



## Marketing Science

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### Valuing Customer Portfolios with Endogenous Mass and Direct Marketing Interventions Using a Stochastic Dynamic Programming Decomposition

Mercedes Esteban-Bravo, Jose M. Vidal-Sanz, Gökhan Yildirim

To cite this article:

Mercedes Esteban-Bravo, Jose M. Vidal-Sanz, Gökhan Yildirim (2014) Valuing Customer Portfolios with Endogenous Mass and Direct Marketing Interventions Using a Stochastic Dynamic Programming Decomposition. Marketing Science 33(5):621-640.  
<https://doi.org/10.1287/mksc.2014.0848>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2014, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Valuing Customer Portfolios with Endogenous Mass and Direct Marketing Interventions Using a Stochastic Dynamic Programming Decomposition

Mercedes Esteban-Bravo, Jose M. Vidal-Sanz

Department of Business Administration, Universidad Carlos III de Madrid, 28903 Getafe, Madrid, Spain  
{mesteban@emp.uc3m.es, jvidal@emp.uc3m.es}

Gökhan Yildirim

Department of Management Science, Management School, Lancaster University, Bailrigg,  
Lancaster LA1 4YX, United Kingdom, g.yildirim@lancaster.ac.uk

The customer relationship management allocation in marketing budgets is potentially misleading when it uses individual customer lifetime value estimations from historical data. Planned marketing interventions would change the purchasing behavior of different customers, and history-based decisions would thus be suboptimal. To cope with this inherent endogeneity, we model the optimal allocation of the marketing mix by accounting simultaneously for mass interventions and direct marketing interventions for each customer. This is a large stochastic dynamic problem that, in general, is computationally rather intractable as a result of the “curse of dimensionality.” We present an algorithm to derive the optimal marketing policies (how the firm should allocate its marketing resources) and the expected present value of those decisions, which maximize the long-term profitability of firms. This allows the firm to value customers/segments and helps the firm to target those that maximize long-term profitability given the optimal marketing resources allocation. We apply the proposed approach in the context of a kitchen appliance manufacturer. The results identify the most effective marketing policies and the endogenous customer values. It is in this context that we also dynamically identify the most profitable customer and the short- and long-term effects of marketing activities on each customer.

**Keywords:** CRM; marketing resource allocation; long-term effect of marketing activities; stochastic dynamic programming; dynamic panel-data models

**History:** Received: September 14, 2012; accepted: December 26, 2013; Preyas Desai served as the editor-in-chief and Duncan Simester served as associate editor for this article. Published online in *Articles in Advance* April 21, 2014.

## 1. Introduction

Customers are a firm's main assets, and marketing departments increasingly adopt customer relationship management (CRM) programs to improve customer acquisition, expenditure, and retention. In essence, CRM involves a systematic allocation of differential resources to customers based on the individual value of these customers to the business. The resources allocated to each customer can be channeled through a mix of alternative interventions and can be complemented by mass actions. Marketing resource allocation has traditionally been based on heuristic rules (see Mantrala 2002), but the benefits of CRM policies are today justified by their impact on firms' returns (Rust et al. 2004). To plan the allocation of resources, managers therefore need to maximize the value of their customer base. Ideally, this concept is measured by the sum of customer lifetime values (CLV)—that is, the sum of net present values of cash flows between a customer and the firm (Gupta et al. 2004, Gupta and Lehmann 2006). The assessment of customers' values and the

effectiveness of a marketing intervention are typically based on the econometric analysis of large customer databases.

CRM requires planning a portfolio of a mix of marketing intervention alternatives. Typically, the literature on budget allocation considers mass interventions from the marketing mix such as advertising, promotion, sales force, prices and price promotions, product, and distribution (for a review, see, e.g., Gupta and Steenburgh 2008, Shankar 2008). The direct marketing literature typically considers a single intervention that is customized or at least tailored to small segments. For example, it is common to use certain pricing decisions (Lewis 2005), catalog mailing (see, e.g., Bitran and Mondschein 1996, Gönül and Shi 1998, Gönül and Ter Hofstede 2006, Simester et al. 2006), couponing decisions (e.g., Bawa and Shoemaker 1987, Rossi et al. 1996), direct mailing (Roberts and Berger 1989), and relationship-oriented magazines (Berry 1995, Bhattacharya and Bolton 1999, McDonald 1998).

Planning optimal CRM interventions by maximizing the expected global CLVs is, by all means, a difficult task. In an attempt to address this, the standard CRM procedure allocates a marketing budget to each individual customer, after ranking customers by their CLV value (Reinartz et al. 2005, Rust et al. 2004, Venkatesan and Kumar 2004). Assessing new marketing interventions using CLVs computed from historical data is potentially misleading. The planned CRM marketing interventions will change the purchasing behavior of different customers, changing their CLVs, turning the customers' rankings upside down, and making our history-based decisions suboptimal. To cope with this inherent endogeneity, the objective of a marketing mix allocation model should be a CLV measure computed as the optimal value achieved when the optimal CRM investment is implemented. The idea is that when the CLV is computed, we should take into account how customers will react to the changes in the CRM policies.

To avoid this endogeneity problem, some authors have tried to optimize the expected CLVs. Rust and Verhoef (2005) optimize each individual customer's profitability year by year (which has been described as myopic planning). Alternatively, other authors have optimized the expected CLV using stochastic dynamic programming (SDP). Although this is a natural approach to solving this problem, SDP is affected by the "curse of dimensionality" (the complexity increases drastically with the size of the problems). Therefore, they consider a partial solution that ignores mass interventions (aimed at all customers) and that focuses on direct individual interventions so that the investment decision for each customer is independent and the standard SDP algorithms can be applied to the problem of the "decoupled decision." In this respect, Gönül and Shi (1998) and Montoya et al. (2010) study direct marketing problems, whereas Khan et al. (2009) estimate the impact of multiple promotional retail instruments on customer behavior (e.g., discount coupons, free shipping offers, a loyalty program), designing a customized promotional schedule that solves a different SDP problem for each customer. Yet how to optimize simultaneously both types of interventions (mass and direct ones) is an unsolved issue because the SDP optimization problems are not separable among customers. Maximizing the expected CLVs of a customer's portfolio with multiple types of personalized and mass marketing interventions, accounting for long-term returns, and solving the endogeneity issue is what Rust and Chung (2006, p. 575) call the "holy grail" of CRM.

In this paper, we provide a fully tailored approach for planning policies that maximize the expected CLV of all customers in the market by accounting for the endogeneity issues. Our approach considers that customer behavior follows a Markov model in which sales

respond to mass and direct marketing interventions and in which marketing expenditures are allocated to maximize the sum of expected CLVs for all its customers. Because such models can become rather intractable in general, we propose a method to address this problem by splitting it into manageable pieces (subproblems) and by coordinating the solutions of these subproblems. With this approach, we obtain two main computational advantages. First, the subproblems are, by definition, smaller than the original problem and therefore are much faster to solve. Second, uncertainty can be easily handled in each of the subproblems. To validate the efficiency of the approach, we provide a proof of convergence and solve several stochastic dynamic CLV models. The numerical results show the effectiveness of the method in solving large-scale problems.

In addition, we present an empirical application and consider the case of a medium-sized international company based in Eastern Europe that manufactures and distributes wholesale built-in electrical appliances for kitchens. Because this is a firm with various forms of sales response, its marketing budget allocation strategy involves general marketing investments (mainly advertising and promotions in professional fairs) and personalized customer investments. In the present paper, we therefore investigate whether these two types of interventions differ across customers. The results show that companies should consider different strategies for different customers to achieve long-term profitability over all time periods.

The rest of this paper proceeds as follows. In §2, we provide a model for dynamically allocating marketing budgets in the context of CRM. The present model simultaneously considers direct marketing interventions tailored to each customer and mass marketing interventions aimed at the customer base. In §4, we present the proposed decomposition methodology. In §5, we illustrate the performance of the algorithm using numerical simulations and provide a proof of convergence. In §6, we present an empirical application of the model to customers of a manufacturer of kitchen appliances. Finally, in §7, we discuss the results and provide some concluding remarks. The appendix provides technical details about the algorithm implementation.

## 2. A Model for Optimal Dynamic Budget Allocation in CRM

Planning marketing interventions in CRM requires managers to allocate budgets dynamically by maximizing the sum of expected CLVs from all customers based on information on history of customer state. To address the optimal budget allocation problem, the firm must carry out two tasks (see, e.g., Gupta et al. 2009).

• *Task 1:* To calculate CLV, a company has to estimate the customer response to marketing strategies. We model customer response as a Markov process (Gupta and Lehmann 2003, 2005; Kamakura et al. 2005; Gupta and Zeithaml 2006).

• *Task 2:* The company will be able to evaluate whether a marketing strategy is optimal by solving its decision problem, which includes all individual customer responses (see, e.g., Rust and Verhoef 2005, Rust and Chung 2006).

In other words, the basic problem that a company faces is determining optimal marketing strategies over time. This optimization problem is done subject to some dynamics that define how these strategies translate into sales and, in turn, into profits. Then, the first task is the study of these dynamics as a dynamic panel sales response model. We assume a market with  $I$  customers. The firm decides among possible strategies sequentially in each period of time  $t \in \{0, 1, 2, 3, \dots\}$ , such as

- $e_{it}$ , the direct marketing interventions on each customer  $i \in \mathbb{I}$  at period of time  $t$ , such as personalized advertising and directed promotional expenditures;
- $A_t$ , the mass marketing interventions at period of time  $t$ , such as the mass marketing investment; and
- $P_t$ , the prices for the different products.

Note that the notation  $e_t$  includes direct marketing interventions of all the customers.

There is a set of consumer states,  $S_t$ , describing the sales-level state of each customer  $i \in \mathbb{I}$  at each period of time  $t$  depending on the decision variables  $A, P, e$ . However, the same approach can be easily adapted to other state variables (e.g., if  $S_t$  are cash flows drawn from each customer). In a Markov process, all the historical information needed to model the customer behavior at each period of time to make a decision is captured in the set of states variables. The (time-invariant) cumulative probability distribution that is conditional on previous decisions and previous state information, is

$$\begin{aligned} F(s' | s, A, P, e) \\ = \Pr(S_t \leq s' | S_{t-1} = s, A_{t-1} = A, P_{t-1} = P, e_{t-1} = e), \end{aligned}$$

which determines the probability of the event  $S_t \leq s'$  given that the consumer was previously in state  $s$ , and the decision variables  $A, P, e$  denote the decision made after observing state  $S_{t-1}$ . We assume that individuals are statistically independent, so that

$$F(s' | s, A, P, e) = \prod_{i \in \mathbb{I}} F_i(s'_i | s_i, A, P, e_i).$$

Two possible approaches for these probability distributions are as follows:

1. *Finite-state problems:* A finite number of possible states  $\{s^1, \dots, s^m\}$  is considered for the sales to each individual. Then  $F_i(s'_i | s_i, A, P, e_i)$  can be characterized

by the transition matrix  $\Pi_i(A, P, e_i)$ , where the  $(r, s)$  element is a probability  $\pi_{r,s}^i$  linking to subsequent states  $s^r$  and  $s^s$ , which depends on the marketing actions  $(A, P, e_i)$ . Markov processes with a finite number of states are known as Markov chains.

2. *Infinite-state problems:* These postulate a sales response model for each customer in the panel,

$$S_{it} = g_i(S_{it-1}, A_{t-1}, P_{t-1}, e_{it-1}, \varepsilon_{it}), \quad (1)$$

where sales responses can vary across customers to allow heterogeneity in the expected responses, and the error term  $\varepsilon_{it}$  as an independent and identically distributed (i.i.d.) random variable with zero mean, statistically independent from  $S_{it-1}, A_{t-1}, P_{t-1}, e_{it-1}$  and cumulative probability distribution  $H_i$ . Then,

$$F_i(s'_i | s_i, A, P, e_i) = \int_{\{\varepsilon_i: g_i(s_i, A, P, e_i, \varepsilon_i) \leq s'_i\}} H_i(d\varepsilon_i).$$

The second approach is more flexible, and usually finite-state problems are regarded as a discrete approximation to processes with continuous states. Although this paper will focus on the general case of infinite states, we try to reconcile the generality of the model (1) with econometric convenience (postulating estimable models for which stationarity can be easily determined). Therefore, we henceforth consider that the company customer response is defined by a standard dynamic panel regression model

$$S_{it} = \rho S_{it-1} + g_i(A_{t-1}, P_{t-1}, e_{it-1}) + \varepsilon_{it}, \quad (2)$$

where the error term  $\varepsilon_{it}$  is assumed to be an i.i.d. random variable with zero mean and constant variance over time and is independent among customers. We will assume that the sales response from the previous period influences the sales responses of the current period at a constant rate  $|\rho| < 1$ . Because sales responses can vary across customers to allow heterogeneity in the expected responses, we assume that there are different transition probabilities for each customer as follows:

$$\begin{aligned} F_i(s'_i | s_i, A, P, e_i) &= \Pr\{\varepsilon_i \leq s'_i - \rho s_i - g_i(A, P, e_i)\} \\ &= H_i(s'_i - \rho s_i - g_i(A, P, e_i)), \end{aligned}$$

where  $H_i(\cdot)$  is the cumulative distribution of the error term  $\varepsilon_{it}$ . The one-lag memory structure imposed here can be relaxed by considering  $p$ -lags autoregressive models in the space of states.

The dynamic model (2) can be estimated using standard econometric techniques for time series cross-section and/or dynamic panels. Firms increasingly store large panel databases with information on their customers, including social information (such as sociodemographic, geographic information, and lifestyle habits) and trade internal data (such as historical transaction records, customer feedback, and Web browsing records); see Bose and Chen (2009). The econometric literature

has developed a battery of linear and nonlinear models for the dynamic analysis of large data panels; marketing researchers have tailored these models for the prediction of future purchases at the customer level (e.g., Schmittlein and Peterson 1994). Using these tools, company managers often estimate the expected CLV for each customer based on its past behavior (generally in a context of *ceteris paribus*, omitting or fixing the marketing mix variables).

Managers are typically profit maximizers, and as such, they face the problem of determining the price and advertising investment that returns the greatest profit. Once the company has estimated the sales responses to different decision variables of interest over time, the company can undertake solving Task 2. In other words, the firm should choose the optimal CRM strategies that maximize the expected sum of its CLVs. This problem is a large-dimensional (discounted) SDP problem. Mathematically, we model a rational forward-looking firm that has to decide on CRM budget allocation strategies over time,  $A_t, P_t, e_t$ , drawing profits

$$r(S_t, A_t, P_t, e_t) := \sum_{i \in I} r_i(S_{it}, A_t, P_t, e_{it}) \quad (3)$$

at each period of time  $t \geq 0$  from all of their customers.<sup>1</sup> In particular, we consider that

$$r_i(S_{it}, A_t, P_t, e_{it}) := (P_t - c_0) \cdot S_{it} - c_i(e_{it}) - c_m(A_t)/I, \quad (4)$$

where  $c_0$  is the unit cost,  $c_i(e_{it})$  is the cost of individual marketing effort  $e_{it}$ , and  $c_m(A_t)$  is the cost of general advertising actions  $A_t$ . (Marketing actions are often in monetary units, so the cost functions are identity functions:  $c_m(A) = A$ ,  $c_i(e_i) = e_i$ .) The objective of the firm is to maximize the expected net present value over the planning horizon,  $E_0[\sum_{t \geq 0} \delta^t r(S_t, A_t, P_t, e_t)]$ , where  $\delta \in (0, 1)$  is the discount factor.

The decisions  $(A_t, P_t, e_t)$  at each period of time are usually constrained to a set  $\mathbb{A}$  of feasible strategies (for example, requiring that the gross investment in marketing allocation must be nonnegative and below a certain limit). We can fix  $(P_t - c_0) = 1$  for all  $t \geq 0$  (introducing constraints  $P_t = 1$  at any time and setting  $c_0 = 0$ ) and then interpret  $S_{it}$  directly as monetary cash flows drawn from individual  $i$ . Notice that the set of feasible strategies can be updated based on the state information  $S_t$ . Therefore, the maximization problem considers  $(A_t, P_t, e_t) \in \mathbb{A}(S_t)$  for each time period, where  $\mathbb{A}(S_t)$  is a nonempty compact set.

**DEFINITION 1 (VALUE FUNCTION).** Conditionally on the initial state  $S_0$ , define the maxim expected value of the discounted expected profit for the firm as

$$V(S_0) := \max_{\{(A_t, P_t, e_t) \in \mathbb{A}(S_t)\}_{t \geq 0}} E_0 \left[ \sum_{t \geq 0} \delta^t r(S_t, A_t, P_t, e_t) \right], \quad (5)$$

which is known as the value function.

**DEFINITION 2 (POLICY FUNCTION).** The value function is characterized by a time-invariant function of the states, known as the policy function  $(A^*(s), P^*(s), e^*(s))$ , such that

$$V(S_0) = E_0 \left[ \sum_{t \geq 0} \delta^t r(S_t, A^*(S_t), P^*(S_t), e^*(S_t)) \right].$$

The idea is that at each time  $t$ , after observing  $S_{t-1}$ , the optimal decisions can be determined with the mapping  $A^*(S_{t-1}), P^*(S_{t-1}), e^*(S_{t-1})$ , and the output  $S_t$  is drawn from

$$\Pr(S_t \leq s' \mid S_{t-1}, A^*(S_{t-1}), P^*(S_{t-1}), e^*(S_{t-1})).$$

Some papers use the alternative notation  $\Pr(S_t \leq s' \mid S_{t-1}, A_t, P_t, e_t)$  for the transition probability instead of  $\Pr(S_t \leq s' \mid S_{t-1}, A_{t-1}, P_{t-1}, e_{t-1})$ . If the control decision is based on the conditioning state  $S_{t-1}$ , the optimization problem does not change, and with both formulations, we get the same policy function  $(A^*(s), P^*(s), e^*(s))$  and the same value function  $V(s)$ .

The computation of the value function  $V(s)$  involves solving an SDP problem in discrete time. Under general regularity conditions, value functions (and policy functions) are characterized by the Bellman equation (Bellman 1955, 1957):

$$V(S_t) = \max_{(A_t, P_t, e_t) \in \mathbb{A}(S_t)} \{r(S_t, A_t, P_t, e_t) + \delta E_t[V(S_{t+1})]\}.$$

Therefore, for each period of time  $t$ , we can compute the expected present discounted value of profits under the current state  $S_t$  as  $V(S_t)$ . The Bellman equation indicates that the value of choosing  $(A_t, P_t, e_t)$  at time  $t$  is equal not only to the immediate expected utility  $r(S_t, A_t, P_t, e_t)$  at time  $t$  but also the additional expected value at time  $t + 1$ , having the augmented information given by  $S_{t+1}$ .

To summarize, managers can optimize their decision-making process as

1. the policy rule  $(A^*(s), P^*(s), e^*(s))$ , which defines the optimal decision based on the sales  $s$  observed in the previous period; or
2. the value function  $V(s)$ , which gives managers the company value derived from the CLVs customer portfolio at the previous period of time.

We should note that there are clear advantages to using this approach: policy rules are simple (ease of understanding for managers) and adaptive (the decisions can be automatically updated as new state information becomes available). Note that they also can be used for simulation. For each period of time  $t$ , given  $S_t$  drawn from the conditional distribution  $F(s \mid S_{t-1}, A_{t-1}, P_{t-1}, e_{t-1})$ , the values  $A_t = A^*(S_t)$ ,  $P_t = P^*(S_t)$ , and  $e_t = e^*(S_t)$  can be used to simulate Monte Carlo scenarios for  $S_{t+1}$  recursively and then to compute numerically the expected path for the optimal policies  $E[A_t]$ ,  $E[P_t]$ , and  $E[e_t]$  and states  $E[S_t]$  as well as confidence intervals.

<sup>1</sup> We use the standard notation “:=” for definitions.

### 3. An Overview of Algorithms for Solving SDP Problems

In dynamic programming, SDP problems may be solved by using Bellman's principle of optimality. There are only a few special types of dynamic problems, however, that can be solved analytically (see, e.g., Ljungqvist and Sargent 2004, Chapter 4). When no analytical representation of the value function is available, an approximation can be obtained numerically through approaches based on value iteration or policy function iteration. In Appendix A, we present a technical overview of these methods for both decision problems with finite states and with continuous states. Although the continuous problems are generally more difficult to handle, both types can be generally computed if the number of state variables is small.

However, the computation of a large SDP problem remains one of the most challenging optimization problems. SDP problems can become intractable as the dimension of the state space increases (that is, the CPU time to calculate a value function increases exponentially in the dimension of the state space). This is known as the *curse of dimensionality* (a term coined by Richard Bellman, who was the first to recognize this problem); it appears in problems with just five- or six-dimensional state spaces (see Bellman 1961, p. 94), and in our context, it implies that we would not be able to solve problems with more than a small number of different clients. However, Bellman himself usually considered more simple Markovian structures, and with more sophisticated models, we often find problems with just four state variables.

These problems are associated with the integral  $E_t[V(S_{t+1})] = \int V(s')F(ds' | S_t, A_t, P_t, e_t)$ , which is needed to evaluate the Bellman equation. The curse of dimensionality can be found in many problems related to the numerical approximation of integrals by discrete summations, regardless of whether we consider deterministic or Monte Carlo methods. For example, this is why it is virtually impossible to nonparametrically estimate density functions with more than three to four variables (although, a very smooth function could be estimated with up to six or seven variables; see Silverman 1986, pp. 93–94). For a dimension larger than five or six, the required samples are huge, and algorithms become so slow that they are infeasible. In its most basic form, the curse of dimensionality means that, for all types of discrete grids that we use, the size required to keep the approximation error below some tolerance grows exponentially with the dimension. Even with dynamic programming for discrete Markov chains (no integral is involved here), we find a numerical problem: the number of elementary algebraic operations required to solve the Bellman equation also grows exponentially with the number of state variables. For a discrete problem with five

customers, a state grid of five elements, and a control grid of seven elements, the size of the transition matrix is  $3,125 \times 3,125 \times 16,807$ . The largest matrix that can be created in MATLAB on 64-bit platforms is with about 231 elements. Alternatively, we could consider sparse matrix representation and/or parallel computing, but then the computational cost goes to search over indices and eventually we exhaust computer memory.

There is a relatively recent literature that can be used to improve the performance of classic SDP algorithms, known as approximate dynamic programming (ADP) methods (sometimes also called reinforced learning). The main idea of ADP is to replace the expectation in the Bellman equation with some sort of refined statistical approximation in an adaptive way. In particular, in each step the challenge of looping between all possible states is substituted with looping some states that some estimations consider important (see Powell 2007, Adelman and Mersereau 2008). These methods have proven successful in some problems (see Simão et al. 2009, 2010) but are known to have failed in many others; they can even diverge or suffer from an overflow. As a general rule, the convergence of ADP algorithms is not guaranteed (Powell 2007, p. 120). In contrast to heuristic ADP approaches, our method has guaranteed convergence.

In this paper, we aim to address this problem by splitting it into manageable pieces (subproblems) and by coordinating the solutions of these subproblems. The subproblems are, by definition, smaller than the original problem and therefore much faster to solve. This methodology lies within the general approach to decomposition algorithms. Based on mathematical programming, these algorithms may be divided into three categories: Dantzig–Wolfe decomposition, Benders decomposition, and augmented Lagrangian relaxation procedures. Both Dantzig–Wolfe decomposition (Dantzig and Wolfe 1960) and Benders decomposition (see Benders 1962, Geoffrion 1972) are efficient schemes for dealing with convex optimization problems, and augmented Lagrangian relaxation (see Ruszczyński 1995) attains an extension to nonconvex problems. In general, these approaches apply to problems that can be completely decoupled or that at least do not have common control variables. However, here we face a problem with constraints that cannot be completely separated because we have common controls, and this greatly complicates the computations.

Some attempts to solve these types of SDP problems combine traditional decomposition algorithms and statistical sampling techniques. Sampling is used to create a scenario tree that represents uncertainty (Heitsch and Römisch 2009), and the original problem is then approximated by a finite deterministic one. As the dimension of the tree grows exponentially with the number of state variables, so does the complexity of

the deterministic problem. Decomposition methods are used to tackle this issue, as is the case in the Benders and Lagrangian schemes (see Birge and Louveaux 1997), but these methods may converge slowly in practice (see Chun and Robinson 1995). In contrast, the current paper first considers the decomposition of the original stochastic problem by using the law of iterated expectations on the value function, and after that, it solves new subproblems by using either value-iteration or policy-iteration algorithms. The approach implements a successive substitution iteration on the value functions of each subproblem, where one subproblem is solved using the previous value for another subproblem. The procedure is generally continued until the changes made by iteration are below some tolerance. In Appendix C, a proof of convergence for the algorithm is developed.

#### 4. Solving SDP Problems Using a Bellman Decomposition Algorithm

In this section we present a new algorithm for solving the SDP problem (5) that circumvents the curse of dimensionality. The key idea of our methodology is to consider a set of smaller independent problems that have optimal solutions that coincide with the corresponding solution for the problem (5). The decomposition approach draws on the partition of the decision variables into individual and aggregate decisions, and this implies that there are subproblems that only consider individual decision variables and subproblems that only consider aggregate decision variables. To split the problem (5) into manageable subproblems, we consider the law of iterated expectations.

To simplify the exposition, we first consider the myopic problem:

$$v^{\text{myop}}(S_{t-1}) = \max_{\{(A, P, e_1, \dots, e_t)\}} \{E_{t-1}[r_1(S_{1t}, A, P, e_1) + r_2(S_{2t}, A, P, e_2) + \dots + r_t(S_{tt}, A, P, e_t)]\},$$

where the expectation is defined with respect to a distribution  $F(S_t | S_{t-1}, e) = \prod_i F(S_{it} | S_{it-1}, A, P, e_i)$ . Let  $A^*, P^*, e^*$  denote the optimal decisions (which are measurable functions of  $S_{t-1}$ ) and  $R_i(S_{it}, e_i) := E[r_i(S_{it}, e_i, A^*, P^*) | S_{it}, e_i]$ . By the law of iterated expectations, we can formally express

$$v^{\text{myop}}(S_{t-1}) = E[R_1(S_{1t}, e_1^*) | S_{1t-1}] + \dots + E[R_t(S_{tt}, e_t^*) | S_{tt-1}] \\ = \sum_{i \in \mathbb{I}} \max_{e_i} E[R_i(S_{it}, e_i) | S_{it-1}].$$

The separable individual subproblems  $\max_{e_i} E[R_i(S_{it}, e_i) | S_{it-1}]$  can be solved individually given  $A^*, P^*$ . The advertising and pricing policies can be

addressed as follows: given  $e^*$ , if there exists a sufficient statistics  $\bar{S} = h(S)$  with low dimension such that

$$E\left[\sum_{i \in \mathbb{I}} r_i(S_{it}, A, P, e_i^*) \mid S_t\right] = E\left[\sum_{i \in \mathbb{I}} r_i(S_{it}, A, P, e_i^*) \mid \bar{S}_t\right] \\ := R(\bar{S}_t, A, P),$$

then it also holds that the myopic problem can be written as

$$v^{\text{myop}}(S_{t-1}) = E_{t-1}[R(\bar{S}_t, A^*, P^*)] \\ = \max_{(A, P)} E[R(\bar{S}_t, A, P) | \bar{S}_{t-1}],$$

where the decision variables  $(A, P)$  common to all individuals are the solution of an aggregate subproblem just defined by the state variables  $\bar{S}_t$  (with a significant reduction of the state dimension). In a sequential implementation, the optimal values of each subproblem are updated according to the most recent computed values (those corresponding to the last subproblems solved). In a parallel implementation, these values are updated from a previous iteration. This procedure is repeated until the optimality criteria for the myopic problem are satisfied.

Translating these ideas into a multistep dynamic decision model defined by objective Equations (3) and (4) and the dynamic constraints (2), we can consider

$$r(S_t, A_t, P_t, e_t) := (P_t - c_0)I\bar{S}_t - \sum_{i \in \mathbb{I}} c_i(e_{it}) - c_m(A_t),$$

where  $\bar{S}_t$  is the average sales at time  $t$  and  $\sum_{i \in \mathbb{I}} S_{it} = I\bar{S}_t$  is the total sales. As a consequence,

$$E[r(S_t, A_t, P_t, e_t^*) | S_t, A_t, P_t] \\ = E[r(S_t, A_t, P_t, e_t^*) | \bar{S}_t, A_t, P_t] := R(\bar{S}_t, A_t, P_t),$$

as  $\bar{S}_t$  is a sufficient statistic for  $r(S_t, A_t, P_t, e_t)$ . Then, the general policies  $\{A_t\}, \{P_t\}$  can be determined based on a single-state variable  $\bar{S}_t$  maximizing  $E[R(\bar{S}_t, A_t, P_t)]$ , which is not affected by the course of dimensionality. In addition,  $\bar{S}_t$  satisfies a simple recursive model

$$\bar{S}_t = \rho \bar{S}_{t-1} + \bar{g}(A_{t-1}, P_{t-1}, e_{t-1}) + \bar{\varepsilon}_t,$$

where  $\bar{\varepsilon}_t = I^{-1} \sum_{i \in \mathbb{I}} \varepsilon_{it}$  and  $\bar{g}(A, P, e) = \sum_{i \in \mathbb{I}} g_i(A, P, e_i)/I$ . Notice also that the transition probability for this artificial-state variable  $\bar{S}_t$  is implicitly determined by the model as  $F(\bar{s}' | \bar{s}, A, P, e) = G_t(\bar{s}' - \rho \bar{s} - \bar{g}(A, P, e))$ , where  $G_t$  is the distribution of  $\bar{\varepsilon}_t$  (it can be computed easily as a convolution). Also, the individual policy  $\{e_{it}\}$  can be determined by maximizing  $R_i(S_{it}, e_{it}) = I \cdot E[r_i(S_{it}, A_t^*, P_t^*, e_{it}) | S_{it}, e_{it}]$ , where  $S_{it}$  satisfies (2) given  $A_t^*$  and  $P_t^*$ .

In practice, the only restrictive condition that we will enforce affects the feasible decision set. We will

assume that the feasible set  $\mathbb{A}(S)$ , where the decisions  $(A, P, e)$  are taken at each period of time, is a Cartesian product of subsets  $\mathbf{G}(\bar{S}) \times \mathbf{e}_1(S_1) \times \cdots \times \mathbf{e}_I(S_I)$ , where  $\mathbf{G}(\bar{S})$  is the feasible set for general decisions  $(A, P)$  that can be updated based on previous average sales, and  $\times_{i \in \mathbb{I}} \mathbf{e}_i(S_i)$  is the feasible set for individual decisions, so that

$$\mathbb{A}(S) := \{(A, P, e) : (A, P) \in \mathbf{G}(\bar{S}), e \in \times_{i \in \mathbb{I}} \mathbf{e}_i(S_i)\}. \quad (6)$$

Budgetary constraints cannot be accommodated here. However, we discuss how this assumption can be relaxed in the concluding section.

In summary, we define the following subproblems:

$$\begin{aligned} V_i(s_i) &= \max_{\{e_{it}\}} E_0 \left[ \sum_{t \geq 0} \delta^t R_i(S_{it}, e_{it}) \mid S_{i0} = s_i \right] \quad \text{for all } i, \\ \bar{V}(\bar{s}) &= \max_{\{A_t, P_t\}} E_0 \left[ \sum_{t \geq 0} \delta^t R(\bar{S}_t, A_t, P_t) \mid \bar{S}_0 = \bar{s} \right], \end{aligned} \quad (7)$$

where  $R_i(S_{it}, e_{it})$  and  $R(\bar{S}_t, A_t, P_t)$  are conditional expectations to the information subset at time  $t$ :

$$\begin{aligned} R_i(S_{it}, e_{it}) &= I \cdot E[r_i(S_{it}, A_t^*, P_t^*, e_{it}) \mid S_{it}, e_{it}], \\ R(\bar{S}_t, A_t, P_t) &= E[r(S_t, A_t, P_t, e_t^*) \mid \bar{S}_t, A_t, P_t], \end{aligned} \quad (8)$$

given  $A_t^*, P_t^*, e_t^*$  the optimal decisions for time  $t$ . Subproblems  $\{V_i(s_i)\}_{i \in \mathbb{I}}$  and  $\bar{V}(\bar{s})$  characterize the value function  $V(\cdot)$  of the original SDP problem (5).

**PROPOSITION 3.** *The value functions defined in (7), with respective transition kernels (10), satisfy*

$$V(s) = I^{-1} \sum_{i \in \mathbb{I}} V_i(s_i) \quad \text{and} \quad V(s) = \bar{V}(\bar{s}). \quad (9)$$

**PROOF.** By the law of iterated expectations, the subproblems  $V_i(s_i)$  for all  $i$  and  $\bar{V}(\bar{s})$  satisfy

$$\begin{aligned} E_0 \left[ \sum_{t \geq 0} \sum_{i \in \mathbb{I}} \delta^t r_i(S_{it}, A_t, P_t, e_{it}) \right] \\ = \sum_{i \in \mathbb{I}} E_0 \left[ \sum_{t \geq 0} \delta^t E[r_i(S_{it}, A_t, P_t, e_{it}) \mid S_{it}, e_{it}] \right] \\ = E_0 \left[ \sum_{t \geq 0} \delta^t E \left[ \sum_{i \in \mathbb{I}} r_i(S_{it}, A_t, P_t, e_{it}) \mid \bar{S}_t, A_t, P_t \right] \right], \end{aligned}$$

where  $A_t = A(S_t)$ ,  $P_t = P(S_t)$ , and  $e_t = e(S_t)$ . These equalities imply that  $V(s) = I^{-1} \sum_{i \in \mathbb{I}} V_i(s_i)$ , and  $V(s) = \bar{V}(\bar{s})$  almost everywhere.  $\square$

The advantage of these subproblems is that they are small by construction. For one set of subproblems, the decision variables are only the direct marketing intervention  $\{e_{it}\}_{t \geq 0}$ . Once the solutions for these subproblems have been computed, the price and mass marketing intervention  $\{P_t, A_t\}_{t \geq 0}$  are updated. This

procedure is repeated until the convergence criteria for the problem (5) are satisfied.

To solve the subproblems separately, we need to know  $R_i(S_{it}, e_{it})$  and  $R(\bar{S}_t, A_t, P_t)$  and the transition probabilities for  $\{V_i(s_i)\}$  and  $\bar{V}(\bar{s})$ , respectively, given by

$$\begin{aligned} \mathbb{F}_i(s'_i \mid s_i, e_i) &= E[F_i(s_i \mid S_{it-1}, A_{t-1}^*, P_{t-1}^*, e_{it-1}) \\ &\quad \mid S_{it-1} = s_i, e_{it-1} = e_i] \quad \text{for all } i, \\ \mathbb{F}(\bar{s}' \mid \bar{s}, A, P) &= E[F(\bar{s}' \mid \bar{S}_{t-1}, A_{t-1}, P_{t-1}, e_{t-1}^*) \\ &\quad \mid \bar{S}_{t-1} = \bar{s}, A_{t-1} = A, P_{t-1} = P]. \end{aligned} \quad (10)$$

However, this computation is not feasible because the optimal policy function  $(A^*, P^*, e^*)$  is unknown. To tackle this issue, we combine a decomposition approach with successive substitution iteration on the value functions of each subproblem, where one subproblem is solved using the previous value of another subproblem. Once the solutions for these subproblems have been computed, the decision variables are updated to their last computed values. The procedure is generally continued until the changes made in each iteration are below some tolerance. The main benefit of this approach is that it offers a convergence proof to the actual solution.

#### 4.1. The Algorithm

The general scheme of the algorithm is stated as follows:

##### Algorithm

1. Initialization: Choose a scenario set of states and a starting policy  $\{A^k(\bar{s}), P^k(\bar{s}), e^k(s)\}$  with  $e^k(s) = (e_1^k(s_1), \dots, e_I^k(s_I))$ . Set  $k = 0$ .
2. Repeat:
  - 2.1. Generate recursively  $\{S_t^k, A_t^k, P_t^k, e_t^k\}_{t=1}^T$ , where  $S_t^k$  is drawn from

$$F(s \mid S_{t-1}^k, A^k(\bar{S}_{t-1}^k), P^k(\bar{S}_{t-1}^k), e^k(S_{t-1}^k)),$$

and compute  $\bar{S}_t^k = h(S_t^k)$  (where  $h(\cdot)$  stands for the average),

- 2.2. With the simulated data compute

$$\begin{aligned} R_i^k(S_{it}, e_{it}) &= I \cdot E[r_i(S_{it}, P_t^k, A_t^k, e_{it}) \mid S_{it}, e_{it}], \\ R^k(\bar{S}_t, A_t, P_t) &= E[r(S_t, P_t, A_t, e_t^k) \mid \bar{S}_t, A_t, P_t]. \end{aligned}$$

and the kernels

$$\begin{aligned} \mathbb{F}_i^k(s'_i \mid s_i, e_i) &= \Pr(S_{it}^k \leq s'_i \mid S_{it-1}^k = s_i, e_{it-1} = e_i), \quad i \in \mathbb{I}, \\ \mathbb{F}^k(\bar{s}' \mid \bar{s}, A, P) &= \Pr(\bar{S}_t^k \leq \bar{s}' \mid \bar{S}_{t-1}^k = \bar{s}, A_{t-1} = A, P_{t-1} = P). \end{aligned}$$

- 2.3. Solve the SDP subproblems

$$\max_{\{e_{it} \in \mathbb{A}_i(S_{it})\}_{t \geq 0}} E \left[ \sum_{t \geq 0} \delta^t R_i^k(S_{it}, e_{it}) \mid S_{i0} = s_i \right] := V_i^k(s_i),$$

in  $\{e_{it}^k\}_{t \geq 0}$  for each  $i \in \mathbb{I}$ , where  $\mathbb{A}_i(S_{it-1}) = \{e_i : 0 \leq e_i \leq \bar{e}_i(S_{it-1})\}$ .



## 2.4. Solve the SDP subproblem

$$\max_{((A_t, P_t) \in \bar{A}(\bar{S}_t))_{t \geq 0}} E \left[ \sum_{t \geq 0} \delta^t R^k(\bar{S}_t, A_t, P_t) \mid \bar{S}_0 = \bar{s} \right] := \bar{V}^k(\bar{s}),$$

where

$$\bar{A}(\bar{S}_t) = \{(p, A) : 0 \leq A \leq \bar{A}(\bar{S}_t), 0 \leq P \leq \bar{P}(\bar{S}_t)\}.$$

2.5. Update  $\{e_{it}^k, A_t^k, P_t^k\}$  to  $\{e_{it}^{k+1}, A_t^{k+1}, P_t^{k+1}\}$ , and set  $k \leftarrow k + 1$ .

3. Until convergence: for some tolerance  $\epsilon > 0$ , when the stopping criteria are satisfied

• Criterion 1:

$$\max \left\{ \sup_t \frac{|A_t^{k+1} - A_t^k|}{1 + \|A_t^k\|_\infty}, \sup_t \frac{|P_t^{k+1} - P_t^k|}{1 + \|P_t^k\|_\infty}, \sup_{t,i} \frac{|e_{it}^{k+1} - e_{it}^k|}{1 + \|e_{it}^k\|_\infty} \right\} < \epsilon,$$

• Criterion 2:

$$\sup_{S_0} \left\{ \frac{|I^{-1} \sum_{i \in \mathbb{I}} V_i^{k+1}(S_0) - \bar{V}^{k+1}(\bar{S}_0)|}{1 + \|I^{-1} \sum_{i \in \mathbb{I}} V_i^{k+1}(S_0)\|_\infty} : \bar{S}_0 = I^{-1} \sum_{i \in \mathbb{I}} S_{0i} \right\} < \epsilon,$$

where the superscript  $k$  denotes the current iteration and  $\|\cdot\|_\infty$  is the supremum norm.

The specific details of the algorithm implementation are presented in Appendix B. Note that any classic method for solving the SDP problem such as value iteration or policy iteration can be applied in Steps 2.3 and 2.4 because the subproblems are small problems with just one state variable, using as an initial point the optimal policy computed in the previous iteration of the algorithm. Also, we deal with some of the same difficulties faced by traditional approaches—that is, how to compute the expectation and then solve the resulting optimization problem. In contrast to traditional approaches, we consider the approximation of the expectations via basis functions and combine collocation methods with numerical simulation. As a consequence, the resulting optimization problem is no longer a difficult problem. The convergence of the algorithm is discussed in Appendix C.

Note that the value function for the original problem  $V(S_1, \dots, S_I)$  and the associated policy functions  $[A, P, e](S_1, \dots, S_I)$  cannot be graphically represented for more than two customers because of the dimension. However, graphical figures for these functions would be intuitive and user-friendly tools for marketing managers. Interestingly, our algorithm overcomes this problem by providing useful and visual tools for managers implementing CRM. After convergence of the algorithm at step  $k^*$  to a numerical solution of the original problem, we can depict graphically in the plane the reduced value function  $V^{k^*}(\bar{S})$  and the associated reduced optimal policy functions  $A^{k^*}(\bar{S})$ ,  $P^{k^*}(\bar{S})$  to provide graphical rules for optimal planning of mass advertising and price (provided that the optimal individual  $e$  is implemented). Furthermore, we can depict in the plane the reduced value function  $V_i^{k^*}(S_i)$

and the associated reduced optimal policy function  $e_i^{k^*}(S_i)$  for the  $i$ th customer, providing a graphical rule for optimal planning of the marketing effort on individual  $i$  (provided that the optimal mass advertising and price have been implemented as well as the effort on other individuals).

After convergence, we can compute the classic policy functions  $[A, P, e](S_1, \dots, S_I)$  and value function  $V(S_1, \dots, S_I)$  by implementing one iteration of the classic value method starting from the reduced value function  $V^{k^*}(\bar{S})$ . The norm of the Bellman equation at this iteration provides a validation for the computed solution.

## 5. Some Numerical Simulations

Let us consider a sales dynamic panel model of the form

$$S_{it} = \rho S_{it-1} + \alpha_i + \beta_{1i} e_{it-1} + \beta_{2i} A_{t-1} + \beta_{3i} P_{t-1}^{\beta_{4i}} + \varepsilon_{it},$$

where  $\{\beta_{1i}, \beta_{2i}\} > 0$ ,  $|\rho| < 1$ ,  $\{e_{it}\}_{t \geq 1}$  is the individual marketing effort,  $\{A_t\}_{t \geq 1}$  is the mass marketing effort,  $\{P_t\}_{t \geq 1}$  is the price, and  $\{\varepsilon_{it}\}_{t \geq 1}$  is the independent white noise process  $N(0, \sigma I)$ . We assume that  $\{e_{it}\}_{t \geq 1}$  and  $\{A_t\}_{t \geq 1}$  are given by a cost function  $c(x) = \gamma x^\phi$ , where  $\gamma > 0$ . Then, given  $\delta \in (0, 1)$ , the firm aims to maximize the expected net present value  $E_0[\sum_{t \geq 0} \delta^t r(S_t, A_t, P_t, e_t)]$  with

$$r(S_t, A_t, P_t, e_t) := (P_t - c_0) \sum_{i \in \mathbb{I}} S_{it} - c(A_t) - \sum_{i \in \mathbb{I}} c(e_{it}).$$

We have implemented our decomposition algorithm using MATLAB 7.6 on an Intel Core vPro i7 with machine precision  $10^{-16}$ . The algorithm stops whenever  $\epsilon = 10^{-8}$ .

First, we consider a simplified model in which prices are considered as given, i.e.,  $\beta_{3i} = 0$ , and using a constant exogenous margin  $m_0$  instead of  $(P_t - c_0)$ . For  $m_0 = 50$ ,  $\rho = 0.2$ ,  $\alpha_i = 60$ ,  $\beta_{1i} = 1.2$ ,  $\beta_{2i} = 1.2$ , and  $\sigma = 5$ , Table 1 reports the number of iterations to

**Table 1** Properties of the Algorithm for Different Problem Sizes in a Model Without Prices

Method	No. of customers	Stopping criteria		No. of iterations	Computational time (in seconds)
		Criterion 1	Criterion 2		
Policy iteration	1	0.0000	0.0000	3	3.8922
	5	0.0000	0.0007	4	11.0790
	25	0.0000	0.0008	4	60.5070
	50	0.0000	0.0009	4	166.9300
	100	0.0000	0.0009	4	687.4300
Value iteration	1	0.0000	0.0000	4	3.5335
	5	0.0000	0.0007	3	8.6160
	25	0.0000	0.0008	4	61.1350
	50	0.0000	0.0009	4	169.0700
	100	0.0000	0.0009	3	545.9600

**Table 2** Properties of the Algorithm for Different Problem Sizes in a Model with Prices

Method	No. of customers	Stopping criteria		No. of iterations	Computational time (in seconds)
		Criterion 1	Criterion 2		
Policy iteration	1	0.0000	0.0527	3	5.9819
	5	0.0000	0.0202	4	29.5548
	25	0.0000	0.0202	4	150.4374
	50	0.0000	0.0202	4	404.8369
	100	0.0000	0.0202	2	873.8870
Value iteration	1	0.0000	0.0115	6	11.2360
	5	0.0000	0.0202	2	16.4847
	25	0.0000	0.0202	2	83.7036
	50	0.0000	0.0324	2	189.4250
	100	0.0000	0.0202	2	663.1727

satisfy the stopping criteria and run times (in seconds) until convergence for different numbers of customers  $I$  as well as both policy iteration and value iteration algorithms to solve Steps 2.3 and 2.4 of the algorithm.

Next, we extend the basic model to the general case for which prices are considered a decision variable. For  $c_0 = 50$ ,  $\rho = 0.2$ ,  $\alpha_i = 60$ ,  $\beta_{1i} = 1.2$ ,  $\beta_{2i} = 1.2$ ,  $\beta_{4i} = -0.5$ ,  $\beta_{3i} = 0.5$ , and  $\sigma = 5$ , Table 2 reports the run times (in seconds) until convergence for different numbers of customers  $I$ . The results show that the proposed algorithm is capable of solving the problem with many customers in a reasonable amount of computer time.

These results suggest that the proposed methodology is an effective and useful tool for solving these types of problems because it breaks down a high-dimensional problem into many low-dimensional ones, hence reducing the curse of dimensionality. It is noteworthy that the standard policy iteration approach cannot solve a problem with more than three customers.

Studying the computational time per step, we have found that the steps that require the most computation are those that solve the subproblems (between 70%–75% of the total computational effort). For each subproblem, we consider either a standard value iteration or policy iteration algorithm that updates the value or policy for all states at the same time. In practice, several well-known numerical techniques can be used to improve performance (e.g., exploiting sparsity in matrix operations, exploiting the information about the parametric model to speed the computation of the integrals). Our simulations are based on a standard

implementation to make it clear how the basic method works.

Table 3 reports run times (in seconds) until convergence, obtained from larger problems with a different number of customers. If we use value iteration, the 2,000 customers' problem is solved in 19 hours and 44 minutes, without using parallelization.

To solve larger-scale problems, we must implement the algorithm in a parallelized way. Scalability is the crucial advantage of decomposition operation research methods. One of the most common procedures to do this consists of implementing the code in High Performance Fortran (HPF), an extension of Fortran 90 that supports parallel computing. Some vendors of Fortran compilers incorporate HPF into their products, but most users have now moved to OpenMP, a parallel processor program that works with Fortran, C, and C++. Therefore, at each step of the algorithm all the individual subproblems would be computed in parallel. Clearly, parallelization is an advantage if the code is run on a parallel computer. Although here are many types of parallel computers—for example, those that use superscalar processors (which can issue multiple instructions per cycle from one instruction stream)—there are huge advantages if the code is run on a computer with multicore processor (which can issue multiple instructions per cycle from multiple instruction streams). Sometimes parallelizations are implemented in a network of computers communicated through the Internet (which is known as distributed computing), but we can use a single computer with many networked processors (known as a massively parallel processor).

### 5.1. Validation

This section provides empirical validation for the algorithm, and it shows a comparison with an ADP algorithm. ADP approaches are widely used because approximate value functions can work well in some problems, but if the approximations do not capture the true behavior of the value function, the algorithms can diverge. There is a wide variety of ADP algorithms. For our validation, we have chosen the standard Q-learning approach (see Figure 8.4 in Powell 2007). We validate the quality of the solution using the Bellman operator, which is similar to the Bellman operator  $\Gamma(v)$  defined in (A1). The true solution is a fixed point  $V$  such that

**Table 3** Run Times Until Convergence for Larger Sizes in a Model with Prices

	No. of customers	No. of iterations	Computational time (in seconds)		No. of customers	No. of iterations	Computational time (in seconds)
Policy iteration	250	5	1,456	Value iteration	250	3	1,079
	500	5	4,738		500	3	3,494
	1,000	4	13,852		1,000	5	16,288
	2,000	5	127,077		2,000	3	71,085

$V = \Gamma(V)$ . Therefore, we have studied the performance of both approaches computing the maximum relative error  $\|(V - \Gamma(V))/(1 + V)\|_\infty$ , albeit the 1 in the denominator can be removed if  $V(S)$  is never null. To do so, after our algorithm converges to the limit point  $V^{k^*}(\tilde{S})$ , we iterate a single step of classic value iteration to compute the typical value function  $V(S)$ .

Table 4 reports the numerical performance of both the ADP and proposed approaches for different numbers of customers. As a general overview, we can observe that the ADP approach is reasonable for small problems ( $I = 1$ ), as is our proposed algorithm. However, the ADP approach is an approximation method, so the ADP solution is an optimal approximation of the actual solution (and the percentage error is about 12%). When  $I = 1$ , our approach just solves the actual Bellman equation ( $V = \Gamma(V)$ ) discretized conveniently, as described in Appendix B. The contribution of the proposed approach can be better observed when considering  $I \geq 2$  customers. A standard ADP algorithm can suffer from the curse of dimensionality because it needs to discretize the state and action spaces.

ADP approaches are essentially approximation methods that replace the expectation of the value function in the Bellman equation with some sort of statistical approximation. For the case of one customer, the quality of the  $Q$ -learning solution is worse than the one provided by our algorithm, which in this case it is just a Bellman iteration (there is no decomposition needed). For  $I = 2$ , the ADP relative error  $(V - \Gamma(V))/V$  is locally larger for some states (the worst state renders an error of 36.99 versus 0.15 in our method), which is a direct consequence of the fact that not all states are visited. We have also computed  $\|(V - \Gamma(V))\|_\infty/\|V\|_\infty$ , which is equal to 0.99 for ADP and to 0.15 for our algorithm, showing that globally both methods work reasonably well for this size problem. But when the number of customers grows, the differences between both approaches become critical. Because ADP approaches do not exploit the structure of the problem, when considering five customers the ADP algorithm simply overflows. If we apply our method with  $I = 5$  customers, the relative error is 0.55 (also  $\|(V - \Gamma(V))\|_\infty/\|V\|_\infty = 0.55$ ). This level of accuracy may suffice for some applications, but perhaps not for all of them. A hybrid method could be applied to improve accuracy, alternating a

single iteration of the classic value function and our decomposition method initialized in the output of this iteration. This strategy could be used to reduce the relative error at a generally affordable cost, an issue which we have left for future research.

## 6. An Empirical Application of a Manufacturer of Kitchen Appliances

In this section we provide an empirical application of the method by considering a medium-sized international wholesale company based in Eastern Europe. (For reasons of confidentiality, the name of the company is withheld.) This company distributes and manufactures a large range of built-in electric kitchen appliances (such as cookers, ovens and hobs, cooker and chimney hoods, external motors, microwaves, dishwashers, washing machines, refrigerators, and related accessories). A large share of the company resources is devoted to general marketing activities (mainly advertising and promotions in professional fairs). The company intends to improve the marketing budget allocation on general advertising (fairs and magazines) and on the budget invested in developing sustainable personalized relationship with its clients (for instance, bearing the cost of entertaining and gifts when visiting clients). Here, we have an aggregate variable that comprises all of this one-to-one marketing effort.

We use a monthly client panel with 7,002 observations from the company spanning from January 2005 to December 2008. The panel is unbalanced, but the vast majority of the clients purchase practically every month within the sample period. Because the company sells a range of products with different sales to each client, they aggregated their data, thus providing us with the monthly net profit drawn from each client. The data set consists of 2,180 products sold to 268 different clients located in four different geographical markets: Asia (which represents 4.24% of the sample), English North America (ENA, comprising the United States and Canada; 19.98%), Europe (40.43%), and Latin America and South America (LASA; 35.35%). Except for the issue of location, there are no other differences between these regions.

The data set includes cash flows from each customer, information on general marketing activities investments (such as magazines advertising or exhibitions), and investments in consumer marketing activities at the client level. On average, the general marketing activities investments of the company in our sample is approximately \$1.3 million, and investments in consumer marketing activities are about \$26,404 and are skewed to the right. Both data series are stable around this average, and in general, the company devotes 97% of the promotional budget to general advertising activities and the remaining 3% to individual advertising activities. In general, the company's policy is to

**Table 4** Validation for Our Method, and Comparison with  $Q$ -Learning

Customers	ADP		Our algorithm	
	$\ (V - \Gamma(V))/V\ _\infty$	CPU (seconds)	$\ (V - \Gamma(V))/V\ _\infty$	CPU (seconds)
$I = 1$	0.1772	53.47	0.0315	7.6491
$I = 2$	39.6	97.80	0.15	18.24
$I = 5$	Out of memory	—	0.55	32.77

apply different levels of individual advertising across its clients. The investment in mass and individual advertising does not vary significantly across location, but for those clients located in ENA, the investment on individual advertising is on average four times larger than the sample. The net return is not affected by the location of the company. Because we do not have sociodemographic characteristics of the different customers, we therefore model heterogeneity by using a dynamic panel with individual random effects. Despite its limitations, this database is particularly suitable for studying the interrelationships between profits and mass marketing and individual investments.

Let  $Y_{i,t}$  denote the financial cash flows (returns) obtained from client  $i$  at the period of time  $t$ , let  $e_{i,t}$  denote the individual marketing effort for client  $i$  at the period of time  $t$ , and let  $A_t$  denote the general marketing effort at period of time  $t$ . Then the company aims to maximize the returns function:

$$r(Y_t, A_t, e_t) = \sum_{i=1}^I Y_{it} - A_t - \sum_{i=1}^I e_{it}. \quad (11)$$

Our first step is to define the transition equation for returns. We consider the dynamic panel specification of a basic Markovian model, i.e.,

$$Y_{i,t} = \rho Y_{i,t-1} + \beta_1 \ln A_{t-1} + \beta_2 \ln e_{i,t-1} + (\eta_i + u_{it}),$$

$$E[u_{i,t} X_{i,t}] = 0, \quad E[u_{i,t}] = 0,$$

for all  $i, t$ , where  $|\rho| < 1$ ,  $u_{i,t}$  is white noise and  $\eta_i$  is a zero mean random coefficient accounting for individual heterogeneity in client profitability levels. The noise  $v_{it} = \eta_i + u_{i,t}$  is autocorrelated due to the stability of  $\eta_i$ , and therefore the ordinary least squares and the within-group estimators are both inconsistent (as  $Y_{i,t-1}$  is a regressor). Taking first differences in the model, we eliminate the specific group effects:

$$\Delta Y_{i,t} = \rho \Delta Y_{i,t-1} + \Delta X'_{i,t-1} \beta + \Delta u_{i,t}, \quad t = 2, \dots, T, \quad (12)$$

where  $X'_{i,t-1} = (\ln A_{t-1}, \ln e_{i,t-1})'$ . The errors  $\{\Delta u_{i,t}\}$  are no longer independent but follow a noninvertible MA(1). This equation can be estimated by instrumental variables (IVs), as proposed by Andersen and Hsiao (1982). It is convenient to use lags of the variable in levels  $Y_{i,t-1}$  as an instrument (as well as lags of other exogenous regressors). Nonetheless, the IV estimator is not efficient because only a few moment conditions are used. Arellano and Bond (1991) proposed a generalized method of moments estimator dealing with this problem. Because Arellano and Bond (1991) estimators can perform poorly in certain cases, the method was refined by Blundell and Bond (1998), who include additional moment conditions (building on previous work by Arellano and Bover 1995). The model used in this paper was estimated in STATA using the Blundell–Bond

**Table 5** Estimates of Model (12)

$Y_{i,t-1}$	Coefficient	Std. error	$z$	$P >  z $
$Y_{i,t-1}$	0.038	0.035	10.84	0.00
$A_{t-1}$	1,338.85	66.17	20.23	0.00
$e_{i,t-1}$	327.67	108.59	3.02	0.003

refinement. We consider the Arellano–Bond estimation approach, which considers lagged levels as well as lagged differences as instruments. In particular, we compute the estimates using the command `xtdpdsys` implemented in STATA. Table 5 reports the estimators of model (12).

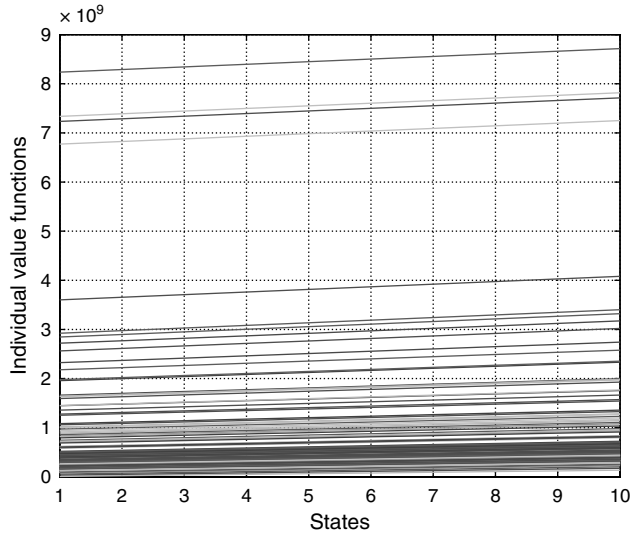
The Wald global significance test is 1,299.38 distributed as a  $\chi^2_3$  with a  $p$ -value of 0.0000. In a model with one lag of the dependent variable,  $k$  strictly exogenous variables, and  $p = T - 2$  periods from which to form moment equations, there are  $k + p \times (p + 1)/2$  moment conditions (see Arellano and Bond 1991). Because the number of clients is  $n = 268$ , there are two strictly exogenous variables  $k = 2$ , and there are panel data of  $T = 48$  periods of time, the number of instrument variable considered here is 1,083.

After the model coefficients have been estimated, because  $T$  is large, we can consistently estimate each specific intercept  $\eta_i$ . For each client we need to take time means on the panel regression equations and then replace  $\sum_{t=1}^T u_{it}/T$  by zero (the expected value), finally getting the estimator of  $\eta_i$ . Then, we compute the values of  $\eta_i$  accounting for individual heterogeneity in client profitability levels as

$$\eta_i = (1 - \rho) \bar{Y}_i - \beta_1 \bar{\ln A} - \beta_2 \bar{\ln e_i},$$

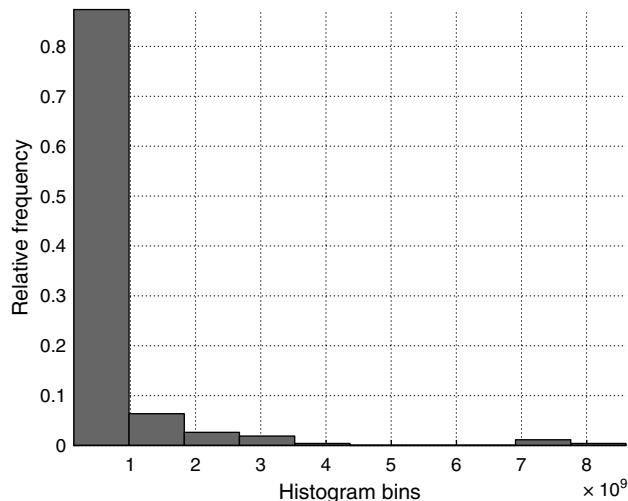
where  $\bar{Y}_i$ ,  $\bar{\ln A}$ , and  $\bar{\ln e_i}$  are the average values of returns, individual marketing effort, and general marketing effort by client  $i$ , respectively. Since this is a vector of 268 coordinates, one per client, these estimates are not reported because of the length. Using this approach, we model the complete heterogeneity of the clients.

Our next step is to compute the optimal general advertising and marketing effort policies. We have applied the proposed decomposition method to the objective (11) and the transition Equation (12) with 268 clients. We consider a state discretization with 10 scenarios (disguised sales levels at company request for confidentiality) for each segment sales variable and 20 equidistant knots for each variable control (policy). We have applied a policy iteration approach to solve each subproblem (although satisfactory results have been found using a value policy iteration approach). The proposed algorithm converges in 17 iterations (about 30 minutes), stopping when the convergence criteria for the global problem is satisfied with  $\epsilon = 10^{-4}$  (and setting the termination tolerance for the subproblems as  $\epsilon = 10^{-3}$ ). The results reveal the efficiency of

**Figure 1** Individual Reduced Value Function (Customer Value Associated with Its Cash Flow States)

the methodology in terms of computing time and accuracy for practical purposes. All figures have been scaled to preserve the nondisclosure agreement for the data. In Figure 1, individual reduced value functions  $\{V_i(s_i)\}$  are shown with respect to the different sales state levels. As a general rule, customers located in Europe have the largest value function in all sales states, whereas customers in LASA and Asia have the smallest expected returns in all sales states. European customers also require more marketing effort per individual than others, whereas Asia requires less marketing effort per individual.

In the estimated model the heterogeneity is confined to the intercept, which is why the reduced value solution are parallel. Figure 2 presents the histogram of intercepts of the individual reduced value functions.

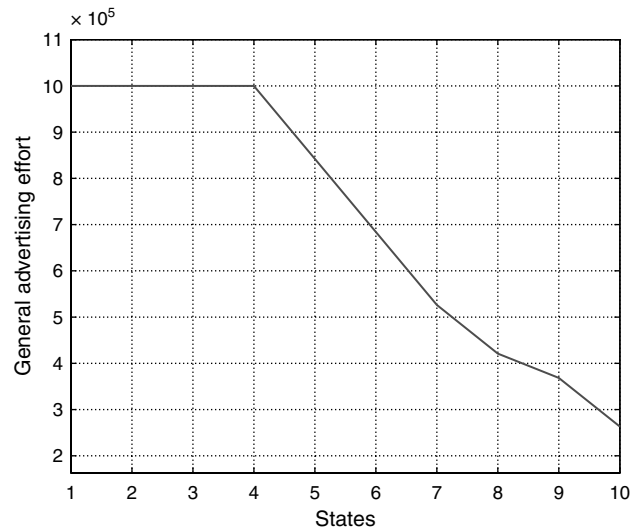
**Figure 2** Distribution of the Individual Reduced Value Function Intercepts

Clearly, each customer brings a different value to the company, which is consistent with the recent literature focusing on one-to-one marketing and the customer lifetime value in a dynamic setting (Lewis 2005, Rust and Verhoef 2005, Khan et al. 2009).

In Figure 3, we show the optimal general marketing policy  $A(\bar{s})$ , demonstrating that the optimal budget allocated to general marketing activities decreases with respect to the states. In contrast, the companies behave optimally if the individual marketing effort essentially remains constant with respect to individual states, although the level of this effort is different for each client—larger in the case of the most profitable clients (located in Europe) and lower for the least profitable clients (located in Asia).

The optimal proportion of effort that the company should devote to the general advertising campaigns and one-to-one marketing is given in Table 6. The optimal decision recommends that a much larger share of communication budget should be invested in individual marketing effort rather than on general advertising, which is not what the company is actually doing. Interestingly, the marketing department is currently pondering this decision (the advisability of moving from general to one-to-one marketing), and our results can help them reformulate their strategy. Notice also that the relative weight of general advertising is reduced when the state variable level is larger.

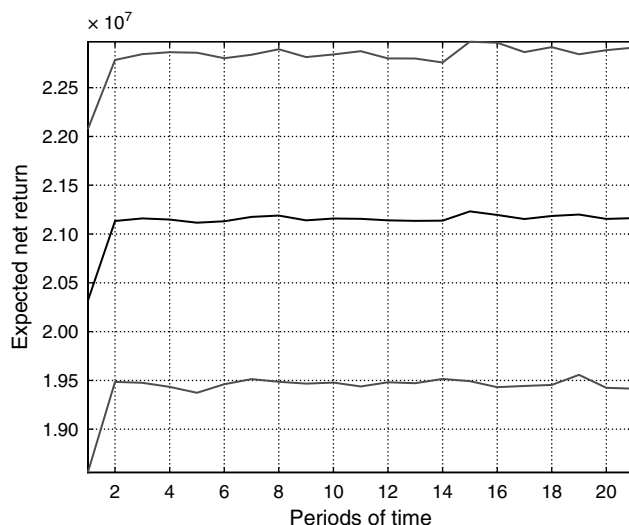
This result contrasts with the current policy of the company (which devotes 97% of the promotional budget to general advertising activities and the remaining 3% to individual advertising activities). Our finding that the company obtains higher expected return by spending more on individual-level marketing relative to mass-level marketing is in line with Khan et al. (2009), who show that customized promotions are

**Figure 3** General Marketing Effort Reduced Policy Function

**Table 6** Distribution of General and Individual Advertising Investment Following the Optimal Policies

Advertising	States									
	1	2	3	4	5	6	7	8	9	10
General (%)	10.35	10.35	10.35	10.35	10.35	9.86	8.86	8.36	7.84	7.32
Individual (%)	89.65	89.65	89.65	89.65	89.65	90.14	91.14	91.64	92.16	92.68

**Figure 4** Expected Returns and Confidence Interval



more profitable than mass promotions. Our optimal allocation also justifies that the company should spend more on one-to-one marketing to boost its long-term customer profitability (Peppers et al. 1999).

Next, based on the optimal policy and the estimated panel model, we have simulated 1,000 paths of optimal cash flow realizations starting at the cash flow for the first period of the sample, and we have computed the expected net return and the 95% confidence interval for 20 periods of time (see Figure 4). The monthly expected returns of the company in our simulation are \$21.1 million with a standard deviation of  $3.31 \times 10^4$ , whereas in the actual data set the average returns are \$20.6 million with a standard deviation of  $6.88 \times 10^6$ . During this period, the company draws an average of a half million dollars less than the expected monthly return based on the optimal SDP policy (but it also has higher variability).

## 7. Conclusions

There is a growing interest among firms in customizing their marketing activities to smaller and smaller units of “individual stores, customers, and transactions” (Bucklin et al. 1998). This implies an enormous number of decisions and is on a scale that requires decision automation tools based on dynamic optimization of small unit panels.

In this paper, we make a computational contribution to solving SDP problems. This allows forward-looking

firms to allocate their marketing budgets by optimizing the CLV of their customer base, simultaneously using customized and mass marketing interventions. The solvability of these models suffers from the curse of dimensionality, and from a modeling standpoint, this limits practitioners. In this sense, we have introduced a novel decomposition methodology for the computation of solutions for large-scale CRM problems. The proposed approach deflates the dimensionality of the models by breaking the problem into a set of smaller independent subproblems. With the numerical results revealing the efficiency of the methodology in terms of computing time and accuracy, it may be concluded that the proposed approach could be applied to many marketing problems with a similar structure.

We have shown that the decomposition method works well in practice, having been successfully applied to assess more than 260 customers of a medium-sized international wholesale company. In addition, we have presented a customer profitability analysis of the company, simultaneously considering the effect of direct marketing and mass marketing interventions at the customer level. We did not have access to specific information about each customer, and our heterogeneity analysis is relatively basic, including all heterogeneity in the individual intercepts. More interesting models can be estimated when this information is available. Note also that the presented framework considers the optimal decision for the firm based on the actual customer base. To extend this framework to a decision-making situation where potential customers could be attracted, we could consider an additional segment or residual customers in which the attractions and churn of new customers could be accounted for. (This segment would be affected by general advertising but not an individual marketing effort.)

Because CRM databases do not often involve panel data across several competitors, no competitive effects have been considered in this paper. To include competition, we should consider a behavioral model for several firms competing for the same customers with mass and customized marketing actions, with the equilibrium being given by the Markov perfect equilibrium (see Dubé et al. 2005). The decomposition algorithm presented in this paper could be a useful tool for addressing the formidable computational effort required to solve this problem. At the moment, however, this a problem best left for future research.

There are other limitations as well as future research opportunities. The key idea of the proposed approach is to split a problem into manageable pieces (subproblems) and coordinate the solutions of these subproblems. Therefore, a crucial assumption seems to be the separability of the feasible decision set into individual and aggregate decisions. The separability of  $\mathbb{A}(S)$  apparently precludes the use of budgetary constraints, such as

$$\mathbb{A}(S) := \left\{ (A, P, e): \sum_{i \in \mathbb{I}} c_i(e_i) + c_m(A) \leq M \right\}, \quad \text{or} \quad (13)$$

$$\mathbb{A}(S) := \left\{ (A, P, e): \sum_{i \in \mathbb{I}} c_i(e_i) + c_m(A) - (P_t - c_0) \sum_{i \in \mathbb{I}} S_i \leq 0 \right\}. \quad (14)$$

However, this limitation can be overcome considering the augmented Lagrangian relaxation approach (see Ruszczyński 1995). The basic idea is to introduce the coupled constraint into the objective function as a barrier term and then to decompose the problem into a set of subproblems, as in the proposed approach here. The separability of the subproblems is obtained by fixing the values of some variables. In particular, for the case of (13), we would consider the augmented Lagrangian objective function:

$$r^a(S_t, \lambda_t, A_t, P_t, e_t) = \sum_{i \in \mathbb{I}} r_i(S_t, A_t, P_t, e_{it}) - \lambda_t \left( \sum_{i \in \mathbb{I}} c_i(e_{it}) - c_m(A_t) - M \right),$$

and we apply our algorithm to the augmented function  $r^a(S_t, \lambda_t, A_t, P_t, e_t)$ , considering  $\lambda_t$  as another control variable (similar to  $A_t, P_t$ ) with a feasibility constraint  $\lambda_t \geq 0$ . We leave the details for future research.

### Acknowledgments

The authors thank Alejandro Balbas and Francisco J. Prieto for their comments that helped improve the manuscript. The authors also thank the editor, associate editor, and three anonymous reviewers for their careful reading of the paper and their valuable comments. Research partly supported by the Ministerio de Ciencia e Innovación (Spain) [Grant ECO2011-30198] and Comunidad de Madrid (Spain) [Grant S0505/TIC-0230].

### Appendix A. Technical Review on Algorithms for SDP Problems

Except for relatively simple cases, the value function cannot be solved analytically, and some numerical algorithms must be implemented. Consider the Bellman operator

$$\Gamma(v) = \max_{(A, P, e) \in \mathbb{A}(s)} \left\{ r(s, A, P, e) + \delta \int v(s') F(ds' | s, A, P, e) \right\} \quad (A1)$$

transforming an arbitrary function  $v(s)$  of the state variables into another function  $\Gamma(v)(s)$ . Obviously, that value function is

a fixed point of  $\Gamma$ , i.e., an element  $V = v^*$  such that  $\Gamma(v^*) = v^*$ . Solving a fixed-point equation is challenging, not just because the unknown is a function but also because  $\Gamma(v)$  involves a maximum.

For a given feasible policy  $(A, P, e)(s)$ , which is not necessarily optimal, we also consider the mapping

$$\Gamma_{(A, P, e)}(v) = \left\{ r(s, (A, P, e)) + \delta \int v(s') F(ds' | s, (A, P, e)) \right\},$$

which is a function of  $s$ . The most commonly used algorithms to compute the value function (and the policy function) work with  $\Gamma(v)$  and  $\Gamma_{(A, P, e)}(v)$ .

- *The value iteration algorithm* iterates  $v_j \leftarrow \Gamma(v_{j-1})$ , where the update is implemented for each state  $s$ , and its convergence is ensured under mild conditions using the fixed-point theorem for contractive mappings (see Appendix C). This was initially proposed by Bellman (1955, 1957) for discrete problems. Unfortunately, the speed can be slow, and when the dimension of the problem is large, it becomes useless.

- *The policy function algorithm* iterates in two steps: (1) policy improvement computing the policy associated with the previous iteration

$$(A, P, e)_j \leftarrow \arg \max_{(A, P, e)} \Gamma_{(A, P, e)}(v_{j-1}),$$

which is computed for each state  $s$ ; and (2) policy valuation computing the value function associated to the previous step  $v_j \leftarrow \Gamma_{(A, P, e)_j}(v_{j-1})$ . Under fairly general conditions, policy iteration can be shown to generate a monotonically improving sequence of functions  $v_j \geq v_{j-1}$ . This method was initially proposed by Howard (1960), and the convergence is studied by Puterman and Brumelle (1979), who also provided an interpretation of this method as a Newton algorithm. It is also possible to consider a hybrid, implementing several steps of value iteration before implementing a policy iteration (see Puterman and Shin 1978). We gain speed using a policy function, but in large problems, the curse of dimensionality takes its toll, and the method also becomes useless.

There is another approach that has been generally considered a theoretical result, but it is increasingly used as a computational algorithm. From the Bellman equation, for any function  $v(s)$  different from the fix point, it holds that  $v(s) \geq \Gamma(v)$ , so we can consider the value function as the solution to the linear program

$$\begin{aligned} \min \quad & \int v(s') F(ds' | s, A, P, e) \\ v(s) \geq & r(s, A, P, e) + \delta \int v(s') F(ds' | s, A, P, e), \\ & \text{for all } s, A, P, e. \end{aligned}$$

Unfortunately, there are an infinite number of constraints when the number of states and feasible controls is infinite. Semi-infinite programming algorithms can handle some of these problems, but the method is not widely used.

Alternatively, some authors work with the first-order conditions associated with the right-hand side of the Bellman equation, leading to an expression called the Euler equation, and try to solve it using the parameterized expectations approach (see Marcet and Marshall 1994). But in general, convergence is not guaranteed.

Clearly, all these algorithms can be directly applied only if the number of states is finite (i.e., the process follows a

Markov chain) and the integral is a summation that can be analytically computed (provided that the number of states is not too high and the summations and maximization decisions for each state can be computed quickly). In the general case with continuous states, however, these algorithms are not feasible and are combined with elements from functional analysis.

Continuous SDP problems are usually solved by combining the ideas of value iteration and policy iteration with collocation methods. The basic idea of collocation methods is to consider a sequence of functions  $\{\phi_k\}_{k \geq 1} \subset B_\infty$  such that any function  $v \in B_\infty$  can be expressed asymptotically as a linear combination of these functions, or more formally, for all  $v \in B_\infty$ ,

$$\inf_{\{\theta_k\}_{k=1}^K} \left\| v(s) - \sum_{k=1}^K \theta_k \phi_k(s) \right\|_\infty \xrightarrow{K \rightarrow \infty} 0.$$

Therefore, we can express  $V(s) \approx \sum_{k=1}^K \theta_k \phi_k(s)$  for some coefficients  $\{\theta_k\}$  and a large enough  $K$ . Several classes of functions can be used for the approximation (e.g., Chebyshev polynomial, splines, neural networks). When the state variable is multidimensional, the base functions are generally obtained by tensor products on univariate basis. The integer  $K$  is exponentially increased with the dimension to obtain a good approximation (this is one type of the curse of dimensionality). Notice that the continuous SDP problem can be approximated by another problem with finite states (only considering a finite partition  $\{\mathcal{S}_k\}$  of the Euclidean state's space  $\mathcal{S}$ , we can approximate  $v$  by simple functions  $\sum_{k=1}^K \theta_k I(s \in \mathcal{S}_k)$ , choosing a representative scenario  $s_k$  for each element of the partition and interpreting  $\theta_k = v(s_k)$ ).

The coefficients  $\{\theta_k\}_{k=1}^K$  are unknown, and the collocation method approximates a functional equation in such a way that the approximated function fits exactly at the prespecified points of the domain. Then, Bellman's equation becomes

$$\sum_{k=1}^K \theta_k \phi_k(s) = \max_{(A, P, e) \in \mathbb{A}(s)} \left\{ r(s, A, P, e) + \delta \sum_{k=1}^K \theta_k \int \phi_k(s') F(ds' | s, A, P, e) \right\}. \quad (\text{A2})$$

Next, we evaluate the linear equation at  $K$  grid points  $\{s_1, \dots, s_K\} \subset \mathcal{S}$  and solve the system in  $\{\theta_k\}_{k=1}^K$ . The system (A2) can be expressed in matrix notation as

$$\Phi \theta = \Gamma(\theta), \quad (\text{A3})$$

where the  $K \times K$  matrix  $\Phi$  has element  $\Phi_{mk} = \phi_k(s_m)$  and the  $K \times 1$  vector  $\Gamma(\theta)$  has the  $m$ th element:

$$\Gamma_m(\theta) = \max_{(A, P, e) \in \mathbb{A}(s_m)} \left\{ r(s_m, A, P, e) + \delta \sum_{k=1}^K \theta_k \int \phi_k(s') F(ds' | s_m, A, P, e) \right\}.$$

The solution of this system is not trivial, and we first need to evaluate the conditional expectations

$$\int \phi_k(s') F(ds' | s_m, A, P, e), \quad (\text{A4})$$

for states  $m = 1, \dots, K$ , often using a numerical integration methods.

The numerical computation of these  $K$  integrals by means of weighted sums requires some type of discretization. There are many procedures to do this: using deterministic Riemann sums, using Monte Carlo approximations based on simulations from  $F(ds' | s_m, A, P, e)$  or refinements such as importance sampling, using other quasi-Monte Carlo methods, or considering weighted deterministic sums based on simple functions as considered by Tauchen (1986); the last is the method that we will apply. In all these methods, the integrals are replaced by a weighted average at discrete points, and the number of required points required to have a good approach increases exponentially with the dimension of the state variables (this the origin of the curse of dimensionality). There is no solution for this problem. Any solution should exploit peculiarities of the model, as we are considering in this paper.

- The *value iteration* method considers the system  $\theta = \Phi^{-1}\Gamma(\theta)$  and iterates the following:

$$\theta \leftarrow \Phi^{-1}\Gamma(\theta)$$

from an initial point  $\theta^0$ .

- The *policy iteration* method uses the Newton iterative updating:

$$\theta \leftarrow \theta - [\Phi - \Gamma'(\theta)]^{-1}[\Phi\theta - \Gamma(\theta)],$$

where  $\Gamma'(\theta)$  is the Jacobian of the collocation function  $\Gamma$  at  $\theta$  that can be computed by applying the *envelope theorem* to the optimization problem in the definition of  $\Gamma(\theta)$  so that

$$\Gamma'_{mj}(\theta) = \delta \int \phi_j(s') F(ds' | s_m, A, P, e).$$

Notice that when the approximation method is based on simple functions, then  $\Phi$  is the identity function, and we can omit this factor. Each time the operator  $\Gamma(\theta)$  is applied, we must solve the maximization problem in  $\Gamma_m(\theta)$  for all states  $s_m \in \{s_1, \dots, s_K\}$ . This can be done in a variety of ways, e.g., using a *global optimization algorithm*. In many applications, the maximization is carried out discretizing the decision space  $\mathbb{A}(s_m)$ . Once we have converged,  $V(s) = \sum_{k=1}^K \theta_k \phi_k(s)$ , and the optimal policy is computed at each state  $s_m \in \{s_1, \dots, s_K\}$ , as the maximizing decision taken at  $\Gamma_m(\theta)$  for the last iteration and the function is computed interpolating these points. The main problem with all the previous techniques is the curse of dimensionality (Bellman 1961). So far, researchers have only been able to solve numerically SDP problems with few state variables. For additional information, see Puterman (1994) and Bertsekas (2005).

There are some variants of value iteration and policy iteration that aim to accelerate the algorithms. A popular one implements a Gauss–Seidel approach, where the update takes into account previous computations of the value function. This is typically implemented when the number of states is countable, or with collocation methods. For example, in a Gauss–Seidel value iteration we consider a sequence of states  $\{s_m\}_{m \in \mathbb{N}}$  and update sequentially all the states  $v_j \leftarrow \Psi(v_{j-1})$ ,



where  $\Psi$  considers the previous updates in the value function. For each  $s_m$ , we compute

$$\Psi(v)(s_m) = \max_{(A, P, e) \in \mathbb{A}(s)} \left\{ r(s_m, A, P, e) + \delta \int_{\{s' < s_m\}} \Psi(v(s')) F(ds' | s_m, A, P, e) + \delta \int_{\{s' \geq s_m\}} v(s') F(ds' | s_m, A, P, e) \right\}.$$

This corresponds to the Gauss–Seidel method to solve the nonlinear system of equations  $\{v(s_m) = \Gamma(v)(s_m)\}_{m \geq 1}$ . Another example is the asynchronous value iteration based on the Gauss–Seidel value iteration, but the states are not updated in order but follow an arbitrary infinite sequence of states such that every state occurs infinitely often. But the method aims to visit all possible states. An alternative option is given by real-time methods, where the sequence of states is generated by simulating the Markovian system, but only the generated states are updated at each stage. At each step  $j$ , if state  $s_m$  is selected, the optimal policy is computed only for that state as the argument maximizing  $r(s_m, A, P, e) + \int v_{j-1}(s') F(ds' | s_m, (A, P, e))$ , and the optimum is used to update the value function at  $s_m$  for all the remainder states  $v_{j+1}(s) = v_j(s)$ . Then another state is generated from the state equation. The algorithm is faster, because the process does not necessarily explore the whole state space, but we do not necessarily obtain a globally defined optimal policy and in general convergence is not guaranteed. Similar ideas hold in the case of policy iteration, when the policy evaluation and policy improvement are only approximately implemented. These methods are generally known as ADP methods, which essentially plays with Monte Carlo simulations to estimate the value function. Instead of updating  $v$  for all possible states  $s$  (as in value iteration), in ADP only a few states  $s_i$  are updated that are drawn from  $F(s' | s_{i-1}, (A, P, e)_{j-1})$ . Given the sampled updates, the value function is estimated using a recursive formulas (analogously to the recursive estimators proposed in the “stochastic approximation” literature). (For a review, see Powell 2007.) In this context, researchers can set decisions to improve the quality of the Monte Carlo estimation (which is called “exploration”) or the quality of the decision (which is called “exploitation”). But as we mentioned before, ADP does not always work well, and its convergence is not generally guaranteed. These algorithms are compatible with our methodology (in the sense that they can be used to solve each subproblem), but we do not emphasize their use.

## Appendix B. Algorithm Implementation

In this section we review the implementation of the algorithm. To understand the details of our approach, first read Appendix A, where we build on these ideas. The key points are the collocation and the discretization to numerically compute the required integrals to apply value iteration or policy function iteration in the subproblems.

**Step 1.** The first step follows the discretization technique. For the most part, we consider a grid of controls,  $\{\mathbf{A}, \mathbf{P}, \mathbf{e}_1, \dots, \mathbf{e}_I\}$ , containing a discretization of the feasible decision set. In particular, we consider relatively large finite intervals for each decision and introduce  $N$  equidistant points for each decision.

**Step 2.1.** The second step is the definition of the scenario nodes and transition probabilities across scenario states. The unconditional distribution can be used to define a grid of representative state values, and the conditional distribution to compute the transition matrix across the elements of the grid. In particular, we consider model (2),  $S_{it} = \rho S_{it-1} + g_i + \varepsilon_{it}$ , where  $\varepsilon_{it} = \eta_i + u_{it} \sim N(0, \sigma_{\varepsilon_i}^2)$  with  $\sigma_{\varepsilon_i}^2 = \sigma_{\eta_i}^2 + \sigma_{u_i}^2$ , and

$$S_{it} | S_{it-1}, A, P, e \sim N(\rho S_{it-1} + g_i, \sigma_{\varepsilon_i}^2),$$

$$\bar{S}_{it} | \bar{S}_{it-1}, A, P, e \sim N\left(\rho \bar{S}_{it-1} + \bar{g}_i, \frac{\sigma_{\varepsilon_i}^2}{I}\right),$$

with  $g_i = g_i(A, P, e_i)$ ,  $\bar{g}_i(A, P, e) = \sum_i g_i(A, P, e_i)/I$ . The stationary marginal distributions of  $S_{it}$  and  $\bar{S}_{it}$  are  $N(g_i(A, P, e_i)/(1 - \rho), \sigma_{\varepsilon_i}^2/(1 - \rho^2))$  and  $N(\bar{g}_i(A, P, e_i)/(1 - \rho), \sigma_{\varepsilon_i}^2/(I \cdot (1 - \rho^2)))$ , respectively. For the  $i$ th customer, we set scenarios in the interval  $[S_i^l, S_i^u]$ , where

$$S_i^l = \min_{A, P, e_i} \left\{ \frac{g_i(A, P, e_i)}{(1 - \rho)} \right\} - 5 \sqrt{\frac{\sigma_{\varepsilon_i}^2}{(1 - \rho^2)}},$$

$$S_i^u = \max_{A, P, e_i} \left\{ \frac{g_i(A, P, e_i)}{(1 - \rho)} \right\} + 5 \sqrt{\frac{\sigma_{\varepsilon_i}^2}{(1 - \rho^2)}}.$$

Therefore, we cover five times the standard deviation from the most extreme mean values. After checking that  $\max\{S_i^l, 0\} < S_i^u$ , we generate  $N$  scenarios distributed uniformly as

$$s_{i1} = \max\{S_i^l, 0\},$$

$$s_{iN} = S_i^u,$$

$$s_{in} = s_{i1} + \frac{s_{iN} - s_{i1}}{N - 1} (n - 1), \quad n = 2, 3, \dots, N - 1.$$

Next we define the product space of states  $\mathcal{S}^I = \prod_{i=1}^I \{s_{i1}, \dots, s_{iN}\}$ . The discrete scenario grid  $\mathcal{S}^I$  can be used to compute the Bellman problem, defining the value functions and the policy functions as mappings defined on  $\mathcal{S}^I$ . Notice that we are considering equidistant points in the grid. If the states variable distribution is highly nonlinear with several modes, it could be convenient to increase the density of points in some locations of special interest. (This is not as relevant with Gaussian data.) The functional base used to approximate the value function is also relevant. Some functional bases provide better approximation using specific points (e.g., wavelets perform better using the dyadic points and Chebyshev polynomials using specific polynomial roots). In our context, based on simple functions, a uniform grid provides a good enough fit.

To implement our algorithm, it is convenient to think of an augmented space of states including mean sales. Consider the mean interval  $[S^l, S^u]$ , with  $S^l = \sum_{i \in I} S_i^l/I$  and  $S^u = \sum_{i \in I} S_i^u/I$ , and generate  $N$  scenarios  $\{\bar{s}_1, \dots, \bar{s}_N\}$  distributed uniformly in  $\max\{S^l, 0\} < S^u$ . Therefore, we can define the augmented space as

$$\mathcal{S}^{I+1} = \left\{ (s, \bar{s}): s = (s_1, \dots, s_I)' \in \mathcal{S}^I, \bar{s} \approx \frac{1}{I} \sum_{i \in I} s_i \right\},$$

where “ $\approx$ ” means that  $\bar{s}$  is the scenario in  $\{\bar{s}_1, \dots, \bar{s}_N\}$  closest to  $\sum_{i \in I} s_i/I$ . Thus, a specific realization of the random vector  $(S_i, \bar{S}_i)$  will be approached by a vector  $(s, \bar{s}) \in \mathcal{S}^{I+1}$ . Given the

structure of the problem, we can define the policy functions  $(A^k, P^k, e^k)$  in the augmented space as a mapping:

$$(A^k, P^k, e^k): \mathcal{S}^{I+1} \ni s \rightarrow (A^k(\bar{s}), P^k(\bar{s}), e_1^k(s_1), \dots, e_I^k(s_I)) \\ \in \{A, P, e_1, \dots, e_I\}.$$

The value function can be approximated in  $\mathcal{S}^{I+1}$  by a simple function:

$$v(s, \bar{s}) = \sum_{n_1, \dots, n_I, n_{I+1}} \theta_{n_1, \dots, n_I, n_{I+1}} \\ \cdot \left\{ \prod_{i=1}^I I(b_{n_{i-1}} < s_i \leq b_{n_i}) \cdot I(b_{n_{I+1}-1} < \bar{s} \leq b_{n_{I+1}}) \right\}.$$

A smooth functional basis could be considered instead of simple functions, e.g., replacing the product of indicator functions by a tensor product of orthonormal polynomials.

*Step 2.2.* To marginalize the effect of some policy controls over the subproblem objective functions (8) and the transition probabilities, we apply the Monte Carlo method. First, given the policy  $(A^k, P^k, e^k)$ , we recursively generate a sample  $\{S_t^k, A_t^k, P_t^k, e_t^k\}_{t=1}^T$  as

$$S_{it}^k = \rho S_{it-1}^k + g_i(A_{t-1}^k, P_{t-1}^k, e_{it-1}^k) + \varepsilon_{it}, \quad i \in \mathbb{I}, \\ \bar{S}_{t-1}^k = I^{-1} \sum_{i \in \mathbb{I}} S_{it}^k,$$

with  $\varepsilon_i \sim N(0, \sigma_{\varepsilon_i}^2 I_T)$  and  $S_{i0}^k = 0$ , and we recursively compute the associated controls as follows:

$$A_t^k = \sum_{n=1}^N A^k(\bar{s}_n) I(b_{n-1} < \bar{S}_{t-1}^k \leq b_n), \\ P_t^k = \sum_{n=1}^N P^k(\bar{s}_n) I(b_{n-1} < \bar{S}_{t-1}^k \leq b_n), \\ e_{it}^k = \sum_{n=1}^N e_i^k(s_{in}) I(b_{i,n-1} < S_{it-1}^k \leq b_{i,n}), \quad i \in \mathbb{I},$$

where  $b_n = (\bar{s}_{n+1} + \bar{s}_n)/2$  and  $b_{i,n-1} = (s_{i,n+1} + s_{i,n})/2$  for  $n = 1, \dots, N-1$ , and we set  $b_0 = b_{i,0} = -\infty$  and  $b_N = b_{i,N} = +\infty$ . The last expressions are used because the policy functions are defined for discrete scenarios—for example, we set  $A_t^k = A^k(\bar{s}_n)$  whenever  $\bar{S}_{t-1}^k \in (b_{n-1}, b_n]$ , which is the interval centered in  $\bar{s}_n$ . For the considered period of study  $T = 1,000$ , we simulate the behavior of sales, mass advertising, direct advertising, and prices for  $T + 100$  times to discard the first 100s, but the figure  $T$  could be increased when the diameter of the feasible decision set or  $N$  increases. The computation of the sample should be done for each iteration, but it only takes a few seconds in all instances.

To properly define the objective function for each subproblem, we compute certain conditional expectations and transition kernels using the simulated sample  $\{S_t^k, A_t^k, P_t^k, e_t^k\}_{t=1}^T$ . First, for all  $i \in \mathbb{I}$  we compute the conditional expectations  $P_{in}^k = E[P_t^k | S_{it}^k = s_{in}]$ ,  $C_{in}^k = E[C_m(A_t^k) | S_{it}^k = s_{in}]$  at the discrete scenarios  $\{s_{in}\}_{n=1}^N$  and  $c_{in}^k = E[c_i(e_{it}^k) | \bar{S}_{it}^k = \bar{s}_n]$  at the scenarios  $\{\bar{s}_n\}_{n=1}^N$ . Then we compute an approximation of the subproblem objective functions (8) evaluated at the discrete scenarios as

$$R_i^k(s_{in}, e_{it}) = I \cdot ((P_{in}^k - c_0) \cdot s_{in} - c_i(e_{it}) - I^{-1} C_{in}^k), \\ R^k(\bar{s}_n, A_t, P_t) = (P_t - c_0) \cdot I \cdot \bar{s}_n - \sum_{i \in \mathbb{I}} c_{in}^k - c_m(A_t).$$

The fastest method to compute the conditional expectations is based on a simple parametric regression model (e.g., specifying  $E[P_t^k | S_{it}^k = s_i] = p(s_i, \beta)$ ). The model is estimated by a least squares method (e.g., minimizing  $\sum_{t=1}^T (P_t^k - p(S_{it}^k, \beta))^2$ ) for direct use (setting  $P_{in}^k = p(s_{in}, \hat{\beta}^k)$ ) for each discrete scenario  $s_{in}$ . The parametric approach works well in our application. Alternatively, we can use a nonparametric estimator. For example, the Nadaraya-Watson estimator of  $E[P_t^k | S_{it}^k = s_{in}]$  is given by

$$E[P_t^k | S_{it}^k = s_{in}] = \frac{\sum_{t=1}^T P_t^k K_{h_T}(S_{it}^k - s_{in})}{\sum_{t=1}^T K_{h_T}(S_{it}^k - s_{in})},$$

where  $K_{h_T}(u) = h_T^{-1} K(u/h_T)$  for an arbitrary kernel density  $K(\cdot)$  (e.g., a standard normal density) and a sequence of positive smoothing parameters  $h_T$  such that  $h_T + (Th_T)^{-1} \rightarrow 0$ . This approach avoids specification assumptions, but it requires larger sample sizes  $T$  than the parametric approach. In addition, an optimal selection of the smoothing parameter is crucial, which is time consuming. However, it might be convenient in some applications.

Next, we compute the marginal transition kernels  $\mathbb{F}^k(s'_i | s_i, e_i)$  and  $\mathbb{F}^k(\bar{s}' | \bar{s}, A, P)$ . There are several possibilities: parametric methods, semiparametric, and nonparametric. The fastest method is based on a parametric regression model,  $E[S_{it}^k | S_{it-1}^k, e_{it-1}^k] = m_i(S_{it-1}^k, e_{it-1}^k, \beta_i)$ ,  $E[\bar{S}_t^k | \bar{S}_{t-1}^k, A_{t-1}^k, P_{t-1}^k] = \bar{m}(\bar{S}_{t-1}^k, e_{it-1}^k, \theta)$ . It estimates the model by a ordinary/nonlinear least squares method. In our applications, we consider this method for a linear parameter model without intercepting where the first regressor is in levels and the controls are in logarithms. Assuming that the errors are conditionally independent of the state variables, we can use the residuals

$$\hat{u}_{it} = S_{it}^k - m_i(S_{it-1}^k, e_{it-1}^k, \hat{\beta}_i), \\ \bar{\hat{u}}_t = \bar{S}_t^k - \bar{m}(\bar{S}_{t-1}^k, A_{t-1}^k, P_{t-1}^k, \hat{\theta}),$$

to estimate the error densities  $g_i(u_{it})$ ,  $\bar{g}(\bar{u}_t)$ . In particular, we have assumed Gaussian distributions  $N(0, \sigma_{u_i}^2)$  and  $N(0, \sigma_{\bar{u}}^2)$ , respectively, and estimating the variances  $\sigma_{u_i}^2$  and  $\sigma_{\bar{u}}^2$  with the mean squared residuals, we get

$$\mathbb{F}_i(s'_i | s_i, e_i) = \frac{1}{\hat{\sigma}_{u_i}} \int_{-\infty}^{s'_i} \phi\left(\frac{z - m_i(s_i, e_i, \hat{\beta}_i)}{\hat{\sigma}_{u_i}}\right) dz \\ = \Phi\left(\frac{s'_i - m_i(s_i, e_i, \hat{\beta}_i)}{\hat{\sigma}_{u_i}}\right), \\ \mathbb{F}(\bar{s}' | \bar{s}, A, P) = \frac{1}{\hat{\sigma}_{\bar{u}}} \int_{-\infty}^{\bar{s}'} \phi\left(\frac{z - \bar{m}(\bar{s}, A, P, \hat{\theta})}{\hat{\sigma}_{\bar{u}}}\right) dz \\ = \Phi\left(\frac{\bar{\hat{u}}_t - (\bar{s}' - \bar{m}(\bar{s}, A, P, \hat{\theta}))}{\hat{\sigma}_{\bar{u}}}\right).$$

Notice that if it is difficult to determine the residuals distribution, we could estimate  $g_i(u_{it})$ ,  $\bar{g}(\bar{u}_t)$  nonparametrically. For example, integrating the Rosenblatt-Parzen kernel density estimator, we obtain a cumulative conditional distribution:

$$\mathbb{F}_i(s'_i | s_i, e_i) \\ = \int_{-\infty}^{s'_i} \left( \frac{1}{T-2} \sum_{t=2}^T K_{h_T}(\hat{u}_{it} - (z - m_i(s_i, e_i, \hat{\beta}_i))) \right) dz,$$

$$\mathbb{F}(\bar{s}' | \bar{s}, A, P) \\ = \int_{-\infty}^{\bar{s}'} \left( \frac{1}{T-2} \sum_{t=2}^T K_{h_T}(\hat{u}_t - (z - \bar{m}(\bar{s}, A, P, \hat{\theta}))) \right) dz,$$

where  $K_{h_T}(u) = h_T^{-1}K(u/h_T)$ . This semiparametric method slows down the algorithm compared with the parametric case. The last alternative is a fully nonparametric estimator such as the cumulated integral of the conditional density estimator formulated by Roussas (1967, 1969) and Chen et al. (2001):

$$\mathbb{F}_i(s'_i | s_i, e_i) \\ = \int_{-\infty}^{s'_i} \frac{\sum_{t=2}^T K_{h_T}(S_{it}^k - z) K_{h_T}(S_{it-1}^k - s_i) K_{h_T}(e_{it-1}^k - e_i)}{\sum_{t=2}^T K_{h_T}(S_{it-1}^k - s_i) K_{h_T}(e_{it-1}^k - e_i)} dz, \\ \mathbb{F}(\bar{s}' | \bar{s}, A, P) \\ = \int_{-\infty}^{\bar{s}'} \left( \sum_{t=2}^T K_{h_T}(\bar{S}_t^k - z) K_{h_T}(\bar{S}_{t-1}^k - \bar{s}) K_{h_T}(A_{t-1}^k - A) K_{h_T}(P_{t-1}^k - P) \right) \\ \cdot \left( \sum_{t=2}^T K_{h_T}(\bar{S}_{t-1}^k - \bar{s}) K_{h_T}(A_{t-1}^k - A) K_{h_T}(P_{t-1}^k - P) \right)^{-1} dz.$$

This method requires very large simulated samples and is sensitive to the selection of the smoothing number that must be optimally determined. In general, we do not recommend it for this algorithm, but it might be useful in some applications.

To apply the collocation method for the Bellman equation associated with each subproblem, we have to integrate the basis functions with respect to  $\mathbb{F}_i(s'_i | s_i, e_i)$  and  $\mathbb{F}(\bar{s}' | \bar{s}, A, P)$ , which requires a numerical integration method. We use Tauchen's method (1986) to approximate the continuous transition kernel  $\mathbb{F}^k(s'_i | s_i, e_i)$  and  $\mathbb{F}^k(\bar{s}' | \bar{s}, A, P)$  by analogous finite-state transition matrix on the states grid  $\{s_1, \dots, s_N\}$ , considering for all  $n, m = 1, \dots, N$ , the transition from  $s_n$  to  $s_m$ :

$$\Pi_{nm}^i(e_i) = \mathbb{F}_i(b_{i,m} | s_{in}, e_i) - \mathbb{F}_i(b_{i,m-1} | s_{in}, e_i), \\ \Pi_{nm}^{\text{mean}}(A, P) = \mathbb{F}(b_m | s_n, A, P) - \mathbb{F}(b_{m-1} | s_n, A, P),$$

where  $b_{i,m} = (s_{i,m+1} + s_{i,m})/2$ ,  $b_m = (\bar{s}_{m+1} + \bar{s}_m)/2$  for  $m = 1, \dots, N-1$ , and we set  $b_{i,0} = b_0 = -\infty$  and  $b_{i,N} = b_N = +\infty$  so that  $\Pi_{n1}^i(A, P, e) = \mathbb{F}_i(b_1 | s_n, e_i)$ ,  $\Pi_{nN}^i(A, P, e) = 1 - \mathbb{F}(b_{N-1} | s_n, A, P)$ , and similarly for  $\Pi_{n1}^{\text{mean}}(A, P)$  and  $\Pi_{nN}^{\text{mean}}(A, P)$ . To apply the collocation value iteration or policy iteration method (for details see Appendix A), the continuous-state expectations of the basis functions (A4) for each subproblem—namely,  $\int \phi_k(s') \mathbb{F}_i(ds' | s_m, e_i)$  and  $\int \phi_k(s') \mathbb{F}(ds' | s_m, A, P)$ —are approximated by the expected values in the analogous discrete Markov chain  $N^{-1} \sum_{n=1}^N \phi_k(s'_n) \Pi_{nm}^i(e_i)$  and  $N^{-1} \sum_{n=1}^N \phi_k(s'_n) \Pi_{nm}^{\text{mean}}(A, P)$ , respectively.

### Appendix C. Convergence Analysis

In this section, we discuss the convergence of the algorithm. Let  $\mathcal{A}$  be the Euclidean set where the decisions are taken. Assume that for each  $s$ , the set  $\mathbb{A}(s) \subset \mathcal{A}$  of admissible controls is a nonempty compact set with Cartesian product structure as described in (6). The admissible state-controls pairs are given by

$$\mathcal{H} := \{(S, A, P, e) : S \in \mathcal{S}, (A, P, e) \in \mathbb{A}(S)\},$$

and as usual, we assume that  $|r(S, A, P, e)|$  is bounded on  $\mathcal{H}$  except for a null probability set. Under these conditions, the convergence of the classical value iteration method is based on central ideas from functional analysis. Consider the Bellman operator  $\Gamma(v)$  defined in (A1). Obviously, that value function is a fixed point of  $\Gamma$ , i.e., an element  $v^*$  such that  $\Gamma(v^*) = v^*$ . The value iteration algorithm considers an arbitrary function  $v_0$ , and recursively computes  $v_j = \Gamma(v_{j-1})$ . Under regularity conditions, the sequence  $\{v_j\}_{j \geq 1}$  converges to a limit that is the value function  $v^*$ .

The argument uses basic concepts of functional analysis. Convergence can be ensured provided that  $\Gamma$  is a contractive operator in a complete metric space. If  $B$  is a complete<sup>2</sup> metric space, an operator  $\Gamma: B \rightarrow B$  is called contractive if  $d(\Gamma(v), \Gamma(v')) \leq cd(v, v')$  for all  $v, v' \in B$  with parameter  $c \in (0, 1)$ . Any contractive operator in a complete metric space has a unique fixed point  $v^*$  and satisfies that  $v^* = \lim_{j \rightarrow \infty} \Gamma^j(v^0)$  for any initial point  $v^0 \in B$ , so that the sequence  $v^j = \Gamma(v^{j-1}) = \Gamma^j(v^0)$  converges to the fixed point, for an introduction see Kolmogorov and Fomin (1970). In particular, we consider the Banach space<sup>3</sup>  $B_\infty$  of bounded and Borel-measurable real-valued functions defined on the Euclidean state's space  $\mathcal{S}$  and endowed with the supremum norm  $\|v\|_\infty = \sup_y |v(y)|$ . If the function  $|r(s, A, P, e)|$  is bounded on  $\mathcal{H}$ , then it is easy to prove that  $\Gamma(v)$  is a contractive operator on  $B_\infty$  with parameter  $\delta \in (0, 1)$ , and the fixed point  $V = \Gamma(V)$  solves the SDP problem,<sup>4</sup> see, e.g., Denardo (1967) and Blackwell (1965). Under stronger conditions on the SDP problem, the value function  $V$  can be proved to be continuous, Lipschitz, once/twice continuously differentiable.

Unfortunately, the implementation of the algorithms is unfeasible with more than three or four state variables because the computation of  $\Gamma(v)$  requires approximation of the numerical integral  $\int v(s') F(ds' | s, A, P, e)$  by an average at selected points, and the number of required points to provide an accurate estimate increases exponentially with the dimension of the state variables.

Next, we discuss the convergence of the presented algorithm. Associated with the value functions defined in (7), we consider the operators:

$$\mathcal{T}_i(V_i, A, P)(s_i) = \max_{\{e_i \in \mathbb{A}_i(s_i)\}} \left\{ R_i(S_{it}, e_{it}) \right. \\ \left. + \delta \int V_i(s'_i) F^{A,P}(s'_i | s_i, e_i) \right\}, \\ \Phi(\bar{V}, e)(\bar{s}) = \max_{\{A, P \in \bar{\mathbb{A}}(\bar{s})\}} \left\{ R(\bar{S}_t, A_t, P_t) \right. \\ \left. + \delta \int \bar{V}(\bar{s}') F^e(d\bar{s}' | \bar{s}, A, P) \right\},$$

where  $F^{A,P}(s'_i | s_i, e_i)$ ,  $F^e(d\bar{s}' | \bar{s}, A, P)$  are defined as in the algorithm Steps 2.1 and 2.3. The arguments that maximize

<sup>2</sup> A metric space,  $B$ , is complete if it is equal to its closure.

<sup>3</sup> A Banach space is a normed linear space, which is complete with respect to the distance  $d(v, v') = \|v - v'\|$  defined from its norm.

<sup>4</sup> There are also extensions for the case where  $r(s, A, P, e)$  is bounded on compact subsets by using other distances (see Rincón-Zapatero and Rodríguez-Palmero 2003).

these two problems are  $\{e_i(s_i)\}_{i=1}^I$  and  $(A(\bar{s}), P(\bar{s}))$ , respectively. The convergence of the decomposition algorithm can be deduced similarly to the proof of convergence of the policy iteration method, using the following arguments:

1. By construction, the solution to the functional equation system,

$$\begin{aligned}\mathcal{T}_i(V_i, A, P)(s_i) &= V_i(s_i), \quad i = 1, \dots, n, \\ \Phi(\bar{V}, e)(\bar{s}) &= \bar{V}(\bar{s}),\end{aligned}$$

satisfies that  $V(s) = I^{-1} \sum_{i=1}^I V_i(s_i, A(\bar{s}), P(\bar{s})) = \bar{V}(\bar{s}, \{e_i(s_i)\})$  a.e., where  $V(s)$  is the value function of the original SDP problem.

2. The algorithm can be considered a recursion defined by a contractive operator. Consider some initial value  $V(s) \in B_\infty$ . Then we can write  $V = I^{-1} \sum_{i=1}^I V_i$  for a vector  $(V_1, \dots, V_I)$  with coordinates  $V_i = \Pi_i V(s)$ , where the operator  $\Pi_i$  is defined as

$$\Pi_i v(s) = E \left[ \sum_{t \geq 0} \delta^t R_{it}(S_{it}, e_{iv}(S_{it})) \mid S_{i0} = s_i \right],$$

$$R_{it}(S_{it}, e_{iv}(S_{it})) = E[I \cdot r_i(S_{it}, e_{iv}(S_{it}), P_v(S_{it}), A_v(S_{it})) \mid S_{it}],$$

and  $A_v(s), P_v(s), e_v(s)$  are the policies rendering the value function  $v(s)$ . These operators satisfy  $\|\Pi_i(v)\|_\infty \leq \|v\|_\infty$ .

The algorithm can be regarded as a sequence obtained by alternating the operators  $(\beta_1, \dots, \beta_I)$  from  $B_\infty \rightarrow B_\infty^I$  defined by  $\beta_i = \mathcal{T}_i \circ \Pi_i V$ , with the operator  $\Phi$ . In other words, it is a recursion defined by the operator  $\Delta = (\Phi \circ I^{-1} \sum_{i=1}^I \beta_i)$  from  $B_\infty \rightarrow B_\infty$ . The operator  $\Delta$  is a contractive operator on  $B_\infty$  since  $\Phi$  and  $\mathcal{T}_i$  are Bellman operators (contractive with parameter  $\delta$ ),

$$\begin{aligned}\|\Delta(v)\|_\infty &= \left\| \Phi \circ \left( \frac{1}{I} \sum_{i=1}^I \beta_i \right)(v) \right\|_\infty \leq \delta \left\| \frac{1}{I} \sum_{i=1}^I \beta_i(v) \right\|_\infty \\ &\leq \delta \frac{1}{I} \sum_{i=1}^I \|\mathcal{T}_i \circ \Pi_i(v)\|_\infty \leq \delta^2 \frac{1}{I} \sum_{i=1}^I \|\Pi_i(v)\|_\infty \leq \delta^2 \|v\|_\infty,\end{aligned}$$

and we can apply a fixed-point theorem to the alternating operator  $\Delta$  to prove convergence to a fixed point satisfying (9).

In the implementation of the algorithm, we use a Monte Carlo simulation and function approximation, and thus we are just considering an approximation of the previous operators. In general, the effect of using Monte Carlo and approximation methods in convergence of value iteration can be addressed using probability tools (see, e.g., Rust 1997, Santos and Rust 2004, Santos and Peralta-Alva 2005, Antos et al. 2008). Discussing this issue would require a more technical paper, and we have not delved into these issues.

## References

- Adelman D, Mersereau AJ (2008) Relaxations of weakly coupled stochastic dynamic programs. *Oper. Res.* 56(3):712–727.
- Anderson TW, Hsiao C (1982) Formulation and estimation of dynamic models using panel data. *J. Econometrics* 18(1):47–82.
- Antos A, Szepesvári C, Munos R (2008) Learning near-optimal policies with Bellman-residual minimization based fitted policy iteration and a single sample path. *Learning Theory* 71(1):89–129.
- Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econom. Stud.* 58(2):277–297.
- Arellano M, Bover O (1995) Another look at the instrumental variable estimation of error-components models. *J. Econometrics* 68(1):29–51.
- Bawa K, Shoemaker RW (1987) The effects of a direct mail coupon on brand choice behavior. *J. Marketing Res.* 24(4):370–376.
- Bellman R (1955) Functional equations in the theory of dynamic programming: Positivity and quasilinearity. *Proc. Natl. Acad. Sci.* 41(10):743–746.
- Bellman R (1957) *Dynamic Programming* (Princeton University Press, Princeton, NJ).
- Bellman R (1961) *Adaptive Control Processes: A Guided Tour* (Princeton University Press, Princeton, NJ).
- Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. *Numer. Math.* 4(1):238–252.
- Berry LL (1995) Relationship marketing in services: Growing interest, emerging perspectives. *J. Acad. Marketing Sci.* 23(4):236–245.
- Bertsekas DP (2005) *Dynamic Programming and Optimal Control*, 3rd ed., Vols. I and II (Athena Scientific, Belmont, MA).
- Bhattacharya CB, Bolton RN (1999) Relationship marketing in mass markets. Sheth JN, Parvatiyar A, eds. *Handbook of Relationship Marketing* (Sage, Thousand Oaks, CA), 327–354.
- Birge JR, Louveaux FV (1997) *Introduction to Stochastic Programming* (Springer, New York).
- Bitran GR, Mondschein SV (1996) Mailing decisions in the catalog sales industry. *Management Sci.* 42(9):1364–1381.
- Blackwell D (1965) Discounted dynamic programming. *Ann. Math. Statist.* 36(1):226–235.
- Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. *J. Econometrics* 87(1):115–143.
- Bose I, Chen X (2009) Quantitative models for direct marketing: A review from systems perspective. *Eur. J. Oper. Res.* 195(1):1–16.
- Bucklin RE, Lehmann DR, Little JDC (1998) From decision support to decision automation: A 2020 vision. *Marketing Lett.* 9(3):235–246.
- Chen X, Linton O, Robinson P (2001) The estimation of conditional densities. Puri ML, ed. *Asymptotic in Statistics and Probability: Papers in Honor of George Gregory Roussas* (VSP International Science Publishers, Leiden, The Netherlands), 71–84.
- Chun BJ, Robinson SM (1995) Scenario analysis via bundle decomposition. *Ann. Oper. Res.* 56(1):39–63.
- Dantzig GB, Wolfe P (1960) Decomposition principle for linear programs. *Oper. Res.* 8(1):101–111.
- Denardo EV (1967) Contraction mappings in the theory underlying dynamic programming. *SIAM Rev.* 9(2):165–177.
- Dubé J-P, Sudhir K, Ching A, Crawford GS, Draganska M, Fox JT, Hartmann C, et al. (2005) Recent advances in structural econometric modeling: Dynamics, product positioning and entry. *Marketing Lett.* 16(3–4):209–224.
- Geoffrion AM (1972) Generalized Benders decomposition. *J. Optim. Theory Appl.* 10(4):237–260.
- Gönül F, Ter Hofstede F (2006) How to compute optimal catalog mailing decisions. *Marketing Sci.* 25(1):65–74.
- Gönül F, Shi MZ (1998) Optimal mailing of catalogs: A new methodology using estimable structural dynamic programming models. *Management Sci.* 44(9):1249–1262.
- Gupta S, Lehmann DR (2003) Customers as assets. *J. Interactive Marketing* 17(1):9–24.
- Gupta S, Lehmann DR (2005) *Managing Customers as Investments: The Strategic Value of Customers in the Long Run* (Wharton School Publishing, Upper Saddle River, NJ).
- Gupta S, Lehmann DR (2006) Customer lifetime value and firm valuation. *J. Relationship Marketing* 5(2/3):87–110.
- Gupta S, Steenburgh TJ (2008) Allocating marketing resources. Working paper, Harvard University, Cambridge, MA. <http://hbswk.hbs.edu/item/5868.html>.

- Gupta S, Zeithaml V (2006) Customer metrics and their impact on financial performance. *Marketing Sci.* 25(6):718–739.
- Gupta S, Lehmann DR, Stuart JA (2004) Valuing customers. *J. Marketing Res.* 41(1):7–18.
- Gupta S, Mela CF, Vidal-Sanz JM (2009) The value of a “free” customer. UC3M Working Paper 09-2-03, University Carlos III de Madrid, Madrid. <http://hdl.handle.net/10016/3883>.
- Heitsch H, Römisch W (2009) Scenario tree modeling for multistage stochastic programs. *Math. Programming* 118(2):371–406.
- Howard RA (1960) *Dynamic Programming and Markov Processes* (MIT Press, Cambridge, MA).
- Kamakura WA, Mela CF, Ansari A, Bodapati A, Fader P, Iyengar R, Naik P, et al. (2005) Choice models and customer relationship management. *Marketing Lett.* 16(3/4):279–291.
- Khan R, Lewis M, Singh V (2009) Dynamic customer management and the value of one-to-one marketing. *Marketing Sci.* 28(6):1063–1079.
- Kolmogorov AN, Fomin SV (1970) *Introductory Real Analysis* (Dover Publications, New York).
- Lewis M (2005) A dynamic programming approach to customer relationship pricing. *Management Sci.* 51(6):986–994.
- Ljungqvist L, Sargent TJ (2004) *Recursive Macroeconomic Theory*, 2nd ed., Vol. 1 (MIT Press, Cambridge, MA).
- Mantrala M (2002) Allocating marketing resources. Weitz BA, Wensley R, eds. *Handbook of Marketing* (Sage, Thousand Oaks, CA), 409–435.
- Marcet A, Marshall DA (1994) Convergence of approximate model solutions to rational expectations equilibria using the method of parameterized expectations. Kellogg Graduate School of Management Working Paper 73, Northwestern University, Evanston, IL.
- McDonald WJ (1998) *Direct Marketing: An Integrated Approach* (Irwin-McGraw-Hill, Boston).
- Montoya R, Netzer O, Jedidi K (2010) Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Marketing Sci.* 29(5):909–924.
- Peppers D, Rogers M, Dorf B (1999) Is your company ready for one-to-one marketing? *Harvard Bus. Rev.* 77(January–February):151–160.
- Powell WB (2007) *Approximate Dynamic Programming: Solving the Curses of Dimensionality* (John Wiley & Sons, Hoboken, NJ).
- Puterman ML (1994) *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (John Wiley & Sons, New York).
- Puterman ML, Brumelle SL (1979) On the convergence of policy iteration in stationary dynamic programming. *Math. Oper. Res.* 4(1):60–69.
- Puterman ML, Shin MC (1978) Modified policy iteration algorithms for discounted Markov decision problems. *Management Sci.* 24(11):1127–1137.
- Reinartz W, Thomas J, Kumar V (2005) Balancing acquisition and retention resources to maximize customer profitability. *J. Marketing* 69(1):63–79.
- Rincón-Zapatero JP, Rodríguez-Palmero C (2003) Existence and uniqueness of solutions to the Bellman equation in the unbounded case. *Econometrica* 71(5):1519–1555. [Corrigendum published in *Econometrica* (2009), 77(1):317–318.]
- Roberts ML, Berger PD (1989) *Direct Marketing Management* (Prentice-Hall, Upper Saddle River, NJ).
- Rossi PE, McCulloch RE, Allenby GM (1996) The value of purchase history data in target marketing. *Marketing Sci.* 15(4):321–340.
- Roussas GG (1967) Nonparametric estimation in Markov processes. *Ann. Inst. Statist. Math.* 21(1):73–87.
- Roussas GG (1969) Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Math. Statist.* 40(4):1386–1400.
- Rust J (1997) Using randomization to break the curse of dimensionality. *Econometrica* 65(3):487–516.
- Rust RT, Chung TS (2006) Marketing models of service and relationships. *Marketing Sci.* 25(6):560–580.
- Rust RT, Verhoef PC (2005) Optimizing the marketing interventions mix in intermediate-term CRM. *Marketing Sci.* 24(3):477–489.
- Rust RT, Lemon KN, Zeithaml VA (2004) Return on marketing: Using customer equity to focus marketing strategy. *J. Marketing* 68(1):109–127.
- Ruszczynski A (1995) On convergence of an augmented Lagrangian decomposition method for sparse convex optimization. *Math. Oper. Res.* 20(3):634–656.
- Santos MS, Peralta-Alva A (2005) Accuracy of simulations for stochastic dynamic models. *Econometrica* 73(6):1939–1976.
- Santos MS, Rust J (2004) Convergence properties of policy iteration. *SIAM J. Control Optim.* 42(6):2094–2115.
- Schmittlein DC, Peterson RA (1994) Customer base analysis: An industrial purchase process application. *Marketing Sci.* 13(1):41–67.
- Shankar V (2008) Strategic allocation of marketing resources: Methods and managerial insights. MSI Working Paper 08-207, Marketing Science Institute, Cambridge, MA. <http://ssrn.com/abstract=1270804>.
- Silverman W (1986) *Density Estimation for Statistics and Data Analysis* (Chapman & Hall, London).
- Simão HP, Day J, George AP, Gifford T, Powell WB, Nienow J (2009) An approximate dynamic programming algorithm for large-scale fleet management: A case application. *Transportation Sci.* 43(2):178–197.
- Simão HP, George A, Powell WB, Gifford T, Nienow J, Day J (2010) Approximate dynamic programming captures fleet operations for Schneider National. *Interfaces* 40(5):1–11.
- Simester DI, Sun P, Tsitsiklis JN (2006) Dynamic catalog mailing policies. *Management Sci.* 52(5):683–696.
- Tauchen G (1986) Finite state Markov-chain approximations to univariate and vector autoregressions. *Econom. Lett.* 20(2):177–181.
- Venkatesan R, Kumar V (2004) A customer lifetime value framework for customer selection and resource allocation strategy. *J. Marketing* 68(4):106–125.