European Journal of Operational Research 198 (2009) 666–674

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

# Using store level scanner data to improve category management decisions: Developing positioning maps ☆

Óscar González-Benito [a,*], María Pilar Martínez-Ruiz [b,1], Alejandro Mollá-Descals [c,2]

[a] Departamento de Administración y Economía de la Empresa, Universidad de Salamanca, Campus Miguel de Unamuno, 37007 Salamanca, Spain
[b] Facultad de Ciencias Sociales, Universidad de Castilla-La Mancha, Avenida de los Alfares, 44. 16002 Cuenca, Spain
[c] Departamento de Comercialización e Investigación de Mercados, Universidad de Valencia, Facultad de Economía, Avenida de los Naranjos, s/n. 46022 Valencia, Spain

ARTICLE INFO

*Article history:*
Received 15 November 2007
Accepted 20 October 2008
Available online 26 October 2008

*Keywords:*
Marketing
Management
Decision making
Market response models
Internal analysis of the market structure
Positioning maps

ABSTRACT

This paper provides evidence of the usefulness of aggregated point-of-sale scanner data to infer the positioning of competing brands, providing valuable information for category management and hence facilitating decision making. Specifically, the authors propose a methodology to study the internal market structure based on market share models with latent heterogeneity when only macro-level time series data (not individual choices) are available. The proposed approach assumes a multidimensional decomposition, latent in the preference structure that is implicit to these types of models. By empirically applying this approach, the authors (1) simultaneously identify both latent dimensions of competing brands and latent segments with different brand preferences, (2) explain the competitive positioning of brands without using disaggregated consumer panel data, and (3) achieve greater predictive performance. The findings offer insights to academics and practitioners interested in improving the practice of category management.



## 1. Introduction

Retailers face a complex task in designing and implementing pricing and deal strategies and tactics for a wide variety of products in today's competitive environment. Careful and calculated selection processes must be undertaken to ensure that retailers choose to market to the right customers at the right time with the right assortments, taking into account the sales and profit impacts of all relevant decisions. Manufacturers also must make informed decisions about the configuration of their product lines and marketing actions, based on insights from brand competition at the point-of-sale. However, while retailers are interested in optimizing the performance of the whole category, manufacturers seek to improve the performance of their brands by enhancing their relative appeal and minimizing cannibalization effects. In this respect, category management has become a core business practice in the consumer packaged goods industry. Marketing decisions should first assess the role of each competing brand within the framework, as defined by the product category. In this scenario, the development of analytic tools to describe and understand intra-category brand competition is of great interest.

Because category management fundamentally is driven by data, information technology (IT) offers an important enabling component. The proliferation of IT in marketing channels has contributed to the greater availability of data about consumer behavior, especially with regard to scanner data recorded in stores through point-of-sale (POS) systems. This development has encouraged marketing researchers to turn to analytic tools that use these data to facilitate the understanding and prediction of market responses to marketing stimuli.

Scanner data provides retailers with data at the household level, that is, information about customers' purchase history. This information is key to facilitating successful customer relationship management. However, the availability of such disaggregated data generally is limited to those customers who use loyalty cards or credit cards. These identification criteria can be cumbersome and result in bias, such as the bias derived from the use of different cards or the use of cards only on some purchase occasions. The resulting sample of consumers may not be representative of the store clientele. In addition, several retail chains do not offer loyalty cards – including Wal-Mart and Aldi – or have suspended their loyalty schemes, such as Safeway UK. This circumstance justifies interest in developing analytical tools focused on aggregate data at the store level, which are straightforwardly available from scanner data and cover the purchases of all customers.

☆ This work has been funded by the Research Project SEJ ECON SEJ2005-06105 GESTIÓN DE LAS PROMOCIONES Y DE LAS MARCAS DE DISTRIBUIDOR PARA UN DESARROLLO EFICIENTE DE LA ESTRATEGIA DE MARKETING DEL MINORISTA, Ministerio de Educación y Ciencia (Spain).

* Corresponding author. Tel.: +34 923 294 400/3008; fax: +34 923 295 715.
*E-mail addresses:* oscargb@usal.es (Ó. González-Benito), MariaPilar.Martinez@uclm.es (M.P. Martínez-Ruiz), Alejandro.Molla@uv.es (A. Mollá-Descals).
[1] Tel.: +34 902 204 100/4242; fax: +34 902 204 130.
[2] Tel.: +34 963 828 834; fax: +34 963 828 333.

0377-2217/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2008.10.015

This paper extends existing research by proposing a methodology to analyze the positioning of competing brands on the basis of aggregated POS scanner data. This methodology enables retailers to identify latent dimensions of competitive positioning, positions of competing brands on these dimensions, and consumer segments based on the importance they give to these dimensions. Together, these elements cover the three key parallel phases of the positioning research process (Hooley and Saunders, 1993): (1) determining competitive dimensions; (2) determining competitors' positions; and (3) determining customer positions.

The main contribution of this proposal therefore is the adaptation of analytical tools developed for disaggregated data to a context in which only aggregated data are available. This proposal relies on two specific research lines within this vast research field: Internal analysis of market structure and modeling of latent heterogeneity.

The internal analysis of the market structure, according to Elrod (1991), infers brands' attributes and consumers' preferences for those attributes from behavioral or judgmental consumer data. Elrod (1991), DeSarbo et al. (1994), Elrod and Keane (1995), and Elrod et al. (2002) demonstrate the multiple methodological traditions oriented toward this goal.

Modeling latent heterogeneity relates to the explicit handling of consumers' heterogeneity in the modeling of market responses, which is a fundamental objective from both an explanatory perspective (i.e., segmentation is key in the development of marketing strategies) and a predictive point of view (i.e., ignoring the diversity of consumers' preferences and sought benefits can bias market response models). Specifically, the latent heterogeneity approach, based on random effects, assumes that consumers' preferences and sensibilities to marketing stimuli differ according to a probability distribution. Among probabilistic models of discrete choice based on the random utility theory, it is possible to distinguish two approaches: The parametric, which assumes a known parametric distribution (Gonul and Srinivasan, 1993), and the semi-parametric, which assumes a discrete distribution (Kamakura and Russell, 1989).

Some contributions attempt to combine analyses of market structure with latent heterogeneity by assuming a multidimensional decomposition of the preference structure that parameterizes probabilistic choice models with random effects. Both the parametric approach (Elrod, 1988; Elrod and Keane, 1995) and the semi-parametric approach (Chintagunta, 1994; Andrews and Manrai, 1999) have been considered. These models indicate the positioning of brands in a set of latent attributes or dimensions, as well as the distribution of consumers' preferences with regard to these dimensions. However, these approaches also employ choice models estimated with disaggregated data at the household level, that is, information about consumers' purchase history.

Instead, the methodology we propose models modeling market structure and latent heterogeneity with aggregate data at the store level. Therefore, it broadens the applicability of aggregate data to category management by simultaneously addressing their implications for both brand positioning and market heterogeneity. Specifically, by adapting Chintagunta's (1994) proposed modeling to market share models, estimated with aggregated data at the store level, we simultaneously identify latent dimensions that characterize competing brands and latent segments with different preferences for these attributes, without needing disaggregated data about consumers' purchasing history. Such information is key for identifying saturated or unexplored positioning profiles and configuring retail assortments and manufacturers' product lines.

Zenor and Srivastava (1993) and Bodapati and Gupta (2004) focus on the estimation of market response models with latent heterogeneity using aggregated data; this work goes further by making the internal analysis of the market structure possible. The simultaneous modeling of the internal analysis of market structure and latent segmentation offers benefits compared with sequential modeling, because it reduces the number of parameters, which makes estimation easier, and identifies a suitable number of dimensions and segments through statistical criteria.

The proposed model also offers an alternative to Chintagunta et al.'s (2002, 2003) methodology to infer brand maps from store level scanner data. Their model interprets aggregate data as an aggregation of consumers' choices, assumes a multidimensional decomposition of the preference structure, and adopts the random effects approach to account for heterogeneity. However, they use a parametric approach, such that heterogeneity in market response is assumed to be normally distributed. Our approach, in contrast, focuses on the less restrictive semi-parametric approach, which makes no a priori assumption about the distribution of response parameters across the population. Another distinguishing feature, straightforwardly adaptable, from Chintagunta et al.'s (2002, 2003) proposal, is our use of daily data from a single store, which enable us to account for the variance of marketing stimuli within each week. Marketing campaigns, such as price promotions, do not always follow a weekly pattern. Moreover, single-store data implies that we can obtain a diagnosis of the internal market structure for each store within a retail chain, which facilitates the development of micromarketing strategies (Montgomery, 1997).

The remainder of this article contains three main sections. First, we describe the methodological approach. Second, we present an empirical application with retail scanner data to exemplify and assess the proposed methodology. Third, the last section summarizes the main conclusions.

## 2. Proposed methodology

The proposed methodology consists of five steps. First, an aggregate interpretation of choice models is proposed through market share models. Second, the multidimensionality assumption is incorporated into the preference structure. These dimensions constitute the axes of the positioning map for competing brands. Third, latent heterogeneity assumptions are included. The assumption of multidimensionality makes sense in this case and enables us to identify the model. Fourth, the estimation procedure is described from aggregated data. Fifth, statistical indicators to compare different model specifications are proposed.

### 2.1. Market response model: Aggregated vs. disaggregated perspective

Probabilistic models of discrete choice based on the random utility theory represent the probability $P_{ir}$ that a consumer $i$ on occasion $r$ selects brand $j$ as a parameterized function $f_j$ of a set of variables $X_{ir}$ characterizing the commercial supply for that consumer and that occasion (for a review, see Manrai, 1995). Two types of parameters are usually distinguished: constant preferences $\alpha_j$ for each brand, and parameters $\beta$ associated with the sensitivity of consumers to the changes in explanatory variables. Subject to this notation, this type of model can be summarized through the following expression:

$$P_{it}(j|\alpha_j,\beta) = f_j(X_{it}|\alpha_j,\beta). \quad (1)$$

The aggregate perspective implies the consideration of the whole market response, that is, the choice behavior of all the consumers aggregated within a period of time $t$. Assume that

(1) The choice behavior of all consumers within the market is guided by the same response parameters ($\alpha_j$,$\beta$).