Operations Research

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

Robust Product Line Design

Dimitris Bertsimas, Velibor Mišić

OPERATIONS RESEARCH
Articles in Advance, pp. 1–18
ISSN 0030-364X (print) | ISSN 1526-5463 (online)

https://doi.org/10.1287/opre.2016.1546

Robust Product Line Design

Dimitris Bertsimas
Sloan School of Management and Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, dbertsim@mit.edu

Velibor V. Mišić
Anderson School of Management, University of California, Los Angeles, Los Angeles, California 90095, velibor.misic@anderson.ucla.edu

The majority of approaches to product line design that have been proposed by marketing scientists assume that the underlying choice model that describes how the customer population will respond to a new product line is known precisely. In reality, however, marketers do not precisely know how the customer population will respond and can only obtain an estimate of the choice model from limited conjoint data. In this paper, we propose a new type of optimization approach for product line design under uncertainty. Our approach is based on the paradigm of robust optimization where, rather than optimizing the expected revenue with respect to a single model, one optimizes the worst-case expected revenue with respect to an uncertainty set of models. This framework allows us to account for parameter uncertainty, when we may be confident about the type of model structure but not about the values of the parameters, and structural uncertainty, when we may not even be confident about the right model structure to use to describe the customer population. Through computational experiments with a real conjoint data set, we demonstrate the benefits of our approach in addressing parameter and structural uncertainty. With regard to parameter uncertainty, we show that product lines designed without accounting for parameter uncertainty are fragile and can experience worst-case revenue losses as high as 23%, and that the robust product line can significantly outperform the nominal product line in the worst case, with relative improvements of up to 14%. With regard to structural uncertainty, we similarly show that product lines that are designed for a single model structure can be highly suboptimal under other structures (worst-case losses of up to 37%), while a product line that optimizes against the worst of a set of structurally distinct models can outperform single model product lines by as much as 55% in the worst case and can guarantee good aggregate performance over structurally distinct models.

Keywords: product line design; robust optimization; parameter uncertainty; structural uncertainty; model uncertainty.
Subject classifications: marketing: choice models, new products.
Area of review: Marketing Science.
History: Received September 2014; revisions received December 2015, May 2016; accepted June 2016. Published online in *Articles in Advance* November 3, 2016.

1. Introduction

A *product line* is a collection of products that are variations of a single basic product, differing with respect to certain attributes. Firms offer product lines to account for heterogeneity in consumer preferences. For example, consider a producer of breakfast cereals. The basic product may be a type of cereal (e.g., corn flakes), and a product line may consist of versions of this cereal that differ with respect to attributes such as the size of the cereal box, the flavor, whether the cereal has low-fat content, whether it has low-calorie content, whether it is nutritionally enriched with certain vitamins, what its retail price is, and so on. The customers of this cereal company may include: weight-conscious adults who may prefer healthy versions of the cereal, college students who may prefer certain flavors and larger packages (with lower cost per unit weight), or fitness-oriented adults who may prefer enriched cereal that has added vitamins and minerals in smaller packages and who may be willing to pay more than other customers. In this case, it is clear that there is no single cereal product that would be highly desirable for all three groups. Rather, it would be more appropriate to introduce several versions of the cereal that, taken together, would appeal to all three categories of customers.

The problem of *product line design* (PLD) is to select the attributes of the products comprising the product line so as to maximize the revenue that will result from the different types of preferences within the customer population. This problem is one of vital importance to the success of a firm. Much research effort has been devoted to modeling and solving the PLD problem.

The key prerequisite to practically all existing PLD approaches is a choice model, which specifies the probability that a random customer selects one of the products in the product line or opts to not purchase any of them, given the set of products that is offered. Almost all approaches to PLD tacitly assume that this choice model is known precisely and is beyond suspicion. In practice, however, this is not the case. Typically, the firm can only estimate the model from data that is obtained via conjoint analysis, wherein a small number of customers from the overall customer population is asked to either rate or choose from different hypothetical products. As such, the firm may face significant *uncertainty* in the choice model. This uncertainty can be of the following two types:

1. *Parameter uncertainty*. For a given parametric choice model that fits the data, the firm may be uncertain about the true parameter values (e.g., the partworths and the segment probabilities) due to the small sample size of the data. The concept of parameter uncertainty and the question of how to make decisions under parameter uncertainty have been well studied in the operations research community in other applications (we survey some of the relevant literature in Section 2), but have received little attention in the context of product line decisions.

2. *Structural uncertainty*. For the given conjoint data, the firm may be uncertain about the right type of parametric model to use to describe the customer population (e.g., a first-choice model or a latent-class multinomial logit model). This type of uncertainty is conceptually different from parameter uncertainty, where the focus is restricted to a single type of model. The concept of structural uncertainty is a new one that was recently proposed and studied in Bertsimas et al. (2016) in the context of designing screening policies for prostate cancer under different models of the evolution of a patient's health. To the best of our knowledge, this type of uncertainty has not been considered previously in the context of product line decisions.

Uncertainty is of significant interest because the optimal product line can change dramatically as one considers different parameter values or different structures of the choice model. Moreover, a product line that is designed for a particular choice model may result in very different revenues if the realized choice model is different from the planned one.

The issue of uncertainty becomes even more significant when one considers the nature of product line decisions. A firm's decision of which products to offer is one that is made infrequently; in most cases, the decision to produce a particular collection of products is one that commits the firm's manufacturing, marketing, and operational resources and cannot be easily reversed or corrected. This relative lack of recourse, along with the strategic importance of the decision, underscores the need for a PLD approach that is immunized to the uncertainty in the underlying choice model.

In this paper, we propose a new optimization approach for PLD that addresses uncertainty in customer choice. Rather than finding the product line that maximizes the revenue under a single, nominal choice model, we propose a robust optimization approach where we specify a set of possible models—the *uncertainty set*—and find the product line that optimizes the worst-case expected revenue, where the worst case is taken over all of the models in the uncertainty set. We make the following contributions:

1. We propose a new type of PLD problem that accounts for choice model uncertainty using robust optimization. Our approach is flexible in that it is compatible with many existing, popular choice models—such as the first-choice model, the latent class multinomial logit (LCMNL), model and the hierarchical Bayes mixture multinomial logit (HB-MMNL) model—and we propose different types of uncertainty sets that account for parametric and structural uncertainty. We also discuss how to trade-off nominal and worst-case revenue by optimizing a weighted combination of the two types of revenues.

2. We demonstrate the value of our approach through computational experiments using a real conjoint data set. In particular:

- We consider parameter uncertainty in the first-choice model, the LCMNL model, and the HB-MMNL model. We show that the nominal product line under each of these types of models is highly vulnerable to parameter uncertainty, and that the robust product line provides a significant edge in the worst case. For example, for the LCMNL model with $K = 3$ customer segments, the revenue of the nominal product line drops by more than 10% in the worst case, while the robust product line leads to worst-case revenues that are 4% higher than the nominal product line.
- We consider structural uncertainty through two examples. In the first example, we consider an uncertainty set of LCMNL models where the models vary in terms of the number of segments K; here, we find that the nominal product line under any one LCMNL model can result in a worst-case revenue that is lower than the nominal revenue by 3%–7%, while the robust product line outperforms the nominal product lines in the worst case by 2%–4%. In the second example, we consider an uncertainty set that consists of LCMNL models and the first-choice (FC) model. Here, we find that the LCMNL models and the FC model lead to strikingly different product lines, and the product line that is optimal for each individual model is highly suboptimal under the other models. In contrast, the robust product line that accounts for all of the structurally distinct models outperforms each nominal product line under the worst-case model.

The rest of the paper is organized as follows. In Section 2, we review the existing literature on PLD. In Section 3, we describe our PLD approach to address uncertainty in customer choice. In Section 4, we present extensive computational evidence that highlights the need to account for uncertainty and the benefits of adopting our approach to address uncertainty. Finally, in Section 5, we conclude the paper and discuss some possible directions for future work.

2. Literature Review

To date, significant research has been conducted in solving the PLD problem. The majority of approaches to the PLD problem model it as an integer optimization problem and assume a first-choice model of customer behavior; examples include Zufryden (1982), Green and Krieger (1985), McBride and Zufryden (1988), Dobson and Kalish

(1988), Kohli and Sukumar (1990), and Dobson and Kalish (1993). The approaches differ in whether the procedure selects the product line from a predefined set of products ("product space" procedures; examples include Green and Krieger 1985; McBride and Zufryden 1988; Dobson and Kalish 1988, 1993) or whether the procedure directly prescribes the attributes and is not restricted to a predefined set ("attribute space" procedures; examples include Zufryden 1982, Kohli and Sukumar 1990). The approaches also differ in the solution method. For example, Zufryden (1982) and McBride and Zufryden (1988) solve their respective PLD problems exactly as integer optimization problems, while Green and Krieger (1985) consider several heuristics, including a greedy heuristic and Kohli and Sukumar (1990) consider a dynamic programming-based heuristic. For a comprehensive comparison of heuristics for the nominal first-choice PLD problem, the reader is referred to the paper of Belloni et al. (2008).

Apart from the first-choice model, other product line research has considered probabilistic choice models, such as the multinomial logit (MNL) model, where customers do not deterministically select their highest utility choice, but rather they randomly select from all of their choices; typically, options with higher utilities are selected more frequently. Examples of product line approaches built on probabilistic choice models include Chen and Hausman (2000), Kraus and Yano (2003), Schön (2010a, 2010b). In Chen and Hausman (2000), the authors consider the problem of selecting a product line from a finite set of products to maximize expected revenue under the MNL model. Although the problem belongs to the class of mixed-integer nonlinear optimization problems, which are typically very difficult to solve, the paper shows that due to the problem structure, it is sufficient to solve the continuous relaxation to obtain an optimal solution to the original problem. Schön (2010a) builds on this modeling approach to incorporate price discrimination for different customer segments, while Schön (2010b) further extends the approach to more general attraction models. Kraus and Yano (2003) consider a share-of-surplus choice model, where the probability of a segment selecting a product is the ratio of the surplus of that product to the total surplus over all of the products. The problem they formulate is a mixed-integer nonlinear optimization problem, which they solve with a combination of simulated annealing and steepest ascent. Kraus and Yano (2003) also additionally address uncertainty in the utilities by modeling the uncertain utilities as random variables and applying a heuristic that attempts to optimize the expected profit. Outside of the academic literature, Sawtooth Software's Advanced Simulation Module (ASM) (Sawtooth Software 2003) allows analysts to perform simulations and to find product lines that optimize one of a number of objectives (such as market share or revenue) with respect to different choice models. These choice models include the first-choice model, the MNL model, and the randomized first-choice (RFC) model. The RFC model is a type of probabilistic choice model, wherein one simulates many customers who choose according to the first-choice rule with a randomly generated utility function; the product utilities are altered for each customer by perturbing the partworths ("attribute error") and then perturbing the utility of each product ("product error").

The approach we propose in this paper differs from the extant PLD literature in two main ways. First, existing approaches to the PLD problem under the first-choice model and under probabilistic choice models assume that the choice model that describes the customer population is known precisely. For example, in the first-choice model, it is assumed that we precisely know the segments comprising the customer population along with their partworths and sizes. As discussed earlier, typically the parameter values that define the choice model may not be known with complete certainty. As we will show later, product lines that do not account for errors in the parameter values that define the choice model can lead to significantly lower revenues when the realized parameter values are different from their planned values. To the best of our knowledge, no other work has highlighted the importance of accounting for parameter uncertainty in PLD or proposed a PLD approach that directly accounts for this uncertainty. Closest in spirit to some of the analyses we conduct is part of the paper of Belloni et al. (2008), where the authors present a post hoc test of how robust the product lines are with respect to error in the partworth utilities. In this experiment, the firm observes partworths that are perturbed from the true partworths (to simulate measurement error) and designs the product line under the perturbed partworths; the authors show that in this setting, the realized revenues are 2%–5% lower than their anticipated values due to this measurement error. In this paper, we consider uncertainty more generally over a wide range of choice models (the first-choice model, the LCMNL model, and the HB-MMNL model) and propose an approach that accounts for uncertainty upfront; as such, our paper complements this existing set of results. With regard to the RFC model implemented in Sawtooth Software's ASM, we emphasize that the RFC model addresses *randomness* in choice in the same way that other probabilistic choice models do; attribute error is analogous to assuming a probability distribution over partworths (as in the LCMNL model, which assumes a discrete mixture distribution or the HB-MMNL model, which assumes a multivariate normal mixture distribution), and product error is analogous to utility error in a random utility framework (as in the MNL model, which assumes standard Gumbel errors, or the multinomial probit model, which assumes normally distributed errors). The RFC model does not address uncertainty in a worst-case way, as we do in the present paper.

Second, virtually all existing approaches to the PLD problem focus on a single parametric structure for the choice model. In contrast, our approach is also designed to accommodate structural uncertainty and account for *multiple* parametric structures. The motivation for this setting is that often we may not be able to isolate a single parametric model as the "best" model, but instead we may have a collection of models that are "good enough" and that could all plausibly describe how the customer population will respond to the product line. The danger of optimizing for a single model is that we do not take the other models into consideration: while the single model product line will guarantee the optimal revenue under the chosen model, it is not guaranteed to deliver good performance under any of the other models. Thus, in our approach, rather than using a single model, we account for all of the models by optimizing a product line with respect to worst-case revenue over the set of models. In this way, the product line is able to guarantee good performance over a range of models, rather than only guaranteeing good performance for a single model. To the best of our knowledge, the idea of optimizing for a set of structurally distinct models has not been proposed in the PLD literature.

The preceding discussion has focused on PLD under uncertainty in the underlying choice model; it is worth noting that other research has considered PLD under other forms of uncertainty. One such uncertainty is engineering uncertainty, where the products must be designed so that they meet certain engineering specifications under varying use conditions; see, for example, Luo (2011). Although we do not consider the engineering aspect of PLD or engineering uncertainty in this paper, it would be interesting to consider in future research.

A related problem to PLD is the single product design (SPD) problem, where the firm seeks to design a single product as opposed to a collection of products; for an overview of the SPD literature, the reader is referred to the literature review of Luo (2011). Within the SPD literature, the papers of Luo et al. (2005) and Besharati et al. (2006) consider robustness in the SPD problem from an integrated engineering and marketing perspective and, in particular, they consider uncertainty in consumer preferences that arises from uncertainty in engineering parameters (for example, consumers may have different use requirements for the product, which can lead to variations in the product's utility), as well as small conjoint sample sizes. The latter form of uncertainty is related to the parameter uncertainty that we study in the present paper and as such, our paper complements this existing work in presenting a methodology for addressing parameter uncertainty in preferences in the design of product lines, as opposed to single products. Our paper is further differentiated from this existing work in robust SPD by also considering structural uncertainty in preferences, which is not considered in Luo et al. (2005) and Besharati et al. (2006).

Outside of the PLD literature, our treatment of parameter uncertainty is closely aligned with the large body of work within the operations research community on robust optimization (Bertsimas et al. 2011). Robust optimization is a technique for optimization under uncertainty, where one wishes to solve an optimization problem that is affected by uncertainty either through its constraints or its objective function. Robust optimization has been successfully applied in a variety of domain areas to address parameter uncertainty, including inventory management (Bertsimas and Thiele 2006), portfolio optimization (Bertsimas and Sim 2004, Goldfarb and Iyengar 2003), power system operations (Bertsimas et al. 2013), and facility location decisions (Baron et al. 2011). The only application of robust optimization techniques to a problem involving parametric choice models that we know of is the paper of Rusmevichientong and Topaloglu (2012). In this paper, the authors consider the problem of assortment optimization—a problem closely related to PLD—under the MNL model with uncertain utilities. They develop theoretical results that show that the robust MNL assortment problem can be solved efficiently; in particular, there exists a revenue-ordered assortment that is optimal, and therefore one can find the globally optimal assortment by enumerating N different assortments, where N is the number of products. Our paper contrasts with this work in a number of ways. First, our paper considers parameter uncertainty under a broader collection of choice models, namely, the first-choice model, the latent class MNL model, and the hierarchical Bayes mixture MNL model. Second, the theoretical results that are developed in Rusmevichientong and Topaloglu (2012) are tailored to the MNL model; in contrast, the uncertainty sets that we propose and the algorithm that we adapt to solve the robust problem (the divide-and-conquer algorithm; see Green and Krieger 1993, Belloni et al. 2008) are general purpose and can be used to address parameter uncertainty in a wide range of choice models. Finally, our paper also considers structural uncertainty, where the uncertainty can range over different parametric model structures, which is not considered in Rusmevichientong and Topaloglu (2012).

Our treatment of structural uncertainty was inspired by the paper of Bertsimas et al. (2016). The paper considers the problem of selecting a prostate cancer screening strategy to maximize a patient's expected quality adjusted life years. This problem is made difficult by the existence of three distinct, yet clinically recognized, models of how a patient's health evolves under different screening strategies. To reconcile these different models, the authors propose optimizing a weighted combination of the worst-case patient outcome and the average outcome over the three models. The authors show that the resulting screening strategy leads to good outcomes under all three models and is superior to the single-model screening strategies in its worst- and average-case performances. The results in the present paper complement those in Bertsimas et al. (2016) by showing that this idea can have significant impact in product line decisions.

3. Model

3.1. Nominal Model

Let n be the number of different attributes of the product. We assume, for simplicity, that each attribute is a binary attribute, i.e., either the product possesses the attribute or not. (This assumption is without loss of generality, as attributes with more than two levels can be modeled by, for example, introducing a binary attribute for each level of the original attribute and requiring that the product only possesses one of the binary attributes.) A product is then a binary vector $\mathbf{a} = (a_1, a_2, \ldots, a_n)$. We let N denote the number of candidate products and we index these products from 1 to N; for $p \in \{1, \ldots, N\}$, $\mathbf{a}^p$ indicates the attribute vector that encodes product p.

We assume that the marginal revenue of product p is given by $r(p)$. To model the market, we assume that each customer makes exactly one choice: he either chooses one of the products in the product line, or he chooses a "no-purchase" option, which corresponds to not purchasing a product from the product line (e.g., because of the existence of a more desirable product by a competing firm). The purchasing behavior of customers is described by a choice model m; the quantity $m(p \mid S)$ denotes the probability that the customer selects product p when offered the product line $S \subseteq \{1, \ldots, N\}$. We use the index 0 to indicate the no-purchase option, so that $m(0 \mid S)$ denotes the probability that the customer does not purchase any of the products in S. The expected revenue of product line S is then given by

$$R(S; m) = \sum_{p \in S} m(p \mid S) \cdot r(p). \tag{1}$$

Let P be the number of products that are to compose the product line (the "width" of the product line). We assume that P is predefined by the firm and is *not* a decision variable. The nominal PLD problem can then be stated as

$$\underset{S \subseteq \{1, \ldots, N\}: |S| = P}{\text{maximize}} \; R(S; m). \tag{2}$$

In words, this is the problem of finding a product line S consisting of P products from the universe of products $\{1, \ldots, N\}$ that leads to the highest possible expected revenue under the choice model m.

There are many possible choices for the choice model m; we present three examples as follows:

1. *First-choice*. We assume that there are finitely many customer types; let K denote the number of customer types. We assume that K is predefined by the firm and is not a decision variable. We let λ^k denote the probability that a random customer is of type k. We thus have that $\sum_{k=1}^{K} \lambda^k = 1$ and $\lambda^k \geqslant 0$ for all $k \in \{1, \ldots, K\}$. For each customer-type $k \in \{1, \ldots, K\}$, we let $u^k(p)$ denote the utility of product $p \in \{1, \ldots, N\}$ and we let $u^k(0)$ denote the utility of the no-purchase option for type k. We assume that the utilities are unique: if $p \neq p'$, then $u^k(p) \neq u^k(p')$, i.e., two products cannot have the same utility. Such an assumption is helpful in ensuring that the first-choice model is well defined. Given that the products are defined by discrete attributes, this assumption is not particularly restrictive.

With these definitions, the choice probability $m(p \mid S)$ can be defined as

$$m(p \mid S) = \sum_{k=1}^{K} \lambda^k \cdot \mathbb{I}\Big\{ p = \underset{p' \in S \cup \{0\}}{\arg\max}\, u^k(p') \Big\}, \tag{3}$$

and the expected revenue $R(S; m)$ is just

$$R(S; m) = \sum_{p \in S} \left(\sum_{k=1}^{K} \lambda^k \cdot \mathbb{I}\Big\{ p = \underset{p' \in S \cup \{0\}}{\arg\max}\, u^k(p') \Big\} \right) \cdot r(p). \tag{4}$$

We assume (as is commonly done in conjoint analysis) that the utility of a product p is a linear function of its attributes; that is,

$$u^k(p) = \sum_{i=1}^{n} u_i^k \cdot a_i^p, \tag{5}$$

where u_i^k is the partworth utility of attribute i for customer-type k. We will use $\mathbf{u}$ to denote the vector of attribute partworths, i.e., $\mathbf{u}^k = (u_1^k, \ldots, u_n^k)$.

2. *Latent class multinomial logit*. We now define the latent class multinomial logit (LCMNL) model. As in the first-choice model, we assume that there are K customer segments. A customer belongs to segment k with probability λ^k. Each product $p \in \{1, \ldots, N\}$ provides some utility $u^k(p)$ to customers in segment k and as before, we assume that $u^k(0)$ is the utility of the no-purchase option. The choice probability $m(p \mid S)$ is then given by

$$m(p \mid S) = \sum_{k=1}^{K} \lambda^k \frac{\exp(u^k(p))}{\sum_{p' \in S} \exp(u^k(p')) + \exp(u^k(0))} \tag{6}$$

and the expected revenue under m is given by

$$R(S; m) = \sum_{p \in S} \sum_{k=1}^{K} \lambda^k \frac{r(p) \cdot \exp(u^k(p))}{\sum_{p' \in S} \exp(u^k(p')) + \exp(u^k(0))}. \tag{7}$$

As in the first-choice model, we will assume that the utility $u^k(p)$ is a linear function of the attributes of product p (cf. Equation (5)) and use $\mathbf{u}$ to denote the vector of attribute partworths, i.e., $\mathbf{u}^k = (u_1^k, \ldots, u_n^k)$.

3. *Mixture multinomial logit*. In the mixture multinomial logit (MMNL) model, we assume that customers choose according to an MNL model, where the partworth vector $\mathbf{u} = (u_1, \ldots, u_n)$ is distributed according to some mixture distribution F. The choice probabilities are then given by

$$m_F(p \mid S) = \int \frac{\exp(u(p))}{\sum_{p' \in S} \exp(u(p')) + \exp(u(0))} \, dF(\mathbf{u}), \tag{8}$$

where $u(p)$ is the utility of product p under the partworth vector $\mathbf{u}$. Note that when F is a distribution with finite support, the choice model reduces to the LCMNL model (6). It is common to assume that $\mathbf{u}$ is normally distributed with mean $\boldsymbol{\beta}$ and covariance matrix $\mathbf{V}$. Thus, given $\boldsymbol{\beta}$ and $\mathbf{V}$, the corresponding choice probability can be written as

$$m_{\boldsymbol{\beta}, \mathbf{V}}(p \mid S) = \int \frac{\exp(u(p))}{\sum_{p' \in S} \exp(u(p')) + \exp(u(0))} \cdot \phi(\mathbf{u}; \boldsymbol{\beta}, \mathbf{V}) \, d\mathbf{u}, \tag{9}$$

where $\phi(\cdot;\boldsymbol{\beta},\mathbf{V})$ is the density function of a multivariate normal random variable with mean $\boldsymbol{\beta}$ and covariance matrix $\mathbf{V}$.

Assuming a mixture distribution that is a multivariate normal, there are two approaches to estimating $\boldsymbol{\beta}$ and $\mathbf{V}$. The first approach is based on expectation maximization (EM) (see Train 2009, 2008), which produces point estimates of $\boldsymbol{\beta}$ and $\mathbf{V}$. The second approach, which is more popular in marketing science and is more widely used in practice, is based on a hierarchical Bayes (HB) formulation of the model (Allenby and Rossi 1998, Rossi and Allenby 2003, Rossi et al. 2012). In the HB approach, $\boldsymbol{\beta}$ and $\mathbf{V}$ are treated as random variables with prior distributions, and the goal is to compute the posterior distributions of $\boldsymbol{\beta}$ and $\mathbf{V}$ given the available choice data from a conjoint analysis of a group of respondents. One of the most common specifications of the parameter priors (see Allenby and Rossi 1998, Rossi et al. 2012) is given by

$$\boldsymbol{\beta} \sim N(\bar{\boldsymbol{\beta}}, \alpha\mathbf{V}), \tag{10a}$$

$$\mathbf{V} \sim IW(v_0, \mathbf{V}_0), \tag{10b}$$

where $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and $IW(\nu, \mathbf{W})$ indicates the inverse Wishart distribution with degrees of freedom ν and scale matrix $\mathbf{W}$. Typically, the parameters of the priors of $\boldsymbol{\beta}$ and $\mathbf{V}$ are specified to ensure that the priors are relatively diffuse. In what follows, we will use HB-MMNL to denote the MMNL model with the above hierarchical Bayes specification (Equations (10a) and (10b)).

Letting Ψ denote the posterior joint probability density function of $(\boldsymbol{\beta}, \mathbf{V})$, it is common to compute the posterior mean of the choice probability $m_{\boldsymbol{\beta},\mathbf{V}}(p \mid S)$ (i.e., by integrating $m_{\boldsymbol{\beta},\mathbf{V}}(p \mid S)$ over Ψ), and to use this as the choice probability:

$$m(p \mid S) = \int\int \frac{\exp(u(p))}{\sum_{p' \in S}\exp(u(p')) + \exp(u(0))} \cdot \phi(\mathbf{u};\boldsymbol{\beta},\mathbf{V}) \cdot \Psi(\boldsymbol{\beta},\mathbf{V})\; d\mathbf{u}\, d\boldsymbol{\beta}\, d\mathbf{V}. \tag{11}$$

For most choice models m, the nominal PLD model (2) is a difficult optimization problem. For example, with the first-choice model, it is known that problem (2) is NP-hard (Kohli and Sukumar 1990). However, there exist many heuristics for solving problem (2) that are able to find high-quality solutions in practical runtimes (Belloni et al. 2008). One such heuristic is the "divide-and-conquer" heuristic (Green and Krieger 1993). In this heuristic, the product line is broken into groups of attributes, and the procedure sequentially optimizes each group of attributes to improve the objective function. We follow the implementation of this heuristic from Belloni et al. (2008) in treating each product as a group of attributes; thus the procedure changes the product line one product at a time until the revenue cannot be improved any further. Algorithm 1 provides a pseudocode description of the heuristic. This heuristic is guaranteed to determine a locally optimal product line, but is not guaranteed to find a globally optimal product line. However, Belloni et al. (2008) showed that this heuristic is able to quickly deliver very high-quality solutions in the first-choice PLD problem, where quality was measured relative to the true optimal solution.

Algorithm 1 (Divide-and-conquer heuristic for the nominal PLD problem (2))

Require: Choice model m, width of product line P, initial product line S
  Set **isLocalOptimal** = **false**
  while ¬**isLocalOptimal do**
    for all $p \in S$ **do**
      Set $\mathscr{S}_p = \{S' \mid S' = S \cup \{p'\} \setminus \{p\}$ for some $p' \in \{1, \ldots, N\} \setminus S\}$
      Set $S^*_p = \arg\max_{S \in \mathscr{S}_p} R(S; m)$
      if $R(S^*_p; m) > R(S; m)$, **then**
        Set $S = S^*_p$
        break
      end if
    end for
    if $R(S^*_p; m) \leqslant R(S; m)$ for each $p \in S$, **then**
      Set **isLocalOptimal** = **true**
    end if
  end while
  return S.

3.2. Robust Model

In the nominal PLD problem (2), we optimize the expected per -customer revenue with respect to the choice model m. In doing so, we tacitly assume that this choice model is known precisely; that is, the choice model m that we plan for in problem (2) is exactly the choice model that will be realized when the product line is offered to the market. In reality, however, we normally do not know the choice model precisely and m is thus only an estimate of the true choice model; in keeping with the nomenclature of robust optimization, m is the *nominal* choice model. This uncertainty constitutes a significant risk to the firm. If the nominal choice model m is not accurate and the choice model that is realized is different from m, then the revenue that is realized from the product line may be significantly lower than the anticipated revenue.

In this section, we propose an approach that directly accounts for the uncertainty in m. Rather than assuming that m is the true choice model, we can instead assume that we know a set of models $\mathscr{M}$, called the *uncertainty set*, which we believe contains the true choice model. This set may consist of models with the same parametric structure but different parameters, or models that have entirely different parametric structures; we discuss ways to construct $\mathscr{M}$ in more detail in Section 3.3. With $\mathscr{M}$ in hand, we proceed as follows: rather than solving a nominal optimization problem where we maximize the expected per-customer revenue with respect to the nominal m, we instead solve a *robust* optimization problem where we maximize the worst-case expected per-customer revenue, where the worst case is taken over all the possible choice models $\tilde{m}$ in the uncertainty set $\mathcal{M}$. Mathematically, the worst-case expected revenue of a product line S is defined as

$$R(S; \mathcal{M}) = \min_{\tilde{m} \in \mathcal{M}} R(S; \tilde{m}). \quad (12)$$

i.e., the worst-case expected per-customer revenue is the *minimum* of the expected per-customer revenue over the possible choice models in $\mathcal{M}$. It is helpful to think of $\tilde{m}$ as being controlled by an adversary ("nature") that wishes to reduce the expected per-customer revenue that the decision maker garners; given the freedom to choose any $\tilde{m}$ from $\mathcal{M}$, nature will select the one that makes the expected per-customer revenue the lowest. For further background in robust optimization, the reader is referred to the review paper of Bertsimas et al. (2011).

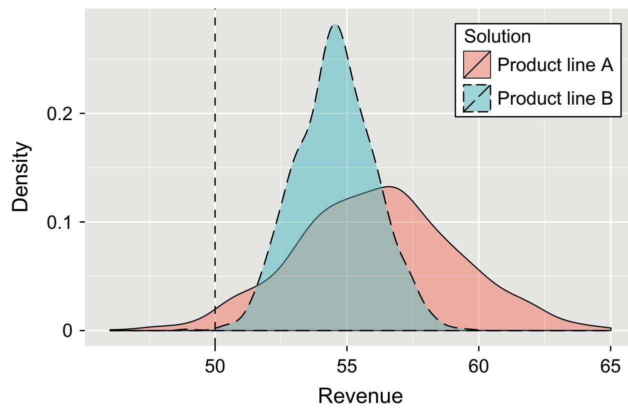
The robust PLD problem can then be defined as

$$\underset{S \subseteq \{1,\ldots,N\}: |S|=P}{\text{maximize}} \; R(S; \mathcal{M}). \quad (13)$$

Note that, like the nominal PLD problem (2), the robust PLD problem (13) is still, in general, a difficult problem to solve to provable optimality. However, many of the same heuristics that are available for solving the nominal PLD problem (2) can be used to solve this problem. In particular, we can use the divide-and-conquer heuristic (Algorithm 1) to solve the problem, where we replace evaluations of the nominal objective function $R(\cdot; m)$ with evaluations of the worst-case objective function $R(\cdot; \mathcal{M})$.

Before we describe the possible forms of the uncertainty set $\mathcal{M}$, it is worthwhile to highlight three important aspects of the robust model. First, we would like to provide some further motivation for the objective in problem (13). The objective function (12) still considers the *expected* revenue of the product line, but unlike the nominal objective function (1), it considers the *worst-case* expected profit of the product line, where the worst case is taken over the uncertainty set $\mathcal{M}$ of choice models. As discussed in Section 2, this contrasts with existing approaches to the PLD problem, all of which consider maximizing expected revenue or market share under a single nominal choice model. The reason for taking our view is that each model in $\mathcal{M}$ represents a plausible outcome with regard to how the customer population might react to the product line and thus, for a product line S, the set $\{R(S; \tilde{m}) \mid \tilde{m} \in \mathcal{M}\}$ represents a set of possible outcomes in terms of the expected per-customer revenue. Different product lines will differ in terms of the set of possible revenue outcomes they induce. A low revenue outcome may have significant negative consequences for the firm's financial viability and its perceived performance. Thus, rather than ensuring a good average revenue outcome over the models in $\mathcal{M}$, the firm may wish to ensure a good revenue outcome under the least favorable model. To illustrate this, consider Figure 1, which displays the hypothetical expected per-customer revenue distributions induced by a set of choice models $\mathcal{M}$ for two different product lines. The dashed vertical line indicates a target value for the expected per-customer revenue. Here, we can see that while product line A ensures a better average performance over the models in $\mathcal{M}$ than product line B, there are many outcomes where the revenue under product line A is below the target revenue, as indicated by the much heavier tail below the target revenue. Thus, in this setting, product line B may be more desirable to the firm than product line A. This is the typical behavior observed when comparing nominal and robust product lines (here, product line A's revenue distribution is representative of that of a nominal product line, while product line B is representative of the robust product line).

Figure 1. (Color online) Hypothetical illustration of revenue distributions under two different product lines.



The second aspect concerns how the solutions of the nominal problem (2) and the robust problem (13) relate to each other in terms of expected per-customer revenue. Let S^{N} be an optimal collection of products for the nominal problem (2) and let S^{R} be an optimal collection of products for the robust problem (13) with uncertainty set $\mathcal{M}$. Then, it follows that the worst-case expected per-customer revenue of the robust solution is always at least as good as the worst-case expected per-customer revenue of the nominal solution; that is,

$$R(S^{\mathrm{R}}; \mathcal{M}) \geqslant R(S^{\mathrm{N}}; \mathcal{M}).$$

To see this, observe that S^{N} is a feasible solution for problem (13); since S^{R} is an optimal solution for problem (13), it follows that S^{N} cannot have a higher value of $R(\cdot; \mathcal{M})$ as this would contradict the optimality of S^{R} for problem (13). A similar result follows in the other direction with respect to the nominal expected per-customer revenue: the nominal expected per-customer revenue of the nominal solution is always at least as good as the nominal expected per-customer revenue of the robust solution; that is,

$$R(S^{\mathrm{R}}; m) \leqslant R(S^{\mathrm{N}}; m).$$

In summary: the robust solution will perform better than the nominal solution in the worst case, while the nominal solution will perform better than the robust solution in the nominal case.

The third aspect of the robust model that is important to consider is the form and the size of the uncertainty set. As the size of the uncertainty set $\mathcal{M}$ increases, the minimization in (12) is taken over a larger set, and thus the corresponding solution of problem (13) will ensure the best worst-case performance over a larger set of possible values of $\tilde{m}$. However, as the uncertainty set increases in size, typically, the performance of the robust solution in the *nominal* case degrades. Thus, there is a trade-off in selecting the uncertainty set: if the set is too large and contains models $\tilde{m}$ that are significantly different from the nominal model m, then the robust solution will be protected against these realizations (relative to the nominal solution), but at the cost of performance under the nominal model m. In contrast, if the set is too small, then the robust solution will have better performance under the nominal model m, but will be vulnerable to extreme realizations of $\tilde{m}$.

3.3. Choices of the Uncertainty Set

We now discuss some ways that we may generate the uncertainty set $\mathcal{M}$.

1. *Finite set of variations of a single model*. Suppose that we fix a given type of model—for example, a latent-class multinomial logit model with K classes. We may then generate B different versions of the same type of model, leading to the uncertainty set

$$\mathcal{M} = \{m_1, \ldots, m_B\}. \tag{14}$$

These variations could be generated in a number of ways:

(a) *Bootstrapping*. We may generate B bootstrapped samples of the conjoint data by sampling with replacement from the respondents in the data set. We can then run the estimation procedure of our desired model class on the choice data of each bootstrapped sample, thus generating B different models.

(b) *Posterior sampling*. Bayesian models like HB will furnish us with a posterior distribution of a parameter $\boldsymbol{\theta}$ that specifies the model; for example, in the MMNL model in Equation (9), the underlying model is specified by the mean and covariance matrix pair $(\boldsymbol{\beta}, \mathbf{V})$, and in estimating the model, one computes the posterior distribution of $(\boldsymbol{\beta}, \mathbf{V})$. In a Bayesian setting, the parameter $\boldsymbol{\theta}$ is unknown, and our uncertainty in this parameter is captured in the posterior distribution of $\boldsymbol{\theta}$ given the available choice data. To form $\mathcal{M}$, we may therefore take B independent samples from the posterior distribution of $\boldsymbol{\theta}$—say, $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^B$—and set each m_b to be the choice model with parameter $\boldsymbol{\theta}^b$.

(c) *Multiple models from the estimation procedure*. It may be the case that the estimation procedure generates multiple models from the same data. For example, the EM procedure (Dempster et al. 1977), which is the typical estimation method of choice for latent-class multinomial logit models and other popular discrete choice models, may converge to B different local minima given different initial starting points that are similar in their log likelihood.

2. *Continuous set of variations of a single model*. Rather than considering a finite collection of B different versions of the same model, we may want to consider sets where one or more parameters that define the model vary continuously within some set. Letting $\boldsymbol{\theta}$ denote the parameter, Θ denote the set of possible values of $\boldsymbol{\theta}$ and $m_{\boldsymbol{\theta}}$ denote the model that corresponds to $\boldsymbol{\theta}$, $\mathcal{M}$ will be defined as

$$\mathcal{M} = \{m_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \Theta\}, \tag{15}$$

and correspondingly, $R(S; \mathcal{M})$ will be defined as

$$\begin{aligned} R(S; \mathcal{M}) &= \min_{\tilde{m} \in \mathcal{M}} R(S; \tilde{m}) \\ &= \min_{\boldsymbol{\theta} \in \Theta} R(S; m_{\boldsymbol{\theta}}). \end{aligned} \tag{16}$$

In the previous choice of the uncertainty set $\mathcal{M}$, where the uncertainty set $\mathcal{M}$ consists of finitely many models, the worst-case revenue $R(S; \mathcal{M})$ can be easily computed: simply compute $R(S; m)$ under each of the finitely many models m in $\mathcal{M}$ and take the minimum of this set of B values. In contrast, when the uncertainty set $\mathcal{M}$ consists of infinitely many models, the function $R(S; \mathcal{M})$ may not be easy to evaluate. As shown in Equation (16), the difficulty of computing $R(S; \mathcal{M})$ depends on how hard it is to optimize $R(S; m_{\boldsymbol{\theta}})$ as a function of the parameter $\boldsymbol{\theta}$ over the set of possible parameter values Θ.

One broad case where $R(S; \mathcal{M})$ is easy to compute is when $R(S; m_{\boldsymbol{\theta}})$ is a linear function of $\boldsymbol{\theta}$ and Θ is a polyhedron (that is, a set defined by finitely many linear equalities and inequalities on $\boldsymbol{\theta}$). In this case, $R(S; \mathcal{M})$ can be computed by solving a linear optimization problem, which is a theoretically tractable problem; this can also be done efficiently in practice provided that the dimension of $\boldsymbol{\theta}$ and the number of constraints defining Θ are not too large. An example of such a case is when we consider uncertainty in customer-type probabilities in the first-choice model. More concretely, we assume that the customer-type probability distribution $\tilde{\boldsymbol{\lambda}}$ is uncertain and belongs to some polyhedral set Λ. Letting $m_{\tilde{\boldsymbol{\lambda}}}$ denote the first-choice model with probability distribution $\tilde{\boldsymbol{\lambda}}$, we can write

$$\begin{aligned} R(S; \mathcal{M}) &= \min_{\tilde{\boldsymbol{\lambda}} \in \Lambda} R(S; m_{\tilde{\boldsymbol{\lambda}}}) \\ &= \min_{\tilde{\boldsymbol{\lambda}} \in \Lambda} \sum_{k=1}^{K} \tilde{\lambda}^k \left(\sum_{p \in S} \mathbb{I}\left\{ p = \operatorname*{arg\,max}_{p' \in S \cup \{0\}} u^k(p') \right\} \cdot r(p) \right). \end{aligned} \tag{17}$$

In Equation (17), we can see that the function being minimized is linear in $\boldsymbol{\lambda}$, and thus computing $R(S; \mathcal{M})$ amounts to solving a linear optimization problem. We now present two choices for the parameter uncertainty set Λ:

(a) *Box uncertainty set.* We assume that for each segment k, $\tilde{\lambda}^k$ is restricted to lie between $\underline{\lambda}^k$ and $\bar{\lambda}^k$. The *box* uncertainty set Λ_{box} is then given by

$$\Lambda_{\text{box}} = \left\{ \tilde{\boldsymbol{\lambda}} \in \mathbb{R}^K \;\middle|\; \begin{array}{l} \underline{\lambda}^k \leqslant \tilde{\lambda}^k \leqslant \bar{\lambda}^k, \quad \forall k \in \{1, \ldots, K\}, \\ \mathbf{1}^T \tilde{\boldsymbol{\lambda}} = 1, \\ \tilde{\boldsymbol{\lambda}} \geqslant \mathbf{0}. \end{array} \right\} \tag{18}$$

(b) *Aggregate deviation uncertainty set.* A limitation of the box uncertainty set is that the worst-case $\tilde{\boldsymbol{\lambda}}$—that is, the $\tilde{\boldsymbol{\lambda}}$ that achieves the worst-case revenue (17)—may be one where multiple $\tilde{\lambda}^k$'s simultaneously take their most extreme values (either $\underline{\lambda}^k$ or $\bar{\lambda}^k$). Such a $\tilde{\boldsymbol{\lambda}}$ may be highly unlikely. Consequently, the resulting robust solution will sacrifice performance on more likely $\tilde{\boldsymbol{\lambda}}$'s to be (unnecessarily) protected against such extreme values of $\tilde{\boldsymbol{\lambda}}$.

An alternative uncertainty set that may be more appropriate, then, is one that not only bounds how much each individual λ^k value deviates from its nominal value, but also bounds the *aggregate* deviation of the λ^k values from their nominal values. The resulting uncertainty set is the *aggregate deviation* uncertainty set, which is defined as

$$\Lambda_{\text{AD}} = \left\{ \tilde{\boldsymbol{\lambda}} \in \mathbb{R}^K \;\middle|\; \begin{array}{l} \sum_{k=1}^{K} |\tilde{\lambda}^k - \lambda^k| \leqslant \Gamma, \\ \underline{\lambda}^k \leqslant \tilde{\lambda}^k \leqslant \bar{\lambda}^k, \quad \forall k \in \{1, \ldots, K\}, \\ \mathbf{1}^T \tilde{\boldsymbol{\lambda}} = 1, \\ \tilde{\boldsymbol{\lambda}} \geqslant \mathbf{0}. \end{array} \right\} \tag{19}$$

where $\sum_{k=1}^{K} |\tilde{\lambda}^k - \lambda^k|$ is the aggregate deviation and Γ is a user-specified bound on this deviation.

In the context of the first-choice model, one may ask if it is possible to account for uncertainty in the partworths of the customer types through a continuous uncertainty set. While it is possible, it turns out to be rather difficult. To illustrate this difficulty, suppose that for each customer-type k, the set $\mathcal{U}^k$ denotes the uncertainty set for the partworth vector $\mathbf{u}^k$. Let $m_{\mathbf{u}^1, \ldots, \mathbf{u}^K}$ denote the choice model corresponding to the partworth vectors $\mathbf{u}^1, \ldots, \mathbf{u}^K$. The worst-case expected revenue is then

$$\begin{aligned} R(S; \mathcal{M}) &= \min_{(\tilde{\mathbf{u}}^1, \ldots, \tilde{\mathbf{u}}^K) \in \mathcal{U}^1 \times \cdots \times \mathcal{U}^K} R(S; m_{\tilde{\mathbf{u}}^1, \ldots, \tilde{\mathbf{u}}^K}) \\ &= \min_{(\tilde{\mathbf{u}}^1, \ldots, \tilde{\mathbf{u}}^K) \in \mathcal{U}^1 \times \cdots \times \mathcal{U}^K} \sum_{k=1}^{K} \lambda^k \\ &\quad \cdot \left(\sum_{p \in S} \mathbb{I}\left\{ p = \operatorname*{arg\,max}_{p' \in S \cup \{0\}} \tilde{u}^k(p') \right\} \cdot r(p) \right) \\ &= \sum_{k=1}^{K} \lambda^k \min_{\tilde{\mathbf{u}}^k \in \mathcal{U}^k} \left(\sum_{p \in S} \mathbb{I}\left\{ p = \operatorname*{arg\,max}_{p' \in S \cup \{0\}} \tilde{u}^k(p') \right\} \cdot r(p) \right). \end{aligned} \tag{20}$$

To compute the worst-case expected revenue in Equation (20), one must compute the minimum over each $\mathcal{U}^k$ of the expression $(\sum_{p \in S} \mathbb{I}\{p = \arg\max_{p' \in S \cup \{0\}} \tilde{u}^k(p')\} \cdot r(p))$. The minimization for each customer-type k is, in general, a difficult problem because the function being minimized is a nonconvex function of $\mathbf{u}^k$. For this reason, in our computational experiments in Section 4, we will not consider how to account for robustness under continuous partworth uncertainty.

3. *Finite set of multiple structurally distinct models.* Upon examining some conjoint data, we may estimate a handful of models that are structurally distinct but provide roughly the same quality of fit to the data—for example, a first-choice model m_{FC}, an LCMNL model with three segment m_{LC3} and an LCMNL model with eight segments m_{LC8}. The uncertainty set is then just the collection of these three models:

$$\mathcal{M} = \{m_{\text{FC}}, m_{\text{LC3}}, m_{\text{LC8}}\}.$$

3.4. Trading Off Nominal and Robust Performance

As presented in Sections 3.1 and 3.2, the firm is faced with a choice between two alternative optimization paradigms: optimizing for the nominal choice model m or optimizing for the worst-case choice model among an uncertainty set $\mathcal{M}$. While the first paradigm is undesirable in that it completely ignores uncertainty, the second paradigm may also be unappealing in that it ignores what may be the most "likely" model that describes the customer population.

One way to bridge the two extremes is to consider robustness as a constraint. In this approach, one would find the product line that maximizes the nominal revenue subject to a constraint that the worst-case revenue is no lower than some predefined amount. Mathematically, this problem can be stated as

$$\underset{S \subseteq \{1, \ldots, N\}: |S| = P}{\text{maximize}} \quad R(S; m) \tag{21a}$$

$$\text{subject to} \quad R(S; \mathcal{M}) \geqslant \underline{R} \tag{21b}$$

where $\underline{R}$ is a desired lower bound on the worst-case revenue.

Although conceptually appealing, problem (21) comes with two practical challenges. First, as noted in Section 3.1, one can use heuristics to solve the robust PLD problem (13) just as one can use them to solve the nominal PLD problem (2). However, to solve the constrained problem (21), existing PLD heuristics would need to be modified to ensure that the solution is feasible. With regard to the divide-and-conquer heuristic, a possible modification would be to consider a two-phase approach: in the first phase, we maximize $R(S; \mathcal{M})$ to try to find a feasible solution; then, assuming that we have found such a feasible solution, we maximize $R(S; m)$, taking care to only consider those solutions that are feasible (i.e., satisfying constraint (21b)) in each iteration.

Second, the firm may not have a single value of $\underline{R}$ in mind, but rather may be interested in seeing how the solution changes over a range of values of $\underline{R}$, to understand the trade-off between nominal and worst-case performance.

An alternative way to proceed in this case is to consider optimizing a weighted combination of the nominal and worst-case revenues, as follows:

$$\underset{S\subseteq\{1,\ldots,N\}:\,|S|=P}{\text{maximize}} \quad (1-\alpha)\cdot R(S;m)+\alpha\cdot R(S;\mathscr{M}). \tag{22}$$

Here, α is a weight between 0 and 1 that determines how much the objective emphasizes the worst-case revenue/de-emphasizes the nominal revenue. Note that unlike problem (21), problem (22) is an unconstrained problem that is amenable to existing PLD heuristics. By solving problem (22) for a range of values of α, it is possible to determine a Pareto efficient frontier of solutions that optimally trade-off worst-case revenue with nominal revenue.

4. Results

We now present the results of an extensive computational study with real conjoint data that illustrate (1) the need to account for uncertainty in PLD and (2) the benefits from adopting the approaches we prescribe. We begin by providing the background on the problem data in Section 4.1. We then present the results of several different experiments, which we summarize as follows:

- In Section 4.2, we consider how to account for uncertainty in customer-type probabilities/segment sizes in the first-choice model. We show that the revenue of the nominal product line can deteriorate quite significantly in the presence of uncertainty, while the robust product line outperforms the nominal product line when they are exposed to moderate to high levels of uncertainty.
- In Section 4.3, we consider how to account for uncertainty in the LCMNL model by constructing the uncertainty set using bootstrapping. We show that the nominal product line under the LCMNL model is extremely susceptible to uncertainty in the LCMNL model parameters and the robust product line provides an edge over the nominal product line in the worst case.
- In Section 4.4, we consider how to account for uncertainty in the HB-MMNL model by constructing the uncertainty set using samples from the posterior distribution of the mean and covariance parameters. We consider nominal solutions based on two different MMNL models—one using the posterior expectation of the choice probabilities (cf. (11)) and one based on a point estimate of the mean and covariance parameters—and we show that each one is outperformed by the robust solution in the worst case.
- In Section 4.5, we consider how to account for structural uncertainty within the LCMNL model. Here, the structural uncertainty comes from the number of segments K, which is unknown to the modeler. We show that when there is uncertainty in the right value of K, committing to any one value of K can result in worst-case revenue losses ranging from 3% to 7% if the true value of K is different. The robust product line, which does not assume a single value of K but protects against a range of values of K, improves on the worst-case performance of the nominal product lines corresponding to the same range of K values by an amount of approximately 2%–4%.
- Finally, in Section 4.6, we consider structural uncertainty under multiple structurally distinct models. The model uncertainty set consists of the first-choice model and two different LCMNL models. We show here that the three nominal solutions lead to significantly lower revenues when a model with a different structure is realized, with worst-case losses ranging from 23% to 37%. In contrast, the robust approach is able to identify a product line that provides essentially the same revenue under all three models, and improves on the worst-case revenue of each nominal product line's revenue by an amount ranging from 13% to 55%.

4.1. Background

For the experiments we conduct here, we use a real conjoint data set from the field test of Toubia et al. (2003). The field test from which the data set is derived is concerned with understanding consumer preferences for a hypothetical laptop bag to be offered by Timbuk2 (Timbuk2 Designs Inc., San Francisco, CA). The data set consists of responses from 330 respondents. The data set contains two relevant pieces of data:

1. *Pairwise comparisons*. As part of the conjoint study, each respondent was required to compare 16 pairs of hypothetical laptop bags and indicate which product was more preferred or whether the respondent was indifferent to the choice.

2. *Estimated partworth utilities for the first-choice model*. Using the pairwise comparison data, the data set provides estimates of each respondent's first-choice partworth vector using the analytic center method of Toubia et al. (2003). We use these partworths in our study of parameter uncertainty under the first-choice model in Section 4.2 and our study of structural uncertainty in Section 4.6.

The hypothetical laptop bag has 10 different attributes, including the price, whether the bag has a handle, and the color of the bag. We use the same revenue structure as Belloni et al. (2008), which previously used the data set of Toubia et al. (2003)—in particular, we assume the same marginal incremental revenues of Belloni et al. (2008). We also assume, as in Belloni et al. (2008), that the price varies from \$70 to \$100 in \$5 increments; this leads to $N = 3{,}584$ possible products.

We assume that the firm is interested in offering $P = 5$ versions of the laptop bag. As in Belloni et al. (2008), we assume that the no-purchase option involves the customer selecting one of three alternative products offered by the competition: (1) a bare bones bag that includes no optional features and that is priced at \$70; (2) a midrange bag that has five of the nine nonprice features and is priced at \$85; and (3) a high-end bag that has all of the features and is priced at \$100.

To solve each nominal and robust PLD problem, we execute the divide-and-conquer heuristic described as Algorithm 1 from 10 randomly chosen starting points, with the appropriate objective function (either a nominal objective function $R(S; m)$ or a worst-case objective function $R(S; \mathcal{M})$), and use the solution with the best objective value.

Our code, with the exception of the HB-MMNL model, was implemented in the Julia technical computing language (Bezanson et al. 2012, Lubin and Dunning 2015). All LCMNL models were estimated using a custom implementation of the EM algorithm (see Train 2008). All HB-MMNL models were estimated using the `Bayesm` package in R (Rossi 2012).

In the experiments that follow, we will focus on two useful metrics for quantifying the benefit of robustness. The first is the worst-case loss (WCL). The WCL is defined for a nominal model m, a nominal product line S^{N} and an uncertainty set $\mathcal{M}$ as

$$\mathrm{WCL}(S^{\mathrm{N}}, m, \mathcal{M}) = 100\% \times \frac{R(S^{\mathrm{N}}; m) - R(S^{\mathrm{N}}; \mathcal{M})}{R(S^{\mathrm{N}}; m)}. \quad (23)$$

The WCL measures how much the revenue of the nominal product line deteriorates when one passes from the nominal revenue under m to the worst-case revenue over $\mathcal{M}$; in other words, it measures how vulnerable the product line is to the worst-case model in $\mathcal{M}$.

The second metric we will consider is the relative improvement (RI) of the robust product line over the nominal product line. The RI is defined for a nominal product line S^{N}, a robust product line S^{R} and an uncertainty set $\mathcal{M}$ as

$$\mathrm{RI}(S^{\mathrm{R}}, S^{\mathrm{N}}, \mathcal{M}) = 100\% \times \frac{R(S^{\mathrm{R}}; \mathcal{M}) - R(S^{\mathrm{N}}; \mathcal{M})}{R(S^{\mathrm{N}}; \mathcal{M})}. \quad (24)$$

The RI measures how much the robust product line improves on the nominal product line in terms of the worst-case revenue, relative to the nominal product line's worst-case revenue.

4.2. Parameter Robustness Under the First-Choice Model

In this section, we demonstrate the importance of accounting for uncertainty in the first-choice model.

We begin by defining the nominal model. We use the form of the model given by Equation (4), where we take each respondent in the data set to represent a customer type. Thus our first-choice model has $K = 330$ customer types. We assume that each customer type is equiprobable, i.e., $\lambda^k = 1/K$. We use the estimated partworths from Toubia et al. (2003) to form the utility function $u^k(\cdot)$ of each customer-type $k \in \{1, \ldots, K\}$.

To model uncertainty, we will assume that the true probability distribution $\boldsymbol{\lambda}$ lies in a box uncertainty set Λ_{box} (cf. Equation (18)) where the lower and upper bars $\underline{\boldsymbol{\lambda}}$ and $\bar{\boldsymbol{\lambda}}$ are parametrized by a scalar $\epsilon \geqslant 0$ in the following way:

$$\underline{\lambda}^k = (1-\epsilon) \cdot \frac{1}{K}, \quad \forall\, k \in \{1, \ldots, K\}, \quad (25)$$

$$\bar{\lambda}^k = (1+\epsilon) \cdot \frac{1}{K}, \quad \forall\, k \in \{1, \ldots, K\}. \quad (26)$$

For a given ϵ, we indicate the corresponding box uncertainty set by $\Lambda_{\mathrm{box},\epsilon}$. Note that in this setting, we can interpret $\Lambda_{\mathrm{box},\epsilon}$ as a set of different weightings of the respondents. For example, with $\epsilon = 2$, each respondent can have a weight $\tilde{\lambda}^k$ as low as 0 and as high as $3/K$ (i.e., the respondent is weighted three times as much as in the nominal probability distribution $\boldsymbol{\lambda}$, where $\lambda^k = 1/K$). Such an uncertainty set could be used to guard against the possibility that some of the respondents are outliers and not representative of the overall customer population.

We then proceed as follows. We solve the nominal PLD problem (2) to obtain a nominal product line S^{N}. Then, for a given value of ϵ, we solve the robust PLD problem (13), where the set of models $\mathcal{M}$ is induced by the uncertainty set $\Lambda_{\mathrm{box},\epsilon}$, to obtain the robust product line $S^{\mathrm{R},\epsilon}$. For each ϵ, we compute the WCL of the nominal product line S^{N} and the RI of the robust product line corresponding to ϵ over the one nominal product line.

Table 1 shows how WCL and RI vary for values of $\epsilon \in \{0.1, 0.2, 0.5, 1, 2, 3, 4\}$. We can see from this table that for small amounts of uncertainty ($\epsilon < 0.5$), the revenue that is garnered under the nominal product line deteriorates moderately (e.g., for $\epsilon = 0.2$, the WCL is 2.38%), and the robust product line provides a slight edge. However, for moderate to large amounts of uncertainty ($\epsilon \geqslant 0.5$), the WCL increases significantly and the robust product line provides an edge over the nominal product line that grows as the amount of uncertainty increases. For example, with $\epsilon = 2$, the WCL is 10.11%, and the robust product line delivers a worst-case revenue that is 7.83% higher than that of the nominal product line.

4.3. Parameter Robustness Under the LCMNL Model

We now consider parameter robustness under the LCMNL model. In this set of experiments, we proceed as follows.

Table 1. WCL of nominal solution and the RI of robust solution over nominal solution for varying values of ϵ.

Uncertainty set $\Lambda_{\mathrm{box},\epsilon}$	$R(S^{\mathrm{N}}; \mathcal{M})$ (\$)	$R(S^{\mathrm{R},\epsilon}; \mathcal{M})$ (\$)	WCL (%)	RI (%)
$\epsilon = 0.0$	72.82	72.82	0.00	0.00
$\epsilon = 0.1$	71.92	71.92	1.23	0.01
$\epsilon = 0.2$	71.02	71.09	2.38	0.10
$\epsilon = 0.5$	68.31	69.12	5.08	1.19
$\epsilon = 1.0$	63.80	67.12	7.83	5.19
$\epsilon = 2.0$	60.70	65.46	10.11	7.83
$\epsilon = 3.0$	57.67	64.88	10.91	12.50
$\epsilon = 4.0$	55.46	64.47	11.47	16.24

Table 2. Comparison of nominal and worst-case revenues for LCMNL model under bootstrapping for $K \in \{1, \ldots, 10\}$.

No. segments K	$R(S^N; m)$ ($)	$R(S^N; \mathcal{M})$ ($)	$R(S^R; \mathcal{M})$ ($)	WCL (%)	RI (%)
$K = 1$	67.97	64.15	64.49	5.62	0.53
$K = 2$	66.40	60.03	60.77	9.59	1.22
$K = 3$	65.34	58.06	60.49	11.14	4.17
$K = 4$	64.87	59.40	59.61	8.43	0.36
$K = 5$	64.90	55.86	58.82	13.93	5.30
$K = 6$	66.19	55.50	58.50	16.16	5.42
$K = 7$	64.87	56.22	58.85	13.34	4.68
$K = 8$	67.02	51.60	58.15	23.02	12.70
$K = 9$	65.14	54.54	57.96	16.28	6.27
$K = 10$	65.49	50.62	57.62	22.70	13.82

For a fixed number of customer classes K, we estimate the nominal LCMNL model m using the pairwise comparison data set from all 330 respondents. Then, we generate a family of B LCMNL models $m_1, \ldots, m_B$ by bootstrapping. Each bootstrapped model is estimated from a bootstrapped data set that is generated by randomly sampling 330 respondents with replacement from the original set of 330 respondents. Note that by applying this procedure, the bootstrapped models effectively allow us to account for the uncertainty in the segment probabilities $\lambda^1, \ldots, \lambda^K$ and the segment-specific partworth vectors $\mathbf{u}^1, \ldots, \mathbf{u}^K$ jointly. We solve the nominal PLD problem (2) with the nominal model m to obtain the nominal product line S^N. We set $\mathcal{M} = \{m_1, \ldots, m_B\}$ and solve the robust PLD problem (13) with $\mathcal{M}$ to obtain the robust product line S^R.

We consider values of $K \in \{1, \ldots, 10\}$. We consider $B = 100$ bootstrapped models. For each estimation (for the nominal model and the B bootstrapped models), we run the EM algorithm from five different randomly generated starting points, and use the model with the highest log likelihood.

Table 2 compares the nominal revenues and the worst-case revenues over $\mathcal{M}$ of the two product lines for each value of K. We can see that under the worst-case model from the bootstrapped collection of models, the realized revenue can deteriorate significantly; for example, with $K = 5$ segments, the expected per-customer revenue is \$64.90 in the nominal case and \$55.86 in the worst case, which is a loss of more than 13%. Furthermore, we can see that in the worst case, the robust product line is able to offer a significant improvement over the nominal product line, ranging from 0.36% ($K = 4$) to as much as 13.82% ($K = 10$).

To visualize the variability of revenues under each product line, Figure 2 plots a smoothed histogram of the revenue under the nominal and robust product lines for $K = 8$. The revenue distribution is formed by the B bootstrapped models in $\mathcal{M}$. We can see that the mean of the robust distribution is less than the mean of the nominal distribution, but the robust distribution has a lighter tail to the left and is more concentrated around its mean. Thus, if we believe that the bootstrapped models in $\mathcal{M}$ are all models that could be realized, then the robust product line will exhibit less risk than the nominal product line.

One interesting insight that emerges from Table 2 is that the RI is generally higher with higher values of K. This makes sense, as the estimated parameter values of more complex models will be more sensitive to the underlying data used to estimate the model. By bootstrapping the data, there will be more variability in the family of models $\mathcal{M}$, and the robust product line S^R may therefore offer a greater edge over the nominal product line S^N in the worst case. At the same time, in practice, more complex models (LCMNL models with higher values of K) may offer a better fit to the data than simpler models (LCMNL models with lower values of K). Therefore, robustness will become more desirable if we believe that we should use a large number of customer classes to describe our customer population.

It is also interesting to consider how these results depend on sample size, specifically the number of respondents in the conjoint study. If we imagine performing the same experiment with a varying number of respondents M, we would expect to see greater variability in the bootstrapped LCMNL models and more dispersed revenue distributions for smaller values of M. To test this, we consider $K = 2$ customer classes and we fix a value of M; we use the

Figure 2. (Color online) Plot of revenues under nominal and robust product lines under bootstrapped LCMNL models with $K = 8$.

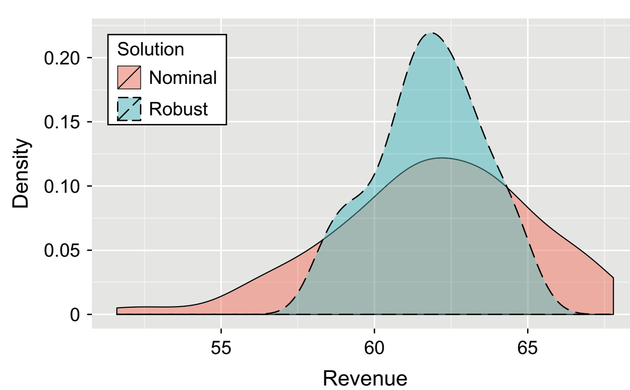
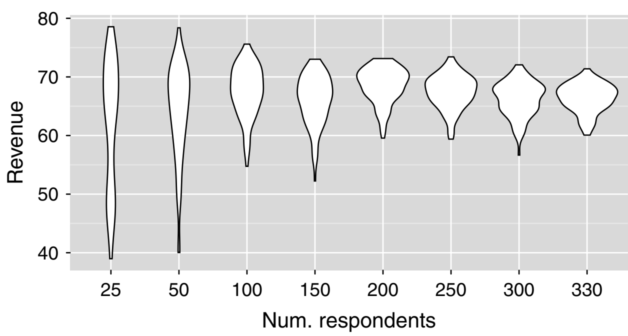


Figure 3. Plot of revenue distributions induced by bootstrapped LCMNL models with $K = 2$ for nominal product line as the number of respondents M varies.



first M respondents in the data set to estimate the LCMNL model and find the nominal product line. We then draw $B = 100$ bootstrap samples, estimate an LCMNL model on each bootstrap sample, and compute the distribution of revenues induced by this bootstrapped collection of models for the nominal product line. Figure 3 plots the revenue distributions obtained from this procedure for values of M ranging from 25 to 330 (the full set of respondents). From this plot, we can see that at low sample sizes, there is high variability in the revenues; at higher sample sizes, this variability is greatly reduced. This suggests that robustness is particularly valuable in settings where the number of respondents is low, and becomes a less salient issue when one has a large number of respondents.

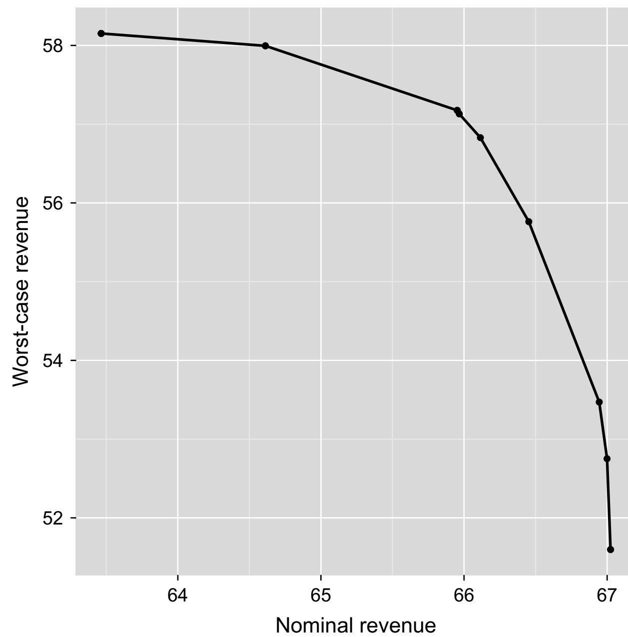
We conclude our study of parametric robustness under the LCMNL model by illustrating the trade-off between nominal and worst-case revenues using the weighted combination approach described Section 3.4. To demonstrate this, we focus on the LCMNL model with $K = 8$ and solve problem (22) for α values in

$$\{0.01, 0.02, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 0.98, 0.99\}.$$

Figure 4 plots each solution of each α value as a point in two-dimensional space, where the x-axis indicates the nominal revenue of the solution and the y-axis indicates the worst-case revenue over $\mathcal{M}$ of the solution; the points are joined together by straight lines to form an approximate Pareto efficient frontier of solutions that are undominated in nominal and worst-case revenue.

From Figure 4, we can see that there is a wide range of solutions between the robust solution (top left end of frontier) and nominal solution (bottom right end of frontier) and that the frontier suggests some profitable trade-offs. For example, we can see that there exist solutions that are very similar to the nominal solution in terms of nominal revenue but whose worst-case revenues are considerably higher. The main takeaway from this plot is that by considering problem (22), the firm can evaluate the range of solutions that

Figure 4. Plot of approximate Pareto efficient frontier of solutions that trade-off nominal revenue and worst-case revenue under the bootstrapped uncertainty set $\mathcal{M}$ for $K = 8$.



are efficient with regard to nominal and worst-case revenue and decide which point on the frontier is most desirable to them.

4.4. Parameter Robustness Under the HB-MMNL Model

In this section, we consider how to incorporate robustness under the HB-MMNL model using posterior sampling. We use the prior specification in Equation (10b) for the priors of the mean $\boldsymbol{\beta}$ and covariance matrix $\mathbf{V}$. We obtain B independent samples $(\boldsymbol{\beta}^1, \mathbf{V}^1), \ldots, (\boldsymbol{\beta}^B, \mathbf{V}^B)$ from the posterior distribution of $(\boldsymbol{\beta}, \mathbf{V})$. We obtain these samples through Markov chain Monte Carlo (MCMC), using the `Bayesm` package in R (Rossi 2012). We ran the MCMC procedure for 100,000 iterations with a burn-in period of 50,000 iterations. Due to the high autocorrelation in the draws of the parameter values, we thin the remaining draws to only retain every 200th draw. The last $B = 100$ of the thinned draws of $(\boldsymbol{\beta}, \mathbf{V})$ are used to form the set $\{(\boldsymbol{\beta}^1, \mathbf{V}^1), \ldots, (\boldsymbol{\beta}^B, \mathbf{V}^B)\}$. We use the default values provided by `Bayesm` for the prior parameters in Equation (10b).

Ideally, we would like to consider the uncertainty set $\mathcal{M}$ that consists of $m_{\boldsymbol{\beta}^b, \mathbf{V}^b}$, as defined in Equation (9), for each sample $b \in \{1, \ldots, B\}$. Unfortunately, each choice model of the form $m_{\boldsymbol{\beta}, \mathbf{V}}$ requires the evaluation of an integral of an MNL choice probability with respect to a multivariate normal density, which cannot be computed in closed form. Thus, for each model $m_{\boldsymbol{\beta}^b, \mathbf{V}^b}$, we approximate the integral by drawing $T = 100$ samples from the multivariate normal density with mean $\boldsymbol{\beta}^b$ and covariance matrix $\mathbf{V}^b$; denoting the approximate choice model by $\hat{m}_{\boldsymbol{\beta}^b,\mathbf{V}^b}$, the choice probability of an option p given the product line S is then just

$$\hat{m}_{\boldsymbol{\beta}^b,\mathbf{V}^b}(p \mid S) = \sum_{t=1}^{T} \frac{1}{T} \frac{\exp(u^{b,t}(p))}{\sum_{p' \in S} \exp(u^{b,t}(p')) + \exp(u^{b,t}(0))}, \tag{27}$$

where $u^{b,t}(\cdot)$ is the utility function corresponding to the tth sample of the partworth vector $\mathbf{u}$ from the multivariate normal distribution corresponding to the bth posterior sample. Our final uncertainty set is thus

$$\mathcal{M} = \{\hat{m}_{\boldsymbol{\beta}^1,\mathbf{V}^1}, \ldots, \hat{m}_{\boldsymbol{\beta}^B,\mathbf{V}^B}\}. \tag{28}$$

We consider two types of nominal models. The first is the approximate posterior expectation (PostExp) MMNL choice model, which is defined as

$$\begin{aligned} m_{\text{PostExp}}(p \mid S) &= \frac{1}{B} \cdot \sum_{b=1}^{B} \hat{m}_{\boldsymbol{\beta}^b,\mathbf{V}^b}(p \mid S) \\ &= \frac{1}{BT} \sum_{b=1}^{B} \sum_{t=1}^{T} \frac{\exp(u^{b,t}(p))}{\sum_{p' \in S} \exp(u^{b,t}(p')) + \exp(u^{b,t}(0))}. \end{aligned} \tag{29}$$

The choice probabilities produced by $m_{PostExp}$ can be viewed as approximations of the true posterior expected choice probabilities given in Equation (11). There are two levels of approximation: the integral over the posterior density of $\boldsymbol{\beta}$ and $\mathbf{V}$ is approximated by the average of B samples from that posterior density, while the inner integral for each posterior sample is approximated by the average of T samples from the corresponding multivariate normal distribution.

The second nominal model we will consider is the approximate point estimate (PointEst) MMNL choice model. Here, we obtain the approximate posterior mean of the mean parameter $\boldsymbol{\beta}$, given by $\boldsymbol{\beta}^* = (1/B) \cdot \sum_{b=1}^{B} \boldsymbol{\beta}^b$, and the covariance matrix, given by $\mathbf{V}^* = (1/B) \cdot \sum_{b=1}^{B} \mathbf{V}^b$. We then plug these point estimates into the definition of the MMNL model in Equation (9). Since the integral that defines Equation (9) cannot be computed in closed form, we approximate the integral with $T = 100$ samples from the corresponding multivariate normal density. Letting $u^{*,t}(\cdot)$ denote the utility function corresponding to the tth sample from the multivariate normal density with mean $\boldsymbol{\beta}^*$ and covariance matrix $\mathbf{V}^*$, we define the approximate point estimate MMNL choice model as

$$m_{\text{PointEst}}(p \mid S) = \frac{1}{T} \sum_{t=1}^{T} \frac{\exp(u^{*,t}(p))}{\sum_{p' \in S} \exp(u^{*,t}(p')) + \exp(u^{*,t}(0))}. \tag{30}$$

This approach of replacing the posterior distribution over $\boldsymbol{\beta}$ and $\mathbf{V}$ by a point estimate is sometimes referred to as the "plug-in" Bayes approach (see Rossi and Allenby 2003).

We optimize each of the resulting objective functions, $R(\cdot; \mathcal{M})$, $R(\cdot; m_{\text{PostExp}})$, and $R(\cdot; m_{\text{PointEst}})$, to obtain the product lines S^{R}, $S^{\text{N}, m_{\text{PostExp}}}$, and $S^{\text{N}, m_{\text{PointEst}}}$, respectively. For each nominal solution, we compute the WCL and the RI of the robust solution under the uncertainty set $\mathcal{M}$.

Table 3. Comparison of solutions under nominal HB models m_{PointEst} and m_{PostExp} to robust solution under uncertainty set $\mathcal{M}$ formed by posterior sampling.

Model m	$R(S^{\text{N},m}; m)$ (\$)	$R(S^{\text{N},m}; \mathcal{M})$ (\$)	WCL (%)	RI (%)
m_{PointEst}	61.33	54.88	10.52	6.25
m_{PostExp}	62.44	56.89	8.89	2.50

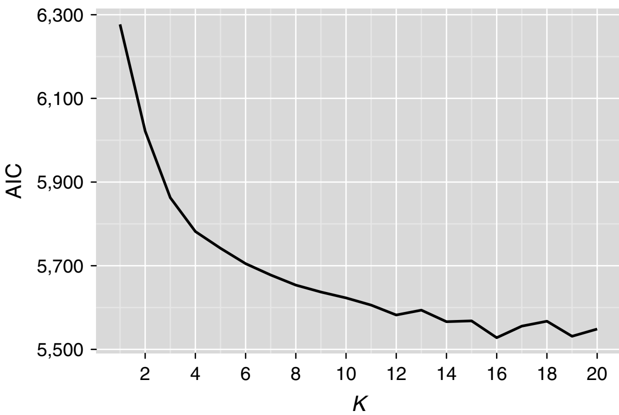
Table 3 shows the nominal and worst-case revenue under $\mathcal{M}$ of the nominal product line under each of the two nominal HB models, as well as the WCL and RI for each nominal product line. The worst-case revenue of the robust product line is \$58.31. From this table, we can see that the nominal product lines obtained using the PostExp model m_{PostExp} and the PointEst model m_{PointEst} are highly susceptible to uncertainty as represented by the uncertainty set $\mathcal{M}$, which is informed by the posterior distribution of the partworth mean vector $\boldsymbol{\beta}$ and the covariance matrix $\mathbf{V}$. At the same time, the robust solution is able to improve the worst-case revenue under $\mathcal{M}$ significantly in each case (more than 6% for the PointEst model and by 2.5% for the PostExp model).

4.5. Structural Robustness Under Different LCMNL Models

One of the difficulties of using the LCMNL model in practice is that we do not know a priori what the right number of customer segments K is. To find the number of segments, one typically estimates the LCMNL model for a range of values of K and then selects the "best" model from this set of models. What is meant by "best" is usually a model that strikes a good balance between fit (as measured by the log likelihood) and complexity (as measured by the number of parameters). While there exist metrics such as the Akaike information criterion (AIC) (Akaike 1974), the Bayesian information criterion (Schwarz 1978), or the consistent Akaike information criterion (CAIC) (Bozdogan 1987) that one can use to quantify the quality of a model and to guide the selection of a best model (e.g., select the model with the best AIC value), these metrics are not infallible and can often lead to different choices of a best model. In the end, one may only be able to delineate a set of "good" models as opposed to a single "best" model.

To illustrate this with the conjoint data set of Toubia et al. (2003), we estimate the LCMNL model using the pairwise comparison data for different values of K in $\{1, 2, \ldots, 20\}$. Figures 5 and 6 display the AIC and CAIC, respectively, against the number of customer segments K. From Figure 6, we can see that the CAIC implies a choice of K of either 3 or 4. (The minimum CAIC is 6,103.17 at $K = 3$; the second lowest is 6,104.89 for $K = 4$.) In contrast, from Figure 5, we can see that AIC continues decreasing as K increases beyond $K = 3$ until it begins to settle down, starting at approximately $K = 12$. Thus, based on the AIC,

Figure 5. AIC for $K \in \{1, \ldots, 20\}$.



it would seem that $K = 12$ would be a more appropriate choice. Based on these plots, we could conclude that any value of K between 3 and 12 is a reasonable value to select.

We now demonstrate how robustness can be used to counteract our uncertainty in the number of segments in the LCMNL model. Given the plots of the AIC and CAIC in Figures 5 and 6, we may decide that the number of segments K should lie in the range between 3 and 12. Thus, letting $m_{\mathrm{LC}K}$ denote the fitted LCMNL model with K customer segments, we would define our uncertainty set to be

$$\mathcal{M} = \{m_{\mathrm{LC3}}, m_{\mathrm{LC4}}, \ldots, m_{\mathrm{LC12}}\}. \tag{31}$$

We would then solve the robust problem (13) with $\mathcal{M}$ to obtain a robust product S^{R}. To compare this solution against the nominal approach, we solve the nominal problem (2) for each $m_{\mathrm{LC}K} \in \mathcal{M}$ to obtain a product line $S^{\mathrm{N},K}$ for each $K \in \{3, \ldots, 12\}$, and evaluate the worst-case revenue $R(S^{\mathrm{N},K}; \mathcal{M})$ for each such product line.

Table 4 provides the results of this comparison. The second column indicates each nominal product line's nominal revenue. The third column indicates each nominal product line's *worst-case* revenue over $\mathcal{M}$. The fourth column displays the WCL of each nominal product line, while the

Figure 6. CAIC for $K \in \{1, \ldots, 20\}$.

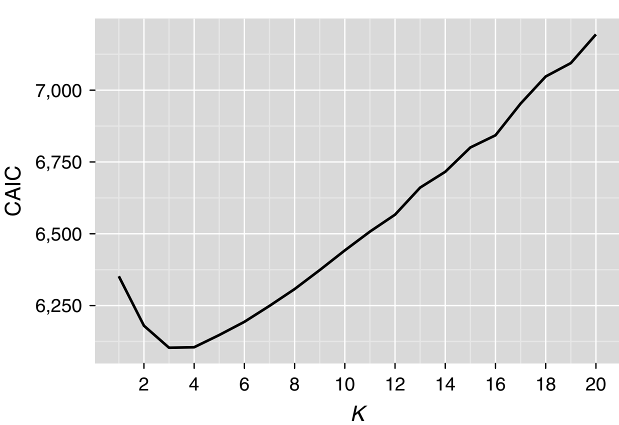


Table 4. Comparison of nominal and worst-case revenues of product lines $S^{\mathrm{N},3}, \ldots, S^{\mathrm{N},12}$.

No. segments K	$R(S^{\mathrm{N},K}; m_{\mathrm{LC}K})$ (\$)	$R(S^{\mathrm{N},K}; \mathcal{M})$ (\$)	WCL (%)	RI (%)
$K = 3$	65.34	61.32	6.17	4.15
$K = 4$	64.87	61.56	5.10	3.74
$K = 5$	64.90	61.35	5.47	4.09
$K = 6$	66.19	62.58	5.46	2.05
$K = 7$	64.87	62.71	3.33	1.83
$K = 8$	67.02	62.33	7.01	2.46
$K = 9$	65.14	62.24	4.45	2.60
$K = 10$	65.49	62.33	4.82	2.46
$K = 11$	66.16	62.48	5.57	2.22
$K = 12$	65.80	61.59	6.41	3.69

fifth column displays the RI of the robust product line over each nominal product line with respect to the worst-case revenue. The worst-case revenue $R(S^{\mathrm{R}}; \mathcal{M})$ of the robust solution S^{R} is \$63.86.

From this table, we can see that the nominal product lines corresponding to $m_{\mathrm{LC3}}, \ldots, m_{\mathrm{LC12}}$ are quite susceptible to uncertainty. This is visible from the values of the WCL; the WCL ranges from 3.33% ($K = 7$) to as much as 7.01% ($K = 8$). At the same time, the robust solution S^{R} provides significant improvement over all of these nominal solutions in the worst case. In particular, for the two models m_{LC3} and m_{LC12} suggested by the CAIC and AIC metrics, respectively, the RIs are 4.15% and 3.69%.

The main takeaway from this analysis is that robust optimization provides an alternative to model selection. When the data set does not clearly imply one single "best" model, but instead implies multiple "reasonably good" models, robustness provides a means of simultaneously optimizing against all of these models, rather than optimizing only against a single one.

4.6. Structural Robustness Under Distinct Models

In Section 4.5, we considered how robust optimization can be used to account for uncertainty in the number of segments in the LCMNL model. In this section, we take this idea a step further by using robust optimization to account for our uncertainty over *structurally distinct* models. In particular, in a given application, we may develop models that make very different assumptions. For example, we may estimate a first-choice model, as in Section 4.2, and an LCMNL model as in Sections 4.3 and 4.5. As in Section 4.5, one model may not clearly stand out as the best model. At the same time, if we make a product line decision based on one of these models, the revenue that is garnered may be significantly lower than anticipated if the market responds according to one of the other models.

To demonstrate the value of robustness, we proceed as follows. We assume that after analyzing the conjoint data set, we identify the following set of models:

$$\mathcal{M} = \{m_{\mathrm{FC}}, m_{\mathrm{LC3}}, m_{\mathrm{LC12}}\}, \tag{32}$$

Table 5. Performance of nominal and robust product lines under the different models in $\mathscr{M}$, as well as the worst-case model.

Product line	Nominal revenue ($)			Worst-case revenue ($)	WCL (%)	RI (%)
	$R(\cdot; m_{\mathrm{FC}})$	$R(\cdot; m_{\mathrm{LC3}})$	$R(\cdot; m_{\mathrm{LC12}})$	$R(\cdot; \mathscr{M})$		
$S^{\mathrm{N},\, m_{\mathrm{FC}}}$	72.82	55.74	56.39	55.74	23.45	13.70
$S^{\mathrm{N},\, m_{\mathrm{LC3}}}$	40.98	65.34	61.32	40.98	37.28	54.64
$S^{\mathrm{N},\, m_{\mathrm{LC12}}}$	48.49	62.93	65.80	48.49	26.31	30.69
S^{R}	63.69	63.43	63.38	63.38	—	—

where m_{FC} is the nominal first-choice model from Section 4.2 constructed using the estimated partworth vectors of all 330 respondents and m_{LC3} and m_{LC12} are the LCMNL models with $K = 3$ and $K = 12$ segments, respectively, estimated from the pairwise comparison data. For each model $m \in \mathscr{M}$, we solve the nominal PLD problem (2) to obtain the nominal product line $S^{\mathrm{N},\, m}$. We compute the revenue of $S^{\mathrm{N},\, m}$ under each model $m' \in \mathscr{M}$, as well as its worst-case revenue over $\mathscr{M}$. Then, we solve the robust PLD problem (13) to obtain the robust product line S^{R}, and compute the revenue of S^{R} under each model $m' \in \mathscr{M}$, as well as its worst-case revenue. We compute the WCL of each model m and the RI of the robust product line S^{R} over the nominal product line for each model m.

Table 5 displays the nominal and worst-case revenues of each of the product lines. From this table, we can see that committing to a single model leads to significantly lower revenues under the other models. For example, if we assume the first-choice model m_{FC}, the nominal revenue is $72.82, but this reduces to $55.74 if the true model is the $K = 3$ segment LCMNL model, representing a loss of more than 20%. If we instead assume the $K = 3$ LCMNL model, the nominal revenue is $65.34, but this reduces to $40.98 if the realized model is the first-choice model (a loss of more than 35%). The WCL of all three nominal product lines is more than 20%, indicating that in the worst case, the revenue of each of these product lines deteriorates by more than 20%.

In contrast to the nominal solution, the robust solution is able to achieve better worst-case performance, ranging from a 13.70% improvement (over the first-choice solution) to 54.64% (over the m_{LC3} solution). Moreover, the revenue of the robust solution is strikingly consistent over all three models, ranging from $63.38 to $63.69. Although the robust solution is not able to simultaneously achieve the best performance that is possible for each of the three models, it is able to achieve good performance for all three models and achieves better aggregate performance than the solution of each individual model. In this way, robust optimization allows us to incorporate all of the models that we deem acceptable, rather than just one.

It is also interesting to examine the product lines themselves. Figures 7–9 display the nominal product lines for m_{FC}, m_{LC3}, and m_{LC12}, respectively, while Figure 10 displays the robust product line. The first column indicates the attribute and each subsequent column represents one of the products; in a given column, a shaded cell indicates that the product has the corresponding attribute. For example, the fourth product in the first-choice product line is priced at $100, and has a handle, a PDA holder, a cellphone holder, and a mesh pocket. For easier comparison, Figure 11 shows the competitive offerings; that is, the products comprising the no-purchase option in the same format.

From these figures, we can see that the three nominal product lines are very different in terms of the actual products. For example, the first-choice product line is structured in a way that it includes products that directly compete with the competitive offerings that comprise the no-purchase option. As a specific example, consider the third product, which is identical to the high-end bag (see Figure 11) in terms of the nonprice features, but is priced at $95 instead

Figure 7. First-choice product line, $S^{\mathrm{N},\, m_{\mathrm{FC}}}$.

Price	$100	$80	$95	$100	$70
Large size	■		■		
Red color			■		
Logo	■		■		
Handle	■		■	■	
PDA holder		■	■	■	
Cellphone holder		■	■	■	
Mesh pocket		■	■	■	
Velcro flap	■	■	■		■
Boot	■	■	■		

Figure 8. LCMNL model with $K = 3$ product line, $S^{\mathrm{N},\, m_{\mathrm{LC3}}}$.

Price	$100	$100	$100	$100	$100
Large size		■		■	■
Red color					■
Logo	■		■		■
Handle	■	■	■	■	■
PDA holder					
Cellphone holder					
Mesh pocket		■			■
Velcro flap	■	■		■	■
Boot	■	■	■	■	■

Figure 9. LCMNL model with $K = 12$ product line, $S^{\mathrm{N},\, m_{\mathrm{LC12}}}$.

Price	$100	$100	$100	$80	$100
Large size		■	■		■
Red color					■
Logo	■		■		■
Handle	■	■	■	■	■
PDA holder					■
Cellphone holder				■	■
Mesh pocket	■	■	■		■
Velcro flap		■	■	■	■
Boot	■	■	■		■

of $100 and is thus more attractive. In contrast to the first-choice product line, the LCMNL product lines have products that are different in terms of features and, in general, are priced higher. This difference in the product lines under the three models manifests itself in the results of Table 5, which shows how each nominal product line is suboptimal under one of the other two nominal models.

Comparing the nominal product lines to the robust product line in Figure 10, we can see that the robust product line contains three products found in the nominal single-model product lines. Specifically, the third product in the first-choice is identical to the third product in the robust product line; the second product in the LCMNL $K = 3$ product line and the fourth product in the robust product line are identical; and the third product in the LCMNL $K = 3$ product line and the fifth product in the robust product line are identical. Furthermore, the first and second products of the robust product line differ by one attribute from the third and first products, respectively, of the LCMNL $K = 12$ product line. Thus the robust product line could be viewed as effectively "blending" the three single-model product lines. This characteristic of the robust product line agrees with our intuition: by using products that are identical or very similar to products in the individual single-model product lines, it seems reasonable to expect that the robust product line should achieve good performance over all three models.

Figure 10. Robust product line, S^{R}.

Price	$100	$100	$95	$100	$100
Large size	■		■	■	
Red color			■		
Logo	■		■		■
Handle	■	■	■	■	■
PDA holder			■		
Cellphone holder			■		
Mesh pocket	■	■	■	■	
Velcro flap	■		■	■	
Boot		■	■	■	■

Figure 11. Competitive products.

Price	$70	$85	$100
Large size			■
Red color			■
Logo			■
Handle			■
PDA holder		■	■
Cellphone holder		■	■
Mesh pocket		■	■
Velcro flap		■	■
Boot		■	■

5. Conclusion

In this paper, we have extended the familiar nominal PLD problem to account for parameter uncertainty and structural uncertainty by adopting a robust optimization approach. With regard to parameter uncertainty, we showed through a number of numerical experiments using a real conjoint data set that parameter uncertainty under three different widely used models has the potential to greatly reduce realized revenues. We further showed how, using the approach we have proposed, it is possible to guard against uncertainty and to limit its impact on revenues. With regard to structural uncertainty, we demonstrated that different model structures can lead to significantly different product line decisions, and that each such product line decision may be highly suboptimal for a different model structure. We then showed how optimizing against the worst-case revenue over all of the structurally different models leads to product lines that achieve good performance over all of the models and improve on each single-model product line in the worst case.

A number of possibilities exist for further exploring the methodology presented here. Apart from the choice model, there exist many other sources of uncertainty. One such source of uncertainty is in the cost information; we may not precisely know the incremental marginal cost of each attribute. Another source of uncertainty is in the no-purchase behavior. In particular, we may not precisely know the competitive offerings that give rise to the no-purchase option; moreover, these competitive offerings may be decided in response to the product line that we offer. One could adapt the worst-case framework that we have described here to accommodate these considerations by populating the uncertainty set with models that correspond to different competitive response scenarios.

Acknowledgments

The authors thank the area editor Harikesh Nair, the former area editor Duncan Simester, the associate editor, and the two anonymous referees for their helpful and insightful comments that greatly improved the paper. The authors gratefully acknowledge the authors of Toubia et al. (2003) for graciously making their data set available, which was used in Section 4. The work of the second author was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada through a PGS-D award.

References

Akaike H (1974) A new look at the statistical model identification. *IEEE Trans. Automatic Control* 19(6):716–723.

Allenby GM, Rossi PE (1998) Marketing models of consumer heterogeneity. *J. Econometrics* 89(1):57–78.

Baron O, Milner J, Naseraldin H (2011) Facility location: A robust optimization approach. *Production Oper. Management* 20(5):772–785.

Belloni A, Freund R, Selove M, Simester D (2008) Optimizing product line designs: Efficient methods and comparisons. *Management Sci.* 54(9):1544–1552.

Bertsimas D, Sim M (2004) The price of robustness. *Oper. Res.* 52(1): 35–53.

Bertsimas D, Thiele A (2006) A robust optimization approach to inventory theory. *Oper. Res.* 54(1):150–168.

Bertsimas D, Brown DB, Caramanis C (2011) Theory and applications of robust optimization. *SIAM Rev.* 53(3):464–501.

Bertsimas D, Silberholz J, Trikalinos T (2016) Optimal healthcare decision making under multiple mathematical models: Application in prostate cancer screening. *Health Care Management Sci.*, ePub ahead of print September 17, https://doi.org/10.1007/s10729-016-9381-3.

Bertsimas D, Litvinov E, Sun XA, Zhao J, Zheng T (2013) Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Trans. Power Systems* 28(1):52–63.

Besharati B, Luo L, Azarm S, Kannan PK (2006) Multi-objective single product robust optimization: An integrated design and marketing approach. *J. Mechanical Design* 128(4):884–892.

Bezanson J, Karpinski S, Shah VB, Edelman A (2012) Julia: A fast dynamic language for technical computing. Preprint, arXiv:1209.5145.

Bozdogan H (1987) Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 52(3):345–370.

Chen KD, Hausman WH (2000) Technical note: Mathematical properties of the optimal product line selection problem using choice-based conjoint analysis. *Management Sci.* 46(2):327–332.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* 39(1):1–38.

Dobson G, Kalish S (1988) Positioning and pricing a product line. *Marketing Sci.* 7(2):107–125.

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

Goldfarb D, Iyengar G (2003) Robust portfolio selection problems. *Math. Oper. Res.* 28(1):1–38.

Green PE, Krieger AM (1985) Models and heuristics for product line selection. *Marketing Sci.* 4(1):1–19.

Green PE, Krieger AM (1993) Conjoint analysis with product-positioning applications. Eliashberg J, Lilien GL, eds. *Handbooks in Operations Research and Management Science*, Vol. 5 (Elsevier, Amsterdam), 467–515.

Kohli R, Sukumar R (1990) Heuristics for product-line design using conjoint analysis. *Management Sci.* 36(12):1464–1478.

Kraus UG, Yano CA (2003) Product line selection and pricing under a share-of-surplus choice model. *Eur. J. Oper. Res.* 150(3):653–671.

Lubin M, Dunning I (2015) Computing in operations research using Julia. *INFORMS J. Comput.* 27(2):238–248.

Luo L (2011) Product line design for consumer durables: An integrated marketing and engineering approach. *J. Marketing Res.* 48(1): 128–139.

Luo L, Kannan PK, Besharati B, Azarm S (2005) Design of robust new products under variability: Marketing meets design. *J. Product Innovation Management* 22(2):177–192.

McBride RD, Zufryden FS (1988) An integer programming approach to the optimal product line selection problem. *Marketing Sci.* 7(2): 126–140.

Rossi PE (2012) *Bayesm: Bayesian Inference for Marketing/Micro-econometrics*. http://CRAN.R-project.org/package=bayesm. R package version 2.2-5.

Rossi PE, Allenby GM (2003) Bayesian statistics and marketing. *Marketing Sci.* 22(3):304–328.

Rossi PE, Allenby GM, McCulloch R (2012) *Bayesian Statistics and Marketing* (John Wiley & Sons, Hoboken, NJ).

Rusmevichientong P, Topaloglu H (2012) Robust assortment optimization in revenue management under the multinomial logit choice model. *Oper. Res.* 60(4):865–882.

Sawtooth Software (2003) Advanced Simulation Module (ASM) for Product Optimization v1.5. *Sawtooth Software Technical Paper Series*. http://www.sawtoothsoftware.com/download/techpap/asmtech.pdf.

Schön C (2010a) On the optimal product line selection problem with price discrimination. *Management Sci.* 56(5):896–902.

Schön C (2010b) On the product line selection problem under attraction choice models of consumer behavior. *Eur. J. Oper. Res.* 206(1): 260–264.

Schwarz G (1978) Estimating the dimension of a model. *Ann. Statist.* 6(2):461–464.

Toubia O, Simester DI, Hauser JR, Dahan E (2003) Fast polyhedral adaptive conjoint estimation. *Marketing Sci.* 22(3):273–303.

Train KE (2008) EM algorithms for nonparametric estimation of mixing distributions. *J. Choice Model.* 1(1):40–69.

Train KE (2009) *Discrete Choice Methods with Simulation*. (Cambridge University Press, New York).

Zufryden FS (1982) Product line optimization by integer programming. *Proc. Annual Meeting of ORSA/TIMS, San Diego, CA*, 100–114.

Dimitris Bertsimas is the Boeing Professor of Operations Research and codirector of the Operations Research Center at the Massachusetts Institute of Technology. His interests include analytics, optimization, and stochastics. This paper is part of his long-term interest in optimization under uncertainty.

Velibor V. Mišić is an assistant professor of Decisions, Operations, and Technology Management at the Anderson School of Management, University of California, Los Angeles. His research interests are in analytics, operations management, marketing science, and optimization under uncertainty, the latter two of which this paper is a part.