

EL PROBLEMA DE ELECCIÓN DE MARCA EN UNA CATEGORÍA COMPLEJA. APLICACIÓN DE UNA RED NEURONAL ARTIFICIAL

Mónica Cortiñas Ugalde¹, Margarita Elorz Domezain y José Miguel Múgica Grijalba
UNIVERSIDAD PÚBLICA DE NAVARRA

RESUMEN

Este trabajo presenta una aplicación de un modelo de red neuronal artificial para explicar la elección entre alternativas en el contexto de una categoría de compra frecuente y de alta penetración en los hogares (el café molido) analizando los datos del punto de venta disponibles para el total de compras en un establecimiento. Se comprueba como los efectos de los instrumentos de marketing en la probabilidad de elección no son homogéneos por alternativas y, además, se constatan efectos acumulativos en las promociones simultaneas. El modelo planteado proporciona, por tanto, información que puede complementar el análisis de otros modelos clásicos como el modelo logit multinomial cuando este se enfrenta a problemas en sus supuestos de partida.

Palabras clave: Modelos de elección, datos de escáner, logit multinomial, redes neuronales artificiales.

ABSTRACT

In this paper we use an Artificial Neural Network Model for the analysis of the brand choice problem in a high-penetration high-purchase frequency product category (ground coffee). Between other results, we find different impacts of the marketing instruments between alternatives and accumulative effects of simultaneous promotions in different alternatives. Therefore the information provided by this model is useful in helping the analyst complementing other classical models like the multinomial logit when these face problems in their initial assumptions.

Keywords: Choice models, scanner data, artificial neural networks, multinomial logit.

¹ Departamento de Gestión de Empresas. Universidad Pública de Navarra. Campus de Arrosadía 31006, Pamplona. Tfno. 948169895, Email: mcortinas@unavarra.es



1. INTRODUCCIÓN

Desde el punto de vista del detallista, el conocimiento del proceso por el cual los consumidores seleccionan una alternativa entre las que permanecen expuestas en el lineal del establecimiento tiene gran relevancia práctica. Las empresas de distribución en España actúan en un entorno cada vez más competitivo; mientras que en los años 90 los hipermercados vivieron su década dorada, a partir de finales de los 90, la competencia entre los distintos formatos se incrementa con las grandes cadenas de distribución tomando posiciones en todos los formatos y con un descenso de la cuota de mercado de hipermercados frente a supermercados (Distribución y actualidad, 2003). En este contexto competitivo, la preocupación de los detallistas por la reducción de costes y el mantenimiento de la cuota de mercado es máxima, por lo que surgen nuevas tendencias de gestión, como las técnicas de Respuesta Eficiente al Consumidor (ECR) que implican, entre otros aspectos, tanto generar una política de inventarios eficiente, que ayude a la reducción de costes, como seleccionar una política de precios y promociones adecuada. Para esto se precisan buenas previsiones de la demanda y valoraciones del impacto de cada uno de los instrumentos del marketing en esta demanda, tanto para cada referencia como al nivel de categoría. Las nuevas tecnologías de la información, como los lectores de códigos de barra instalados en los puntos de venta y la creación de bases de datos de clientes, han facilitado enormemente la recogida de datos actualizados para dar respuesta a estas cuestiones. Ante la ingente cantidad de datos generados por estos sistemas, se hace necesario disponer de modelos adecuados para que el minorista pueda convertirlos en información útil y pueda incorporar esa información a la gestión de su oferta.

En aras a proporcionar un instrumento adecuado para la toma de decisiones en este contexto de gestión detallista, en este trabajo se analiza el problema de elección de marca, utilizando para ello una red neuronal artificial (en adelante RNA) y

centrándonos en dos aspectos: (1) capacidad de las RNA para estimar las cuotas de mercado para una categoría de producto dada y (2) información que este modelo proporciona acerca del impacto de los instrumentos de marketing en estas cuotas de mercado. Como base para la evaluación de los resultados se toma un modelo clásico en el tratamiento del problema de elección, como es el modelo logit multinomial. El modelo logit multinomial es un modelo ampliamente contrastado en el contexto de problemas de elección, por lo que los resultados que proporciona en cuanto a capacidad de ajuste a los datos pueden contemplarse como requerimientos mínimos para los modelos alternativos.

Respecto a la información que proporciona el modelo RNA acerca del efecto de los instrumentos de marketing en las cuotas de mercado, si existen efectos como interacciones entre distintos precios (BUCKLIN y GUPTA, 1999), efectos umbral, o asimetrías en las respuestas de las distintas alternativas a cada uno de los instrumentos, los supuestos de muchos de los modelos clásicos se incumplen. Por ejemplo, una promoción en precios puede no ser efectiva hasta que no se supera un determinado umbral de valor como un 15% también puede tener un efecto máximo en el 20% a partir del cual el efecto en cuota de mercado disminuye. También es posible que para una referencia dada, el efecto de una promoción sea superior cuando simultáneamente se realiza una promoción de otra alternativa. En el caso de las RNA, no se realizan hipótesis a priori sobre la relación entre los instrumentos de marketing y la respuesta de mercado, sino que estas relaciones se derivan de un conjunto de datos dado definiendo la estructura que proporciona un mejor ajuste de este conjunto de datos. Por tanto, un modelo RNA puede ser un instrumento útil para investigar la posible existencia de estos efectos y mostrarse más efectivo en el estudio de un problema de elección.

Este trabajo se estructura en seis partes. En el punto dos se introduce el problema de elección de marca y su análisis mediante el modelo RNA. El punto tres realiza la revisión de las aplicaciones



RNA en el área de marketing y especialmente en el análisis de la respuesta de mercado. Estas aplicaciones, centradas principalmente en el análisis de la capacidad predictiva de estos modelos, sugieren que las RNA pueden proporcionar un ajuste mayor que otros modelos tradicionales como modelos de regresión lineal para ventas, análisis discriminante y el modelo logit multinomial. Sin embargo, en la mayoría de los casos, no se analiza el efecto de las variables independientes, por lo que las fuentes de mejora en la estimación no quedan claras. En el punto cuatro se describen los objetivos e hipótesis de la investigación, la base de datos y los modelos empíricos utilizados en su aplicación al problema de elección de marca. A continuación, en el punto cinco, se describe el proceso de estimación de los cuatro modelos empíricos y los resultados obtenidos en la aplicación de estos modelos. Estos resultados muestran una buena capacidad de ajuste de los dos modelos empíricos de RNA y la existencia de efectos tanto lineales como no lineales para las promociones de precios, además de efectos de interacción entre descuentos. Por último, en el punto seis, se señalan las principales conclusiones del trabajo, sus implicaciones para la gestión y las posibles futuras extensiones al mismo.

2. EL PROBLEMA DE ELECCIÓN DE MARCA EN UNA CATEGORÍA COMPLEJA

La respuesta de mercado esperada para unas determinadas actividades de marketing se forma por la agregación de las respuestas de los individuos que componen un determinado mercado ante estas actividades. Por tanto, la modelización de respuestas de mercado está íntimamente relacionada con la teoría subyacente sobre comportamiento de estos individuos. Las ventas totales dentro de una determinada categoría y las cuotas de mercado para distintas alternativas de una categoría son, en este sentido, el resultado del proceso de decisión de compra de los individuos que componen el mercado.

Siguiendo a ROBERTS Y LILLEN (1993), la modelización del proceso de decisión de compra se encuadra dentro de un conjunto de modelos de comportamiento del consumidor para cada una de las distintas etapas del proceso básico de compra de los consumidores, desde el nacimiento de la necesidad hasta el comportamiento post-compra. Por tanto, la modelización de la decisión de compra implica una teoría subyacente sobre el comportamiento del consumidor durante la etapa de elección entre alternativas y las anteriores. Para establecer esta teoría es necesario definir cuatro elementos: el individuo que toma la decisión, las alternativas disponibles, los atributos que caracterizan a las alternativas y una regla de decisión (BEN-AKIVA Y LERMAN, 1985). En general, se considera que las alternativas en el conjunto de elección pueden caracterizarse completamente mediante el conjunto de atributos. La decisión de compra será por tanto más compleja cuanto mayor sea el número de atributos que diferencian a las alternativas, cuanto mayor sea el número de alternativas disponibles, cuanto mayor sea la complejidad de la regla de elección utilizada y dependiendo de las características del individuo decisor.

Uno de los modelos más utilizados en este contexto es el modelo logit multinomial. En este modelo, se caracteriza la utilidad atribuida por el individuo a una alternativa i en el conjunto de elección C por la expresión (BEN-AKIVA Y LERMAN, 1985):

$$U_i = V_i + \epsilon_i \quad [2.1]$$

Donde: V_i es el componente sistemático de la función y ϵ_i es el componente aleatorio que recoge los factores no observados.

La probabilidad de elección de esta alternativa es la probabilidad de que su utilidad sea mayor o igual que la utilidad del resto de las alternativas en el conjunto de elección. Bajo ciertos supuestos, esta probabilidad toma la forma:

$$P_C(i) = e^{V_i} / \sum_{j \in C} e^{V_j} \quad [2.2]$$

Donde: c es el conjunto de elección.



Normalmente, se asume que la función V para un individuo tiene una forma aditiva y linealmente separable (MCFADDEN, 1986) de la forma:

$$V_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik} \quad [2.3]$$

Donde: V_i es la utilidad sistemática de la alternativa i

X_{ik} es el vector de k atributos de la alternativa i para el individuo n

$\beta = [\beta_1, \beta_2, \dots, \beta_k]$ es el vector de parámetros a estimar

Este modelo se estima por el procedimiento de máxima verosimilitud. Una de las propiedades de este modelo, es la que se deriva del supuesto de errores independientemente distribuidos. Esta propiedad, denominada propiedad de independencia de las alternativas irrelevantes, establece que, para un individuo, la relación entre las probabilidades de elección de dos alternativas es constante e independiente de las probabilidades de elección del resto de las alternativas en el conjunto de elección. Como resultado de esta propiedad, variaciones en el conjunto de elección no provocan variaciones en las probabilidades de elección relativas entre alternativas. Además, una variación en los valores de los atributos de una alternativa tiene los mismos efectos en las probabilidades de elección del resto de las alternativas.

Los supuestos del modelo logit multinomial pueden en muchos casos no ser realistas. Por ejemplo, en el caso de la categoría del café molido, tenemos dos atributos principales, la marca y el tipo de café, que puede ser natural, mezcla o algunas variedades especiales, además del precio del producto y las actividades promocionales. Los cambios en el precio de uno de los productos pueden afectar en mayor medida a las compras en otro de su misma variedad que al resto de productos. Además, los productos de una misma marca pueden seguir una misma estrategia de precios y promociones que hagan que sea difícil distinguir los efectos de cada uno de ellos. En el cuadro 1 se muestra un ejemplo para las compras de café molido en un hipermercado durante un año. Existen 14 alternativas disponi-

bles y se muestran las cuotas semanales de cada una de ellas comparando las que se obtienen cuando la alternativa con mayor cuota (la variedad mezcla de Bonka) está en promoción o no. Se observa como el efecto de la promoción de la alternativa líder afecta en mayor medida a algunas alternativas como las alternativas de Marcilla y Saimaza mientras que otras alternativas como 154, Brasilia o Soley resultan menos afectadas. Las marcas nacionales que compiten más directamente sufren los mayores efectos de las promociones. Por otro lado, la política de promociones simultaneas de las distintas variedades de las mismas marcas hace que los efectos de los descuentos en las alternativas de la misma marca (Bonka natural, alternativa 4) sean también positivos.

En este trabajo, se propone para el análisis de este problema un modelo flexible que permita incluir este tipo de efectos sin predeterminarlos de antemano. Las RNA son una familia de modelos que surgen inicialmente en campos de investigación relacionados con la forma de funcionamiento del cerebro humano. Sin embargo, la línea de investigación predominante en estos modelos está relacionada con sus buenas capacidades para representar distintos patrones de datos teniendo como objetivo no tanto desarrollar modelos plausibles desde un punto de vista biológico, como el que estos modelos sirvan para aproximar distintos fenómenos que van desde el reconocimiento de caracteres escritos a mano al diagnóstico de ciertas enfermedades a partir de sus síntomas (RUMELHART Y MCCLELLAND, 1986). Estos modelos han tenido una amplia aplicación dentro de distintas áreas de gestión, principalmente en el área de finanzas (una revisión de estas aplicaciones se encuentra en KRYCHA Y WAGNER, 1999) y en el análisis de problemas de marketing (HILERA Y MARTÍNEZ, 1995; DAVIES, GOODE, MAZANEC Y MOUTINHO, 1999). Entre sus principales ventajas están su aprendizaje adaptativo mediante la utilización de parte de los datos como entrenamiento, su tolerancia a fallos (por ejemplo datos extremos o errores en los datos), la capacidad para operar en tiempo real en el caso de utilizar el hardware apro-



piado y su capacidad para trabajar con gran cantidad de datos (LUQUE, 2003). Las RNA recogen muy distintos tipos de modelos pero una de las clasificaciones más usuales distingue entre las llamadas de aprendizaje supervisado, en las que

existe un resultado deseado que es el que se intenta modelizar, y las de aprendizaje no supervisado, en las que se trata de extraer patrones presentes en los datos sin necesidad de asociarlos a un resultado concreto.

CUADRO 1
Cuotas de mercado durante un año cuando existe o no promoción de la alternativa Bonka mezcla

Alternativa	Cuota media	Cuota no prom.	Cuota prom.	% variación
1 Bonka mezcla	18,3%	10,3%	25,0%	143,7%
2 Marcilla mezcla	15,0%	19,1%	11,6%	-39,4%
3 Marcilla natural	10,4%	13,2%	8,1%	-38,3%
4 Bonka natural	11,6%	6,7%	15,7%	136,2%
5 Brasilia mezcla	9,1%	9,0%	9,1%	1,1%
6 Saimaza mezcla	6,1%	8,3%	4,3%	-48,4%
7 154 mezcla	6,7%	6,6%	6,9%	5,3%
8 Soley mezcla	5,1%	5,5%	4,9%	-11,3%
9 Marcilla especial	3,6%	5,2%	2,2%	-57,1%
10 Saimaza natural	3,3%	4,5%	2,4%	-47,8%
11 Bonka especial	4,5%	4,4%	4,7%	7,0%
12 Bahía mezcla	2,4%	3,2%	1,7%	-48,6%
13 154 natural	2,2%	2,0%	2,3%	14,0%
14 Bahía natural	1,6%	2,1%	1,2%	-43,1%
Media	7,1%	7,1%	7,1%	-1,9%

Dentro del primer grupo, las redes de aprendizaje supervisado, uno de los modelos más habituales es la red multicapa alimentada hacia delante. Estas redes (también llamadas Multilayer Feedforward Artificial Neural Nets, en adelante MFANN) son básicamente un mecanismo para modelizar las relaciones entre unas variables de entrada (las denominadas variables independientes) y unas variables de salida (las denominadas variables dependientes) determinadas. En este tipo de red, las unidades en la primera capa (la capa de entrada) captan los valores de las variables explicativas en el sistema y envían ese valor x_i a la capa oculta intermedia. Las unidades en la capa oculta ponderan esos valores iniciales recibidos mediante unos pesos determinados y calculan un nuevo valor basado en la transformación del valor inicial mediante una función de transformación. A continuación envían ese valor a la

capa de salida. Por último, las unidades de salida calculan el resultado final, de la misma forma que las unidades ocultas. El resultado final de la red se puede expresar como una función de los valores iniciales y los vectores de pesos:

$$O_k = F_1(\sum_j w_{kj}^2 H_j) = F_1(\sum_j w_{kj}^2 F_2(\sum_i w_{ji}^1 x_i)) \quad [2.4]$$

Donde: O_k : Valor de la variable dependiente k
 w_{kj}^2 : Peso de conexión de la unidad oculta j a la unidad de salida k
 H_j : Valor de la unidad oculta h
 w_{ji}^1 : Pesos de conexión de la unidad de entrada i a la unidad oculta j
 X_i : valor de la unidad de entrada i
 F_1, F_2 : funciones de conexión

El procedimiento de estimación en una MFANN consiste en la selección de los pesos de



conexión entre las capas, de forma que el error cometido en las unidades de salida sea el mínimo. Este proceso de selección de los pesos se denomina aprendizaje y es un proceso iterativo en el que cada uno se modifica en función de la comparación entre el resultado obtenido para unos pesos dados y el resultado deseado. El procedimiento de aprendizaje en esta red puede considerarse, por tanto, similar al de estimación de los parámetros en un modelo econométrico. Existen diversos mecanismos para realizar este aprendizaje, siendo el más conocido en método de retropropagación de errores que después se ha modificado en diversos algoritmos para hacerlo más eficiente².

En el caso de aplicar este tipo de red al problema de elección de marca, las unidades de entrada en la red son los distintos atributos de las alternativas disponibles, mientras que las unidades de salida de la red son las probabilidades de elección de cada una de las alternativas.

3. LAS REDES NEURONALES Y LA PREVISIÓN DE VENTAS Y CUOTAS DE MERCADO

Las aplicaciones de las RNA al análisis de problemas de marketing siguen una tendencia creciente. Como ejemplos, podemos citar los trabajos de WRAY et al (1994) y BEJOU et al (1996), que utilizan una RNA para analizar las relaciones comprador-vendedor en un contexto de empresa-a-empresa, KUMAR et al (1995) y OLMEDA (1996), que analizan la decisión de aceptar o no un nuevo producto por distribuidores, CURRY y MOUTINHO (1993), comparando las RNA con los modelos de ecuaciones estructurales, REINEKE FLIN et al (1995) que comparan las RNA y la regresión a la hora de predecir el comportamiento de compra basándose en datos de encuesta o DAVIES, GOODE, MOUTINHO Y OGBONNA (2001) analizando la formación de la satisfacción del cliente con el super-

mercado minorista. Sin embargo, las aplicaciones de RNA más numerosas aparecen en el campo de segmentación de mercados. Por ejemplo, ZAHAVI y LEVIN (1997a) y (1997b) analizan la aplicabilidad de las redes neuronales para la planificación de mailings en campañas de marketing directo, MAZANEC (1992) compara las RNA con el análisis discriminante para segmentar el mercado de turistas y BOONE y ROEHM (2002) comparan una RNA con los modelos de mixturas en la segmentación de una base de datos de clientes de un detallista.

Centrándonos en los trabajos que pretenden estimar la respuesta de mercado a los instrumentos de marketing mediante distintos tipos de RNA, podemos dividirlos en dos grupos, los que intentan explicar los resultados para una única alternativa dada realizando comparaciones con técnicas de regresión y los que intentan modelizar elecciones teniendo en cuenta al mismo tiempo diferentes alternativas, más próximos a este trabajo. La mayor parte de los trabajos de estos dos grupos comparan los modelos únicamente en cuanto a su capacidad de ajuste a los datos, ya sean ventas o cuotas de mercado, y no en cuanto a las relaciones entre acciones de marketing y resultados.

En el primer grupo de trabajos, que explican los resultados para una referencia, podemos clasificar los trabajos de HRUSCHKA (1993), WIERENGA y KLUYTMANS (1994 y 1996), GRUCA et al (1995) y AINSCOUGH y ARONSON (1999). Los resultados de este primer grupo de trabajos indican que las RNA pueden obtener mejores resultados que modelos de regresión con las mismas variables en la predicción de ventas y cuotas de mercado de un producto de consumo. Las razones de esta mejora en la estimación se atribuyen a la flexibilidad de los modelos RNA frente a los modelos de regresión sobre todo en cuanto a su capacidad para modelizar relaciones no lineales y su mayor tolerancia a la presencia de multicolinealidad en los datos como muestran los estudios de simulación. Sin embargo, ninguno de estos trabajos analiza empí-

² En el Anexo se detallan los aspectos relativos a la implementación de este modelo en el presente trabajo.



ricamente las relaciones entre las variables independientes y las dependientes ni se comprueba la posible existencia de estos efectos no lineales. Además, en estos trabajos se toma una perspectiva de fabricante interesado en los resultados de una única referencia de forma aislada en lugar de los resultados para una categoría de producto, por lo que no se consideran efectos cruzados entre alternativas.

El segundo grupo de trabajos analiza el problema de elección entre distintas alternativas tenidas en cuenta simultáneamente mediante RNA. En este grupo de trabajos, más próximos al presente problema, se incluyen los trabajos de DASGUPTA et al (1994), WEST et al (1997), DE CARVALHO et al (1998), WARTENBERG y DECKER (1996), AGRAWAL y SCHORLING (1996), y BENTZ y MERUNKA (2000). Los tres primeros trabajos analizan el problema de elección entre únicamente dos alternativas mientras que en los otros tres, las alternativas consideradas son más de dos. Los dos últimos trabajos son los más relevantes en este caso, ya que realizan comparaciones empíricas entre el modelo logit multinomial y RNA. Este segundo grupo de trabajos presentan resultados positivos en cuanto a capacidad de ajuste de las RNA frente a los modelos logit binomial y multinomial, aunque no en todos los casos. Además, en cuanto a los efectos de las variables independientes, algunos de estos trabajos analizan los posibles efectos no lineales (WEST et al, 1997 y BENTZ y MERUNKA, 2000) para distintos tipos de variables con metodologías alternativas. En ambos casos se confirma la existencia de estos efectos. WARTENBERG y DECKER (1996) también analizan la naturaleza de la competencia entre alternativas encontrando asimetrías que no pueden detectarse con el modelo alternativo presentado. Sin embargo, este grupo de trabajos presenta algunas lagunas. Aunque los resultados de ajuste frente al modelo logit multinomial son superiores para la red neuronal, no se han investigado completamente las fuentes de esta mejora en ajuste. BENTZ y MERUNKA (2000) analizan únicamente una de las posibles fuentes de esta mejora, los efectos no lineales, a través de un

modelo restringido para permitir la existencia de este tipo de efectos. En el presente trabajo, sin embargo, se presenta un modelo más amplio, en el que no se realizan supuestos iniciales sobre la estructura del proceso de elección.

4. OBJETIVOS, HIPÓTESIS Y MODELOS EMPÍRICOS

4.1. Objetivos e hipótesis

Como se ha puesto de manifiesto en la revisión anterior de la literatura, las RNA han obtenido resultados superiores al modelo de contraste en la mayoría de los casos, aunque en ocasiones estos resultados no son concluyentes. Los resultados de los estudios de simulación muestran que las RNA pueden ser útiles en la estimación de cuotas de mercado proporcionando un ajuste superior a los modelos de contraste cuando los supuestos de los modelos lineales de regresión y los modelos logit que se utilizan como comparación no se cumplen, es decir, la forma funcional subyacente es no lineal, existen problemas de multicolinealidad, y, en el caso de los modelos logit, se incumple la propiedad de independencia de las alternativas irrelevantes. En categorías de producto en las que el número de alternativas disponibles para el consumidor es elevado, el proceso de elección es más complejo. En este contexto, es más probable que la red neuronal artificial supere el modelo logit multinomial en la estimación de las cuotas de mercado (AGRAWAL y SCHORLING, 1996). El primer objetivo de este trabajo es la evaluación de la capacidad de un modelo RNA para la estimación de las cuotas de mercado en una categoría de producto de compra compleja.

Respecto a la información que proporciona el modelo sobre el efecto de las variables independientes en las dependientes, en el caso del modelo logit multinomial se obtiene un conjunto de parámetros para los que, si se cumplen los supuestos de partida, es posible calcular valores de significatividad que permiten el contraste de



hipótesis. En el caso de la RNA, el número de parámetros estimados, (los pesos de la red), es muy superior al del modelo logit multinomial. Por ejemplo, una RNA con 14 variables independientes, 14 dependientes y 5 unidades ocultas tiene un total de 159 parámetros. Se han desarrollado estadísticos de contraste para estos pesos de este tipo de red mediante distribuciones asintóticas (e.g. KUNG y WHITE, 1992) pero la interpretación de estos parámetros para una variable dada resulta muy complicada debido a este alto número de parámetros. Para simplificar la interpretación de los resultados del modelo, en lugar de analizar directamente los pesos calculados, se van a calcular las derivadas de los resultados de la red con respecto a cada una de las variables independientes (WEST, BROCKET Y GOLDEN, 1997). Este procedimiento permite el análisis de los efectos de cada una de estas variables de forma separada y proporciona una visión descriptiva de los principales efectos presentes en los datos. Por ejemplo, VAN HEERDE et al (2001) analizan la naturaleza de la curva que relaciona los descuentos en precios con las ventas mediante una regresión no paramétrica. Sus resultados muestran que, en el caso de las promociones en precios de tres categorías de productos de consumo, la respuesta en ventas presenta tanto efectos de tipo umbral y niveles de saturación como efectos de interacción entre los descuentos de distintas alternativas. El segundo objetivo de este trabajo será la comprobación de la posible existencia de efectos no lineales, efectos umbral e interacciones (p.e.: mayor influencia de algunas alternativas en otras o efectos mayores de las promociones cuando se producen simultáneamente promociones en dos alternativas) entre los distintos instrumentos de marketing a través

del análisis de las derivadas parciales de los resultados en el modelo RNA.

4.2. Modelos empíricos y base de datos

La aplicación empírica de la RNA y el modelo alternativo (el logit multinomial) se va a realizar en la categoría de producto del café molido. En este trabajo únicamente se van a tener en cuenta, como determinantes para la elección, las características de las alternativas de la categoría y no características de los clientes que realizan la compra, aunque el comportamiento de elección de estos clientes pueda presentar diferencias según estas características (BERNÉ, CORTIÑAS, ELORZ y MÚGICA, 2004)³. Como características de estas alternativas, la base de datos recogida en los sistemas de escáner del punto de venta incluye únicamente información acerca del precio unitario de cada producto, la marca y el tipo de producto. La información sobre publicidad o colocación de productos puede estar disponible mediante un proceso de recogida de datos manual, factible para pequeños periodos de tiempo pero altamente costoso para el análisis de varias categorías de productos o durante largos periodos de tiempo por lo que utilizaremos únicamente la información disponible directamente para los detallistas, a través del escáner del establecimiento. El precio unitario recogido en cada transacción, se descompone en dos variables, que van a utilizarse en modelos alternativos: precio regular y descuento (GUPTA, 1988). De la consideración del precio unitario como variable única o como dos variables diferenciadas (precio regular y descuento) tanto en un modelo logit multinomial o en una MFANN surgen 4 modelos empíricos a estimar:

³ Los datos disponibles para la gestión detallista únicamente presentan información acerca de las características de los individuos en el caso de los individuos que disponen de una tarjeta del establecimiento y, en estos casos, únicamente para las compras que han realizado con esta tarjeta. El enfoque detallista de este trabajo hace que nos centremos únicamente en las características disponibles para todas las elecciones.



CUADRO 2
Modelos empíricos y variables utilizadas

Modelo	Nombre	Variables incluidas en el modelo		
		<i>Precio regular</i>	<i>Porcentaje Descuento</i>	<i>Precio Unitario</i>
1	MNLogit(a)	X	X	
2	MNLogit(b)			X
3	RNA(a)	X	X	
4	RNA(b)			X

La especificación de los modelos logit multinomiales con estas variables será:

Modelo MNLogit(a): $P(i) = \frac{\exp(\alpha_i + \beta_1 PR_i + \beta_2 Dto_i)}{\sum_1^K \exp(\alpha_k + \beta_1 PR_k + \beta_2 Dto_k)}$ [4.1]

Donde: P(i): Probabilidad de elección de la alternativa i [0-1]
 PR_i, PR_k: Precio regular de la alternativa i, k
 Dto_i, Dto_k: Porcentaje de descuento en precio de la alternativa i, k
 K: número de alternativas en el conjunto de elección
 α_i, β₁, β₂: Parámetros a estimar

Modelo MNLogit(b): $P(i) = \frac{\exp(\alpha_i + \beta_1 PU_i)}{\sum_1^K \exp(\alpha_k + \beta_1 PU_k)}$ [4.2]

Donde: P(i): Probabilidad de elección de la alternativa i [0-1]
 PU_i, PU_k: Precio unitario de la alternativa i, k
 K: número de alternativas en el conjunto de elección
 α_i, β₁: Parámetros a estimar

La MFANN utilizada en este caso utiliza dos tipos de funciones de transferencia: la función en la primera capa es de tipo logístico y en la segunda capa de tipo lineal. En el Anexo de este trabajo se detallan las decisiones tomadas en la definición del modelo RNA. La especificación resultante de los dos modelos de redes neuronales es:

Modelo RNA(a): $Cuota(k) = \frac{\sum_j w_{kj}^2 (1/1 + \exp(-\sum_j w_{ji}^1 X_i))}{\sum_j w_{kj}^2 (1/1 + \exp(-\sum_j w_{ji}^1 X_i))}$ [4.3]

$X = [PR_1, PR_2, PR_i, \dots, PR_K; Dto_1, Dto_2, Dto_i, \dots, Dto_K]$ [4.4]

Donde: Cuota(k): Cuota diaria de la alternativa k
 Número de unidades ocultas
 w²_{kj} w¹_{ji}: peso de la unidad j en la unidad k, peso de la unidad i en la unidad j
 X_i: valor de la unidad de entrada i, componente i del vector de variables X
 PR_i: Precio regular de la alternativa i
 Dto_i: Porcentaje de descuento en precio de la alternativa i
 K: número de alternativas en el conjunto de elección

Modelo RNA(b): $Cuota_i = \frac{\sum_j w_{kj}^2 (1/1 + \exp(-\sum_j w_{ji}^1 X_i))}{\sum_j w_{kj}^2 (1/1 + \exp(-\sum_j w_{ji}^1 X_i))}$ [4.5]

X = [PU₁, PU₂, PU_i, ..., PU_K] [4.6]
 Donde: PU_k: Precio unitario de la alternativa k
 El resto de variables son las mismas que en el caso anterior

Para comparar los modelos logit multinomial y las redes neuronales se utilizan las estimaciones de ambas modelizaciones. Sin embargo, en el modelo logit multinomial la variable estimada es la probabilidad de elección para cada una de las alternativas, mientras que en la red neuronal artificial se estiman directamente las cuotas semanales resultantes para las alternativas. COOPER y NAKANISHI (1988) muestran la relación existente entre la cuota de mercado de una marca y su probabilidad de elección. Esta relación depende del número de unidades compradas por elección, de la probabilidad media del mercado y de la probabilidad de elección individual. La mayor o menor



relación entre la cuota de las distintas marcas y las probabilidades de elección a nivel agregado dependerá del grado de homogeneidad y correlación que exista entre las probabilidades de elección y el volumen de compra de los consumidores que adquieren sus productos en el mercado. En este caso, al no disponer de datos sobre la frecuencia de compra ni sobre las probabilidades de elección individuales para las compras totales, las probabilidades de elección se equiparan a la cuota estimada para cada una de las alternativas.

La base de datos de trabajo comprende todas las compras de la categoría de café molido realizadas en un hipermercado de la zona este española durante 1999. Este hipermercado pertenece a una cadena de establecimientos regional, situada entre las 20 mayores españolas por volumen de facturación y entre las 25 primeras por número de empleados. La selección de un único establecimiento como base de trabajo permite, en primer lugar, homogeneizar en mayor medida el conjunto de elección presente durante el periodo ya que, aunque las grandes marcas nacionales pueden ofertarse durante un largo periodo de tiempo, la selección de marcas regionales y locales presentes en el establecimiento puede ser más variable por lo que el conjunto de elección puede presentar grandes variaciones en periodos de tiempo largo. En segundo lugar, tomando una perspectiva de gestión, algunos autores plantean la necesidad de establecer estrategias de precios a nivel de establecimiento, en lugar de a nivel de cadena, debido a las diferencias en las sensibilidades al precio que se pueden encontrar en los establecimientos de una misma cadena en distintas zonas (MONTGOMERY, 1997, HOCH et al 1995 entre otros).

La base de datos contiene inicialmente 28.842 observaciones. De estas observaciones iniciales, se eliminan aquellas alternativas que están presentes menos de tres semanas del total en el año en el establecimiento y aquellas que representan menos

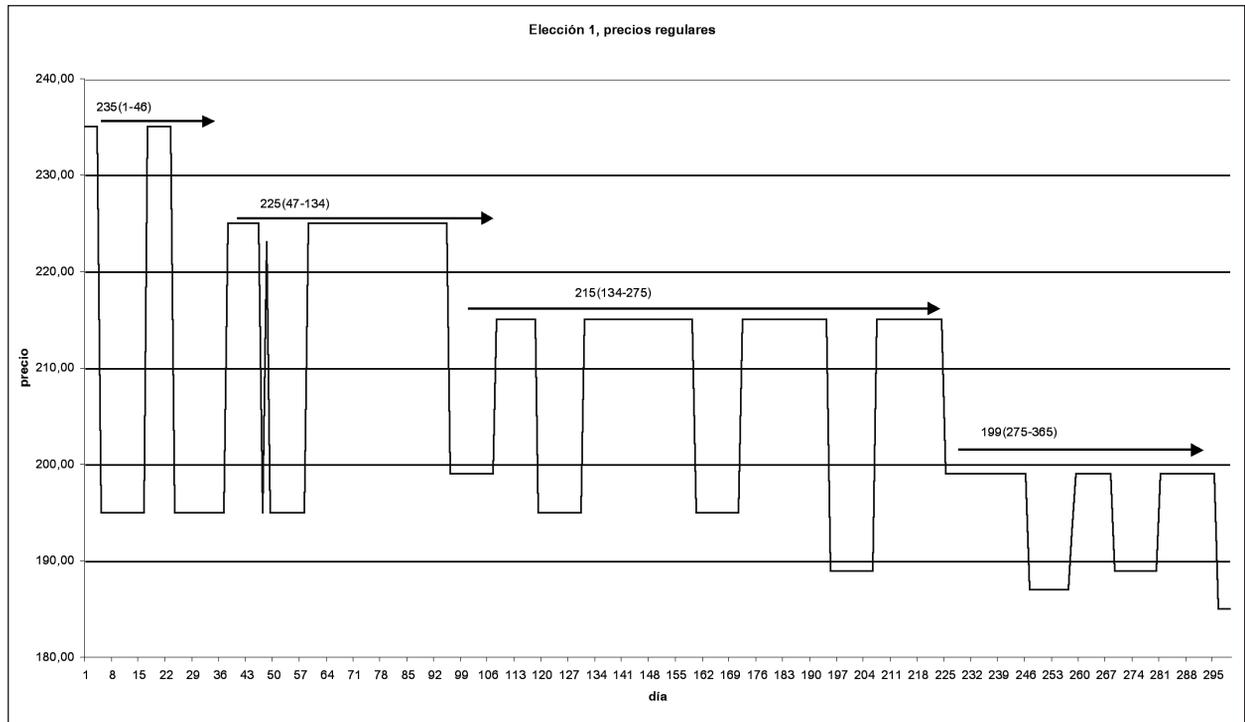
de un 1% de las compras totales de lo que resultan 25.565 observaciones de compra para 14 alternativas de producto. Las 14 alternativas se presentan en el mismo formato (250 gr.) de modo que el precio unitario que se toma por alternativa corresponde al precio del paquete y el precio regular se calcula por inspección visual de los datos teniendo en cuenta las disminuciones del precio unitario que se mantienen por periodos breves de tiempo (una o dos semanas) a partir de las cuales el precio unitario vuelve a los valores anteriores. El porcentaje de descuento se calcula a partir de este precio regular. En la figura 1 se muestra un ejemplo de construcción de los precios regulares para la alternativa 1.

Las alternativas corresponden, por una parte, a 3 variedades de café distintas (café mezcla, natural y especial) y, distinguiendo por marcas, a 7 marcas (4 marcas nacionales, Bonka, Marcilla, Saimaza y Brasilia y otras marcas locales, Café 154, Bahía y Soley). Las dos marcas nacionales con mayor cuota ofrecen las tres variedades contempladas mientras que el resto de las marcas presentan una o dos variedades. La variedad de café con mayor cuota en el establecimiento es la variedad mezcla, seguida del café natural. En el cuadro 3 se muestran algunos descriptivos para las 14 alternativas.

La alternativa de mayor precio medio es la que corresponde a la variedad especial de Bonka. Podemos observar como los precios para las variedades natural y mezcla de una misma marca presentan unos valores medios muy similares y como las marcas nacionales presentan precios medios mayores que el resto de las marcas en el conjunto de elección. El descuento medio se sitúa en el 10% del precio regular pero se presentan bastantes diferencias entre alternativas. Dos de ellas no se promocionan durante el periodo (Brasilia y Soley mezcla). Además, las alternativas de una misma marca tiende a promocionarse simultáneamente y con un porcentaje de descuento similar.



FIGURA 1
Precios regulares y descuentos para la alternativa Bonka mezcla



CUADRO 3
Descriptivos

Alternativa	Precio Unit	Precio Reg	Descuento	Cuota
1 Bonka mezcla	206	216	7,6%	18,31%
2 Marcilla mezcla	217	225	9,7%	14,97%
3 Marcilla natural	217	225	9,7%	10,41%
4 Bonka natural	207	217	7,0%	11,61%
5 Brasilia mezcla	209	209	–	9,06%
6 Saimaza mezcla	226	233	12,8%	6,12%
7 154 mezcla	199	225	11,7%	6,74%
8 Soley mezcla	205	206	–	5,13%
9 Marcilla especial	215	222	8,7%	3,58%
10 Saimaza natural	225	233	13,4%	3,34%
11 Bonka especial	270	271	4,4%	4,52%
12 Bahía mezcla	189	191	15,2%	2,39%
13 154 natural	184	186	9,4%	2,18%
14 Bahía natural	188	191	15,9%	1,62%
Media	211	218	10,5%	7,14%



5. PROCESO DE ESTIMACIÓN Y SELECCIÓN DE MODELOS Y RESULTADOS

5.1. Proceso de estimación y selección de modelos

Para la estimación de los 4 modelos empíricos planteados, las observaciones para cada compra se agregan al nivel de semana debido a la poca variabilidad de los precios y los descuentos diarios, obteniendo 53 observaciones de compras semanales para cada una de las 14 alternativas, que incluyen la información de los precios unitarios, precios regulares y descuentos.

Los dos modelos logit multinomial se estiman por el procedimiento de máxima verosimilitud. Una vez estimados los modelos, comparamos la bondad del ajuste correspondiente a cada uno de ellos, utilizando el método de comparación SBIC (Criterio de Información Bayesiana de Schwarz). Al igual que en el caso del criterio AIC (criterio de

Información de Akaike), el SBIC selecciona el modelo que considera que combina un valor alto del logaritmo de la función de verosimilitud con un número pequeño de parámetros estimados⁴. El modelo seleccionado en este trabajo, según esta conclusión, es el que presenta un menor valor para el estadístico SBIC. La definición para este criterio es⁵:

$$SBIC = -2[LL(B) - (K/2) Ln(T)] \quad [4.7]$$

Donde: LL(B): logaritmo de la función de verosimilitud de los parámetros estimados
 K: número de parámetros a estimar
 T: número de observaciones, total de unidades compradas (T=27.969)

Según este criterio, que presenta una conclusión similar al criterio AIC, el mejor modelo en el caso del logit multinomial es el modelo MNL(a), en el que se incluyen efectos separados para el precio regular y el porcentaje de descuento en precio (Cuadro 4).

CUADRO 4
Modelos MNL(a) y MNL(b). Ajuste

Modelo	Variables	LL(B)	T	K	AIC	SBIC
MNL(a)	R.P. P.D.	-62.561	27.969	15	125.152	125.275
MNL(b)	U.P.	-62.963	27.969	14	125.954	126.069

Para la estimación de los modelos de redes neuronales, en primer lugar se realiza una selección de la función de salida en la capa final y del método de entrenamiento. Una vez seleccionada la función de salida lineal y el método de entrenamiento de Levenberg-Marquardt se selecciona el número de unidades de la capa oculta para ambos

modelos mediante un proceso iterativo. Se parte de la estructura más sencilla, con una sólo una unidad oculta y progresivamente se añade complejidad hasta detenerse en la estructura con un valor medio menor para el error cuadrático en la estimación. Cada uno de los modelos se estima 10 rondas de forma que se obtienen valores para el

⁴ Los dos criterios (AIC y SBIC) tienen en común que consideran mejor modelo aquel que combina un valor alto de la función de verosimilitud con un número pequeño de parámetros. ALLENBY (1989) demuestra que, para muestras mayores de 8 observaciones el criterio SBIC es más adecuado que el AIC, aunque ambos criterios se muestran en los resultados.

⁵ La formulación utilizada en este trabajo para el SBIC es la de KAMAKURA et al, 1996. Puesto que las definiciones alternativas son crecientes o decrecientes respecto a las originales y dado que estas se utilizan para seleccionar modelos alternativos, la definición concreta no altera los resultados de la selección. La formulación del criterio AIC utilizada es la propuesta por KUMAR y SASHI (1989): $AIC = -2[LL(B) - k]$ donde LL(B): logaritmo de la función de verosimilitud de los parámetros estimados y K: número de parámetros a estimar.



promedio y la desviación típica de este error tanto en el conjunto de datos total como en el de entrenamiento y validación. En ambos casos, el número de unidades seleccionado es 5. También se observa como el error para el modelo (a) se sitúa en todos los casos por debajo del error en el modelo (b). En el Anexo se dan más detalles sobre este proceso de estimación

Una vez definida la estructura, en el cuadro 4 se compara el ajuste de ambos modelos. El criterio

de selección de modelos utilizado en este caso es el promedio del error cuadrático medio⁶ ya que esta es la función objetivo a minimizar utilizada durante la estimación.

$$PECM = \frac{\sum_1^N \sum_1^T (y_t - o_t)^2}{T} \quad [4.8]$$

Donde: N: Número de rondas de estimación

T: Número de observaciones

y_t: Valor estimado para la observación t

o_t: Valor observado para la observación t

CUADRO 5
Modelos RNA(a) y RNA(b). Ajuste

	<i>RNA(a)</i> <i>Número de unidades desde 1 hasta 20</i>		<i>RNA(b)</i>
Media	0,0011		0,0033
Desviación típica	0,0003		0,0017
Z Wilcoxon		-3,8242	
Sig		0,0001	
<i>Estructura: 5 unidades</i>			
PECM	0,0008		0,0009
Desviación típica	0,0001		0,0002
Valor		-1,6827	
Sig		0,0009	

La utilización del precio regular y el descuento (26 variables) en el precio en lugar del precio unitario (14 variables) aumenta el número de parámetros a estimar para la estructura con 5 unidades ocultas desde 159 hasta 229 parámetros. Tanto el test de Wilcoxon para muestras relacionadas a lo largo de todas las estimaciones como el test para las estimaciones con 5 unidades ocultas muestran como el modelo (a) es superior al (b). Por tanto, en ambos casos, el modelo elegido es aquel que utili-

za la información sobre el precio regular y el porcentaje de descuento en precios.

5.2. Resultados

El cuadro 5 muestra, como primera aproximación al ajuste de los modelos, las correlaciones de las estimaciones de los cuatro modelos con las cuotas de mercado semanales de todas las alternativas.

⁶ En este caso no se utiliza el criterio SBIC ya que, como señalan Qui y Zhang (2001), la penalización del número de parámetros que lleva asociado este criterio en el área de redes neuronales artificiales conlleva la selección del modelo más simple en todos los casos.



CUADRO 6
Correlaciones de Pearson. Cuotas de Mercado y Estimaciones

	Cuota	RNA(a)	RNA (b)	MNLogit(a)	MNLogit(b)
Cuota	1				
RNA(a)	0,945***	1			
RNA(b)	0,909***	0,923***	1		
Logit(a)	0,913***	0,930***	0,909***	1	
Logit(b)	0,901***	0,916***	0,899***	0,983***	1

Todas las correlaciones son significativas al 1%. En la primera columna se encuentran las correlaciones de los cuatro modelos con la cuota de mercado. El modelo con mayor correlación es la estimación mediante red neuronal que incluye los precios regulares y descuentos de cada alternativa, seguido del modelo logit con las mismas variables, es decir, los dos modelos seleccionados. El modelo de red con precios unitarios presenta una correlación mayor con las cuotas que el modelo logit con esta variable. Por tanto, en pri-

mer lugar, el modelo de red (a) parece realizar estimaciones aceptables para las cuotas de mercado.

El segundo objetivo del trabajo está relacionada con el tipo de efectos de los instrumentos de marketing (precio regular y descuento) en las cuotas de mercado y los efectos cruzados entre alternativas. El cuadro 7 muestra los parámetros estimados en el modelo logit multinomial elegido tomando como base la alternativa que tiene la menor cuota de mercado (Bahía natural).

CUADRO 7
Parámetros estimados

Alternativa	Coficiente	Valor t
1 Bonka mezcla	2,710	48,721
2 Marcilla mezcla	2,804	47,223
3 Marcilla natural	2,428	40,340
4 Bonka natural	2,246	39,471
5 Brasilia mezcla	2,388	43,239
6 Saimaza mezcla	1,934	29,034
7 154 mezcla	1,070	17,393
8 Soley mezcla	1,689	29,302
9 Marcilla especial	1,370	21,278
10 Saimaza natural	1,298	18,431
11 Bonka especial	3,038	32,959
12 Bahía mezcla	0,420	6,751
13 154 natural	0,299	4,692
Precio Regular	-0,022	-24,863
Descuento	10,795	96,042



De acuerdo con la especificación del modelo, se calcula un único coeficiente para todas las alternativas, para el precio regular y el porcentaje de descuento. Todos los coeficientes estimados son significativos al 1% y tienen el signo esperado. Los valores constantes para la utilidad de cada alternativa (estimaciones de α_i) pueden interpretarse como valores de preferencia para cada alternativa. La alternativa con mayor valor (3,038) es la variedad especial de Bonka. Los valores para el resto de las constantes tienen una ordenación de mayor a menor similar a las de las cuotas de mercado. Para comprobar si se mantiene o no la propiedad de independencia de las alternativas irrelevantes, el modelo logit se vuelve a estimar con sólo una parte de las alternativas, las resultantes de eliminar las variedades especiales y las marcas que sólo presentan una variedad (8 alternativas). Los parámetros del nuevo modelo se comparan con el anterior mediante el test de HAUSMAN Y

MCFADDEN (1984) que rechaza la hipótesis nula de independencia de las alternativas irrelevantes⁷.

En el caso del modelo RNA(a), los efectos de las variables independientes se evalúan mediante el cálculo de derivadas de las 14 variables dependientes con respecto a las independientes (ver Anexo para las expresiones de las derivadas) además de los valores constantes para cada una de estas variables. En el cuadro 8, se muestran estos valores constantes, además del efecto del precio regular y descuento de cada alternativa en la cuota estimada para esa alternativa evaluado en los valores medios de estas variables. Los valores de las constantes, en este caso, no son comparables a los obtenidos en el modelo logit multinomial. Mientras que los parámetros estimados en el caso de logit muestran el efecto constante de cada alternativa en la función de utilidad de los individuos, los parámetros para la red son efectos constantes directamente sobre las cuotas de mercado.

CUADRO 8
RNA(a). Constantes y efectos propios medios

Alternativa	Constante	Propio PR (efecto medio)	Propio dto (efecto medio)
1 Bonka mezcla	0,314	-0,008	0,026
2 Marcilla mezcla	0,400	-0,000	0,920
3 Marcilla natural	0,347	-0,001	0,747
4 Bonka natural	0,203	-0,001	0,923
5 Brasilia mezcla	-0,023	-0,009	—*
6 Saimaza mezcla	-0,299	0,052	0,461
7 154 mezcla	0,384	-0,029	0,193
8 Soley mezcla	-0,340	0,007	—*
9 Marcilla especial	0,083	-0,001	0,155
10 Saimaza natural	-0,121	-0,025	0,378
11 Bonka especial	0,058	-0,001	0,573
12 Bahía mezcla	-0,111	-0,003	0,021
13 154 natural	0,153	-0,002	0,232
14 Bahía natural	-0,048	-0,001	0,004

* Las alternativas 5 y 8 no se promocionan durante el periodo.

⁷ El resultado del test en este caso es 168,79 que sigue una distribución chi-cuadrado con 6 grados de libertad (la diferencia entre 14 y 8 alternativas) por lo que se rechaza la hipótesis nula de independencia de alternativas irrelevantes al 1%.



Las marcas Bonka y Marcilla muestran los mayores valores para estas constantes con valores positivos, al igual que la marca 154 (una marca local). En estos casos, las cuotas que obtienen las alternativas son superiores a las que se esperarían por los efectos de los precios regulares y descuentos en cada momento. En el caso de las alternativas con constantes negativas, la cuota efectiva estimada es menor a la explicada mediante los efectos de los precios regulares y porcentajes de descuentos.

El resto de los parámetros presentados en el cuadro 8 corresponden a los efectos medios de precios y descuentos de una alternativa en la cuota estimada para esa alternativa. Cada valor indica la modificación esperada en la cuota de la alternativa para una modificación en uno de los instrumentos de marketing partiendo del valor medio (se ha tomado la mediana). En el caso de los precios regulares, todos los parámetros estimados, salvo dos casos, presentan el signo esperado. Estos dos casos con signos inesperados se corresponden con la variedad mezcla de las marcas Saimaza y Soley (alternativas 6 y 8). Una posible explicación para este fenómeno es la alta correlación de estos dos precios regulares con el precio regular de la alternativa 10, Saimaza natural (0,899 y 0,999 respectivamente); cuando el precio de la alternativa 10 aumenta, las cuotas de mercado de las alternativas 6 y 8 aumentan al mismo tiempo que sus precios. Como se esperaba, el efecto de aumentos en los precios regulares en la cuota de mercado es menor en el caso de Bonka y Marcilla salvo en el caso de la alternativa 1 que se sitúa en valores intermedios. Por tanto, puede decirse que las marcas nacionales se enfrentan a una menor sensibilidad a los precios que el resto de las marcas.

Los efectos del porcentaje de descuento en precios, muestran el signo positivo esperado. El porcentaje de descuento tiene el mayor efecto en el caso de las variedades mezcla y natural de Marcilla y las variedades mezcla y especial de Bonka. Por tanto, en el caso de estas marcas nacionales, las variaciones de los precios regulares tienen poca repercusión sobre las cuotas de mercado mientras que las promociones en precios tienen fuertes efectos. Sin embargo, las promociones en precios tienen muy poco efecto en el caso de Bahía para las variedades mezcla y natural y para el café especial de Marcilla. En el caso de esta alternativa (alternativa 9), el efecto del precio regular también es muy bajo, por lo que el efecto de la constante en esta alternativa es muy alto. Esta alternativa es un café especial para filtros lo que podría hacer que una relativa situación de monopolio para un nicho de consumidores. Las dos variedades de Bahía no experimentan prácticamente variaciones en cuota ante promociones de precios.

El siguiente paso es detallar los resultados de la red neuronal con respecto al porcentaje de descuento en precios. Desde el punto de vista detallista, la toma de decisiones sobre promociones en precios resulta un instrumento más rápido y sencillo de utilizar que las modificaciones en el precio regular. La estructura de red analizada permite detallar los efectos de estos instrumentos tanto para distintos puntos de valor (distintos efectos dependiendo del nivel de partida) como analizar los efectos cruzados entre distintas alternativas.

En el cuadro 9 y las figuras 2(a) y 2(b) se muestran las derivadas de las cuotas de cada alternativa con respecto a su propio porcentaje de descuento medio evaluadas en distintos puntos.



CUADRO 9
RNA(a). Porcentaje Descuento Propio

Alternativa	0%	0-5%	5-10%	10-15%	15-20%	>20%
1 Bonka mezcla	0,543	0,541	0,026	0,209	0,007	–
2 Marcilla mezcla	0,381	0,698	0,920	0,925	0,127	0,138
3 Marcilla natural	0,186	0,624	0,747	0,962	0,034	0,065
4 Bonka natural	0,294	0,373	0,923	0,946	0,821	–
6 Saimaza mezcla	0,351	0,457	0,394	0,461	0,473	0,204
7 154 mezcla	–	–	0,200	0,193	0,056	0,050
9 Marcilla especial	0,029	0,058	0,155	0,162	–	–
10 Saimaza natural	0,182	–	0,199	0,378	0,370	0,170
11 Bonka especial	0,548	0,573	0,515	–	–	–
12 Bahía mezcla	-0,019	–	–	0,021	-0,001	–
13 154 natural	0,368	0,179	0,232	0,131	–	–
14 Bahía natural	0,022	–	–	–	–	0,004

La observación de ambas figuras da una primera idea de que el tipo de respuesta ante descuentos de precios es diferente entre marcas. En algunos casos, parece que el efecto del descuento puede ser de tipo lineal, como en el caso de las alternativas 9, 11, que son las dos variedades especiales de las Bonka y Saimaza o las alternativas Saimaza y 154 mezcla. En otros casos, el efecto de las promociones de precios en las cuotas es demasiado bajo como para extraer conclusiones, como en el caso de las alternativas Bahía mezcla y natural. Sin embargo, para otras alternativas, el efecto de las promociones en precios desagregado por niveles presenta una forma claramente no lineal. En general, los gráficos presentan dos tipos de efectos. En el caso de las alternativas Bonka mezcla y 154 natural, el valor mayor para la derivada del precio se encuentra en el nivel de “no descuento” y disminuye a partir de este nivel. En estos casos,

la ganancia de la cuota de mercado parece conseguirse por el hecho de realizar la promoción y el efecto marginal del volumen de descuento es cada vez menor. Sin embargo, los efectos marginales del descuento para las alternativas Marcilla mezcla y natural y Bonka natural siguen una tendencia diferente. En estos tres casos, el efecto de un aumento del descuento desde el nivel 0 es bastante bajo y este efecto aumenta progresivamente hasta llegar al máximo para descuentos entre el 10 y el 15%. A partir de este punto, el efecto marginal del descuento disminuye bruscamente en el caso de las alternativas 2 y 3, y más moderadamente en el caso de la alternativa 4. Para estas alternativas, parece existir un nivel de descuento óptimo para el que el efecto marginal es máximo y a partir del cual se produce un efecto saturación. Por tanto, el efecto de las promociones en precios no es homogéneo entre alternativas.



FIGURA 2(a)
RNA(a): Efectos porcentaje de descuento, alternativas 1 a 7

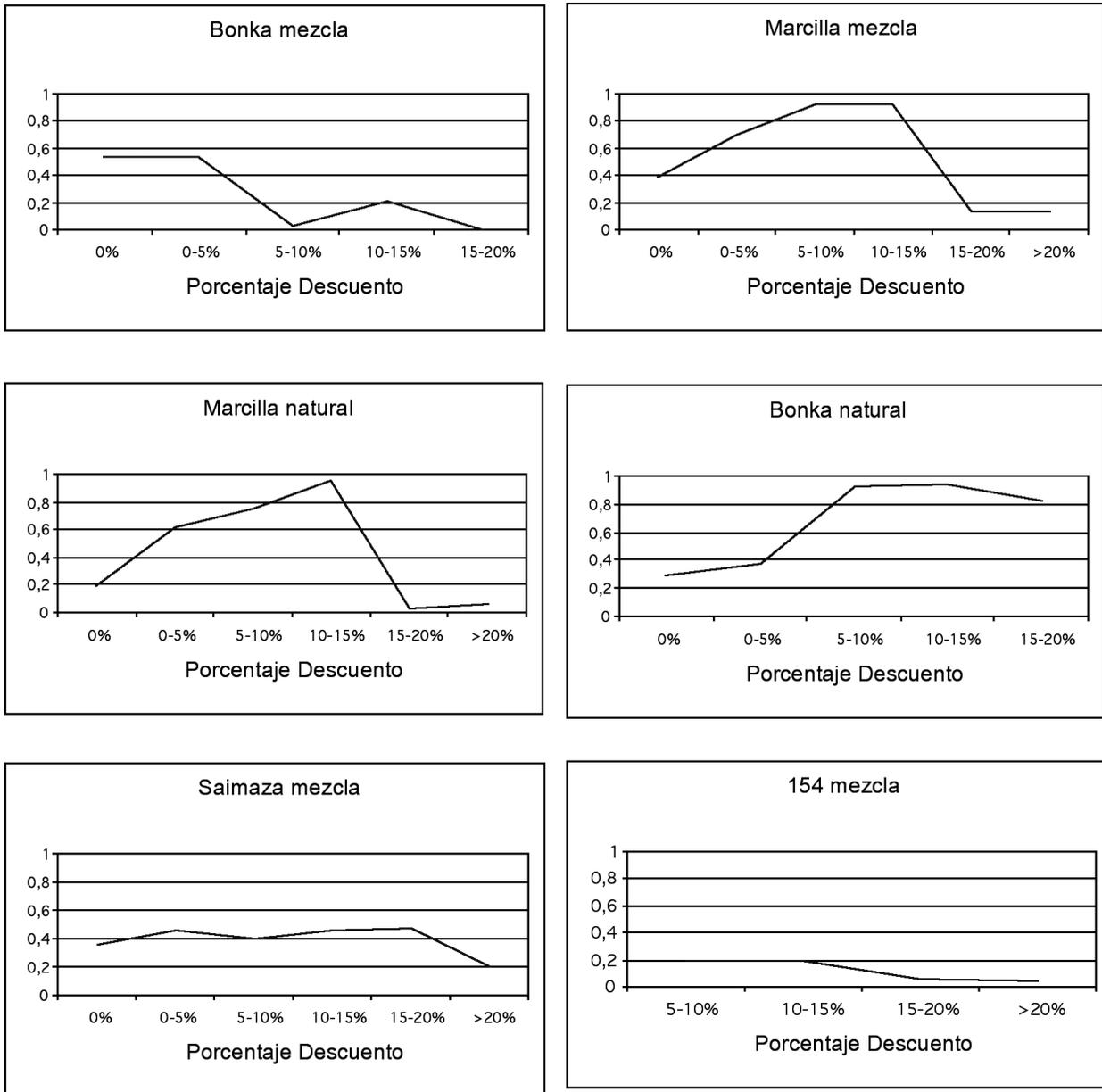
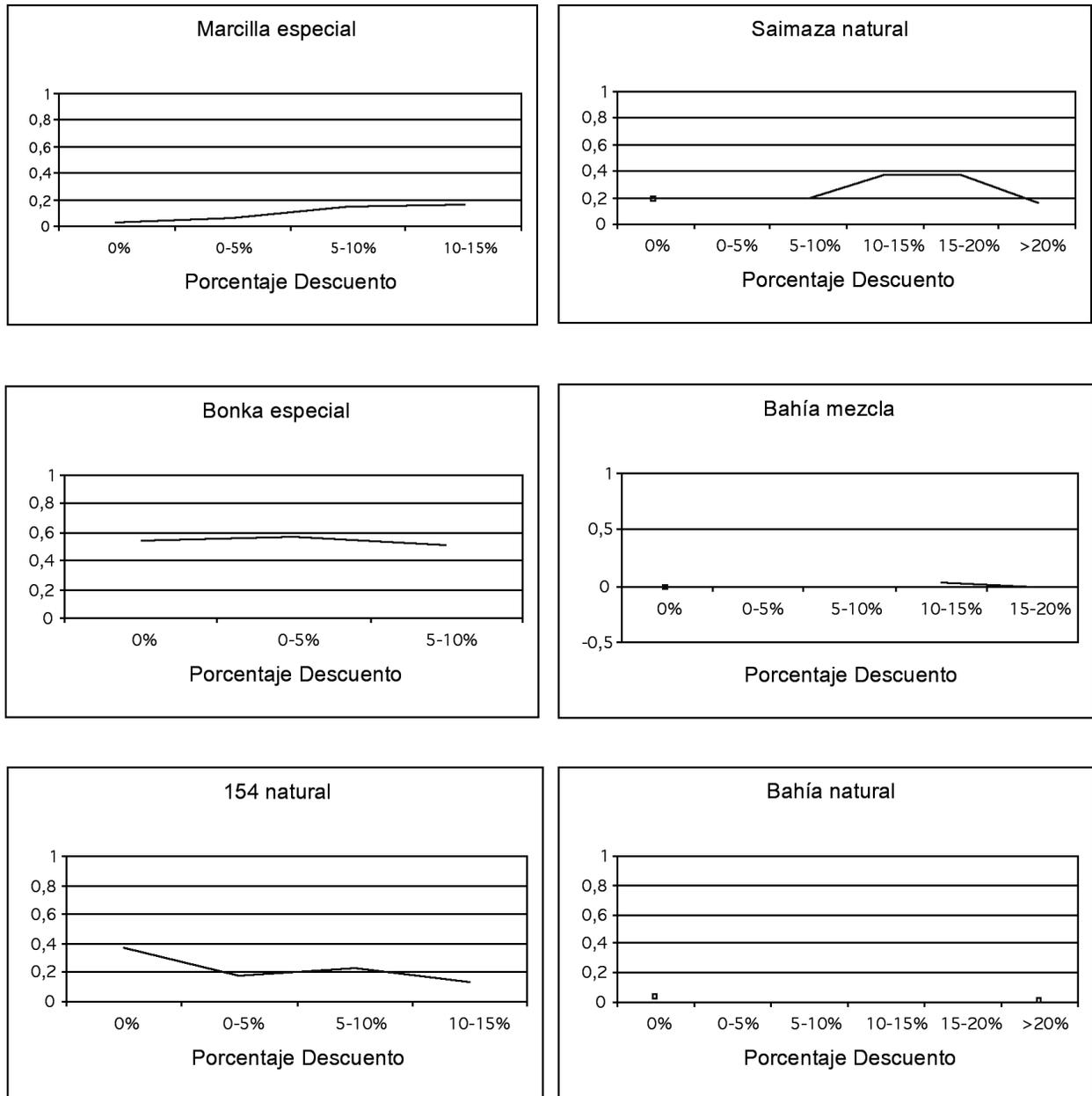


FIGURA 2(b)
RNA(a): Efectos porcentaje de descuento, alternativas 8 a 14



Por último, se presentan algunos resultados referidos a los efectos cruzados y de interacción entre marcas. Como se señalaba al describir la base de datos, en este caso, en general, las promociones se establecen para las marcas y no para las

alternativas. En el cuadro 10 se presentan, los efectos cruzados de los descuentos agregados para las marcas, sobre las cuotas de las alternativas 1 y 2 (las alternativas con mayor cuota)⁸.

CUADRO 10
RNA(a). Efectos agregados por marcas. Alternativas 1 y 2

Descuento (suma)	Cuota Alt. 1 (Bonka mezcla)	Cuota Alt.2 (Marcilla mezcla)
Bonka	1,3617	-0,6879
Marcilla	-0,6414	1,7174
Saimaza	-0,7616	-0,2306
Bahía	-0,2988	0,1351
154	-0,0072	-0,046

En ambos casos, el efecto de descuentos en la propia marca es claramente positivo. Sin embargo, los descuentos en el resto de las marcas no afectan por igual a ambas marcas. Mientras que en el caso de la alternativa 1 (Bonka) los descuentos de Marcilla y Saimaza tienen efectos similares, la cuota de la alternativa 2 (Marcilla) se ve mucho más afectada por descuentos en la marca Bonka que en la marca Saimaza. La marca 5 (Bahía) afecta de forma positiva a la cuota de la alternativa 2 debido a que, para las 18 semanas en que esta alternativa está en promoción, la correlación entre ambos descuentos es muy alta (0,974). Sin embargo, esta

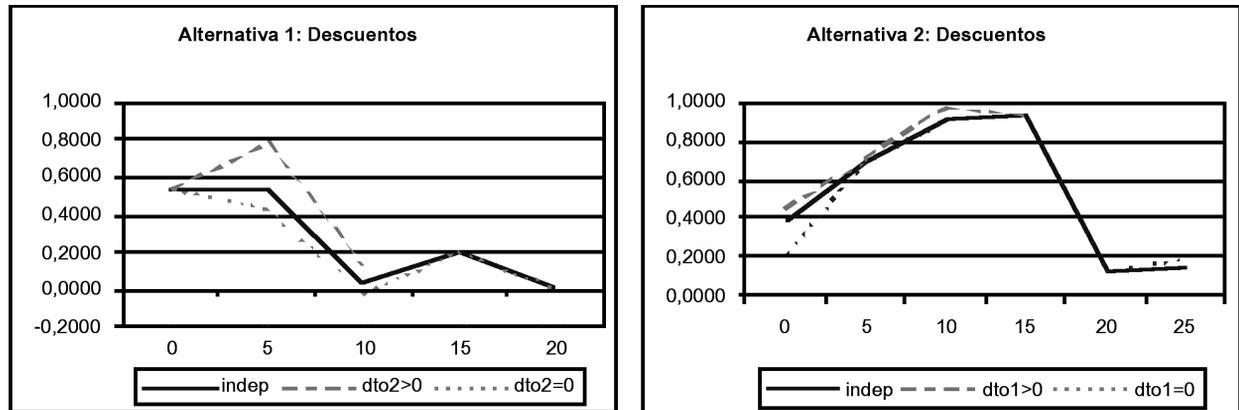
marca tiene un efecto bajo en la cuota de la alternativa 1. Los descuentos en la marca 154 afectan en menor medida a la cuota de las dos alternativas aunque el efecto es superior para la alternativa Marcilla.

Para terminar, se analizan los efectos de interacción para las dos alternativas con mayor cuota (Bonka y Marcilla mezcla). En este caso, se comprueba el efecto de un descuento en precio en la propia alternativa dependiendo de que la otra se promocioe o no simultáneamente. En la figura 3 se muestran para las alternativas 1 y 2, los efectos de interacción en promociones en la marca.

⁸ En este caso, no se han calculado los efectos en cada punto, que pueden o no ser lineales, para facilitar el análisis y se presentan únicamente los resultados que corresponden al punto medio de descuento, por lo que estos resultados serían comparables a los que aparecen en la tabla 8.



FIGURA 3
Interacciones de descuentos. Alternativas 1 (Bonka mezcla) y 2 (Marcilla mezcla)



En el caso de la alternativa 1 (Bonka mezcla), el análisis inicial mostraba cómo el mayor efecto marginal se producía por el hecho de introducir la promoción y cómo el incremento del descuento afectaba en menor medida a la cuota de mercado. Al introducir el efecto del descuento en la alternativa 2 (Marcilla Mezcla), estas conclusiones se modifican. Cuando las alternativas 1 y 2 se promocionan simultáneamente, el efecto inicial de la promoción sigue siendo similar, pero el efecto en la cuota aumenta al incrementar el valor del descuento. En el caso de la alternativa 2, el efecto en la cuota de mercado al introducir una promoción en precios, aunque el porcentaje de descuento sea muy bajo, se incrementa considerablemente cuando la alternativa 1 está siendo promocionada simultáneamente. Sin embargo, a partir de estos bajos niveles, las ganancias en cuota al incrementar el volumen de descuentos son similares independientemente de que ambas alternativas se promocionen al mismo tiempo.

6. CONCLUSIONES

La primera conclusión que se puede extraer de este trabajo se refiere a la información que proporciona el precio de un producto al explicar la capacidad de elección. En el caso de los cuatro

modelos empíricos planteados en este trabajo, los resultados son mejores cuando se divide el precio en dos componentes: el precio regular y el porcentaje de descuento sobre este precio que cuando únicamente se proporciona información sobre el precio unitario. Parece existir un efecto promoción que es capaz de explicar más que la simple comparación de los precios a pagar por el comprador independientemente del modelo utilizado para reflejar esta relación. Los consumidores habituales en el establecimiento pueden tener un sistema de precios de referencia que puedan comparar fácilmente los precios actuales con este precio referencial para valorar el porcentaje de descuento efectivo.

En segundo lugar, el modelo MFANN presenta unos resultados de ajuste adecuados para modelizar las cuotas de mercado semanales en un establecimiento en una categoría de producto compleja, los resultados son superiores a los del modelo logit multinomial, modelo ampliamente contrastado en este contexto. Por tanto, este tipo de RNA resulta útil en el contexto de elección dentro de una categoría de producto aunque sería necesario obtener una mayor cantidad de datos y evaluar la capacidad predictiva del modelo para obtener conclusiones definitivas sobre este aspecto.

La superioridad del ajuste de las redes neuronales frente a técnicas estadísticas clásicas es un



resultado encontrado con frecuencia en la literatura. Estas técnicas permiten la estimación de funciones altamente no flexibles por lo que, en muchas ocasiones, superan las estimaciones de modelos más restringidos. Sin embargo, el alto número de parámetros estimado en estos modelos hace que en ocasiones hayan sido contemplados como “cajas negras” adecuados en problemas de predicción pero de utilidad reducida en la explicación de respuestas de mercado. Aunque la definición de la forma funcional estimada por el modelo es compleja, es posible estimar los efectos en los resultados de cada una de las variables explicativas y este análisis puede indicar las causas de la superioridad en la estimación. En este trabajo se ha utilizado este enfoque para, mediante la evaluación de derivadas de los resultados en la red, obtener una visión más detallada de las relaciones entre variables dependientes e independientes.

En primer lugar, se analizan las relaciones entre los instrumentos de promoción utilizados en cada alternativa y sus cuotas semanales. Se comprueba como los efectos no son homogéneos por alternativas. Aunque no disponemos de intervalos de confianza para las derivadas por niveles, el análisis gráfico muestra para las alternativas con mayores cuotas dos tipos de efectos. En primer lugar, de modo similar a los resultados de VAN HEERDE et al (2001), en algunas alternativas parece existir un nivel umbral alrededor del 5% a partir del cual el efecto marginal del descuento aumenta, y un nivel alrededor del 15% en el que este efecto marginal se hace máximo. En estos casos, es necesario realizar un nivel mínimo de promoción en descuento para que la promoción sea efectiva. En otras ocasiones, de forma similar a los resultados de AGRAWAL y SCHORLING (1996), el efecto marginal máximo en la cuota se produce para los niveles de descuento más bajos, a partir de los cuales el efecto disminuye. En estos casos, parece existir un efecto promoción de precios frente a ausencia de promoción, por el que los compradores reaccionan a la existencia o no de una promoción, independientemente del nivel de descuento de la misma.

Las implicaciones de estos resultados para la política de promociones son claras: el porcentaje de descuento aplicado para cada alternativa debe tener en cuenta estos factores. Para aquellas alternativas en las que el efecto marginal del descuento aumenta con el porcentaje de descuento la fijación de un descuento mayor tiene un efecto superior que en aquellos casos en los que el efecto de la promoción es de tipo umbral, en los que por el hecho de introducir la promoción se consiguen los mayores efectos.

En segundo lugar, se analizan tanto los efectos cruzados entre las distintas alternativas como los efectos de interacción entre instrumentos. Respecto a los efectos cruzados entre alternativas, la simultaneidad entre promociones de distintas variedades para una misma marca, limita este análisis a efectos entre marcas. Los efectos en el punto medio indican que, como cabe esperar, para el caso de las dos alternativas presentadas, las promociones de la propia marca afectan positivamente a su cuota mientras que el resto lo hace de forma negativa. Sin embargo, aunque las dos alternativas correspondan a la variedad mezcla de las marcas Bonka y Marcilla, los efectos del resto de las marcas no se reparten de forma similar. La cuota de una la alternativa mezcla de Bonka se ve afectada de forma más homogénea por los descuentos en el resto de las marcas, mientras que en el caso de la segunda (variedad mezcla de Marcilla), la marca Bonka tiene la mayor influencia en su cuota de mercado. En el primer caso, la alternativa tiene un mayor número de competidores directos, mientras que en el segundo, la alternativa tiene una directa competidora y se ve menos afectada por el resto de los descuentos. Esto puede indicar que los compradores de estas dos marcas las contemplan de forma diferente. Mientras que, en un caso, los compradores habituales pueden ser más sensibles al precio y estar dispuestos a cambiar de marca cuando se realiza una promoción de otra alternativa, en el otro, los compradores habituales pueden percibir una única alternativa como posible sustituto y reaccionar únicamente ante las promociones de esta alternativa.



Además, se analizan los efectos de interacción entre los instrumentos promocionales de las alternativas Bonka y Marcilla mezcla. La relación entre el descuento y cuota de mercado para estas alternativas se ve modificada en función de que, simultáneamente, se realice o no una promoción de su principal competidora. Para la primera alternativa, si la segunda no realiza una promoción, el mayor efecto marginal en la cuota se produce al realizar un pequeño descuento, pero si la segunda alternativa se promociona al mismo tiempo, el efecto marginal aumenta al aumentar el volumen de descuento. En el caso de la segunda alternativa, si la primera alternativa no se promociona, el efecto de introducir un pequeño porcentaje de descuento es bajo y se consigue un alto rendimiento marginal en la cuota para niveles de descuento superiores al 10%. Sin embargo, cuando la primera alternativa se encuentra en promoción, el efecto marginal en la cuota de introducir un porcentaje bajo de descuento aumenta. Según estos resultados, la política de promociones óptima para ambas alternativas tiene que tener en cuenta las promociones en la otra y, además, esta influencia no es simétrica entre alternativas. Por tanto, la política de promociones no puede establecerse de forma independiente para cada alternativa, la competencia entre marcas no es simétrica entre alternativas y la existencia de promociones en otras alternativas pueden afectar al nivel de estos efectos umbral.

La principal limitación del trabajo se relaciona con la ausencia de criterios para valorar la significatividad de los efectos individuales de las variables. Como líneas futuras de investigación y para superar estas limitaciones, se plantea la posibilidad de combinar esta técnica con otras técnicas paramétricas introduciendo alguno de los efectos aquí analizados en un modelo paramétrico a contrastar o evaluar la posibilidad de desarrollar intervalos de confianza empíricos a través de técnicas como el bootstrapping. La segunda limitación se relaciona con el número total de datos, sería deseable disponer de una mayor cantidad de observaciones semanales que permita valorar también la capacidad predictiva del modelo.

7. REFERENCIAS

- AGRAWAL D. y SCHORLING, C. (1996), "Market Share Forecasting: An Empirical Comparison of Artificial Neural Networks and Multinomial Logit Model", *Journal of Retailing*, Vol. 72(4), pp. 383-407.
- AINSCOUGH T.L. y ARONSON, J.E. (1999), "An Empirical Investigation and Comparison of Neural Networks and Regression for Scanner Data Analysis", *Journal of Retailing and Consumer Services*, Vol. 6(4), pp. 205-217.
- ALLENBY, G.M. (1989), "A Unified Approach to Identifying, Estimating and Testing Demand Structures with Aggregate Scanner Data", *Marketing Science*, Vol. 8(3), pp. 265-280.
- BEJOU, D., B. WRAY, B. y INGRAM, T.N. (1996), "Determinants of Relationship Quality: An Artificial Neural Network Analysis", *Journal of Business Research*, Vol. 36(2), pp. 137-143.
- BEN-AKIVA, M. y LERMAN, S.R. (1985), *Discrete choice analysis: theory and application to travel demand*. MIT Press.
- BENTZ, Y. y MERUNKA, D. (2000), "Neural Networks and the Multinomial Logit for Brand Choice Modelling: A Hybrid Approach", *Journal of Forecasting*, Vol. 19, pp. 177-200.
- BERNÉ, C, CORTIÑAS, M., ELORZ, M y MÚGICA, J.M. (2004), "The Use of a Retail Store Database for Brand Choice Analysis", *The International Review of Retail, Distribution and Consumer Research*, Vol. 14(1), pp. 19-29.
- BOONE, D.S. y ROEHM, M (2002), "Retail Segmentation using Artificial Neural Networks", *International Journal of Research in Marketing*, Vol. 19(3), pp. 287-301.
- BUCKLIN, R.E. y GUPTA, S. (1999), "Commercial Use of UPC Scanner Data: Industry and Academic Perspectives", *Marketing Science*, Vol. 18(3), pp. 247-273.
- COOPER, L.G. y NAKANISHI, M. (1988), *Market Share Analysis. Evaluating Competitive Marketing Effectiveness*, Kluwer Academic Publishers.
- CURRY, B. y MOUTINHO, L. (1993), "Neural Networks in Marketing: Modelling Consumer



Response to Advertising Stimuli”, *European Journal of Marketing*, Vol. 27(7), pp. 5-26.

DASGUPTA, C.G., DISPENSA, G.S. y GHOSE, S. (1994), “Comparing The Predictive Performance of a Neural Network Model with Some Traditional Market Response Models”, *International Journal of Forecasting*, Vol. 10, pp. 235-344.

DAVIES, F., M. GOODE, J. MAZANEC Y MOUTINHO L. (1999), “LISREL and Neural Network Modelling: Two Comparison Studies”, *Journal of Retailing and Consumer Services*, Vol. 6, pp. 249-261.

DAVIES, F.M., M. GOODE, L. MOUTINHO Y OGBONNA. E. (2001), “Critical factors in Consumer Supermarket Behavior: A Neural Network Approach”, *Journal of Consumer Behavior*, Vol. 1(1), pp. 35-49.

DE CARVALHO, M.C.M., DOUGHERTY, M.S., FOWKES, A.S. y WARDMAN, M.R. (1998), “Forecasting Travel Demand: A Comparison of Logit and Artificial Neural Networks Methods”, *Journal of the Operational Research Society*, Vol. 49, pp. 717-722.

DISTRIBUCIÓN ACTUALIDAD. (2003), “Hipermercados en España”, *Distribución Actualidad*, Vol. 319, Octubre.

GRUCA, T.S., KLEMEZ, B.R. y PETERSEN, E.A.F. (1995), “Inside the Black Box: Neural Network Modeling of Aggregate Market Response”, *Working Paper*, Abril, University of Iowa.

GUPTA, S. (1988), “Impact of Sales Promotions on When, What and How Much to Buy”, *Journal of Marketing Research*, Vol. 25, pp. 342-355.

HAUSMAN, J. y MCFADDEN, D. (1984), “Specification Tests for the Multinomial Logit Model”, *Econometrica*, Vol. 52, pp.1219-1240.

HEERDE, H.J. VAN, LEEFLANG, P.S. y WITTINK, D.R. (2001), “Semiparametric Analysis of the Deal Effect Curve”, *Journal of Marketing Research*, Vol. 38(2), pp. 197-216.

HILERA, J.R. y MARTÍNEZ, V.J. (1995), *Redes Neuronales Artificiales, Fundamentos, Modelos y Aplicaciones*, Ed. Ra-Ma, Madrid.

HOCH, S.J., KIM, B., MONTGOMERY, A.L. y ROSSI, P.E. (1995), “Determinants of Store-Level Pri-

ce Elasticity”, *Journal of Marketing Research*, Vol. 32, pp. 17-29.

HORNICK K., M. STINCHCOMBE Y WHITE, H. (1989), “Multilayer Feedforward Networks Are Universal Approximators”, *Neural Networks*, Vol. 2, pp. 359-366.

HRUSCHKA, H. (1993), “Determining Market Response Functions by Neural Network Modeling: A Comparison to Econometric Techniques”, *European Journal of Operational Research*, Vol. 66, pp. 27-35.

HU, M.H., SHANKER, M Y HUNG., M.S (1999), “Estimation of Posterior Probabilities Of Consumer Choices With Neural Networks Classifiers”, *International Journal Of Research In Marketing*, Vol. 16, pp. 307-317.

ILASHEM, S. (1992), “Sensitivity Analysis for Feedforward Artificial Neural Networks with Differentiable Activation Functions” en *Proceedings of the International Joint Conference on Neural Networks*, Baltimore, IEEE, New Jersey, vol I, 419-424.

KAMAKURA, W.A., KIM, B.D. y LEE, J. (1996), “Modelling Preference and Structural Heterogeneity in Consumer Choice”, *Marketing Science*, Vol. 10(2), pp. 172-183.

KRYCHA, K.A. y WAGNER, U. (1999), “Applications of Artificial Neural Networks in Management Science: A Survey”, *Journal of Retailing and Consumer Services*, Vol. 6, pp. 185-203.

KUMAR, A., RAO, R. y SONI, H. (1995), “An Empirical Comparison of Neural Network and Logistic Regression Models”, *Marketing Letters*, Vol. 6(4), pp. 251-263.

KUMAR, A. y SASHI, C.M. (1989), “Confirmatory Analysis of Aggregate Hierarchical Structures: Inferences from Brand Switching Behavior”, *Journal of Marketing Research*, Vol 26, pp. 444-453.

KUNG, C.M. y WHITE, H. (1992), “Artificial Neural Networks: An Econometric Perspective”, *Econometric Reviews*, Vol. 13(1), pp. 1-91.

LUQUE MARTÍNEZ, T. (2003), *Nuevas Herramientas de Investigación de Mercados*, Ed. Civitas Ediciones, Madrid.



McFADDEN, D. (1986), "The Choice Theory Approach to Market Research", *Marketing Science*, Vol. 5(4), pp. 275-297.

MAZANEC, J. (1992), "Classifying Tourists into Marketing Segments: A Neural Network Approach", *Journal of Travel and Tourism Management*, Vol. 1, pp. 39-59.

MONTGOMERY, A.L. (1987), "Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data", *Marketing Science*, Vol. 16(4), pp. 315-317.

OLMEDA, I. (1996), "Modelos Discriminantes para la Decisión de Introducción de Nuevos Productos por Distribuidores", Actas VII Encuentro de Profesores de Marketing, Zaragoza.

QI, M. y ZHANG, G.P. (2001), "An Investigation of Model Selection Criteria for Neural Network Time Series Forecasting", *European Journal of Operational Research*, Vol. 132, pp. 666-680.

REINECKE FLIN, L., EASTMAN, J.K. y NEWELL S.J. (1995) "An Exploratory Study of the Application of Neural Networks to Marketing: Predicting Rock Music Shopping Behaviour", *Journal of Marketing Theory and Practice*, Vol. 3(2), pp. 75-85.

ROBERTS, J.H. y LILIEN, G.L. (1993), "Explanatory and Predictive Models of Consumer Behaviour", en *Marketing*, ed. J. Eliashberg y G.L. Lilien, North-Holland. Amsterdam.

RUMMELHART, DAVID E., JAMES L. MCCLELLAND & THE PDP GROUP. (1986), *Parallel Distributed Processing: Explorations In The Micro-*

structure Of Cognition, Ed. The Mit Press, Cambridge.

WARTENBERG, F. y DECKER, R. (1996), "Analysis Of Sales Data, A Neural Net Approach", En *From Data To Knowledge*, Gaul, W. y D. Pfeifer Eds., Springer, Berlin

WEST, P.M., BROCKET, P.L. y GOLDEN, L.L. (1997), "A Comparative Analysis of Neural Networks and Statistical Methods for Predicting Consumer Choice", *Marketing Science*, Vol. 16(4), pp. 370-391.

WIERENGA, B. y KLUYTMANS, J. (1994), "Neural Nets versus Marketing Models in Time Series Analysis: A Simulation Study" en *Proceedings of the 23th Annual Conference of the European Marketing Academy (EMAC 94)*, Maastricht.

WIERENGA, B. y KLUYTMANS, J. (1996), "Prediction with Neural Nets in Marketing Time Series Data", *Working Paper, Management Report Series*, nº 258, Erasmus University of Rotterdam.

WRAY, B., PALMER, A. y BEJOU, D. (1994), "Using Neural Network Analysis to Evaluate Buyer-Seller Relationship", *European Journal of Marketing*, Vol. 28(10), pp. 32-48.

ZAHAVI, J. y LEVIN, N. (1997a), "Applying Neural Computing to Target Marketing", *Journal of Direct Marketing*, Vol. 11(1), pp. 5-22.

ZAHAVI, J. y LEVIN, N. (1997b), "Issues and Problems in Applying Neural Computing to Target Marketing", *Journal of Direct Marketing*, Vol. 11(4), pp. 63-75.



ANEXO. IMPLEMENTACIÓN DE LA RED MULTICAPA ALIMENTADA HACIA DELANTE

Para la aplicación empírica de este trabajo se seleccionó en primer lugar el tipo de RNA a utilizar. Las redes MFANN fueron las seleccionadas por ser las más utilizadas en la literatura para el tratamiento de este tipo de problemas. Para ajustar esta red, se necesita tomar diversas decisiones relativas a su implementación. En primer lugar, es necesario seleccionar un método de entrenamiento. El algoritmo clásico es el denominado método de retropropagación de errores pero este método puede ser bastante lento. Dos de los métodos de estimación alternativos son el algoritmo de Levenberg-Marquardt y el algoritmo de gradiente conjugado de escala. En segundo lugar, existen diversas alternativas para las funciones de transferencia en las capas ocultas y de salida (HILERA y MARTÍNEZ, 1995). En este caso, para la función de la capa de salida se plantearon 2 alternativas: la función logística o lineal. La función lineal presenta la ventaja de aproximar la estructura de la red a la del logit multinomial haciendo más comparables los modelos y la ventaja adicional de simplificar el cálculo de derivadas para interpretar los resultados de la red pero puede afectar a la calidad de las estimaciones. Por último, es necesario seleccionar el número de capas ocultas y el número de unidades en cada capa que tendrá cada una de las redes. El número de unidades en una capa oculta es el que permite asegurar las propiedades de aproximación de la MLANN para el caso de funciones continuas

(HORNICK, STINCHCOMBE y WHITE, 1989), pero cuanto mayor sea este número mayor será el número de parámetros a estimar. En el caso de que existan discontinuidades en la función que relaciona las entradas con las salidas puede ser necesario introducir capas ocultas adicionales pero en este caso no se ha considerado esta posibilidad para facilitar el cálculo de derivadas y la interpretación de los resultados.

Para seleccionar estos aspectos, se realiza un proceso experimental, partiendo de una red inicial con la estructura más sencilla y se estiman sucesivamente redes más complejas combinando un número de unidades ocultas creciente, los dos métodos de entrenamiento alternativos y las dos alternativas de función de salida (WEST, BROCKET, y GOLDEN, 1997). El aprendizaje de las redes se detiene por el método de validación cruzada. Para cada combinación de método de entrenamiento, función de salida y número de unidades se estiman 10 redes neuronales en las que se para las que se calcula el error cuadrático medio y su desviación típica. El paquete informático utilizado para las estimaciones es el módulo de RNA de Matlab. Las estimaciones de cada combinación “método de entrenamiento-función de salida-número de unidades” se comparan en cuanto a este error cuadrático medio y se utiliza la estructura que genera un error menor. El cuadro 11 resume las decisiones de formulación.



CUADRO 11
Decisiones de formulación

Estructura		
Tipo	Multicapa Alimentada Hacia Delante	
	Modelo (a)	Modelo (b)
Número de capas	3	3
Unidades en cada capa		
Capa de entrada	14 precios regulares y 12 descuentos	14 precios
Capa oculta	5	5
Capa de salida	14 cuotas	14 cuotas
Funciones		
Función capa oculta	Logística	Logística
Función salida	Lineal	Lineal
Entrenamiento		
Algoritmo aprendizaje	Levenberg-Marquardt	
Parámetros		
Valor μ inicial	0.001	0.001
Incremento μ	10	10
Disminución μ	0.1	0.1
Función de ajuste	Error Cuadrático Medio	Error Cuadrático Medio
Validación cruzada		
Conjunto de entrenamiento	70% del conjunto de datos	
Conjunto de validación	30% del conjunto de datos	

Por último, una vez estimada la red y en lo que se refiere a la interpretación de los resultados, algunos autores han utilizado medidas de resúmenes de pesos como el máximo efecto de los pesos para cada variable (KUMAR et al, 1995) o la predicción media para cada combinación de variables independientes (HU et al, 1999). Sin embargo, otra alternativa a estos métodos es valorar el efecto de las variables independientes en la variable dependiente, mediante la evaluación de las derivadas del resultado de la red con respecto a cada una de estas variables independientes (ILASHEM, 1992, WEST BROCKET, y GOLDEN, 1997, WARTENBERG y DECKER, 1996). La valoración de estas derivadas en los distintos niveles de las variables independientes permite, por un lado, valorar la importancia general de cada una de las variables en los resultados obtenidos y, por otro lado, analizar los

efectos no lineales que se presentan en el modelo, a través de la comparación de las derivadas en los distintos niveles.

En el caso de una MFANN con una única capa oculta y con funciones de activación logística y lineal, el proceso para el cálculo de las derivadas es el que se describe a continuación. Para cada unidad de salida K , tomando el vector de resultados \mathbf{O}_k para cada uno de los patrones de entrada en el caso, calculamos la sensibilidad para cada variable x_i como la derivada parcial con respecto a x_i . Para calcular esta derivada parcial aplicamos la regla de la cadena de forma que:

$$\partial \mathbf{O}_k / \partial x_i = \sum_j \partial \mathbf{O}_k / \partial H_j \cdot \partial H_j / \partial x_i \tag{A.1}$$

$$\text{Dado: } \mathbf{O}_k = f^2(\mathbf{w}^{(2)}_k \mathbf{H}) \tag{A.2}$$



donde $\mathbf{w}^{(2)}$ es la matriz de pesos $w_{jk}^{(2)}$ para $j=0,..J$ (número de unidades de salida y 0 igual a constante o peso del sesgo), \mathbf{H} es el vector de activaciones de la capa oculta y $f^{(2)}$ la función de activación en la capa de salida

$$H_j = f^{-1}(x'w^{(1)}_j) \quad [A.3]$$

donde x es el vector de variables independientes, $w^{(1)}_j$ es el vector de pesos que une la capa de entrada con la unidad oculta j incluido el peso de la unidad de sesgo y $f^{(1)}$ es la función de activación en la capa oculta

Según estas expresiones, la derivada parcial del resultado del modelo con respecto a una de las variables independientes sería:

$$\partial O_k / \partial x_i = \sum_j w_{ji}^{(1)} w_{jk}^{(2)} f^{(1)'}(x'w^{(1)}_j) f^{(2)'}(w^{(2)}\mathbf{H}) \quad [A.4]$$

En el caso de que la capa de salida tenga una función de activación de tipo lineal y la función de

activación en la capa oculta sea de tipo logístico, distinguimos entre $f^{(1)}=1/(1+\exp(-z))$ y $f^{(2)}(w^{(2)}\mathbf{H})=w^{(2)}\mathbf{H}$ y la expresión para la sensibilidad se transforma en:

$$\partial O / \partial x_i = \sum_j w_{ji}^{(1)} w_{jk}^{(2)} f'(x'w^{(1)}_j) (1-f(x'w^{(1)}_j)) \quad [A.5]$$

A partir de esta expresión general y una vez finalizado el entrenamiento, la sensibilidad media del resultado para un input determinado se calcula para el caso en el que las variables dependientes son categorías como:

$$S_{ik} = 1/N(k) \sum_1^{N(k)} \partial O_k / \partial x_i \quad [A.6]$$

donde S_{ik} es la sensibilidad del input i en la categoría k y $N(k)$ es el número de vectores con el valor k para la variable i . En el caso de que las variables sean de tipo continuo podemos agruparlas por intervalos y realizar el mismo cálculo.

