Journal of the Operational Research Society (2013), 1–12

www.palgrave-journals.com/jors/

Detecting inefficiently managed categories in a retail store

Marcel Goic[1]*, Máximo Bosch[2] and Juan P Castro[1]

[1]University of Chile, Santiago, Chile; and [2]Universidad de las Américas, Santiago, Chile

Growth in operational complexity is a worldwide reality in the retail industry. One of the most tangible expressions of this phenomenon is the vast increase in the number of products offered. To cope with this problem, the industry has developed the 'category management' approach, in which groups of products with certain common characteristics are grouped together into 'categories', managed as if they were independent business units. In this paper, we propose a model to evaluate relative category performance in a retail store, considering they might have different business objectives. Our approach is based on Data Envelopment Analysis techniques and requires a careful definition of the resources that categories use to contribute to achieving their business objectives. We illustrate how to use our approach by applying it to the evaluation of several categories in a South American supermarket. The empirical results show that, even for very conservative assumptions, the model has a significant discriminatory power, identifying 25% of the sample as not operating efficiently. Although efficiency scores might exhibit a relatively large dispersion, the set of efficient units is robust to data variations.

Journal of the Operational Research Society advance online publication, 18 December 2013; doi:10.1057/jors.2013.146

Keywords: category management; retail productivity; data envelopment analysis

Introduction

Most retailers are facing increasing complexity in their operations. A mid-size supermarket offers approximately 20 000 stock-keeping units (SKUs) while hypermarkets display more than 80 000 units, creating a very challenging managerial problem: how to make decisions concerning price, assortment, space allocation and promotion for thousands of different SKUs. The common approach adopted within the industry to cope with this problem is category management (CM), in which products with a certain degree of similarity are grouped into clusters called 'categories'. The goal of CM is to have business units that can be managed quasi-independently (Dussart, 1998; Nielsen, 2005). In a typical supermarket, products are organized in up to hundreds categories, each one formed by several dozens of SKUs, although large variations are observed.

The CM process includes solving strategic problems such as deciding which products belong to each category, defining category objectives or *roles* (such as drawing customers into the store), conveying quality to customers through variety and satisfying consumers' requirements in relation to price adequately, as well as improving business returns (Hoch and Lodish, 1998). Once the strategic decisions have been made, there are complex tactical decisions that need to be taken within each category, such as pricing, assortment and promotional planning, shelf space allocation, among others.

A key element in the successful implementation of a CM strategy is the evaluation of the performance of each category. A proper assessment is required not only to evaluate category managers but also to ensure better allocation of resources in the store. For example, each store has a relatively fixed floor capacity and store managers need to decide how much space to assign to each category. An approach widely used to conduct this evaluation is the comparison of a series of indices including financial (eg sales, margins), operational (eg inventory to sales ratio, inventory turnover) and marketing (eg penetration, market share) metrics. However, given that different categories have different roles, it is very difficult to use the simple comparison of indices to conduct the evaluation and allocate resources among different categories in a store. For example, a category devoted to generating traffic to the store might exhibit very low margins but large numbers of baskets[1] including its items, while a category devoted to enhancing the variety of the assortment might exhibit a high margin but very low penetration. How can we identify whether some categories are poorly managed when they have different objectives? How can we allocate resources among categories that have been defined as having different objectives? In this article, we propose a methodology to conduct this multi-objective performance analysis. Our proposed solution aims to help retail chains in the complex task of evaluating category managers in relation to many categories and stores. In a large supermarket, over 100 categories are

**Correspondence: Marcel Goic, Department of Industrial Engineering, University of Chile, Republica 701, Santiago, RM 8370439, Chile.*
E-mail: mgoic@dii.uchile.cl

[1]We use the term 'basket' to refer to the list of all products that a customer buys in a single trip to the store as it has been used widely in retail literature (eg Bell and Lattin, 1998; Russell and Petersen, 2000).

usually defined, so a large chain may have to analyse many thousands. In this context, it becomes necessary to define a fast and easily understood solution for the formal evaluation of category performance.

In the following section, we provide a detailed description of the managerial problem that motivates this research. We then describe how retail productivity analysis has been addressed in the literature and present our methodological approach, discussing its strengths and limitations with respect to other available methods. The model and its application in an empirical case are described in the following two sections. Finally, we conclude with a final discussion and directions for future research.

Problem description

This research entails the evaluation of different categories in a retail store in which each category is managed as a quasi-independent unit with its own business objectives. This is precisely the scheme proposed by CM which is widely adopted in the industry. The periodic evaluation of category performance is a key component of the successful implementation of this strategy for two reasons. First, it provides a metric for the evaluation of category managers who are in charge of assortment, price, promotion, display and planogram decisions for the category. This is particularly relevant given the common practice of having category captains, where one of the manufacturers takes the lead in determining marketing efforts for the whole category (Bandyopadhyay *et al*, 2009). Second, store managers have several shared resources that need to be allocated across categories. Such allocations cannot be undertaken without a proper assessment of their impact on category productivity.

A major difficulty in comparing categories in a store is that each one of them has its own defined objectives. For example, Dhar *et al* (2001) classify category roles according to frequency of purchase and percentage of household purchase in the category, resulting in four roles: staples, niches, variety enhancers and fill-ins. Similarly, Hoch and Lodish (1998) describe five closely related roles: traffic builders, transaction builders, cash generators, profit contributors and image creators. These classifications clearly convey that it is cumbersome to evaluate the operation of these heterogeneous units by simply comparing efficiency indices as has been common practice in the industry.

Using indices to conduct the analysis has significant limitations. The most direct limitation is the inability to provide a proper benchmark against which to contrast the resulting indices. Suppose, for example, we want to compare the relative efficiency between a category that has been defined as *traffic builder* and another defined as *variety enhancer*. Given their objectives, we might expect the former to have low margins and that a large number of baskets will include its items. On the other hand, in the latter category, we expect to have low penetration but high margins per transaction. With this pattern, each category manager can claim efficiency and nothing can be claimed about the relative efficiency of these two units unless very strong assumptions are made.

Probably the most natural solution to the benchmarking problem is to classify all categories according to their roles and make index comparisons only within each group. Although this is a step in the right direction, it suffers from at least two important weaknesses: (1) several categories are not easily classified in a group and have more than one role (eg those with mild penetration and an average gross margin) and (2) even assuming that all categories can easily be assigned to a group, how can a store manager allocate shared resources between groups? Another possibility is to compare each category with the same category in other stores. This approach would require information gathering from multiple stores and proper adjustments for intrinsic market differences across stores. It could give us a relative efficiency score for each category; however, it would not be informative in relation to managerial decisions that have to be taken at the store and not at the chain level, such as floor space or local display allocation.

In this article, we describe how data envelopment analysis (DEA) models can be used to build a computational tool to guide the performance analysis of multiple categories with multiple goals in a retail store.

Efficiency analysis in the retail industry

The assessment of relative performance has been an issue of importance in the retail industry for decades. Starting from the early analyses of simple output/input ratios (Ingene, 1982), the marketing literature has proposed several methods to evaluate the productivity of units of different natures. The intrinsically multidimensional character of the evaluation introduces a methodological challenge that cannot easily be overcome by the simple use of efficiency ratios. One of the most widely used approaches in conducting performance evaluation is the application of regression approaches in which an aggregated measure of the inputs used or outputs generated is estimated as a function of several factors. For example, Ratchford and Stoops (1988) estimate the usage of labour as a function of quantity sold in several categories. Kamakura *et al* (1996) estimate the total cost of bank branches using labour, floor area and the volume of several services as explanatory variables. Gauri *et al* (2009) use expected market share as a metric to evaluate store performance while controlling for both demographic and store features. More recently, Assaf *et al* (2011) have extended traditional regression frontier approaches to accommodate a Bayesian framework. All these applications assume that either inputs (eg cost) or outputs (eg market share) can be summarized into a single variable. We believe that for the goal of our investigation, this assumption is very restrictive. Categories in a retail store have different objectives and imposing a functional form to aggregate all resources in a single metric might be quite controversial. For example, while a cost regression function could be appropriated to describe the performance of a *profit* *contributor* category, major assumptions need to be made to formulate a cost function to describe *traffic builders* or *image creator* categories. In addition, retail accounting practices make it difficult to distribute fixed costs across categories to provide detailed cost information at that level.

Another drawback of the use of regression analysis to evaluate category performance is the relatively large requirement for disaggregated data. To generate reliable estimation of frontier functions using a regression approach, we could use multiple periods of historical data. However, decision making in the retail industry is very dynamic and structural changes are frequently made in a store. The reduction of floor space, the introduction of seasonal categories and store layout redesigns are constantly being evaluated and implemented by store managers implying structural changes that limit the availability of valid historical data. Finally, standard regression analysis describes average performance and not an efficient frontier, severely limiting its usefulness in guiding resource allocation. To conform to our research goals, we would need to impose a special structure on the distribution of the error term of the regression (Ferrier and Knox Lovell, 1990) adding unnecessary barriers to the routine use of our model in the industry.

A second major technique that has been used to describe retail performance is DEA. This is a ‘data-oriented’ approach for evaluating the performance of a set of peer entities called decision-making units (DMUs) that convert multiple inputs into multiple outputs (Cooper *et al*, 2004). A unit is considered efficient if there is no other unit (or combination of units) that generates the same amount of products with fewer resources, or conversely that generates more products with the same use of resources. The literature reports DEA applications used to evaluate the performance of DMUs in many industries such as schools, hospitals and production plants, among others. For excellent reviews of DEA applications, see Charnes *et al* (1994) and Tavares (2002).

The literature also reports several applications of DEA models in the retail industry. Donthu and Yoo (1998) were one of the first to propose the use of DEA models to evaluate the performance of retail outlets. They illustrate the methodology by characterizing the stores of a fast food restaurant chain in relation to three inputs (store size, manager experience and promotions) and two outputs (sales and customer satisfaction) and show that DEA can effectively discriminate between efficient and inefficient units. By varying the set of inputs and outputs and the assumption concerning the production function to provide a better description of the context, DEA models have been applied successfully to the analysis of the relative performance of retail stores in different markets (Barros and Alves 2003; Perrigot and Barros, 2008; for a more extensive review of the applications of efficiency analysis in the retail industry, see Barros, 2006).

Grewal *et al* (1999) extend previous performance analyses of retail outlets by explicitly considering regional and assortment differences across stores. By comparing the performance of stores selling different assortments, this article demonstrates how DEA models can be used to compare units selling different sets of products as we do in our research. Finally, Keh and Chub (2003) desegregate the process of transforming labour and capital into retail outputs in three stages, providing valuable managerial information on the nature of the inefficiencies. Other recent developments have also extended basic DEA models to conduct evaluations in cases in which the available data are imprecise (Lin, 2011). The methodology has even been used not only to assess the static performance of different units but also to evaluate the dynamics of performance, such as the evolution of productivity over time (de Jorge Moreno, 2010) and the effect of regulations on observed performance (de Jorge Moreno, 2006).

In all these studies, the performance analysis is undertaken at the retail store level and to the best of our knowledge, there is no previous study providing a multivariate approach in the evaluation of category performance in the retail industry. The key premise of CM in considering categories as quasi-independent business units leads us to believe that the application of DEA in the analysis of their performance is quite natural. In essence, under a CM regime, categories use store resources just as a store uses chain resources to provide retail services. However, moving from store to category analysis implies several methodological challenges. In a traditional DEA application, DMUs correspond to units designed to reach the same goal but in different places and environments. This is the case of bank branches or workgroups within the same company. In contrast, the performance of a category needs to be defined in terms of the different objectives that the category can have. We postulate that a proper choice of inputs and outputs will enable us to use a DEA approach to detect categories that are inefficiently managed in comparison to others in the same retail store.

Given our research goals, DEA presents several advantages with respect to other alternative approaches. First, DEA is a multi-criteria methodology that allows for the simultaneous performance analysis of units that use multiple inputs to generate multiple outputs. Second, DEA is a nonparametric technique and therefore we do not need to impose any functional form of the production function to describe how inputs and outputs are related. This is particularly useful for our application in which categories have different objectives and therefore the underlying mechanisms of production might differ considerably. Finally, DEA does not require many points of historical data to conduct performance evaluation and therefore can be applied continuously even after using the results to introduce structural changes in the store configuration.

Dyson *et al* (2001) enumerate several issues that may limit the validity of DEA results, discussing the homogeneity of the units under assessment, the choice of inputs and outputs and the measurement of the variables, among others. For our application, we consider it worthwhile to discuss the homogeneity of the units and measurement errors in some detail. DEA has been criticized for capitalizing on measurement errors which might make the evaluation sensible to outliers (Kamakura *et al*, 1996). To ameliorate this risk, we propose the evaluation of categories using information from the previous month. Using this time horizon, we reduce the possibility of evaluating categories operating in unusual circumstances, such as an unexpected out-of-stock event resulting from exogenous demand shocks. Furthermore, we conduct sensitivity analysis to study the impact of data variations in the detection of inefficiencies.

Regarding the homogeneity assumption, differences in environment and scale of production are not a major concern in our application. In fact, all categories serve the same set of potential consumers. Most of the literature analysing efficiency in the retail industry devote significant attention to modelling efforts that control for the heterogeneity of branch locations. This is because some stores are located in neighbourhoods with customers who have a larger disposable income or with different proportions of residential properties. When DMUs differ in the underlying conditions under which they operate, the literature proposes adjusting for non-homogeneity using regression techniques (Sexton *et al*, 1994; Haas and Murphy, 2003). The basic idea is to regress efficiency scores on site characteristics to determine the expected output for each unit. Then, DEA is used to analyse deviations with respect to those expectations. In the problem we analysed, all categories are homogeneous in their environments because they face the same customers, even though not all customers buy in all categories. Regression analysis could be used to adjust for other sources of heterogeneity, such as average package size, participation in inflation baskets or the intensity of local competition for the category. However, in our application, the cost of acquiring all this information at the category level is prohibitive. On the other hand, the availability of detailed information on other characteristics of the categories could enhance the ability of our model to detect underperforming units.

By evaluating units that sell different products and pursue different objectives, we might also face the risk of comparing units that have intrinsic differences in their production mechanisms. However, according to our interviews with category and store managers, it seems clear that the task involved in the administration of different categories is essentially the same. In as much as the set of outputs carefully takes into account all possible goals of the store, the comparison is fair for all units. If one category is detected as inefficient, there are other categories that can achieve the same business objectives using fewer resources. In this sense, we do not aim to find the absolute production frontier, general best practices or critical success factors that are common to all categories (Thomas *et al*, 1998). Instead we want to detect categories for which resource allocation can be modified the better to achieve the store mission materialized in multiple goals. Our main interest is in finding categories that are producing less than the attainable level of output in a single period of time and not in finding structural drivers of retail productivity. However, the method is also able to provide guidelines with regard to this aspect.

Modelling

Let us consider a set of n categories, in which each unit i is described by a matrix X, where $\{x_{ji}\}$ indicates the amount of input j used (j=1…m), and by a *m*atrix Y, where $\{y_{ki}\}$ indicates the amount of output k produced (k=1…s). Then, the production problem can be solved by finding for each unit p the relative weights of inputs and outputs assigned to maximize the ratio input/output, maintaining the ratio of each bounded unit. The primary formulation can be viewed in its dual form in Equations (1)–(5). This formulation is preferred because it has a better computational efficiency and because of the relative ease of finding an economic interpretation (Cooper *et al*, 2006). The equations are as follows:

$$\min \theta_p \tag{1}$$

$$\text{subject to} \quad \theta_p x_p = X\lambda + s^- \tag{2}$$

$$y_p = Y\lambda - s^+ \tag{3}$$

$$1 = e\lambda \tag{4}$$

$$\theta_p \geqslant 0,\ \lambda \geqslant 0, s^- \geqslant 0,\ s^+ \geqslant 0 \tag{5}$$

where e is a vector of ones. This dual formulation is commonly referred as the input-oriented BCC model and differs from the standard CCR formulation in relation to the convexity constraint (4) which allows for variable economies of scale. We consider that imposing constant returns to scale might be too restrictive despite the fact that supermarket categories tend to be fairly similar in relation to assigned space and number of SKUs sold. In this model, the value of θ_p corresponds to a performance index which indicates that object p is efficient only if θ_p=1. The λ variables can be interpreted as the weight of unit i in the linear combination used to benchmark unit p. Finally, variables s^- and s^+ correspond to excess input and output shortfalls of the unit and need to be zero to denote efficiency. Many alternative DEA models have been proposed in the literature, but a complete exploration of them is beyond the scope of this research. Our choice of the model is based on the simplicity of its implementation and interpretation.

The key challenge in modelling this problem is determining the set of inputs and outputs that characterize all supermarket categories. To determine the final list of variables to include in the evaluation, we considered two main criteria: first, the ability to capture all the possible roles in which a category might be contributing to store performance as a whole; second, the accessibility of the data required to compute the metrics. Thus, we propose a set of four inputs and four outputs that we now discuss in turn.

Inputs

Inputs are defined as the resources available to a DMU for the maximization of its performance. Following Dyson *et al*'s (2001) suggestion to consider factors that cover the full range of resources used, we consider here both store and exogenous resources for which the categories compete.

(1) *Space* corresponds to the available shelf space for the category. More space provides two positive impacts on

category performance. First, it confers more visibility that results in more sales. Second, it reduces the probability of on-shelf stock outs. Given that the depth of shelves is basically constant for all categories in our data set, we measured space as the total area (m^2) of the shelf assigned to the category. As product sizes vary across categories, more space does not directly imply more sales. However, using this metric is justified by the ability of the store managers to reallocate shelf space among categories. In fact, when deciding on layout, store managers are interested in knowing whether the space assigned is contributing to generating profits. A potentially complementary measure of space is the number of facings used by the category. For simplicity, we did not consider this variable because for many products with irregular shapes this evaluation is cumbersome.

(2) *Promotional efforts* can be carried out by the retailer or the manufacturer. However, in the industry we are analysing, most promotions are financed by the latter based on an agreement between both parties. Usually, manufacturers accord a yearly (or semester) promotional budget for each store and implement this by sacrificing margin which is transferred directly to the price, attracting more customers to the store. Consumption levels throughout the year are also agreed. The most common type of promotion is price reduction, that is, the discount the manufacturer gives the retailer for a certain volume (or the volume sold in a predetermined time) of the SKU being promoted. In this way, promotions for a given SKU consume an available and limited resource. We considered price discounts, measured as the manufacturer's sacrifice in net margin multiplied by the number of promotional days. For multiple SKUs promoted in the category, we computed the weighted average based on their market shares in the category.

(3) *Features*. Supermarkets regularly publish catalogues in which they feature special deals. For an individual SKU, presence in these catalogues has a remarkable effect on sales. However, there is a print space constraint that limits the number of products that can be included. This input was measured as the number of category items featured in the period. If very different sizes of advertisements were possible, we suggested using a measure based on square inches of print space. The store we evaluated in our empirical application does not use newspaper advertising and therefore it is not included in this input.

(4) *Number of SKUs*. Kök *et al* (2008) argue that due to fixed store space and financial resources, assortment planning requires a trade-off between three elements: how many different categories does the retailer carry (called a retailer's breadth), how many SKUs do they carry in each category (called depth), and how much inventory do they stock of each SKU? The breadth *versus* depth trade-off is a fundamental strategic choice and decided at store or chain levels, and so only the number of SKUs given an assigned space and the amount of inventory to stock are category decisions. As consumers are heterogeneous in their preferences, we expect that adding a SKU in a category should increase sales and should also generate a better perception of variety. However, carrying more SKUs implies higher operational costs. These operational costs are usually decomposed in the assortment literature in two parts: one related to the quantity stocked and the other a fixed cost for each product included (see, eg, Dobson and Kalish, 1993). The use of the number of SKUs as an input measurement follows the same reasoning. Assuming direct inventory costs are similar for all items in the category, the operational cost is directly related to the number of items in the assortment. In the absence of a direct measure of inventory cost, we use the number of products in the assortment as a proxy of the cost. We consider that this is a reasonable approximation as we only consider units with similar cost structures excluding, for example, fresh seafood and delicatessen products which require additional personnel to serve customers.

Outputs

Outputs are results that have been defined by store managers as desirable. They include not only direct economic results but others that may be related to the store market positioning. In this case we consider the following:

(1) *Sales* comprise the total sales volume over the period for all the SKUs in the category measured in US$.

(2) *Penetration*. Some products are important because most customers include them in their baskets. We measure penetration as the number of shopping baskets in the period that include at least one SKU from the category. Penetration is the usual objective for many categories.

(3) *Margin*. The economic contribution of the category is measured well by the net margin that is the sale value minus total costs, including operating costs. However, retail accounting systems do not always have the capability to prorate fixed costs to a category level, so we use the gross margin. This is the sale value minus the replenishment cost.

(4) *Share*. We include this measure as an indicator of the store's performance compared with its competitors. Consulting companies such as ACNielsen elaborate periodic reports on the share of all categories of all supermarkets in the store's relevant geographic area. Using this information, we can establish whether the performance of a category is similar to that of the competition related to the share of sales, that is, the relative importance of sales in the category with respect to competition nearby. Thus, a large share implies that customers use the focal store to buy products in the category more than from direct competitors. To facilitate interpretation, we normalize with respect to the average category share.

(5) *Perceived variety* has been recognized as one of the most important factors in determining store patronage. Several approaches to the measurement of an assortment's variety have been proposed in the literature, including

the availability of a favourite product, the amount of shelf space devoted to the category (Broniarczyk *et al*, 1998) and attribute-based models (Hoch *et al*, 1999). Considering our goal of using data that are easily accessible to store managers, we selected an attribute-based metric in which 'category trees', widely used in CM, provide a direct definition of attributes which can be used to measure variety without asking customers every time a new product is introduced in the assortment. More precisely, we used a weighted entropy index defined as follows (van Herpen and Pieters, 2002):

$$\text{Weighted Entropy} = -\sum_{i \in I} \gamma_i p_i \ln p_i \qquad (6)$$

where I is the total number of attribute levels and p_i denotes the proportion of products in the assortment with attribute level i. According to this definition, entropy is greater when more attribute levels occur in relatively balanced proportions, indicating higher levels of variety in the assortment. Thus, if a category carries several flavours and several package sizes, then it would have greater entropy compared with one carrying a few flavours of only one package size. Parameters γ_i are the relative importance of the level in the category. For example, in the soft drink category, cola might be more important than orange and lemon flavours, or in the oil category, sunflower oil might be more important than olive oil. These parameters need to be estimated and we do this using the analytic hierarchy process (AHP) using evaluations of three experts for each category, as described in Appendix A. This measurement technique is used to determine the relative importance of a set of activities or criteria representing judgments of multiple agents and has been employed in several marketing applications (Wind and Saaty, 1980). These parameters only need to be re-estimated once in a while, not for every application of the DEA model.

As we discussed earlier, an important requirement for a fair comparison between categories with different roles is including every possible objective that a category might be pursuing as an output in the evaluation. In accordance with the CM literature, we recognize five category roles (Hoch and Lodish, 1998) and our list of outputs properly covers them all. *Traffic builder* objectives are closely related to penetration and sales, *transaction builder* categories should excel in share and sales, *cash generator* objectives are captured by sales, and *profit contributors* should exhibit high margins. Finally, *image creators* might pursue several objectives, such as enhancing the shopping experience, perception of price or perceived quality. Shopping experience is usually related to decision variables not included in our analysis: display quality and general store ambience. To measure this, customer ratings on experience might be necessary; however, they are not commonly available. If they were, they could be incorporated into another output indicator. Perception of price is usually measured at the store level and not at the category level. Thus, we considered perceived variety as the only image creator objective; however, if information about other objectives were available, they could be considered.

Our focus on detecting inefficiency gives us some degree of flexibility to include a relatively large number of inputs and outputs. When using DEA, increasing the number of factors reduces the ability to find inefficient units. A relatively large number of variables provides a relative strong test of efficiency. When a unit is detected as inefficient in a high-dimensional space of inputs and outputs, it is a robust indication of a productivity gap. Dyson *et al* (2001) suggest that as a rule of thumb to achieve a reasonable level of discrimination, the number of units to be analysed needs to be at least $2 \times m \times s$ where m is the number of inputs and s is the number of outputs. Our selection of variables fulfils this recommendation in a very conservative manner. In our application we found that even with our relatively large number of factors, we were able to identify several inefficient units.

Model strengthening using weight restrictions

Considering that categories have multiple goals where they can perform well, we face the risk that no category is detected as inefficient. One way to improve the discriminatory power of the model is to incorporate weight restrictions in the model. As pointed out by Allen *et al* (1997), restrictions in DEA formulations can be incorporated in many different ways. To illustrate the effect of imposing these constraints in our setting, we restrict our attention to the case of direct restrictions on relative values of output weights of the form $v_k \leqslant \alpha v_h$, meaning that the relative importance that a DMU assigns to output k cannot exceed the importance of output h by more than α times. If the matrix $W=\{w_{kh}\}$ represents the dual variables associated with the weight constraints of outputs k and h ($k, h=1 \ldots s$), then including these conditions implies that in the dual formula we must modify Equation (3) associated with primal output weights as follows:

$$y_p = Y\lambda + W - \alpha W s' - s^+ \qquad (7)$$

There are two main factors that limit the use of these constraints in evaluating category performance in a retail store. First, weight constraints might neglect category roles, forcing units to excel in generating all outputs. In fact, one of the main motivations in restricting weights is to avoid some DMUs being assessed only on a small subset of their inputs and outputs. However, this is a feature we consider useful in using DEA to analyse the data. If weight constraints are imposed for all pairs of outputs (k, h), all units are evaluated using a composite outcome that considers all outputs in the model. In this case, it would be hard for a category that specializes in excelling in some of the store goals to be identified as operating efficiently. For example, *traffic builders* or *traffic creators* could be flagged as inefficient for not generating solid margins.

The second factor limiting the use of weight constraints is the need to elicit the relative importance of inputs and outputs. In some cases, weights can be derived from economic or engineering characteristics of the production technology (Dyson and Thanassoulis, 1988). In others, they represent managers' views regarding the value of each variable in the operation of the firm (Podinovski, 2004). According to our interviews with store and category managers, there is no clear economic relationship between different outputs. Moreover, it is difficult to make judgments about how important a store goal is with respect to others. For example, for decision makers, it is not clear how important *perceived variety* is against total *sales*, or *penetration* against *share*. The only exception for this is *margin*, against which decision makers feel they could define acceptable ranges of importance. Margin is a direct metric of short-term profitability and therefore they can interpret weight restrictions as the profits they are sacrificing in order to improve other store goals. Thus, we include in our analysis these forms of weight constraints for values of α=1.25 and α=2.0, reflecting the lower and upper bounds of α according to managers' opinions.

Empirical application

We applied our methodology at a Chilean supermarket belonging to a retail chain with 18 stores in seven cities, with local market shares ranging from 40 to 60% depending on the city. The evaluation was undertaken for a 2500 m^2 store with 18 cashiers, carrying more than 40 000 SKUs. Our data set contained monthly information on the 40 categories from the grocery sections of the store. We excluded the bakery, delicatessen and fresh seafood sections to compare only units with similar cost structures. In addition, all the sections we did not consider in the analysis require dedicated personnel to either cut, clean or weigh the products. From the 54 categories in the grocery section, we combined six categories to facilitate space measurement and we removed eight seasonal categories that are only available for a few weeks in the year. Table 1 displays the inputs and outputs we used for all the categories included in our empirical application.

Literature on DEA has discussed the implications of including zeros in the data (Charnes *et al*, 1991). In such cases, a unit could be identified as efficient without using all inputs or the product of all outputs. As pointed out by Thompson *et al* (1993), a careful analysis of the nature of the zeros in the data is required to avoid misleading conclusions about the efficiency scores. In our application, the appearance of zeros in the Promotion and Feature inputs is justified because they represent fundamental characteristics of the underlying decision-making process. While promotion is mainly devoted to increasing sales and market share and is mainly used for *transaction builder* or *cash generator* categories, features are meant to bring customers to the store and are therefore mainly used in relation to *traffic builder* categories. For other DEA applications with structural uses of zeros in data, see Byrnes *et al* (1984) and Thompson *et al* (1990).

Table 2 displays efficiency scores θ_p resulting from the direct application of model described in Equations (1)-(5), together with the scores of the model with weight constraints. In spite of our very conservative modelling approach in which we allowed for variable economies of scale and included a relatively large number of variables with respect to the number of units being analysed, the results indicate that 10 categories (25%) are detected as inefficient in producing the business goals that the supermarket pursues. Although the number of efficient categories is larger than in other DEA applications, it is important to recall that we are not interested in describing the production frontier or finding best practices. In our application, we try to identify categories that are underperforming in achieving store goals in order to inform store managers about short-term resource reallocations. As previously mentioned, the routine application of this model is in accordance with a continuous improvement management approach. Thus, the empirical results of our model show a valuable discriminatory power in identifying the units where managerial adjustment would have a positive impact on overall store productivity.

When looking at the efficiency results of the weight-constrained model, we can observe that the scores are almost invariant to these restrictions. With the exception of coffee and detergents, there is no change in the inefficient categories identified. As discussed in the Modelling section, we imposed the restriction that the relative importance assigned to any output cannot exceed α times the margin generated by the category. Thus, categories we identify as efficient achieve their goals without greatly sacrificing short-term profits.

An essential step in the identification of improvement opportunities is the proper assessment of the robustness of the efficiency scores. We started by analysing how different metrics to measure inputs and outputs affect efficiency scores. The normalization of variables using percentages or absolute values does not produce differences in the scores. To study how the evaluation is affected by data variations, we conducted a sensitivity analysis using the smoothed bootstrap technique suggested by Simar and Wilson (1998). Our choice of this technique is based on simplicity and data requirements. A formal description of the procedure is outlined in Appendix B. Box plots of the bootstrap samples for all categories detected as inefficient are displayed in Figure 1. By inspecting the distribution of the efficiency scores, we confirm that the inefficiencies detected are robust to variations in the data. We also note that some categories, such as legumes and chlorine, present large dispersion, providing further evidence of the need for sensitivity analysis in the general case.

We are interested in providing managerial recommendations to improve productivity. DEA models also provide guidance in this direction. Following Cooper *et al*'s (2006) suggestion, we ran a second model in which we fixed the efficiency value of θ_p and maximized the sum of slack variables subject to Equations (2)-(5). This helps us to discriminate between *technical* or weak efficiency and *Pareto-Koopmans* or strong efficiency, but more importantly it helps us to find the projection that indicates the reduction of resources that the category would need to accomplish to be efficient in producing the same amount of outputs. Let θ_p^* represent the

Table 1 Inputs and outputs

Category	*Inputs*				*Outputs*				
	Space	*Promotion*	*Feature*	*No. SKUs*	*Sales*	*Penetration*	*Margin*	*Share*	*Variety*
Oils	13.6	279.07	3	140	17 987.85	4760	2223.72	68.84	2.09
Rice	12.4	126.61	3	125	10 270.55	4314	1327.96	78.97	2.24
Sugar	4.9	1.15	0	31	10 480.78	5010	1550.38	60.77	1.24
Coffee	7.4	101.89	1	113	8271.34	1810	902.34	93.12	2.27
Pasta	18.5	47.39	8	269	11 995.95	8114	1815.57	85.45	2.51
Baking supplies	4.9	5.95	0	43	1406.42	836	212.27	95.62	2.26
Flour	7.4	239.54	0	18	4250.43	1542	606.71	74.34	1.2
Milk	34.6	31.75	4	104	19 695.04	3116	2317.41	89.2	1.26
Tea	19.8	131.17	0	225	10 262.2	4482	1341.41	89.3	2.75
Milk modifier	4.9	5.53	0	101	3310.87	964	492.73	99.08	1.82
Cereals	14.8	38.51	0	225	10 427.15	3060	1296.35	91.39	2.34
Baby food	5.2	0	12	101	1930.89	656	278.13	100	1.2
Cocktail	6.2	10.44	4	276	4938	1740	982.81	97	2.01
Snack	19.8	121.61	4	339	15 891.2	6672	3118.2	97	2
Mexican food	1.2	4.5	0	55	431.42	66	82.21	97	0.76
Condiments	19.8	226.06	6	559	20 866.09	11 064	3457.95	81.08	3.18
Canned Fruit	7.4	139.39	2	228	7175.45	1812	1021.54	79.02	2.06
Canned Seafood	12.4	179.24	3	302	14 212.8	3942	2249.23	87.65	2.14
Canned Vegetables	1.2	56.5	1	166	5404.31	1546	940.25	69.79	2.19
Creams	2.5	11.51	0	39	3928.21	1108	624.37	97.37	1.9
Grain legumes	7.4	16.92	1	75	2559.65	1004	406.33	70.39	1.88
Jams & jelly	6.2	27.51	3	127	3945.29	2386	615.97	84.65	2.52
Desserts & gelatin	9.9	10.41	0	190	2612.12	2310	413.76	97.93	2.72
Diet	12.4	6.51	0	110	3419.81	738	628.61	94.87	2.49
Fruit—dried	7.4	0	0	295	4773.24	624	885.43	92.11	3.15
Mashed potatoes	1.2	21.5	1	18	1027.63	466	153.94	86.08	0.54
Powder drinks	9.9	1.65	0	223	6893.74	8094	877.33	99.21	2.83
Tomato sauces	6.2	6.95	3	163	5089.59	3900	833.03	85.85	1.95
Soups	6.2	1.62	0	170	3561.44	3086	586.01	92.86	1.55
Chocolates	7.4	857.16	0	636	15 851.18	4408	3030.91	98.59	3.1
Cookies	27.2	422.33	0	440	18 979.08	14 694	3370.89	97.47	3
Candies & gums	9.9	7.03	0	445	9107.44	3730	1792.44	94.86	2.8
Detergents	27.2	405.29	4	295	29 114.14	4974	1423.54	97.83	1.95
Disposable diapers	23.4	64.94	2	393	13 975.8	1370	1290.93	95.29	1.84
Paper products	32.9	607.52	6	172	39 539.82	15 012	4418.04	90.36	1.79
Wrapping bags	4.9	88.84	0	26	4888.38	1756	558.79	80.97	0.68
Wax	8.6	11.8	0	111	3060.27	1214	585.86	84.3	1.84
Chlorine	14.8	51.6	0	60	5099	2466	873.13	84.28	1.5
Fresheners/deodorizers	3.7	0	1	192	3929.68	774	701.49	96.29	1.81
Insecticide	3.7	71.86	1	105	7460.16	1190	1346.21	96.81	1.63

optimal efficient score of the first stage and s^{-*} the optimal value of the slack variable of the second stage. Then the required amount of input for the unit p ($\hat{x}_p$) is given by:

$$\hat{x}_p = \theta_p^* x_p - s_p^{-*} \tag{8}$$

Note that for efficient categories θ_p*=1 and s_p^-*=0 and therefore $\hat{x}_p = x_p$. A value of $\theta_p^* < 1$ implies a *radial* or *scale* inefficiency, meaning that all inputs should be reduced by the fraction θ_p* while keeping the outputs constant to be considered efficient. A value of $s_j^-* > 0$ implies a *mix* inefficiency, meaning that input j can be reduced further, changing the relative proportion of inputs that are being used. Reductions in percentages of inputs for all inefficient categories are displayed in the left part of Table 3 showing that model detects both scale and mixture inefficiencies. For example, when looking at the wax category, we observe that the model recommends a reduction in all inputs of 52.48% from their current levels.[2] On the other hand, when looking at the nappies category, we observe a recommendation of 9.35% in scale reduction, but on top of that a large reduction in the number of SKUs being offered. As is illustrated by these two examples, the distinction between radial and mix inefficiencies is particularly relevant in our application. While mix inefficiencies suggest changes in decisions that are internal to the category operation, radial inefficiencies suggest that the relative importance of the category in the store needs to

[2]Feature cannot be reduced because in the period analysed, there were no products in this category in the catalogue.

Table 2 Category efficiency

Categories	*Base*	*Weight constraints*		*Categories*	*Base*	*Weight constraints*	
	θ_p	θ_p (α=1.25)	θ_p α=2.00)		θ_p	θ_p (α=1.25)	θ_p (α=2.00)
Oils	1	1	1	Legumes	0.414	0.381	0.406
Rice	0.916	0.916	0.916	Jams & jelly	1	1	1
Sugar	1	1	1	Desserts & gelatin	0.976	0.801	0.845
Coffee	1	0.956	1	Diet	1	1	1
Pasta	1	1	1	Dried fruit	1	1	1
Baking supplies	1	1	1	Mashed potatoes	1	1	1
Flour	1	1	1	Powder drinks	1	1	1
Milk	1	1	1	Tomato sauces	0.779	0.752	0.762
Tea	1	1	1	Soups	1	1	1
Milk modifiers	1	1	1	Chocolates	1	1	1
Cereals	1	1	1	Cookies	1	1	1
Baby food	1	1	1	Candies & gums	1	1	1
Cocktail	0.881	0.845	0.881	Detergents	1	0.74	0.854
Snack	1	1	1	Diapers	0.906	0.612	0.664
Mexican food	1	1	1	Paper products	1	1	1
Condiments	1	1	1	Wrapping bags	1	1	1
Canned fruit	0.478	0.463	0.468	Wax	0.416	0.409	0.415
Canned seafood	0.875	0.873	0.874	Chlorine	0.424	0.406	0.41
Canned vegetables	1	1	1	Fresheners/deodorizers	1	1	1
Creams	1	1	1	Insecticide	1	1	1

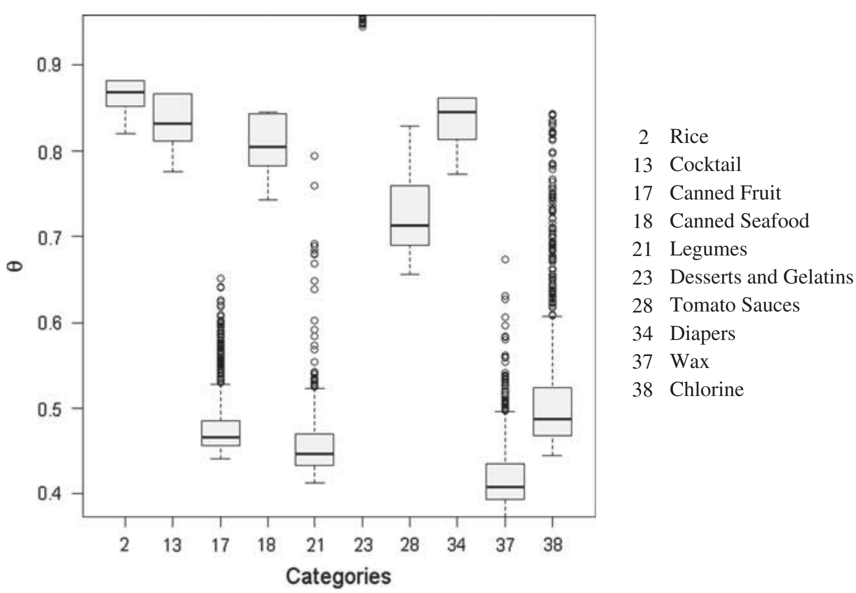

Figure 1 Bootstrap sample box-plots for inefficient DMUs.

be revised. Thus, radial inefficiencies provide a guide to store managers for reallocating common resources to satisfy the multiple goals defined by a CM scheme to a greater extent.

We have proposed the use of the results of the model to adjust input levels, but we could also use it to provide output goals for the categories. The definition of this benchmark is analogous to (8) and numerical results are displayed in the right part of Table 3. For example, given the space and number of SKUs in the nappies category, the goal should be to increase perceived variety by 19.10%.

Discussion and future research

Efficiency in resource usage is one of the fundamental strategic components in the retail industry. The widely used strategy of CM requires the continuous monitoring and evaluation of

Table 3 Category efficiency: Reductions

Inefficient DMUs	*Space*	*Promotion*	*Feature*	*No. SKUs*	*Sales*	*Penetration*	*Margin*	*Share*	*Variety*
Rice	7.57	7.57	17.37	7.57	0.82	2.18	—	8.24	—
Cocktails	9.89	9.88	99.60	31.30	11.66	10.84	—	—	8.03
Canned fruit	47.08	47.08	53.15	47.08	—	17.51	2.25	—	—
Canned seafood	11.50	11.50	0.12	13.83	—	43.14		—	—
Legumes	49.14	49.13	91.70	49.13	28.82	37.75	24.23	29.83	—
Desserts & gelatins	13.28	75.62	—	2.14	110.19	157.24	74.99	—	—
Tomato sauces	18.42	18.43	98.40	44.82	26.90	—	13.35	2.24	—
Nappies	13.50	9.35	9.35	46.45	—	314.01	16.71	—	19.10
Wax	52.48	52.48	—	52.48	24.81	35.90	—	5.30	
Chlorine	77.00	79.45	—	40.30	21.68	—	7.56	—	9.03

category performance. However, academics and practitioners have been surprisingly silent on the procedures for conducting such evaluations beyond the very limited use of simple efficiency ratios. In this article, we discuss a DEA-based methodology to evaluate category performance in a retail store. Our proposed approach allows us to compare business units with different goals, such as those we find in retail stores under a CM regime, and to detect those that are underperforming in achieving store business objectives. Thus, our proposed methodology can help store managers not only to identify sources of inefficiencies in terms of resource allocations, but also relieves them from assigning a rigid definition of category roles. The simultaneous evaluation of multiple objectives creates the opportunity for categories to contribute to more than a single business goal.

A key component in the methodology is the selection of variables that characterize the operation of the categories under evaluation. We propose a set of inputs and outputs that we believe are representative of the operations of supermarkets and that properly cover category roles reported in the literature. We believe that our list constitutes a good baseline for describing category operations, but it needs to be revised case by case to account for the specific conditions and goals of the store under analysis. For example, a retailer might consider that perception of price is a key component of its marketing strategy. In this case, the supermarket could run a regular survey among its clients asking them to rate the price of each category compared with competitive stores and include such a variable as an output in the evaluation. Our variable selection process is also limited by the availability of data. For example, were accurate inventory costs available for all categories, we could use that information to complement that provided by the number of SKUs in the category.

We illustrate the use of our model by applying it to the evaluation of several categories in a South American supermarket. Our results show that the proposed methodology has significant discriminatory power to detect categories that are inefficiently managed. In the empirical application, we found that 25% of the units were detected as inefficient despite being very conservative in our modelling and variable selection.

Several alternatives are available to increase the requirement to certify efficiency in terms of strengthening the formulation; most notably, we can incorporate restrictions to the weight coefficients that define the efficiency ratio. For example, we could impose upper bounds to the weights of one variable or require that the weight of one characteristic needs to be smaller than other to account for strategic orientations. When imposing weight constraints on outputs relative to margin, we found that the identity of inefficient categories remains almost unchanged. The set of inefficient categories is also robust to data variations, but efficient scores can exhibit significant dispersion and therefore sensitivity analysis needs to be performed. We use a smoothed bootstrap technique but other complementary tools, such as windows analysis or metric approaches, could also be used (Cooper *et al*, 2006).

Category performance needs continuous monitoring. We propose to apply our model on a regular basis and to keep a record of previous efficiency indexes. We consider evaluation results should be used as a filter that enables the store manager to identify easily those categories that are underperforming. The movement from here to actual decisions needs to consider the nature of the inefficiency and the corresponding corrective actions. For example, major decisions, such as redefining category floor space, require thorough analysis of performance over several periods to mitigate the influence of seasonality and demand shock factors. Other decisions, such as assortment reduction, can be taken after analysing only a few periods. Output benchmarks should always be provided to category managers to inform them of realistic targets for all outputs.

In terms of methodology, we identify several avenues for future research. Our model conducts analysis for categories in a single store because we consider that most CM decisions are done at that level. As we discussed in the problem definition, the pure comparison of each category with their counterparts in other stores does not provide a helpful output for decision makers. However, we recognize that across-store comparisons provide valuable information that can enhance our performance assessment. To include this information, a hierarchical model needs to be accommodated. Here each weight coefficient defining the efficiency ratio could be decomposed into two parts: a category component that captures commonalities across stores and store-specific components that account for different competitive environments or local strategic orientations.

We have described the basic components of what we believe constitutes a helpful tool for store managers. Implementation is simple, the data requirements are in line with current systems and the resulting linear problems are computationally fast to solve. However, a full decision support system would require additional components. Over and above a friendly set of user reports, the system should allow the user to include additional constraints, add and remove variables and track the time series of past efficiency scores. The availability of all these elements would certainly bolster the ability of retailers to undertake continuous improvements in their productivity, thus positively impacting long-term profitability.

Acknowledgements—The first two authors gratefully acknowledge partial funding by FONDEF (project D06I1015). The first author also acknowledges the Millennium Institute on Complex Engineering Systems.

References

Allen R, Athanassopoulos A, Dyson RG and Thanassoulis E (1997). Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions. *Annals of Operations Research* **73**(0): 13–34.

Assaf GA, Barros C and Sellers-Rubio R (2011). Efficiency determinants in retail stores: A Bayesian framework. *Omega* **39**(3): 283–292.

Barros CP (2006). Efficiency measurement among hypermarkets and supermarkets and the identification of the efficiency drivers: A case study. *International Journal of Retail & Distribution Management* **34**(2): 135–154.

Barros CP and Alves CA (2003). Hypermarket retail store efficiency in Portugal. *International Journal of Retail & Distribution Management* **31**(11): 549–560.

Bandyopadhyay S, Rominger A and Basaviah S (2009). Developing a framework to improve retail category management through category captain arrangements. *Journal of Retailing and Consumer Services* **16**(4): 315–319.

Bell DR and Lattin JM (1998). Shopping behavior and consumer preference for store price format: Why 'large basket' shoppers prefer EDLP. *Marketing Science* **17**(1): 66–88.

Broniarczyk SM, Hoyer WD and McAlister L (1998). Consumers' perceptions of the assortment offered in a grocery category: The impact of item reduction. *Journal of Marketing Research* **35**(2): 166–176.

Byrnes P, Färe R and Grosskopf S (1984). Measuring productive efficiency: An application to Illinois strip mines. *Management Science* **30**(6): 671–681.

Charnes A, Cooper WW and Thrall RM (1991). A structure for classifying and characterizing efficiency and inefficiency in data envelopment analysis. *Journal of Productivity Analysis* **2**(3): 197–237.

Charnes A, Cooper WW, Lewin A and Seiford LM (1994). *Data Envelopment Analysis: Theory, Methodology, and Application*. Kluwer Academic Publishers: Norwell, MA.

Chong J-K, Ho T-H and Tang CS (2001). A modeling framework for category assortment planning. *Manufacturing and Service Operations Managements* **3**(3): 191–210.

Cooper WW, Seiford LM and Zhu J (2004). Data envelopment analysis: History, models and interpretations. In: Cooper WW, Seiford LM and Zhu J (eds). *Handbook on Data Envelopment Analysis*. Springer: New York, pp 1–39.

Cooper WW, Seiford LM and Tone K (2006). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-solver Software*. Springer: New York.

de Jorge Moreno J (2006). Regional regulation analysis of performance in Spanish retailing. *International Journal of Retail & Distribution Management* **34**(10): 773–793.

de Jorge Moreno J (2010). Productivity growth of European retailers: A benchmarking approach. *Journal of Economic Studies* **37**(3): 288–313.

Dhar SK, Hoch SJ and Kumar N (2001). Effective category management depends on the role of the category. *Journal of Retailing* **77**(2): 165–184.

Dobson G and Kalish S (1993). Heuristics for pricing and positioning a product-line using conjoint and cost data. *Management Science* **39**(2): 160–175.

Donthu N and Yoo B (1998). Retail productivity assessment using data envelopment analysis. *Journal of Retailing* **74**(1): 89–105.

Dussart C (1998). Category management: Strengths, limits and developments. *European Management Journal* **16**(1): 50–62.

Dyson RG and Thanassoulis E (1988). Reducing weight flexibility in data envelopment analysis. *Journal of the Operational Research Society* **39**(6): 563–576.

Dyson RG, Allen R, Camanho AS, Podinovski VV, Sarrico CS and Shale EA (2001). Pitfalls and protocols in DEA. *European Journal of Operation Research* **132**(2): 245–259.

Ferrier GD and Knox Lovell CA (1990). Measuring cost efficiency in banking: Econometric and linear programming evidence. *Journal of Econometrics* **46**(X): 229–245.

Gauri DK, Pauler JG and Trivedi M (2009). Benchmarking performance in retail chains: An integrated approach. *Marketing Science* **28**(3): 502–515.

Grewal D, Levy M, Mehrotra A and Sharma A (1999). Planning merchandising decisions to account for regional and product assortment differences. *Journal of Retailing* **75**(2): 405–424.

Haas DA and Murphy FH (2003). Compensating for non-homogeneity in decision-making units in data envelopment analysis. *European Journal of Operational Research* **144**(3): 530–544.

Hall P (1986). On the number of bootstrap simulations required to construct a confidence interval. *The Annals of Statistics* **14**(4): 1453–1462.

Hoch SJ and Lodish LM (1998). *Store brands and category management*. Wharton School University of Pennsylvania Working Paper.

Hoch SJ, Bradlow ET and Wansink B (1999). The variety of an assortment. *Marketing Science* **18**(4): 527–546.

Ingene CA (1982). Labor productivity in retailing. *The Journal of Marketing* **46**(4): 75–90.

Kamakura WA, Lenartowicz T and Ratchford BT (1996). Productivity assessment of multiple retail outlets. *Journal of Retailing* **72**(4): 333–356.

Keh HT and Chub S (2003). Retail productivity and scale economies at the firm level: A DEA approach. *The International Journal of Management Science* **31**(2): 75–82.

Kök AG, Fisher M and Vaidyanathan R (2008). Retail assortment planning: Review of literature and industry practice. In: Agrawal N and Smith SA (eds). *Retail Supply Chain Management. Quantitative Models on a Empirical Studies*. Springer: New York.

Lin HT (2011). Efficiency analysis of chain stores: A case study. *Journal of the Operational Research Society* **62**(7): 1268–1281.

Nielsen AC, Heller A and Karolefski J (2005). *Consumer-Centric Category Management: How to Increase Profits by Managing Categories based on Consumer Needs*. John Wiley & Sons Inc.: Hoboken, NJ.

Perrigot R and Barros CP (2008). Technical efficiency of French retailers. *Journal of Retailing and Consumer Services* **15**(4): 296–305.

Podinovski VV (2004). Production trade-offs and weight restrictions in data envelopment analysis. *Journal of the Operational Research Society* **55**(12): 1311–1322.

Ratchford BT and Stoops G (1988). A model and measurement approach for studying retail productivity. *Journal of Retailing* **64**(3): 241–263.

Russell GJ and Petersen A (2000). Analysis of cross category dependence in market basket selection. *Journal of Retailing* **76**(3): 367–392.

Saaty TL (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences* **1**(1): 83–98.
Sexton TR, Sleeper S and Taggart RE (1994). Improving pupil transportation in North Carolina. *Interfaces* **24**(1): 87–103.
Simar L and Wilson PW (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science* **44**(1): 49–61.
Sheather SJ (2004). Density estimation. *Statistical Science* **19**(4): 588–597.
Tavares G (2002). *A bibliography of data envelopment analysis*. Rutcor Research Report, Rutgers Center for Operations Research.
Thomas RR, Barr RS, Cron WL and Slocum JW (1998). A process for evaluating retail store efficiency: A restricted DEA approach. *International Journal of Research in Marketing* **15**(5): 487–503.
Thompson RG, Langemeier LN, Lee C-T, Lee E and Thrall RM (1990). The role of multiplier bounds in efficiency analysis with application to Kansas farming. *Journal of Econometrics* **46**(1): 93–108.
Thompson RG, Dharmapala PS and Thrall RM (1993). Importance for DEA of zeros in data, multipliers, and solutions. *Journal of Productivity Analysis* **4**(4): 379–390.
van Herpen E and Pieters R (2002). The variety of an assortment: An extension to the attribute-based approach. *Marketing Science* **21**(3): 331–341.
Wind Y and Saaty TL (1980). Marketing applications of the analytic hierarchical process. *Management Science* **26**(7): 641–658.

Appendix A

Analytic hierarchical process

The starting point for applying the AHP is to structure the problem in a hierarchy (Saaty, 2008). In our problem, the hierarchy is naturally given by the product tree where each attribute in the category represents a layer and each attribute level is a branch in the tree (Chong *et al*, 2001). To find the relative weight γ_i of attribute importance, a small panel of three experts needs to complete the relative importance of attribute in each level in the hierarchy. For example, to assess the relative importance of flavours in carbonated drinks, experts need to complete a table such as the one in the left-hand panel of Table A1; this translates into Matrix A, displayed in right-hand panel of the same table.

Table A1 Illustration of relative weight assessments

Flavour	*Cola*	*Orange*	*Lemon*
Cola	1	3	7
Orange	1/3	1	5
Lemon	1/7	1/5	1

$$A = \begin{bmatrix} 1 & 3 & 7 \\ 1/3 & 1 & 5 \\ 1/7 & 1/3 & 1 \end{bmatrix}$$

Then, the relative importance of each attribute level is given by:

$$\gamma_i = \sqrt[n]{\prod_{j=1}^{n} a_{ij}} \Big/ \sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{kj}} \quad \text{(A.1)}$$

By applying this equation to our example, we get (γ_{cola}, γ_{orange}, γ_{lemon})=(0.649, 0.279, 0.072). The same procedure is applied to all levels in the tree to derive the importance of each attribute level in the category.

Appendix B

Smoothed bootstrap technique

The smoothed bootstrap technique applies kernel density estimation to the empirical distribution of θ_p^* and then samples from the resulting continuous distribution (Simar and Wilson, 1998).

1. Compute efficiency scores θ_p^* for each category p.
2. Generate a random sample $\{\beta_1^*, \ldots, \beta_n^*\}$ from F, the empirical distribution of θ_p^*

$$F(t) = \begin{cases} n^{-1}, & t = \theta_p^*,\, p = 1, \ldots, n \\ 0 & \text{otherwise} \end{cases} \quad \text{(B.1)}$$

3. Generate a smooth sample

$$\tilde{\theta}_p^* = \begin{cases} \beta_p^* + h\varepsilon_p, & \beta_p^* + h\varepsilon_p \leqslant 1 \\ 2 - \beta_p^* - h\varepsilon_p & \text{otherwise} \end{cases} \quad \text{(B.2)}$$

   where h is the bandwidth of a kernel estimation of the density of θ_p^*. To determine the actual value of h, we use Silverman's Rule of Thumb (Sheather, 2004) which in our application corresponds to h=0.0735. Also, ε_p is a random draw from a standard normal distribution.
4. Define the bootstrap sample by correcting the variance as follows:

$$\tilde{\theta}_p^* = \overline{\beta}^* + \frac{1}{\sqrt{1 + h^2/\hat{\sigma}_\theta^2}} \left(\tilde{\theta}_p^* - \overline{\beta}^*\right) \quad \text{(B.3)}$$

   where $\hat{\sigma}_\theta^2$ is the variance of the smooth sample $\left\{\tilde{\theta}_p^*\right\}$
5. Compute a matrix of input X_b where each column p is given by

$$x_{pb} = \left(\theta_p^* \big/ \tilde{\theta}_{pb}^*\right) x_p \quad \text{(B.4)}$$

6. Compute the bootstrap estimate of efficiency index θ_{pb}^* for each category by solving the optimization problem given by Equations (1)-(5).
7. Repeat steps 2-6 for b=1, … , B. In our application and following the recommendation of Hall (1986), we used B=1000.

Received 6 March 2012;
accepted 26 September 2013 after two revisions