Multiperiod Multiproduct Advertising Budgeting: Stochastic Optimization Modeling

C. Beltran-Royo[*] L. F. Escudero[†] H. Zhang[‡]

09/02/2015

Abstract

We propose a stochastic optimization model for the Multiperiod Multiproduct Advertising Budgeting problem, so that the expected profit of the advertising investment is maximized. The proposed model is a convex optimization problem that can readily be solved by plain use of standard optimization software. It has been tested in a case study derived from a real advertising campaign. In the case study, the expected profit of the stochastic approach has been favorably compared with the expected profit of the deterministic approach. This provides a quantitative argument in favor of the stochastic approach for managerial decision making in a data-driven framework.

Key words: Marketing, advertising budgeting, advertising adstock, stochastic optimization, convex optimization.

1 Introduction

In this paper we address the Multiperiod Multiproduct Advertising Budgeting (MAB) problem (notice that we drop one M for short). More specifically, we wish to simultaneously optimize the advertising campaigns of different products [9], considering a multiperiod planning horizon and having the *cumulative advertising effect* [27] as the inter-period linking variables.

This work can be considered as the second part of [4] where: First, the relevance of the MAB problem and its state-of-the-art were discussed. Second, a deterministic MAB optimization model was introduced. Third, its theoretical properties were studied (convexity, separability and optimality conditions). The purpose of this second part is to introduce a stochastic MAB optimization model and compare it with its deterministic counterpart introduced in [4]. In the remainder of the paper, we will replace the term 'optimization model' by 'model' for short. In deterministic models, each unknown parameter is substituted by an estimate, that is, all the problem parameters are considered to be known. However, in real situations many of the problem parameters are unknown at the moment of deciding the advertising budget. In this paper we will enhance our previous MAB model by explicitly dealing with this uncertainty: Each unknown parameter will be incorporated into the MAB model as a random variable instead of as a single estimate. Thus, the new stochastic MAB model proposes an optimal

[*]Corresponding author: cesar.beltran@urjc.es, Statistics and Operations Research, Rey Juan Carlos University, Madrid, Spain.

[†]Statistics and Operations Research, Rey Juan Carlos University, Madrid, Spain.

[‡]School of Management, University of Shanghai for Science and Technology, Shanghai, China.

advertising budget adapted not only to the estimated parameter value, but to a set of representative values of the uncertain parameters, taking into account the likelihood of each parameter value.

Stochastic models of the advertising budgeting problem under uncertainty have been proposed in the literature from different perspectives. In [30] a stochastic game theory approach is used to analyze the optimal advertising spending in a duopolistic market where the share of each firm depends on its own and its competitors advertising decisions. In [3] a Bayesian dynamic linear model is used for studying the wear out effects of different themes of advertising (for example, price advertisements versus product advertisements) in order to improve the effectiveness of advertising budget allocation across different themes. In [15] a Markov decision process approach is used for modeling the multistage advertising budgeting problem. Other approaches that have been proposed to solve the advertising budgeting problem under uncertainty are based on multicriteria fuzzy optimization [28], chance constraint goal optimization [5] and robust optimization [2]. Other works that consider the stochastic side of the advertising budgeting problem in a multiperiod environment are [3, 30], among others. A recent survey about dynamic advertising (deterministic and stochastic approaches) can be found in [23].

As we have mentioned, different stochastic aspects of the advertising budgeting problem have been considered in literature. But, as far as we know, a stochastic version of the MAB problem has not yet been addressed. Thus, for example, [13] considers a multiproduct model but for a single period and with a deterministic approach, [35] considers a multiperiod model but for a single product and also with a deterministic approach, [15] considers a stochastic multiperiod model but for a single product. However, as we will illustrate in our case study, a relevant expected profit improvement can be achieved if the parameter uncertainty in MAB models is taken into account (stochastic models) compared to ignoring it (deterministic models). In this paper we propose and analyze a stochastic optimization version of the MAB problem, that improves two previous advertising budgeting models presented in [4] and [13], respectively. The proposed model has appealing properties: it is a convex optimization model and it is a stochastic optimization model (it takes into account the likelihood of the representative values of the uncertain parameters). There are other choices to deal with the uncertainty in the MAB problem. Thus for example, the chance constraint approach allows for incorporating probabilistic constraints which are interesting from a modelling point of view, but, in general, the resulting deterministic equivalent optimization problem is non-convex. Another possible approach is robust optimization which deals with uncertainty by considering uncertainty sets for the model parameters, but this approach, in contrast with the approach we use, does not take into account the likelihood of each parameter value given by its probability distribution. See [24].

As it has been stated above, the aim of the stochastic MAB model presented here consists of obtaining the optimal advertising budget and its related allocation, by considering that the probability distribution of the uncertain parameters is known (or at least we can approximate it). Our model aims to addressing the following questions: a) What is the optimal multiproduct advertising budget for the whole planning horizon? b) Given an advertising budget, how can we optimally allocate it along the planning horizon? c) Is it advisable to consider stochastic MAB models? or on the contrary, is it enough to consider deterministic ones?

The MAB model introduced in [4] is simple but realistic enough to be used in the advertising industry. From a mathematical point of view, it corresponds to a convex optimization problem which is numerically tractable and allows for computing the global optimal solutions with moderate computational effort. We will analyze under which conditions these desirable properties are inherited by the new stochastic MAB model.

The remainder of the paper is organized as follows. Section 2 introduces by example the single-stage stochastic optimization problem with deterministic feasible set. In Section 3 a stochastic version of the MAB problem is introduced and analyzed. A case study is introduced in Section 4 to illustrate the effectiveness of the stochastic model as well as the theoretical concepts of Section 3. Section 5 concludes and outlines future research. Finally, in Appendix A we list and prove the theoretical results of the paper and in Appendix B we give some basic features of regression models and Bayesian inference.

2 Single-stage stochastic optimization with deterministic feasible set

Stochastic optimization literature considers an optimization horizon with one, two or more decision stages (single-stage, two-stage and multistage models, respectively). Single-stage models [24, 31] consider only one decision stage, that is, the decision is to be made 'here and now' and the model does not account for any corrective (recourse) actions as in the two-stage and multistage cases [6, 34]. As we mentioned in Section 1, a stochastic version of the MAB problem has not yet been addressed. As a first step in this direction, in this paper we propose a single-stage stochastic MAB model. We show that, in the context of advertising budgeting, for a slightly higher conceptual and computational effort, the single-stage stochastic model may significantly improve the deterministic one in two aspects: expected profit and accuracy (see Section 4). However, one could also be interested in going further to study two-stage or multistage models with recourse, since it would be natural to re-solve the MAB problem over the rolling horizon and update the decisions. As a sequel of this paper, given the important nature of the issue, we are planning to study the advantage of using a two-stage or multistage stochastic approach, as pointed out in Section 5.

In this section we introduce the single-stage stochastic optimization problem with deterministic feasible set, that is, with deterministic parameters in the constraints. Let us consider the decision vector $x \in \mathcal{D} \subset \mathbb{R}^n$, where $\mathcal{D}$ is called the feasible set, and the random vector ξ whose probability distribution is supported on a set $\Xi \subset \mathbb{R}^m$. Given the cost function $F(x, \xi)$ we define the stochastic problem P_S as

$$\min_{x \in \mathcal{D}} F_S(x),$$

where $F_S(x) := \mathbb{E}[F(x, \xi)]$. Notice that the uncertain parameters are only in the cost function and not in $\mathcal{D}$. In some situations problem P_S is approximated by a deterministic model named the expected value problem P_{EV} that can be expressed as

$$\min_{x \in \mathcal{D}} F_{EV}(x),$$

where $F_{EV}(x) := F(x, \mathbb{E}[\xi])$. In this context let $x^*_{(\cdot)}$ and $F^*_{(\cdot)}$ denote the optimal solution and cost of problem $\mathrm{P}_{(\cdot)}$, respectively. The deterministic and stochastic MAB models, to be presented in Section 3, have the structure of problems P_{EV} and P_S, respectively, and for this reason we review some basic but useful properties to formulate and solve those two problems (see Appendix A, Proposition 1).

In the context of two-stage stochastic linear optimization let EV, EEV and RP denote the optimal cost of the Expected Value problem, the Expected result of using the EV solution and the optimal cost of the Recourse Problem, respectively; see e.g. [6, 24]. Given that the problem of concern in this section does not consider a recourse (corrective action), let us use the term SP (optimal cost of the Stochastic Problem) instead of RP. These concepts can be adapted in our case as follows,

$$\mathrm{EV} = F^*_{EV}, \qquad \mathrm{EEV} = F_S(x^*_{EV}), \qquad \mathrm{SP} = F^*_S.$$

With this notation, Proposition 1 (see Appendix A) recovers the well known result in stochastic linear minimization EV $\leq$ SP $\leq$ EEV. Another useful concept is VSS (Value of the Stochastic Solution), such that VSS = EEV $-$ SP. It measures the cost of ignoring uncertainty in choosing a decision; see again e.g. [6, 24]. Let us illustrate these concepts by means of an example.

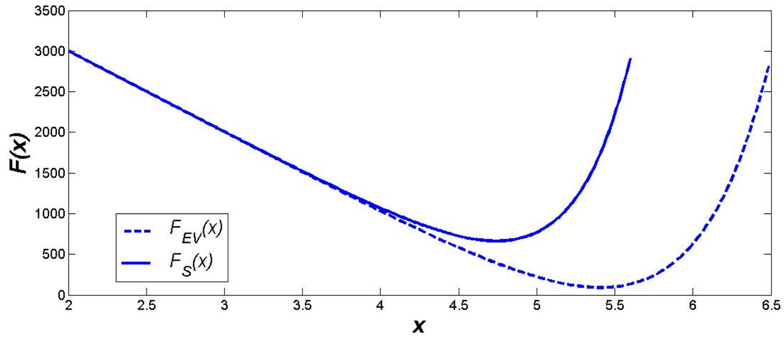

Figure 1: In Example 1, $F_{EV}(x)$ is a bad approximation to $F_S(x)$ for optimization purposes.

Example 1 *The objective of this example is to show that the expected value function F_{EV} may be a bad approximation to the stochastic function F_S for optimization purposes. Consider the cost function*

$$F(x, \xi) = k_1 + k_2\, e^{\xi_1 x} - \xi_2\, x,$$

where $k_1 \in \mathbb{R}$ and $k_2 > 0$ are constants, and ξ_1 and ξ_2 are normal random variables $N(\bar{\xi}_1, \sigma_1^2)$ and $N(\bar{\xi}_2, \sigma_2^2)$, respectively. It is easy to prove that $F(x, \xi)$ is convex in ξ, then by Proposition 1 (see Appendix A) we get

$$k_1 + k_2\, e^{\bar{\xi}_1 x} - \bar{\xi}_2\, x \quad \leq \quad \mathbb{E}\left[k_1 + k_2\, e^{\xi_1 x} - \xi_2\, x\right] \quad x \in \mathbb{R}.$$

Consider problems P_S and P_{EV} associated with $F(x, \xi)$ and $\mathcal{D} = \mathbb{R}$. In this case we have:

$$\begin{aligned} F_S(x) &= \mathbb{E}\left[k_1 + k_2\, e^{\xi_1 x} - \xi_2\, x\right] \\ &= k_1 + k_2\, \mathbb{E}\left[e^{\xi_1 x}\right] - \bar{\xi}_2\, x \\ &= k_1 + k_2\, e^{\bar{\xi}_1 x + 0.5\sigma_1^2 x^2} - \bar{\xi}_2\, x, \end{aligned}$$

where we have used that $\mathbb{E}\left[e^{\xi_1 x}\right]$ corresponds to the moment-generating function of the normal random variable ξ_1 and therefore it is equal to $e^{\bar{\xi}_1 x + 0.5\sigma_1^2 x^2}$ (see [11] for details).

However, in the expected value problem, the function $F_S(x)$ is approximated by

$$F_{EV}(x) = k_1 + k_2\, e^{\bar{\xi}_1 x} - \bar{\xi}_2\, x.$$

Thus, the approximation

$$\mathbb{E}\left[e^{\xi_1 x}\right] = e^{\bar{\xi}_1 x + 0.5\sigma_1^2 x^2} \approx e^{\bar{\xi}_1 x},$$

may produce poor optimization results as it is depicted in Figure 1 for the instance with the following parameters: $k_1 = 5000, k_2 = 0.01, \bar{\xi}_1 = 2, \sigma_1^2 = 0.1, \bar{\xi}_2 = 1000, \sigma_2^2 = 100$. After solving P_{EV} and P_S it results

$$\begin{aligned} (x_{EV}^*, F_{EV}^*) &= (5.4,\ 90), \\ (x_S^*, F_S^*) &= (4.7,\ 663), \\ F_S(x_{EV}^*) &= 1750. \end{aligned}$$

Since EV = 90, SP=663 and EEV = 1750, the chain of inequalities EV $\leq$ SP $\leq$ EEV holds. Furthermore, although the deterministic and the stochastic optimal solutions are similar (5.4 versus 4.7), the corresponding expected costs are very different (1750 versus 663), giving a VSS equal to 1087. Of course, this is only a toy example, but it illustrates the well known fact that the expected value function F_{EV} may be a bad approximation to the stochastic function F_S for optimization purposes. ∎

3 Formulation of the MAB problem

The objective of the Multiperiod Multiproduct Advertising Budgeting (MAB) problem is to maximize the profit of the *sales due to advertising.* Notice that we distinguish between baseline sales (sales that one would expect without advertising) and sales due to advertising. Let us briefly review some key concepts in the advertising industry (see, for example, [20] for more details).

- *Advertising media:* The technology through which the advertising takes place. Although it is still dominated by traditional technologies, namely television, radio, print publications, etc., internet based advertising is rapidly gaining market share [1, 10, 12].
- *Insertion:* A single placement of an ad in an advertising media.
- *Reach:* The proportion of the target audience exposed to at least one insertion of the advertisement [9]. This proportion is called the reached audience.
- *Frequency:* The average number of times a person from the reached audience is exposed to an advertisement.
- *Exposure:* Exposure to an advertisement involves reach and frequency and can be measured in Gross Rating Points: GRPs $\equiv$ reach $\times$ frequency. For example, a purchase of 100 GRPs could mean that 100% of the target audience is exposed once to an advertisement or that 50% of the target audience is exposed twice [20]. Normally, advertising is measured in GRPs and not in euros [3]. Managers evaluate the effectiveness of their campaigns in terms of demand generated per GRP and most media buying is done in terms of GRPs. The reason is that GRPs provide a more accurate picture of advertising input than advertising expenditures since it is not clear how much advertising exposure can be purchased for a given budget.
- *Marketing mix:* Marketing mix variables ('marketing mix' for short), correspond to price, sales promotions, advertising copy, advertising channel, timing, and other brand-specific marketing factors.
- *Market segmentation:* The distinct consumer groups, each one characterized by the same needs and behaviors [8].

Market response models provide a basis for fine tuning the marketing mix, which in order to be effective has to take into account the market segmentation. The largest category of market response models are those dealing with sales and market share as dependent variables. Companies want to know what influences their sales (the *sales drivers* or, for short, *drivers*). Their objective is to set the marketing mix in order to optimize their sales. One of the limitations of the MAB model that we present is that it does not take the product price as a sales driver, that is, as a decision variable (since prices are input data). All the other above mentioned drivers (sales promotions, advertising copy, advertising channel and timing) can be taken into account in our MAB model.

We introduced in [4] a deterministic model for the MAB problem. A more realistic approach to this problem consists of allowing randomness of the main parameters of the model, namely advertising saturation levels, advertising diminishing returns, cross product effects and white noise of the profit function. With this objective in mind, in this section the stochastic optimization version of the MAB problem is introduced, directly derived from the deterministic model in [4].

3.1 Notation

	Indexes:	
t	Index for periods, $t \in \mathcal{T} = \{1, \ldots, T\}$	
j	Index for products, $j \in \mathcal{J} = \{1, \ldots, J\}$	
i	(Auxiliary) index for products, $i \in \mathcal{I} = \mathcal{J}$	
k	Index for advertising channel, $k \in \mathcal{K} = \{1, \ldots, K\}$	
r	Index for components of ξ, $r \in \mathcal{R} = \{1, \ldots, R\}$	
s	Index for scenarios, $s \in \mathcal{S} = \{1, \ldots, S\}$	
$\mathcal{TJK}$	Stands for $\mathcal{T} \times \mathcal{J} \times \mathcal{K}$ (analogously for other combinations of index sets: $\mathcal{JK}$, $\mathcal{TJ}$, etc.)	
	Deterministic parameters:	
c_{tjk}	Cost of advertising channel jk in period $t \in \{1, \ldots, T+1\}$, $c_{tjk} > 0$ (note that, for simplicity, to refer to the investment in the k-th advertising channel for product j, we use the expression 'advertising channel jk')	$jk \in \mathcal{JK}$
δ_{jk}	Retention rate of the advertising effect from period to period for channel jk, $\delta_{jk} \in \,]0, 1[$	$jk \in \mathcal{JK}$
p_{tj}	Profit per unit of product j in period t, $p_{tj} > 0$	$tj \in \mathcal{TJ}$
$\widetilde{y}_{0jk}$	Initial accumulated advertising effect of channel jk ('adstock'), $\widetilde{y}_{0jk} \geq 0$	$jk \in \mathcal{JK}$
	Stochastic parameters:	
α_{tjk}	Advertising saturation level in period t of channel jk, $\alpha_{tjk} > 0$	$tjk \in \mathcal{TJK}$
β_{tjk}	Advertising diminishing return to scale in period t of channel jk, $\beta_{tjk} > 0$	$tjk \in \mathcal{TJK}$
γ_{tijk}	Sales of product i in period t induced by one unit invested in advertising channel jk ('cross product effect')	$tijk \in \mathcal{TIJK}$
ε	Stochastic error of the profit function such that $\mathbb{E}[\varepsilon] = 0$	
ξ	Random vector that accounts for all the stochastic parameters of problem MAB ($\alpha_{tjk}, \beta_{tjk}, \gamma_{tijk}$ and ε)	
ξ_r	Component of ξ such that $\xi = (\xi_1, \ldots, \xi_R)^\intercal$	$r \in \mathcal{R}$
ξ^s	Scenario or realization of ξ such that its probability is w^s (weight)	$s \in \mathcal{S}$
$\widehat{\xi}$	Random vector defined by $\{(\xi^s, w^s)\}_{s \in \mathcal{S}}$	
$\bar{\xi}$	Expectation of $\widehat{\xi}$	
	Functions:	
S_{tjk}	Sales of product j in period t due to advertising channel jk	$tjk \in \mathcal{TJK}$
C_{tijk}	Cross product effect: sales of product $i \neq j$ in period t due to advertising channel jk	$tijk \in \mathcal{TIJK}$
P	Profit function	
	Sets:	
$\mathcal{D}$	Feasible set for problems MAB$_S$ and MAB$_{EV}$	

	Decision variables:	
g_{tjk}	Investment (GRPs) in advertising channel jk in period t (GRP stands for 'Gross Rating Points')	$tjk \in \mathcal{TJK}$
y_{tjk}	Accumulated advertising effect of channel jk in period t ('adstock')	$t \in \{0, \ldots, T\}, jk \in \mathcal{JK}$
z_{1jk}	Value of the initial adstock level for advertising channel jk	$jk \in \mathcal{JK}$
z_{Tjk}	Value of the final adstock level for advertising channel jk	$jk \in \mathcal{JK}$
x	Vector that accounts for all the decision variables of problem MAB (g_{tjk}, y_{tjk} and z_{tjk})	

3.2 Adstock function

Advertising on different media (television, newspapers, internet, etc.) tries to increase consumption in two ways. On the one hand, it tries to influence the consumer immediate brand choice and, on the other hand, it tries to increase brand awareness, in order to make easier the brand choice for future advertising. Therefore, the advertising effect on consumer purchase behavior spreads over time and the advertising investment in one period is accumulated with past advertising effects. In this respect, it is common to use the function so-called *adstock*, which is a mathematical model that combines the past and current advertising effects. According to [32], the adstock function models the impact that advertising has over consumer awareness and in turn on sales volume. It means stocking the advertising effect by integrating prior advertising expenditures into a stock function, say $y_t(g)$, and considering the carry-over effect over time. As reported in [20], a typical choice is the (linear) basic adstock model of Broadbent [7]:

$$y_0(g) = \widetilde{y}_0 \tag{1}$$
$$y_t(g) = \delta\, y_{t-1}(g) + g_t \qquad t \in \mathcal{T}, \tag{2}$$

where we have that $\widetilde{y}_0$ is the initial value of the adstock, y_t is the adstock at time t, δ is the retention rate of the advertising effect and g_t is the investment in advertising at time t. Observe that to define y_t we could write $y_t(g_1, \ldots, g_t)$ instead, to indicate that the current advertising effect depends on past and current advertising investments. To simplify the notation, we just write $y_t(g)$. Notice that (2) is the discrete time version of the Nerlove-Arrow continuous time model for adstock [27]. Later we will append a pair of indexes jk, to indicate the advertising channel in equations (1)-(2), such that we will have $y_{0jk}(g), \widetilde{y}_{0jk}$, etc. for all $jk \in \mathcal{JK}$.

In order to model the adstock, other functions can be used such as, for example, the logistic (S-curve) model. The advantage of the logistic model and other nonlinear adstock models, over the basic adstock model, is that they can capture diminishing returns and saturation levels [32]. The advantage of the basic adstock model corresponds to its effectiveness: it is simple and captures the accumulative and decay effects of advertising over time. Nevertheless, in our MAB model we capture diminishing returns and saturation levels using the sales response function S_{tjk} combined with the basic adstock model (see equation (6) and comments therein).

3.3 Profit function

In this section a stochastic MAB profit function is introduced directly derived from the deterministic one in [4], where further modeling details can be found (Section 2.2: 'The multiproduct sales response function'). We define the stochastic MAB profit function $P(g, \xi)$ as follows:

$$P(g, \xi) \;=\; \sum_{tjk \in \mathcal{TJK}} P_{tjk}(g, \alpha, \beta, \gamma) + \varepsilon, \tag{3}$$

where

$$g = (g_{tjk})_{tjk \in \mathcal{TJK}}, \tag{4}$$

$$\xi = \big((\alpha_{tjk})_{tjk \in \mathcal{TJK}},\ (\beta_{tjk})_{tjk \in \mathcal{TJK}},\ (\gamma_{tijk})_{tijk \in \mathcal{TIJK}},\ \varepsilon\big), \tag{5}$$

$$P_{tjk}(g,\alpha,\beta,\gamma) = p_{tj}\ S_{tjk}(g,\alpha,\beta) + \sum_{i \in \mathcal{I}} p_{ti}\ C_{tijk}(g,\gamma) -\ c_{tjk}\ g_{tjk} + z_{tjk}(g) \quad tjk \in \mathcal{TJK},$$

$$S_{tjk}(g,\alpha,\beta) = \alpha_{tjk}(1 - e^{-\beta_{tjk}\ y_{tjk}(g)}) \quad tjk \in \mathcal{TJK}, \tag{6}$$

$$C_{tijk}(g,\gamma) = \gamma_{tijk}\ y_{tjk}(g) \quad tijk \in \mathcal{TIJK}, \tag{7}$$

$$y_{tjk}(g) = \delta_{jk}\ y_{t-1,jk}(g) + g_{tjk} \quad tjk \in \mathcal{TJK}, \tag{8}$$

$$y_{0jk} = \widetilde{y}_{0jk} \quad jk \in \mathcal{JK}, \tag{9}$$

$$z_{1jk}(g) = -c_{1jk}\ \delta_{jk}\widetilde{y}_{0jk} \quad jk \in \mathcal{JK}, \tag{10}$$

$$z_{tjk}(g) = 0, \quad 1 < t < T, jk \in \mathcal{JK}, \tag{11}$$

$$z_{Tjk}(g) = c_{T+1,jk}\ \delta_{jk}\ y_{Tjk}(g) \quad jk \in \mathcal{JK}. \tag{12}$$

The following comments are in order:

- In equation (3) the error ε of the profit function is required both by (possibly) omitted variables in the model and by truly random disturbances, as pointed out in [20].

- In equation (5), coefficients γ_{tijk} with $i = j$ are assumed both to exist and to be 0 for notational simplicity.

- Equation (6) models the sales of product j induced by advertising channel jk in period t. In general the single product *sales response function* S_{tjk} corresponds to increasing concave functions which model diminishing returns and advertising saturation levels. A typical choice, among others, is the model in equation (6), which is called the 'modified exponential' function (see [4] and [20] for details). Observe that in $S_{tjk}(g,\alpha,\beta)$ we could write α_{tjk} and β_{tjk} instead, but we drop the subindexes to simplify the notation. Analogously in equation (7).

- Equation (7) models the cross product effects: sales on product $i \neq j$ due to advertising in channel jk in period t. Notice that cross product effects are modeled as linear functions although they may be slightly nonlinear (decreasing and convex) such that the resulting profit function is not concave. As is well known, concavity of the objective function is a desirable property in a maximization problem since it guarantees global optimality (assuming a convex feasible domain). On the other hand, the cross product effects are often small relative to the direct advertising effects, as in the case study presented in Section 4. Thus, in our view, to approximate the cross advertising effects by linear functions assuming that the cross effects are small, represents a balance between model accuracy and model tractability (see [4] for details).

- Equation (8) corresponds to the adstock function.

- Equation (9) sets the initial adstock level for advertising channel jk.

- Equation (10) accounts for the value of the initial adstock level for advertising channel jk.

- In equation (11), for notational convenience, we use these dummy terms and set them equal to 0.

- Equation (12) accounts for the value of the final adstock level for advertising channel jk.

From a statistical point of view the profit function in (3) corresponds to a *nonlinear regression* model where the profit, say $\widetilde{P}$, is the dependent variable and the advertising investment vector g is the set of independent variables. These variables are related by means of the regression function such that $\widetilde{P} = P(g, \xi)$. From equations (3)-(12) it is clear that it is a nonlinear regression model. As usual, the regression parameters are unknown and one way to model them is by means of a probabilistic model in the Bayesian inference approach (see Appendix B). In our case, we consider the random vector ξ which accounts for all the regression parameters. As a consequence, profit $\widetilde{P}$ is a random variable since its value depends on the realization of ξ. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables, i.e. $\mathbb{E}[\widetilde{P} \mid g]$. This conditional expectation is computed by using the regression model as follows $\mathbb{E}[\widetilde{P} \mid g] = \mathbb{E}[P(g, \xi)]$. Thus, in order to compute the advertising investment that gives the best expected profit we will solve the following optimization problem

$$\max_{g \in \mathcal{G}} \mathbb{E}[\widetilde{P} \mid g] = \max_{g \in \mathcal{G}} \mathbb{E}[P(g, \xi)], \tag{13}$$

where $\mathcal{G}$ is the set of feasible advertising investments. Notice that this problem falls into the class of single-stage stochastic optimization problems with deterministic feasible set (assuming that there is no uncertainty in the definition of $\mathcal{G}$). In contrast, in [4] we approximated this problem by solving the (deterministic) expected value counterpart

$$\max_{g \in \mathcal{G}} P(g, \mathbb{E}[\xi]).$$

The previous two problems have the same structure as problems P_S and P_{EV}, respectively, introduced in Section 2, where it was shown that problem P_{EV} may be a very poor approximation to problem P_S.

3.4 Stochastic optimization based on scenarios

If a stochastic optimization problem, as for example (13), becomes too difficult it is common to approximate it in terms of scenarios [6], a methodology closely related to the Sample Average Approximation method [25, 34]. Roughly speaking, in the scenario based approach the uncertainty of the problem is approximated by a set of scenarios. More precisely, let ξ denote the random vector of the stochastic parameters of the MAB model. Let ξ^s denote the scenario or realization of the random vector ξ for all $s \in \mathcal{S}$, where $\mathcal{S}$ is the index set of the scenarios that are considered (the cardinality of $\mathcal{S}$ is assumed to be finite). Let w^s denote the probability (weight) of scenario $s \in \mathcal{S}$. In this way the set $\{(\xi^s, w^s)\}_{s \in \mathcal{S}}$ defines a random vector $\widehat{\xi}$ with finite support. That is, in the scenario based approach, the random vector ξ is approximated by the random vector $\widehat{\xi}$. We denote by $\bar{\xi}$ the expectation of $\widehat{\xi}$ (i.e. $\bar{\xi} = \mathbb{E}[\widehat{\xi}]$). Obviously, $\bar{\xi}$ only is an approximation to $\mathbb{E}[\xi]$. It will be useful to compute the expectation of each component $\widehat{\xi}_r$ of $\widehat{\xi}$, for all $r \in \mathcal{R}$, which can be done by means of the formula

$$\mathbb{E}[\widehat{\xi}_r] = \sum_{s \in \mathcal{S}} w^s \xi_r^s \qquad r \in \mathcal{R}. \tag{14}$$

Furthermore, any β_{tjk} corresponds to a ξ_r for a unique $r \in \mathcal{R}$, such that notation and meaning of $\xi_r^s, \widehat{\xi}_r, \bar{\xi}_r$ will be translated as $\beta_{tjk}^s, \widehat{\beta}_{tjk}, \bar{\beta}_{tjk}$. Analogously for α_{tjk} and γ_{tijk}.

3.5 Objective function

In this section a stochastic MAB objective function, based on scenarios, is introduced directly derived from the profit function (3). We define the stochastic MAB objective function as follows:

$$\begin{aligned} F(x,\xi^s) = -\sum_{tjk\in\mathcal{TJK}} \Big\{ & p_{tj}\; S_{tjk}(y,\alpha^s,\beta^s) + \sum_{i\in\mathcal{I}} p_{ti}\; C_{tijk}(y,\gamma^s) \\ & -\; c_{tjk}\; g_{tjk} + z_{tjk} \Big\} - \varepsilon^s \qquad\qquad s\in\mathcal{S}, \end{aligned}$$

where

$$\begin{aligned} x &= \big((g_{tjk})_{tjk\in\mathcal{TJK}},\quad (y_{tjk})_{tjk\in\mathcal{TJK}},\quad (z_{tjk})_{tjk\in\mathcal{TJK}}\big), \\ \xi^s &= \big((\alpha^s_{tjk})_{tjk\in\mathcal{TJK}},\quad (\beta^s_{tjk})_{tjk\in\mathcal{TJK}}, \\ &\qquad (\gamma^s_{tijk})_{tijk\in\mathcal{TIJK}},\quad \varepsilon^s\big) && s\in\mathcal{S}, \\ S_{tjk}(y,\alpha^s,\beta^s) &= \alpha^s_{tjk}(1-e^{-\beta^s_{tjk}\; y_{tjk}}) && s\in\mathcal{S},\ tjk\in\mathcal{TJK}, \\ C_{tijk}(y,\gamma^s) &= \gamma^s_{tijk}\; y_{tjk} && s\in\mathcal{S},\ tijk\in\mathcal{TIJK}, \\ z_{1jk} &= -c_{1jk}\; \delta_{jk}\widetilde{y}_{0jk} && jk\in\mathcal{JK}, \\ z_{tjk} &= 0, && 1<t<T, jk\in\mathcal{JK}, \\ z_{Tjk} &= c_{T+1,jk}\; \delta_{jk}\; y_{Tjk} && jk\in\mathcal{JK}. \end{aligned}$$

Notice that this objective function has been written as the opposite of the profit function (3) in order to be used to define a minimization problem. Furthermore, it is defined in the x-space which accounts for vector (g,y,z). In contrast, profit function (3) was defined in the g-space. Both function could be equivalently used to define the MAB optimization problem. However, the version in the g-space is more appropriate for regression analysis, where g correspond to the independent variables and the profit is the dependent variable (see Section 3.3). On the other hand, the version in the x-space seems more appropriate to define the optimization problem: the objective function is less involved since adstock functions $y_{tjk}(g)$ in (6) and (7) can be included as variables y_{tjk}, linked with vector g by means of linear constraints (17). In this way it easier to analyze the structure of the resulting optimization problem (15)-(22), in particular the objective function (23).

3.6 Stochastic optimization model

Let the Stochastic Multiperiod Multiproduct Advertising Budgeting problem MAB$_S$ be defined as a *single-stage* stochastic optimization problem with deterministic feasible set. Notice that in a different way as traditionally it is done in stochastic optimization for multiperiod problems [6, 24], all (multi-period) decisions are taken at the beginning of the planning horizon as it is a practice in the advertising sector. That is, it is a stochastic optimization problem without recourse.

Taking into account the cost function $F(x,\xi^s)$, defined in the previous section, and the dynamics of the adstock function (1)-(2), problem MAB$_S$ can be modeled as a single-stage stochastic optimization with deterministic feasible set (nonlinear objective function and linear constraints).

$$\min_{x} \quad F_S(x) = \mathbb{E}[F(x,\widehat{\xi})] \; = \sum_{s\in\mathcal{S}} w^s F(x,\xi^s), \tag{15}$$

$$\begin{aligned}
\text{s.t.} \quad & y_{0jk} = \widetilde{y}_{0jk} && jk \in \mathcal{JK}, && (16)\\
& y_{tjk} - \delta_{jk}\, y_{t-1,jk} - g_{tjk} = 0 && tjk \in \mathcal{TJK}, && (17)\\
& z_{1jk} = -c_{1jk}\,\delta_{jk}\widetilde{y}_{0jk} && jk \in \mathcal{JK}, && (18)\\
& z_{tjk} = 0 && 1 < t < T, jk \in \mathcal{JK}, && (19)\\
& z_{Tjk} - c_{T+1,jk}\,\delta_{jk}\, y_{Tjk} = 0 && jk \in \mathcal{JK}, && (20)\\
& Ax \le b, && && (21)\\
& \underline{x} \le x \le \overline{x}. && && (22)
\end{aligned}$$

where $x = (g, y, z)$ and equation (21) accounts for possible linear constraints for x, which has lower and upper bounds (equation (22)). In Proposition 2 of Appendix A we show that under mild assumptions this objective function can be stated as follows

$$\begin{aligned}
F_S(x) \;\; = \;\; - \sum_{tjk\in\mathcal{TJK}} & \Big\{ p_{tj}\,\bar{\alpha}_{tjk} \left(1 - \mathbb{E}[e^{-\widehat{\beta}_{tjk}\, y_{tjk}}]\right) + \sum_{i\in\mathcal{I}} p_{ti}\,\bar{\gamma}_{tijk}\, y_{tjk} \\
& - c_{tjk}\, g_{tjk} + z_{tjk} \Big\}.
\end{aligned} \tag{23}$$

where $\mathbb{E}[e^{-\widehat{\beta}_{tjk}\, y_{tjk}}] = \sum_{s\in\mathcal{S}} w^s e^{-\beta^s_{tjk}\, y_{tjk}}$.

If we define the feasible set $\mathcal{D}$ by equations (16)-(22) then problem MAB_S can be written as

$$\min_{x\in\mathcal{D}} F_S(x),$$

i.e., problem MAB_S has the same structure as problem P_S considered in Section 2. Also observe that, although problem MAB_S considers T periods (i.e., it is a *multiperiod* problem) the value of the decision variables does not depend on each scenario ξ^s. Therefore, it is a *single-stage* stochastic optimization problem with deterministic feasible set. In spite of the problem being a single-stage one, it takes into account all the scenarios ξ^s, in contrast with the deterministic approach, that simply takes into account the expected value of the parameters. Finally, Problem MAB_S is a convex optimization problem, which is a convenient feature for optimization and computational aspects (see Appendix A, Proposition 3).

3.7 Deterministic optimization model

The stochastic parameters of problem MAB_S are replaced with their expected value to obtain the so-called Expected Value problem, that by definition is a deterministic one. So, the Expected Value Multiperiod Multiproduct Advertising Budgeting problem MAB_{EV} can be expressed as

$$\min_{x\in\mathcal{D}} \quad F_{EV}(x) = F(x,\bar{\xi}),$$

where $\bar{\xi} = \mathbb{E}[\widehat{\xi}] = \sum_{s\in\mathcal{S}} w^s \xi^s$.

This objective function can be stated as follows

$$\begin{aligned}
F_{EV}(x) \;\; = \;\; - \sum_{tjk\in\mathcal{TJK}} & \Big\{ p_{tj}\,\bar{\alpha}_{tjk} \left(1 - e^{-\bar{\beta}_{tjk}\, y_{tjk}}\right) + \sum_{i\in\mathcal{I}} p_{ti}\,\bar{\gamma}_{tijk}\, y_{tjk} \\
& - c_{tjk}\, g_{tjk} + z_{tjk} \Big\}.
\end{aligned} \tag{24}$$

Notice that the term $\mathbb{E}[e^{-\widehat{\beta}_{tjk}\ y_{tjk}}]$ of F_S (equation (23)) is approximated in F_{EV} by the term $e^{-\bar{\beta}_{tjk}\ y_{tjk}}$ (equation (24)).

Model MAB$_{EV}$ corresponds to the one introduced in [4] and has the same structure as problem P$_{EV}$ analyzed in Section 2. Notice that, in this deterministic version the random vector $\widehat{\xi}$ of the stochastic version, is replaced by its expectation $\bar{\xi} = \mathbb{E}[\widehat{\xi}]$, so we minimize $F(x, \bar{\xi})$. In contrast, in the stochastic MAB model we minimize $\mathbb{E}[F(x, \widehat{\xi})]$ which is, in general, different from $F(x, \bar{\xi})$. Furthermore, problem MAB$_{EV}$ is also a convex optimization problem (see Appendix A, Proposition 4). Finally it is important to mention that under mild assumptions we have that $F^*_{EV} \leq F^*_S \leq F_S(x^*_{EV})$ (see Appendix A, Proposition 5).

4 Case study

4.1 Case description

An experimental case is analyzed to show the improvement that the stochastic model can bring to the deterministic one of the multiperiod multiproduct advertising budgeting (MAB) problem under parameter uncertainty. Notice that the level of improvement depends on the instance considered. In any case, regarding the expected profit, the stochastic approach will always be at least as good as the deterministic one (see Appendix A, Proposition 5). The objective of the current case study is to show by example that in the MAB problem, for a slightly higher conceptual and computational effort, the stochastic approach may significantly improve the deterministic approach in two aspects: expected profit and accuracy. As it was pointed out in the introduction, considering the inherent uncertainty of some MAB parameters, the aim of this work is to answer the following questions: a) What is the optimal multiproduct advertising budget for the whole planning horizon? b) Given an advertising budget, how can we optimally allocate it along the planning horizon? c) Is it important to consider stochastic models? or on the contrary, is it enough to consider deterministic ones?

The computations have been conducted on a laptop under Windows XP, with a processor Intel Core Duo 2.40 GHz and with 3.48 GB of RAM. The implementation has been written in Matlab (R2008b) and the constrained convex minimization problems MAB$_S$ and MAB$_{EV}$, have been solved by function `fmincon` from the Matlab Optimization Toolbox (V4.1) with default parameters.

Table 1: Problem dimensions and profit per unit of product.

Parameter	Value	Units
T	12	month
I	2	product
K	2	advertising channel
p_{t1}	1.75	euro / unit of product P1
p_{t2}	1.40	euro / unit of product P2

Instances MAB$_S$ and MAB$_{EV}$ that are considered in this computational experiment were derived from a real case addressed at the consulting company Bayes Forecast[1] to plan the advertising campaign for a leading fast moving consumer goods company. The instance we present here, considers a twelve months planning horizon (T = 12), two products that for confidentiality we denote by P1 and P2 (I = 2) and two advertising channels (K = 2). The first channel corresponds to TV advertising and the

[1] Bayes Forecast S.L., Madrid (Spain), www.bayesforecast.com

Table 2: Advertising retention rate δ_{jk}, initial adstock $\widetilde{y}_{0jk}$ and advertising diminishing return $\bar{\beta}_{tjk}$.

j	δ_{j1}	δ_{j2}	$\widetilde{y}_{0j1}$	$\widetilde{y}_{0j2}$	$\bar{\beta}_{tj1}$	$\bar{\beta}_{tj2}$
1	0.660	0.552	300	300	0.010	0.010
2	0.588	0.552	50	50	0.005	0.005

Table 3: Advertising saturation levels.

t	$\bar{\alpha}_{t11}$	$\bar{\alpha}_{t12}$	$\bar{\alpha}_{t21}$	$\bar{\alpha}_{t22}$
1	345,000	270,000	86,400	105,600
2	389,850	305,100	83,700	102,300
3	493,350	386,100	113,400	138,600
4	510,600	399,600	137,700	168,300
5	731,400	572,400	199,800	244,200
6	838,350	656,100	251,100	306,900
7	897,000	702,000	278,100	339,900
8	969,450	758,700	259,200	316,800
9	734,850	575,100	224,100	273,900
10	386,400	302,400	191,700	234,300
11	427,800	334,800	189,000	231,000
12	407,100	318,600	218,700	267,300

Table 4: Cost of the advertising channels.

t	c_{t11}	c_{t12}	c_{t21}	c_{t22}
1	480	528	432	475
2	480	528	432	475
3	640	704	576	634
4	640	704	576	634
5	480	528	432	475
6	480	528	432	475
7	640	704	576	634
8	640	704	576	634
9	480	528	432	475
10	480	528	432	475
11	640	704	576	634
12	640	704	576	634

Table 5: Cross product effect levels.

i	$\bar{\gamma}_{ti11}$	$\bar{\gamma}_{ti21}$	$\bar{\gamma}_{ti12}$	$\bar{\gamma}_{ti22}$
1	0	-0.00010	0	-0.00015
2	-0.00020	0	-0.00015	0

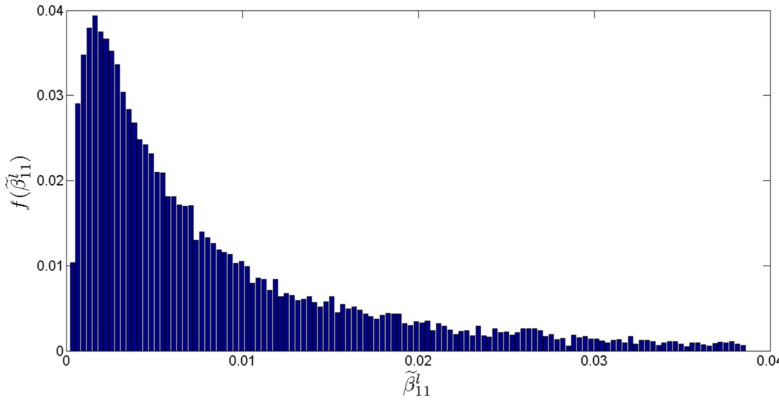

Figure 2: Probability mass function of $\widetilde{\beta}_{11}$.

second channel corresponds to in-store promotions. The original real case considered a higher number of products and advertising channels, however we considered a reduced number of them for a clearer exposition. Furthermore, the data from the real case was modified for confidentiality and research reasons. Thus, for example, the costs in Table 4 were modified from the original data in order to define artificial alternating low and high price periods. The objective was to observe the adaptation of the two optimization approaches to dynamic GRP prices. Table 1 shows the problem dimensions as well as the unit profit per product, being the same for all periods in the planning horizon. Table 2 gives the advertising retention rate δ_{jk} (no units), the initial adstock $\widetilde{y}_{0jk}$ (GRPs) and the advertising diminishing return $\bar{\beta}_{tjk}$ (GRPs^{-1}) related to function (25), being the same for all periods in the planning horizon. Notice that only the expected value $\bar{\beta}_{tjk}$ is given, since the corresponding empirical distributions consider 300 realizations for each β_{tjk} and therefore are too large to be included in the paper. Nevertheless, they can be obtained under request or downloaded from Section 'Publications' at `http://bayes.etsii.urjc.es/∼cbeltran/CV/`. Table 3 shows the advertising saturation level $\bar{\alpha}_{tjk}$ (units of product j) related to function (25). Table 4 gives the cost (euros / GRP) of advertising channels in the periods along the planning horizon. Finally, Table 5 shows the cross product effect level $\bar{\gamma}_{tijk}$ (units of product i / GRP) related to functions (23) and (24).

In the MAB model, sales due to advertising is modeled as a function of the accumulated advertising effect (adstock) y as follows,

$$S(y, \alpha, \beta) = \alpha \left(1 - e^{-\beta\, y}\right). \tag{25}$$

This single product sales response function is known as the 'modified exponential' function [20].

Note: Function S_{tjk} was introduced in Section 3, and here the indexes tjk are dropped for simplicity of exposition. The positive parameter α corresponds to the advertising saturation level. This means that no matter how much marketing effort is expended, the sales due to advertising will not be higher than α. The positive parameter β regulates the advertising diminishing return to scale. On the other hand, γ, the cross product sales effect between products P1 and P2, is due to substitution in this case study, i.e., advertising on, say P1, will increase P1 sales but will reduce P2 sales and viceversa. This effect is known as *cannibalization* [17]. Under cannibalization, the cross product effect parameter γ is negative (see Table 5).

4.2 Stochastic modelling issues

As stated in Section 3.3, the sales functions S_{tjk} and C_{tijk} are based on the random vector ξ, which accounts for the stochastic parameters $\alpha_{tjk}, \beta_{tjk}, \gamma_{tijk}$ and ε, and has probability density function ρ_1.

In order to set up problems MAB$_{EV}$ and MAB$_S$, the first step is to generate a set of scenarios $\{\xi^s\}_{s\in\mathcal{S}}$ (sample of the random vector ξ). In this case study we took $S = 5000$. By using Bayesian inference one can obtain a sample of the so-called *posterior probability distribution* of ξ, which corresponds to a discrete estimation of ρ_1. Although a technical description of the Bayesian approach is out of the scope of this work, in Appendix B we have reported some of its basic features in the context of regression models. In particular, we have stressed the (dis)advantages of the Bayesian inference compared to the classical (frequentist) one.

For the deterministic problem MAB$_{EV}$ we estimate the expected value of $\alpha_{tjk}, \beta_{tjk}$ and γ_{tijk} by $\bar{\alpha}_{tjk}, \bar{\beta}_{tjk}$ and $\bar{\gamma}_{tijk}$, respectively. In this case study, these estimations are computed from the 5000 scenarios such that $\bar{\alpha}_{tjk} = \sum_{s\in\mathcal{S}} w^s \alpha^s_{tjk}$ (the other estimations are computed analogously). On the other hand, the stochastic problem MAB$_S$, is based on the expected values of $\alpha_{tjk}, \gamma_{tijk}$ and $e^{-\beta_{tjk}y_{tjk}}$. The first two are estimated as before. The third one is estimated as follows

$$\mathbb{E}\left[e^{-\beta_{tjk}\ y_{tjk}}\right] \approx \sum_{s\in\mathcal{S}} w^s e^{-\beta^s_{tjk}\ y_{tjk}}.$$

If the number of scenarios and the number of parameters β_{tjk} is large, then the CPU time to solve problem MAB$_S$ can be considerable. One way to reduce this time is by means of the alternative approximation

$$\mathbb{E}\left[e^{-\beta_{tjk}\ y_{tjk}}\right] \approx \mathbb{E}\left[e^{-\widetilde{\beta}_{tjk}\ y_{tjk}}\right] = \sum_{l\in\mathcal{L}} \pi^l_{tjk}\ e^{-\widetilde{\beta}^l_{tjk}\ y_{tjk}}.$$

where $\widetilde{\beta}_{tjk}$ is a finite support random variable that approximates the continuous random variable β_{tjk}. The support of $\widetilde{\beta}_{tjk}$ is $\{\widetilde{\beta}^l_{tjk}\}_{l\in\mathcal{L}}$ and the corresponding probability values are $\{\widetilde{\pi}^l_{tjk}\}_{l\in\mathcal{L}}$. Roughly speaking, the way to construct the probability mass function of $\widetilde{\beta}_{tjk}$ corresponds to represent the histogram of the observations $\{\beta^s_{tjk}\}_{s\in\mathcal{S}}$ (for details see Algorithm 1 in Appendix A). As pointed out in Algorithm 1, the set of probability values $\{\pi^l_{tjk}\}_{l\in\mathcal{L}}$ approximates the marginal density function of β_{tjk} for all $tjk \in \mathcal{TJK}$. As an example, Figure 2 depicts the probability mass function of $\widetilde{\beta}_{11}$. Note: Since in this case study β_{tjk} does not depend on the specific time period t, index t can be dropped to just write $\widetilde{\beta}_{jk}$. The elements that define $\widetilde{\beta}_{jk}$ are: the index set $\mathcal{L} = \{1, \ldots, L\}$, its support $\{\widetilde{\beta}^l_{jk}\}_{l\in\mathcal{L}}$ and the corresponding probability values $\pi^l_{jk} = P\big(\widetilde{\beta}_{jk} = \widetilde{\beta}^l_{jk}\big)$ for all $l \in \mathcal{L}$. In this case study we have considered $L = 300$ realizations for each $\widetilde{\beta}_{jk}$.

Notice that the two problems, MAB$_{EV}$ and MAB$_S$, are built from almost the same data (Table 1 to Table 5). The difference is that problem MAB$_{EV}$ is based on the expected values $\bar{\beta}_{tjk}$ in Table 2, whereas problem MAB$_S$ considers the inherent uncertainty by means of the random variables $\widetilde{\beta}_{tjk}$.

4.3 Determining the optimal budget and its allocation

We interpret 'budget' as the total sum of money set aside or needed for a purpose, such that once a budget is decided it should not be exceed. A company with several departments can be interested in determining the optimal budget for each department. At a first stage, the company can compute the optimal spending of a given department by solving an optimization problem with no budget constraint. At a second stage, this optimal spending can be imposed as the department budget.

Thus, in order to determine the optimal budget and its allocation in our case study, next we solve the MAB$_{EV}$ and MAB$_S$ instances defined by the data listed in Section 4.1, with the bound constraint $g \geq 0$. Table 6 shows the main results of these two instances, under labels EV and SP, respectively, defined in Section 2. In particular it shows the optimal budget (optimal spending) and the optimal profit as well as the related CPU time. As an example, the optimal budget allocation corresponding

Table 6: Optimal budget and optimal profit.

	EV	EEV	SP	Variation
Budget (euros)	3,818,334	-	4,350,039	+13.93%
Deterministic profit ($-F_{EV}$)	23,276,709	-	-	-
Expected profit ($-F_S$)	-	16,870,731	17,608,877	+4.38%
CPU time (s)	2	-	54	-

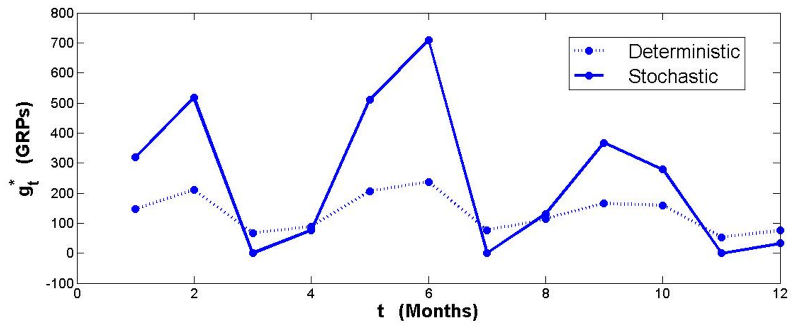

Figure 3: Optimal budget allocation g_t^* for advertising channel 1 of product P1: Deterministic approach versus stochastic approach.

to advertising channel 1 of product P1, is depicted in Figure 3, where it can be observed that the stochastic approach is more sensitive to price changes than the deterministic approach. The former allocates more GRPs in low price periods (months 1-2, 5-6 and 9-10) and less GRPs in high price periods (months 3-4, 7-8 and 11-12).

In this context, the Expected profit of using the Expected Value solution x^*_{EV}, for short EEV, is computed as $-F_S(x^*_{EV})$ where F_S is the stochastic objective function (23). Remember that the objective function F was defined as the opposite of the profit function P in order to define a minimization problem. Table 6 shows that the advertising budget proposed by the stochastic model is 13.93% (531,705 euros) higher than the advertising budget proposed by the deterministic model. It also shows that the stochastic approach produces an expected profit 4.38% (738,146 euros) higher than the expected profit of the deterministic approach given by the EEV. Of course, the stochastic improvement in expected profit observed in this case study does not guarantee this level of improvement for all the MAB instances. Notice that in this case study the deterministic approach is erroneous, thus misleading for managerial purposes, since the expected profit estimation has an error of almost 38% (the expected profit estimation is EV = 23,276,709 euros and the corresponding true expected profit is EEV = 16,870,731 euros).

It is worthy to mention that the inequality chain EEV $\leq$ SP $\leq$ EV (maximization version) is fulfilled

$$16,870,731 \leq 17,608,877 \leq 23,276,709,$$

in accordance with Proposition 5 (see Appendix A). Notice that in the proposition we have the opposite orientation of the inequalities since there we are in a minimizing cost context and here we are in a maximizing profit context. In this instance, the value of the stochastic solution (VSS) is equal to 531,705 euros.

So far we have compared the stochastic versus the deterministic approaches by comparing the corresponding expected profits $\mathbb{E}[P(x^*_S, \widehat{\xi})]$ and $\mathbb{E}[P(x^*_{EV}, \widehat{\xi})]$, as numerical values. It can also be useful

Table 7: Statistical parameters for the sample of profits in euros (deterministic and stochastic approaches).

	Deterministic	Stochastic
Min	11,814,360	12,483,785
Max	21,071,046	21,242,252
Mean	16,855,996	17,619,252
Median	16,895,003	17,678,295
Standard deviation	1,305,596	1,093,411

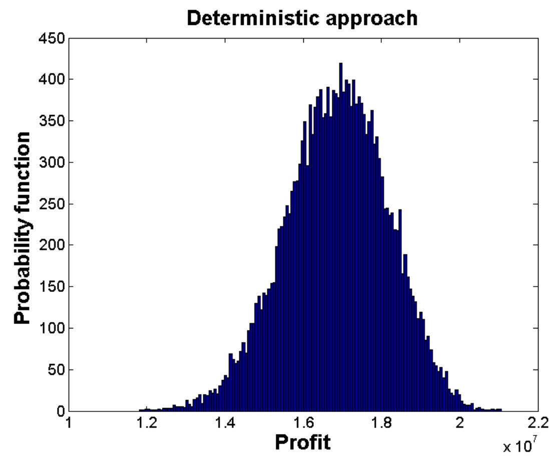

Figure 4: Sample probability distribution of the profit (euros) that may be obtained with the deterministic approach (sample expected profit 16,855,996 euros).

to compare the corresponding profits $P(x_S^*, \widehat{\xi})$ and $P(x_{EV}^*, \widehat{\xi})$, as random variables, whose empirical probability distribution can be seen in Figures 4 and 5, respectively. These distributions have been obtained by computing the profits associated to a sample of size 20000 of the random vector ξ. That is, for each sample vector ξ^k, we have computed the corresponding sample profits $P(x_S^*, \xi^k)$ and $P(x_{EV}^*, \xi^k)$ whose histogram can be seen in those figures, respectively. The sample expected profits for the deterministic and stochastic approaches thus computed are 16,855,996 and 17,619,252 euros, respectively which are not far from the model expected profits in Table 6, 16,870,731 and 17,608,877 euros, respectively. Other relevant sample parameters can be found in Table 7.

For a more precise comparison of the previous random profits, their empirical Cumulative Distribution Function (CDF) can be used (see Figure 6). Since the deterministic approach CDF is above the stochastic approach CDF, we can conclude that decision x_S^* has first-order stochastic dominance over decision x_{EV}^* [19]. Therefore decision x_S^* can be ranked as superior to decision x_{EV}^* from a probabilistic point of view (on top of a better expected profit). Notice that the these CDFs give useful further information as for example, the percentage of scenarios with shortfall (i.e., scenarios whose profit is below a given threshold) or the expected shortfall.

4.4 Determining the optimal allocation for a given budget

To compute the optimal budget one assumes that there is no limit on the available spending, as in the previous section. However, very often managers have to allocate a limited advertising budget. This

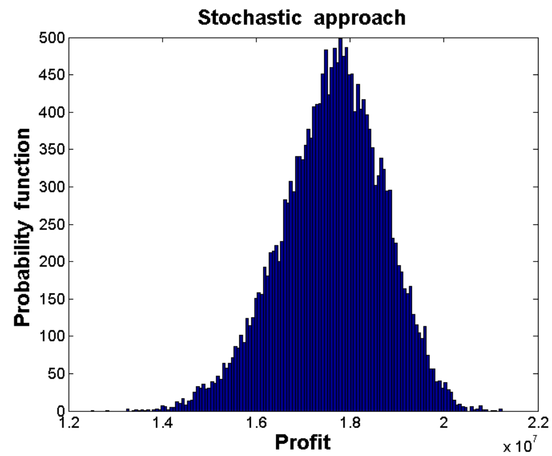

Figure 5: Sample probability distribution of the profit (euros) that may be obtained with the stochastic approach (sample expected profit 17,619,252 euros).

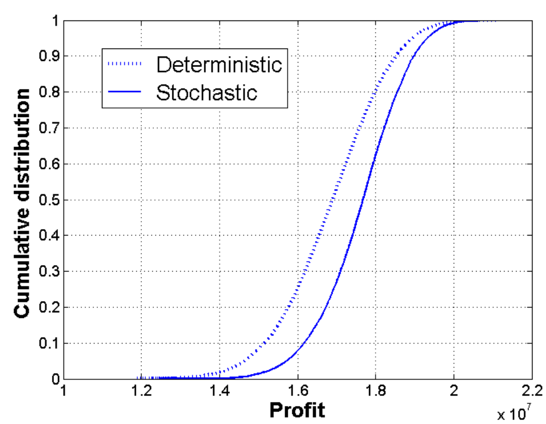

Figure 6: It is more likely to obtain a low profit by using the deterministic approach.

Table 8: Expected profit for the optimal budget and the reduced budget (stochastic approach).

	Optimal budget	Reduced budget	Variation
Budget	4,350,039	2,175,020	-50.00%
Expected profit	17,608,877	16,576,823	-5.86%
CPU time (s)	54	40	

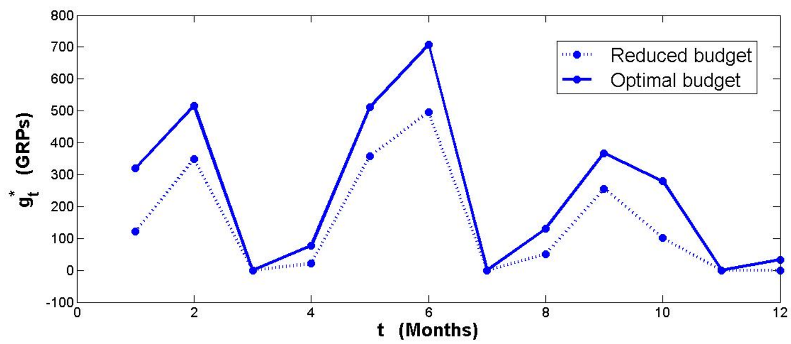

Figure 7: Optimal budget allocation g_t^* for advertising channel 1 of product P1 considering the optimal and the reduced budget, respectively.

can be done by imposing the following budget constraint in the MAB model:

$$\sum_{tjk \in \mathcal{TJK}} c_{tjk}\, g_{tjk} \leq b. \tag{26}$$

Notice that this is a particular case of constraint (21). Under this model, the advertising budget b is allocated among all the advertising channels for the different products and along all the periods. Of course, many other kind of constraints could be considered: the company could be interested in limiting the advertising budget within each time period, it could be interested in imposing a threshold for the adstock variables at the end of the planning horizon, it could impose a budget for each advertising channel, etc.

For example, if we were limited to 50% of the optimal budget as computed in Section 4.2, problem MAB$_S$ should be solved with $b = 2,175,020$ euros in (26). Table 8 shows the results obtained for both cases: optimal budget versus reduced budget. Observe that by reducing 50% (2,175,020 euros) the optimal budget for the 12 months, the optimal expected profit is reduced by 5.86% (1,032,054 euros). As an example, Figure 7 depicts the budget allocation corresponding to advertising channel 1 of product P1 obtained by the two approaches.

4.5 Model sensitivity

In this section we study the sensitivity of the model under parameter perturbations. More specifically we study the effect that parameter perturbations produce on the optimal profit of each model, namely EV, SP and EEV.

Table 9 shows the results of the test instances under consideration. Instance I-01 is the reference one and corresponds to the MAB instance analyzed in Section 4.3. Instances I-02 and I-03, with labels ‘p +25%’ and ‘p -25%’, correspond to the related modifications of I-01, such that all the profit

Table 9: Model sensitivity (results in euros): I-01 is the reference instance. The other ones have been generated by perturbing in $+25\%$ or -25% the indicated group of parameters in I-01.

Instance	Parameter	Perturbation	EEV	SP	EV
I-01			16,870,731	17,608,877	23,276,709
I-02	p	$+25\%$	22,125,135	23,154,999	30,060,925
I-03	p	-25%	11,766,665	12,248,365	16,590,051
I-04	c	$+25\%$	15,964,539	16,626,972	22,395,841
I-05	c	-25%	17,925,287	18,775,761	24,252,711
I-06	$\widetilde{y}_0$	$+25\%$	16,870,731	17,608,877	23,276,709
I-07	$\widetilde{y}_0$	-25%	16,870,731	17,608,877	23,276,709
I-08	α	$+25\%$	22,252,197	23,301,224	30,169,256
I-09	α	-25%	11,672,690	12,145,273	16,501,491
I-10	β	$+25\%$	18,347,819	19,114,247	24,135,405
I-11	β	-25%	14,874,176	15,563,217	22,001,988
I-12	γ	$+25\%$	16,759,756	17,484,094	23,177,511
I-13	γ	-25%	16,983,945	17,736,402	23,377,048
I-14	δ	$+25\%$	18,703,392	19,654,983	24,831,262
I-15	δ	-25%	15,450,388	16,033,161	21,958,752
Average			16,895,879	17,644,355	23,285,491
				+4.43%	+37.82%

parameters $\{p_{tj}\}_{tj \in \mathcal{TJ}}$ have been increased and decreased by 25%, respectively, and all the other parameters remain unchanged. Instances I-04 to I-15 correspond to the $\pm 25\%$ changes indicated in the table under the heading Parameter.

Regarding the CPU time for problem solving, the EV and SP solutions have been obtained in a few seconds and around one minute, respectively, for each of the instances. Observe that the inequality EEV $\leq$ SP $\leq$ EV holds for all of the instances, in accordance with Proposition 4 in Appendix A. Therefore, as we already pointed out, problem EV has two clear disadvantages compared to problem SP, namely, a worse expected profit (given by EEV) and an erroneous estimation of this expected profit (given by EV). As reported in the last line of Table 9, on average and in this context, a SP manager would have an expected profit 4.43% better than an EV manager. Also on average, the expected profit estimation given by an EV manager would have an error of 37.82%, since notice that the expected profit estimation is EV = 23,285,491 euros and the true expected profit is EEV = 16,895,879 euros.

5 Concluding remarks

The main contribution of this paper is to introduce a stochastic model for the Multiperiod Multiproduct Advertising Budgeting (MAB) problem. We call it problem MAB$_S$ and it is intended to solve the MAB problem under uncertainty, that is, by taking into account the randomness of the problem parameters. As far as we know, a *stochastic* version of the MAB problem has not been considered in literature.

Problem MAB$_S$ addressed in this work corresponds to a *single-stage* stochastic optimization problem with deterministic feasible set. From a theoretical point of view, it has been shown that problem MAB$_S$ is convex, thus, numerically tractable and with global optimal solutions. We have also proven that the optimal expected profit given by the stochastic approach is at least as good as the expected profit given by the deterministic approach.

From a practical point of view, first, it has been shown that the stochastic model, in comparison with its deterministic counterpart, allows for a better allocation of the advertising investment along the planning horizon. A MAB$_S$ instance derived from a real advertising campaign has been used as a pilot case in the computational experiment, where the stochastic model has improved by 4.38% the optimal expected profit computed by the deterministic one. Second, it has been assessed that the deterministic approach may be erroneous, thus misleading for managerial purposes. In the pilot case, the expected profit estimation given by the deterministic model has shown an error of almost 38% compared to the true expected profit (given by the expected profit of using the expected value solution). Third, it has been shown by example that the MAB$_S$ model can be used not only to compute the optimal budget, but for optimally allocating any budget combined with other managerial constraints. For example, it has been observed that, by reducing 50% the optimal budget, the advertising expected profit drops 5.86%.

Therefore, we can conclude that it is important for advertising budgeting to consider effective stochastic models as the one presented in this work. In this case, for a slightly higher conceptual and computational effort, the stochastic model MAB_S may significantly improve the deterministic model MAB_{EV} in two aspects: higher expected profit and more reliable results.

Finally, we would like to point out some limitations of the MAB$_S$ model here presented. The first limitation is that this stochastic model is risk neutral. That is, it is based on the expected profit and it does not incorporate any risk measure to cope with the cases of high variability of the profit over the scenarios as for example conditional Value-at-Risk [29, 33, 34] and stochastic dominance [16, 18, 19] among others. The second limitation is that the model does not take the product price as a sales driver [37], that is, as a decision variable (notice that prices are input data in the current version of the model). The third limitation is that the model does not take into account the competitive or cooperative aspects of advertising budgeting [26]. As a sequel of this paper, we are planning to improve this version of the MAB model regarding these three aspects. Furthermore, we will study the advantage of using a two-stage or a (restricted) multistage approach to the problem by clustering consecutive periods into stages and taking stage-based decisions allowing then recourse actions.

6 Appendix A: Theoretical results

In Proposition 1 we review some classical results which are basic but useful in this context (notice that we use some notation defined in Sections 2 and 3). The other propositions concern the MAB problem introduced in this paper.

Proposition 1

1. *(Jensen's inequality [21]) Let ξ be a random vector such that $\mathbb{E}\big[\xi\big] = \bar{\xi}$ and $G(\xi)$ be a convex function. Then*
$$G(\bar{\xi}) \leq \mathbb{E}[G(\xi)].$$
2. *Let ξ be a random vector and $F(x,\xi)$ be a convex function in ξ. Then*
$$\min_{x\in\mathcal{D}} F(x,\bar{\xi}) \leq \min_{x\in\mathcal{D}} \mathbb{E}\big[F(x,\xi)\big].$$
3. *If $F(x,\xi)$ is a convex function in ξ then:*
$$F^*_{EV} \leq F^*_S \leq F_S(x^*_{EV}).$$

4. *If $F(x,\bar{\xi})$ is a convex function in x, then P_{EV} is a convex optimization problem (i.e., minimization of a convex cost function and convex feasible set).*

5. *If $F(x,\xi)$ is a convex function in x for all $\xi \in \Xi$, then P_S is a convex optimization problem.*

Proposition 2 *If each $(\alpha_{tjk}, \beta_{tjk})$, for all $tjk \in \mathcal{TJK}$, is a pair of independent random variables, then $\mathbb{E}\big[P(x,\xi)\big]$ can be computed as follows*

$$
\begin{aligned}
\mathbb{E}\big[P(x,\xi)\big] \quad = \quad & \sum_{tjk \in \mathcal{TJK}} \Bigg\{ p_{tj}\ \mathbb{E}[\alpha_{tjk}] \left(1 - \mathbb{E}\left[e^{-\beta_{tjk}\ y_{tjk}}\right]\right) + \sum_{i \in \mathcal{I}} p_{ti}\ \mathbb{E}[\gamma_{tijk}]\ y_{tjk} \\
& \qquad - c_{tjk}\ g_{tjk} + z_{tjk} \Bigg\},
\end{aligned} \tag{27}
$$

where $\bar{\alpha}_{tjk} = \mathbb{E}[\alpha_{tjk}]$ and $\bar{\gamma}_{tjk} = \mathbb{E}[\gamma_{tjk}]$. Furthermore, if $\widehat{\xi}$ is an approximation of ξ based on the set of scenarios $\{\xi^s\}_{s \in \mathcal{S}}$ with respective probability weights $\{w^s\}_{s \in \mathcal{S}}$ such that

$$
\xi^s = \big((\alpha^s_{tjk})_{tjk \in \mathcal{TJK}}, \quad (\beta^s_{tjk})_{tjk \in \mathcal{TJK}}, \quad (\gamma^s_{tijk})_{tijk \in \mathcal{TIJK}}, \quad \varepsilon^s\big) \quad s \in \mathcal{S},
$$

then:

$$
\begin{aligned}
\mathbb{E}\big[P(x,\widehat{\xi})\big] \quad = \quad & \sum_{tjk \in \mathcal{TJK}} \Bigg\{ p_{tj}\ \bar{\alpha}_{tjk} \ \left(1 - \mathbb{E}\left[e^{-\widehat{\beta}_{tjk}\ y_{tjk}}\right]\right) + \sum_{i \in \mathcal{I}} p_{ti}\ \bar{\gamma}_{tijk}\ y_{tjk} \\
& \qquad - c_{tjk}\ g_{tjk} + z_{tjk}(g) \Bigg\},
\end{aligned} \tag{28}
$$

where $\bar{\alpha}_{tjk} = \sum_{s \in \mathcal{S}} w^s \alpha^s_{tjk}$, $\bar{\gamma}_{tijk} = \sum_{s \in \mathcal{S}} w^s \gamma^s_{tijk}$ and $\mathbb{E}\left[e^{-\widehat{\beta}_{tjk}\ y_{tjk}}\right] = \sum_{s \in \mathcal{S}} w^s e^{-\beta^s_{tjk}\ y_{tjk}}$.

Proof: Equation (27) follows by considering two basic results in probability. The first one is that $\mathbb{E}[\cdot]$ is a linear operator such that $\mathbb{E}[k_1 X + k_2 Y] = k_1 \mathbb{E}[X] + k_2 \mathbb{E}[Y]$ for any pair k_1, k_2 of constants and any pair X, Y of random variables. The second one is that $\mathbb{E}[XY] = \mathbb{E}[X]\ \mathbb{E}[Y]$ for any pair X, Y of independent random variables. Equation (28) comes from the fact that $\mathbb{E}[\widehat{\xi}\,] = \sum_{s \in \mathcal{S}} w^s \xi^s$. ■

Proposition 3 *Problem MAB$_S$ is a convex optimization problem.*

Proof: For a fixed ξ^s, the cost function $F(x,\xi^s)$ is convex in x if α^s_{tjk} and β^s_{tjk}, are positive for all $tjk \in \mathcal{TJK}$ (see [4]). This condition is satisfied since, α^s_{tjk} and β^s_{tjk}, for all $tjk \in \mathcal{TJK}$, are positive parameters in problem MAB$_S$. Then, $F(x,\xi^s)$ is a convex function in x for all $s \in \mathcal{S}$ and by Proposition 1, problem MAB$_S$ is a convex optimization problem (notice that the feasible set $\mathcal{D}$ is convex since it is defined by linear constraints). ■

Proposition 4 *Problem MAB$_{EV}$ is a convex optimization problem.*

Proof: Analogous to the proof of Proposition 3. ■

Proposition 5 *If each $(\alpha_{tjk}, \beta_{tjk})$, for all $tjk \in \mathcal{TJK}$, is a pair of independent random variables then for problems MAB$_{EV}$ and MAB$_S$ it results*

$$
F^*_{EV} \leq F^*_S \leq F_S(x^*_{EV}).
$$

Proof: Given the scenario-based random vector $\widehat{\xi} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}, \widehat{\varepsilon})^\intercal$ and its component $\widehat{\beta} = (\widehat{\beta}_{tjk})_{tjk \in \mathcal{TJK}}$, we define the following auxiliary function

$$Q(x, \widehat{\beta}) = \sum_{tjk \in \mathcal{TJK}} \Big\{ p_{tj}\ \bar{\alpha}_{tjk}\ \Big(1 - e^{-\widehat{\beta}_{tjk}\ y_{tjk}}\Big) + \sum_{i \in \mathcal{I}} p_{ti}\ \bar{\gamma}_{tijk}\ y_{tjk} \\ -c_{tjk}\ g_{tjk} + z_{tjk} \Big\}.$$

It is clear that $\mathbb{E}[P(x, \widehat{\xi})] = \mathbb{E}[Q(x, \widehat{\beta})]$ for all x. Therefore $F_S(x) = -\mathbb{E}[Q(x, \widehat{\beta})]$. Similarly, $P(x, \mathbb{E}[\widehat{\xi}]) = Q(x, \mathbb{E}[\widehat{\beta}])$ for all x. Therefore $F_{EV}(x) = -Q(x, \mathbb{E}[\widehat{\beta}])$. This implies that Q could be used instead of P to define problems MAB$_S$ and MAB$_{EV}$. On the other hand, it is easy to see that $-Q(x, \widehat{\beta})$ is a convex function in $\widehat{\beta}$. The proof can be completed by applying Proposition 1. ■

Algorithm 1 *(Finite support random variable approximation to a continuous random variable)* Let us consider the index sets $\mathcal{L} = \{1, \dots, L\},\ \mathcal{R} = \{1, \dots, R\},\ \mathcal{S} = \{1, \dots, S\}$, the continuous random vector $\xi = (\xi_1, \dots, \xi_R)^\intercal$ and the set of realizations (scenarios) $\{\xi^s\}_{s \in \mathcal{S}}$ with corresponding probability weights $\{w^s\}_{s \in \mathcal{S}}$. In this context, each random variable ξ_r for all $r \in \mathcal{R}$, can be approximated by a finite support random variable $\widetilde{\xi}_r$ constructed form the set $\{(\xi_r^s, w_r^s)\}_{s \in \mathcal{S}}$ as follows:

1. Set L, the cardinal of the support of $\widetilde{\xi}_r$.
2. Define the interval $\mathbb{I} = [\ \min\{\xi_r^s\}_{s \in \mathcal{S}},\ \max\{\xi_r^s\}_{s \in \mathcal{S}}\ ]$.
3. Partition $\mathbb{I}$ into L non-intersecting subintervals of equal length such that
$$\mathbb{I} = \bigcup_{l \in \mathcal{L}} \mathbb{I}^l.$$
4. Define $\widetilde{\xi}_r^l$ as the middle point of interval $\mathbb{I}^l$ for all $l \in \mathcal{L}$.
5. Define $\pi_r^l = \sum_{s \in \mathcal{S}} \{w^s \mid \xi_r^s \in \mathbb{I}^l\}$.
6. Then the set $\{(\widetilde{\xi}_r^l, \pi_r^l)\}_{l \in \mathcal{L}}$ defines a finite support random variable that we name $\widetilde{\xi}_r$ and that approximates the continuous random variable ξ_r such that
$$P\big(\xi_r \in \mathbb{I}^l\big) \approx P\big(\widetilde{\xi}_r = \widetilde{\xi}_r^l\big) = \pi_r^l \qquad l \in \mathcal{L}.$$

■

Notice that the set of probability values $\{\pi_r^l\}_{l \in \mathcal{L}}$ is an approximation to the marginal probability density function of ξ_r that we denote by f_r, for all $r \in \mathcal{R}$. As pointed out in [11], the information about the degree of stochastic dependence of the components of vector ξ is not incorporated into the marginal density functions f_r. As a consequence, although the marginal distributions of $\{\xi_r\}_{r \in \mathcal{R}}$ can be derived from their joint distribution, it is not possible to reconstruct the joint distribution of $\{\xi_r\}_{r \in \mathcal{R}}$ from their marginal distributions without additional information. Thus, although in the original scenarios $\{\xi^s\}_{s \in \mathcal{S}}$ the components ξ_r may be correlated across $r \in \mathcal{R}$, the marginal probability density functions treat each ξ_r as an independent random variable for all $r \in \mathcal{R}$.

7 Appendix B: Regression models and Bayesian inference

From a statistical point of view, the sales function in equation (6) corresponds to a regression model *nonlinear in the parameters* where the sales, say $\widetilde{S}$, is the dependent variable and the advertising investment g is the independent variable. Notice that, although regression literature normally uses the notation Y and x, we will use $\widetilde{S}$ and g instead, in order to maintain the same notation through the paper. Next we give some basic properties regarding regression models and Bayesian inference. Usually one distinguishes two types of regression models: linear and nonlinear models in the parameters [14]. To simplify the exposition we restrict ourselves to the simple regression case, i.e., one independent variable g, knowing that these results can be generalized to multiple regression.

The simple regression model *linear in the parameters* corresponds to

$$\widetilde{S} = \alpha + \beta g + \varepsilon,$$

where α and β are the unknown regression parameters and ε is the Gaussian random error, such that $\varepsilon \sim N(0, \sigma^2)$. In order to emphasize the functional dependence of g we will write $\widetilde{S} \mid g = \alpha + \beta g + \varepsilon$.

In regression models one is interested in $\mathbb{E}[\widetilde{S} \mid g]$, the expected value of the dependent variable for a given value of the independent variable. Taking into account that $\mathbb{E}[\varepsilon] = 0$, this expected value can be estimated as follows

$$\mathbb{E}[\widetilde{S} \mid g] \approx \hat{\alpha}^* + \hat{\beta}^* g, \tag{29}$$

where $\hat{\alpha}^*$ and $\hat{\beta}^*$ are the least square estimates of the unknown parameters α and β, respectively, that can be computed as follows. Given the data points $(g_1, \widetilde{s}_1), \ldots, (g_n, \widetilde{s}_n)$, compute:

$$\begin{aligned}
\hat{\alpha}^* &= \frac{\sum_{i=1}^{n} (\widetilde{s}_i - \bar{s})(g_i - \bar{g}_i)}{\sum_{i=1}^{n} (g_i - \bar{g})^2} \\
\hat{\beta}^* &= \bar{s} - \hat{\alpha}^* \bar{g} \\
\bar{g} &= \frac{1}{n} \sum_{i=1}^{n} g_i, \qquad \bar{s} = \frac{1}{n} \sum_{i=1}^{n} \widetilde{s}_i
\end{aligned}$$

(see Theorem 11.1.1 in [11]). In this section we will write $A \approx B$, to indicate that B is an estimation (approximation) of the true value A, which is unknown.

The simple regression model *nonlinear in the parameters* corresponds to

$$\widetilde{S} \mid g = h(g, \gamma) + \varepsilon,$$

where γ is the vector of unknown regression parameters, h is a nonlinear function of γ and ε is the Gaussian random error such that $\varepsilon \sim N(0, \sigma^2)$. In some cases, as pointed out in [14], a nonlinear regression model can be transformed into a linear one, and therefore can be analyzed as if it was linear. In other cases it is not possible such a transformation and the model has to be treated as nonlinear. For example, this is the case of the sales function in equation (6):

$$\widetilde{S} \mid g = \alpha(1 - e^{-\beta g}) + \varepsilon$$

where $\gamma = (\alpha, \beta)^\intercal$ is the vector of unknown regression parameters.

As in the linear case, one is interested in the expected value $\mathbb{E}[\widetilde{S} \mid g]$. This expected value can be estimated by

$$\mathbb{E}[\widetilde{S} \mid g] \approx h(g, \hat{\gamma}^*), \tag{30}$$

where $\hat{\gamma}^*$ is the least square estimate of the unknown vector of parameters γ, that can be computed as follows. Given the data points $(g_1, \widetilde{s}_1), \ldots, (g_n, \widetilde{s}_n)$, compute:

$$\hat{\gamma}^* = \arg\min_{\hat{\gamma}} \sum_{i=1}^{n} \left(\widetilde{s}_i - h(g_i, \hat{\gamma})\right)^2 \tag{31}$$

In contrast with the linear case, the least square approach applied to the nonlinear one has some drawbacks [14]:

1. In general, either there is no closed-form expression for the best-fitting parameters $\hat{\gamma}^*$ or it is mathematically involved to calculate it.
2. Usually numerical optimization algorithms are applied to determine the best-fitting parameters and there may be many local minima of the function to be optimized in (31).

So far, in this section we have used the so-called frequentist inference. In order to overcome its drawbacks for the nonlinear case above mentioned, one can use the so-called Bayesian inference. The main difference between both approaches lies is that in the Bayesian approach the unknown parameters are treated as random variables in every statistical inference problem. In contrast, the frequentist approach considers that it is not appropriate to assign a probability distribution to a parameter but claim instead that the true value of the parameter is a certain fixed number whose value happens to be unknown to the experimenter [11]. Therefore, by using Bayesian inference, γ and σ, are viewed as a random vector $\boldsymbol{\gamma}$ and a random variable $\boldsymbol{\sigma}$, respectively. Notice that in this section we write unknown parameters viewed as random variables in boldface. In this case, we have

$$\widetilde{S} \mid g \approx h(g, \boldsymbol{\gamma}) + \varepsilon,$$

whose expected value can be estimated as

$$\mathbb{E}[\widetilde{S} \mid g] \approx \mathbb{E}[h(g, \boldsymbol{\gamma})] \tag{32}$$

Notice that we have used that $\mathbb{E}[\varepsilon] = 0$.

Let us define the random variables $S_i = \widetilde{S} \mid g_i$ for $i = 1, \ldots, n$. Suppose that $\widetilde{S}_1, \ldots, \widetilde{S}_n$ are independent given $g_1, \ldots, g_n$, $\boldsymbol{\gamma}$ and $\boldsymbol{\sigma}$, with $\widetilde{S}_i$ having the normal distribution with mean $\mathbb{E}[h(g_i, \boldsymbol{\gamma})]$ and variance $\boldsymbol{\sigma}^2$. Let us also define the vector of unknown parameters $\boldsymbol{\xi} = (\boldsymbol{\gamma}^\intercal, \boldsymbol{\sigma})^\intercal$. According to [20], Bayesian inference can be used to combine initial information with new data. The initial information could result form previous studies, theoretical considerations, etc. Initial information about unknown parameters is expressed as a prior probability density function $\rho_0(\xi)$. The new sample information is represented by its likelihood function $f_n(\widetilde{s} \mid \xi)$. Bayes' theorem is then used to obtain a posterior probability density function $\rho_1(\xi \mid \widetilde{s})$, which incorporates both the initial information and sample information:

$$\rho_1(\xi \mid \widetilde{s}) = \frac{f_n(\widetilde{s} \mid \xi)\, \rho_0(\xi)}{g_n(\widetilde{s})},$$

where $\widetilde{s} = (\widetilde{s}_1, \ldots, \widetilde{s}_n)^\intercal$ is one observation of the random vector $(\widetilde{S}_1, \ldots, \widetilde{S}_n)^\intercal$ and g_n is its marginal density function. Further details can be found in [11, 22].

Compared to the frequentist approach, the Bayesian approach has the following (dis)advantages [36] in the context of nonlinear regression:

1. One advantage is that the expectation $\mathbb{E}[\widetilde{S} \mid g]$ can be estimated straightforward by sampling from the posterior probability density function of $\boldsymbol{\gamma}$ according to equation (32). In frequentist inference, according to equation (30), one estimates this expectation by solving the least squares optimization problem (31), which in general is mathematically involved or requires numerical optimization with possible local optima, as already said.

2. Another advantage is the use of a prior probability density function, which allows the modeler to incorporate the expert knowledge of the problem. However, how to specify an adequate prior density function may also be seen as a disadvantage.

3. The second disadvantage of Bayesian inference is the need to evaluate multiple integrals as for example the computation of the marginal density function

$$g_n(\widetilde{s}) = \int f_n(\widetilde{s} \mid \xi)\, \rho_0(\xi)\, d\xi,$$

which very often is analytically complex or intractable in the case of nonlinear regression. With the increasing computing power, this drawback has been overcome by using sampling methods, in particular the so-called Markov Chain Monte Carlo (MCMC) method. In general, these methods are computationally intensive, therefore time consuming.

A complete comparison of the frequentist and Bayesian approaches to nonlinear regression modelling can be found in [36].

Acknowledgements: We thank the reviewers for their suggestions and comments, which helped to improve this paper. Thanks are due to the support of the consulting company Bayes Forecast (www.bayesforecast.com). We also thank the support of: grant S2009/esp-1594 from the 'Comunidad de Madrid' (Spain), grants MTM2012-36163-C06-06 and MTM2012-31514 from the Spanish government and grant 71401106 from the National Natural Science Foundation of China.

References

[1] Md. T. Ahmed and C. Kwon. Optimal contract-sizing in online display advertising for publishers with regret considerations. *Omega*, 42:201–212, 2014.

[2] A. Albadvi and H. Koosha. A robust optimization approach to allocation of marketing budgets. *Management Decision*, 49(4):601–621, 2011.

[3] F. M. Bass, N. Bruce, S. Majumdar, and B. P. S. Murthi. Wearout effects of different advertising themes: A dynamic bayesian model of the advertising-sales relationship. *Marketing Science*, 26(2):179–195, 2007.

[4] C Beltran-Royo, H. Zhang, L. A. Blanco, and J. Almagro. Multistage multiproduct advertising budgeting. *European Journal of Operational Research*, 225(1):179–188, 2013.

[5] U. K. Bhattacharya. A chance constraints goal programming model for the advertising planning problem. *European Journal of Operational Research*, 192:382–395, 2009.

[6] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, 2nd edition, 2011.

[7] S. Broadbent. One way TV advertisements work. *Journal of the Market Research Society*, 21(3):139–165, 1979.

[8] A. Buratto, L. Grosset, and B. Viscolani. Advertising channel selection in a segmented market. *Automatica*, 42(8):1343 – 1347, 2006.

[9] P. J. Danaher and R. T. Rust. Determining the optimal return on investment for an advertising campaign. *European Journal of Operational Research*, 95:511–521, 1996.

[10] J. Deane and A. Agarwal. Scheduling online advertisements to maximize revenue under variable display frequency. *Omega*, 40:562–570, 2012.

[11] M. H. DeGroot and M. J. Schervish. *Probability and Statistics*. Addison Wesley, 2002.

[12] A. Deza and M. R. Huang, K. aand Metel. Chance constrained optimization for targeted Internet advertising. *Omega*, 53:90–96, 2015.

[13] P. Doyle and J. Saunders. Multiproduct advertising budgeting. *Marketing Science*, 9(2):97–113, 1990.

[14] N. R. Draper and H. Smith. *Applied regression analysis 3rd ed.* John Wiley and Sons, 1998.

[15] R. Du, A. Hu, and S. Ai. Stochastic optimal budget decision for advertising considering uncertain sales responses. *European Journal of Operational Research*, 183:1042–1054, 2007.

[16] L. F. Escudero, A. Garin, M. Merino, and G. Perez. On time stochastic dominance induced by mixed integer-linear in multistqage stochastic programs. *Submitted*, 2014.

[17] M. Garrat, R. Alcaraz, and S. Cohen. From brilliant to actionable: It takes technical brilliance and constant questioning to achieve the truly actionable in marketing ROI. In *Esomar - Congress*, 2011.

[18] R. Gollmer, U. Gotzes, and R. Schultz. A note on second-order stochastic dominance constraints induced by mixed-integer linear recourse. *Mathematical Programming*, 126(1):179–190, 2011.

[19] R. Gollmer, F. Neise, and R. Schultz. Stochastic programs with first-order dominance constraints induced by mixed-integer linear recourse. *SIAM Journal on Optimization*, 19(2):552–571, 2008.

[20] D. M. Hanssens, L. J. Parsons, and R. L Schultz. *Market response models: Econometric and time series analysis*. Kluwer academic publishers, second edition, 2001.

[21] J.-B. Hiriart-Urruty and C. Lemarechal. *Fundamentals of convex analysis*. Springer, 2001.

[22] P. D. Hoff. *A first course in Bayesian statistical methods*. Springer, 2009.

[23] J. Huang, M. Leng, and L. Liang. Recent developments in dynamic advertising research. *European Journal of Operational Research*, 220(3):591–609, 2012.

[24] P. Kall and J. Mayer. *Stochastic Linear Programming*. Kluwer academic publishers, 2005.

[25] A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.

[26] S. X. Li, Z. Huang, J. Zhu, and P. Y. K. Chau. Cooperative advertising, game theory and manufacturer-retailer supply chains. *Omega*, 30(5):347–357, 2002.

[27] M. Nerlove and K. J. Arrow. Optimal advertising policy under dynamic conditions. *Economica*, 29(114):129–142, 1962.

[28] B. Perez-Gladish, I. Gonzalez, A. Bilbao-Terol, and M. Arenas-Parra. Planning a TV advertising campaign: A crisp multiobjective programming model from fuzzy basic data. *Omega*, 38:84–94, 2010.

[29] G. Ch. Pflug and W. Romisch. *Modeling, measuring and managing risk.* World Scientific Singapore, 2007.

[30] A. Prasad and S. P. Sethi. Competitive advertising under uncertainty: A stochastic differential game approach. *Journal of Optimization Theory and Applications*, 123:163–185, 2004.

[31] A. Prekopa. *Stochastic Programming*. Springer, 1995.

[32] K. A. A. D. Raj, B. Kanagasabapathi, S. Shrivastava, K. Krishnan, and M. Shah. Performance evaluation of adstock models using market drivers in the consumer packaged goods (CPG) industry. *International Journal of Electronic Marketing and Retailing*, 5(2):173–186, 2012.

[33] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[34] A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory*, volume 9. Society for Industrial Mathematics, 2009.

[35] S. Sriram and M. U. Kalwani. Optimal advertising and promotion budgets in dynamic markets with brand equity as a mediating variable. *Management Science*, 53(1):46–60, 2007.

[36] J. Wakefield. *Methods and Models in Statistics*, chapter Non-linear regression modelling and inference, pages 119–153. London: Imperial College Press, 2004.

[37] J. Zhang, W. K. Chiang, and L. Liang. Strategic pricing with reference effects in a competitive supply chain. *Omega*, 44:126–135, 2014.