Digital Signal Processing 33 (2014) 156–168

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Statistical nonlinear analysis for reliable promotion decision-making

Cristina Soguero-Ruiz [a,*], Francisco Javier Gimeno-Blanes [b], Inmaculada Mora-Jiménez [a], María del Pilar Martínez-Ruiz [c], José Luis Rojo-Álvarez [a,d]

[a] Department of Signal Theory and Communications, Universidad Rey Juan Carlos, Fuenlabrada, Spain
[b] Department of Signal Theory and Communications, Universidad Miguel Hernández, Elche, Spain
[c] Area of Commercialization and Market Research, Universidad de Castilla La Mancha, Albacete, Spain
[d] Prometeo, Electric and Electronic Department, Universidad de las Fuerzas Armadas ESPE, Ecuador

ARTICLE INFO

Article history:
Available online 4 July 2014

Keywords:
Nonlinear methods
Nonparametric bootstrap
Aggregated data
Promotion sales
Retailer
Decision support

ABSTRACT

New economic conditions have led to innovations in retail industries, such as more dynamic retail approaches based on flexible strategies. We propose and compare different approaches incorporating nonlinear methods for promotional decision-making using retail aggregated data registered at the point of the sale. Specifically, this paper describes a reliable quantification tool as an effective information system leveraged on recent and historical data that provides managers with an operative vision. Furthermore, a new set of indicators are proposed to evaluate the reliability and stability of the data model in the multidimensional feature space by using nonparametric resampling techniques. This allows the user to make a clearer comparison among linear, nonlinear, static, and dynamic data models, and to identify the uncertainty of different feature space regions, for example, corresponding to the most frequent deal features. This methodology allows retailers to use aggregated data in suitable conditions that will result in acceptable confidence intervals. To test the proposed methodology, we used a database containing the sales history of representative products registered by a Spanish retail chain. The results indicate that: (1) the deal effect curve analysis and the time series linear model do not provide enough expressive capacity, and (2) nonlinear promotional models more accurately follow the actual sales pattern obtained in response to the implemented sales promotions. The quarterly temporal analysis conducted enabled the authors to identify long-term changes in the dynamics of the model for several products, especially during the early stage of most recent economic crisis, consistent with the information provided by the reliability indices in terms of the feature space. We conclude that the proposed method provides a reliable operative tool for decision support, allowing retailers to alter their strategies to accommodate consumer behavior.

1. Introduction

Traditionally, scholars have argued that customers tend to adapt their buying behavior during adversity and economic downturn [1,2]. This effect has been evidenced during the 2008 economic crisis, with a significant effect in the consumer goods industry. As a consequence, there has been a substantial change in commercial and marketing strategies to adapt to the consumer's behavior. In particular, retailers have reallocated commercial efforts from other marketing instruments toward increasing the direct price-deals, as a key tool to attract consumer attention. However, despite the extensive literature devoted to finding accurate promotional models in terms of pricing effects [3,4], no consensus has been reached on whether the promotional model yields better margins or benefits for either at retailers or manufacturers.

Researchers have proposed several hypotheses for promotional modeling. In particular, the deal effect curve (DEC) [3,5–7] presents a static model for shaping the price-demand elasticity, whereas conventional time series (TS) analysis and the well-known Box–Jenkins methodology add a dynamic approach to retail sales forecasting analysis [8,9]. However, linear or static models can only loosely explain complex interactions among products and sales. Scholars have extensively studied nonlinear machine learning approaches in attempt to develop methods that better follow the human behavior, for example, artificial neural networks and support vector machines [10–13].

From a retail manager's viewpoint, sales forecasting is essential not only to set the right pricing for an individual product [14] but also to define the promotional structure that maximizes benefits within a category as a whole [15]. The same rational applies

* Corresponding author.
E-mail address: cristina.soguero@urjc.es (C. Soguero-Ruiz).

http://dx.doi.org/10.1016/j.dsp.2014.06.014
1051-2004/

to individual customer behavior with regard to the total impact of a certain promotional strategy [15–19]. As a consequence, promotional models built on market-level data are considered as the best suited to describe the market behavior. Executive decisions are mainly based on this kind of information, especially for those retail chains accounting for a significant market share. Although it is evident that aggregated retail sales forecasting could potentially improve store sales prognosis [10], nevertheless, many authors have warned against the biasing risk during the aggregation process [20].

For a decision-making tool to be an efficient instrument for promotional retail management, it must be designed to be operative and reliable. To be operative, the retail management tool should be able to handle data models that: (1) can be better described TS dynamics, static paradigms, or even by both; and (2) can be better represented by linear or by nonlinear dynamics. To be reliable, the tool must be more robust when working with aggregated data than working with store level data, but also must ensure an adequate aggregation process. In addition, the tool should provide a simple way for the researcher to visualize its statistical properties in the feature space.

In this paper, we propose an operative and reliable analysis tool for promotional decision making based on retail aggregated data. The main contribution from a digital signal and data processing viewpoint is the proposal of a new set of indicators for evaluating the reliability and stability of a data model in terms of multidimensional feature space rather than a single merit figure for the model (e.g., the mean squared error). These indicators allow the user to identify the uncertainty of different feature space regions, for example, unusual promotion conditions. Using the statistical processing available, we can study the performance of different algorithms and different feature spaces. The use of aggregate data in suitable conditions yields moderate and acceptable confidence intervals in these feature spaces.

2. Background

In this section, we present a brief marketing literature review. We note relevant data aggregation precedents and summarize conventional static, dynamic, and learning-based nonlinear promotional sales models.

Data aggregation at chain level. Previous research has considered three levels of aggregation: store, chain, and market levels. At the *store-level*, data can characterize consumers' behavior (by considering buying habits such as products, and units to evaluate loyalty and churn rates), as well as brand or product sales (by aggregating sales). Household information for each product category can also be used to analyze the individual brand sales behavior and pricing effects can also be analyzed [21]. Further aggregation at *chain-level*, or even at *market-level*, integrates the information for brands or categories to provide accumulative effects [22].

According to published research analysis, each level of aggregation may introduce bias, which depends to a great extent on the aggregation method, thus limiting the generalization capabilities of the forecasting model. In [23], the authors analyze bias effect by comparing sales estimates at both store and chain level, and conclude that bias may be related to heterogeneous marketing strategies within stores. The authors also note that relevant information, such as marketing strategies followed by competitive retailers, is not reported or registered through scanner datasets. Other studies use different approaches to address model heterogeneity and bias among stores. For example, [24] proposes a random coefficient demand model to avoid bias when data aggregated across stores with heterogeneous promotional activity are considered. However, bias may not be fully removed due to substitutive effects, competing products and heterogeneity; therefore, in the current study, we followed the methodology in [25], in which bias can be mitigated by aggregating data across stores with homogeneous marketing activities.

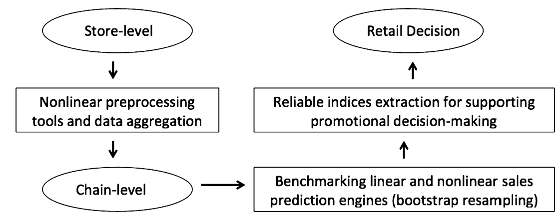


Fig. 1. Schematic of the proposed chain-level analysis.

DEC promotional modeling. The relationship between sales and temporary price discounts can be shown in the DEC representation [26], which accounts for the static sales elasticity and pricing promotion effects. The DEC shape provides information about different phenomena, such as direct discount effects, cross-effects generated from other products promotions, and concurrent sales initiatives on different promotional media [5]. Direct discount evaluation is often the first DEC analysis for any product, even though exogenous variables effects must always be considered because products do not exist in isolated markets. For a more detailed promotional sales model development, additional information must be incorporated to the DEC analysis. Therefore, research has proposed such modeling techniques as linear and nonlinear regression [27,26], to provide a more comprehensive description of the DEC. Furthermore, complex DEC shapes and cross-effects have been better represented with nonlinear statistical learning methods [5,28–31].

TS for promotional modeling. Promotional activities typically exhibit a strong temporal dependence, which suggests that certain models taking into account temporal variations could yield better results than static DEC. In this setting, the statistically well-founded TS analysis, has received a great deal of attention in the last decade of the twenty-first century, due to the vast amount of data available from electronic records and media (e.g., scanner data), which allows both the cross-sectional and longitudinal analyses [32]. Researchers have used TS techniques for forecasting marketing variables and for evaluating specific situations [33]. New tools based on TS have proliferated in recent years to support general decision-making and especially in marketing activities [32]. For example, autoregressive moving-average (ARMA) modeling provides a well-developed general framework to analyze time series. It can be further extended to take into account exogenous variables (so-called ARMAx models) to improve their predictive capabilities. A multivariate version of ARMA models, the vector ARMA, allows adjusting models in which the dependent variable can be explained by multiple TS [33].

Statistical learning for promotional modeling. Other researchers have proposed nonlinear statistical learning algorithms, including such classic nonparametric methods as k nearest neighbors (k-NN) and kernel estimators [34], as well as the new learning techniques such as neural networks and support vector machines [12,13,35–38]. It is worthwhile to note that nonlinearity, nonnormal errors, and heteroscedasticity are automatically harmonized by these kinds of methods. Although it is not always explicit, a significant part of the promotional marketing analysis literature suggests that nonlinear methods can often provide better models for promotional dynamics.

3. Proposed chain level analysis

In line with the research presented in the preceding section, we propose a three stages chain-level analysis as shown in Fig. 1. First, we use a signal preprocessing method to aggregate data

Table 1
Summary of symbols and their descriptions. The order in the table is according to lowercase and uppercase in Latin and Greek alphabet, respectively.

Notation	Description
e_i	Model residual for the i-th product
$f_{\hat{S}_i(\boldsymbol{\Psi}_i)}(\hat{S}_i(\boldsymbol{\Psi}_i))$	Statistical distribution of sold units as a function of the input space
$f^*_{\hat{S}_i(\boldsymbol{\Psi}_i)}(\hat{S}_i(\boldsymbol{\Psi}_i))$	Estimation of the statistical distribution of hypersurface
$k\text{-NN}_{\Xi^{n_0}}$, $k\text{-NN}_{\Psi^{n_0}}$	k-NN estimator with Ξ^{n_0}, Ψ^{n_0}
$\boldsymbol{q}, \boldsymbol{q}^*$	Residuals and Bootstrap residuals for *Model 1*
$\mathbf{r}, \boldsymbol{r}^*$	Residuals and Bootstrap residuals for *Model 2*
$s_{i,k}(n)$	Sold units for the i-th product at store k in week n
$t(\cdot)$	Merit figure calculation operator
$BL_i^R(\boldsymbol{\Psi}_i)$	Estimated baseline sale units in region R for the i-th product
$BLRI_i^R(\boldsymbol{\Psi}_i)$	Baseline Relative Index in region R for the i-th product
$CI_{\hat{S}_i}$	Confidence Interval (CI) for the estimated sales
$CI^{u,R}_{(\hat{S}_i)}$	Upper limit of CI for the estimated sales in region R
$CI^{l,R}_{(\hat{S}_i)}$	Lower limit of CI for the estimated sales in region R
$DR_i^R(\boldsymbol{\Psi}_i)$	Dynamic Range in region R for the i-th product
$DRI_i^R(\boldsymbol{\Psi}_i)$	Dynamic Range Index in region R for the i-th product
E_p	Expectation conditioned to price
F	Method used for estimation (DEC, AR, ARx or k-NN)
MAE_I	Mean absolute error for the i-th product
$P_{i,k}(n)$	Price for the i-th product at store k in week n
$P_i(n)$	Central price proposed by headquarters for the i-th product in week n
$\hat{P}_i(n)$	Estimated central price for the i-th product in week n
R_i	Difference between maximum and minimum estimated sold units for the i-th product
$S_i(n)$	Aggregated sold units for the i-th product in week n
$\hat{S}_i(n)$	Estimated aggregated sold units for the i-th product in week n
$\hat{S}_i(\boldsymbol{\Psi}_i)$	Estimated aggregated sold units for the i-th product depending on the feature space
Φ	Estimation operator (median)
$\mathbf{V}_i = \{\boldsymbol{\Psi}_i(n), S_i(\boldsymbol{\Psi}_i)\}$	Set of observations consisting of the feature vector and corresponding sold units
$VC^R_{\hat{S}_i}(\boldsymbol{\Psi}_i)$	Variation coefficient index for the i-th product in region R
$X_{i,k}(n)$	Uncertainty term for i-th product at store k in week n
ϕ_t, θ_j	AR, ARx model parameters
$\mu^R_{\hat{S}_i}(\boldsymbol{\Psi}_i)$, $\sigma^R_{\hat{S}_i}(\boldsymbol{\Psi}_i)$	Mean, Standard deviation of estimated sales in region R for the i-th product
$\hat{Q}^*(b)$	Replication estimator of the merit figure
$\varepsilon(\hat{P}_i)$	Neighborhood of the estimated price for the i-th product
Θ_i	Feature extraction operator for prices series for the i-th product
Θ^{n_0}	Feature extraction operator for prices series for a temporal depth n_0
Ξ_i	Feature extraction operator for sold units series for the i-th product
Ξ^{n_0}	Feature extraction operator for sold units series for a temporal depth n_0
$\boldsymbol{\Psi}_i = [\boldsymbol{\Theta}_i^T, \boldsymbol{\Xi}_i^T]^T$	Feature vector for the i-th product
$\boldsymbol{\Psi}_i^R$	Feature vector in region R for the i-th product
$\boldsymbol{\Psi}^{n_0}$	Feature vector for a temporal depth n_0

from store-level to chain-level, based on *a-priori* considerations and simple morphological analysis. Second, we propose a generic promotional sales forecasting tool that can account for static and time-varying dynamics models in a given product, while maintaining a simple and compact mathematical form. Decisions about different plausible models are determined by their comparative benchmarking by means of nonparametric resampling statistical tests (formally introduced in [12]). Finally, the new proposed statistical indices are defined in the feature space and calculated for each product using resampling techniques.

3.1. General forecasting promotional model

Table 1 summarizes the notation used throughout the paper. We will generally consider data available at discrete time n, mostly consisting of prices and sold units in a weekly time period. Accordingly, $P_{i,k}(n)$ ($s_{i,k}(n)$) denotes the price (the number of sold units) for the i-th product at store k during week n, where $i \in \{1,\dots,I\}$, $k \in \{1,\dots,K\}$ and $n \in \{1,\dots,N\}$, with I, K and N being the total number of products, stores and weeks, respectively. Recall that $P_i(n)$ represents the price proposed by headquarters (HQ), which should be identical for the same product and week in all stores, however, day-by-day knowledge shows that prices are often different at each store due to promotional local decisions. This variability may be related to human errors during scanning process at cashier, special discounts applied due to damaged items, errors in the information systems, or even changes in prices due to local strategies. Store and central prices can be related by the following expression,

$$P_{i,k}(n) = P_i(n) + X_{i,k}(n) \tag{1}$$

where X is an uncertainty term.

Because we are interested in decision making according to central prices, we will need to approximate them by an adequate estimation operator Φ, this is,

$$\hat{P}_i(n) = \Phi\{P_{i,k}(n)\} \tag{2}$$

In contrast, sold units can be readily aggregated at the chain-level ($S_i(n)$) by the accumulative sum across stores, that is,

$$S_i(n) = \sum_{k=1}^{K} s_{i,k}(n) \tag{3}$$

After data preprocessing, a general forecasting model for the i-th product can be written as

$$S_i(n) = \hat{S}_i(n) + e_i(n) \tag{4}$$

$$\hat{S}_i(n) = F\big(\boldsymbol{\Theta}_i\{\hat{P}_i(n)\}, \boldsymbol{\Xi}_i\{S_i(n)\}\big) \tag{5}$$

where $\hat{S}_i(n)$ are the forecasted values of aggregated sold units at time n for the i-th product; e_i are the model residuals; operator F stands for the method used for estimation, such as DEC analysis, linear TS, or nonlinear statistical learning algorithms; and $\boldsymbol{\Theta}_i$, $\boldsymbol{\Xi}_i$, denote the features extracted from prices and sold units series for the i-th product. Note that operator F gives a data description in the so-called *feature space*, defined by $\boldsymbol{\Theta}_i$ and $\boldsymbol{\Xi}_i$ feature (column) vectors, which can be concatenated in a simple feature vector given by $\boldsymbol{\Psi}_i = [\boldsymbol{\Theta}_i^T, \boldsymbol{\Xi}_i^T]^T$. In the following, we will use both $\hat{S}_i(n)$ and $\hat{S}_i(\boldsymbol{\Psi}_i(n))$ to denote the estimated number of sold units at time n for the i-th product.

The DEC model with just self-product effects can be readily expressed by $\boldsymbol{\Psi}_i(n) = [\hat{P}_i(n)]$, as follows,

$$\hat{S}_i(\hat{P}_i) = E_P\{S_i(n) \mid \hat{P}_i(n) \in \varepsilon(\hat{P}_i)\} \tag{6}$$

where E_P denotes expectation conditioned to price. This averaging operator E_P is used to smooth the number of sold units with respect to observed pairs of price and sold units within a neighborhood of a given price, $\varepsilon(\hat{P}_i)$. Note that operator F in Eq. (5) is given by the price expectation within the $\hat{P}_i$ neighborhood.

AR(p) and *ARx*(p, q) models, with p autoregressive terms and q exogenous input terms, can be written as follows,

$$\hat{S}_i(n) = \sum_{t=1}^{p} \phi_t S_i(n-t) \tag{7}$$

$$\hat{S}_i(n) = \sum_{t=1}^{p} \phi_t S_i(n-t) + \sum_{j=0}^{q} \theta_j \hat{P}_i(n-j) \tag{8}$$

where ϕ_t, θ_j are the model parameters [39,40], and Eq. (5) is readily adapted by generating the following feature spaces,

$$\boldsymbol{\Theta}_i\{\hat{P}_i(n)\} = [\hat{P}_i(n), \ldots, \hat{P}_i(n-q)]^T \tag{9}$$

$$\boldsymbol{\Xi}_i\{S_i(n)\} = [S_i(n-1), \ldots, S_i(n-p)]^T \tag{10}$$

Thus, we can simply use $F(\boldsymbol{\Theta}_i, \boldsymbol{\Xi}_i) = \boldsymbol{\phi}^T \boldsymbol{\Xi}_i + \boldsymbol{\theta}^T \boldsymbol{\Theta}_i$ for the ARx model accounting for past prices as exogenous variables, and $F(\boldsymbol{\Theta}_i, \boldsymbol{\Xi}_i) = \boldsymbol{\phi}^T \boldsymbol{\Xi}_i$ for the AR model, which is only built on the self-dynamics of the observed time series without information about prices.

In addition, nonlinear data models can readily be taken into account with the model nomenclature in Eq. (5). For the current study, we decided to use the k-NN technique as a nonlinear method for promotional sales forecasting, due to its extreme simplicity and acceptable performance in many applications. The k-NN estimator in n_0 is a nonparametric procedure that just consider the k nearest data to $\boldsymbol{\Psi}_i(n_0)$, according to a given similarity or distance measurement [41], where k has to be previously fixed during the design procedure. Conventional distances are L_1 and L_2 norms, though different measurements have been proposed according to the nature of the data [42].

The k-NN estimator assumes that data close in the feature space $\boldsymbol{\Psi}$ provide similar values for the independent variable. Therefore, to estimate the number of sold units at any time n_0, $\hat{S}_i(n_0)$, the k-NN estimator uses a local neighborhood $\kappa(n_0)$ to provide the estimation as

$$\hat{S}_i(n_0) = F_{\kappa(n_0)}\{S_i(n)/n \in \kappa(n_0)\} \tag{11}$$

where $F_{\kappa(n_0)}$ is the weighted average operator that depends on distance and parameter k, and it is given by: $F_{\kappa(n_0)} = \frac{\sum_{l=1}^{k} w_l S_i(n_l)}{\sum_{l=1}^{k} w_l}$, where $w_l = 1/d_l$ depends on the distance to the l-th nearest neighbor (d_l).

3.2. Merit figures and generalization

We chose the most convenient representation from among the previously described promotional model (DEC, TS and nonlinear methods) for each product. We calculated the model quality in terms of mean and scatter. For the first one, we used the mean absolute error (*MAE*), given by

$$MAE_i = \frac{1}{N} \sum_{n=1}^{N} |\hat{S}_i(n) - S_i(n)| \tag{12}$$

For characterizing the scatter, the confidence interval (*CI*) is computed. *CI* refers to the range where the true value of a random variable lies with a given confidence level (often and here, 95%). Specifically, we calculated the variation of confidence interval (ΔCI) as follows:

$$\Delta CI_i = CI_i^u - CI_i^l \tag{13}$$

where CI_i^u and CI_i^l denote the upper and lower limits of CI_i, respectively.

The generalization capabilities of the developed models must be carefully guarantied, for which cross-validation techniques can be used. Here, we consider *hold-out* and *leave one out cross validation* (LOOCV) techniques, as far as they have been shown to give almost unbiased estimators of the generalization performance of statistical data models [43]. In the former, data are split into two subsets, one for training and another for validation (often 50% each). In the latter, the original database is divided into N subsets with the same size (as many subsets as available instances), one for validation and the remaining $N-1$ for training, and the process is repeated N times (each corresponding to a different sample being used out-of-sample for validation purposes), the model performance being eventually given by the averaged out-of-sample performance of each of the N partitions.

3.3. Statistical estimation with bootstrap resampling

In current work, we use the estimated distribution of sold units or performance indicators in two different scenarios, namely, for model comparison and for quality and reliability evaluation. In both cases, there is no clear evidence that the underlying statistical distribution can be expected to follow a known mathematical form; rather, different products in different scenarios will follow different dynamics and statistical laws. In this scenario, Bootstrap resampling techniques can provide a useful method for empirical nonparametric estimation of the probability density function (*pdf*) of statistical entities from a set of observations [44].

Bootstrap resampling for data model comparison. Bootstrap resampling has been successfully used before for estimating the performance of different methods and for selecting the design parameters of machine learning techniques in promotional modeling problems [12]. A clear cut-off test can be built, allowing us to benchmark the significance of the performance difference between different promotional models, for instance, linear vs nonlinear models, or models with different feature spaces. We established the two compared models *Model*$_1$ and *Model*$_2$, and a statistical decision criterion (i.e., a nonparametric hypothesis test) to determine whether the difference (in terms of some error magnitude, e.g. *MAE*, of the sales promotions out-of-sample estimations) between *Model*$_1$ and *Model*$_2$ are statistically significant or not. The use of a paired resampling scheme, in which the same resamples are to be considered in both benchmarked models, allows us to observe a paired comparison of errors in which the in-sample variability of the resampling is trivially controlled.

The nonparametric Bootstrap hypothesis test for model comparison can be summarized as follows:

1. The out-of-sample *MAE* of the residuals are given by $\boldsymbol{q} = [q_1, q_2, ..., q_N]$ and $\boldsymbol{r} = [r_1, r_2, ..., r_N]$ for $Model_1$ and $Model_2$, respectively.
2. A Bootstrap resample is a new set obtained from sampling with replacement of the elements of the original set ($\boldsymbol{q}$ and $\boldsymbol{r}$ in our case) yielding populational resamples $\boldsymbol{q}^*$ and $\boldsymbol{r}^*$, respectively. The resampling process is repeated B times, with b indexing the resampling number ($b = 1, \dots, B$). The b-th resamples $\boldsymbol{q}^*(b)$ and $\boldsymbol{r}^*(b)$ contain elements of $\boldsymbol{q}$ and $\boldsymbol{r}$, respectively, appearing zero, one, or several times. Recall that superscript * is used to highlight any terms from the Bootstrap resampling.
3. A Bootstrap replication of a quality estimator ($\hat{Q}^*(b)$) is constrained to the elements in the Bootstrap resample, i.e, $\hat{Q}^*(b) = t(\cdot)$, where $t(\cdot)$ is the merit figure calculation operator. In this work, models are compared according to two statistical operators, ΔMAE and ΔCI, given by

$$\Delta MAE^*(b) = MAE\big(\boldsymbol{q}^*(b)\big) - MAE\big(\boldsymbol{r}^*(b)\big) \tag{14}$$

$$\Delta CI^*(b) = \Delta CI\big(\boldsymbol{q}^*(b)\big) - \Delta CI\big(\boldsymbol{r}^*(b)\big) \tag{15}$$

4. A simple hypothesis test is given by H_0: $\Delta MAE = 0$ vs H_1: $\Delta MAE \neq 0$, and similarly for ΔCI.

Thus, the null hypothesis is not accepted if the 95% *CI* of the distribution ΔMAE or ΔCI does not overlap zero. A significantly negative (positive) difference statistic indicates that $Model_1$ outperforms $Model_2$ (and vice-versa), as seen in [12,13]. The use of both statistical measurements allows us to detect improvements in performance in terms of error magnitude and scatter.

Bootstrap resampling for model reliability indicators. The performance and reliability indicators described in next subsection are based on results obtained after applying Bootstrap resampling techniques, which were first proposed as nonparametric procedures for estimating the *pdf* of a random variable from an informative set of observations [44]. These indicators are obtained from the statistical distribution of sold units as a function of features being used; in other words, it is necessary to estimate $f_{S_i(\boldsymbol{\Psi}_i)}(S_i(\boldsymbol{\Psi}_i))$.

Let $\mathbf{V}_i = \{\boldsymbol{\Psi}_i, S_i(\boldsymbol{\Psi}_i)\}$ be a set of N observations consisting of the sold units and their corresponding feature vector in each week for the i-th product. Thus, $f_{S_i(\boldsymbol{\Psi})}(S_i(\boldsymbol{\Psi}))$ is approximated by an empirical estimation, obtained from sampling with replacement observations in $\mathbf{V}_i$. First, we construct a new set $\mathbf{V}_i^* = \{\boldsymbol{\Psi}_i^*, S_i^*(\boldsymbol{\Psi}_i)\}$, where superscript * represents, in general terms, any observation from the Bootstrap resampling process. Therefore, set $\mathbf{V}_i^*$ contains elements of $\mathbf{V}_i$ which are included none, one, or several times. We repeat the resampling process B times, yielding $\{\mathbf{V}_i^*(b)\}_{b=1}^B$. A Bootstrap replication of an estimation u_i is given by its calculation constrained to the observations in the Bootstrap resamples, thus is, $u_i^*(b) = F(\mathbf{V}_i^*(b))$.

In our scenario, this approach allows us to replicate multidimensional feature spaces by simply Bootstrapping the available observations, thereby yielding the necessary statistical distribution for their characterization. Then, we can estimate $f^*_{S_i(\boldsymbol{\Psi}_i)}(S_i(\boldsymbol{\Psi}_i))$ for the statistical distribution $f_{S_i(\boldsymbol{\Psi}_i)}(S_i(\boldsymbol{\Psi}_i))$. The data we use are sparse, that is, every product is not sold in all stores at each week. Therefore, we addressed the interpolation of sold units using the estimation $f^*_{S_i(\boldsymbol{\Psi}_i)}(S_i(\boldsymbol{\Psi}_i))$ obtained for each resample b. This allows to readily estimate the new set of quality and reliability indices proposed in this paper and described in the following section.

3.4. Model quality and reliability indicators

In contrast to conventional approaches, which use a single value of a merit figure to evaluate model performance, we propose here to use a new set of indicators to characterize both quality and reliability in a given region $R \in \boldsymbol{\Psi}_i$ ($\boldsymbol{\Psi}_i^R$). Note that different values for the same indicator can be obtained for different regions. As an intuitive example, indicators may be less reliable in regions with scarce, noisy or non-informative data.

A set of four quality indicators will be calculated using the B estimations $u_i^*(b) = F(\mathbf{V}_i^*(b))$, obtained through the B resamples $\{\mathbf{V}_i^*(b)\}_{b=1}^B$. These indicators are proposed to measure the reliability of model F, and are defined as follows.

(1) Variation Coefficient (VC) index. VC measures dispersion in relation to mean value. It is a useful statistic for comparing the degree of variation between two datasets, even when their means are drastically different. The lower the *VC*, the more reliable our predictions are. It can be written as

$$VC^R_{\hat{S}_i}(\boldsymbol{\Psi}_i) = \frac{\sigma^R_{\hat{S}_i}(\boldsymbol{\Psi}_i)}{\mu^R_{\hat{S}_i}(\boldsymbol{\Psi}_i)} \tag{16}$$

where $\sigma^R_{\hat{S}_i}(\boldsymbol{\Psi}_i)$ and $\mu^R_{\hat{S}_i}(\boldsymbol{\Psi}_i)$ are the standard deviation and mean, respectively, of sales estimations in region R.

(2) Confidence Intervals variation (ΔCI). We particularize ΔCI previously defined in Eq. (13) to calculate the reliability of the estimated sales in region R for the i-th product. The narrower the confidence interval, the lower the variability is, hence ΔCI can be used as a reliability measurement in a region R of the feature space, denoted as

$$\Delta CI^R_{\hat{S}_i}(\boldsymbol{\Psi}_i) = CI^{R,u}_{\hat{S}_i}(\boldsymbol{\Psi}_i) - CI^{R,l}_{\hat{S}_i}(\boldsymbol{\Psi}_i) \tag{17}$$

where $CI^{R,u}_{\hat{S}_i}(\boldsymbol{\Psi}_i)$ and $CI^{R,l}_{\hat{S}_i}(\boldsymbol{\Psi}_i)$ denote the upper and lower limits of the confidence interval, respectively.

(3) Baseline Relative Index (BLRI). A key concept in marketing research is baseline sales, typically defined as the sales of a given product when there are neither marketing promotions for this product, nor promotions for other interacting products [45,46]. Marketing managers widely use baseline sales to assess the profitability and effectiveness of marketing activities by investigating how promotions can affect baseline sales over time. In this setting, we created a new index to assess the accuracy of a promotional model in not only absolute but also in relative terms respect to the baseline. This can be achieved by normalizing the number of estimated sold units with respect to the baseline sales, and we define it as

$$BLRI^R_i(\boldsymbol{\Psi}_i) = \frac{\hat{S}^R_i(\boldsymbol{\Psi}_i) - BL^R_i(\boldsymbol{\Psi}_i)}{BL^R_i(\boldsymbol{\Psi}_i)} \tag{18}$$

where $BL^R_i(\boldsymbol{\Psi}_i)$ is the estimated baseline sales in region R of the feature space for the i-th product. Note that $BLRI = 0$ indicates that the estimated number of sold units is similar to baseline, whereas $BLRI > 1$ indicates that estimated sold units are greater than baseline.

(4) Dynamic Range Index (DRI). It is based on dynamic range DR_i, defined as the difference between the maximum and minimum values of a variable, for us, estimated sales for the i-th product. We define it as follows:

$$DRI^R_i(\boldsymbol{\Psi}_i) = \frac{\hat{S}^R_i(\boldsymbol{\Psi}_i) - DR^R_i(\boldsymbol{\Psi}_i)}{DR^R_i(\boldsymbol{\Psi}_i)} \tag{19}$$

This way, *DRI* provides an idea of the accuracy in terms of the forecasting variability. The greater the *DRI*, the lower the variability is.

The four previous indices allow us to check for the reliability and uncertainty of a given model for promotional sales depending on the feature space. Note that the two first (two last) indices are absolute (relative) magnitudes, and that the statistical distribution of $\hat{S}_i$ has to be estimated in the feature space.

4. Database analysis and preprocessing

Our database contains the weekly consolidated information from all electronically recorded data from scanners at cash registers for a Spanish store chain. The information is from 118 stores ($K = 118$), during 105 weeks ($N = 105$) between 2008 and 2009 for 10 products ($I = 10$).

We selected *Laundry detergent* as the best category to be analyzed in this study for several reasons: (1) it is an easily storable product with almost no expiration date; (2) almost all customers buy products of this category, and it is considered a basic household product; and (3) it was one of the largest products in the database in terms of sales. We assembled a database consisting of six brands in this category, including a private label (Table 2), and sold units were almost equal across stores (Table 3). The largest store in terms of sold units accounted only for the 2.72% of the total sales; however, in light out of discussion in Section 2, this scenario is adequate for aggregation purposes.

Table 2
Price level, brand and regular and promotional sold units (%) for each product.

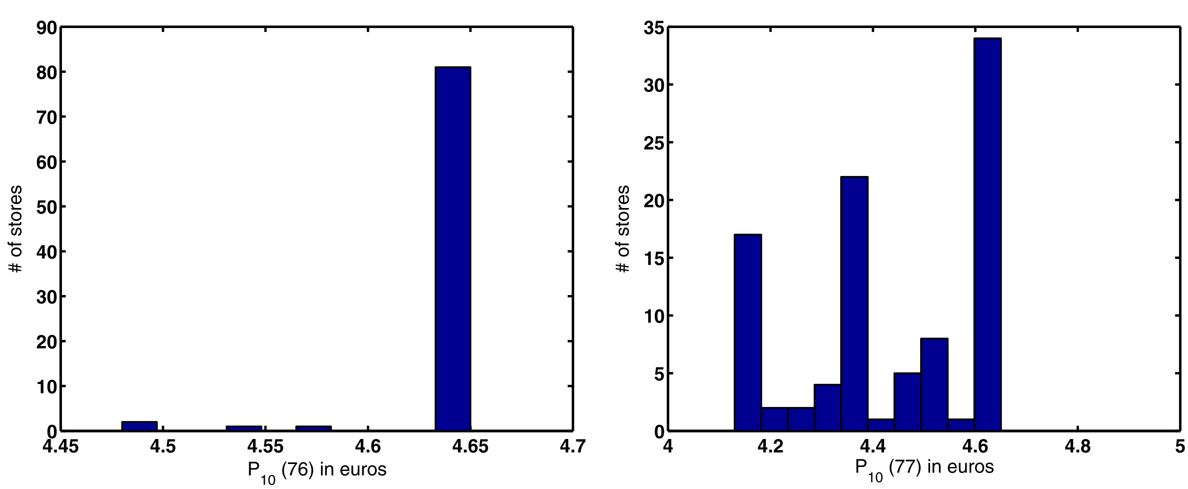
Product	Price level	Brand	Regular sold units (%)	Promotional sold units (%)
Product 1	0.3426	Brand 1	43.82%	51.99%
Product 2	0.2860	Brand 1	43.82%	51.99%
Product 3	0.2489	Brand 2	7.62%	12.20%
Product 4	0.2181	Brand 2	7.62%	12.20%
Product 6	0.1810	Brand 3	5.22%	5.38%
Product 5	0.1522	Brand 3	5.22%	5.38%
Product 7	0.1455	Brand 4	1.86%	1.62%
Product 10	0.1372	Private label	17.19%	8.74%
Product 8	0.1004	Brand 5	2.24%	0.00%
Product 9	0.0836	Brand 5	2.24%	0.00%

Table 3
Top ten retail stores in terms of sold units.

Stores	Regular sold units (%)	Promotional sold units (%)
Store 1	2.72%	2.61%
Store 2	2.43%	2.07%
Store 3	2.16%	1.95%
Store 4	2.02%	1.88%
Store 5	1.80%	1.20%
Store 6	1.71%	1.16%
Store 7	1.65%	1.75%
Store 8	1.60%	2.35%
Store 9	1.59%	1.37%
Store 10	1.56%	1.83%

Promotional activities are carried out by HQ, which means that prices are assumed to be identical for each product in every single store. Consequently, we considered chain-level decisions and global strategies as the main source of promotional activity, rather than store-level marketing strategies. However, database information showed remarkable variability in terms of pricing being applied across stores (Fig. 2), which reveals that real databases always incorporate uncontrolled effects on actual pricing, regardless how strict an HQ's pricing is. As an example, Fig. 2 represents bar graphs for prices versus stores for one product in our database. These graphs demonstrate that the existence of a well-known statistical distribution shape is not a sustainable assumption. The high variability shown in these graphs suggests that information provided by cross-stores should be considered.

Apart from that, a subsequent step in preprocessing aimed to identify whether the analyzed week could be considered as a promotional or a regular pricing-week. This categorization was performed for each product, by setting the week as promotional (or regular) when at least 40% of the stores had promotional prices (or regular) prices. Fig. 3(a) and Fig. 3(b) show that prices are scarce and corrupted by impulsive noise for both promotional and regular prices. To overcome these problems and get the same prices for all stores according to HQ pricing policy, a twofold preprocessing is performed. First, the well-known median filter has been used as the estimation operator, denoted as $\Phi\{P_{i,k}(n)\} = median_k\{P_{i,k}(n)\}$. This filter is robust with respect to the statistical distribution of the uncertainty term in Eq. (1). We empirically chose a size window of 5 elements, and the filter was applied every week for all the available stores (i.e. one dimensional filtering). Second, the mode of every week was computed in order to have only one price per week. Two examples of the preprocessing results are shown in Fig. 3(c) and Fig. 3(d) for regular and promotional prices, respectively.



(a) (b)

Fig. 2. Number of stores with a given price for Product 10 during two consecutive weeks, $n = 76$ (a) and $n = 77$ (b).

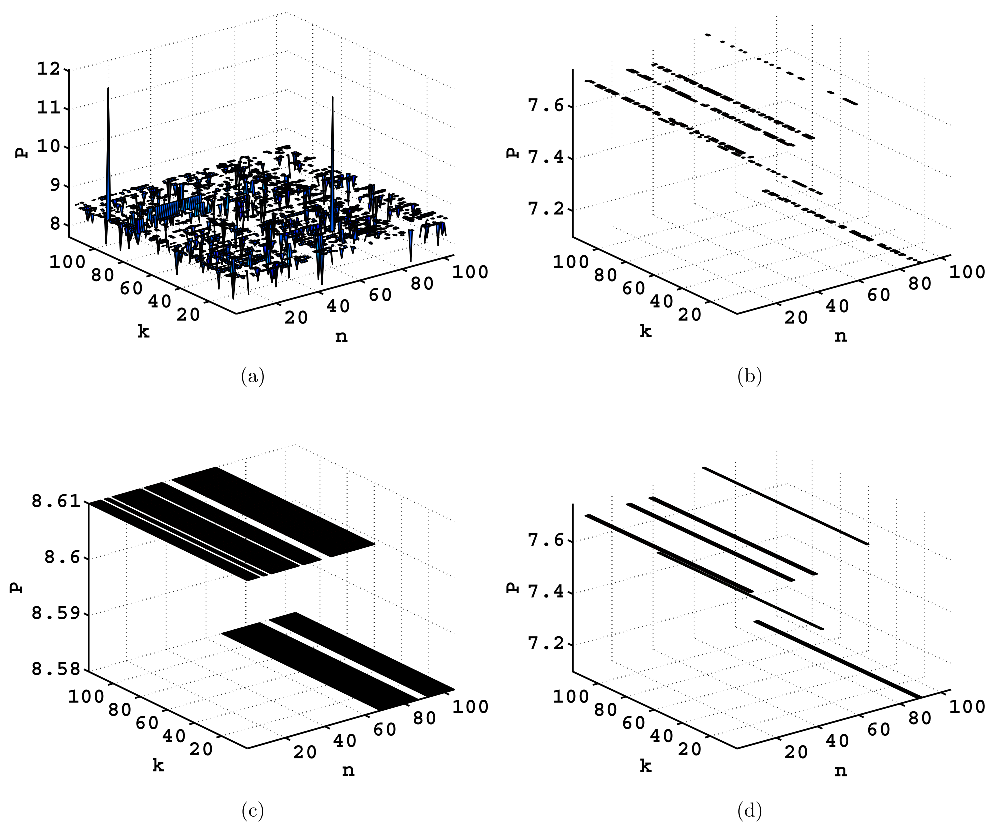


Fig. 3. Prices for Product $i = 1$, where central prices are stated as regular (a), (c) or promotional (b), (d). The prices before (a), (b) and after (c), (d) preprocessing are also shown.

5. Experiments and results on household data

In this section, we describe a set of experiments for analyzing the suitability of the methodology presented in Fig. 1 to the laundry category product database. First, we benchmark and compare the quality of models with increasing algorithmic complexity (DEC, TS, nonlinear models), as well as with different feature spaces in *Experiments A*, *B*, and *C*. Second, we present a quarter analysis in order to study the model stability with respect to changes in the marketing environment in *Experiment D*. Last, *Experiment E* presents reliability indices for the best combination model-feature space for several representative products.

5.1. Experiment A. DEC static model

We estimated the DEC static model by considering only each price index own-effect after preprocessing. The sold units' estimator requires a smoothing operator (Eq. (6)); to do so, we implemented two methods. The average, obtained as a function of the discrete set of prices, and the k-NN estimator, which provides a statistically more effective effect, limiting the impact of outliers. The k-NN method depends on the number of neighbors considered for local-averaged estimation, explored in the range $(1, 30)$ and selected the one minimizing *MAE* (Eq. (12)). In the examples shown in Fig. 4, we obtained the minimum error for $k = 11$ for Product 3, and for $k = 27$ for Product 6.

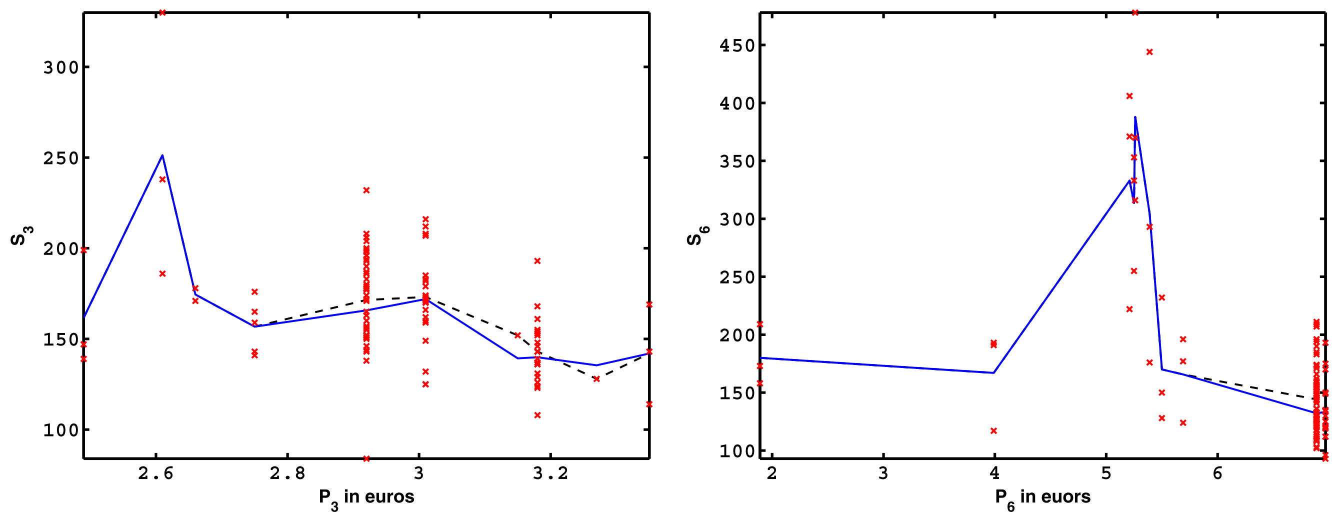
Recall that this DEC estimation does not take into account neither simultaneous promotional effects in related products nor temporal structure. Some examples of DEC estimation are shown in Fig. 4, in which the uneven scatter distribution of price-sold units pairs shows that promotions for each product are scarce compared with regular prices, especially in Fig. 4(b). It is noteworthy that in some situations, units sold at regular prices exceed that for promotional prices. For the rest of experiments section, DEC model based on the k-NN estimator is considered for several reasons: (1) we experimentally checked that it is robust to outliers; and (2) sold units were estimated for the whole range of prices, not only for a discrete set of prices, being able benchmarking results with those provided by other models proposed in this work. Table 4 shows the average obtained from bootstrap resampling for merit figures *MAE* and ΔCI when using DEC k-NN estimator. It can be seen that this method performs similarly in terms of mean and scatter. Results interpretation based on consumer behaviors suggest that, in general, sales estimations for products with a higher number of promotions have worse quality (e.g., Product 1). Note also that good performance is achieved for the private label (Product 10).

DEC approach can be suitable for promotional sales forecasting, though several limitations can be observed. First, the expected effects of demands with respect to prices are not clearly evidenced, even when we use a high rotation category without seasonal effects. Second, according to the observed data, similar prices can yield very different number of sold units, and sometimes lower prices seem to result in lower demand. However, it could be argued that static models for non-perishable products can obviate temporal conditions. For example, consumers could easily defer laundry detergent purchase, or they could accumulate a number of laundry detergents in a single purchase when prices are low.

However, on the whole, these apparently contradictory results raise doubts about the suitability of the DEC promotional model for fitting the products in our database.

5.2. Experiment B. Intrinsic and exogenous TS dynamic models

We considered two different TS promotional models, namely, an AR description (Eq. (7)), and an ARx promotional model (Eq. (8)). In both cases, we used a two hold-out technique (50% for train-



(a) (b)

Fig. 4. Examples of sold units and prices per week, Product 3 (a) and Product 6 (b), including the observed pairs of data (red points, one per week), the price-averaged DEC estimator (dashed), and the k-NN DEC estimator (solid), with $k = 11$ for Product 3 and $k = 27$ for Product 6. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Bootstrap test for k-NN DEC. For each cell, the table shows free parameter k in parentheses (first row), average of *MAE* (second row) and ΔCI (third row) obtained from bootstrap resampling.

Model	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$	$i = 10$
DEC	(1)	(1)	(11)	(7)	(30)	(27)	(24)	(30)	(3)	(13)
	78.38	32.67	20.78	156.80	273.32	34.92	30.87	34.52	29.15	28.27
	187.39	83.47	55.11	675.75	1079.84	103.51	101.20	78.53	90.67	81.01

Table 5
Individual and paired Bootstrap tests for TS (AR, ARx). Individual: For each cell, the table shows p-th and q-th selected orders (first row), the average *MAE* (second row), and 95% *CI* (third row) from Bootstrap resampling. Paired Bootstrap between DEC and TS methods, and between AR and ARx: average of ΔMAE (first row) and ΔCI (second row). For each product, boldface emphasize that in the comparison $Model_1$ vs $Model_2$, best performance is achieved with $Model_1$ (negative values) or with $Model_2$ (positive values).

	Model	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$	$i = 10$
Individual	AR (p)	(10)	(10)	(10)	(9)	(10)	(10)	(10)	(10)	(10)	(9)
		132.26	59.46	30.14	221.99	373.92	72.92	67.85	37.48	20.80	42.16
		156.87	126.48	72.04	1050.68	1376.73	227.44	150.44	85.32	54.01	102.58
	ARx (p, q)	(10, 10)	(10, 10)	(10, 10)	(10, 10)	(10, 10)	(10, 10)	(10, 10)	(10, 10)	(10, 10)	(10, 10)
		64.75	33.20	43.25	233.05	313.11	56.87	93.91	97.88	56.04	24.79
		246.07	74.89	172.68	860.55	876.86	183.60	332.07	598.53	296.87	64.77
Paired	DEC vs AR	**−53.72**	**−26.88**	**−9.28**	**−65.54**	**−98.90**	**−37.81**	**−37.01**	−2.77	8.24	**−13.94**
		30.38	**−43.00**	−17.39	−391.09	−290.65	**−121.53**	**−49.34**	**−6.85**	**36.48**	−21.20
	DEC vs ARx	**13.84**	−0.63	**−22.24**	**−77.27**	**−39.54**	**−22.01**	**−63.65**	**−63.61**	**−26.80**	3.58
		−129.16	33.12	2.98	−60.97	153.75	92.95	2.86	19.01	−42.78	4.97
	AR vs ARx	**67.29**	**26.40**	**−13.05**	−11.66	60.51	**15.83**	−26.57	**−61.73**	**−35.24**	**17.43**
		−89.21	**51.89**	**−100.50**	199.32	**499.59**	43.64	**−176.22**	**−516.22**	**−244.36**	**37.95**

ing) to estimate their out-of-sample performance. We explored orders p and q up to 10 lags, selecting the ones which minimize *MAE* (Eq. (12)). Results are not presented due to space limitations.

Table 5 shows the p-th and q-th selected orders in terms of *MAE* for the AR and ARx models, respectively; and the average obtained from bootstrap resampling for merit figures *MAE* and ΔCI for each product. We obtained non-parsimonious models with high orders for both p and q, which highlights a mismatch between the model proposed by TS and the data dynamics. For some products (3, 4, 7, 8, and 9), the time series of the sales volume seemed self-related and with limited correlation with the exogenous variable (prices), whereas for the other products, the performance improved significantly when the exogenous variable was considered. Nonparametric paired Bootstrap resampling method was applied to test whether the differences in the benchmarking comparison in the table were statistically significant.

Furthermore, Table 5 presents the following comparisons: (1) DEC versus AR; (2) DEC versus ARx; and (3) AR versus ARx. From the first comparison, we can conclude that there were significant performance differences in ΔMAE for most products (except Products 8 and 9), indicating that DEC yielded significantly better quality for the estimations, and the scatter was lower when DEC was considered for Products 2, 6, 7 and 8. The second comparison indicated significant performance differences in ΔMAE for Products 3, 4, 5, 6, 7, 8 and 9, showing that DEC yielded better quality for the estimation, in contrast, for Product 1 ARx yielded significantly better quality. However, the scatter was lower when we used DEC for Product 1, indicating significantly better predictions in terms of scatter. The third comparison indicated significant performance

Table 6
Individual and paired Bootstrap tests for k-NN. Individual: For each cell, the free parameter k (first row), the average *MAE* (second row), and 95% CI (third row) from Bootstrap resampling. Paired Bootstrap between k-NN and DEC, k-NN and TS, and k-NN with different temporal depth: average of ΔMAE (first row) and ΔCI (second row). For each product, boldface emphasize that in the comparison $Model_1$ vs $Model_2$, best performance is achieved with $Model_1$ (negative values) or with $Model_2$ (positive values).

	Model	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$	$i=9$	$i=10$
Individual	k-NN$_{\Xi^1}$ (k)	(2)	(30)	(9)	(1)	(30)	(30)	(22)	(30)	(2)	(20)
		75.25	33.74	21.13	88.89	292.55	35.85	30.10	33.60	28.92	25.25
		170.12	87.95	56.96	350.75	1251.68	135.27	95.47	74.84	87.22	61.13
	k-NN$_{\Psi^1}$ (k)	(2)	(2)	(3)	(8)	(17)	(5)	(1)	(3)	(2)	(20)
		67.42	47.67	21.70	208.75	268.92	67.68	38.00	39.57	33.93	28.67
		141.17	107.98	53.45	580.44	1053.12	152.35	85.50	93.84	73.47	73.87
Paired	DEC vs k-NN$_{\Xi^1}$	**6.77**	−2.35	0.07	4.06	23.76	**−7.08**	**2.20**	0.90	0.00	−0.69
		14.11	−8.12	1.83	−17.45	36.80	−57.74	8.51	−10.43	−4.25	7.73
	AR vs k-NN$_{\Xi^1}$	57.62	26.10	9.12	133.21	83.07	36.44	38.01	**3.63**	**−8.26**	16.56
		−15.14	**45.81**	15.67	**700.05**	123.79	**105.61**	**54.34**	11.10	**−33.67**	**40.72**
	ARx vs k-NN$_{\Psi^1}$	−3.98	**−14.72**	**21.71**	24.16	47.60	−10.78	**56.31**	**59.47**	**22.64**	−3.82
		103.81	−24.77	**118.69**	260.35	−166.01	44.98	**240.22**	**501.94**	**225.12**	−8.58
Paired	k-NN$_{\Xi^1}$ vs k-NN$_{\Psi^1}$	**8.08**	**−14.13**	−0.59	**−120.94**	**24.07**	**−32.01**	**−7.66**	**−5.92**	**−4.89**	**−3.54**
		28.57	−19.67	3.33	−237.22	200.80	−18.69	10.58	**−19.08**	13.96	−13.30
	k-NN$_{\Psi^1}$ vs k-NN$_{\Psi^2}$	3.71	0.77	−0.00	0.64	−11.29	−0.52	−0.38	0.19	−0.31	−0.59
		13.75	−0.82	0.22	−11.95	−40.21	−0.71	−17.87	0.18	0.13	−7.22

differences were found in ΔMAE for the Products 3, 7, 8 and 9, demonstrating that AR methods performed significantly better for the estimation, whereas for Products 1, 2, 6 and 10 ARx yielded a better quality. The scatter was lower when AR was considered for Products 2, 5 and 10, yielding significantly better predictions in terms of scatter, and ARx for Products 1, 3, 7, 8 and 9. In summary, no clear trend in terms of general behavior and prediction of the promotional models could be observed in this set of models.

5.3. Experiment C. Improvements from nonlinear methods

We used the k-NN method for nonlinear promotional modeling. Its design depends on a free parameter, k, which stands for the number of neighbors considered for local-averaged estimation. In this study we explored the range $(1, 30)$ and selected the one which provided the minimum *MAE*.

We considered different feature vectors to characterize temporal evolution in terms of exogenous and/or endogenous variables. The notation for the feature space in this experiment, for a temporal depth n_0, in terms of the feature space, is as follows:

$$\Xi^{n_0} = \left[\hat{P}(n), \cdots, \hat{P}(n-n_0)\right] \tag{20}$$

$$\Theta^{n_0} = \left[S(n-1), \cdots, S(n-n_0)\right] \tag{21}$$

$$\Psi^{n_0} = \left[\Xi^{n_0}, \Theta^{n_0}\right] \tag{22}$$

According to Eq. (20), which addresses different temporal depths for past prices, five models were benchmarked, i.e. Ξ^1, Ξ^2, Ξ^3, Ξ^4, Ξ^5. We do not present all the paired Bootstrap tests here due to space limitation; nevertheless, the results showed that the estimated performance improved when we included in the model information over two consecutive weeks, current and past. Thus, Ξ^1 was a suitable feature space for nonlinear promotional models, when we considered only the exogenous variable. We performed a similar analysis using a set of consecutive temporal depths n_0 for exogenous and endogenous variables simultaneously, and observed better results when considered information over coupled consecutive weeks, this is, for Ψ^1.

Table 6 shows the individual Bootstrap tests in terms of *MAE* and *CI* when considering Ξ^1 and Ψ^1 since both are the feature spaces that provide the minimum error.

Table 6 also shows the results when comparing static, dynamic, linear and nonlinear models with paired Bootstrap tests. First, regarding static and dynamic models (DEC vs k-NN$_{\Xi^1}$), we observed significant performance differences in ΔMAE for Products 1 and 7, for which the dynamic approach yielded better estimation quality, whereas for Product 6, the static model performed better. Secondly, we compared linear and nonlinear models, namely, AR vs Ξ^1, obtaining better quality in terms of scatter for Products 2, 4, 6, 7 and 10 when the nonlinear model is considered. Regarding linear and nonlinear models with past sold units and prices, we observed significantly better estimation quality in ΔMAE for Products 3, 7, 8 and 9, whereas nonlinear models showed better results for Product 2. It is noteworthy that the scatter was lower when nonlinear models were considered for Products 1, 3, 7, 8 and 9.

Thus, we can conclude that, for modeling the sold units of laundry detergent products in this database, the consideration of modeling promotional sales with nonlinear methods yields a significant performance improvement for most of the products. Therefore, we benchmarked new feature vectors with different temporal depths, which indicated that including additional past sold units and prices did not improve sales forecast. Table 6 also shows that, for Products 2, 4, 6, 7, 8 and 9, it was significantly better to exclude the endogenous variable (sales).

5.4. Experiment D. Nonlinear analysis with quarterly temporal data

Results obtained from our chain-level analysis thus far suggest that nonlinear models, with feature spaces given by temporal depth of one lag for sales and prices, are in general appropriate for promotional dynamics representation. An operative tool for promotional decision support should provide with the following capabilities: (1) The retailer must be able to detect and adapt the promotional models to consumer's changing behavior; and (2) the retailer must have available knowledge about the actual reliability and stability of each promotional model. To this end, *Experiment D* analyzes the changes detection in dynamics of promotional models for the products in our database in terms of available historical observations. We used a simple performance analysis of the model in terms of quarterly observed data when progressively included in the model.

First, we compared two promotional models for a given product. In $Model_1$, we forecasted the sold units in a given quarter by only considering historical data of this same quarter, whereas in $Model_2$, we used past and/or future quarter data. A paired Bootstrap test with confidence level of 95% is run where the null hypothesis was that both models exhibited the same performance in terms of *MAE*, and the alternative was that one model performed

Table 7
Quarterly temporal analysis. $\hat{P}(Qa:Qb)$ indicates that the historical data are from weeks included from quarter *Qa* to quarter *Qb*. After using a paired Bootstrap test, each cell represents those products whose quality estimations are statistically better considering (boldface) or not (normal) the information of other quarters during the past two years (first part of the table), or past year (last part of the table) for estimating quarterly sold units (left column). Symbol − indicates that there are not statistically differences for any product.

	$\hat{P}(Q1)$	$\hat{P}(Q1:Q2)$	$\hat{P}(Q1:Q3)$	$\hat{P}(Q1:Q4)$	$\hat{P}(Q1:Q5)$	$\hat{P}(Q1:Q6)$	$\hat{P}(Q1:Q7)$	$\hat{P}(Q1:Q8)$
$\hat{S}(Q1)$	–	–	9	9	1	1	1	1
$\hat{S}(Q2)$	5	5	5, 6, 10	5, 10, **2**	1, 6, 10	1, 6, 10, **7**	1, 6, 10, **7**	1, 6, **7**
$\hat{S}(Q3)$	5	5	–	–	–	**6**	1, **6**	1, **4**, **6**
$\hat{S}(Q4)$	2, 5	2, 5	2, 5	2, 5, **6**	2, 5, **4**	2, 5, 10, **1**, **7**	2, 5, 10, **7**	2, 10, **7**
$\hat{S}(Q5)$	1, 5, **10**	1, 2, 5, 10	1, 2, 5, 7, 9, 10	1, 2, 5, 7, **3**	1, 2, 5, 7, **3**	1, 2, 5, **8**	1, 2, **8**	1, 2, **4**, **8**
$\hat{S}(Q6)$	1, 5	1, 5	1, 5, 10	1, 5, 10	1, 5, 10	1, 5, 10	1, 5, 10	1, 5, 10
$\hat{S}(Q7)$	1, 5	1, 5	1, 5	1, 5, **2**	1, 5, **2**	1, 5	1, 5	1, 5, **2**
$\hat{S}(Q8)$	1, 5, **6**	1, 5, 6	1, 5, 6, **4**, **10**	1, 5, 6	1, 5, 6, **3**	1, 5, 6, 9, **4**	1, 2, 5, 6, 9, **10**	1, 2, 5, 6, 9, **10**

	$\hat{P}(Q4)$	$\hat{P}(Q4:Q5)$	$\hat{P}(Q4:Q6)$	$\hat{P}(Q4:Q7)$	$\hat{P}(Q4:Q8)$
$\hat{S}(Q4)$	–	–	1, 10, **6**	1, 4, 10, **2**, **6**	1, 4, 10, **2**, **6**
$\hat{S}(Q5)$	1	1,	**8**	**8**	**8**
$\hat{S}(Q6)$	1, **3**, **9**	1, **6**, **9**	1, **6**, **9**	–	**6**
$\hat{S}(Q7)$	1, **2**	1, **2**	1	1	1, **2**
$\hat{S}(Q8)$	1	1, **3**	9, **2**, **4**	6, 9, **2**, **10**	6, 9, **2**, **10**

significantly better than the other. Table 7 summarizes the results after using a paired Bootstrap test, each cell represents those products whose quality estimations are statistically better considering (boldface) or not (normal) the information of other quarters during the past two years (first part of the table), or past year (last part of the table) for estimating quarterly sold units (left column). Symbol − indicates that there are not statistically differences for any product. Remark that notation $\hat{P}(Qa:Qb)$ indicates that the historical data are from weeks included from quarter *Qa* to quarter *Qb*, and data include eight quarters spanning two years. In addition, note that the beginning of the economic downturn in Spain could be dated to approximately 2008 (the first year of observations in our database, first quarter referred as *Q*1 in the table); therefore, we provide separate information for the cases in which only the 2009 is considered and those in which both 2008 and 2009 are included.

For data available from two years, Products 1, 2, and 5 exhibited in general better performance when we considered only the current quarterly information. For Product 1, this behavior remained when considering data in the previous year. For Product 2, the inclusion of past and future data resulted in a performance improvement only when we considered the most recent year. For Product 5, there were no performance differences regardless of which cases were used. For Product 7, we obtained better results when we included past and future quarters of data for the two years analysis and worse results for the most recent year term. For Product 8, it was always better to include additional quarters of data regardless of whether we used one or two years.

According to these results, it could be argued that transitory dynamics during the first year may provoke less effective promotional models. This rationale is consistent with our finding that results were less accurate when we included observations from the first year. However, the equivalent situations triggering to different results in separated products would suggest the existence of different sensitivities. To deal with this, the following experiment analyzes the usefulness of the proposed reliability indices in the feature space for each product.

5.5. Experiment E. Reliability indices in feature spaces

As described in Subsection 3.4, the statistical distribution of estimated sold units $\hat{S}$ for the i-th product in the space defined by feature vector Ψ, i.e., $f_{\hat{S}_i(\boldsymbol{\Psi}_i)}\hat{S}_i(\boldsymbol{\Psi}_i)$, can be readily estimated by using Bootstrap resampling, and it is denoted as $f^*_{\hat{S}_i(\boldsymbol{\Psi}_i)}\hat{S}_i(\boldsymbol{\Psi}_i)$. Its statistical average is a hypersurface of the sales as a function of the feature space, and more general, it provides useful information for both reliability and decision-making point of views by allowing us to obtain the indices previously defined in Eqs. (16), (17), (18), and (19).

For this database, we checked that better results were obtained when sold units predictions were made with a nonlinear model considering two consecutive weeks. With this forecasting model, we checked the reliability and stability of results when working with one or two years. This experiment presents the proposed indices in the feature space for two illustrative example products, namely, Product 1 and Product 5. Fig. 5 shows the predicted sales units $\hat{S}_i$ as a function of the feature space for both products, when using two years (a), (c) and one year (b), (d). As previously described, changes in the dynamics and the promotions in the available time periods were determinant for the model forecasting capabilities. *VC* was larger in general for Product 1 with two years data, but also it was larger, in general, in promotional regions of the feature space (e), (f). For Product 5, *VC* was quite constant throughout the feature space, but lower when only considering the last year. ΔCI was strongly dependent of the region in the feature space (i), (k), and higher for promotional regions. *BLRI* for Product 1 indicated a higher efficiency, relative to the baseline, of the promotional activity when data from one year was used (m), (n). *DRI* was larger in promotional regions for Product 5 using the data from two years, whereas it was reduced and became near constant for the data from one year (s), (t).

The significant differences among estimated sold units and reliability indices in one year versus two years, indicating that memory effect may not be justified to be larger than one year, and so, proposed models would not necessary provide stability and reliability capabilities, adding information beyond one year historical data.

It is also remarkable the higher stability and more solid behavior of Product 5 in all analyzed statistics over a wider range of different prices. This situation contrasted against Product 1, indicating that for this kind of product significant effects are led by price change.

6. Conclusions

We propose a promotional chain-level analysis and data modeling based on retail aggregated data to support retail marketing decisions. Fig. 1 depicts the necessary steps for the reliability and stability analysis of promotional models from a chain-level point of view. First, we aggregated retail data at the store-level using simple morphological preprocessing tools, to come to a market-level decision. Second, we benchmarked and optimized linear and nonlin-

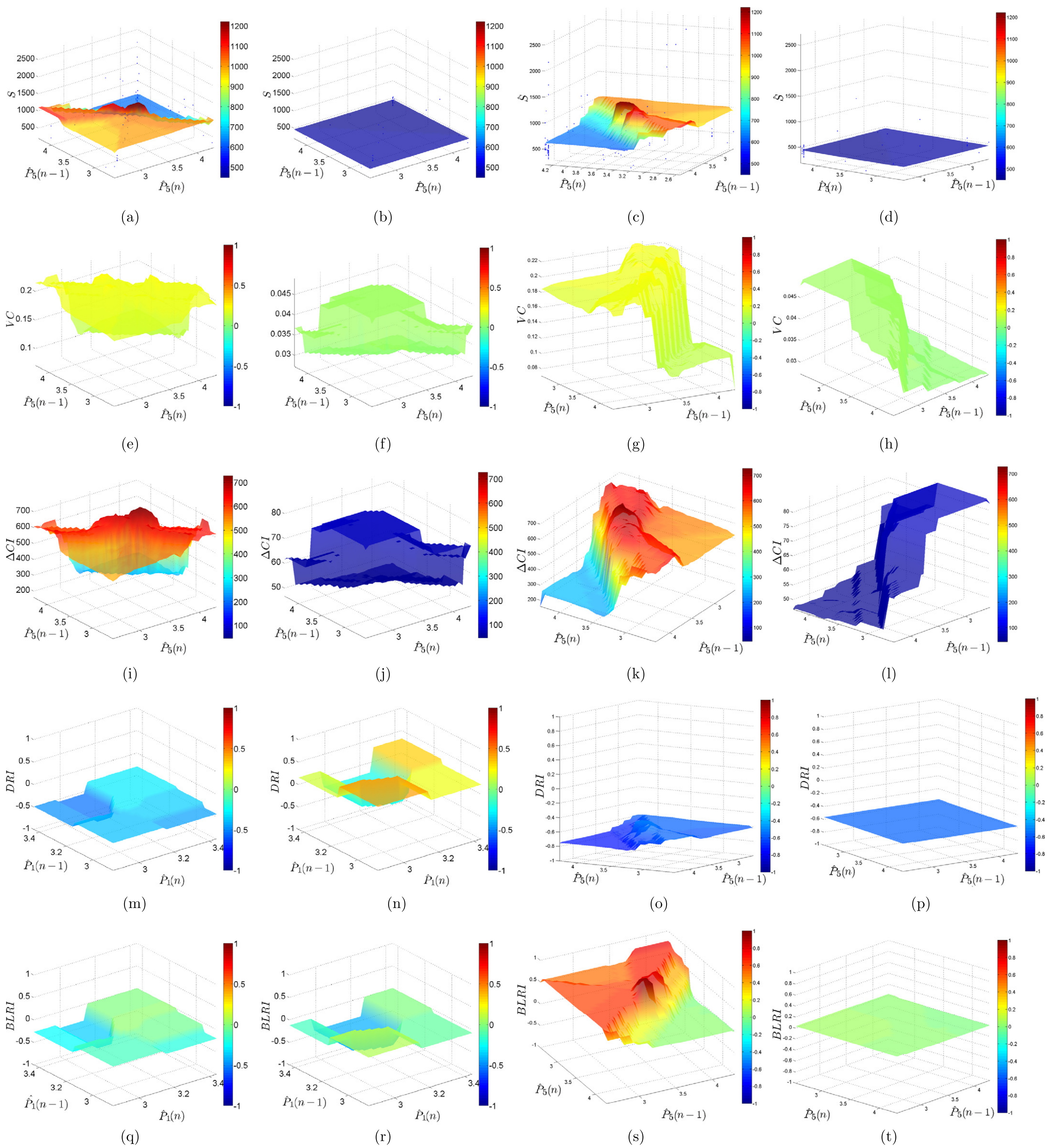


Fig. 5. Estimated sold units and reliability indices. Columns 1 (two years) and 2 (one year) for Product 1; Columns 3 (two years) and 4 (one year) for Product 5. Estimated sold units: (a)–(d); *VC*: (e)–(h); ΔCI: (i)–(l); *BLRI*: (m)–(p); *DRI*: (q)–(t).

ear prediction engines using nonparametric Bootstrap resampling based on performance statistics. Third, we took into account the reliability and stability of the promotional models built with the products in the database using a set of new indicators based on bias and scatter measurements in the feature space.

The economic downturn that began in 2008 is one of the largest in history, at least in some countries; thus, it is more necessary than ever for retailers to effectively evaluate short to medium term promotional effects. It is possible that traditional promotional models do not accurately reflect the actual complexity in the real time because of the increasing amount of concurrent aspects that affect consumer behavior. Therefore, researchers should focus on new models that can capture and statistically represent this new scenario.

Limitations of DEC static analysis. Our experiments showed that the DEC analysis could not provide consistent results in terms of unequivocal demand for a certain pricing level in our data (Fig. 4). Note that laundry detergent category is not subjected to expiration date, thus, it could be argued that households may stockpile the product if prices justify doing so. This assumption is one of several behind the "buy two and get one free" promotional offers, which are common for many long-expiration date products. The present study shows that the static DEC model does not provide a direct statistical match between the endogenous (demand) and exogenous (price) variables.

Limitations of TS linear models. The existence of communication networks and consumer networking may change the market dynamics, resulting in a large and increasing number of concurrent effects. Accordingly, retailers have used dynamic TS models based on historical and current data to guide the promotional strategies. In our data, TS linear models with intrinsic and exogenous variables had an acceptable fit to the data, but only at the expense of non-parsimonious models. Although previous research shows that future demand is forecasted better by considering memory for both pricing and demand data, paradoxically, our data set contained several products for which the incorporation of the exogenous variable into the AR analysis significantly worsened forecasting performance.

Scope of nonlinear models. In general, non-linear models' predictive capability was more effective than that of linear models. Although the results were not uniform for all the products, we obtained consistently parsimonious promotional nonlinear models that yielded acceptable forecasting performance in all the products with up to one lag for both the exogenous and endogenous variables. In some cases, TS linear models performed similarly to nonlinear ones, which is consistent with the fact that linear TS models are specific cases of nonlinear models.

We observed most behavioral singularities in the premium product (higher price, Product 1) and in the most competitive products (private label and low prices, Products 7, 8, 9, and 10).

Model stationarity, weeks and quarters. The strong dependence of promotional models with respect to time implies that their forecasting capabilities are tied to the data stationarity. From a marketing viewpoint, this implication is a natural consequence of demand being consolidated weekly an also the minimum retail promotional periods being typically one week. However, consumer behaviors should exhibit certain monthly trends, especially considering that Spanish payroll cycles are typically monthly and more intensive shopping activity usually occurs at the beginning of the month. In addition, retailers' budgets, control, and results are typically provided on a monthly basis, whereas many other economic indicators, that may be observed to support real and practical marketing strategies, are released quarterly (e.g., Gross Domestic Product or employment rate). In our data, a clear change in the consumer behavior took place in 2008, which affected the stationarity properties of the recorded products.

Reliability indices in feature spaces. Our experiments clearly showed that adequate promotional models can be built within parsimonious feature spaces that accommodate temporal dimensions, considering jointly the endogenous lag effect with the price demand sensitivity. However, promotional regions in the feature space will have a different representation in terms of example feature vectors, because promotional periods are usually shorter than regular prices periods. Therefore, special attention should be paid to the uncertainty of the predicted sales in those promotional regions. It is still possible to calculate reliability indices in the same feature spaces, and often allowing its visualization, thus supporting promotional decision making.

Future work. The present analysis points to several worthwhile avenues for further research. This work can be readily extended to other product categories which results may be different due to the special characteristics of the selected products. In addition, the joint modeling of different complementary or alternative categories could incorporate further information on cross-effects in these related products. Finally, more advanced machine learning techniques could be used and evaluated for improving the predictive capabilities of nonlinear promotional models.

Acknowledgments

This work has been partly supported by Spanish Research Project TEC2010-19263 (Spanish Government) and by the Prometeo Project of the Secretariat for Higher Education, Science, Technology and Innovation of the Republic of Ecuador. CSR is supported by Spanish Grant FPU12/04225.

References

[1] S. Ang, S. Leong, P. Kotler, The Asian apocalypse: crisis marketing for consumers and businesses, Long Range Plan. 33 (1) (2000) 97–119.
[2] M. Martínez-Ruiz, A. Jiménez-Zarco, A. Izquierdo-Yusta, The effects of the current economic situation on customer satisfaction and retail patronage behavior, Total Qual. Manag. Bus. Excell. 23 (11–12) (2012) 1207–1225.
[3] R. Blattberg, R. Briesch, E. Fox, How promotion work, Manag. Sci. 14 (3) (1995) 122–132.
[4] D. Hanssens, L. Parsons, R. Schultz, Market Response Models: Econometric and Time Series Analysis, Kluwer Academic Press, Boston, 2001.
[5] H.J.V. Heerde, P.S.H. Leeflang, D.R. Wittink, Semiparametric analysis to estimate the deal effect curve, J. Mark. Res. 38 (2) (2001) 197–215.
[6] A. Airbarg, N. Arora, Brand portfolio promotions, J. Mark. Res. 45 (4) (2008) 391–402.
[7] C. Soguero-Ruiz, F.J. Gimeno-Blanes, I. Mora-Jiménez, M.P. Martínez-Ruiz, J.L. Rojo-Álvarez, Deal effect curve and promotional models – using machine learning and bootstrap resampling test, in: ICPRAM, vol. 2, 2012, pp. 537–540.
[8] G. Box, G. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, 1970.
[9] S. Kapoor, P. Madhok, S. Wu, Modeling and forecasting sales data by time series analysis, J. Mark. Res. 18 (1) (1981) 94–100.
[10] I. Alon, M. Qi, R. Sadowski, Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods, J. Retail. Consum. Serv. 8 (2001) 147–156.
[11] J. Kuvulmaz, S. Usanmaz, S. Engin, Time-series forecasting by means of linear and nonlinear models, in: Proc. 4th Mexican International Conference on Artificial Intelligence, vol. 3789, 2005, pp. 504–513.
[12] C. Soguero-Ruiz, F.J. Gimeno-Blanes, I. Mora-Jiménez, M.P. Martínez-Ruiz, J.L. Rojo-Álvarez, On the differential benchmarking of promotional efficiency with machine learning modelling (ii): Practical applications, Expert Syst. Appl. 39 (17) (2012) 12784–12798.
[13] C. Soguero-Ruiz, F.J. Gimeno-Blanes, I. Mora-Jiménez, M.P. Martínez-Ruiz, J.L. Rojo-Álvarez, On the differential benchmarking of promotional efficiency with machine learning modeling (i): principles and statistical comparison, Expert Syst. Appl. 39 (17) (2012) 12772–12783.
[14] C. Cuthbertson, J. Reynolds, Retail Strategy, Routledge, Oxford, 2012.
[15] R. Rooderkerk, H. Van Heerde, T.H.A. Bijmolt, Optimizing retail assortments, Mark. Sci. 32 (5) (2013) 699–715.
[16] O. González-Benito, M. Martínez-Ruiz, A. Mollá-Descals, Spatial mapping of price competition using logit-type market share models and store-level scanner-data, J. Oper. Res. Soc. 60 (1) (2009) 52–62.
[17] O. González-Benito, M. Martínez-Ruiz, A. Mollá-Descals, Using store level scanner data to improve category management decisions: developing positioning maps, Eur. J. Oper. Res. 198 (2) (2009) 664–674.
[18] O. González-Benito, M. Martínez-Ruiz, A. Mollá-Descals, Retail pricing decisions and product category competitive structure, Decis. Support Syst. 40 (1) (2010) 110–119.
[19] M. Ataman, H. Van Heerde, C. Meta, The long-term effects of marketing strategy on brand sales, J. Mark. Res. 47 (5) (2010) 866–882.
[20] G. Zotteria, M. Kalchschmidtb, A model for selecting the appropriate level of aggregation in forecasting processes, Int. J. Prod. Econ. 108 (2007) 74–83.
[21] A. Bodapati, S. Gupta, The recoverability of segmentation structure from store-level aggregate data, J. Mark. Res. 41 (3) (2004) 351–364.
[22] E. Foekens, P. Leeflang, D. Wittink, A comparison and an exploration of the forecasting accuracy of a loglinear model at different levels of aggregation, Int. J. Forecast. 10 (2) (1994) 245–261.
[23] M. Christen, S. Gupta, J.C. Porter, R. Staelin, D.R. Wittink, Using market-level data to understand promotion effects in a nonlinear mode, J. Mark. Res. 34 (3) (1997) 322–334.
[24] T. Steven, Avoiding aggregation bias in demand estimation: a multivariate promotional disaggregation approach, Quant. Mark. Econ. 4 (2006) 383–405.

[25] R. Link, Are aggregate scanner data models biased? J. Advert. Res. Sep./Oct. (1995) RC8–RC12, ARF.
[26] H. van Heerde, P.S.H. Leeflang, D.R. Wittink, How promotions work: scan*pro-based evolutionary model building, Schmalenbach Bus. Rev. 54 (2002) 198–220.
[27] D. Bell, J. Chiang, V. Padamanabhan, The decomposition of promotional response: an empirical generalization, Mark. Sci. 18 (4) (1999) 504–546.
[28] B. Liu, F. Kong, X. Yang, Pepp: profits estimation in prices promotion, in: Proc. of 2004 International Conf. on Machine Learning and Cybernetics, vol. 2, 2004, pp. 1146–1151.
[29] M.-P. Martínez-Ruiz, A. Mollá-Descals, M.A. Gómez-Borja, J.L. Rojo-Álvarez, Evaluating temporary retail price discounts using semiparametric regression, J. Retail. Consum. Serv. 3 (13) (2006) 193–204.
[30] M.-P. Martínez-Ruiz, A. Mollá-Descals, M.A. Gómez-Borja, J.L. Rojo-Álvarez, Using daily store-level data to understand price promotion effects in a semiparametric regression model, J. Retail. Consum. Serv. 3 (13) (2006) 193–204.
[31] T. Wang, Y. Li, S. Zhao, Application of SVM based on rough set in real estate prices prediction, in: 4th International Conf. on Wireless Communications, Networking and Mobile Computing, 2008, pp. 1–4.
[32] G. Marnik, M. Dominique, Time-series models in marketing: past, present and future, Int. J. Res. Mark. 17 (2000) 183–193.
[33] C. Horvath, M. Kornelis, P.S.H. Leeflang, What marketing scholars should know about time series analysis: time series applications in marketing, Research Report, University of Groningen, Research Institute SOM (Systems, Organizations and Management).
[34] E. Alpaydin, Introduction to Machine Learning, The MIT Press, Cambridge, MA, 2010.
[35] R. Kuo, K. Xue, Fuzzy neural networks with application to sales forecasting, Fuzzy Sets Syst. 2 (1999) 123–143.
[36] G. Pillo, V. Latorre, S. Lucidi, E. Procacciz, An application of learning machines to sales forecasting under promotions, Working paper 1.
[37] J. Yao, N. Teng, H. Poh Chew, L. Tan, Forecasting and analysis of marketing data using neural networks, J. Inf. Sci. Eng. 14 (1998) 843–862.
[38] S. McIntyre, D. Achabal, C. Miller, Applying case-based reasoning to forecasting retail sales, J. Retail. 69 (4) (1993) 372–398.
[39] P. Brockwell, R. Davis, Introduction to Time Series and Forecasting, Springer, New York, 2002.
[40] K. Jonasson, S.E. Ferrando, Evaluating exact VARMA likelihood and its gradient when data are incomplete, ACM Trans. Math. Soft. (TOMS) 35 (1) (2008).
[41] L. Devroye, L. Gyorfi, G. Lugosi, Probabilistic Theory of Pattern Recognition, Stochastic Modeling and Applied Probability, Springer, New York, 1996.
[42] T.M. Mitchell, Machine Learning, McGraw-Hill Higher Education, Boston, 1997.
[43] R. Duda, P. Hart, D. Stork, Pattern Classification, Wiley-Interscience, New York, 2001.
[44] B. Efron, R. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall, Boca Raton, FL, 1997.
[45] P. Kopalle, C. Mela, L. Marsh, The dynamic effect of discounting on sales: empirical analysis and normative pricing implications, Mark. Sci. 18 (3) (1999) 317–332.
[46] T. Ando, Bayesian state space modeling approach for measuring the effectiveness of marketing activities and baseline sales from POS data, in: Proc. of IEEE Int. Conf. on Data Mining, 2006, pp. 21–32.

Cristina Soguero-Ruiz received the Telecommunication Engineering degree, the B.Sc. degree in Business Administration and Management, and the M.Sc. degree in Biomedical Engineering from the University Rey Juan Carlos, Madrid, Spain, in 2011 and 2012, respectively. She is currently working towards the Ph.D. degree in the Joint Doctoral Program in Multimedia and Communications in conjunction with University Rey Juan Carlos and University Carlos III in Spain. The Program is awarded with the Spanish Quality Mention. She is supported by FPUs Spanish Research and Teaching Fellowship (3rd place in TEC area). Her main research interests include statistical learning theory, digital signal processing, with application to e-Health and marketing.

Fco. Javier Gimeno-Blanes, Telecommunications engineer, Advanced Studies in Taxation and Business Administration and Ph.D. in Communications Technologies. He is currently Professor at the University Miguel Hernández, ViceDean of the Polytechnic School of Elche, and Adjunct Vice-President for Technology Development. He has worked in telecom industry for 10 years holding management positions as Director of Strategic Planning, responsible for Strategic Planning at Telefónica DataCorp, member of the Chairman's Office of Telefónica Holding and Corporate Development at Grupo Fuertes. Primary research lines are in the Bio-signals processing, business planning, business valuation, financial risk management, pricing and marketing forecasting.

Inmaculada Mora-Jiménez received the Telecommunication engineering degree from Universidad Politécnica de Valencia, Spain, in 1998, and the Ph.D. degree from Universidad Carlos III de Madrid, Spain, in 2004. Currently, she is an Associate Professor in the Department of Signal Theory and Communications at Universidad Rey Juan Carlos in Madrid, Spain. Her main research interests include statistical learning theory, neural networks, and their applications to image processing, bioengineering, and communications.

María Pilar Martínez-Ruiz (Ph.D., University of Castilla–La Mancha, Spain) is Associate Professor of Marketing at Marketing-Department, University of Castilla–La Mancha, Spain. Her research interests fall in the areas of consumer behavior in the food service industry, pricing and marketing communications. Professor Martínez-Ruiz has published in Expert Systems with Applications, European Journal of Operational Research, International Journal of Market Research, Decision support Systems and Electronic commerce, European Journal of Marketing, Journal of Retailing and Consumer Services, Total Quality Management and Business Excellence, and other journals.

José Luis Rojo-Álvarez received the Telecommunications Engineer degree (1996, University of Vigo), and the Ph.D. in Telecommunications (2000, Polytechnic University of Madrid). Since 2006, he has been an Associate Professor in Rey Juan Carlos University, Madrid, Spain. He has published more than 70 papers in JCR journals and more than 150 conference communications. He has participated in more than 50 projects (with public and private funding). He received in 2009 the I3 Program to the Excellent Research Trajectory from Spanish Ministry of Science. His research interests include statistical learning and digital signal processing with applications to biomedical engineering and marketing.