

# ESTIMACIÓN DE DEMANDA A NIVEL DE SKU UTILIZANDO UN ENFOQUE BASADO EN ATRIBUTOS: APLICACIÓN AL CASO SUPERMERCADOS<sup>1</sup>

*NICOLAS FRITIS, JUAN JOSÉ JOVE, CAROLINA OLGUIN*  
*Departamento de Ingeniería Industrial, Universidad de Chile*  
*nfritis@ing.uchile.cl, jjove@ing.uchile.cl, maolguin@ing.uchile.cl*

*MAXIMO BOSCH P., MARCEL GOIC*  
*Departamento de Ingeniería Industrial, Universidad de Chile.*  
*mbosch@dii.uchile.cl, mgoic@dii.uchile.cl*

## RESUMEN

En este trabajo se propone un modelo de estimación de demanda basado en los atributos que describen una categoría de productos (por ejemplo marcas, tamaños, sabores, etc.). El modelo se basa en crear relaciones analíticas entre los productos y los atributos que los componen, de esta manera basta estimar los parámetros del modelo solo utilizando información de atributos y a partir de cálculos se llega a los parámetros a nivel de SKU.

El modelo se valida con un conjunto de datos simulados para luego ser calibrado con 181 días de datos reales de dos subcategorías distintas: pastas de diente y lavalozas. Entre los resultados se encuentra que los ajustes a nivel de atributos son considerablemente superiores a los obtenidos a nivel de SKU. Por otra parte, los parámetros que entrega el modelo se utilizan para recobrar elasticidades a nivel de atributos y a nivel de SKU, y además se utilizan para realizar estimaciones de demanda, las que tienen un error promedio porcentual del orden de un 30%. Dichos pronósticos superan a los obtenidos al utilizar otras técnicas, como alisamiento exponencial simple, y alisamiento exponencial con estacionalidad, entre otros.

Palabras clave: Modelos econométricos, atributos, retail, pronósticos.

## ABSTRACT

In this work a based demand estimation model on the attributes is proposed that describe a category of products (for example brands, size, flavors). The model is based on creating analytical relations between the products and the attributes that compose them, in this manner suffices to reckon the parameters of the alone model utilizing data of attributes and from calculate arrives at the parameters to SKU level. The model is validated with an assembly of data simulated in order then to be calibrated with 181 days of real data of two different sub-categories. Among the results is found that the adjustments to attribute level are considerably over them obtained to SKU level. On the other hand, the parameters that delivers the model are utilized to recover elasticities to attribute level and to SKU level, and besides they are utilized for carry out estimations of demand, the ones that have an error percentage average on the order of a 30%. Said forecasts surpass to them obtained al to utilize other techniques, like single exponential smooth, and with seasonal nature exponential smooth, among others.

Key words: Econometric models, attributes, retail, forecasts.

---

<sup>1</sup> Estudio financiado por proyecto FONDEF D03I1057

# 1 INTRODUCCIÓN

Hoy en día la mayoría de las categorías de productos consisten en cientos de SKU (*stock keeping units*), lo que sumado a todos los efectos que tienen sobre los clientes las variables precio y promoción, entre otras, crea un gran desafío para quienes desean estudiar las respuestas de los consumidores hacia la góndola.

Básicamente, existen dos opciones para los analistas que desean recobrar los parámetros de respuestas hacia cada una de los SKU. La primera es estimar un modelo por cada SKU, la implementación de esta clase de modelos no resulta adecuada, pues las elecciones de los consumidores generalmente no dependen de un solo producto, sino de un conjunto de productos que comparten ciertas características. A esto hay que sumarle el tema de los quiebres de stock, que provoca que los consumidores cambien constantemente sus elecciones. Además, aparecen problemas cuando existe volatilidad en el conjunto de productos debido a la constante entrada y salida de SKUs en la categoría (Bucklin and Gupta, 1999). Otro problema son los requerimientos computacionales de estos modelos, los que pueden llegar a ser significativos, especialmente si consideramos construir modelos sobre varias categorías, o si necesitamos correr simulaciones que requieren repetidas estimaciones de estos modelos.

Esto ha motivado a los investigadores a desarrollar modelos que permitan la descripción de SKUs pero sin considerarlas como su unidad de análisis (Fader and Hardie, 1996). Dichos modelos se basan en que en las categorías con un gran número de alternativas pueden ser descritas de acuerdo a un conjunto de atributos (marca, tamaño, sabor, forma, etc.). Estos métodos permiten agregar SKUs dentro de dichas opciones, reduciendo la cantidad de parámetros a estimar y también trabajando sobre conjuntos de datos más estables sin perder el punto de vista del SKU.

En lo que sigue se analizan los aspectos que motivan este estudio para luego especificar los objetivos del mismo. Posteriormente se realizará una revisión bibliográfica del tema, se especificará la metodología a seguir, y se detallará el modelo implementado. El análisis de resultados se realiza en dos partes: primero se analizará una aplicación del modelo con datos simulados y luego se aplicará el modelo a datos reales provenientes de un supermercado. Posteriormente, se realiza un análisis de sensibilidad y se discuten diversos aspectos referentes al modelo para finalmente revisar las principales conclusiones de este trabajo.

## MODELO

El modelo esta basado en uno desarrollado por Bell, Bronfer y Chintagunta. El software en que se programó es Matlab, por considerarse una herramienta poderosa, amistosa, ampliamente utilizada, y que cuenta con todas las herramientas necesarias (paquetes estadísticos, de pronósticos, procesamiento de datos, etc.) para el desarrollo del modelo.

A continuación se describen las principales características del modelo.

### Consideraciones generales

Para efectos de una mejor comprensión, se describen las variables con las que se trabaja:

**Productos (J):** Corresponden a la unidad mínima de análisis. Se debe conocer su precio y la cantidad vendida de cada producto en cada periodo de tiempo. Se describen con el subíndice **J**.

**Atributos (A):** Se entiendo por atributo a las características físicas que describen de manera única, cada uno de los productos disponibles en góndola. Por ejemplo, en el caso de las pastas de dientes, se puede hablar de los atributos marca, tamaño y tipo de envase, sabor, etc. Muchos de estos atributos, especialmente marca y tamaño, son comunes a prácticamente todas las categorías de productos, otros, varían de una categoría a otra. Se describen con el subíndice **A**.

**Nivel (L):** Se utiliza el término “nivel de atributo” para denotar cada uno de las distintas ofertas, que juntas, constituyen un atributo; por ejemplo, Pepsodent, Colgate y Aquafresh, son un subconjunto de niveles bajo el atributo **marca**, en la categoría de pastas dentales. Se describen con el subíndice **L**.

### Datos de entrada

El modelo requiere de los siguientes datos: Definición de un bien de referencia (de ahora en adelante *outside good*), *share* de cada producto, matriz de descripciones producto – atributo y características de la góndola (precio, promociones, etc.). Cada uno de ellos debe estar referido a una unidad de tiempo específica, la cual puede ser días, semanas, meses, etc. El detalle de los datos necesarios es el siguiente:

### Outside good

Representa la alternativa de no consumir (i.e. utilidad igual a 0), y se utiliza para referenciar las utilidades de los productos a su valor. Se puede estimar de dos maneras:

- Estimando el mercado potencial del producto y calculando cual es la diferencia entre las ventas reales y las potenciales en cada unidad de tiempo.
- Eligiendo algún SKU para referenciar las utilidades hacia ella. En este caso es necesario referenciar además, todas las características de la góndola con respecto al SKU escogido.

Independiente de la manera de estimar el *outside good*, nos referiremos a él con el subíndice **0**.

### Market shares

Representa el porcentaje de las ventas que tiene cada producto, incluido el *outside good*. Se calcula dividiendo las ventas de cada producto por las ventas totales, en cada unidad de tiempo. Nos referiremos a los *shares* con **s** al hablar de de productos, y **S** al hablar de *shares* a nivel de atributos.

### Características de la góndola

Representan la presencia o ausencia de elementos que pueden influir sobre las ventas de los productos. Se utiliza entre ellas el precio de cada producto en cada unidad de tiempo y se pueden agregar otro tipo de datos como la presencia de publicidad, promociones, etc. Para efectos explicativos se trabajará solo con el precio, aunque se puede utilizar cualquier otro tipo de variable (promociones, publicidad, etc.), siempre y cuando se encuentre centrado sobre la media. Además se requiere conocer que SKUs están efectivamente disponibles en góndola en cada periodo de tiempo.

### Relación productos - atributos

Describen la pertenencia (o no pertenencia) de un producto a un determinado nivel de un atributo, en otras palabras, es una manera de describir cada producto como un conjunto de atributos. La notación usual es mediante matrices producto - atributo.

### Modelo de regresión

Se plantea el siguiente modelo de regresión:

$$\underbrace{\ln(S_{at}^l) - \ln(S_{ot}^l)}_{\Psi_a^l} = \alpha_a^l + \gamma_a^l (\bar{P}_{at}^l - \bar{P}_a^l) \quad (1)$$

Donde:

- $S_{at}^l$  : *Share* del nivel **l** del atributo **a** en **t**. Se calculo como la suma de los *shares* de los productos que pertenecen al nivel **l** en **t**.

- $S_{ot}$ : *Share* del *outside good* en  $t$ . Se obtiene estimando el mercado potencial del producto y calculando cual es la diferencia entre las ventas reales y las potenciales en cada unidad de tiempo.
- $\bar{P}_{at}^l$ : Precio promedio del nivel **I** del atributo **a** en  $t$ . Se calcula como:

$$\bar{P}_{at}^l = \sum_{i=1}^{N_a^l} \frac{S_{it}}{S_{at}^l} p_{it} \quad (2)$$

- $\bar{P}_a^l$ : Promedio del precio del nivel **I** del atributo **a**. Se calcula tomando el promedio sobre  $t$  de  $\bar{P}_{at}^l$ .
- $\alpha_a^l$ : Efecto fijo del nivel **I** del atributo **a**. **Parámetro a estimar en la regresión.**
- $\gamma_a^l$ : Efecto del precio sobre la utilidad del nivel **I** del atributo **a**. **Parámetro a estimar en la regresión.**

### Primera regresión – Recuperación de los interceptores

Una vez calculados todos los datos, se aplica una regresión de mínimos cuadrados ordinarios sobre el modelo la ecuación 5.

Al realizar la regresión, para cada uno de los atributos y cada uno de los niveles, se obtiene todos los  $\alpha_a^l$ . Como se conocen las participaciones de mercado para cada producto y para nivel de cada atributo, se esta en condiciones de recobrar el interceptor para cada producto aplicando (la demostración se encuentra en el anexo 1):

$$\beta_j = \alpha_a^l + \frac{1}{T} \sum_t \left[ \ln\left(\frac{S_{jt}}{S_{at}^l}\right) \right] \quad (3)$$

### Segunda regresión – Recuperación de los efectos de góndola

Hasta ahora no se tiene una manera de recobrar los efectos de marketing a nivel de SKU directamente utilizando una formula como se hace en el caso de los efectos fijos, por lo que es necesario realizar una segunda regresión (esta vez a nivel de SKU) para poder recuperar dichos efectos.

Sin embargo no necesitamos partir de cero, habiendo estimado el modelo a nivel de atributos calculamos el efecto fijo como se muestra en la ecuación 7, luego en el modelo a nivel de SKU, sustraemos dichos efectos fijos sobre las transformaciones logarítmicas de las participaciones de mercado diferenciadas. Así, escribimos la siguiente regresión:

$$\underbrace{\ln(s_{jt}) - \ln(s_{0t}) - \beta_j}_{\tilde{y}_{jt}} = \gamma_j \cdot (p_{jt} - \bar{p}_{jt}) \quad (4)$$

Al realizar esta regresión para cada producto se obtienen los parámetros del precio correspondientes.

### Estimación de demanda

La estimación de demanda procede en dos pasos, primero, utilizando el modelo de atributos se estiman los *shares* para cada SKU, y luego, con una estimación de las ventas totales de la categoría, se estima la demanda en unidades para cada SKU. El detalle de cada uno de estos pasos se explica a continuación:

#### Calculo de *shares*

Una vez estimados los parámetros del modelo, dado un futuro nivel de precios se pueden estimar las utilidades de cada SKU y a partir de eso obtener el *share* estimado de cada producto. Esto se hace recordando que el modelo supone que los *shares* y las utilidades se relación a través de la siguiente ecuación (Modelo *multinomial logit*):

$$s_j = \frac{\exp(u_j)}{1 + \sum_{k \in J} \exp(u_k)} \quad (5)$$

### Estimación de ventas totales

Para estimar la demanda por cada producto en términos de unidades vendidas, solo basta multiplicar cada *share* por las ventas totales de la subcategoría y dividir por el tamaño del SKU, es decir:

$$Ventas_j = \frac{s_j \cdot Ventas\_totales}{Tamaño_j} \quad (6)$$

Para esto, se requiere contar con una estimación de las ventas totales de la categoría, para lo cual se utilizará una técnica de alisamiento exponencial con estacionalidad que se explica a continuación. Sean:

$D_t$  : Demanda observada en el periodo t

$R_t$  : Factor de estacionalidad en el periodo t

L : Largo de la estacionalidad

Entonces, se definen  $A_t$  (promedio calculado para el periodo t) y  $F_{t+k}$  (pronostico para el periodo t + k) como:

$$A_t = \alpha \cdot \left( \frac{D_t}{R_{t-L}} \right) + (1 - \alpha) \cdot A_{t-1} \quad 0 < \alpha < 1 \quad (7)$$

$$R_t = \gamma \cdot \left( \frac{D_t}{A_t} \right) + (1 - \gamma) \cdot R_{t-L} \quad 0 < \gamma < 1 \quad (8)$$

$$F_{t+k} = A_t \cdot R_{t-L+k} \quad (9)$$

En donde  $\alpha, \gamma$  se calibran de manera que minimicen un error de ajuste sobre el conjunto de validación.

El índice de estacionalidad representa el grado en el cual la estacionalidad afecta a un segmento particular del año. El cálculo envuelve una comparación del valor esperado de un período específico con respecto a la media general, en otras palabras mide en que grado el promedio de un período en particular se encuentra por arriba (o por debajo) de la media. Por lo tanto, para obtener una estimación precisa del índice de estacionalidad, se calcula el promedio del primer período del ciclo, el segundo período, etc., y se dividen por la media general. La formula para determinar el factor de estacionalidad es:

$$R_i = \frac{D_i}{D} \quad (10)$$

Donde:

$D_i$  : Venta promedio del periodo estacional  $i$

D : Promedio general de ventas

$i$  : Período estacional del ciclo (en este caso días de la semana)

## RESULTADOS

El modelo se prueba con datos reales de dos categorías distintas: Pastas de diente (42 SKU) y lavalozas (56 SKU).

Se realizaron estimaciones de demanda diarias tal como se muestra en la ecuación (12). Se utilizaron 181 días de datos de los cuales 153 días correspondieron a calibración y 28 a validación.

Una tabla resumen con los errores promedio de la estimación para cada día de validación se muestra a continuación para cada una de las categorías en estudio:

**Tabla 1 Resumen de errores de pronóstico para la subcategoría pastas de diente**

	MAD	MAD Pond	MIN MAD	MAX MAD	MAPE	MAPE Pond	Error total
Día 1	5.28	10.07	0.25	21.16	41.74%	26.39%	29.57%
Día 2	4.56	10.99	0.23	39.69	92.30%	46.39%	44.85%
Día 3	3.73	8.06	0.11	18.8	61.77%	38.19%	36.83%
Día 4	3.25	6.6	0.14	19.01	50.85%	31.77%	32.86%
Día 5	3.39	6.9	0.03	16.67	83.37%	34.89%	37.58%
Día 6	4.96	9.07	0.11	22.63	111.24%	41.00%	45.36%
Día 7	5.65	12.75	0.08	43.98	44.77%	38.04%	39.90%
Día 8	5.2	6.9	0.08	21.93	52.57%	27.90%	32.36%
Día 9	4.49	9.33	0.02	34.51	42.14%	36.77%	42.16%
Día 10	4.63	8.47	0.16	31.16	40.36%	32.41%	38.35%
Día 11	3.45	5.21	0.03	19.24	48.06%	24.42%	30.66%
Día 12	4.55	5.86	0.07	29.84	61.43%	31.93%	40.79%
Día 13	4.54	8.72	0.15	23.58	58.35%	32.22%	32.38%
Día 14	4.44	10.22	0.25	26.57	55.90%	31.55%	32.44%
Día 15	5.49	21.28	2.25	26.63	76.47%	33.74%	32.05%
Día 16	3.97	15.22	0.92	27.16	63.10%	33.27%	39.06%
Día 17	3.23	6.95	0.00	10.01	71.86%	29.10%	32.87%
Día 18	3.34	9.18	0.20	21.31	55.06%	33.54%	36.31%
Día 19	3.83	7.63	0.04	8.94	91.77%	44.70%	46.74%
Día 20	3.35	11.91	0.06	18.21	53.29%	32.91%	31.98%
Día 21	8.25	14.22	2.00	27.82	68.21%	44.01%	53.91%
Día 22	6.20	29.67	0.23	44.89	41.05%	43.07%	38.75%
Día 23	4.81	7.66	0.00	13.36	97.72%	33.68%	42.91%
Día 24	3.94	12.81	0.00	45.94	53.01%	29.26%	37.40%
Día 25	2.84	8.05	0.00	19.90	42.86%	26.23%	26.90%
Día 26	3.96	10.57	0.14	44.52	41.61%	32.16%	39.29%
Día 27	4.73	17.21	0.09	41.57	52.06%	39.13%	33.67%
Día 28	9.69	31.47	0.01	68.69	62.16%	40.78%	41.96%
<b>Promedio</b>	<b>4.63</b>	<b>11.54</b>	<b>0.27</b>	<b>28.13</b>	<b>61.25%</b>	<b>34.62%</b>	<b>37.50%</b>

**Tabla 2 Resumen de errores de pronóstico para la subcategoría lavalozas**

	MAD	MAD Pond	MIN MAD	MAX MAD	MAPE	MAPE Pond	Error total
Día 1	3.49	5.58	0.29	18.38	67.35%	34.50%	29.94%
Día 2	3.00	5.42	0.06	27.27	57.80%	34.56%	34.82%
Día 3	3.26	6.06	0.05	23.45	51.99%	38.90%	35.31%
Día 4	2.57	3.89	0.10	14.54	74.39%	34.79%	32.01%
Día 5	3.03	4.96	0.23	10.55	62.89%	36.54%	33.96%
Día 6	3.25	7.84	0.04	12.33	46.75%	35.28%	32.73%
Día 7	3.13	9.50	0.02	12.17	45.66%	28.13%	25.93%
Día 8	3.31	6.08	0.00	11.56	73.60%	29.51%	27.40%
Día 9	2.97	4.59	0.03	16.1	59.38%	32.31%	30.24%
Día 10	2.74	4.55	0.06	26.45	52.20%	33.46%	31.62%
Día 11	2.75	5.07	0.07	11.19	53.39%	31.71%	29.54%
Día 12	3.08	4.29	0.04	24.83	56.62%	41.29%	39.17%
Día 13	5.13	11.95	0.09	21.18	55.88%	38.18%	38.16%
Día 14	2.77	4.68	0.00	39.12	53.59%	25.24%	23.96%
Día 15	7.57	19.41	0.20	99.72	47.14%	41.98%	39.74%
Día 16	5.19	13.85	0.01	91.82	119.37%	43.41%	41.14%
Día 17	4.27	6.47	0.00	41.54	58.69%	40.38%	37.99%
Día 18	2.88	4.91	0.21	21.77	54.48%	30.84%	28.33%
Día 19	3.11	7.43	0.08	20.65	51.08%	34.46%	30.87%
Día 20	2.94	4.45	0.01	8.54	54.64%	27.03%	25.87%
Día 21	4.84	9.19	0.02	8.31	51.21%	29.92%	30.18%
Día 22	5.20	8.41	0.03	15.69	40.93%	29.93%	28.98%

Día 23	4.28	6.42	0.02	17.9	45.48%	35.41%	34.31%
Día 24	4.35	7.46	0.01	36.75	76.55%	41.61%	39.50%
Día 25	2.95	4.78	0.01	14.7	52.06%	33.84%	31.92%
Día 26	2.57	4.04	0.06	10.1	49.39%	27.07%	24.34%
Día 27	3.69	6.56	0.02	9.71	49.38%	32.03%	29.08%
Día 28	5.76	11.57	0.09	107.28	60.13%	34.38%	30.94%
<b>Promedio</b>	<b>3.72</b>	<b>7.12</b>	<b>0.07</b>	<b>27.63</b>	<b>57.93%</b>	<b>34.17%</b>	<b>32.07%</b>

Para comprobar que tan buenas son las estimaciones, se compararon estos resultados con los de un alisamiento exponencial simple, un alisamiento exponencial con estacionalidad, y una serie de regresión con siete rezagos. Los resultados fueron los siguientes:

**Tabla 3 Resumen de errores de pronóstico para distintas metodologías subcategoría pastas de diente**

	MAD	MAD POND	MIN MAD	MAX MAD	MAPE	MAPE POND	ERROR Porcentual
<b>Modelo de atributos</b>	4.63	11.54	0.27	28.13	61.25%	34.62%	37.50%
<b>Alisamiento exponencial simple</b>	6.23	21.54	0.08	51.98	76.67%	46.60%	44.85%
<b>Alisamiento exponencial Con estacionalidad</b>	5.11	15.55	0.17	36.4	65.41%	34.98%	36.05%
<b>Serie de regresión</b>	5.27	16.3	0.16	39.58	69.81%	37.20%	37.76%

**Tabla 4 Resumen de errores de pronóstico para distintas metodologías subcategoría lavalozas**

	MAD	MAD POND	MIN MAD	MAX MAD	MAPE	MAPE POND	ERROR Porcentual
<b>Modelo de atributos</b>	3.72	7.12	0.07	27.63	57.93%	34.17%	32.07%
<b>Alisamiento exponencial simple</b>	4.38	8.7	0.06	30.38	63.95%	39.81%	38.72%
<b>Alisamiento exponencial Con estacionalidad</b>	3.87	7.51	0.07	24.28	57.53%	35.56%	33.79%
<b>Serie de regresión</b>	7.73	15.08	0.18	44.95	145.89%	69.37%	63.83%

A pesar que no existe un mejor modelo que sea el que minimice el error en cada caso, los resultados muestran que el modelo de atributos es el que aparece como el más preciso (es el mejor en 9 de las 14 comparaciones) en ambas categorías.

Nótese además que nuestro modelo de atributos se calibra para toda la categoría completa, mientras que otros enfoques mostrados requieren de estimar un modelo por cada SKU presente en la categoría.

## CONCLUSIONES

Se ha propuesto un modelo basado en atributos que permite estimar demanda a nivel de SKU. El modelo trabaja en dos etapas, en la primera se estiman los *shares* para cada SKU y en la segunda se realiza un pronóstico para las ventas totales de la categoría. Juntamos ambas estimaciones se obtiene un pronóstico de ventas para cada una de las SKU de las subcategoría.

El modelo se aplicó a un conjunto de datos reales provenientes de dos subcategorías de un mismo supermercado. Los resultados probaron que el modelo ajusta bien a ambas subcategorías aportando una manera sencilla de estimar las respuestas de los consumidores hacia la góndola.

Los resultados del modelo se utilizan para estimar demanda logrando errores promedio de aproximadamente un 30%, lo que es bastante bueno considerando que para explicar las variaciones en las ventas solo se utiliza información histórica de ventas y precios de cada SKU.

La gran ventaja de nuestro modelo es que permite examinar parámetros nivel de atributos, mientras que un enfoque clásico a nivel de SKU no lo permitiría. La ventaja de esto es que, a través de relaciones entre los parámetros a nivel de atributos y los calculados para SKU, se puedan generar estimaciones de utilidad y sensibilidades al precio para SKUs que actualmente no son parte de la góndola, entregando una valiosa herramienta para analizar respuestas hacia nuevos productos.

## **Bibliografía**

Bucklin and Sunil Gupta (1999) "Commercial Use of Scanner Data: Industry and Academic Perspectives," *Marketing Science*, 18 (3), 247-273.

Camerer and Ho (1999) "A Parsimonious Model of SKU Choice: Familiarity-based Reinforcement and Response Sensitivity" *Journal of Marketing Research*, 17 (5), 142-176.

David R. Bell, Andre Bonfrer and Pradeep K. Chintagunta (2005) "Recovering SKU-Level Preferences and Response Sensitivities from Market Share Models Estimated on Item Aggregates", *Journal of Marketing Research*, Vol. XLII (May), 169-182.

David Simchi-Levi, Philip Kaminsky, Edith Simchi-Levi (1999) "Designing and managing the supply chain : concepts, strategies and case studies", New York : McGraw-Hill.

Fader, Peter S. and Bruce G. S. Hardie (1996) "Modeling Consumer Choice Among SKUs," *Journal of Marketing Research*, 33, (4), 442-452.

Ho, Teck-Hua, and Chong, Juin-Kuan (2003) "A Parsimonious Model of Stockkeeping Unit Choice," *Journal of Marketing Research*, 40 (3), 351-365.

Lee G. Cooper, Masao Nakanishi (1993) "Market – Share Analysis", Kluwer Academic Publishers, 103-175.

Naresh K. Malhotra, David F. Birks (2003) "Marketing Research: An Applied Approach", 714-715.

Nevo, Aviv. (2000) "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand", *Journal of Economics and Management Strategy*, 9(4), 513-548

Russell, Gary. J. and Ruth N. Bolton (1988) "Implications of Market Structure for Elasticity Structure" *Journal of Marketing Research*, 25 (3), 229-241.

## ANEXO 1 - DEMOSTRACION DE LA RECUPERACIÓN DE INTERCEPTORES

### Conjuntos

J: Conjunto de productos (SKUs)  $J = 0..J$ , donde 0 representa el bien de referencia o *outside good*

T: Conjunto de periodos de tiempo (días, semanas, meses, etc.)

A: Conjunto de atributos

A<sub>l</sub>: Conjunto de niveles del atributo *a*

### Parámetros

$u_{jt}$  : Utilidad del producto *j* en *t*.

$\Psi_{lt}^a$  : Utilidad del nivel *l* del atributo *a* en *t*

$s_{jt}$  : *Share* del producto *j* en el periodo *t*.

$S_{lt}^a$  : *Share* del nivel *l* del atributo *a* en el periodo *t*.

### Demostración

Se parte del supuesto de un modelo *multilogit* (MNL), con un bien de referencia (de ahora en adelante *outside good*) el cual representa la utilidad de no consumir (i.e. utilidad igual a 0), de esta manera se tiene:

$$\text{A nivel de SKU : } s_{jt} = \frac{\exp(u_{jt})}{1 + \sum_k \exp(u_{kt})} \quad (11)$$

$$\text{A nivel de atributos : } S_{lt}^a = \frac{\exp(\Psi_{lt}^a)}{1 + \sum_a \sum_l \exp(\Psi_{lt}^a)} \quad (12)$$

Como el bien de referencia es uno solo, debiera ser claro que:

$$s_{0t} = \frac{1}{1 + \sum_a \sum_l \exp(\Psi_{lt}^a)} = \frac{1}{1 + \sum_k \exp(u_{kt})} \quad (13)$$

Intuitivamente esto quiere decir que sumar las utilidades de todos los SKUs es equivalente a sumar sobre todos los niveles de todos los atributos, y que dicha suma representa la utilidad total de la góndola, es decir:

$$\text{Utilidad de la góndola} = \sum_k \exp(u_{kt}) = \sum_a \sum_l \exp(\Psi_{lt}^a) \quad (14)$$

Entonces, utilizando (1) y (2) se tiene que:

$$\frac{s_{jt}}{\exp(u_{jt})} = \frac{S_{lt}^a}{\exp(\Psi_{lt}^a)} \quad (15)$$

Que es la ecuación que relaciona utilidades a nivel de atributos y a nivel de SKU a través de los *shares*. Para continuar, a continuación se definen los siguientes modelos de regresión:

$$\text{A nivel de SKU : } u_{jt} = \beta_j + \gamma(p_{jt} - \bar{p}_j) + \varepsilon_{jt} \quad (16)$$

$$\text{A nivel de atributos: } \Psi_{lt}^a = \alpha_l^a + \gamma(P_{lt}^a - \bar{P}_l^a) + \xi_{lt}^a \quad (17)$$

Con (5) se tiene:

$$\frac{s_{jt}}{S_{lt}^a} = \frac{\exp(u_{jt})}{\exp(\Psi_{lt}^a)} \Rightarrow \ln\left(\frac{s_{jt}}{S_{lt}^a}\right) = u_{jt} - \Psi_{lt}^a \quad (18)$$

Reemplazando las utilidades por su expresión en la regresión (Ecuación (6) y (7)):

$$\ln\left(\frac{s_{jt}}{S_{lt}^a}\right) = \beta_j + \gamma(p_{jt} - \bar{p}_j) + \varepsilon_{jt} - \alpha_l^a - \gamma(P_{lt}^a - \bar{P}_l^a) - \xi_{lt}^a \quad (19)$$

Tomando el promedio sobre el tiempo:

$$\frac{1}{T} \sum_t \ln\left(\frac{S_{jt}}{S_{it}^a}\right) = \beta_j + \underbrace{\gamma(\bar{p}_j - \bar{p}_j)}_{=0} + E(\epsilon_{jt}) - \alpha_i^a - \underbrace{\gamma(\bar{P}_i^a - \bar{P}_i^a)}_{=0} - E(\xi_{it}^a) \quad (20)$$

Como los errores tienen media 0, entonces se concluye:

$$\beta_j = \alpha_i^a + \frac{1}{T} \sum_t \ln\left(\frac{S_{jt}}{S_{it}^a}\right) \quad (21)$$

Que es la relación entre efectos fijos a nivel de atributos y de SKU que se buscaba. Nótese que es posible la introducción de cualquier número de variables explicativas (siempre centradas sobre la media) sin que estas afecten la recuperación final.