho índice se encontrará comprendido entre -1 , cuando la familia no tenga un comportamiento variado y compre únicamente en un establecimiento ($p_{1,t} = 1, p_{2,t} = \dots = p_{m,t} = 0$) y $-1/m = -0.09$, que es el valor que tomará cuando el porcentaje de compra en cada uno de los establecimientos sea el mismo ($p_{1,t} = \dots = p_{m,t} = 1/m$) y, por lo tanto, el grado de variación será máximo.

Navarra, Burgos, Álava, La Rioja y Palencia; el Área Metropolitana de Barcelona constituye el área Nielsen 7 y, finalmente, el Área Metropolitana de Madrid se corresponde con el área Nielsen 8.

FIGURA 1. Evolución mensual del índice IHH



En la Figura 1 se muestra la evolución mensual de la distribución del índice IHH a lo largo del periodo comprendido entre Abril de 2003 y Abril de 2004. Se observa, en primer lugar, que en cada mes existe un grupo de familias (en torno a un 37%) cuyo grado de comportamiento variado es mínimo, es decir, que compra en un único establecimiento. Así mismo se observa la presencia de un pequeño grupo de familias para las cuales $IHH_t = 0$, es decir, que no compran dicho mes. Ello se debe a que en ese mes dichas familias están de vacaciones y, por lo tanto, no han comprado en los establecimientos de compra habituales. Para analizar si este hecho tiene alguna influencia sobre el comportamiento variado de las familias se han incluido un conjunto de indicadores $I_t = I(IHH_t=0)$ que toman un valor igual a 1 si en el mes t una familia dada no ha realizado compra alguna en los establecimientos considerados y 0 en otro caso.

Este comportamiento del índice IHH provoca que, para describir su evolución a lo largo del tiempo, tengamos que utilizar modelos de regresión con variables

dependientes de tipo mixto. En la literatura uno de los más utilizados es el modelo de decisiones separadas (Lancaster, 2004) que proporciona una gran flexibilidad a la hora de capturar diferentes clases de efectos de las variables explicativas sobre la variable independiente.

Con respecto a las variables independientes del modelo, la variable clase social distingue si se trata de una familia perteneciente a la clase social alta, media o baja.

El tamaño del hogar toma el valor 1 si se trata de un hogar unipersonal, 2 si es bi-personal y así sucesivamente.

Para medir el volumen total de la cesta de la compra de la unidad de consumo, consideramos el gasto total *per cápita* anual realizado por la familia y se toman logaritmos con el fin de aumentar el grado de normalidad de dicha variable.

La frecuencia de compra refleja el número de veces –número de viajes– que ha acudido la familia a realizar la compra en todo el año.

El área geográfica es el área Nielsen a la que pertenece el hogar. En España, una opción de reparto del territorio en áreas comerciales es la que proporcionan las denominadas áreas Nielsen. Desde hace cuarenta años, la consultora AC Nielsen, analiza la oferta y la demanda comercial en España desde un reparto de la península en zonas geográficas que presentan una agrupación no coincidente con la de las comunidades autónomas. El criterio original de distribución atendió a un reparto en cuatro cuadrantes del mapa peninsular y a los circuitos comerciales entonces utilizados por las empresas. Hoy en día, su reconocimiento se ha generalizado de tal modo que los circuitos comerciales son los que se adaptan a la existencia de estas áreas. La localización de un hogar en una u otra de estas zonas puede estar recogiendo aspectos del hogar que proceden del entorno externo.

Respecto a las características del responsable de la compra, y siguiendo los criterios de segmentación utilizados por AC Nielsen, la edad del responsable de la compra se ha segmentado en tres categorías que se obtienen tomando como puntos de corte las edades de 35 y 54 años.

Por otro lado, la situación laboral del responsable de la compra hace alusión a si el comprador trabaja fuera del hogar (valor 1) o no trabaja fuera del hogar (valor 0).

Finalmente, siguiendo los criterios de segmentación utilizados por AC Nielsen, el tamaño poblacional se ha recodificado tomando los valores 0, 1, 2 y 3 en función de si el hábitat en el que reside la familia es menor o igual a 10.000 habitantes, entre 10.000 y 50.000, entre 50.000 y 200.000 y superior a 200.000, respectivamente.

En las Tablas 1 y 2 se presenta un análisis descriptivo de las características demográficas y de compra de los hogares de la muestra. En concreto, la Tabla 1 muestra los estadísticos descriptivos de carácter longitudinal y la Tabla 2 los descriptivos correspondientes a las variables independientes de carácter transversal. Como puede observarse, el gasto mensual presenta un comportamiento estacionario de 162 € de media, siendo el mes de diciembre en el que se realizó el mayor gasto medio y agosto el mes en el que se realizó el menor gasto medio. Por otro lado, la frecuencia de compra oscila en torno a 4-5 veces al mes, siendo los meses de julio y agosto los de menor frecuencia media de compra y el mes de octubre el de mayor.

TABLA 1. Análisis descriptivo de las variables independientes de carácter longitudinal

Periodo Temporal	Gasto Mensual (€)		Frecuencia de compra		I _t =1
	Media	Desv. Típica	Media	Desv. Típica	Porcentaje
28/04/03 - 28/05/03	159.16	110.95	4.61	3.09	2.43
28/05/03 - 28/06/03	162.37	112.93	4.41	2.90	2.03
28/06/03 - 28/07/03	155.42	120.78	4.21	2.86	5.75
28/07/03 - 28/08/03	136.81	107.08	3.65	2.71	15.37
28/08/03 - 28/09/03	159.50	111.42	4.41	2.91	3.32
28/09/03 - 28/10/03	172.69	124.82	4.73	3.04	2.28
28/10/03 - 28/11/03	168.56	123.14	4.62	2.96	1.59
28/11/03 - 28/12/03	188.60	129.55	4.62	2.92	2.08
28/12/03 - 28/01/04	160.67	123.48	4.41	2.75	2.38
28/01/04 - 28/02/04	161.49	111.96	4.62	3.08	1.09
28/02/04 - 28/03/04	165.74	111.28	4.60	2.98	2.03
28/03/04 - 28/04/04	167.87	125.01	4.44	2.88	2.48

En la Tabla 2 se muestran las variables explicativas de carácter trasversal así como su correspondiente codificación ya que, con el fin de capturar posibles efectos no lineales, se optó por utilizar una codificación tipo indicador tomando como categorías de referencia las indicadas en dicha tabla. El perfil más frecuente corresponde a una familia de clase media (52.6%) que vive en las áreas Nielsen de Este de España (áreas 1, 2 y 3, con un 45.1%), en una población de más de 200.000 habitantes (34.3%), con un número de miembros igual 4 (30.9%) y sin niños menores de 12 años (64.8%). El

responsable de la compra tiene una edad comprendida en el intervalo de entre 35 y 54 años (51.9%) y no trabaja fuera del hogar (81%).

TABLA 2. Análisis descriptivo de las características estáticas del panel de hogares

	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%

	< 35 años *	Entre 35 y 54 años	> 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%

	< 10000*	Entre 10000 y 50000	Entre 50000 y 200000	Más de 200000	Total
Tamaño Población	15.7%	22.5%	27.6%	34.3%	100%

	No*	Sí	Total
Hijos<12 Años	64.8%	35.2%	100%
Trabajo del comprador fuera del hogar	81.0%	19.0%	100%

Nota: Para cada variable categórica la clase de referencia se señala con *

3.2. Los datos

Los datos corresponden a una muestra de N familias (N = 2017) en las que se observa la evolución de su comportamiento variado a lo largo de T periodos de tiempo (T = 12) así como los valores p covariables X_1, \dots, X_p que son las variables explicativas de dicho comportamiento.

Dichos datos vienen, por lo tanto, dados por $D = \{(x'_{it}, y_{it}, IHH_{it}) ; i = 1, \dots, N; t = 1, \dots, T\}$ con $x_{it} = (1, x_{it1}, \dots, x_{itp})'$ los valores de las covariables de la familia i-ésima, IHH_{it} es el índice Hirschman-Herfindahl de comportamiento variado de la familia i-ésima durante el mes t e y_{it} es un indicador de comportamiento variado de la familia en cada periodo de tiempo que viene dado por $y_{it} = 1$ si $IHH_{it} < 1$ (comportamiento variado) e igual a 0 si $IHH_{it} = 1$ (comportamiento no variado).

3.3. Modelo de decisiones separadas

El modelo utilizado es uno de decisiones separadas –que distingue entre la decisión de comportamiento variado de cada familia y el grado de variación presente en dicho comportamiento– viene dado por las expresiones:

$$P(y_{it} = 1) = \Phi(\boldsymbol{\gamma}\mathbf{x}_{it}) \quad (3.1)$$

$$h_{it} = -\log(-IHH_{it}) = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_i \text{ con } \varepsilon_i \sim N(0, \sigma_i^2); i = 1, \dots, N \text{ si } h_{it} > 0 \quad (3.2)$$

donde: Φ es la función de distribución de una normal estándar, $\boldsymbol{\gamma} = (\gamma_0, \dots, \gamma_p)'$ y $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)'$ son los coeficientes de regresión que determinan el tipo de influencia ejercida por las covariables X_1, \dots, X_p sobre la decisión de comportamiento variado de las familias y el grado de dicho comportamiento, respectivamente. Entre las covariables X_1, \dots, X_p incluiremos el indicador I_t de si cada familia no ha comprado en el periodo previo al periodo actual t , tomando como punto de partida $I_{i0} = 0 \forall i$. La introducción de esta variable nos permitirá analizar hasta qué punto haber estado de vacaciones influye significativamente sobre un posible cambio en las pautas habituales de comportamiento variado de una familia.

Observar, en particular, que el modelo permite analizar explícitamente si existen diferencias entre las familias que tienen un comportamiento variado de compra y las que no, mediante el estudio de los valores de los coeficientes $\boldsymbol{\gamma}$. Este tipo de información resulta relevante a la hora de determinar los perfiles de aquellas familias leales a un único establecimiento, la cual puede ser muy útil a los gerentes de las empresas para diseñar políticas de fidelización más adecuadas. Por su parte, los coeficientes $\boldsymbol{\beta}$ miden la influencia ejercida por las variables explicativas sobre el grado de comportamiento variado dentro del grupo de las familias que muestran un comportamiento variado, influencia que no tiene por qué ser del mismo tipo que la influencia ejercida sobre la parte discreta del modelo.

4. Estimación y resultados

4.1. Estimación de los parámetros del modelo

La aproximación metodológica adoptada en la estimación y selección del modelo es bayesiana que, tal y como demuestran los recientes trabajos de Lancaster (2004) y Rossi *et al.* (2005), proporciona un mayor grado de fiabilidad al análisis efectuado al no depender el proceso de estimación de parámetros de resultados asintóticos.

4.2. Resultados

Dado que la metodología estadística es bayesiana, la estimación de los parámetros del modelo propuesto se realiza a partir de la distribución *a posteriori* de los mismos. Para realizar inferencias acerca de los parámetros del modelo hemos utilizado las medianas y cuantiles 2.5% y 97.5% *a posteriori* los cuales proporcionan una estimación puntual y un intervalo de credibilidad bayesiano del 95% para dichos parámetros, respectivamente.

En la Tabla 3 se muestran las estimaciones puntuales de los parámetros γ y β así como los límites de sus intervalos de credibilidad al 95%, obtenidos a partir de las medianas y los cuantiles 2,5% y 97,5% de su distribución *a posteriori*, respectivamente.

En relación a la influencia ejercida por las características demográficas de los hogares, se observa:

- (i) La existencia de diferencias entre los hogares de clase social alta y el resto. Los primeros tienen una mayor probabilidad de comprar en un único establecimiento. Esto es, tienden a presentar un escenario de compra sujeto a menor variación. No se observan sin embargo, diferencias significativas entre los hogares de clase social media y baja; como tampoco se aprecian diferencias significativas en el grado de comportamiento variado entre las distintas clases sociales. Estos resultados se confirman en todos los periodos de tiempo (mes a mes) en los que se ha realizado la estimación. Con estos resultados no podemos rechazar la primera de las hipótesis planteadas (H1).

(ii) El tamaño del hogar ejerce una influencia significativa sobre el grado de comportamiento variado, de forma que, a mayor tamaño, menor tiende a ser la variación del escenario de compra del hogar. Consecuentemente, los resultados obtenidos corroboran las tesis de Mägi (2003) y no podemos rechazar la hipótesis H2.

TABLA 3. Estimaciones de los coeficientes del modelo de decisiones separadas

VARIABLES	Modelo de Decisiones Separadas					
	Modelo Probit (γ)			Modelo Regresión Lineal (β)		
	Q2.5	Mediana	Q97.5	Q2.5	Mediana	Q97.5
Constante	-0,9544	-0,7968	-0,6597	0,6775	0,7541	0,8066
Clase social media	-0,0739	-0,0242	0,0202	-0,0177	0,0044	0,0281
Clase social alta	-0,1514	-0,1016	-0,0515	-0,0229	0,0037	0,0300
Tamaño hogar 2	-0,1562	-0,0724	0,0100	-0,0829	-0,0580	-0,0217
Tamaño hogar 3	-0,0124	0,0678	0,1516	-0,0799	-0,0526	-0,0203
Tamaño hogar 4	-0,0199	0,0682	0,1504	-0,1086	-0,0804	-0,0514
Tamaño hogar > 4	-0,0409	0,0455	0,1335	-0,1326	-0,0989	-0,0597
35 < Edad < 54 años	0,0356	0,0908	0,1516	0,0011	0,0240	0,0449
Edad > 55 años	-0,0282	0,0347	0,0940	-0,0087	0,0136	0,0365
Trabaja fuera del hogar	-0,0608	-0,0110	0,0362	-0,0122	0,0067	0,0249
Miembros hogar < 12 años	-0,0775	-0,0345	0,0082	-0,0029	0,0118	0,0274
Log(gastos)	0,0216	0,0465	0,0718	-0,0633	-0,0538	-0,0409
Frecuencia	0,1763	0,1843	0,1917	0,0041	0,0075	0,0104
Nielsen2	-0,3232	-0,2580	-0,1949	-0,0285	0,0086	0,0368
Nielsen3	-0,1559	-0,0997	-0,0382	-0,0537	-0,0280	-0,0018
Nielsen4	0,0695	0,1426	0,2165	-0,0400	-0,0071	0,0263
Nielsen5	0,1562	0,2209	0,2953	-0,0606	-0,0289	0,0052
Nielsen6	-0,0379	0,0295	0,0999	-0,0732	-0,0334	0,0010
Nielsen7	-0,0129	0,0572	0,1269	-0,0578	-0,0243	0,0125
Nielsen8	-0,0240	0,0475	0,1131	-0,0023	0,0202	0,0503
10000 < Población ≤ 50000	0,1047	0,1622	0,2178	-0,0254	0,0079	0,0341
50000 < Población ≤ 200000	0,0593	0,1135	0,1741	-0,0453	-0,0145	0,0140
Población > 200000	0,0097	0,0655	0,1252	-0,0740	-0,0405	-0,0146
I(IHH = 0)	-0,1429	-0,0431	0,0591	-0,0515	-0,0156	0,0220

Nota: Se señalan en negrita los coeficientes significativos para un nivel de credibilidad del 95%.

Respecto a la influencia ejercida por las características demográficas del responsable de compra, se observa:

(i) Una influencia significativa del tramo de edad en que se encuentre el comprador en la variación del escenario de compra. De este modo, no podemos rechazar la hipótesis H3. Además, el tipo de influencia ejercida tiene forma de “U” invertida $\neg\cap$, de modo que el comportamiento variado de los compradores jóvenes y de los mayores de 53 años es menor que el de las que pertenecen al segmento de edad media. Se corroboran así los resultados obtenidos por Berné *et al.* (1996). Los responsables de la compra denominados jóvenes son los que presentan escenarios menos variados (significatividad al 90%). La edad media sin embargo parece estar sujeta a condicionantes capaces de superar los supuestos mayores costes de oportunidad temporales que soporta este grupo de edad, de manera que la variación relativa de su escenario de compra es mayor.

(ii) Los efectos ejercidos por la situación laboral del responsable de la compra no son significativamente diferentes de cero. Así, disponer de menos tiempo libre no influye significativamente sobre el grado de comportamiento variado del escenario de compra. No podemos por lo tanto, aceptar la hipótesis H4.

(iii) La presencia en el hogar de algún miembro menor de 12 años, parece influir en la manifestación de un comportamiento de compra extremo, ya que aumenta la probabilidad de que estos hogares presenten, bien escenarios de compra sin variación, bien escenarios de compra muy variados. Este efecto no lineal es significativo al 90%.

Las dos variables independientes relacionadas con el comportamiento de compra, volumen y frecuencia de la misma, ejercen una influencia significativa sobre el comportamiento variado:

(i) El volumen total del gasto del hogar *per cápita* ejerce un efecto no monótono, observándose que, si los gastos mensuales son relativamente elevados, es más probable que resuelva la compra en más de un establecimiento. Es decir, es más probable que el hogar pertenezca al grupo de hogares que presentan un escenario de compra variado. Ahora bien, la variación del mismo tiende a disminuir conforme aumenta el volumen de gasto. Estos resultados se constatan mes a mes. En consecuencia, no podemos aceptar la hipótesis H6.

(ii) Cuanto mayor es el número de viajes de compra, más probable es que el hogar presente un escenario de compra variado. Además, este resultado se confirma mes a mes. Por tanto, no podemos rechazar la hipótesis H7.

Finalmente, los resultados sobre el efecto de las variables relativas a la localización de la primera residencia del hogar son los siguientes:

(i) El área geográfica de la península a la que pertenece la primera residencia del hogar, ejerce una influencia significativa sobre su comportamiento variado, lo que permite no rechazar la hipótesis H8. En particular, los hogares que viven en las áreas Nielsen 2 y 3, esto es, en el litoral sur y levantino, tienen menos probabilidades de comprar en más de un establecimiento, mientras que los situados en el Noroeste (áreas Nielsen 4 y 5) tienen más probabilidades de presentar un escenario de compra variado. Además, los que se localizan en el área Nielsen 3 tienden a presentar un escenario de compra menos variado que el resto.

(ii) El tamaño de la población en la que se localiza el hogar también influye significativamente en la variación de su escenario de compra, si bien el efecto ejercido no es monótono. Los que viven en las ciudades con menor número de habitantes tienen una mayor probabilidad de pertenecer al grupo de hogares que no manifiesta un comportamiento variado en su escenario, mientras que los que viven en ciudades con un número intermedio de habitantes tienen menos probabilidades que los hogares que viven en ciudades con mayor demografía. Puesto que es de esperar que un mayor número de habitantes atraiga a un mayor número de establecimientos y una oferta más competitiva, una posible razón de esta relación es que el grado de complejidad incremental de la compra variada en estas ciudades, fundamentalmente debido a los relativamente mayores costes temporales soportados (distancias y compromisos), no llegue a compensar los posibles beneficios derivados de la compra cruzada. Esta idea se corrobora en las estimaciones para el grupo de hogares con escenario variado, ya que su grado de variación tiende a ser menor cuanto mayor es el tamaño de la población en la que se vive. Con todo, no podemos aceptar la hipótesis H9.

Por último, los signos negativos de los coeficientes del indicador $I(IHH_{t-1}=0)$ indican que después de un período (t) sin compra el grado de comportamiento variado del escenario de compra correspondiente tiende a ser menor. Sin embargo, el efecto de dicho indicador no resulta significativo.

A modo de resumen, la Tabla 4 muestra el signo esperado en cada una de las nueve relaciones hipotetizadas, así como los efectos encontrados en los modelos probit (3.1) y de regresión (3.2), que componen el modelo de Decisiones Separadas.

TABLA 4. Signos esperados y contraste de hipótesis

HIPÓTESIS	RELACIÓN HIPOTETIZADA	CONTRASTE	EFFECTO RESULTANTE
<i>H1 (Clase social)</i>	↓	ACEPTADA	Efecto Probit Clase Alta
<i>H2(Tamaño del hogar)</i>	Significativa pero no hipotetizada	ACEPTADA	Efecto Regresión ↓
<i>H3 (Edad)</i>	Significativa pero no hipotetizada	ACEPTADA	Efectos Probit y Regresión ∩
<i>H4 (Responsable de compra trabaja fuera del hogar)</i>	↓	RECHAZADA	No significativo
<i>H5 (Hijos pequeños en el hogar)</i>	↓	RECHAZADA	Efecto Probit ↓ Efecto Regresión ↑ al 90%
<i>H6 (Gasto)</i>	↓	RECHAZADA	Efecto Probit ↑ Efecto Regresión ↓
<i>H7(Frecuencia)</i>	↑	ACEPTADA	Efectos Probit y Regresión ↑
<i>H8(Localización geográfica)</i>	No hipotetizada	ACEPTADA	Efectos Probit y Regresion significativos
<i>H9 (Tamaño de la población)</i>	↑	RECHAZADA	Efecto Probit ∩ Efecto Regresión ↓ en grandes ciudades

Nota : ↑: Efecto monótono no decreciente

↓ : Efecto no creciente

∩: Efecto no lineal en forma de “U” invertida.

4.3. Bondad de ajuste y comparación de modelos

La bondad de ajuste del modelo se analizó aplicando el procedimiento descrito en la sección A.2.3 del apéndice, calculando los intervalos predictivos bayesianos del 95% y el 99% para los valores de $\{h_{it} : (i,t) \in \{1, \dots, N\} \times \{1, \dots, T\}\}$ y evaluando sus cubrimientos empíricos utilizando la expresión (A.8). Dichos cubrimientos son iguales a un 97.42% y 99.57%, respectivamente, los cuales no difieren significativamente de sus niveles nominales.

Así mismo se comparó el modelo de decisiones separadas con un modelo Tobit que supone que la influencia ejercida por las covariables del problema es la misma tanto para determinar si existe o no comportamiento variado como para determinar su grado de variación. Se utilizaron como criterios de comparación la raíz del error cuadrático medio (RMSE), el error absoluto medio (MAD) y el logaritmo de la densidad predictiva a posterior (LPRED) dados por las expresiones (A.11), (A.12) y (A.13) del apéndice.

Los resultados del proceso de comparación se muestran en la Tabla 5 en la que se observa que nuestro modelo se ajusta mejor a los datos en todos los criterios de comparación.

TABLA 5. Comparación de modelos

Criterios/Modelos	Tobit	Decisiones Separadas
RMSE	0.3581	0.3208
MAD	0.2876	0.2790
LPRED	-1605.2	-1460.3

5. Conclusiones

Esta investigación considera, como punto de partida que, en el mercado minorista de alimentación, las unidades de consumo tienen un escenario de compra habitual formado por un conjunto determinado de establecimientos entre los cuales el hogar reparte su presupuesto total. En concreto, en este trabajo se ha realizado un análisis dinámico del comportamiento variado de las familias españolas determinando qué características demográficas, geográficas y de comportamiento de compras de las mismas, de las tradicionalmente consideradas en la literatura, ejercen una influencia significativa sobre el mismo.

Dada la naturaleza mixta de la variable dependiente, se ha utilizado un modelo estadístico bayesiano de decisiones separadas el cual proporciona un marco muy flexible y general para analizar este problema al permitir, por un lado, analizar si existen diferencias entre las familias con y sin comportamiento variado de compra y, por el otro, la captura de efectos no monótonos de las variables explicativas sobre la variable dependiente del modelo. La metodología estadística bayesiana proporciona, además, una mayor fiabilidad a las estimaciones obtenidas al no depender las mismas de resultados asintóticos.

Utilizando la información proporcionada por una base de datos construida a partir de un panel de hogares de AC Nielsen, la cual contiene información del historial de compra de una muestra exhaustiva de familias españolas, se ha encontrado que la frecuencia y el gasto mensual realizado en las compras, las características demográficas de las familias como la edad del responsable de la compra, su clase social y el número

de miembros del hogar así como el área geográfica en la que viven, ejercen una influencia significativa en la selección de dicho escenario. Además, el tipo de influencia ejercido difiere según sea el tipo de escenario elegido (variado o no variado) o, en caso de ser variado, el grado de variación del mismo.

La contribución más relevante del análisis empírico desarrollado es, por tanto, haber comprobado la existencia de algunos antecedentes del comportamiento variado tomando como referencia el modelo actual de compra caracterizado por un escenario de compra habitual que presenta variación mensual. De este modo, se ha iniciado una línea de investigación que es una extensión de la línea general de comportamiento del consumidor y más concretamente, del comportamiento variado de consumo.

Este estudio presenta, sin embargo, sendas limitaciones derivadas del hecho de haber utilizado una fuente secundaria de datos² e información agregada con frecuencia mensual. Un análisis que podría resultar adecuado sería utilizar modelos en tiempo continuo que permitieran analizar el comportamiento de las familias con un nivel de desagregación mayor estudiando, en particular, la influencia del tiempo entre compras sobre las decisiones de compras de las familias. También podría resultar interesante estudiar otros períodos temporales y otras formas de variación en el escenario (cambio de establecimiento/formato/grupo de distribución en el tiempo).

El modelo utilizado supone, además, que el efecto ejercido por las características socio-demográficas de las familias es homogéneo para todas las familias. Esta hipótesis podría ser poco realista y sería interesante debilitarla con el fin de analizar la posible existencia de grupos de comportamiento diferenciados en las familias. Para ello, podrían utilizarse modelos jerárquicos bayesianos que constituyen una herramienta útil para este tipo de problemas.

Finalmente, una mayor profundidad en el análisis de las sinergias que puede conseguir la implementación combinada de estrategias ofensivas, defensivas, estrategias de variedad y estrategias multiformato de las empresas detallistas, desde la línea de investigación aquí iniciada y para otros contextos de consumo, puede conseguir más implicaciones útiles para la gestión del minorista.

² En particular, merece la pena destacar que los puntos de corte utilizados para las variables edad, presencia de niños en el hogar son propios de la base de datos que se está utilizando, lejos de ser un estándar en la literatura de marketing. Lo mismo sucede con el reparto geográfico de la península en Áreas Nielsen.

6. Implicaciones para la gestión

La gestión minorista debe identificar los perfiles de sus clientes para decidir la estrategia más conveniente a la hora de dirigirse a ellos. Puesto que se ha observado que, en un contexto de compra de productos de uso frecuente para el hogar, es mayor el predominio del grupo de hogares que manifiestan un comportamiento variado (63%) respecto al comportamiento de lealtad a un único establecimiento comercial (37%), la identificación de los dos grupos por parte de los minoristas puede ofrecer muchos beneficios para la rentabilidad de los activos minoristas.

Una vez identificados los segmentos dentro de la base de clientes de una compañía, dos estrategias pueden ser utilizadas por el minorista para disminuir el impacto de la competencia existente en el sector. Primero, estrategias defensivas, dirigidas a incrementar la satisfacción y a mantener al segmento de clientes que se muestran leales a un único establecimiento. Segundo, estrategias ofensivas, cuyo objetivo de captación de clientes nuevos (que muestran un comportamiento variado elevado) esté ligado a la localización de nuevas tiendas en el área que tengan formatos diferenciados de los ya existentes y que sean capaces de capturar parte del gasto que no captura el establecimiento principal por restricciones de formato y desviaciones en el surtido.

Para la primera de dichas estrategias se deduce de nuestro estudio que pertenecer a un estatus social más elevado, tener un menor tamaño del hogar, pertenecer a los segmentos de edad de menos de 35 años y superior a 55 años, un volumen de gasto *per cápita* menor y una menor frecuencia –viajes de compra– para resolver la misma, parece ser el perfil de un hogar que muestra un escenario de compra con variación nula (esto es, que manifiestan lealtad a un único establecimiento). Si, además, estos hogares se ubican en las áreas Nielsen 2 y 3 y tienen un tamaño poblacional inferior a 10.000 habitantes, las empresas tienen más posibilidades de éxito al implementar estrategias de fidelización a un único establecimiento comercial.

Por otro lado, los hogares que muestran una mayor variación temporal en el escenario de compra son aquellos hogares uni o bipersonales que tienen un gasto *per cápita* elevado, una baja frecuencia de compra, en el que la persona responsable de la compra tiene una edad comprendida entre 35 y 54 años, y pertenecen a poblaciones de tamaño intermedio (entre 10.000 y 200.000 habitantes). Puesto que los hogares que forman parte del grupo que manifiesta un comportamiento variado temporal reparten su

presupuesto mensual entre diversos establecimientos y, por esa razón, las empresas minoristas pueden encontrar interés en complementar las estrategias defensivas con estrategias multiformato. Así, podría decirse que las estrategias de variedad y estrategias multiformato cobran un especial protagonismo para la gestión minorista pues al ofrecer una amplia gama de formatos comerciales pueden lograr que un determinado cliente permanezca leal a la empresa, aún visitando diversos establecimientos comerciales, debido a que las razones que motivan la elección de los establecimientos del escenario de compra se encuentra motivada por diversos factores.

Entre los dos extremos se situarán aquellos hogares con una combinación de los factores antedichos. Las técnicas de segmentación de mercados tienen en ellos variables potencialmente explicativas capaces de ofrecer información interesante para tomar decisiones más acertadas sobre la(s) estrategia(s) de marketing a seguir en relación a su mezcla de marketing de servicios minoristas.

Referencias bibliográficas

- ARRONDO, E. (1999): *Perfiles de Comportamiento y Gestión de la Satisfacción y Lealtad del Consumidor: Un Análisis Empírico del Comercio Minorista de Alimentación*, Tesis Doctoral, Abril, San Sebastián.
- BERNÉ MANERO, C.; GARGALLO VALERO, P.; MARTÍNEZ CARABALLO, N. y SALVADOR FIGUERAS, M. (2008): “Análisis del Comportamiento Variado de Compra de los Hogares Españoles: una Aproximación Bayesiana”, *Revista Estadística Española* (en prensa).
- BERNÉ MANERO, C. y MARTÍNEZ CARABALLO, N. (2008): “Determinantes del Comportamiento Variado del Consumidor en el Escenario de Compra”, *Revista Europea de Dirección y Economía de la Empresa*, 17 (en prensa).
- BERNÉ, C.; MOLLÁ, A. y MÚGICA, J. M. (2004): “The Influence of Purchase Frequency on Varied Consumption”, *33rd European Marketing Academy Conference*, Murcia, May.
- 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, pp. 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, pp. 151-164.
- CASELLA, G. y ROBERT, G. (1996): “Rao-Blacwellisation of Sampling Schemes”. *Biometrika*, 83, 81-94.
- EAST, R.; HAMMOND, K.; HARRIS, P. y LOMAX, W. (2000): “First-Store Loyalty and Retention”, *Journal of Marketing Management*, 16 (4), pp. 307–325.
- EAST, R.; HARRIS, P.; LOMAX, W.; WILLSON, G. y PERKINS, D. (1997): “First-Store Loyalty to US and British Supermarkets”, Kingston Business School, *Occasional Paper Series*, 27, July, pp. 1-15.
- EAST, R.; HARRIS, P.; WILLSON, G. y LOMAX, W. (1995): “Loyalty to Supermarkets”, *The International Review of Retail, Distribution and Consumer Research*, 5 (1), pp. 99-109.
- FOX, E. J.; MONTGOMERY, A. L. y LODISH, L. (2004): “Consumer Shopping and Spending across Retail Formats”, *Journal of Business*, 77 (2), pp. S25-S60.
- GIVON, M. (1984): “Variety Seeking Through Brand Switching”, *Marketing Science*, 3 (1), pp. 1-22.
- HOWARD, J. A. y SHETH, J. N. (1969): *The Theory of Buyer Behaviour*. New York: John Wiley and Sons.

- KAHN, B. E. (1995): "Consumer Variety-Seeking among Goods and Services. An Integrative Review", *Journal of Retailing and Consumer Services*, 2 (3), pp. 139-148.
- KAHN, B. E. y MCALISTER, L. (1997): *Grocery Revolution: The New Focus on the Consumer*, Reading, MA: Addison Wesley, Longman
- LANCASTER, T. (2004): *An Introduction to Modern Bayesian Econometrics*. Blacwell 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), pp. 97-106.
- MASON, N. (1991): "An Investigation into Grocery Shopping Behaviour in Britain", AC Nielsen, Nielsen House, Heading-on, Oxford OX3 9RX.
- MCGOLDRICK, P. J. y ANDRÉ, E. (1997): "Consumer Misbehaviour: Promiscuity or Loyalty in Grocery Shopping", *Journal of Retailing and Consumer Services*, 4 (2), pp. 73-81.
- MEULENBERG, M. T. G. (1989): "Variation in Food Consumption: Some Aspects of Measurement and Empirical Findings for the Netherlands", *Tijdschrift voor Sociaal – Wetenschappelijk Onderzoek van de Landbouw*, 4, pp. 3-24.
- 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.
- PESSEMIER, E. (1985): "Varied Individual Behavior: Some Theories, Measurement and Models"; *Multivariate Behavioral Research*, 20, pp. 69-94.
- RESA, S. (2007): "Hábitos de Compra y Consumo en Navidad", *Distribución y Consumo*, Septiembre-Octubre, pp. 87-98.
- RHEE, H. y BELL, D. R. (2002): "The Inter-Store Mobility of Supermarket Shoppers", *Journal of Retailing*, 78 (4), pp. 225-237.
- ROBERT, C. y CASELLA, G. (2004): *Monte Carlo Statistical Methods*. Second Edition. Springer Texts in Statistics.
- 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, pp. 1-17.
- SEGARRA ROCA, P. (2007): *Influencia de la Heterogeneidad del Mercado en la Intención de Comportamiento del Consumidor: Respuestas a la Actividad Relacional en la Distribución de Gran Consumo*, Tesis Doctoral, Universitat Rovira I Virgili, Junio, Barcelona.

- SOBERON-FERRER, H. y DARDIS, R. (1991): "Determinants of Household Expenditures for Services", *Journal of Consumer Research*, 17 (4), pp. 385-397.
- THEIL, H. (1967): *Economics and Information Theory*, Amsterdam: North-Holland.
- THEIL, H. y FINKE, R. (1983): "The Consumer's Demand for Diversity", *European Economic Review*, 23, pp. 395-400.
- 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.

APÉNDICES

En los siguientes apéndices se calculan las distribuciones a posteriori de los parámetros de los modelos (3.1)-(3.2). Además, se muestra cómo calcular los cuantiles de dichas distribuciones que luego se utilizan para realizar inferencias acerca de dichos parámetros y como elaborar predicciones del comportamiento variado de una familia de características dadas. Finalmente se describe el criterio de bondad de ajuste del modelo a los datos utilizado en el trabajo.

A.1 Distribución a priori

Dado que la aproximación adoptada en el trabajo es bayesiana, necesitamos explicitar la distribución a priori sobre los parámetros del modelo. En nuestro caso la distribución es la utilizada habitualmente en la literatura (Lancaster, 2004) y viene dada por:

$$\boldsymbol{\gamma} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_\gamma) \quad (\text{A.1})$$

$$\boldsymbol{\beta} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_\beta) \quad (\text{A.2})$$

$$\tau_i = \frac{1}{\sigma_i^2} \sim \text{Gamma}\left(\frac{n_0}{2}, \frac{n_0 s_0}{2}\right); i=1, \dots, N \quad (\text{A.3})$$

donde: las matrices $p \times p$ $\boldsymbol{\Sigma}_\gamma$, $\boldsymbol{\Sigma}_\beta$ son definidas positivas y las constantes positivas n_0 y s_0 se suponen conocidas y donde se supone que las distribuciones (A.1), (A.2) y (A.3) son independientes.

En el estudio realizado se tomaron $n_0 = 0.01$, $s_0 = 1$, y $\boldsymbol{\Sigma}_\beta = \boldsymbol{\Sigma}_\gamma = 10000\mathbf{I}_{24}$ que corresponden a una distribución no informativa sobre los parámetros del modelo. Dado el elevado tamaño muestral disponible (2017 familias), un estudio de sensibilidad demostró que dichos valores no influyeron, de manera esencial, en los resultados obtenidos.

A.2 Distribución a posteriori

Si $\boldsymbol{\tau} = (\tau_1, \dots, \tau_N)'$, $L(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\tau} | \mathbf{D})$ denota la función de verosimilitud del modelo (3.1)-(3.2) y $\pi(\boldsymbol{\gamma})$, $\pi(\boldsymbol{\beta})$ y $\pi(\boldsymbol{\tau})$ son las densidades de las distribuciones a priori (A.1)-(A.3), respectivamente, la densidad a posteriori de dichos parámetros se obtiene a partir del Teorema de Bayes y viene dada por:

$$\begin{aligned} \pi(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\tau} | \mathbf{D}) &\propto L(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\tau} | \mathbf{D}) \pi(\boldsymbol{\gamma}) \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\tau}) \propto \\ &\prod_{i=1}^N \prod_{t=1}^T \left(\frac{\tau_i^{1/2} \exp\left[-\frac{\tau_i}{2} (\mathbf{h}_{it} - \boldsymbol{\beta}' \mathbf{x}_{it})^2\right]}{\Phi(\tau_i^{1/2} (\boldsymbol{\beta}' \mathbf{x}_{it}))} \right)^{(1-I_{it})y_{it}} \prod_{i=1}^N \prod_{t=1}^T [\Phi(\boldsymbol{\gamma}' \mathbf{x}_{it})]^{(1-I_{it})y_{it}} [1 - \Phi(\boldsymbol{\gamma}' \mathbf{x}_{it})]^{(1-I_{it})(1-y_{it})} \\ &\times \exp\left[-\frac{1}{2} \boldsymbol{\gamma}' \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{\gamma}\right] \exp\left[-\frac{1}{2} \boldsymbol{\beta}' \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta}\right] \prod_{i=1}^N \tau_i^{\frac{n_0-1}{2}} \exp\left[-\frac{n_0 s_0}{2} \tau_i\right] I_{(0, \infty)}(\tau_i) \end{aligned} \quad (\text{A.4})$$

Esta distribución no es tratable analíticamente y, por lo tanto, recurriremos a métodos MCMC (*Markov Chain Monte Carlo*) y, más concretamente, el *Gibbs sampling* con aumento de datos para calcularla (Lancaster, 2004; Rossi *et al.*, 2005). Para ello, introducimos las variables latentes $\boldsymbol{\lambda} = \{\lambda_{it}; i=1, \dots, N; t=1, \dots, T\}$ dadas por las expresiones

$$\lambda_{it} = \boldsymbol{\gamma}' \mathbf{x}_{it} + v_{it} \text{ con } v_{it} \sim N(0,1) \quad i=1, \dots, N; t=1, \dots, T$$

verificando que $\lambda_{it} \geq 0$ si $y_{it} = 1$ y $\lambda_{it} < 0$ si $y_{it} = 0$; $i=1, \dots, N$ y donde el indicador $I_{it} = 1$ si no se dispone del dato h_{it} por tratarse de un periodo vacacional de la familia i -ésima de la muestra y vale 0 en otro caso.

El cálculo de la distribución (A.4) se realiza a partir de la distribución:

$$\pi(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda} | \mathbf{D}) \propto \prod_{i=1}^N \prod_{t=1}^T \left(\frac{\tau_i^{1/2} \exp\left[-\frac{\tau_i}{2} (\mathbf{h}_{it} - \boldsymbol{\beta}' \mathbf{x}_{it})^2\right]}{\Phi(\tau_i^{1/2} (\boldsymbol{\beta}' \mathbf{x}_{it}))} \right)^{(1-I_{it})y_{it}} \times$$

$$\begin{aligned}
& \times \prod_{i=1}^N \prod_{t=1}^T \exp \left[-\frac{1}{2} (1 - I_{it}) (\lambda_{it} - \boldsymbol{\gamma}' \mathbf{x}_{it})^2 \right] \left(I_{[0, \infty)}(\lambda_{it}) \right)^{(1-I_{it})y_{it}} \left(I_{(-\infty, 0)}(\lambda_{it}) \right)^{(1-I_{it})(1-y_{it})} \\
& \times \exp \left[-\frac{1}{2} \boldsymbol{\gamma}' \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{\gamma} \right] \exp \left[-\frac{1}{2} \boldsymbol{\beta}' \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta} \right] \prod_{i=1}^N \tau_i^{\frac{n_0}{2} - 1} \exp \left[-\frac{n_0 s_0}{2} \tau_i \right] I_{(0, \infty)}(\tau_i) \quad (\text{A.5})
\end{aligned}$$

donde: $I_A(x)$ denota la función indicador en A . Observar, en particular, que la distribución (A.4) es la distribución marginal de (A.5).

Dado que (A.5) tampoco es tratable analíticamente utilizaremos el *Gibbs sampling* para calcularla. Dicho algoritmo nos proporcionará una muestra de (A.5) a partir de la cual podremos realizar inferencias acerca de las componentes de la distribución (A.4). Para implementar el *Gibbs sampling* necesitamos conocer las distribuciones completamente condicionadas de (A.5). En la siguiente subsección se describe la forma de dichas distribuciones así como la forma de extraer una muestra de cada una de ellas.

A.2.1. Distribuciones completamente condicionadas de (A.5)

Vienen dadas por las siguientes expresiones

1) Distribución $\boldsymbol{\beta} \mid \mathbf{D}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \boldsymbol{\gamma}$

Se tiene que $\boldsymbol{\beta} \mid \mathbf{D}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \boldsymbol{\gamma} = \boldsymbol{\beta} \mid \mathbf{D}, \boldsymbol{\tau}$ donde:

$$\begin{aligned}
\pi(\boldsymbol{\beta} \mid \mathbf{D}, \boldsymbol{\tau}) & \propto \prod_{i=1}^N \prod_{t=1}^T \left(\frac{\exp \left[-\frac{\tau_i}{2} (h_{it} - \boldsymbol{\beta}' \mathbf{x}_{it})^2 \right]}{\Phi(\tau_i^{1/2}(\boldsymbol{\beta}' \mathbf{x}_{it}))} \right)^{(1-I_{it})y_{it}} \exp \left[-\frac{1}{2} \boldsymbol{\beta}' \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta} \right] \propto \\
& \propto \frac{\exp \left[-\frac{1}{2} (\boldsymbol{\beta} - \mathbf{m}_{\boldsymbol{\beta}})' \mathbf{S}_{\boldsymbol{\beta}}^{-1} (\boldsymbol{\beta} - \mathbf{m}_{\boldsymbol{\beta}}) \right]}{\prod_{i=1}^N \prod_{t=1}^T [\Phi(\tau_i^{1/2}(\boldsymbol{\beta}' \mathbf{x}_{it}))]^{y_{it}(1-I_{it})}}
\end{aligned}$$

donde:

$$\mathbf{m}_\beta = \mathbf{S}_\beta \left(\sum_{i=1}^N \tau_i \left(\sum_{t=1}^T (1 - I_{it}) y_{it} \mathbf{x}_{it} \mathbf{h}_{it} \right) \right)$$

$$\mathbf{S}_\beta = \left(\sum_{i=1}^N \tau_i \left(\sum_{t=1}^T (1 - I_{it}) y_{it} \mathbf{x}_{it} \mathbf{x}'_{it} \right) + \Sigma_\beta^{-1} \right)^{-1}$$

Para extraer una muestra de esta distribución utilizamos un paso de Hastings-Metropolis tomando como densidad de salto $N_p(\mathbf{m}_\beta, \mathbf{S}_\beta)$.

2) Distribución $\tau \mid \mathbf{D}, \beta, \lambda, \gamma$

Se tiene que $\tau \mid \mathbf{D}, \beta, \lambda, \gamma = \prod_{i=1}^N \tau_i \mid \mathbf{D}, \beta$ con:

$$\pi(\tau_i \mid \mathbf{D}, \beta) \propto \prod_{t=1}^T \left(\frac{\tau_i^{1/2} \exp\left[-\frac{\tau_i}{2} (\mathbf{h}_{it} - \beta' \mathbf{x}_{it})^2\right]}{\Phi(\tau_i^{1/2} (\beta' \mathbf{x}_{it}))} \right)^{y_{it}(1-I_{it})} \tau_i^{\frac{n_0}{2}-1} \exp\left[-\frac{n_0 s_0}{2} \tau_i\right] I_{(0, \infty)}(\tau_i) \propto$$

$$\tau_i^{\frac{n_0 + \sum_{t=1}^T (1-I_{it}) y_{it}}{2} - 1} \exp\left[-\frac{\tau_i}{2} \left(n_0 s_0 + \sum_{t=1}^T y_{it} (1-I_{it}) (\mathbf{h}_{it} - \beta' \mathbf{x}_{it})^2 \right)\right] I_{(0, \infty)}(\tau_i)$$

$$\propto \frac{\tau_i^{\frac{n_0 + \sum_{t=1}^T (1-I_{it}) y_{it}}{2} - 1} \exp\left[-\frac{\tau_i}{2} \left(n_0 s_0 + \sum_{t=1}^T y_{it} (1-I_{it}) (\mathbf{h}_{it} - \beta' \mathbf{x}_{it})^2 \right)\right] I_{(0, \infty)}(\tau_i)}{\prod_{t=1}^T [\Phi(\tau_i^{1/2} (\beta' \mathbf{x}_{it}))]^{(1-y_{it})(1-I_{it})}} \quad i=1, \dots, N$$

Para extraer una muestra de cada una de estas distribuciones utilizamos el método de las variables auxiliares (*slice sampling*) tomando como densidad de muestreo

$$\text{Gamma} \left(\frac{n_0 + \sum_{t=1}^T (1 - I_{it}) y_{it}}{2}, \frac{n_0 s_0 + \sum_{t=1}^T y_{it} (1 - I_{it}) (\mathbf{h}_{it} - \beta' \mathbf{x}_{it})^2}{2} \right) \text{ para } i=1, \dots, N \quad (\text{A.6})$$

Siguiendo este método se generarían $u_{it} \sim U\left(0, \frac{1}{\left[\Phi(\tau_i^{1/2}(\boldsymbol{\beta}'\mathbf{x}_{it}))\right]}\right)$ para cada par (i,t) tales

que $y_{it} = 1$, $I_{it} = 0$ con $i=1,\dots,N$; $t=1,\dots,T$. Posteriormente se generaría τ_i de la distribución (A.6) truncada en $(\tau_{i,\min}, \tau_{i,\max})$ donde:

$$\tau_{i,\min} = \max_{i:y_{it}=1, I_{it}=0 \wedge \boldsymbol{\beta}'\mathbf{x}_{it} > 0 \wedge u_{it} \geq 1} \left\{ \left(\max \left\{ \frac{\Phi^{-1}(1/u_{it})}{\boldsymbol{\beta}'\mathbf{x}_{it}}, 0 \right\} \right)^2 \right\}$$

$$\tau_{i,\max} = \min_{i:y_{it}=1, I_{it}=0 \wedge \boldsymbol{\beta}'\mathbf{x}_{it} < 0 \wedge u_{it} \geq 1} \left\{ \left(\frac{\Phi^{-1}(1/u_{it})}{\boldsymbol{\beta}'\mathbf{x}_{it}} \right)^2 \right\}$$

3) Distribución $\boldsymbol{\lambda} | \mathbf{D}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\gamma}$

Se tiene que $\boldsymbol{\lambda} | \mathbf{D}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\gamma} \sim \prod_{i=1}^N \prod_{t=1}^T \lambda_{it} | \mathbf{D}, \boldsymbol{\gamma}$ con $\lambda_{it} | \mathbf{D}, \boldsymbol{\gamma} \sim N(\boldsymbol{\gamma}'\mathbf{x}_{it}, 1)$ truncada en

$[0, \infty)$ si $I_{it}=0$ e $y_{it} = 1$ y en $(-\infty, 0)$ si $I_{it} = y_{it} = 0$ para $i=1,\dots,N$; $t=1,\dots,T$

4) Distribución $\boldsymbol{\gamma} | \mathbf{D}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}$

Se tiene que $\boldsymbol{\gamma} | \mathbf{D}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda} \sim \boldsymbol{\gamma} | \mathbf{D}, \boldsymbol{\lambda} \sim N_p(\mathbf{m}_\gamma, \mathbf{S}_\gamma)$ con:

$$\mathbf{m}_\gamma = \mathbf{S}_\gamma \left(\sum_{i=1}^N \sum_{t=1}^T (1 - I_{it}) \mathbf{x}_{it} \lambda_{it} \right) \text{ y } \mathbf{S}_\gamma = \left(\sum_{i=1}^N \sum_{t=1}^T (1 - I_{it}) \mathbf{x}_{it} \mathbf{x}_{it}' + \boldsymbol{\Sigma}_\gamma^{-1} \right)^{-1}$$

Utilizando estas distribuciones el siguiente algoritmo obtendría una muestra de la distribución a posteriori (A.5).

A.2.2. Algoritmo

Paso 0: Extraer una muestra $(\boldsymbol{\beta}^{(0)}, \boldsymbol{\tau}^{(0)}, \boldsymbol{\gamma}^{(0)})$ de la distribución a priori (A.1)-(A.3).

Fijar $niter_{\max}$ número máximo de iteraciones. Colocar el contador de iteraciones $s = 1$.

Paso 1: Para $i=1, \dots, N$; $t=1, \dots, T$ extraer $\lambda_{it}^{(s)}$ de $N(\boldsymbol{\gamma}^{(s-1)'} \mathbf{x}_{it}, 1)$ truncada en $[0, \infty)$ si $I_{it}=0$ e $y_{it} = 1$ y en $(-\infty, 0)$ si $I_{it} = y_{it} = 0$

Paso 2: Extraer $\boldsymbol{\gamma}^{(s)}$ de $N_p(\mathbf{m}_{\boldsymbol{\gamma}}^{(s)}, \mathbf{S}_{\boldsymbol{\gamma}})$ donde $\mathbf{S}_{\boldsymbol{\gamma}} = \left(\sum_{i=1}^N \sum_{t=1}^T (1 - I_{it}) \mathbf{x}_{it} \mathbf{x}_{it}' + \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \right)^{-1}$ y

$$\mathbf{m}_{\boldsymbol{\gamma}}^{(s)} = \mathbf{S}_{\boldsymbol{\gamma}} \left(\sum_{i=1}^N \sum_{t=1}^T (1 - I_{it}) \mathbf{x}_{it} \lambda_{it}^{(s)} \right)$$

Paso 3: Extraer $\boldsymbol{\beta}^*$ de $N_p(\mathbf{m}_{\boldsymbol{\beta}}^{(s-1)}, \mathbf{S}_{\boldsymbol{\beta}}^{(s-1)})$, donde:

$$\mathbf{S}_{\boldsymbol{\beta}}^{(s)} = \left(\sum_{i=1}^N \tau_i^{(s)} \left(\sum_{t=1}^T (1 - I_{it}) y_{it} \mathbf{x}_{it} \mathbf{x}_{it}' \right) + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \right)^{-1} \text{ y } \mathbf{m}_{\boldsymbol{\beta}}^{(s)} = \mathbf{S}_{\boldsymbol{\beta}}^{(s)} \left(\sum_{i=1}^N \tau_i^{(s)} \left(\sum_{t=1}^T (1 - I_{it}) y_{it} \mathbf{x}_{it} h_{it} \right) \right)$$

Extraer $u \sim U(0, 1)$. Si $u \leq \min\{1, \alpha\}$ tomar $\boldsymbol{\beta}^{(s)} = \boldsymbol{\beta}^*$. En otro caso tomar $\boldsymbol{\beta}^{(s)} = \boldsymbol{\beta}^{(s-1)}$

donde

$$\alpha = \frac{\prod_{i=1}^N \prod_{t=1}^T \left[\Phi \left(\sqrt{\tau_i^{(s-1)}} \left(\boldsymbol{\beta}^{(s-1)' } \mathbf{x}_{it} \right) \right) \right]^{y_{it}(1-I_{it})}}{\prod_{i=1}^N \prod_{t=1}^T \left[\Phi \left(\sqrt{\tau_i^{(s-1)}} \left(\boldsymbol{\beta}^{*'} \mathbf{x}_{it} \right) \right) \right]^{y_{it}(1-I_{it})}}$$

Paso 4: Generar $u_{it}^{(s)} \sim U \left(0, \frac{1}{\left[\Phi \left(\sqrt{\tau_i^{(s-1)}} \left(\boldsymbol{\beta}^{(s)' } \mathbf{x}_{it} \right) \right) \right]} \right)$ para cada par (i, t) tales que y_{it}

$= 1, I_{it} = 0$ con $i=1, \dots, N$; $t=1, \dots, T$. Para $i=1, \dots, N$ generar $\tau_i^{(s)}$ de

$$\text{Gamma} \left(\frac{n_0 + \sum_{t=1}^T (1 - I_{it}) y_{it}}{2}, \frac{n_0 s_0 + \sum_{t=1}^T y_{it} (1 - I_{it}) (\mathbf{h}_{it} - \boldsymbol{\beta}^{(s)'} \mathbf{x}_{it})^2}{2} \right)$$

truncada en $(\tau_{i,\min}^{(s)}, \tau_{i,\max}^{(s)})$ donde:

$$\tau_{i,\min}^{(s)} = \max_{i: y_{it}=1, I_{it}=0 \wedge \boldsymbol{\beta}^{(s)'} \mathbf{x}_{it} > 0 \wedge u_{it}^{(s)} \geq 1} \left\{ \left(\max \left\{ \frac{\Phi^{-1}(1/u_{it}^{(s)})}{\boldsymbol{\beta}^{(s)'} \mathbf{x}_{it}}, 0 \right\} \right)^2 \right\}$$

$$\tau_{i,\max}^{(s)} = \min_{i: y_{it}=1, I_{it}=0 \wedge \boldsymbol{\beta}^{(s)'} \mathbf{x}_{it} < 0 \wedge u_{it}^{(s)} \geq 1} \left\{ \left(\frac{\Phi^{-1}(1/u_{it}^{(s)})}{\boldsymbol{\beta}^{(s)'} \mathbf{x}_{it}} \right)^2 \right\}$$

Paso 5: Si $s = \text{niter}_{\max}$ parar. En caso contrario hacer $s = s+1$ e ir al paso 1.

Como resultado de este algoritmo se obtiene una muestra

$$\left\{ (\boldsymbol{\beta}^{(s)}, \boldsymbol{\tau}^{(s)}, \boldsymbol{\gamma}^{(s)}, \boldsymbol{\lambda}^{(s)}) ; s = \text{niter}_0 + 1, \dots, \text{niter}_{\max} \right\} \quad (\text{A.7})$$

de la distribución a posteriori (A.5) (y, por consiguiente de (A.4) si nos quedamos solo con las componentes $\boldsymbol{\beta}$, $\boldsymbol{\tau}$ y $\boldsymbol{\gamma}$) donde niter_0 es el número de iteraciones que se estima necesario para alcanzar la convergencia del algoritmo³. En nuestro caso se tomó $\text{niter}_{\max} = 10000$ iteraciones alcanzándose la convergencia tras $\text{niter}_0 = 1000$ iteraciones. A partir de la muestra (A.7) se obtuvieron estimaciones puntuales de los parámetros utilizando la

³ Métodos para determinar dicho valor pueden encontrarse en Robert y Casella (2004).

mediana e intervalos de credibilidad bayesianos del $100(1-\alpha)\%$ ($0 < \alpha < 1$) utilizando los cuantiles $\alpha/2$ y $1-\alpha/2$.

A.2. 3. Bondad de ajuste y comparación de modelos

Los procesos de bondad de ajuste y comparación de modelos se realizaron desde un punto de vista predictivo, comparando las distribuciones predictivas $\{h_{it} \mid \mathbf{D}, \mathbf{x}_{it} : i=1, \dots, N; t = 1, \dots, T \text{ tales que } I_{it} = 0\}$ con los valores observados de h_{it} . A tal fin, y dado el tratamiento analítico de dichas distribuciones es complicado, se aplicó el método de Monte Carlo y, más concretamente, el muestreo de composición, para obtener una muestra de dichas distribuciones. Para ello se utilizó la muestra (A.7) de la distribución a posteriori (A.5) y se aplicó el siguiente algoritmo de predicción:

Para $i=1, \dots, N; t = 1, \dots, T$ tales que $I_{it} = 0$ y para $s = niter_0+1, \dots, niter_{max}$ realizar los pasos 1) y 2) siguientes:

- 1) Generar $\lambda_{it}^{(s)}$ de $N(\boldsymbol{\gamma}^{(s)'} \mathbf{x}_{it}, 1)$. Si $\lambda_{it}^{(s)} < 0$ poner $h_{it}^{(s)} = 0$. En otro caso ir al paso 2.
- 2) Generar $h_{it}^{(s)}$ de una distribución $N\left(\boldsymbol{\beta}^{(s)'} \mathbf{x}_{it}, \frac{1}{\tau_i^{(s)}}\right)$ truncada en $(0, \infty)$.

Como resultado de este algoritmo se obtiene una muestra

$$\{h_{it}^{(s)}; s = niter_0 + 1, \dots, niter_{max}\} \quad (A.8)$$

de las distribuciones predictivas $h_{it} \mid \mathbf{D}, \mathbf{x}_{it}$ para $i=1, \dots, N; t = 1, \dots, T$ tales que $I_{it} = 0$.

La bondad de ajuste del modelo (3.1)-(3.2) se evaluó analizando el cubrimiento empírico $CUB(100*(1-\alpha))$ ($0 < \alpha < 0.5$) de los los intervalos predictivos bayesianos dados

por $\left[h_{it}\left(\frac{\alpha}{2}\right), h_{it}\left(1-\frac{\alpha}{2}\right) \right]$ donde $h_{it}(\alpha)$ denota el cuantil α de la muestra (A.8). El

cubrimiento empírico para un nivel del $100*(1-\alpha)\%$ viene dado por la expresión:

$$\text{CUB}(100*(1-\alpha)) = 100 \frac{\sum_{i=1}^N \sum_{t=1}^T (1-I_{it}) I_{\left[h_{it}\left(\frac{\alpha}{2}\right), h_{it}\left(1-\frac{\alpha}{2}\right) \right]}(h_{it})}{\sum_{i=1}^N \sum_{t=1}^T (1-I_{it})} \quad (\text{A.9})$$

Así mismo, los criterios de comparación de modelos utilizados en la sección 4.2 vienen dados por las expresiones:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N \sum_{t=1}^T (1-I_{it}) (h_{it} - \hat{h}_{it})^2}{\sum_{i=1}^N \sum_{t=1}^T (1-I_{it})}} \quad (\text{A.10})$$

$$\text{MAD} = \frac{\sum_{i=1}^N \sum_{t=1}^T (1-I_{it}) |h_{it} - \hat{h}_{it}|}{\sum_{i=1}^N \sum_{t=1}^T (1-I_{it})} \quad (\text{A.11})$$

donde $\hat{h}_{it} = E[h_{it} | \mathbf{D}, \mathbf{x}_{it}] ; i=1, \dots, N$.

El cálculo de $\hat{h}_{it} = E[h_{it} | \mathbf{D}, \mathbf{x}_{it}] ; t=1, \dots, T ; i=1, \dots, N$ se llevó a cabo mediante el estimador tipo “Blackwell-Rao” (Casella y Robert, 1996) dado por las expresiones

$$\frac{1}{\text{niter}_{\max} - \text{niter}_0} \sum_{s=\text{niter}_0+1}^{\text{niter}_{\max}} E[h_{it} | \boldsymbol{\theta}^{(s)}, \mathbf{x}_{it}]$$

donde:

$$E[h_{it} | \boldsymbol{\theta}, \mathbf{x}_{it}] = E[h_{it} | \boldsymbol{\beta}, \tau_i, \boldsymbol{\gamma}, \mathbf{x}_{it}] = \Phi(\boldsymbol{\gamma}' \mathbf{x}_{it}) E[h_{it} | \boldsymbol{\beta}, \tau_i, h_{it} > 0, \mathbf{x}_{it}] =$$

$$= \Phi(\boldsymbol{\gamma}' \mathbf{x}_{it}) \left(\boldsymbol{\beta}' \mathbf{x}_{it} + \frac{1}{\Phi(\sqrt{\tau_i} (\boldsymbol{\beta}' \mathbf{x}_{it}))} \frac{\exp\left[-\frac{\tau_i}{2} (\boldsymbol{\beta}' \mathbf{x}_{it})^2\right]}{\sqrt{2\pi\tau_i}} \right)$$

Así mismo, se utilizó el logaritmo de la densidad predictiva a posteriori

$$\text{LPRED} = \log(f(\mathbf{h}|\mathbf{D})) \quad (\text{A.12})$$

donde $f(\mathbf{h}|\mathbf{D})$ es la función de densidad predictiva a posteriori que viene dada por:

$$f(\mathbf{h}|\mathbf{D}) = \int f(\mathbf{h} | \boldsymbol{\theta}, \mathbf{D}) \pi(\boldsymbol{\theta} | \mathbf{D}) d\boldsymbol{\theta}$$

con $\mathbf{h} = (h_{it})'$ y $\boldsymbol{\theta}$ el vector de parámetros del modelo.

Dicha densidad se calculó mediante la expresión:

$$\frac{1}{\text{niter}_{\max} - \text{niter}_0} \sum_{s=\text{niter}_0+1}^{\text{niter}_{\max}} f(\mathbf{h} | \boldsymbol{\theta}^{(s)}, \mathbf{X})$$

donde:

$$f(\mathbf{h}|\boldsymbol{\theta}, \mathbf{X}) = f(\mathbf{h}|\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \mathbf{X}) =$$

$$= \prod_{i=1}^N \prod_{t=1}^T \left(\frac{\tau_i^{1/2} \exp\left[-\frac{\tau_i}{2} (h_{it} - \boldsymbol{\beta}' \mathbf{x}_{it})^2\right]}{\Phi(\sqrt{\tau_i} (\boldsymbol{\beta}' \mathbf{x}_{it}))} \right)^{y_{it}(1-I_{it})} \prod_{i=1}^N \prod_{t=1}^T [\Phi(\boldsymbol{\gamma}' \mathbf{x}_{it})]^{y_{it}(1-I_{it})} [1 - \Phi(\boldsymbol{\gamma}' \mathbf{x}_{it})]^{(1-y_{it})(1-I_{it})}$$