Uncertainty Representation and Risk Management for Direct Segmented Marketing

Victor S. Y. Lo,[a] Dessislava Pachamanova,[b,*] and Nalan Gulpinar[c]

[a]*Head of Data Science and Artificial Intelligence, Workplace Solutions, Fidelity Investments, Boston, MA, USA.*

[b,*]*Corresponding author. Professor, Mathematics and Sciences Division, Babson College, Wellesley, MA, USA. Email: dpachamanova@babson.edu*

[c]*Associate Professor, Operations Group, Warwick Business School, The University of Warwick, Coventry, UK.*

Correspondence should be addressed to Dessislava Pachamanova, Professor, Division of Mathematics and Sciences, 319 Babson Hall, Babson College, Wellesley, MA 02457, USA; Email: **dpachamanova@babson.edu**; Telephone: 781-239-6671.

Dr. Victor S.Y. Lo is a seasoned Big Data, Marketing, and Risk leader with over two decades of consulting and corporate experience employing data-driven solutions in a wide variety of business areas, including Customer Relationship Management, Market Research, Advertising Strategy, Risk Management, Financial Econometrics, Insurance, Transportation, and Human Resources. He is actively engaged with Big Data Analytics, causal inference, and is a pioneer of Uplift/True-lift modeling. Victor has managed teams of analysts in multiple organizations. He is currently Head of Data Science & Artificial Intelligence in Workplace Solutions at Fidelity Investments. Previously he managed analytics teams in Personal Investing, Corporate Treasury, and Healthcare & Total Well-being at Fidelity. Prior to Fidelity, he was VP and Manager of Modeling and Analysis at FleetBoston Financial and Senior Associate at Mercer Management Consulting (now Oliver Wyman). For academic services, Victor has been a visiting research fellow and corporate executive-in-residence at Bentley University. He has been serving on the steering committee of the Boston Chapter of INFORMS and on the editorial board for two journals. He has co-authored a graduate level econometrics book and published numerous articles in Data Mining, Marketing, Statistics, and Management Science literature.

Dr. Dessislava Pachamanova is Professor and Zwerling Family Endowed Research Scholar at Babson College, Wellesley, Massachusetts, USA. Her research and consulting span multiple fields, including financial modeling, robust optimization, predictive analytics, operations, and simulation. She has authored and coauthored tens of articles in operations research, finance, accounting, engineering, marketing and management journals, numerous book chapters, as well as three books in computational finance and risk management. Dr. Pachamanova holds an A.B. from Princeton University and a Ph.D. from the Sloan School of Management at the Massachusetts Institute of Technology.

Dr. Nalan Gulpinar is an Associate Professor in Operational Research and Management Sciences at Warwick Business School, The University of Warwick, the UK. She received BSc and MSc degrees in Mathematics and a PhD in Operational Research (OR). Before joining Warwick, she was first a research fellow, then a research lecturer in the Department of Computing at Imperial College London. Her research projects have been of multi-disciplinary nature owing to the different broad contexts in which mathematical modeling and operational research techniques are applied to develop robust and optimal decisions. The main focus has been decision-making under uncertainty and risk management with applications to finance, healthcare, and supply chain management.

Uncertainty Representation and Risk Management for Direct Segmented Marketing

Abstract: Mining for truly responsive customers has become an integral part of customer portfolio management, and, combined with operational tactics to reach these customers, requires an integrated approach to meeting customer needs that often involves the application of concepts from traditionally distinct fields: marketing, statistics, and operations research. This article brings such concepts together to address customer value and revenue maximization as well as risk minimization for direct marketing decision making problems under uncertainty. We focus on customer lift optimization given the uncertainty associated with lift estimation models, and develop risk management and operational tools for the multiple treatment (recommendation) problem using stochastic and robust optimization techniques. Results from numerical experiments are presented to illustrate the effect of incorporating uncertainty on the performance of recommendation models.

Summary Statement of Contribution: This paper discusses the concept of lift in the context of revenue management for marketing campaigns, and introduces a risk management framework based on methods from the predictive analytics, stochastic programming, and robust optimization fields. The framework can be used to mitigate errors in the customer engagement process and to reduce volatility in the revenues realized from marketing efforts.

Keywords: Uplift modeling, lift, risk management, marketing revenue management, targeted offers, estimation uncertainty, stochastic programming, robust optimization

1. Introduction

The dynamic nature of customer needs, driven by rapid technological innovations and competition, has created new market risks and increased the need to enhance the delivery of customer value through integration of marketing with other company activities such as revenue management (Lummus et al. 2003, Juttner et al. 2007). Marketers are held accountable for how marketing investments perform (Bick 2009, Stewart 2009, Ryals et al. 2007, McDonald 2006), and today marketing activities aim not only for rigorous understanding of customer needs, choices and behavior towards to certain products and services, but also for combining that understanding with operational excellence in fulfilling these needs.

Solving the marketing campaign problem (MCP) is one aspect of this phenomenon. (For examples of different variations of the MCP formulation, see Asllani and Halstead (2011), Beltran-Royo et al. (2016), and Deza et al. (2015).) Given N potential customers and M campaigns (also referred to as *treatments* or *service recommendations*), the goal in the MCP is to determine the optimal assignment of treatments to customers so as to maximize the value realized from the company's customer portfolio, maximizing sales in the process, and typically with a limited marketing budget. The goal of maximizing customer value translates into the goal of maximizing shareholder value (Ryals et al. 2007), which is different from the goal of maximizing market share or customer satisfaction (Lukas et al. 2005). However, it can be argued that maximizing customer value is the goal that best maximizes the value of a firm's stakeholders (Doyle 2000, Ryals et al. 2007).

To justify marketing spend, treatments should be assigned to customers based on a customer score that evaluates the likelihood that a customer n will respond positively to a treatment m. This score is referred also as the *lift* and is estimated using statistical procedures. The MCP can be stated as an optimization problem in which the objective is to maximize the total lift, subject to budget constraints. Such a problem formulation results in a recommendation for the optimal assignment and falls into the category of prescriptive analytics models. (See, for example, LaValle et al. 2011 for an overview of how companies employ predictive and prescriptive analytics technique to realize value.) Note that the objective of maximizing lift is different from the objective of targeting customers who have a high likelihood of purchasing the product, as some customers may purchase the product without receiving a treatment, thus wasting marketing efforts and spend. Some customers may also be likely to buy the product but change their mind if contacted. A strong case for the importance of estimating the lift as opposed to simply using the likelihood of response has been made both in the literature and in practice (see, for example, Lo (2002), Lo (2008), Kane et al. (2014), Siegel 2011 and Chapter 7 in Siegel 2013).

Although assigning the optimal campaigns to individual customers is the goal, in many important practical applications, marketing campaign optimization is done in stages. In the first stage, customer segments are determined based on the available data and the specific application. (We will use "segments" and "clusters" interchangeably in this article because segments are often estimated using cluster analysis.) Common customer segmentation models utilize recency, frequency, and monetary value (RFM) variables, contact and response histories, or demographic attributes. In the second stage, marketing campaigns are designed for the different segments, and the same treatment is assigned to all customers in a customer segment, implicitly assuming that customers behave statistically homogeneously within a segment (Bitran and Mondschein 1996, Storey and Cohen 2002, Bertsimas and Mersereau 2006, Simester et al. 2006, Ryals et al. 2007). Examples of natural applications of customer segmentation are situations in which the information about individual customers is limited, such as organic search in the context of Internet marketing (that is, customers coming from a search engine such as Google or Bing). Even in situations in which detailed information about individual customers is available, the marketing campaigns are often designed to target customer segments. This is because it may be too expensive to reach individual customers, or because the mode of communication with customers (e.g., a mass mailing or a TV ad) may require it. In this paper, we consider a segmentation-based version of the MCP and address practical issues that arise in the application of the optimization methodology for assignment of treatments to segments. (We note that in the limit, a segment could be an individual customer, so our discussion extends to the individual case.)

The relationship between marketing spend and the returns from a company's customer portfolio is volatile, and "…[c]ompanies can be profitable in an accounting sense and yet still destroy shareholder value, because risk has not been adequately taken into account." (Ryals et al.

2007). Some researchers have explored the issue of risk in realized customer cash flows (e.g., Tarisi et al. 2011); some have considered risk attitudes of an individual marketing manager (Brockhaus 1980); and some have looked at organizational risk attitudes (Pennings and Wansink 2004). However, the issue of incorporating risk in marketing spend decisions should be receiving more attention (Ryals et al. 2007).

This article looks at the issue of risk management in the context of the MCP, but from a statistical perspective. Statistical risk measures can be translated to risk preferences (Artzner et al. 1999, Szegö 2002). Specifically, we address the problem of the uncertainty in estimated lift values, which can lead to dramatically different optimal treatment assignment strategies. The uncertainty in lift values has multiple sources: random errors arising from estimating lifts using statistical procedures and limited cluster sizes; errors due to changes in the populations of the different clusters; and errors due to systematic changes in the economy or customer behavior (Lo and Pachamanova 2015). We focus on errors arising from statistical procedures and propose a comprehensive framework for incorporating estimation risk in prescriptive models for treatment assignments, mapping models of uncertainty to risk measures and evaluating the characteristics of the resulting assignments. We bring together methodologies for incorporating uncertainty in prescriptive models from the fields of stochastic programming and robust optimization and show how they apply in the context of a practical solution to the MCP under uncertainty.

Although this paper focuses on the MCP in terms of maximizing lift, the insights from our work carry over to the problem of optimal revenue management for marketing campaigns; see, for example, Gubela et al. (2017). Specifically, expected revenue can be expressed as a multiple of lift, and the analysis of frameworks for including lift uncertainty is directly applicable to the analysis of revenue uncertainty for marketing campaigns. As we will show, minimizing estimation uncertainty maps into risk minimization for customer revenues under different risk measures, placing our work in the realm of revenue maximization and risk minimization for marketing spend. Perhaps the closest previous work on this topic is Ryals et al. (2007), who introduce a framework for marketing spend revenue maximization and risk minimization based on Modern Portfolio Theory from the finance literature. Our approach, however, is philosophically different, and has roots in the control, statistical, and operations research literature rather than the financial literature. The problem of revenue maximization for marketing campaigns itself is part of the larger problem of revenue management for organizations, and goes hand-in-hand with issues in operations and supply chain management (Rhee and Mehra 2006, Hildebrandt and Wagner 2000, Shah et al. 2013, Curcuru 2011).

The paper is organized as follows. In Section 2, we discuss the idea behind and the estimation of lift. In Section 3, we describe the segmentation-based MCP. Section 4 reviews approaches from stochastic programming and robust optimization that can be used to incorporate uncertainty in the marketing campaign problem. Section 5 introduces models for the uncertainty in lifts. Section 6 presents the results of computational experiments that study the characteristics of the optimal assignments of offers to customers using the different approaches and evaluate their performance statistically. Section 7 summarizes our framework, explains how it can be implemented in practice, and interprets the results from the computational experiments in that context. Section 8 concludes with managerial implications.

2. Lift Estimation

Predictive modeling is often applied at the individual customer level to understand the characteristics of customers who are likely to respond to a marketing campaign. Traditional response models are designed to identify likely *responders* regardless of whether these customers are targeted. Such models are based on statistical estimation and machine learning algorithms, such as logistic regression and decision trees. They typically assign a score, or a probability of response, to a customer, with high scores indicating high likelihood of response. (See, for example, Shmueli et al. 2016.) The problem with traditional response models is that they do not measure the incremental impact of a marketing campaign on customer response, and it is ultimately the incremental impact on customer response that should enter the calculation of the effect of marketing spend on revenue.

Finding customers whose decisions will be *positively influenced* by a marketing campaign has been referred to as *uplift modeling* (Radcliffe and Surry 1999), *true lift* (Lo 2002) and *net lift* (Lund 2012, Kubiak 2012). The idea is to differentiate between four different types of customers, thus making the most efficient use of marketing budgets. The four types of customers are (Figure 1):

- *Sure things*: Those customers would purchase regardless of whether they are targeted;
- *Lost causes*: Those customers would not purchase regardless of whether they are targeted;
- *Do-not-disturbs*: Those customers have a negative reaction to the marketing contact and will not purchase if targeted, although they would have purchased had they not been targeted;
- *Persuadables*: Those customers purchase only if contacted. They are the only efficient target.

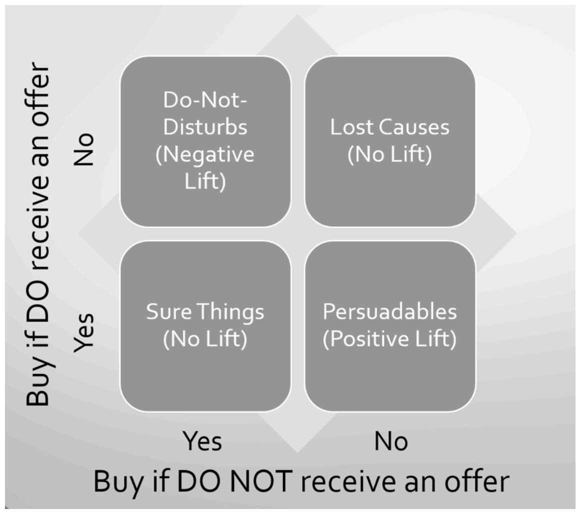


Figure 1. Four types of customer groups. (Based on Siegel 2011 and Radcliffe 2007.)

To estimate lift, customers are split into two groups: treatment (*T*) and control (*C*). A common approach is then to fit separate response models to each group, and determine lift for a customer as the difference between the response score obtained from the two separate models. A response model could be, for example, a logistic regression model with customer response as the target (output) variable. If $\hat{p}(R|T)$ is the estimated probability of response (R) among the treated population, and $\hat{p}(R|C)$ is the estimated probability of response among the control group, then the lift $\hat{\pi}_i$ for a specific customer *i* is estimated as

$$\hat{\pi}_i = \hat{p}_i(R|T) - \hat{p}_i(R|C)$$

This model is referred to as the *Two Model Approach*. When there are multiple treatments, a separate response model is required for each treatment group, and the lift for each treatment and each customer is estimated as the difference between the corresponding treatment score and the control group score.

An alternative approach, the *Treatment Dummy Approach*, is to fit a single response model but introduce dummy variables and capture lift through terms that interact with the treatment dummy variables. The lift is then estimated by calculating the model score assuming the treatment dummy variable equals 1, and subtracting from it the model score calculated after assuming the treatment dummy variable equals 0. Specifically, if the probability of response $\hat{p}_i(R)$ for customer i, calculated from a logistic regression model, is a function of the treatment dummy variable, T_i, and a vector of independent variables or predictors X_i,

$$\hat{p}_i(R) = \frac{e^{\alpha+\boldsymbol{\beta}'\boldsymbol{X}_i+\gamma T_i+\boldsymbol{\delta}'\boldsymbol{X}_i T_i}}{1+e^{\alpha+\boldsymbol{\beta}'\boldsymbol{X}_i+\gamma T_i+\boldsymbol{\delta}'\boldsymbol{X}_i T_i}}$$

then the lift is estimated by subtracting the score estimated when the treatment variable T_i is set to 0 from the score estimated when the treatment variable T_i is set to 1:

$$\hat{\pi}_i = \hat{p}_i(R|T=1) - \hat{p}_i(R|T=0)$$

$$= \frac{e^{\alpha+\boldsymbol{\beta}'\boldsymbol{X}_i+\gamma T_i+\boldsymbol{\delta}'\boldsymbol{X}_i T_i}}{1+e^{\alpha+\boldsymbol{\beta}'\boldsymbol{X}_i+\gamma T_i+\boldsymbol{\delta}'\boldsymbol{X}_i T_i}} - \frac{e^{\alpha+\boldsymbol{\beta}'\boldsymbol{X}_i}}{1+e^{\alpha+\boldsymbol{\beta}'\boldsymbol{X}_i}}$$

Here α is a scalar, $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are vectors of coefficients of the appropriate dimensions, and γ is a scalar.

There are several other methods for estimating lift in practice. Interested readers are referred to Zhao et al. (2017) for a comprehensive review.

Compared to estimating a response score based on a single model, estimating the lift may introduce a higher degree of variability, because the variance of the lift estimate is the sum of the variances of the treatment response and the control response (since the treatment and control random samples are independent), approximately doubling the variance of the estimate relative to the traditional single response estimate; see comments from Athey and Imbens (2015) and Medvedev (2016). Specifically, the coefficient of variation (CV) of the lift estimate, defined as the standard deviation divided by the mean, could be much larger than the CV of the treatment or control response rate. This is because the mean difference between treatment and control response rates (the denominator of the CV for the lift estimate) tends to be much smaller than the treatment or control response rate itself, which, together with the higher standard deviation (the numerator of the CV for the lift estimate), results in a higher CV for the lift estimate. It is important to recognize the presence of increased estimate variability when using the lift estimate for decision making, and hence imperative to incorporate considerations for uncertainty in the treatment allocation models.

3. The Segmentation-Based MCP: Problem Formulation

Consider M different campaigns (treatments, service recommendations) that are designed to target K customer segments, or clusters. Let N_k represent the number of individuals within each cluster k, $k = 1,\ldots,K$. Each campaign m has a cost c_m and campaign budget B_m. Let η_{km} denote the number of individuals in cluster k to receive treatment m, and c_{km} be the cost of treatment m for each individual in cluster k. For simplicity and as is often the case in practice, we assume that the cost is not individual but segment-specific.

As we explained in Section 2, lifts are typically estimated on the individual customer level. Let $\tilde{\pi}_{nm}$ denote the lift from targeting customer n with campaign m. We use the tilde symbol to represent the uncertainty associated with estimating the lift using statistical procedures. Assume that representative cluster-level lift scores $(\tilde{\phi}_{k1}, \ldots, \tilde{\phi}_{kM})$ are calculated for each cluster $k = 1, \ldots, K$, e.g., by averaging the individual customer lifts in that cluster. Given the uncertainty in the individual lift scores, the cluster-level lift scores $(\tilde{\phi}_{k1}, \ldots, \tilde{\phi}_{kM})$ are uncertain as well.

To formulate the cluster-based MCP (CMCP) mathematically, we take into consideration various operational constraints, business constraints, and customer contact policy restrictions that need to be satisfied by a feasible assignment. The aggregate business constraints determine campaign and communication budgets, channel capacity limits, and minimum or maximum cell size.

- The campaign budget B_m limits the cost incurred for each campaign m :

$$\sum_{k=1}^{K} c_{km}\eta_{km} \leq B_m, \qquad m = 1, \dots, M$$

- Sometimes, there is a restriction that the total cost over all campaigns cannot exceed a global budget B:

$$\sum_{k=1}^{K}\sum_{m=1}^{M} c_{km}\eta_{km} \leq B$$

- Often, there is a constraint on the number of customers who could be offered a treatment:

$$\sum_{k=1}^{K} \eta_{km} \leq C_m, \qquad m = 1, \dots, M$$

In summary, the cluster-based MCP model maximizes the sum of the aggregate cluster lifts subject to various business constraints. It can be formulated as a single stage stochastic linear program, which we will refer to as the cluster marketing campaign problem (CMCP):

(CMCP)

$$\max_{\boldsymbol{\eta}} \sum_{k=1}^{K}\sum_{m=1}^{M} \tilde{\phi}_{km}\, \eta_{km}$$

s.t.

$$\sum_{k=1}^{K}\sum_{m=1}^{M} c_{km}\eta_{km} \leq B \qquad \text{(Budget Constraint)}$$

$$\sum_{m=1}^{M} \eta_{km} \leq N_k, \qquad k = 1, \dots, K \quad \text{(Number of individuals within each cluster)}$$

$$\sum_{k=1}^{K} \eta_{km} \leq C_m, \qquad m = 1, \dots, M \quad \text{(Number of individuals receiving each treatment)}$$

$$\eta_{km} \geq 0, \qquad k = 1, \dots, K, \quad m = 1, \dots, M.$$

The decision variables η_{km} here will be treated as non-negative real numbers that can be rounded up to approximate integer values. Thus, this model can be solved using linear programming software. Let us denote by ב the set to which the values of η_{km} that satisfy the constraints in the formulation above belong. In the next section, we suggest approaches for accounting for the uncertainty in the cluster-level lift estimates $\tilde{\phi}_{km}$.

As mentioned earlier, the problem of maximizing expected total lift is directly related to the problem of expected marketing campaign revenue maximization. Specifically, if the expected lift for cluster k and treatment m is $\hat{\phi}_{km}$, the expected revenue for that cluster is $\hat{r}_{km}\hat{\phi}_{km}\eta_{km}$, where $\hat{r}_{km}$ is the expected revenue for treatment *m* per customer in cluster *k*.

4. Incorporating Parameter Uncertainty into the MCP

The quality of the optimal solution in an optimization problem is highly dependent on the quality of the inputs to the optimization problem. In the particular applications we are discussing in this article, we are using model estimates of the lifts $\hat{\pi}_{ij}$ to calculate cluster-level lifts $\hat{\phi}_{km}$, which are a product of statistical estimation and are not necessarily accurate. Solving optimization problems with parameter uncertainty has long been a subject of research in engineering (Du and Chen 2000, Ben-Tal and Nemirovski 2001, Fang and Li 2009) and financial (Garlappi et al. 2006, Cont et al. 2010, Fabozzi et al. 2007, 2010, Ibragimov et al. 2015) applications; however, it has not received the same level of attention in the marketing literature. In this section, we outline two main approaches for dealing with parameter uncertainty.

When different scenarios can be generated for the inputs to the problem, one can employ *stochastic programming* techniques. Stochastic programming has been around for decades (Dantzig 1955, Wallace and Ziemba 2005, Shapiro et al. 2009, Birge and Louveaux 2011, King and Wallace 2012). It solves the optimization problem by optimizing an objective function that is a statistic calculated over the scenarios (such as an average or a given quantile) and may contain other terms and constraints, such as a probabilistic constraint that requires that the optimal solution satisfies a given condition in a particular percentage of the scenarios.

Instead of only scenarios, one can consider more general uncertainty sets for the input parameters. A branch of optimization under uncertainty that solves for the optimal solution assuming that the uncertain parameters can take any value within the prespecified uncertainty sets is *robust optimization*. There are multiple ways to formulate the problem and different choices of uncertainty sets (Ben-Tal and Nemirovski 1998, Tutuncu and Koenig 2004, Ben-Tal et al. 2009, Bertsimas et al. 2011, Wiesemann et al. 2014, Bertsimas et al. 2018). Robust optimization for certain types of problems and uncertainty sets reduces to solving a new version of the original problem, called the robust counterpart, in which the values of the uncertain parameters are replaced with parameters from the uncertainty set formulation. Robust optimization has some overlap with stochastic programming but has historically evolved separately as a field.

Let us consider a nominal formulation of the (CMCP), which we will refer to as (CMCP-N), with the following objective:

(CMCP-N)

$$\max_{\eta \in \beth} \sum_{k=1}^{K} \sum_{m=1}^{M} \hat{\phi}_{km} \, \eta_{km}$$

where $\hat{\phi}_{km}$ are some expected (nominal) values of the cluster-level lifts. (We will explain how such estimates can be calculated in Section 5.) The (CMCP-N) formulation will serve as a benchmark for the performance comparison of the other formulations under uncertainty.

Next, we illustrate the application of stochastic programming and robust optimization approaches to problem formulation (CMCP) in more detail. Specifically, we show examples of how uncertainty can be represented, and discuss mapping the different methodologies for optimization under uncertainty to risk measures for the targeted offers problem.

4.1. Stochastic Programming

The conventional approach to decision making under uncertainty is based on expected value optimization. This requires a representation of uncertainty that is expressed in terms of a multivariate continuous distribution. The underlying decision model can then be generated with internal sampling or a discrete approximation of the distribution. For instance, the expected value of the lift scores using a probability distribution for the lift scores, $E[\tilde{\phi}_{km}]$, needs to be estimated. For a discrete number of future realizations of uncertain parameters with known probabilities, the decision-making problem is described as a scenario-based stochastic programming problem.

Given a set S of discrete scenarios $s \in S$ with corresponding probabilities of occurring q_s, the stochastic campaign optimization model maximizing the expected value of total response rates subject to various business constraints and customer contact policy restrictions can be formulated as:

(SCMCP-EV):

$$\max_{\eta \in \beth} \sum_{s \in S} q_s \sum_{k,m} \phi_{km}^{s} \eta_{km}^{s}$$

If we calculate the nominal estimates $\hat{\phi}_{km}$ as expected values, the expected value stochastic programming formulation (SCMCP-EV) is equivalent to the nominal formulation (CMCP-N). A variety of modifications can be implemented to control the effect of uncertainty in the lift score estimates. Such modifications include chance constraint formulations in which a percentile of the possible distribution (rather than the expected value) for the uncertain expression $\sum_{k=1}^{K} \sum_{m=1}^{M} \tilde{\phi}_{km}\, \eta_{km}$ is maximized; maximum regret formulations in which the maximum regret (the worst-case regret) is minimized, and worst-case scenario formulations in which the worst case realization of the uncertain expression over the set of scenarios is explicitly taken into consideration (Shapiro et al. 2009).

As an example, let us consider a maximum regret formulation. A maximum regret optimization approach finds a feasible solution minimizing (over all scenarios) the maximum deviation of the value of the solution from the optimal value of the corresponding scenario.

Let z^s be the optimal value of objective function for each scenario $s \in S$. The maximum lift for a specific scenario $s \in S$ would be computed as

$$z^s = \max_{\eta \in \beth} \sum_{k,m} \phi_{km}^s \eta_{km}^s$$

The regret of a global solution η_{km} over a specific scenario $s \in S$ is defined as the difference between z^s and $\sum_{k,m} \phi_{km}^s \eta_{km}$, or $\left(z^s - \sum_{k,m} \phi_{km}^s \eta_{km} \right)$.

The maximum regret $R_{max}(\eta)$ of a solution η_{km} can be formulated as the following optimization problem over all possible scenarios $s \in S$:

$$R_{max}(\eta) = \max_{s \in S} \left\{ z^s - \sum_{k,m} \phi_{km}^s \eta_{km} \right\}$$

The min-max regret optimization problem minimizes the maximum regret $R_{max}(\eta)$. In other words, for the robust deviation decision, we can formulate the min-max regret problem as follows:

(SCMCP-MR):

$$\min_{\boldsymbol{\eta} \in \beth} R_{max}(\eta) = \min_{\boldsymbol{\eta} \in \beth} \max_{s \in S} \left\{ z^s - \sum_{k,m} \phi_{km}^s \eta_{km} \right\}$$

which can be rewritten as

$$\min_{\boldsymbol{\eta} \in \beth, v} \; v$$

s. t.

$$v \geq z^s - \sum_{k,m} \phi_{km}^s \eta_{km}, \qquad s \in S$$

4.2. Robust Optimization

Robust optimization deals with data uncertainty but does not require a known distribution of the underlying uncertain parameters. It takes a worst-case approach to the decision-making problem formulations. The robust optimization approach solves an optimization problem assuming that the uncertain input data belong to an uncertainty set, and finds the optimal solution if the uncertainties take their worst-case values within that uncertainty set. The shape and the size of the uncertainty set can be used to vary the degree of conservativeness of the solution and to represent an investor’s risk preferences.

Robust optimization tries to find the optimal solution when the parameters in an optimization problem are not fixed, but are allowed to vary in pre-specified uncertainty sets. In practice, the robust optimization approach often reduces to solving the optimization problem when the uncertainties take on “worst-case values”. In a maximization problem, the robust optimization approach would involve maximizing the objective function under some kind of worst-case scenario for the coefficients in the problem. In financial decision making, the use of robust optimization has been justified by the empirically observed tendency for people to make choices that minimize the effect of the worst-case outcome (Schmeidler and Gilboa 2004).

Let U denote an uncertainty set where uncertain parameters such as the cluster lift scores belong. Then the robust optimization formulation of the objective in the targeted offers direct marketing problem can be stated as follows:

(RCMCP):

$$\max_{\boldsymbol{\eta} \in \beth} \min_{\tilde{\phi}_{km} \in U} \sum_{k=1}^{K} \sum_{m=1}^{M} \tilde{\phi}_{km} \eta_{km}$$

Uncertainty sets could be any sets, including collections of scenarios. When they are collections of scenarios, the robust optimization formulation has a structure similar to the problems considered in Section 4.1. One can also use characteristics or summary measures of the distribution of uncertainties $\tilde{\phi}_{km}$, such as means, standard deviations, covariances, support, etc.

The simplest example is when the input parameters in the optimization problem are allowed to take values within interval uncertainty sets, i.e., when we specify an upper and lower bound for each uncertain parameter (lift value in this case). If we can establish a lower bound for each uncertain parameter, we can then simply replace the lift values by their associated lower bounds ("worst case values") while keeping everything else the same, resulting in a standard deterministic linear programming model.

Many uncertainty sets used in the robust optimization literature are based on limiting the normed distances of the uncertain parameters from some nominal values, such as the point estimates obtained from statistical procedures. As we will show in Section 4.3, the choice of norm can be related to a risk measure used by the modeler to protect against uncertainty. Depending on the choice of norm, one can end up with more or less tractable robust counterparts. The robust counterparts can be derived and preserve the complexity of the original problem only for some uncertainty sets.

In the rest of this section, we introduce robust optimization formulations for the MCP with two uncertainty sets that are most commonly used in applications of robust optimization: the interval and the ellipsoidal (Euclidean norm) uncertainty sets. Because CMCP is a linear problem, the robust counterpart of CMCP can be derived in closed form for these uncertainty sets.

4.2.1. Interval Uncertainty

As we mentioned in the introduction to this section, interval uncertainty sets are the simplest uncertainty sets. Suppose $\phi_{km} \in \left[\underline{\phi}_{km}, \overline{\phi}_{km}\right]$, $k = 1, \dots, K, m = 1, \dots, M$. Such an interval can be obtained, for example, if we know the support of the distribution of the nominal cluster lift values, or if we calculate 95% confidence intervals for the values of the nominal lifts calculated from data. Formally, the interval uncertainty set can be defined as

$$U_{interval} = \left\{\phi_{km} \mid \phi_{km} \in \left[\underline{\phi}_{km}, \overline{\phi}_{km}\right], k = 1, \dots, K, m = 1, \dots, M\right\}$$

Then, the problem (RCMCP) can be rewritten as

(RCMCP-IN)

$$\max_{\boldsymbol{\eta}\in\beth, v} \quad v$$

s.t.

$$\min_{\underline{\phi}_{km} \leq \phi_{km} \leq \overline{\phi}_{km},\ k=1,\dots,K, m=1,\dots,M} \quad \sum_{k=1}^{K}\sum_{m=1}^{M} \phi_{km}\eta_{km} \geq v$$

Because η_{km} are nonnegative, the minimum in the constraint above is attained when all ϕ_{km} are at their minimum values $\underline{\phi}_{km}$. So, the robust counterpart of the original problem under interval uncertainty sets can be calculated in closed form:

(RCMCP-IN):

$$\max_{\boldsymbol{\eta}\in\beth} \sum_{k=1}^{K}\sum_{m=1}^{M} \underline{\phi}_{km}\eta_{km}$$

The interval uncertainty sets is often too conservative in practice, and it is rare that all uncertain coefficients take their worst-case values at the same time. A popular variation of this uncertainty set used in practice is the Bertimas and Sim (2004) uncertainty set, which specifies that only up to Γ of the uncertain coefficients in the optimization problem can take their worst-case values within the given intervals. It turns out that the Bertsimas and Sim (2004) uncertainty set is a special case of a *normed* uncertainty set,

$$U_{D-norm} = \left\{\boldsymbol{\phi} \middle| \,||\left|\mathbf{A}\left(vec(\boldsymbol{\phi}) - vec(\widehat{\boldsymbol{\phi}})\right)\right|\,||_d \leq \theta\right\}$$

where $vec(\boldsymbol{\phi})$ is a vector obtained by stacking the entries ϕ_{km}, $vec(\widehat{\boldsymbol{\phi}})$ is a vector obtained by stacking the nominal (expected) values $\widehat{\phi}_{km}$, **A** is a matrix of appropriate dimensions, and $|||.|||_d$ denotes the *D*-norm, which for an integer *d* is the sum of the absolute values of the largest *d* entries of a vector ***s*** (see, for example Gotoh and Uryasev 2016). The Bertsimas and Sim (2004) uncertainty set is a special case of the normed uncertainty set above when **A** = **I**, where **I** is the identity matrix. Often, the matrix **A** that is used is the inverse of the square root of the covariance matrix of the uncertain estimates $\widetilde{\boldsymbol{\phi}}$, $\boldsymbol{\Sigma}^{-1/2}$.

4.2.2. Ellipsoidal Uncertainty Set

The ellipsoidal uncertainty set is also a special case of the normed uncertainty set introduced in the previous section when the norm used in the definition of an uncertainty set is the Euclidean norm. (The Euclidean norm is the square root of the sum of the squares of the entries of a vector *s*). Usually, the matrix **A** is set to the inverse of the square root of the covariance matrix of uncertain coefficients, which we will denote by $\boldsymbol{\Sigma}^{-1/2}$. The ellipsoidal uncertainty set is perhaps the most widely used uncertainty set in the robust optimization literature, and its use has roots in the robust control literature (Ben-Tal et al. 2009). The reason for its popularity is that, as we mentioned in the previous subsections, the interval uncertainty sets in Section 4.2.1 can be too conservative for most practical purposes, and the mean and the covariance structure of the underlying data are often available. A correlation structure for the uncertain nominal lifts can be generated from data, for example, by using bootstrapping (see Lo and Pachamanova 2015), and the uncertainty set can instead be specified as

$$U_{ell} = \left\{\boldsymbol{\phi} | \; ||\boldsymbol{\Sigma}^{-1/2} \left(vec(\boldsymbol{\phi}) - vec\left(\widehat{\boldsymbol{\phi}}\right)\right)||_2 \leq \theta, \boldsymbol{\phi} \in \mathbb{R}^{K\times M}\right\}$$

Here $\boldsymbol{\phi}$ is the stacked vector of cluster-level lift estimates ϕ_{km} for $k = 1, \dots, K, m = 1, \dots, M$. The robust counterpart of a linear constraint when the uncertain coefficients are assumed to vary in an uncertainty set defined by a norm is a sum of a term involving the nominal values of the uncertain coefficients and a term involving a penalty (θ) and the dual norm of the norm in the uncertainty set definition. Because the Euclidean norm is self-dual, the robust counterpart of a linear constraint with uncertain coefficients restricted to lie in the ellipsoidal uncertainty set has a penalty term involving an Euclidean norm.

In other words, given the ellipsoidal uncertainty set for the coefficients $\widetilde{\boldsymbol{\phi}}$, the robust counterpart of the objective function constraint

$$\min_{\boldsymbol{\phi}\in U_{ell}} \sum_{k=1}^{K}\sum_{m=1}^{M} \phi_{km}\eta_{km} \geq v$$

becomes a deterministic nonlinear inequality. Because the Euclidean norm is self-dual, the robust counterpart of the above constraint is

$$\sum_{k=1}^{K}\sum_{m=1}^{M} \hat{\phi}_{km}\eta_{km} - \theta\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{\eta}\right\|_2 \geq v$$

The original problem under ellipsoidal set uncertainty can be stated as follows:

(RCMCP-EL)

$$\max_{\boldsymbol{\eta}\in\beth} \quad v$$

s.t.

$$\sum_{k=1}^{K}\sum_{m=1}^{M} \hat{\phi}_{km}\eta_{km} - \theta\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{\eta}\right\|_2 \geq v$$

There is a variety of other uncertainty sets that can be used to represent the uncertainty in the inputs to the CMCP. For example, Chen, Sim and Sun (2007) suggested an uncertainty set that becomes equivalent to the ellipsoidal uncertainty set if the underlying distributions are symmetric, but can otherwise take into consideration asymmetry in the probability distributions of the uncertain coefficients by taking advantage of so-called forward and backward deviations. Each of these uncertainty sets can be mapped to financial risk measures.

4.3. Risk Representation for the Marketing Campaign Problem under Uncertainty

Ryals et al. (2007) explored risk management of marketing revenues under a particular risk measure – the variance, which is a deviation-based risk measure, standard in finance. Other risk measures such as the quantile-based Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) are concerned with the probability or magnitude of losses instead of deviation from the expected value (see, for example, Jorion 2000 and Rockafellar and Uryasev 2002).

Defining uncertainty sets about the uncertain parameters in the CMCP translates into imposing financial risk measures on the total lift and revenues realized through the marketing campaigns. In the literature, uncertainty sets in robust optimization have been mapped to financial risk measures in order to reflect decision makers' preferences (Natarajan et al. 2009, 2010), to extend traditional financial risk measures (El Ghaoui et al. 2003, Natarajan et al. 2008, Bertsimas and Brown 2009), to incorporate features of underlying data for the uncertain coefficients (Goldfarb and Iyengar 2003, Ben-Tal et al. 2013), and to improve computational tractability (Bertsimas et al. 2018). The stochastic and robust optimization framework introduced in this paper therefore allows for much flexibility in the definition of objectives for risk management in the targeted marketing context.

Specifically, assuming that the uncertain coefficients lie in the ellipsoidal uncertainty set of a particular size described in Section 4.2.2 translates into imposing a penalty for the standard deviation of the "portfolio" of uncertain lift estimates. This problem is related to the revenue variance minimization problem analyzed in Ryals et al. (2007). When the uncertainty set is specified as scenarios, one obtains a stochastic programming formulation (Section 4.1), which can be written as an exact Value-at-Risk optimization formulation over representative scenarios from the distribution of possible lifts (Shapiro et al. 2009). When the uncertainty set is specified as the asymmetric uncertainty set of Chen, Sim and Sun (2007), mentioned in Section 4.2.2, it corresponds to a new, improved tail measure that approximates the Value-at-Risk risk measure and can be used to approximate the CVaR of the distribution of the underlying uncertainties as well. A framework for linking these uncertainty sets to data that may be available for the uncertainties in the problem is provided in Bertsimas et al. (2018), and some suggestions in the context of uplift modeling are provided in the following section.

5. Modeling Lift Estimate Uncertainty

How does one generate the input values to the stochastic and robust formulations? One could, for example, use the average lift for treatment *m* within a cluster *k* as the nominal estimate for the cluster lift for that treatment. Specifically,

$$\hat{\phi}_{km} = \sum_{i=1}^{N_k} \hat{\pi}_{im}$$

where N_k are the number of customers in cluster *k* and $\hat{\pi}_{im}$ is the individual lift for customer *i* and treatment m.

If one can assume that general laws, such as the Central Limit Theorem, apply, one could calculate, for example, the standard deviation of each estimate $\hat{\phi}_{km}$ as the standard deviation of the individual lifts for that treatment (*m*) in that cluster (*k*) divided by the square root of the number of observations in the cluster. However, it may not be possible to argue that the lifts of customers within a segment, no matter how carefully the segment is defined, can be considered observations from the same distribution and that the Central Limit Theorem would apply. Instead, we suggest using bootstrapping to generate scenarios. The scenarios generated from bootstrapping can then be used in the stochastic problem formulations, to estimate moments of the cluster level lift score for each treatment $E(\tilde{\phi}_{km}) = \hat{\phi}_{km}$ and covariances $Cov(\tilde{\phi}_{km}, \tilde{\phi}_{k'm'})$, to determine the worst-case value of the cluster-level lifts, etc.

Generating scenarios has the additional advantage that it allows us to capture potential correlation structures behind treatments within each cluster and between clusters. As an example, recall that the worst-case interval formulation in Section 4.2.1 involves the worst-case values of the cluster-level lifts for each treatment. If those are estimated based on joint scenarios generated for treatments within clusters, the worst-case estimates will not be as conservative as the worst-case values for individual estimates. This is because it is possible that within a cluster, the worst-case value for one treatment is not attained at the same time as the worst-case value for another treatment, and assuming that all worst-case values occur at the same time would be unnecessarily conservative. Taking into consideration how the values actually occur leads to better robust formulations performance in practice.

6. Computational Experiments

We use a real data set from an online retailer to illustrate the effect of incorporating uncertainty in the cluster lift estimates with the different optimization-under-uncertainty techniques described in Section 4. There are two treatments, men's and women's merchandise. Individual customer lift estimates are calculated using the Two Model Approach described in Section 2 for a no-mail control group and the two treatment groups from an email marketing campaign. Cluster analysis in SAS is then performed on the customers in the data set using the individual lift scores for men's merchandise and the lift score for women's merchandise as inputs to a k-means clustering algorithm. Originally, 10 clusters are obtained. Two clusters (the original clusters 1 and 3) are merged because of similarity and small sizes, and the clusters are renumbered from 1 to 9. The cluster sizes for which treatments need to be determined are shown in Table 1.

Cluster	Number of Observations
1	4,180
2	5,650
3	60,220
4	12,370
5	8,940
6	29,240
7	28,070
8	4,100
9	37,080
Total	**189,850**

Table 1. Cluster sizes.

The individual lift estimates are aggregated to obtain cluster-level lift estimates for each treatment. The bubble chart in Figure 2 shows the average lift per cluster for each treatment. It can be observed that Cluster 8 has high average lift for both men's and women's merchandise, whereas Cluster 2 actually has negative lift for the two treatments. If the nominal lift estimates were accurate, as much of the marketing efforts as possible should be directed towards Cluster 8, and none should be allocated to Clusters 2 or 1.

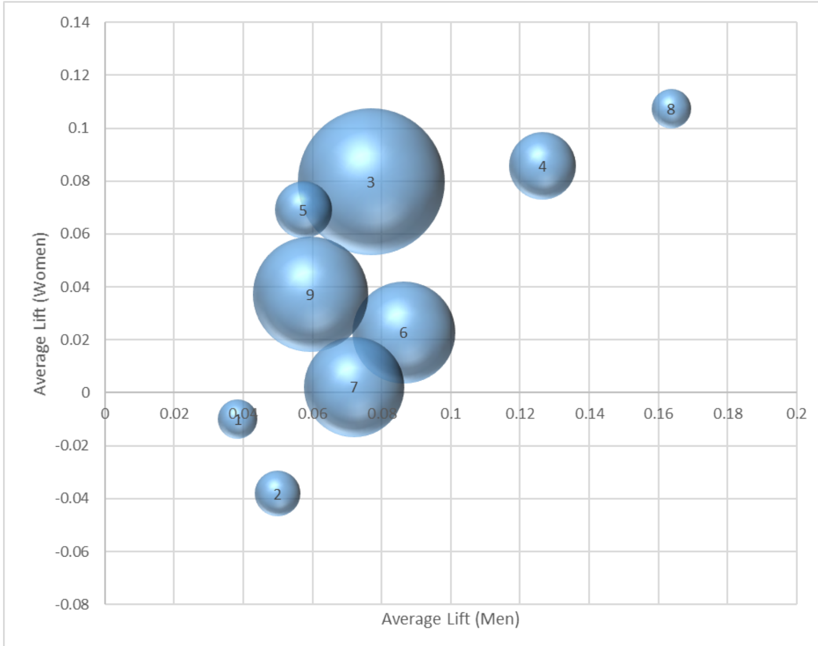


Figure 2. Bubble chart for average observed lift by cluster. Bubble size corresponds to the number of observations in the cluster.

We conduct the following experiments to study the effect of different ways of incorporating uncertainty on the realized lift. We use Python to aggregate and process the data, and AMPL with solvers IBM CPLEX and BARON to solve the optimization problems.

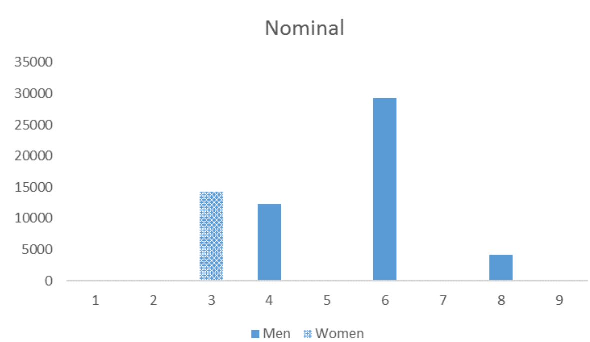
Training data is generated by drawing randomly 100 scenarios per cluster based from the online retailer dataset 50 times. These 50 datasets are used to calculate 50 means for the treatment lifts of the clusters. The average of these 50 means is used as the nominal estimate for the lifts in the optimization formulations, and the 50 scenarios for the means are used as possible scenarios in the various optimization formulations that require scenario data. The 50 scenarios are also used to generate the covariance matrix of the nominal lift estimates and determine the worst-case values for the lift estimates.

We assume that the cost per treatment is \$1 and that the available budget is \$60,000. The stochastic approaches described in Sections 4.1 and 4.2 are implemented, and we report results for the following models:

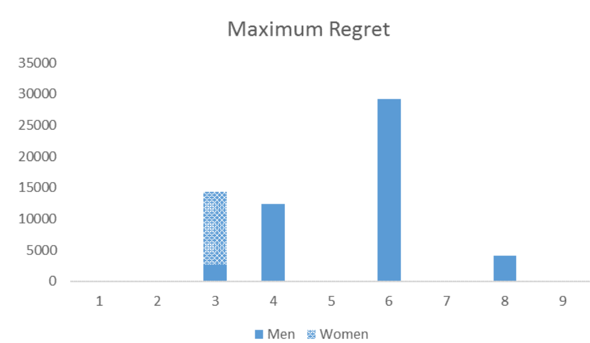
- The Nominal formulation (CMCP-N) from the introduction to Section 4, which will serve as a benchmark because it does not incorporate considerations for uncertainty;
- The Maximum Regret approach (SCMCP-MR), Section 4.1;
- The Interval uncertainty robust optimization formulation (RCMCP-IN), Section 4.2.1;
- The Ellipsoidal uncertainty robust optimization formulation (RCMCP-EL), Section 4.2.2.

6.1. Characteristics of Optimal Cluster Allocation

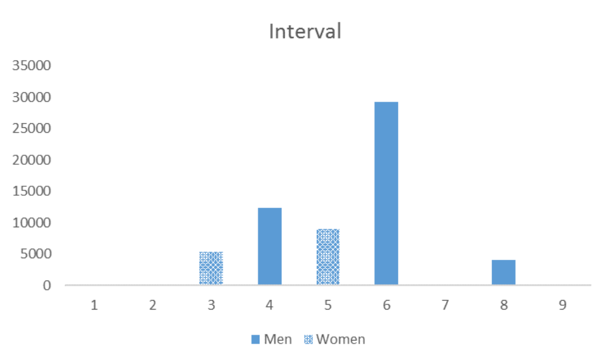
We first discuss the cluster allocations resulting from the different models. Figure 3 shows the cluster allocations for (a) (CMCP-N), (b) (SCMCP-MR), (c) (RCMCP-IN), (d) (RCMCP-EL) with value for the robustness budget theta of 20, (e) (RCMCP-EL) with theta equal to 150, and (f) (RCMCP-EL) with theta equal to 300.



(a)

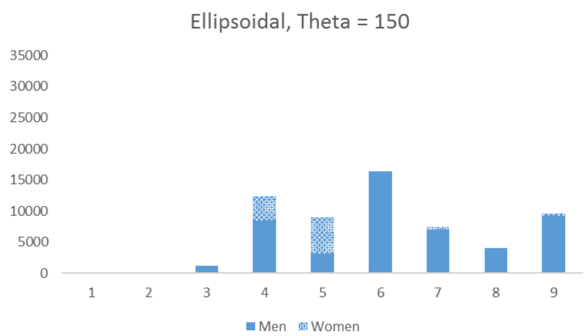


(b)



(c)

(d)



(e)

(f)

Figure 3. Cluster allocations for (a) (CMCP-N), (b) (SCMCP-MR), (c) (RCMCP-IN), (d) (RCMCP-EL) with $\theta = 20$, (e) (RCMCP-EL) with $\theta = 150$, and (f) (RCMCP-EL) with $\theta = 300$.

Some interesting observations can be made based on the results in Figure 3.

- The optimal cluster allocation for the Nominal problem (a) agrees with our observations from Figure 2. Because the costs of marketing to the different clusters are assumed to be the same, the algorithm logically selects to market the treatment (in most cases, Men's Merchandise) to the clusters with the highest lift until the budget is exhausted. It allocates 12370, 29240, and 4100 to men in Clusters 4, 6, and 8, respectively, which are the corresponding clusters' limits. Only after exhausting these possibilities does the algorithm allocate the remaining 14,290 to women in Cluster 3.
- The stochastic and robust approaches result in more diversification than the Nominal approach. With the exception of the Maximum Regret approach, all other stochastic approaches allocate to more clusters than the Nominal approach, and even the Maximum Regret approach, which allocates to the same number of clusters as the Nominal approach (4 clusters), chooses to diversify within the clusters and allocates to both men and women in Cluster 3.
- The degree of diversification for the Ellipsoidal approach can be controlled by the "robustness budget" parameter theta. The higher the value of theta, the more penalty there is for uncertain estimates deviating from their nominal values, and the more

different the Ellipsoidal and the Nominal allocations look. Specifically, when theta is relatively small (20), the cluster allocation (d) looks similar to the Nominal allocation with the addition of Cluster 7. When theta is 150 (e), the allocation is of smaller amounts to multiple clusters and treatments, spreading the risk. When theta is 300 (f), the penalty for uncertainty is so high, that it is optimal not to allocate the entire budget and to spread the allocation more evenly among the clusters.

6.2. Performance Evaluation

Do the different cluster allocations for the stochastic formulations actually result in risk reduction compared to the Nominal approach? To answer this question, a test data set of 200 scenarios is generated from the original data. Using the optimal allocations resulting from the various approaches in Section 6.1, we calculate the realized lift in each of the test scenarios. The summary statistics for the realized lifts for the different approaches are shown in Table 2.

	Nominal	Interval	Max Regret	Ellipsoidal20	Ellipsoidal150	Ellipsoidal300
Mean	5876.7065	5780.9885	5871.6872	5827.2433	5217.3279	1603.3474
Min	5037.6129	4998.5835	5022.7839	4971.9561	4423.2438	1408.0376
5th Per	5233.4083	5188.0957	5239.4815	5224.3975	4813.1027	1508.5897
25th Per	5575.3214	5439.5284	5576.4901	5537.0804	5017.5484	1558.0839
Median	5871.5743	5771.7755	5865.5381	5819.8640	5215.7888	1602.8412
75th Per	6154.0621	6047.4792	6143.4234	6093.9303	5398.0646	1652.8411
95th Per	6498.1474	6423.8280	6502.5265	6506.6897	5620.9669	1698.6911
Max	6897.7758	6779.3620	6920.7589	6770.4810	5851.9300	1751.7663
StDev	410.3164	393.8975	408.2891	400.2428	260.5810	62.1641
IQR	578.7408	607.9508	566.9333	556.8499	380.5162	94.7573
Range	1860.1629	1780.7786	1897.9750	1798.5249	1428.6863	343.7287
Coeff of Var	6.9821	6.8137	6.9535	6.8685	4.9945	3.8771

Table 2. Summary statistics for the realized lifts of the optimal strategies resulting from the different approaches based on the test set scenarios: mean, minimum, 5th percentile, 25th percentile (first quartile), median, 75th percentile (third quartile), 95th percentile, maximum, standard deviation, interquartile range (IQR), range (maximum – minimum), and coefficient of variation (standard deviation divided by mean and reported as a percentage).

It can be observed from Table 2 that the stochastic approaches reduce the realized average lift: all realized average lifts in the first row of the table are less than the realized average lift of the Nominal strategy. With the exception of Maximum Regret, all stochastic approaches also perform more poorly than the Nominal approach when it comes to minimum realized lift or 5^{th} percentile. However, the stochastic approaches are successful at reducing the spread of the distribution of realized lifts. The standard deviations realized from the stochastic strategies are lower than the standard deviation realized from the Nominal approach. The coefficient of variation, which is the standard deviation scaled by the realized mean, is also lower for the stochastic approaches than for the Nominal approach. This means that the risk per realized unit of reward is lower when uncertainty it taken into consideration during the optimization process.

Two visualizations – in Figure 4 and in Figure 5 – help to illustrate the effect of taking into consideration uncertainty in lift optimization. It can be observed that the methods for optimization under uncertainty tend to reduce the spread of possible outcomes; however, they do so at the expense of reduced average performance.

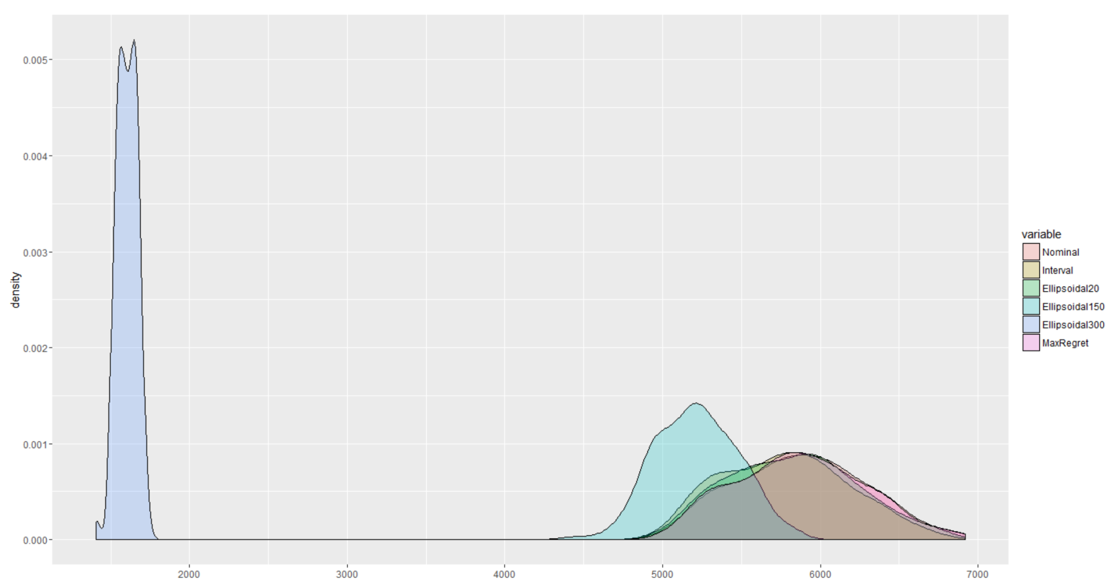


Figure 4. Overlapping density plots of the simulated outcomes from applying the nominal and other strategies on the test set.

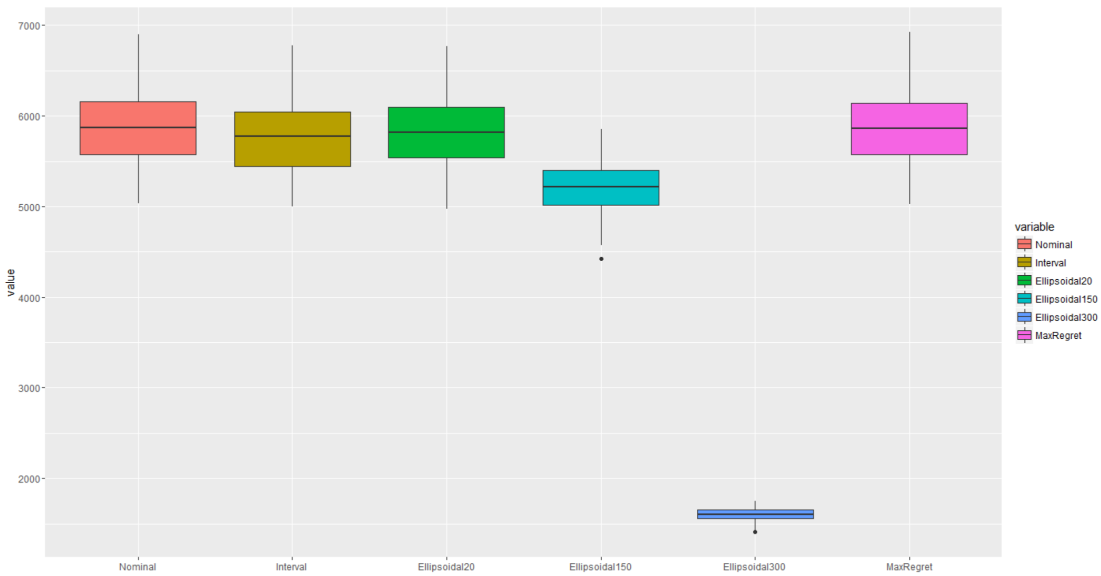


Figure 5. Boxplots for the distributions of the realized lifts with the different strategies on the scenarios from the test set.

It is helpful to analyze also the extent to which the stochastic optimization approaches account for the worst-case loss (WCL). This is important because, as explained in Section 3, expected lift optimization is directly related to expected revenue optimization. Specifically, if the expected lift for cluster k and treatment m is $\hat{\phi}_{km}$, the expected revenue for that cluster is $\hat{r}_{km}\hat{\phi}_{km}\eta_{km}$, where $\hat{r}_{km}$ is the expected revenue for treatment m per customer in cluster k. Underperforming in terms of expected revenue should be a source of concern.

We use the following metric for WCL: for each optimization model, we calculate the difference between the average realized total lift and the worst realized total lift as a percentage of the average realized lift based on the simulated observations in the test set. This measures how far the expected outcome is from the worst-case outcome under the optimal strategy for that approach – a similar concept to the concept of drawdown in investments (see http://www.investopedia.com/terms/d/drawdown.asp). Table 3 summarizes the results from the test set. It can be observed that the Interval uncertainty set approach as well as a calibrated Ellipsoidal approach (in this case, the Ellipsoidal approach with theta of 150) perform best under this metric. Specifically, if a marketing manager uses the Ellipsoidal approach with theta of 150, the maximum loss in revenues he can expect is 12.20% from the predicted expected revenue based on the model.

	Nominal	Interval	Max Regret	Ellipsoidal20	Ellipsoidal150	Ellipsoidal300
WCL (%)	14.3305	13.5748	14.7212	15.2410	12.2000	14.5038

Table 3. Worst-case realized loss calculation.

7. Discussion

In this section, we discuss in more detail the current state of uplift modeling in organizations, list our contributions in context, and explain how our framework can be implemented in practice.

7.1. Uplift versus traditional customer response modeling in organizations

Many organizations collect large amounts of customer data in order to understand their needs, predict their future behavior, and optimize future contacts. Early usage of analytics was to apply predictive modeling (also called supervised learning) to target customers who are likely to take a desirable action regardless of whether or not they receive an intervention or treatment, as documented in marketing analytics textbooks such as Jackson and Wang (1994) and Roberts and Berger (1999). Such practice almost guarantees that the model targets are better than random targets in terms of response rate by design.

Marketing measurement teams (typically separated from data science teams in corporations) are more focused on measuring campaign success. They apply A/B testing or randomized experiments to measure whether the campaign (treatment) generates incremental value for the targets over not receiving the campaign. For example, in a customer cross-sell campaign, a traditional predictive modeling approach would consist of developing a model to differentiate between those who responded from those who did not respond to a previous campaign, and then applying such model to a future campaign. A randomized experiment, on the other hand, would have two target groups: model targets (say, using the top 3 deciles in terms of model score) and random targets (for comparison and potential model fine-tuning). In each of the two target groups, customers would be split randomly into treatment (receiving the marketing campaign) and control (not receiving the campaign) groups, so any difference in measurement result can be attributable to the treatment (campaign).

A traditional predictive model is designed to focus on customers who are likely to take the desirable action, so one usually sees success over random targets. However, those customers who are likely to take the desirable action (as found by the top 3 model deciles in this example) may respond naturally, regardless of whether they receive the marketing campaign, resulting in no actual lift over control. Table 4 shows an illustrative example. Here there is a difference between the model group and the random group – the top 3 deciles of the model targets have a higher likelihood of purchase than the random targets. However, there is no difference in purchases between the treatment and the control groups within each target group; hence, the campaign does not make a difference and there is no lift. This happens because optimizing lift is not the objective of the traditional predictive model, i.e., *what is modeled does not match what is measured*.

	Model Targets (Top 3 Deciles)	**Random Targets**
Treatment	6.70%	2.50%
Control	6.70%	2.50%
Lift	**0.00%**	**0.00%**

Table 4. Example of Campaign Measurement Result from Traditional Predictive Modeling.

Since the idea of changing the objective was brought up by Radcliffe and Surry (1999) and Lo (2002), the data science marketing practice in industry has started to change its focus from traditional predictive modeling to uplift modeling in order to maximize impact over control. Uplift modeling has since spread from marketing and sales to political campaigns, e.g., targeting of swing voters in the 2012 Obama re-election campaign, as documented in Stedman (2013) and Siegel (2013). Independently, the healthcare industry has also been researching a similar set of methodologies to target patients who are likely to be impacted by a medical treatment (see, for example, Yong 2015).

7.2. Incorporating Model Uncertainty

Although predictive models (traditional or uplift) are routinely employed in marketing, most academic literature and industry practices ignore the model estimation risk caused by uncertainty of model estimates. Such risk tends to be more significant in uplift modeling because of its requirement to estimate the difference between treatment and control response rates, which is often relatively small and has high variability. Failure to address model estimation risk when using output from such models as inputs for treatment optimization or human decision making can result in highly uncertain outcomes and loss of valuable opportunities. This article brings awareness of this issue to managers and practitioners as well as researchers.

7.3. Our contribution

We outlined a framework for taking into consideration the uncertainty in assessing customer segment responsiveness when managing marketing campaigns, and studied the effect of various assumptions on the allocation of marketing spend. We emphasized the importance of measuring responsiveness through lift, which measures the differential effect of a campaign, and considered an optimization problem that maximizes the total lift (equivalently, the total marketing revenue) in order to decide on the optimal allocation of marketing spend to customer segments. We showed how to use the robust optimization and stochastic programming methodologies to take lift estimation errors into consideration in the allocation. The main idea was to consider either alternative scenarios for the possible values of lift, or "uncertainty sets" around the estimated values of lift that could be linked to the amount of variability in the estimate and are consistent with the financial risk measure framework suggested in Ryals et al. (2007). Imposing this protection against uncertainty required rewriting the optimization problem but did not make it substantially more difficult to solve with today's advanced optimization software. Rewriting the optimization problem to make marketing spend allocation more robust to uncertainty also required user specification of the desired level of protection against such uncertainty, expressed through a parameter specified in the optimization. This parameter, referred to as a "robustness budget", reflects different levels of aversion to uncertainty that could be expressed by a marketing manager, and its function is similar to the function of the coefficient of risk aversion of the marketing manager discussed in Ryals et al. (2007).

Our computational experiments showed the effect of varying the assumptions of uncertainty and the value of the robustness budget on the optimal allocation. A "nominal" allocation using only point estimates of a customer segment's responsiveness (or expected revenues) would take a greedy approach and allocate as much as possible to the segments with the highest expected return. This is a procedure often used in practice – a simple ranking of the estimated response or revenue for different segments. A "robustified" allocation would be more cautious, and would distribute the marketing spend in such a way as to diversify between segments in case the inputs provided to the algorithm predicting the amount of responsiveness were incorrect. The higher the value of the robustness budget, the more diversified the allocation is among segments, reducing overall risk as measured by the spread of possible outcomes and the worst-case loss.

The summary statistics and metrics presented in the computational experiments are important for the assessment of the financial viability of the application of techniques from stochastic and robust optimization in uplift modeling and risk management. Because expected revenues and drawdown in revenues are proportional to the expected lift and drawdown in lift, respectively, substantial savings can be realized from testing and calibrating robust models to improve the risk characteristics of marketing campaigns.

7.4. Implementation of the robust framework in practice

Given the importance of the MCP, specialized commercial marketing optimization software packages are available to handle the MCP optimization problem. Examples include MarketSwitch[1] and SAS Marketing Optimization.[2] Such customized software typically integrates directly with campaign management tools and uses proprietary mathematical algorithms. Alternatively, the MCP can be solved using popular open source (free) modeling languages such

---

[1] See http://www.experian.com/decision-analytics/marketswitch-optimization.html.

[2] See https://www.sas.com/en_us/software/marketing-optimization.html.

as R,[3] Python[4] and Julia.[5] The statistical estimation and optimization in our proposed robust framework are not implemented in commercial packages; however, a modeler using the open source modeling languages could easily add on the robust modeling capability to existing code. R, Python, Julia and similar modeling environments have a range of useful statistical and optimization libraries that can be used as building blocks for implementing the nominal and the robust versions of the lift estimation algorithm and the optimization formulations suggested in this article.

8. Implications for Managers

The focus on customer-centric marketing over the last two decades has created an imperative to understand customer characteristics and needs, and better data collection and analytics are important factors that have made that possible (Shah et al. 2006, Kumar et al. 2006, Pop 2017). At the same time, two critical principles for using data and analytics effectively in decision making are not always applied in practice: (1) training models and decision makers to optimize metrics that actually matter, and (2) incorporating considerations for the expected degree of model error given the data quality or statistical procedure used. This article showed how these two principles can be integrated with marketing campaign management.

Once an organization has adopted models that optimize the right metrics, accountability for model measurement error can be incorporated within organizational processes so that marketing managers can be alerted to the level of uncertainty when ranking opportunities. The tolerance towards model estimation error can be calibrated. In the context of marketing spend allocation that balances risk and return between various segments, Ryals et al. (2007) warn about the danger of individual managers' risk preferences not aligning with organizational risk preferences. In the context of marketing spend allocation that balances considerations for uncertainty in the statistical estimates of a segment's responsiveness, the risk preference

---

[3] See http://r-project.com/.

[4] See https://www.python.org/.

[5] See https://www.juliaopt.org/.

(robustness) level can be set based on characteristics of the data, and can be managed consistently across the organization.

In recent years, calls for facilitating strategic integration of marketing with other functional areas of an organization have increased (Piercy and Rick 2014, Kumar et al. 2006, Woodburn 2004, Hulbert et al. 2003). Marketing can increase its influence within organizations by understanding and adopting concepts that are traditionally used by other areas such as operations management and finance. The marketing campaign management framework outlined in this article brings together such concepts: predictive models, appropriate metrics, model accuracy, revenue management, and risk management.

References

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203-228.

Asllani, A. and D. Halstead. (2011). Using RFM Data to Optimize Direct Marketing Campaigns: A Linear Programming Approach. *Academy of Marketing Studies Journal*. 15(2):59-76.

Atamturk, A. (2004). Polyhedral Methods in Discrete Optimization. In *Trends in Optimization*. S. Hosten, J. Lee, R. Thomas (eds.), Proceedings of Symposia in Applied Mathematics, American Mathematical Society, 61:21-37

Athey, S. and G.W. Imbens (2015). Machine Learning Methods for Estimating Heterogeneous Causal Effects, *Working Paper*, Graduate School of Business, Stanford University.

Beltran-Royo, C., L.F. Escudero, and H. Zhang. (2016). Multiperiod Multiproduct Advertising Budgeting: Stochastic Optimization Modeling. *Omega*. 59(A): 26-39.

Ben-Tal, A., & Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, 23(4), 769-805.

Ben-Tal, A., El Ghaoui, L., & Nemirovski, A. (2009). *Robust optimization*. Princeton University Press.

Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., & Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2), 341-357.

Ben-Tal, A., & Nemirovski, A. (2001). *Lectures on modern convex optimization: analysis, algorithms, and engineering applications* (Vol. 2). SIAM.

Bertsimas, D., & Brown, D. B. (2009). Constructing uncertainty sets for robust linear optimization. *Operations Research*, 57(6), 1483-1495.

Bertsimas, D., Brown, D. B., & Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM review*, 53(3), 464-501.

Bertsimas, D., Gupta, V., & Kallus, N. (2018). Data-driven robust optimization. *Mathematical Programming*, 167(2), 235-292.

Bertsimas, D., Mersereau, A. J. (2007). A learning approach for interactive marketing to a customer segment. *Operations Research*, 55(6), 1120-1135.

Bertsimas, D., Sim, M. (2004). The price of robustness. *Operations Research*, 52(1), 35-53.

Bick, G. N. (2009). Increasing shareholder value through building customer and brand equity. *Journal of Marketing Management*, 25(1-2), 117-141.

Bitran, G., S. Mondschein. (1996) Mailing decisions in the catalog sales industry. *Management Science*. 42(9):1364–1381.

Birge, J. R., & Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science & Business Media.

Brockhaus Sr, R. H. (1980). Risk taking propensity of entrepreneurs. *Academy of Management Journal*, 23(3), 509-520.

Chen, X., M. Sim, P. Sun. (2007). A robust optimization perspective of stochastic programming. *Operations Research*. 55(6): 1058–1071.

Cont, R., Deguest, R., & Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, 10(6), 593-606.

Curcuru, N. (2011). The Relationship between revenue management and marketing: Why it's crucial for these departments to work together. *HospitalityNet*. Available from https://www.hospitalitynet.org/opinion/4054064.html.

Dantzig, G. B. (1955). Linear programming under uncertainty. *Management Science.* 1(3): 197-206.

Deza, A., K. Huang, and M. R. Metel. 2015. Chance Constrained Optimization for Targeted Internet Advertising. *Omega*. 53: 90–96.

Doyle, P. (2000). Valuing marketing's contribution. *European Management Journal*, 18(3), 233-245.

Du, X., & Chen, W. (2000). Towards a better understanding of modeling feasibility robustness in engineering design. *Journal of Mechanical Design*, 122(4), 385-394.

Fabozzi, F. J., Kolm, P. N., Pachamanova, D. A., & Focardi, S. M. (2007). *Robust portfolio optimization and management*. John Wiley & Sons.

Fabozzi, F. J., Huang, D., & Zhou, G. (2010). Robust portfolios: contributions from operations research and finance. *Annals of Operations Research*, 176(1), 191-220.

Fang, Z. X., & Li, H. Q. (2009). Robustness of engineering structures and its role in risk mitigation. *Civil Engineering and Environmental Systems*, 26(3), 223-230.

Garlappi, L., Uppal, R., & Wang, T. (2006). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies*, 20(1), 41-81.

Ghaoui, L. E., Oks, M., & Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4), 543-556.

Goldfarb, D., & Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of*

Operations Research, 28(1), 1-38.

Gotoh, J. Y., & Uryasev, S. (2016). Two pairs of families of polyhedral norms versus $ $\ell _p $ $ ℓ p-norms: proximity and applications in optimization. *Mathematical Programming*, 156(1-2), 391-431.

Gubela, R. M., Lessmann, S., Haupt, J., Baumann, A., Radmer, T., & Gebert, F. (2017). Revenue Uplift Modeling. Conference Paper, Thirty Eighth International Conference on Information Systems (ICIS), Seoul, South Korea.

Hildebrandt, L., Wagner, U. (2000). Marketing and operations research – a literature survey. *OR Spectrum*. 22: 5–18.

Hulbert, J. M., Capon, N., & Piercy, N. (2003). Total integrated marketing: Breaking the bounds of the function. Simon and Schuster.

Ibragimov, M., Ibragimov, R., & Walden, J. (2015). *Heavy-tailed distributions and robustness in economics and finance* (Vol. 214). Basel: Springer.

Jackson, R. & Wang, P. (1996). *Strategic Database Marketing*. Chicago: NTC Publishing.

Jorion, P. (2000). Value at Risk: The New Benchmark for Managing Financial Risk. McGraw-Hill, New York.

Juttner, U., Christopher, M. & Baker, S. (2007). Demand chain management-integrating marketing and supply chain management. *Industrial Marketing Management*. 36(3): 377–392.

King, A. J., & Wallace, S. W. (2012). *Modeling with stochastic programmin*g. Springer Science & Business Media.

Kubiak, R. (2012). Net lift model for effective direct marketing campaigns at 1800flowers. com. In *SAS Global Forum*, Paper (pp. 108-2012).

Kumar, V., Lemon, K. N., & Parasuraman, A. (2006). Managing customers for value: An overview and research agenda. Journal of Service Research, 9(2), 87-94.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan management review*, 52(2), 21.

Lo, V.S.Y. (2002) The True-lift Model - A Novel Data Mining Approach to Response Modeling in Database Marketing. *ACM SIGKDD Explorations*, 4, 2, 78-86.

Kane, K., Lo, V. and J. Zheng. (2014) Mining for the Truly Responsive Customers and Prospects Using True-Lift Modeling: Comparison of New and Existing Methods. *Journal of Marketing Analytics*, 2(4): 218-238.

Lo, V.S.Y. (2008) New Opportunities in Marketing Data Mining. In *Encyclopedia of Data Warehousing and Mining*, 2nd edition, J. Wang (ed.). Idea Group Publishing.

Lo, V.S.Y. and D.A. Pachamanova. (2015). From Predictive Uplift Modeling to Prescriptive Uplift Analytics: A Practical Approach to Treatment Optimization While Accounting for Estimation Risk. *Journal of Marketing Analytics*. 3(2): 79–95.

Lukas, B. A., Whitwell, G. J., & Doyle, P. (2005). How can a shareholder value approach improve marketing's strategic influence?. *Journal of Business Research*, 58(4), 414-422.

Lummus, R. R., Duclos, L. K. & Vokurka, R. J. (2003). The impact of marketing initiatives on the supply chain. *Supply Chain Management: An International Journal*. 8(4): 317–323.

Lund, B. (2012) Direct Marketing Profit Model. In: *Proceedings of Midwest SAS Users Group*, Paper CI-04.

Lusch, R. F., & Webster Jr, F. E. (2011). A stakeholder-unifying, cocreation philosophy for marketing. *Journal of Macromarketing*, 31(2), 129-134.

McDonald, M. (2006). Do your marketing strategies create or destroy shareholder value?. *Journal of Database Marketing & Customer Strategy Management*, 13(2), 102-107.

Medvedev, Y. (2016) Personal Lending: Customer Credit and Pricing Optimization, *SAS Global Forum,* paper 2600-2016.

Natarajan, K., D. Pachamanova and M. Sim. (2008). Incorporating Asymmetric Distributional Information in Robust Value-at-Risk Optimization. *Management Science.* 54(3): 573-585.

Natarajan, K., D. Pachamanova and M. Sim. (2009). Constructing Risk Measures from Uncertainty Sets. *Operations Research.* 57(5): 1129-1141.

Natarajan, K., Sim, M., & Uichanco, J. (2010). Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance*, 20(4), 695-731.

Pennings, J. M., & Wansink, B. (2004). Channel contract behavior: The role of risk attitudes, risk perceptions, and channel members' market structures. *The Journal of Business*, 77(4), 697-724.

Piercy, N. C., & Rich, N. (2004). Strategic marketing and operations relationships: the case of the lean enterprise. *Journal of Strategic Marketing*, 12(3), 145-161.

Pop, O-M. (2017). Ecosystems: Goods-dominant vs Service-dominant logic. Available at https://blog.hypeinnovation.com/ecosystems-goods-dominant-vs-service-dominant-logic.

Radcliffe, N.J. (2007) Generating Incremental Sales: Maximizing the Incremental Impact of Cross-Selling, Up-Selling and Deep- Selling Through Uplift Modelling. Stochastic Solutions Limited

Radcliffe, N. J., & Surry, P. D. (1999). Differential response analysis: Modeling true response by isolating the effect of a single action. *Credit Scoring and Credit Control VI*. Edinburgh, Scotland.

Rhee, M., Mehra, S. (2006). Aligning operations, marketing, and competitive strategies to enhance performance: An empirical test in the retail banking industry. *Omega*. 34(5): 505–515.

Roberts, M.L. and Berger, P.D. (1999). Direct Marketing Management. New Jersey: Prentice-Hall.

Rockafellar, R. T., S. Uryasev. (2002). Conditional value-at-risk for general loss distributions.

Journal of Banking and Finance. 26(7): 1443–1471.

Ryals, L., Dias, S., & Berger, M. (2007). Optimising marketing spend: return maximisation and risk minimisation in the marketing portfolio. *Journal of Marketing Management*, 23(9-10), 991-1011.

Shah, D., Rust, R. T., Parasuraman, A., Staelin, R., & Day, G. S. (2006). The path to customer centricity. *Journal of service research*, 9(2), 113-124.

Shah, N., Soni, H. and Patel, K. (2013). Optimizing inventory and marketing policy for non-instantaneous deteriorating items with generalized type deterioration and holding cost rates. *Omega*. 41(2): 421–430.

Schmeidler, D., & Gilboa, I. (2004). Maxmin expected utility with non-unique prior. In *Uncertainty in Economic Theory* (pp. 141-151). Routledge.

Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2009). *Lectures on stochastic programming: modeling and theory*. Society for Industrial and Applied Mathematics.

Shmueli, G., Bruce P., Patel N. (2016) Data Mining for Business Analytics: Concepts, Techniques, and Applications with XLMiner. 3rd Ed. J. Wiley & Sons.

Siegel, Eric (2011) Uplift Modeling: Predictive Analytics Can't Optimize Marketing Decisions Without It. Prediction Impact white paper sponsored by Pitney Bowes Business Insight.

Siegel, E. (2013) *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die.* New Jersey: Wiley.

Simester, D., P. Sun, J. N. Tsitsiklis. (2006) Dynamic catalog mailing policies. *Management Science*. 52(5): 683–696.

Stewart, D. W. (2009). Marketing accountability: Linking marketing actions to financial results. *Journal of Business Research*, 62(6), 636-643.

Stedman, C. (2013), How Uplift Modeling Helped Obama's Campaign – and Can Aid

Marketers. *Predictive Analytics Times*, available at: https://www.predictiveanalyticsworld.com/patimes/how-uplift-modeling-helped-obamas-campaign-and-can-aid-marketers/2613/.

Storey, A. and M. Cohen. (2002). Exploiting Response Models: Optimizing Cross-Sell and Up-Sell Opportunities in Banking. *In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 325-331, New York, NY.

Szegö, G. (2002). Measures of risk. *Journal of Banking & Finance*, 26(7), 1253-1272.

Tarasi, C. O., Bolton, R. N., Hutt, M. D., & Walker, B. A. (2011). Balancing risk and return in a customer portfolio. *Journal of Marketing*, 75(3), 1-17.

Wallace, S. W., & Ziemba, W. T. (Eds.). (2005). Applications of stochastic programming. Society for Industrial and Applied Mathematics.

Wiesemann, W., Kuhn, D., & Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, 62(6), 1358-1376.

Woodburn, D. (2004). Engaging marketing in performance measurement. *Measuring Business Excellence*, 8(4), 63-72.

Yong, Florence H. (2015) Quantitative Methods for Stratified Medicine. PhD Dissertation, Department of Biostatistics, Harvard T.H. Chan School of Public Health, Harvard University.

Zhao, Y., Fang, X., and Simchi-Levi, D. (2017). Uplift Modeling with Multiple Treatments and General Response Types. *Conference Paper*, 2017 SIAM International Conference on Data Mining (SDM 2017). Available at https://arxiv.org/pdf/1705.08492.pdf.