Revenue-Utility Tradeoff in Assortment Optimization under the Multinomial Logit Model with Totally Unimodular Constraints*

Mika Sumida[1], Guillermo Gallego[2], Paat Rusmevichientong[3], Huseyin Topaloglu[1], James Davis[4]

[1]School of Operations Research and Information Engineering, Cornell Tech, New York, NY 10044
[2]Department of Industrial Engineering and Decision Analytics, HKUST, Hong Kong
[3]Marshall School of Business, University of Southern California, Los Angeles, CA 90089
[4]Uber Technologies Inc., 1455 Market Street, San Francisco, CA 94103
ms3268@cornell.edu, ggallego@ust.hk, rusmevic@marshall.usc.edu, topaloglu@orie.cornell.edu, jamesmariodavis@gmail.com

May 28, 2019

We examine the revenue-utility assortment problem with the goal of finding an assortment that maximizes a linear combination of the expected revenue of the firm and the expected utility of the customer. This criterion captures the tradeoff between the firm-centric objective of maximizing the expected revenue and the customer-centric objective of maximizing the expected utility. The customers choose according to the multinomial logit model, and there is a constraint on the offered assortments characterized by a totally unimodular matrix. We can solve the revenue-utility assortment problem by finding the assortment that maximizes only the expected revenue, after adjusting the revenue of each product by the same constant. To find an optimal assortment, we use a parametric linear program to generate a collection of candidate assortments that is guaranteed to include an optimal solution to the revenue-utility assortment problem. This collection of candidate assortments also allows us to construct an efficient frontier that shows the optimal expected revenue-utility pairs as we vary the weights in the objective function. Furthermore, we develop a procedure that limits the number of candidate assortments placed under consideration while maintaining the solution quality. Through extensive examples, we demonstrate a broad range of applications that fit within our framework.

Key words: choice modeling, multinomial logit, revenue-utility tradeoff, totally unimodular constraints

1. Introduction

In the revenue management literature, discrete choice models continue to receive attention as an attractive option for modeling demand, because these models capture the substitution possibilities among products. By using discrete choice models, we can develop demand models

* This manuscript extends and overrides the *unpublished* work by Davis et al. (2013), in which the authors focused on maximizing the expected revenue under the multinomial logit model with totally unimodular constraints.

that capture the fact that customers choose and substitute among products and that if a certain product is unavailable, then some customers may substitute another product, whereas others may decide to leave the system without making a purchase. A growing body of literature indicates that capturing the customer choice process by using discrete choice models can yield better operational decisions; see, for example, Talluri and van Ryzin (2004), Vulcano et al. (2010), and Feldman et al. (2019). However, most of this literature focuses on the firm-centric objective of maximizing expected revenue, leaving customer-centric objectives relatively untouched.

We study an optimization problem whose objective is to find an assortment that maximizes a linear combination of the expected revenue of the firm and the expected utility of the customer. This criterion captures the tradeoff between the firm-centric objective of expected revenue and the customer-centric objective of expected utility. Customers choose among the products according to the multinomial logit model, and there are constraints on the offered assortment characterized by a totally unimodular matrix. Our totally unimodular constraints encompass numerous assortment and pricing applications with different operational constraints. We refer to our optimization problem as the revenue-utility assortment problem.

We show that we can obtain an optimal solution to the revenue-utility assortment problem by finding an assortment that maximizes only the expected revenue, after adjusting the revenue of each product by the same amount. To find an optimal assortment, we formulate a parametric linear program (LP) that generates a collection of candidate assortments that is guaranteed to include an optimal solution to the revenue-utility assortment problem. Moreover, this collection of candidate assortments includes an optimal solution to the revenue-utility assortment problem for every value of the weights we put on the expected revenue and the expected utility, and this property allows us to construct an efficient frontier that shows the optimal revenue-utility pairs as the weights vary. Moreover, we develop an approach that considers only a limited number of candidate assortments and simultaneously maintains a prespecified solution quality.

Main Contributions: To gain an overview of our problem setup, let $\mathcal{N} = \{1, 2, \ldots, n\}$ denote the set of available products. An assortment is represented by a vector $\boldsymbol{x} = (x_1, \ldots, x_n) \in \{0, 1\}^n$, where $x_i = 1$ if and only if we offer product i. The set of feasible assortments is $\mathcal{F} = \{\boldsymbol{x} \in \{0, 1\}^n : \mathbf{A}\boldsymbol{x} \le \boldsymbol{b}\}$, where $\mathbf{A}$ is a totally unimodular matrix. We give numerous examples that formulate operationally useful assortment and pricing problems through totally unimodular constraints. Customers choose among the products according to the multinomial logit model. Our goal is to find an assortment that maximizes a linear combination of the expected revenue and the expected utility, where we put a weight of one on the expected revenue and a weight of λ on the expected utility. When we need to explicitly refer to the weight on the expected utility, we refer to our assortment problem as the $(1, \lambda)$-weighted revenue-utility assortment problem.

Novel formulation. Our objective function and constraints incorporate novel features. The objective function considers both the firm-centric expected revenue and the customer-centric expected utility. Through totally unimodular constraints, we capture a variety of practically relevant applications, in which we can impose bounds on the number of offered products, incorporate display location effects, and formulate assortment problems that choose the offered products as well as their prices. In our problem formulation, we build on the fact that the multinomial logit model is compatible with the random utility maximization principle, according to which a customer associates random utilities with the alternatives and chooses the alternative with the largest utility; see, for example, McFadden (1974). Naturally, if the utilities of all alternatives are shifted by a constant, then the choice process of the customers does not change. To ensure that the objective function of our revenue-utility assortment problem remains invariant to such a constant shift, we focus on the expected utility of the customer net of the expected utility she would have received if she had left without making a purchase. Our approach is equivalent to normalizing the mean utility of the no-purchase option to zero.

Characterization of an optimal assortment. We construct a lower bound on the objective function of the revenue-utility assortment problem that requires computing only the expected revenue, but *after* adjusting the revenue associated with each product by the same additive constant (Lemma 3.1). We show that this lower bound is tight at the optimal solution, in the sense that we can obtain an optimal solution to the revenue-utility assortment problem by finding an assortment that maximizes only the expected revenue, after adjusting the revenue associated with each product by the same additive constant (Theorem 3.2). This characterization establishes a new and critical connection between the expected revenue-utility and the standard expected revenue objective functions. An immediate consequence of this structural property is that when there is no constraint, an optimal solution to the revenue-utility assortment problem is a revenue-ordered assortment that offers a certain number of products with the largest revenues (Corollary 3.3). Unfortunately, computing the "right" adjustment in the revenues of the products requires maximizing a nonconcave function with many local maxima, which is challenging. However, as discussed in the next paragraph, we build on this connection to develop a solution method for the revenue-utility assortment problem.

Efficient solution methods. We develop an approach to solving the revenue-utility assortment problem that is based on solving a parametric LP. In this LP, we vary a parameter over the real line to generate a collection of candidate assortments such that the collection is guaranteed to include an optimal solution to the revenue-utility assortment problem (Theorem 4.1). Using r_i and v_i to denote the revenue and preference weight, respectively, of product i, scaling all the revenues and preference weights so that they take integer values, and letting m be the number of constraints, we show that the number of candidate assortments in the collection is at most $\min\left\{1+n\,\max_{i\in\mathcal{N}} v_i,\; 2+2\,n\,\max_{i\in\mathcal{N}} r_i\,v_i,\; (m+n)^{1+m}\right\}$ (Theorem 4.2). The first two terms of the minimum show that for fixed revenues and preference weights, the number of candidate assortments grows linearly with the number of products; the last term shows that for a fixed number of constraints, the number of candidate assortments grows polynomially with the number of products. An important feature of our collection of candidate assortments is that it is *independent* of the weight λ in the $(1,\lambda)$-weighted revenue-utility assortment problem. Thus, once we construct the collection of candidate assortments, we can use the same collection to solve the $(1,\lambda)$-weighted revenue-utility assortment problem for all values of λ simultaneously. This allows us to construct an efficient frontier that shows the optimal expected revenue-utility pairs as the weight λ varies.

We also develop an approach that balances solution quality and computational effort. Letting $V_{\min}$ and $V_{\max}$ be the smallest and largest preference weights, respectively, for a given grid size $\rho>0$, our approach generates a collection of $O\left(\frac{1}{\rho}\log\left(nV_{\max}/V_{\min}\right)\right)$ candidate assortments and ensures that the collection includes a solution whose objective value is at least $1/(1+\rho)$ of the optimal value. By adjusting the value of ρ, we can strike a balance between the solution quality and the number of candidate assortments, the latter quantity being a measure of computational effort.

An LP for expected revenue maximization. When $\lambda = 0$, the objective function of the $(1,\lambda)$-weighted revenue-utility assortment problem reduces to the expected revenue criterion. Even in this simpler setting, we offer a novel contribution by showing that if there are constraints on the assortment characterized by a totally unimodular matrix, then we can maximize the expected revenue by solving an LP with $n+1$ variables and $n+m+1$ constraints (Theorem 5.2). Our analysis uses the LP duality and can potentially be applied to other assortment problems, providing connections between LP and assortment optimization.

Applications. We describe five practical problem classes that can be formulated using totally unimodular constraints. First, we consider a variety of cardinality constraints that limit the number of products in the offered assortment. Second, we consider assortment problems with display location effects, in which the attractiveness of each product depends on its attributes as well as on the location where the product is displayed. Display location effects are a common consideration in retail, because getting a prime location can boost the attractiveness of the products displayed on a shelf or a web page. Third, we consider pricing problems in which there is a finite menu of possible prices and the attractiveness of a product depends on its price. In our formulation of the pricing problem, the relationship between the attractiveness of a product and its price can be arbitrary. Fourth, we consider pricing problems with a price ladder constraint, in which there is an inherent ordering in the qualities of the products and the prices of the products must adhere to the same ordering. Fifth, we consider assortment problems with product precedence constraints, such that a particular product cannot be offered unless certain related products are also offered. These five applications can be modeled using totally unimodular constraints. To demonstrate the effectiveness of our solution methods, we conduct numerical experiments on assortment problems with display location effects and pricing problems.

Related Literature: Our paper is related to research on assortment problems under the multinomial logit model, the goal of which is to find an assortment that maximizes the expected revenue. Talluri and van Ryzin (2004) and Gallego et al. (2004) examine the problem without any constraints, and both studies show that an optimal assortment is revenue ordered. Rusmevichientong et al. (2009) present a polynomial-time approximation scheme for instances in which each product has a space requirement and there is a limit on the total space consumption of the offered products. Rusmevichientong et al. (2010) focus on cardinality constraints on the offered assortment and develop an efficient algorithm for computing an optimal assortment. Bront et al. (2009), Mendez-Diaz et al. (2014), Rusmevichientong et al. (2014), and Desir et al. (2016) focus on the assortment problem under a mixture of multinomial logit models in which there are multiple customer types and customers of different types choose according to different multinomial logit models. The authors of these studies characterize the computational complexity of the problem and provide heuristics, integer programming formulations, and approximation methods. Gallego et al. (2015) show that the assortment problem under the multinomial logit model can be formulated as an LP even when products consume combinations of resources and there are constraints on the expected consumption of each resource, but their approach does not consider constraints on what assortments can be offered.

The work presented in this paper is an outgrowth of our earlier work, which was circulated as an unpublished technical report (Davis et al. 2013); in that report, we focused on finding an assortment that maximizes only the expected revenue without considering the expected utility. We believe the work in this paper is unique and amplifies our unpublished work substantially, because this paper is one of very few studies that take a customer-centric view of assortment optimization, allowing us to manage the tradeoff between firm-specific and customer-specific objectives. Moreover, much of the assortment optimization work, including that of Davis et al. (2013), exploits the fact that the expected revenue under the multinomial logit model can be written as a fraction of two linear functions; see Davis et al. (2014), Feldman and Topaloglu (2015), and Li et al. (2015) for work under other choice models that build on a similar fractional structure of the expected revenue function. This structure immediately breaks down when we include the expected utility in the objective function. Our algorithmic approach, which constructs a lower bound on the objective function of the revenue-utility assortment problem and parametrically maximizes this lower bound, differs substantially from the approaches used in the previous literature. Other work on assortment optimization under the multinomial logit model includes Wang (2012), Abeliuk et al. (2016), Aouad et al. (2018), Sen et al. (2018), Wang and Sahin (2018), Aouad et al. (2019), and Flores et al. (2019).

We are aware of only two other papers on assortment problems with customer-centric objectives. Ashlagi and Shi (2016) formulate the problem of designing school choice menus as a large-scale optimization problem. Their column generation subproblem has an objective function similar to ours. However, they use an entirely different argument to characterize an optimal assortment, and they do not consider efficient algorithms under constraints on the offered assortment or pricing variants. In the context of drug design, Truong (2014) formulates an assortment problem that minimizes the difference between the expected cost and the expected utility. Her expected cost function is similar to our expected revenue, but since she focuses on minimizing the difference between the expected cost and the expected utility, the structure of her objective function is different. Moreover, she does not consider pricing variants or constraints on the offered assortment.

Our pricing application uses discrete price menus. Pricing models traditionally assume a parametric relationship between the price and the preference weight of a product. For example, if the mean utility of a product is linear in its prices, then the preference weight of product i, as a function of its price p, is given by $e^{\alpha_i - \beta_i p}$ for constants (α_i, β_i). Under these parametric forms, the expected revenue is smooth in the prices; see, for example, Song and Xue (2007), Dong et al. (2009), Li and Huh (2011), Gallego and Wang (2014), Li and Huh (2015), Li and Webster (2017), and Chen and Gallego (2019). Our application to the pricing problem in Section 6.3 allows the preference weight of an item to depend on its price in an arbitrary fashion, without any restriction. Furthermore, since we work with discrete price menus, we can limit attention to operationally appealing prices, such as those in increments of a dollar or those that have 99 cents as the final digits.

Organization: In Section 2, we formulate our assortment problem. In Section 3, we show that the revenue-utility assortment problem can be solved by finding an assortment that maximizes the expected revenue, after adjusting the revenues of all products by the same additive amount. In Section 4, we use this observation to formulate a parametric LP to generate a collection of candidate assortments. In Section 5, we develop a method that balances the number of candidate assortments with the solution quality. In Section 6, we discuss applications with totally unimodular constraints. We present numerical experiments in Section 7 and offer conclusions in Section 8.

2. Problem Formulation

Let $\mathcal{N} = \{1, 2, \ldots, n\}$ denote the set of products. The revenue associated with product i is $r_i \geq 0$. We use $\boldsymbol{x} = (x_1, \ldots, x_n) \in \{0,1\}^n$ to capture the subset of products that we offer to the customers, where $x_i = 1$ if and only if we offer product i. We refer to the vector $\boldsymbol{x}$ simply as the assortment that we offer. The customers make a choice within the assortment that we offer according to the multinomial logit model. Under the multinomial logit model, a customer associates a random utility with each product i, which has the Gumbel distribution with location and scale parameters $(\mu_i, 1)$. Similarly, a customer associates a random utility with the no-purchase option, which also has the Gumbel distribution with location and scale parameters $(\mu_0, 1)$. The customer chooses the available alternative that provides the largest utility; this alternative may be one of the products in the offered assortment or the no-purchase option. Letting $v_i = e^{\mu_i}$ denote the preference weight of product i and $v_0 = e^{\mu_0}$ denote the preference weight of the no-purchase option, if we offer the assortment $\boldsymbol{x}$, then the customer chooses product i with probability

$$\phi_i(\boldsymbol{x}) \ = \ \frac{v_i\, x_i}{v_0 + \sum_{j \in \mathcal{N}} v_j\, x_j} \ .$$

Under the assortment $\boldsymbol{x}$, the expected utility that the customer obtains from the chosen alternative is $\log(v_0 + \sum_{i \in \mathcal{N}} v_i\, x_i) + Q$, where Q is the Euler-Mascheroni constant (McFadden 1974).

We have two goals in mind when choosing the assortment to offer. First, we want to maximize the expected revenue obtained from the customer. When the customer chooses product i, we obtain a revenue of r_i, so if we offer the assortment $\boldsymbol{x}$, then the expected revenue obtained from the customer is $\sum_{i \in \mathcal{N}} \phi_i(\boldsymbol{x})\, r_i = \sum_{i \in \mathcal{N}} r_i\, v_i\, x_i / \big(v_0 + \sum_{j \in \mathcal{N}} v_j\, x_j\big)$. Second, we want to maximize the expected utility that the customer receives from the chosen alternative, *net* of the expected utility that she obtains when she must choose the no-purchase option. The expected utility of the no-purchase option is $\log v_0 + Q$. Therefore, if we offer the assortment $\boldsymbol{x}$, then the customer's expected utility from the alternative she chooses, net of the expected utility she obtains when she must choose the no-purchase option, is $\log\big(v_0 + \sum_{i \in \mathcal{N}} v_i\, x_i\big) + Q - \log v_0 - Q = \log\Big(1 + \sum_{i \in \mathcal{N}} \frac{v_i}{v_0}\, x_i\Big)$. Letting $V(\boldsymbol{x}) = \sum_{i \in \mathcal{N}} \frac{v_i}{v_0} x_i$ for notational brevity and using $\boldsymbol{r} = (r_1, \ldots, r_n)$ to denote the vector of product revenues, if we offer the assortment $\boldsymbol{x}$, then the expected revenue and the net expected utility of the customer are, respectively, given by

$$\mathsf{Rev}(\boldsymbol{x}; \boldsymbol{r}) \ = \ \frac{\sum_{i \in \mathcal{N}} r_i\, v_i\, x_i}{v_0 + \sum_{i \in \mathcal{N}} v_i\, x_i} \qquad \text{and} \qquad \mathsf{Util}(\boldsymbol{x}) = \log(1 + V(\boldsymbol{x})).$$

We focus on the net expected utility rather than the expected utility for the following reason. If we multiply the preference weight of all alternatives by a constant, then the choice probability of each alternative remains the same. Therefore, if we estimate the parameters of the multinomial logit model from the data, then we can estimate them only up to a multiplicative constant; see Section 3.5 in Train (2003). If we multiply the preference weights of all alternatives by α, then the expected utility of the customer increases from $\log(v_0 + \sum_{i \in \mathcal{N}} v_i\, x_i) + Q$ to $\log \alpha + \log(v_0 + \sum_{i \in \mathcal{N}} v_i\, x_i) + Q$. Thus, the expected utility of the customer depends on a multiplicative constant that applies to the preference weights of all alternatives, but we cannot estimate such a multiplicative constant from the data. On the other hand, even when we multiply the preference weights of all alternatives by α, the expected net utility of the customer remains the same as $\log\left(1 + \sum_{i \in \mathcal{N}} \frac{v_i}{v_0}\, x_i\right)$. Thus, by focusing on the net expected utility, we obtain a measure of utility that is insensitive to a multiplicative constant that applies to all preference weights. Lastly, we need to find assortments that maximize the expected revenue under different product revenues, so we make explicit the dependence of $\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r})$ on the product revenues $\boldsymbol{r}$.

The set of feasible assortments that we can offer is given by $\mathcal{F} = \{\boldsymbol{x} \in \{0,1\}^n : \mathbf{A}\,\boldsymbol{x} \le \boldsymbol{b}\}$, where m is the number of constraints, $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a totally unimodular matrix, and $\boldsymbol{b} \in \mathbb{R}^m$ is an integral vector. We express our set of feasible assortments by using "less than or equal to" constraints, but negating or duplicating a row of a totally unimodular matrix preserves its total unimodularity. Thus, we can accommodate "greater than or equal to" or "equal to" constraints by replacing a "greater than or equal to" constraint with the negative of a "less than or equal to" constraint and by replacing an "equal to" constraint with a pair of "less than or equal to" and "greater than or equal to" constraints; see Proposition 2.2 in Chapter III.1 in Nemhauser and Wolsey (1988). Our goal is to find a feasible assortment that maximizes a linear combination of the expected revenue and the net expected utility. Let $\lambda \ge 0$ be a parameter that controls the tradeoff between the expected revenue and the expected utility. We want to solve the following optimization problem:

$$\begin{aligned} Z^*_\lambda &= \max_{\boldsymbol{x} \in \mathcal{F}} \Big\{ \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) + \lambda\, \mathsf{Util}(\boldsymbol{x}) \Big\} \\ &= \max_{\boldsymbol{x} \in \mathcal{F}} \left\{ \frac{\sum_{i \in \mathcal{N}} r_i\, v_i\, x_i}{v_0 + \sum_{i \in \mathcal{N}} v_i\, x_i} + \lambda\, \log(1 + V(\boldsymbol{x})) \right\}. \end{aligned} \tag{Revenue-Utility}$$

Next, we characterize an optimal solution to the Revenue-Utility problem and use the characterization to create a collection of candidate assortments that contains an optimal solution.

3. Characterization of an Optimal Assortment

In the Revenue-Utility problem, both the expected revenue $\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r})$ and the expected utility $\mathsf{Util}(\boldsymbol{x})$ depend on the offered assortment $\boldsymbol{x}$. The next lemma provides a lower bound on the objective function of the Revenue-Utility problem that requires computing only the expected revenue, but not the expected utility. In this lemma and throughout the rest of the paper, we let $\boldsymbol{e} \in \mathbb{R}^n$ be a vector of all ones, $V_{\min} = \min_{i\in\mathcal{N}} v_i/v_0$ and $V_{\max} = \max_{i\in\mathcal{N}} v_i/v_0$. Moreover, we assume that $\mathcal{F}$ includes a nonempty assortment; otherwise, the Revenue-Utility problem is trivial.

Lemma 3.1 (Lower Bound) *Noting that the objective function of the* Revenue-Utility *problem is* $\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) + \lambda\log(1+V(\boldsymbol{x}))$, *we have that for all* $\boldsymbol{x}\in\mathcal{F}$ *and* $t\in[V_{\min}, nV_{\max}]$,

$$\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) + \lambda\log(1+V(\boldsymbol{x})) \;\geq\; \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda(1+t)\,\boldsymbol{e}) \,+\, \lambda\,(\log(1+t)-t)\,.$$

Proof: Consider increasing the revenues of all products by some amount $\alpha\in\mathbb{R}$. Using the definitions of $\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r})$ and $V(\boldsymbol{x})$, for any $x\in\mathcal{F}$, we have

$$\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\alpha\,\boldsymbol{e}) = \frac{\sum_{i\in\mathcal{N}}(r_i+\alpha)\,v_i\,x_i}{v_0+\sum_{i\in\mathcal{N}} v_i\,x_i} = \frac{\sum_{i\in\mathcal{N}} r_i\,v_i\,x_i}{v_0+\sum_{i\in\mathcal{N}} v_i\,x_i} + \frac{\alpha\sum_{i\in\mathcal{N}}\frac{v_i}{v_0}\,x_i}{1+\sum_{i\in\mathcal{N}}\frac{v_i}{v_0}\,x_i} = \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) + \frac{\alpha\,V(\boldsymbol{x})}{1+V(\boldsymbol{x})}\,.$$

Because $\log(1+a)$ is concave in a and its derivative at a is $1/(1+a)$, by the subgradient inequality, we have $\log(1+b)\leq\log(1+a)+\frac{1}{1+a}(b-a)$ for all $a\in\mathbb{R}_+$ and $b\in\mathbb{R}_+$, which is equivalent to $\log(1+a)\geq\log(1+b)-b+\frac{(1+b)\,a}{1+a}$. For any $\boldsymbol{x}\in\mathcal{F}$ and $t\in[V_{\min}, nV_{\max}]$, using this inequality with $a=V(\boldsymbol{x})$ and $b=t$, we get $\log(1+V(\boldsymbol{x}))\geq\log(1+t)-t+\frac{(1+t)\,V(\boldsymbol{x})}{1+V(\boldsymbol{x})}$. So, we have

$$\begin{aligned}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) + \lambda\log(1+V(\boldsymbol{x})) \;&\geq\; \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) + \lambda\left\{\log(1+t)-t+\frac{(1+t)\,V(\boldsymbol{x})}{1+V(\boldsymbol{x})}\right\}\\ &=\; \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda(1+t)\,\boldsymbol{e}) \,+\, \lambda\,(\log(1+t)-t)\,,\end{aligned}$$

where the last equality follows from the identity at the beginning of the proof with $\alpha=\lambda(1+t)$. ∎

Since $\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda(1+t)\,\boldsymbol{e})+\lambda\,(\log(1+t)-t)$ is a lower bound on the objective function of the Revenue-Utility problem for all $x\in\mathcal{F}$ and $t\in[V_{\min}, nV_{\max}]$, we can maximize this function over all $\boldsymbol{x}\in\mathcal{F}$ and $t\in[V_{\min}, nV_{\max}]$ to obtain a lower bound on the optimal objective value of the Revenue-Utility problem. In other words, we can solve the problem

$$\max_{t\in[V_{\min}, nV_{\max}]}\left\{\,\max_{\boldsymbol{x}\in\mathcal{F}}\Big\{\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda(1+t)\,\boldsymbol{e})\Big\}+\lambda\,(\log(1+t)-t)\,\right\}. \qquad (\textsf{Parametric})$$

The next theorem shows that the above Parametric lower bound on the optimal objective value of the Revenue-Utility problem is actually tight. Furthermore, we can use an optimal solution to the Parametric problem to obtain an optimal solution to the Revenue-Utility problem, yielding a characterization of an optimal solution to the Revenue-Utility problem.

Theorem 3.2 (Revenue-Utility Solution) *If* $(t^*, \boldsymbol{x}^*)$ *is an optimal solution to the* Parametric *problem, then* $\boldsymbol{x}^*$ *is also an optimal solution to the* Revenue-Utility *problem. Furthermore, the* Parametric *and* Revenue-Utility *problems have the same optimal objective value.*

Proof: Let $\widehat{\boldsymbol{x}}$ be an optimal solution to the Revenue-Utility problem providing the optimal objective value Z^*_λ. Since $\boldsymbol{x}^*$ is a feasible solution to the Revenue-Utility problem, we have $Z^*_\lambda \geq \mathsf{Rev}(\boldsymbol{x}^*; \boldsymbol{r}) + \lambda \log(1+V(\boldsymbol{x}^*))$. Using Lemma 3.1 with $\boldsymbol{x} = \boldsymbol{x}^*$ and $t = t^*$, we have

$$
\begin{aligned}
\mathsf{Rev}(\boldsymbol{x}^*; \boldsymbol{r}) + \lambda \log(1+V(\boldsymbol{x}^*)) &\geq \mathsf{Rev}(\boldsymbol{x}^*; \boldsymbol{r} + \lambda\,(1+t^*)\,\boldsymbol{e}) + \lambda\,(\log(1+t^*) - t^*) \\
&\overset{(a)}{\geq} \mathsf{Rev}(\widehat{\boldsymbol{x}}\,; \boldsymbol{r} + \lambda\,(1+V(\widehat{\boldsymbol{x}}))\,\boldsymbol{e}) + \lambda\,(\log(1+V(\widehat{\boldsymbol{x}})) - V(\widehat{\boldsymbol{x}})) \\
&\overset{(b)}{=} \mathsf{Rev}(\widehat{\boldsymbol{x}}\,; \boldsymbol{r}) + \frac{\lambda\,(1+V(\widehat{\boldsymbol{x}}))\,V(\widehat{\boldsymbol{x}})}{1+V(\widehat{\boldsymbol{x}})} + \lambda\,(\log(1+V(\widehat{\boldsymbol{x}})) - V(\widehat{\boldsymbol{x}})) \\
&= \mathsf{Rev}(\widehat{\boldsymbol{x}}\,; \boldsymbol{r}) + \lambda \log(1+V(\widehat{\boldsymbol{x}})) \;=\; Z^*_\lambda,
\end{aligned}
$$

where (a) holds because $\mathcal{F}$ includes a nonempty assortment, yielding $\widehat{\boldsymbol{x}} \neq \boldsymbol{0} \in \mathbb{R}^n_+$, in which case we get $V(\widehat{\boldsymbol{x}}) \in [V_{\min}, nV_{\max}]$, so $(V(\widehat{\boldsymbol{x}}), \widehat{\boldsymbol{x}})$ is a feasible but not necessarily an optimal solution to the Parametric problem, whereas (b) follows by the identity at the beginning of the proof of Lemma 3.1 with $\alpha = \lambda\,(1+V(\widehat{\boldsymbol{x}}))$. Since $Z^*_\lambda \geq \mathsf{Rev}(\boldsymbol{x}^*; \boldsymbol{r}) + \lambda \log(1+V(\boldsymbol{x}^*))$, all of the above inequalities hold as equalities, in which case $\mathsf{Rev}(\boldsymbol{x}^*; \boldsymbol{r} + \lambda\,(1+t^*)\,\boldsymbol{e}) + \lambda\,(\log(1+t^*) - t^*) = Z^*_\lambda$, so the Parametric and Revenue-Utility problems have the same optimal objective value. Similarly, since all of the above inequalities hold as equalities, we have $\mathsf{Rev}(\boldsymbol{x}^*; \boldsymbol{r}) + \lambda \log(1+V(\boldsymbol{x}^*)) = Z^*_\lambda$, which implies that $\boldsymbol{x}^*$ is an optimal solution to the Revenue-Utility problem. ■

By Theorem 3.2, letting t^* be an optimal solution to the outer maximization problem in the Parametric problem, we can obtain an optimal solution to the Revenue-Utility problem by solving the problem $\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x}; \boldsymbol{r} + \lambda\,(1+t^*)\,\boldsymbol{e})$. Thus, we can solve the Revenue-Utility problem by finding an assortment that maximizes *only the expected revenue*, as long as we shift all product revenues by $\lambda\,(1+t^*)$. However, finding an optimal solution to the outer maximization in the Parametric problem to determine t^* is difficult. In Figure 3, we plot the objective function of the outer maximization as a function of t, which involves multiple local maxima. Instead of trying to find the global maximum, our approach in Section 4 generates a collection of candidate assortments without knowing the value of t^* such that this collection is guaranteed to contain an optimal solution to the problem $\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x}; \boldsymbol{r} + \lambda\,(1+t^*)\,\boldsymbol{e})$. The collection of candidate assortments thus includes an optimal solution to the Revenue-Utility problem as well. By checking the objective value associated with each candidate assortment, we determine an optimal solution to the Revenue-Utility problem.

We close this section with a corollary to Theorem 3.2, which shows that a revenue-ordered assortment is optimal to the Revenue-Utility problem when there is no constraint.

Corollary 3.3 (Revenue-Ordered) *If there is no constraint, then a revenue-ordered assortment solves the* Revenue-Utility *problem; that is, if $\mathcal{F} = \{0,1\}^n$ and the products are indexed such that $r_1 \geq r_2 \geq \ldots \geq r_n$, then there exists an optimal solution $\boldsymbol{x}^*$ to the* Revenue-Utility *problem such that $x^*_i = 1$ for all $i \leq \mathsf{s}^*$ and $x^*_i = 0$ for all $i > \mathsf{s}^*$ for some $\mathsf{s}^* \in \mathcal{N}$.*

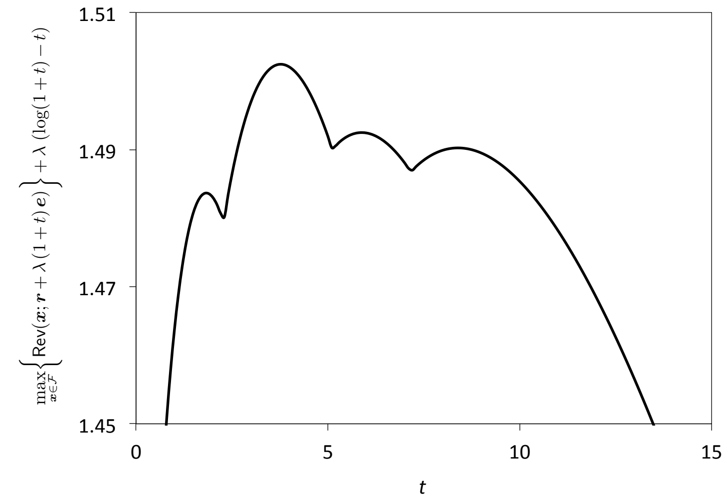

Figure 1 The objective function of the outer maximization in the Parametric problem as a function of $t\in[V_{\min}, nV_{\max}]$, with $n=6$, $\mathcal{F}=\{0,1\}^n$, $\lambda=0.37$, $(r_1,\ldots,r_6)=(1.89,1.71,1.65,0.67,0.45,0.34)$, $(v_1,\ldots,v_6)=(0.24,0.54,1.05,1.94,2.11,2.51)$, and $v_0=1$.

Proof: It is well-known that a revenue-ordered assortment maximizes the expected revenue when there is no constraint; that is, if $r_1\ge r_2\ge\ldots\ge r_n$, then an optimal solution $\boldsymbol{x}^*$ to the problem $\max_{\boldsymbol{x}\in\{0,1\}^n}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r})$ is of the form $x_i^*=1$ for all $i\le\mathsf{s}^*$ and $x_i^*=0$ for all $i>\mathsf{s}^*$ for some $\mathsf{s}^*\in\mathcal{N}$; see, for example, Talluri and van Ryzin (2004). By the discussion that follows Theorem 3.2, an optimal solution to $\max_{\boldsymbol{x}\in\{0,1\}^n}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t)\,\boldsymbol{e})$ for some $t\in[V_{\min}, n\,V_{\max}]$ is also an optimal solution to the Revenue-Utility problem. In the last problem, the revenue of product i is $r_i+\lambda\,(1+t)$. Since adding a constant to the revenue of each product does not change the ordering of the revenues, a revenue-ordered assortment is optimal for the last problem as well. ∎

Corollary 3.3 generalizes existing results on the optimality of revenue-ordered assortments for the unconstrained expected revenue maximization problem. Even when the objective function includes the expected utility in addition to the expected revenue, revenue-ordered assortments remain optimal. Next, we focus on generating our collection of candidate assortments.

4. Constructing Candidate Assortments

From the discussion in the previous section, we know that an optimal solution to the Revenue-Utility problem can be obtained by solving the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t^*)\,\boldsymbol{e})$ for some $t^*\in[V_{\min}, n\,V_{\max}]$. Note that $\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t^*)\,\boldsymbol{e})=\frac{\sum_{i\in\mathcal{N}}(r_i+\lambda\,(1+t^*))\,v_i\,x_i}{v_0+\sum_{i\in\mathcal{N}}v_i\,x_i}\ge\gamma$ if and only if $\sum_{i\in\mathcal{N}}(r_i-\gamma+\lambda\,(1+t^*))\,v_i\,x_i\ge v_0\,\gamma$. Motivated by this observation, we consider the following LP:

$$\mathsf{LP}(\gamma)\;=\;\max_{\boldsymbol{x}\in\mathbb{R}^n_+}\left\{\sum_{i\in\mathcal{N}}(r_i-\gamma)\,v_i\,x_i\;\middle|\;\mathbf{A}\boldsymbol{x}\le\boldsymbol{b},\quad x_i\le 1\quad\forall\,i\in\mathcal{N}\right\}.\qquad\text{(Candidate LP)}$$

Since the constraint matrix $\mathbf{A}$ is totally unimodular, it is a standard result that the Candidate LP has an optimal solution with all decision variables taking binary values; see Proposition 2.2 in Chapter III.1 in Nemhauser and Wolsey (1988). Therefore, for each $\gamma \in \mathbb{R}$, there exists an optimal solution $\boldsymbol{x}_{\mathsf{LP}}(\gamma) \in \mathcal{F}$ to the Candidate LP problem. We consider the collection of candidate assortments given by $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma \in \mathbb{R}\}$. Solving the Candidate LP for all $\gamma \in \mathbb{R}$ is an application of a parametric LP; see Chapter 5.5 in Bertsimas and Tsitsiklis (1997). The optimal objective value of the Candidate LP, given by $\mathsf{LP}(\gamma)$, is continuous, piecewise linear, decreasing, and convex in γ. The number of breakpoints of the function $\mathsf{LP}(\cdot)$ gives the cardinality of $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma \in \mathbb{R}\}$. The next theorem shows that $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma \in \mathbb{R}\}$ contains an optimal solution to the Revenue-Utility problem.

Theorem 4.1 (Collection of Candidate Assortments) *There exists an optimal solution to the* Revenue-Utility *problem that is in the collection of assortments* $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma \in \mathbb{R}\}$.

Proof: Let t^* be an optimal solution to the outer maximization in the Parametric problem, and let $\gamma^* = \max_{\boldsymbol{x} \in \mathcal{F}} \mathsf{Rev}(\boldsymbol{x}; \boldsymbol{r} + \lambda\,(1+t^*)\,\boldsymbol{e})$. We set $\boldsymbol{x}^* = \boldsymbol{x}_{\mathsf{LP}}(\gamma^* - \lambda\,(1+t^*))$; that is, $\boldsymbol{x}^*$ is an optimal solution to the Candidate LP when solved with $\gamma = \gamma^* - \lambda\,(1+t^*)$. We will prove by contradiction that $\boldsymbol{x}^*$ is an optimal solution to the problem $\max_{\boldsymbol{x} \in \mathcal{F}} \mathsf{Rev}(\boldsymbol{x}; \boldsymbol{r} + \lambda\,(1+t^*)\,\boldsymbol{e})$. Suppose on the contrary that $\boldsymbol{x}^*$ is not an optimal solution to the problem $\max_{\boldsymbol{x} \in \mathcal{F}} \mathsf{Rev}(\boldsymbol{x}; \boldsymbol{r} + \lambda\,(1+t^*)\,\boldsymbol{e})$, so that $\gamma^* > \mathsf{Rev}(\boldsymbol{x}^*; \boldsymbol{r} + \lambda\,(1+t^*)\,\boldsymbol{e}) = \frac{\sum_{i \in \mathcal{N}} (r_i + \lambda\,(1+t^*))\,v_i\,x_i^*}{v_0 + \sum_{i \in \mathcal{N}} v_i\,x_i^*}$. Focusing on the first and third expressions in this chain of inequalities and rearranging the terms, we get $\sum_{i \in \mathcal{N}} (r_i - \gamma^* + \lambda\,(1+t^*))\,v_i\,x_i^* < v_0\,\gamma^*$. Moreover, each $\boldsymbol{x} \in \mathcal{F}$ is a feasible solution to the Candidate LP when solved with $\gamma = \gamma^* - \lambda\,(1+t^*)$, but $\boldsymbol{x}^*$ is an optimal solution to this LP, which implies that $\sum_{i \in \mathcal{N}} (r_i - \gamma^* + \lambda\,(1+t^*))\,v_i\,x_i \leq \sum_{i \in \mathcal{N}} (r_i - \gamma^* + \lambda\,(1+t^*))\,v_i\,x_i^*$ for all $\boldsymbol{x} \in \mathcal{F}$. Therefore, for each $\boldsymbol{x} \in \mathcal{F}$, we have

$$\sum_{i \in \mathcal{N}} (r_i - \gamma^* + \lambda\,(1+t^*))\,v_i\,x_i \;\leq\; \sum_{i \in \mathcal{N}} (r_i - \gamma^* + \lambda\,(1+t^*))\,v_i\,x_i^* \;<\; v_0\,\gamma^*.$$

Focusing on the first and third expressions above and solving for γ^*, we get $\frac{\sum_{i \in \mathcal{N}} (r_i + \lambda\,(1+t^*))\,v_i\,x_i}{v_0 + \sum_{i \in \mathcal{N}} v_i\,x_i} < \gamma^*$ for each $\boldsymbol{x} \in \mathcal{F}$. This contradicts the fact that $\gamma^* = \max_{\boldsymbol{x} \in \mathcal{F}} \mathsf{Rev}(\boldsymbol{x}; \boldsymbol{r} + \lambda\,(1+t^*)\,\boldsymbol{e})$!

By the argument in the previous paragraph, we have established that $\boldsymbol{x}^*$ is an optimal solution to the problem $\max_{\boldsymbol{x} \in \mathcal{F}} \mathsf{Rev}(\boldsymbol{x}; \boldsymbol{r} + \lambda\,(1+t^*)\,\boldsymbol{e})$. It follows from Theorem 3.2 that $\boldsymbol{x}^*$ is also an optimal solution to the Revenue-Utility problem. Lastly, since $\boldsymbol{x}^* = \boldsymbol{x}_{\mathsf{LP}}(\gamma^* - \lambda\,(1+t^*))$, we have $\boldsymbol{x}^* \in \{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma \in \mathbb{R}\}$. Therefore, there exists an optimal solution to the Revenue-Utility problem that is in the collection of assortments $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma \in \mathbb{R}\}$. ∎

We can use a parametric LP to construct the collection $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma \in \mathbb{R}\}$ of assortments such that this collection includes an optimal solution to the Candidate LP for all $\gamma \in \mathbb{R}$. By Theorem 4.1, an optimal solution to the Revenue-Utility problem is guaranteed to be in this collection. To emphasize that this collection has a finite number of assortments, we write

$$\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma \in \mathbb{R}\} = \{\boldsymbol{x}^\ell_{\mathsf{Cand}} \in \mathcal{F} : \ell = 1, \ldots, L\}, \qquad \text{(Collection of Candidate Assortments)}$$

where L is the number of candidate assortments, and for each ℓ, $\boldsymbol{x}^\ell_{\mathsf{Cand}}$ is a candidate assortment, corresponding to an optimal solution to the Candidate LP for some γ.[1] The above expression highlights the discrete nature of the collection of candidate assortments. The next theorem shows that L is upper bounded by the minimum of three terms. The first two terms show that for fixed preference weights and revenues, L is linear in the number of products n; the third term shows that for a fixed number of constraints m, L is polynomial in n. Note that as long as the preference weights and revenues of the products are rationals, we can assume that the preference weights and revenues are integers, because by the discussion in Section 2, we can scale the preference weights and revenues by the same constant.

Theorem 4.2 (Number of Candidate Assortments) *If v_i and $r_i\, v_i$ are integers for all $i \in \mathcal{N}$, then $L \leq \min\left\{1 + n\, \max_{i\in\mathcal{N}} v_i,\ 2 + 2\,n\, \max_{i\in\mathcal{N}} r_i\, v_i,\ (m+n)^{1+m}\right\}$.*

Proof: The bound $(m+n)^{1+m}$ follows by counting the number of extreme point solutions of an LP. Including the slack variables for the two sets of constraints, the Candidate LP has $2n+m$ decision variables and $n+m$ constraints, in which case a naive argument shows that the number of extreme point solutions of this LP is $\binom{2n+m}{n+m} = O((2n+m)^n)$, but since n of the $n+m$ constraints are bounds on the decision variables, we can use a more refined argument to establish the bound of $(m+n)^{1+m}$. We defer the details to Appendix A. Here, we prove the bounds $1 + n\, \max_{i\in\mathcal{N}} v_i$ and $2 + 2\,n\, \max_{i\in\mathcal{N}} r_i\, v_i$ using a technique adapted from Carstensen (1983). Note that the function $\gamma \mapsto \mathsf{LP}(\gamma)$ is continuous, piecewise linear, decreasing, and convex in γ. As γ ranges over $\mathbb{R}$, the number of breakpoints of $\mathsf{LP}(\cdot)$ gives the number of possible optimal solutions to the Candidate LP. For each $\boldsymbol{x} \in \mathcal{F}$, let $\ell_{\boldsymbol{x}}(\gamma) = \sum_{i\in\mathcal{N}} (r_i - \gamma)\, v_i\, x_i$, which is linear in γ. In this case, we have

$$\mathsf{LP}(\gamma) = \max\left\{\ell_{\boldsymbol{x}}(\gamma) \ : \ \boldsymbol{x} \in \mathcal{F}\right\},$$

where we use the fact that there exists a binary-valued optimal solution to the Candidate LP. Thus, $\mathsf{LP}(\cdot)$ is the pointwise maximum of the lines $\{\ell_{\boldsymbol{x}}(\cdot) : \boldsymbol{x} \in \mathcal{F}\}$. If two lines $\ell_{\boldsymbol{x}'}(\cdot)$ and $\ell_{\boldsymbol{x}''}(\cdot)$ have the same slope, then we can eliminate one of the lines from the set $\{\ell_{\boldsymbol{x}}(\cdot) : \boldsymbol{x} \in \mathcal{F}\}$ without changing the function $\mathsf{LP}(\cdot)$. Therefore, the number of lines in the set $\{\ell_{\boldsymbol{x}}(\cdot) : \boldsymbol{x} \in \mathcal{F}\}$ that is necessary to describe the function $\mathsf{LP}(\cdot)$ is the number of different slopes for these lines. The slope of the line

[1] If for some γ there are multiple optimal solutions, we break ties using a deterministic tie-breaking rule.

$\ell_{\boldsymbol{x}}(\cdot)$ is $-\sum_{i\in\mathcal{N}} v_i\,x_i$, which is an integer between $-\sum_{i\in\mathcal{N}} v_i$ and zero. Therefore, there are at most $1+\sum_{i\in\mathcal{N}} v_i \le 1+n\max_{i\in\mathcal{N}} v_i$ different slopes for the lines in the set $\{\ell_{\boldsymbol{x}}(\cdot):\boldsymbol{x}\in\mathcal{F}\}$, yielding at most $1+n\max_{i\in\mathcal{N}} v_i$ breakpoints.

Similarly, if three lines $\ell_{\boldsymbol{x}'}(\cdot)$, $\ell_{\boldsymbol{x}''}(\cdot)$, and $\ell_{\boldsymbol{x}'''}(\cdot)$ have the same intercept, then we can eliminate one of the lines from the set $\{\ell_{\boldsymbol{x}}(\cdot):\boldsymbol{x}\in\mathcal{F}\}$ without changing the function $\mathsf{LP}(\cdot)$. Thus, the number of lines in the set $\{\ell_{\boldsymbol{x}}(\cdot):\boldsymbol{x}\in\mathcal{F}\}$ that is necessary to describe the function $\mathsf{LP}(\cdot)$ is twice the number of different intercepts for these lines. The intercept of the line $\ell_{\boldsymbol{x}}(\cdot)$ is $\sum_{i\in\mathcal{N}} r_j\,v_j\,x_j$, which is an integer between zero and $\sum_{i\in\mathcal{N}} r_i\,v_i$, so we get at most $1+n\max_{i\in\mathcal{N}} r_i\,v_i$ different intercepts, yielding at most $2+2\,n\max_{i\in\mathcal{N}} v_i$ breakpoints. ■

Note that since the bound $(m+n)^{1+m}$ is based on counting the extreme point solutions of an LP, it holds even when v_i and $r_i\,v_i$ are not integers.

Efficient Frontier for the Revenue-Utility Tradeoff: The parameter λ in the Revenue-Utility problem controls the tradeoff between the expected revenue and the expected utility. Often, we want to understand how the optimal expected revenue and expected utility, along with the optimal solution to the Revenue-Utility problem, change as a function of the parameter λ. Letting $\boldsymbol{x}^*_\lambda$ be an optimal solution to the Revenue-Utility problem as a function of λ, we want to compute $\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r})$ and $\mathsf{Util}(\boldsymbol{x}^*_\lambda)$ for all $\lambda\ge 0$ simultaneously. Here, the key observation is that the parameter λ does not play any role in the Candidate LP. Thus, the Collection of Candidate Assortments given by $\{\boldsymbol{x}^\ell_{\mathsf{Cand}}:\ell=1,\dots,L\}$ includes an optimal solution to the Candidate LP for every possible value of $\gamma\in\mathbb{R}$, and this collection of candidate assortments is *independent* of the value of the parameter λ. Therefore, when solving the Revenue-Utility problem for any value of $\lambda\ge 0$, we can use the same candidate assortments; thus, for all $\lambda\ge 0$, $\boldsymbol{x}^*_\lambda = \arg\max\big\{\mathsf{Rev}(\boldsymbol{x}^\ell_{\mathsf{Cand}};\boldsymbol{r})+\lambda\,\mathsf{Util}(\boldsymbol{x}^\ell_{\mathsf{Cand}}):\ell=1,\dots,L\big\}$.

Figure 2 shows the efficient frontier of attainable expected revenue-utility pairs for a certain problem instance. The crosses correspond to the pairs $\{(\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r}),\mathsf{Util}(\boldsymbol{x}^*_\lambda)):\lambda\ge 0\}$. Naturally, larger expected utility comes at the expense of smaller expected revenue.

5. A Discretization Method for Dealing with Many Candidate Assortments

When the number of candidate assortments is large, a natural question is whether we can come up with a smaller collection of candidate assortments while ensuring that we can obtain a near-optimal solution to the Revenue-Utility problem by using the smaller collection. In this section, we give a discretization method that allows us to consider only a small number of candidate assortments while controlling the quality of the solutions we obtain by doing so. By Theorem 3.2, if $(t^*,\boldsymbol{x}^*)$

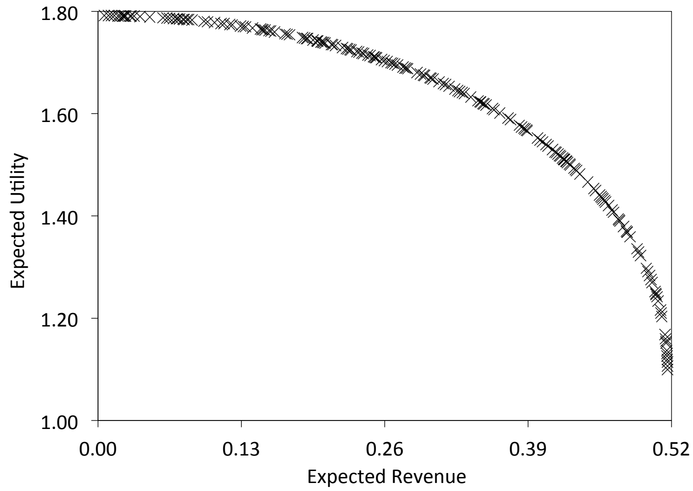

Figure 2 The efficient frontier for a particular problem instance with $n = 100$, $\mathcal{F} = \{\boldsymbol{x} \in \{0,1\}^n : \sum_{i \in \mathcal{N}} x_i \leq 50\}$, $v_0 = 1$, and for each $i \in \mathcal{N}$, r_i is uniformly generated over $[0,1]$ and $v_i = 1 - r_i^2$.

is an optimal solution to the Parametric problem, then we can obtain an optimal solution to the Revenue-Utility problem by solving the problem $\max_{\boldsymbol{x} \in \mathcal{F}} \mathsf{Rev}(\boldsymbol{x}; \boldsymbol{r} + \lambda\,(1+t^*)\,\boldsymbol{e})$. In our discretization method, we build a one-dimensional geometric grid that covers the interval $[V_{\min}, nV_{\max}]$. For each value of t in the grid, we solve the problem $\max_{\boldsymbol{x} \in \mathcal{F}} \mathsf{Rev}(\boldsymbol{x}; \boldsymbol{r} + \lambda\,(1+t)\,\boldsymbol{e})$ to get a candidate assortment. We check the objective value of each candidate assortment for the Revenue-Utility problem and pick the best one. The formal description of our discretization method is presented below.

Discretization Method for Revenue-Utility Maximization

Initialization: Pick a grid size $\rho > 0$. Using $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ to denote the round up and round down functions, define the geometric grid over the interval $[V_{\min}, nV_{\max}]$ as

$$\mathsf{Grid} = \left\{ (1+\rho)^k \ : \ k = \lceil \log V_{\min} / \log(1+\rho) \rceil, \ldots, \lfloor \log(nV_{\max}) / \log(1+\rho) \rfloor \right\} \cup \left\{ V_{\min}, nV_{\max} \right\}.$$

Description of the Method: For each $t \in \mathsf{Grid}$, let $\widehat{\boldsymbol{x}}_t$ be an optimal solution to the problem $\max_{\boldsymbol{x} \in \mathcal{F}} \ \mathsf{Rev}(\boldsymbol{x}, \boldsymbol{r} + \lambda\,(1+t)\,\boldsymbol{e})$.

Output: Return the assortment $\widehat{\boldsymbol{x}}$ from the collection $\{\widehat{\boldsymbol{x}}_t : t \in \mathsf{Grid}\}$ with the largest objective value for the Revenue-Utility problem; that is, $\widehat{\boldsymbol{x}} = \arg\max_{t \in \mathsf{Grid}} \ \mathsf{Rev}(\widehat{\boldsymbol{x}}_t, \boldsymbol{r}) + \lambda \log(1 + V(\widehat{\boldsymbol{x}}_t))$.

Our main results are stated in the next two theorems. The first theorem shows that the output of the discretization method provides a $1/(1+\rho)$-approximate solution to the Revenue-Utility problem. The second theorem shows that for each $t \in \mathsf{Grid}$, $\widehat{\boldsymbol{x}}_t$ can be computed by solving an LP.

Theorem 5.1 (Performance) *The output* $\widehat{\boldsymbol{x}}$ *is a* $1/(1+\rho)$*-approximation to the* Revenue-Utility *problem; that is,* $\mathsf{Rev}\big(\widehat{\boldsymbol{x}};\boldsymbol{r}\big)+\lambda\log(1+V(\widehat{\boldsymbol{x}}))\ge Z^*_\lambda/(1+\rho)$.

Proof: Let $\boldsymbol{x}^*$ be an optimal solution to the Revenue-Utility problem. Since $V(\boldsymbol{x}^*)\in[V_{\min},nV_{\max}]$, let $\widehat{t}\in\mathsf{Grid}$ be such that $\widehat{t}\le V(\boldsymbol{x}^*)\le(1+\rho)\,\widehat{t}$. By the definition of $\widehat{\boldsymbol{x}}$, we get

$$\begin{aligned}\mathsf{Rev}\left(\widehat{\boldsymbol{x}}\,;\boldsymbol{r}\right)+\lambda\,\log\big(1+V(\widehat{\boldsymbol{x}})\big) &\ge \mathsf{Rev}\left(\widehat{\boldsymbol{x}}_{\widehat{t}}\,;\boldsymbol{r}\right)+\lambda\,\log\big(1+V(\widehat{\boldsymbol{x}}_{\widehat{t}})\big)\\ &\overset{(a)}{\ge} \mathsf{Rev}\big(\widehat{\boldsymbol{x}}_{\widehat{t}}\,;\boldsymbol{r}+\lambda\,(1+\widehat{t}\,)\,\boldsymbol{e}\big)+\lambda\,\big(\log(1+\widehat{t}\,)-\widehat{t}\,\big)\\ &\overset{(b)}{\ge} \mathsf{Rev}\big(\boldsymbol{x}^*\,;\boldsymbol{r}+\lambda\,(1+\widehat{t}\,)\,\boldsymbol{e}\big)+\lambda\,\big(\log(1+\widehat{t}\,)-\widehat{t}\,\big),\end{aligned}$$

where (a) follows from using Lemma 3.1 with $\boldsymbol{x}=\widehat{\boldsymbol{x}}_{\widehat{t}}$ and $t=\widehat{t}$, whereas (b) follows because, by definition of the discretization method, $\widehat{\boldsymbol{x}}_{\widehat{t}}=\arg\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x}\,;\boldsymbol{r}+\lambda\,(\,1+\widehat{t}\,)\,\boldsymbol{e})$.

We have $\mathsf{Rev}(\boldsymbol{x}^*;\boldsymbol{r}+\lambda\,(\,1+\widehat{t}\,)\,\boldsymbol{e})=\mathsf{Rev}(\boldsymbol{x}^*;\boldsymbol{r})+\frac{\lambda\,(\,1+\widehat{t}\,)\,V(\boldsymbol{x}^*)}{1+V(\boldsymbol{x}^*)}$ by the identity at the beginning of the proof of Lemma 3.1, and thus,

$$\begin{aligned}\mathsf{Rev}\big(\boldsymbol{x}^*\,;\boldsymbol{r}+\lambda\,(1+\widehat{t}\,)\,\boldsymbol{e}\big)+\lambda\,\big(\log(1+\widehat{t}\,)-\widehat{t}\,\big) &= \mathsf{Rev}\left(\boldsymbol{x}^*;\boldsymbol{r}\right)+\lambda\left\{\frac{(1+\widehat{t}\,)V(\boldsymbol{x}^*)}{1+V(\boldsymbol{x}^*)}-\widehat{t}\,\right\}+\lambda\,\log\big(1+\widehat{t}\,\big)\\ &\overset{(c)}{\ge} \mathsf{Rev}\left(\boldsymbol{x}^*;\boldsymbol{r}\right)+\lambda\,\log\big(1+\widehat{t}\,\big)\\ &\overset{(d)}{\ge} \mathsf{Rev}\left(\boldsymbol{x}^*;\boldsymbol{r}\right)+\lambda\,\log\left(1+\frac{V(\boldsymbol{x}^*)}{1+\rho}\right),\end{aligned}$$

where (c) holds because $\frac{a}{1+a}$ is increasing in a and $V(\boldsymbol{x}^*)\ge\widehat{t}$, so we get $\frac{V(\boldsymbol{x}^*)}{1+V(\boldsymbol{x}^*)}\ge\frac{\widehat{t}}{1+\widehat{t}}$, and (d) follows from the fact that $(1+\rho)\,\widehat{t}\ge V(\boldsymbol{x}^*)$.

For all $a\in\mathbb{R}_+$ and $\rho\in\mathbb{R}_+$, we have the inequality $(1+a)^{1+\rho}\ge1+(1+\rho)\,a$. Using this inequality with $a=V(\boldsymbol{x}^*)/(1+\rho)$, we get $\left(1+\frac{V(\boldsymbol{x}^*)}{1+\rho}\right)^{1+\rho}\ge1+V(\boldsymbol{x}^*)$. Therefore, we have

$$\begin{aligned}\mathsf{Rev}(\boldsymbol{x}^*;\boldsymbol{r})\,+\,\lambda\,\log\left(1+\frac{V(\boldsymbol{x}^*)}{1+\rho}\right) &\ge \mathsf{Rev}(\boldsymbol{x}^*;\boldsymbol{r})\,+\,\frac{\lambda\,\log(1+V(\boldsymbol{x}^*))}{1+\rho}\\ &\overset{(e)}{\ge} \frac{1}{1+\rho}\Big\{\mathsf{Rev}(\boldsymbol{x}^*;\boldsymbol{r})\,+\,\lambda\,\log(1+V(\boldsymbol{x}^*))\Big\}\\ &= \frac{Z^*_\lambda}{1+\rho},\end{aligned}$$

where (e) holds because $\mathsf{Rev}(\boldsymbol{x}^*;\boldsymbol{r})\ge0$. By the three chains of displayed inequalities above, we have $\mathsf{Rev}\big(\widehat{\boldsymbol{x}};\boldsymbol{r}\big)+\lambda\,\log(1+V(\widehat{\boldsymbol{x}}))\ge Z^*_\lambda/(1+\rho)$, as desired. ■

Since our discretization method constructs one candidate assortment for each point in Grid and there are $O\Big(\frac{\log(nV_{\max})-\log(V_{\min})}{\log(1+\rho)}\Big)=O\left(\frac{1}{\rho}\log(nV_{\max}/V_{\min})\right)$ points in Grid, we can find a $1/(1+\rho)$-approximate solution to the Revenue-Utility problem by checking the objective values of $O\left(\frac{1}{\rho}\log(nV_{\max}/V_{\min})\right)$ candidate assortments. In the discretization method, we need to solve the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\ \mathsf{Rev}(\boldsymbol{x},\boldsymbol{r}+\lambda\,(1+t)\,\boldsymbol{e})$ for each value of $t\in\mathsf{Grid}$. This problem finds an assortment that maximizes only the expected revenue under totally unimodular constraints on offered assortments. Setting $\alpha=\lambda\,(1+t)$ for notational brevity, the last problem is of the form $\max_{\boldsymbol{x}\in\mathcal{F}}\ \mathsf{Rev}(\boldsymbol{x},\boldsymbol{r}+\alpha\,\boldsymbol{e})$. Using the decision variables $\boldsymbol{y}=(y_1,\ldots,y_n)\in\mathbb{R}_+^n$ and $y_0\in\mathbb{R}_+$, the next theorem shows that the optimal objective value of the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x},\boldsymbol{r}+\alpha\,\boldsymbol{e})$ can be obtained by solving the following LP:

$$Z_{\mathsf{Rev}}=\max_{(\boldsymbol{y},y_0)\in\mathbb{R}_+^{n+1}}\left\{\sum_{i\in\mathcal{N}}(r_i+\alpha)\,v_i\,y_i\ \middle|\ \mathbf{A}\boldsymbol{y}\le y_0\,\boldsymbol{b},\ y_i\le y_0\ \forall\, i\in\mathcal{N},\ v_0\,y_0+\sum_{i\in\mathcal{N}}v_i\,y_i=1\right\}.\quad(\mathsf{Revenue\ LP})$$

Moreover, after obtaining the optimal objective value of the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\alpha\,\boldsymbol{e})$, we can obtain an optimal solution to this problem simply by solving the Candidate LP.

Theorem 5.2 (LP for Revenue Optimization) *Let γ^* denote the optimal objective value of the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\alpha\,\boldsymbol{e})$. Then, the optimal objective value of the* Revenue LP *is also γ^*; that is, $Z_{\mathsf{Rev}}=\gamma^*$. Furthermore, if $\boldsymbol{x}^*$ is an optimal solution to the* Candidate LP *with $\gamma=\gamma^*-\alpha$, then $\boldsymbol{x}^*$ is also an optimal solution to the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\alpha\,\boldsymbol{e})$.*

Proof: We first claim that $v_0\,\gamma^*=\mathsf{LP}(\gamma^*-\alpha)$, where $\mathsf{LP}(\gamma^*-\alpha)$ is the optimal objective value of the Candidate LP with $\gamma=\gamma^*-\alpha$. Let $\widetilde{\boldsymbol{x}}$ be an optimal solution to $\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x},\boldsymbol{r}+\alpha\,\boldsymbol{e})$, so

$$\gamma^*\ =\ \mathsf{Rev}(\widetilde{\boldsymbol{x}}\,;\,\boldsymbol{r}+\alpha\,\boldsymbol{e})\ =\ \frac{\sum_{i\in\mathcal{N}}(r_i+\alpha)\,v_i\,\widetilde{x}_i}{v_0+\sum_{i\in\mathcal{N}}v_i\,\widetilde{x}_i}\ .$$

Focusing on the first and third expressions above and rearranging the terms, we get $v_0\,\gamma^*=\sum_{i\in\mathcal{N}}(r_i-\gamma^*+\alpha)\,v_i\,\widetilde{x}_i$. Since $\widetilde{\boldsymbol{x}}$ is a feasible, but not necessarily an optimal, solution to the Candidate LP with $\gamma=\gamma^*-\alpha$, we have $\mathsf{LP}(\gamma^*-\alpha)\ge\sum_{i\in\mathcal{N}}(r_i-\gamma^*+\alpha)\,v_i\,\widetilde{x}_i=v_0\,\gamma^*$. Hence, $v_0\,\gamma^*\le\mathsf{LP}(\gamma^*-\alpha)$. On the other hand, letting $\boldsymbol{x}^*$ be an optimal solution to the Candidate LP with $\gamma=\gamma^*-\alpha$, since $\boldsymbol{x}^*$ is a feasible, but not necessarily an optimal, solution to the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\alpha\,\boldsymbol{e})$, we have $\gamma^*\ge\mathsf{Rev}(\boldsymbol{x}^*;\boldsymbol{r}+\alpha\,\boldsymbol{e})=\frac{\sum_{i\in\mathcal{N}}(r_i+\alpha)\,v_i\,x_i^*}{v_0+\sum_{i\in\mathcal{N}}v_i\,x_i^*}$, in which case focusing on the first and third expressions in this chain of inequalities and rearranging the terms, we get $v_0\,\gamma^*\ge\sum_{i\in\mathcal{N}}(r_i-\gamma^*+\alpha)\,v_i\,x_i^*=\mathsf{LP}(\gamma^*-\alpha)$. Hence, $v_0\,\gamma^*\ge\mathsf{LP}(\gamma^*-\alpha)$. This proves the claim.

Note that Z_{Rev} is the optimal objective value of the Revenue LP. First, we show that $Z_{\mathsf{Rev}}\ge\gamma^*$. Let $\widetilde{\boldsymbol{x}}$ be an optimal solution to the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\alpha\,\boldsymbol{e})$. Setting $\widetilde{V}=v_0+\sum_{i\in\mathcal{N}}v_i\,\widetilde{x}_i$ for notational brevity, we define the solution $(\widetilde{\boldsymbol{y}},y_0)$ to the Revenue LP as $y_i=\widetilde{x}_i/\widetilde{V}$ for all $i\in\mathcal{N}$ and $\widetilde{y}_0=1/\widetilde{V}$. Since $\widetilde{\boldsymbol{x}}\in\mathcal{F}$, we have $\mathbf{A}\widetilde{\boldsymbol{x}}\le\boldsymbol{b}$, in which case it is straightforward to check that the solution $(\widetilde{\boldsymbol{y}},\widetilde{y}_0)$ is feasible to the Revenue LP. Moreover, for the Revenue LP, this solution provides an objective value of $\sum_{i\in\mathcal{N}}(r_i+\alpha)\,v_i\,\widetilde{y}_i=\sum_{i\in\mathcal{N}}(r_i+\alpha)\,v_i\,\widetilde{x}_i/\widetilde{V}=\mathsf{Rev}(\widetilde{\boldsymbol{x}};\boldsymbol{r}+\alpha\,\boldsymbol{e})=\gamma^*$. Hence, we have $Z_{\mathsf{Rev}}\geq\gamma^*$.

Second, we show that $Z_{\mathsf{Rev}}\leq\gamma^*$. Using the dual variables $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_m)$, $\boldsymbol{\sigma}=(\sigma_1,\ldots,\sigma_n)$, and γ, the duals of the Revenue LP and the Candidate LP with $\gamma=\gamma^*-\alpha$ are

$$Z_{\mathsf{Rev}} \;=\; \min_{(\boldsymbol{\mu},\boldsymbol{\sigma},\gamma)\in\mathbb{R}_+^{m+n}\times\mathbb{R}} \left\{\gamma \;\middle|\; \sum_{\ell\in\mathcal{M}} a_{\ell i}\,\mu_\ell+\sigma_i+v_i\,\gamma\geq(r_i+\alpha)\,v_i \quad \forall\, i\in\mathcal{N},\quad \sum_{\ell\in\mathcal{M}} b_\ell\,\mu_\ell+\sum_{i\in\mathcal{N}}\sigma_i=v_0\,\gamma\right\},$$

$$\mathsf{LP}(\gamma^*-\alpha) \;=\; \min_{(\boldsymbol{\mu},\boldsymbol{\sigma})\in\mathbb{R}_+^{m+n}} \left\{\sum_{\ell\in\mathcal{M}} b_\ell\,\mu_\ell+\sum_{i\in\mathcal{N}}\sigma_i \;\middle|\; \sum_{\ell\in\mathcal{M}} a_{\ell i}\,\mu_\ell+\sigma_i\geq(r_i-\gamma^*+\alpha)\,v_i \quad \forall\, i\in\mathcal{N}\right\},$$

where we use $\mathcal{M}=\{1,\ldots,m\}$ to index the rows of the matrix $\mathbf{A}$ and let $a_{\ell i}$ be the $(\ell,i)^{\text{th}}$ entry of $\mathbf{A}$ and b_ℓ be the ℓ^{th} entry of $\boldsymbol{b}$. As long as $\mathcal{F}$ includes a nonempty assortment, the Revenue LP and the Candidate LP are feasible and bounded, so strong duality holds and the optimal objective values of the duals are equal to those of the primals. Let $(\boldsymbol{\mu}^*,\boldsymbol{\sigma}^*)$ be an optimal solution to the dual of the Candidate LP above. Therefore, we have $\sum_{\ell\in\mathcal{M}} a_{\ell i}\,\mu_\ell^*+\sigma_i^*\geq(r_i-\gamma^*+\alpha)\,v_i$ for all $i\in\mathcal{N}$ and $\sum_{\ell\in\mathcal{M}} b_\ell\,\mu_\ell^*+\sum_{i\in\mathcal{N}}\sigma_i^*=\mathsf{LP}(\gamma^*-\alpha)=v_0\,\gamma^*$, where the last equality uses the fact that $\mathsf{LP}(\gamma^*-\alpha)=v_0\,\gamma^*$, as shown at the beginning of the proof. The last inequality and chain of equalities show that the solution $(\boldsymbol{\mu}^*,\boldsymbol{\sigma}^*,\gamma^*)$ is feasible to the dual of the Revenue LP above. Furthermore, for the dual of the Revenue LP above, this solution provides an objective value of γ^*. Hence, $Z_{\mathsf{Rev}}\leq\gamma^*$. Thus, we get $Z_{\mathsf{Rev}}=\gamma^*$, so the optimal objective value of the Revenue LP is γ^*.

Lastly, we show that if $\boldsymbol{x}^*$ is an optimal solution to the Candidate LP with $\gamma=\gamma^*-\alpha$, then $\boldsymbol{x}^*$ is also an optimal solution to the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\alpha\,\boldsymbol{e})$. Using the fact that $v_0\,\gamma^*=\mathsf{LP}(\gamma^*-\alpha)$, we have $v_0\,\gamma^*=\mathsf{LP}(\gamma^*-\alpha)=\sum_{i\in\mathcal{N}}(r_i-\gamma^*+\alpha)\,v_i\,x_i^*$. Solving for γ^* in this chain of equalities, we get $\gamma^*=\frac{\sum_{i\in\mathcal{N}}(r_i+\alpha)\,v_i\,x_i^*}{v_0+\sum_{i\in\mathcal{N}}v_i\,x_i^*}=\mathsf{Rev}(\boldsymbol{x}^*;\boldsymbol{r}+\alpha\,\boldsymbol{e})$. Thus, since $\gamma^*=\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\alpha\,\boldsymbol{e})$, $\boldsymbol{x}^*$ is an optimal solution to the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\alpha\,\boldsymbol{e})$. ■

In certain applications, we may be interested in finding an assortment that maximizes only the expected revenue. By Theorem 5.2, it follows that in such cases we can simply solve an LP.

6. Applications of Totally Unimodular Constraints

In this section, we give examples of assortment optimization settings that fit our formulation with totally unimodular constraints. For each of these settings, we can use our approach to solve the Revenue-Utility problem, yielding an assortment that maximizes the linear combination of the expected revenue and expected utility. Of course, setting $\lambda=0$ in the Revenue-Utility problem, we obtain an assortment that maximizes only the expected revenue.

6.1 Cardinality Constraints and Their Variants

In certain applications, due to limited space in a physical store or on a web page, we are interested in limiting the cardinality of the offered assortment. Using b to denote the upper bound on the number of products that we can offer, the set of feasible assortments is $\mathcal{F} = \{\boldsymbol{x} \in \{0,1\}^n : \sum_{i\in\mathcal{N}} x_i \le b\}$. Here, the constraint matrix $\mathbf{A} = (1,\ldots,1)$ has a single row consisting of all ones. Therefore, $\mathbf{A}$ is totally unimodular, in which case we can use our approach to find an assortment that maximizes the linear combination of the expected revenue and the expected utility. It turns out that our approach can handle slightly more general cardinality constraints, which we call nested cardinality constraints. In particular, consider the case in which we have a collection of subsets of products $\{S_k \subseteq \mathcal{N} : k = 1,\ldots,K\}$, where for any pair of subsets, either one subset includes the other or their intersection is empty; that is, for all $k, \ell = 1,\ldots,K$, we have $S_k \subseteq S_\ell$ or $S_\ell \subseteq S_k$ or $S_k \cap S_\ell = \varnothing$. In nested cardinality constraints, the cardinality of the products that we can offer within each subset S_k is limited to b_k. Thus, the set of feasible assortments is $\mathcal{F} = \{\boldsymbol{x} \in \{0,1\}^n : \sum_{i\in S_k} x_i \le b_k \ \ \forall\, k = 1,\ldots,K\}$. Using the fact that $S_k \subseteq S_\ell$ or $S_\ell \subseteq S_k$ or $S_k \cap S_\ell = \varnothing$, we can arrange the columns of this constraint matrix in such a way that each row includes only consecutive ones. Such a matrix is called an interval matrix, and it is totally unimodular; see Corollary 2.10 in Chapter III.1 in Nemhauser and Wolsey (1988).

For example, if we choose an assortment of shirts to offer with S_1 being the set of all available shirts, S_2 being the set of all long-sleeved shirts, and S_3 being the set of all short-sleeved shirts, then $S_2 \subseteq S_1$, $S_3 \subseteq S_1$, and $S_2 \cap S_3 = \varnothing$. In this case, the nested cardinality constraints ensure that the number of offered shirts is at most b_1, the number of offered long-sleeved shirts is at most b_2, and the number of offered short-sleeved shirts is at most b_3. If, in addition, S_4 is the set of all blue long-sleeved shirts, then $S_4 \subseteq S_1$, $S_4 \subseteq S_2$, and $S_4 \cap S_3 = \varnothing$, in which case the nested cardinality constraints ensure that the number of offered blue long-sleeved shirts is at most b_4.

6.2 Display Location Effects

Consider the case in which the preference weight of each product depends on the location where the product is displayed. In brick and mortar retail, when a product is displayed at a prominent location, it is more likely to be noticed by customers than when it is displayed at an inconspicuous location. In online retail, when a product is displayed at the top of the search results, it is more likely to be chosen by customers than when it is displayed at the bottom. To model the display location effects, we use $\mathcal{N} = \{1,2,\ldots,n\}$ to index the items that we can offer to the customers. If we display item i at location ℓ, then its preference weight is $v_{i\ell}$. We use $\mathcal{N}$ also to index the possible locations at which we can display the items. If the number of possible locations is less than the number of items, then we can define additional locations with $v_{i\ell} = 0$ for each item i and for each additional location ℓ, in which case displaying an item at an additional location is equivalent to not displaying the item at all. We capture our assortment decisions by $\boldsymbol{x} = \{x_{i\ell} : i \in \mathcal{N},\ \ell \in \mathcal{N}\} \in \{0,1\}^{n\times n}$, where $x_{i\ell} = 1$ if and only if we offer item i at location ℓ. Therefore, the set of products is $\mathcal{N} \times \mathcal{N}$, and offering the product (i,ℓ) corresponds to displaying item i at location ℓ.

The expected revenue and the expected utility from our assortment decisions, along with the set of feasible assortments, are given by

$$\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) \;=\; \frac{\sum_{(i,\ell)\in\mathcal{N}\times\mathcal{N}} r_i\, v_{i\ell}\, x_{i\ell}}{v_0 + \sum_{(i,\ell)\in\mathcal{N}\times\mathcal{N}} v_{i\ell}\, x_{i\ell}}, \qquad \mathsf{Util}(\boldsymbol{x}) = \log\Big(1 \;+ \sum_{(i,\ell)\in\mathcal{N}\times\mathcal{N}} \frac{v_{i\ell}}{v_0}\, x_{i\ell}\Big)$$

$$\mathcal{F} = \left\{ \boldsymbol{x} \in \{0,1\}^{n\times n} \;:\; \sum_{\ell\in\mathcal{N}} x_{i\ell} \le 1 \;\; \forall\, i \in \mathcal{N}, \;\; \sum_{i\in\mathcal{N}} x_{i\ell} \le 1 \;\; \forall\, \ell \in \mathcal{N} \right\},$$

where the first constraint ensures that each item is displayed in at most one location and the second constraint ensures that each location is used by at most one item. Here, the constraint matrix is the constraint matrix of an assignment problem, which is known to be totally unimodular; see Corollary 2.9 in Chapter III.1 in Nemhauser and Wolsey (1988). Note that if the locations have a natural sequence, as in online search results, then our formulation allows skipping a location, but if $v_{i1} \ge v_{i2} \ge \ldots \ge v_{in}$ for all $i \in \mathcal{N}$, so that locations with smaller indices are more preferable, then it is straightforward to show that it is not optimal to skip any of the locations.

6.3 Pricing with Discrete Price Menus

In our problem setup up to this point, the prices of the products are fixed. Consider the case in which the price of each product is a decision variable rather than being fixed. The preference weight of each product depends on its price. Given a finite set of possible price levels, we want to choose the assortment of products to offer and the corresponding prices. We use $\mathcal{N} = \{1,2,\ldots,n\}$ to index the items that we can offer to the customers and $\mathcal{K} = \{1,2,\ldots,K\}$ to index the possible price levels that we can choose for the items. The price that corresponds to price level k is r_k, so the set of possible prices for the items is $\{r_k : k \in \mathcal{K}\}$. If we use the price level k for item i, then its preference weight is v_{ik}. Note that we do not require a specific functional form between the preference weight of an item and its price. The price-demand relationship can be arbitrary. Our notation indicates that the set of possible prices for each item is the same, but it is straightforward to extend our formulation to incorporate different sets of possible prices for different items. To capture our assortment decisions, we use the vector $\boldsymbol{x} = \{x_{ik} : i \in \mathcal{N},\ k \in \mathcal{K}\} \in \{0,1\}^{n\times K}$, where

$x_{ik}=1$ if and only if we offer item i at price level k. In this case, the set of products is $\mathcal{N}\times\mathcal{K}$, and offering the product (i,k) corresponds to offering item i at price level k.

The expected revenue and the expected utility from our assortment decisions and the set of feasible assortments are given by

$$\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) \;=\; \frac{\sum_{(i,k)\in\mathcal{N}\times\mathcal{K}} r_k\, v_{ik}\, x_{ik}}{v_0+\sum_{(i,k)\in\mathcal{N}\times\mathcal{K}} v_{ik}\, x_{ik}}, \qquad \mathsf{Util}(\boldsymbol{x})=\log\Big(1\;+\sum_{(i,k)\in\mathcal{N}\times\mathcal{K}}\frac{v_{ik}}{v_0}\,x_{ik}\Big)$$

$$\mathcal{F}=\left\{\boldsymbol{x}\in\{0,1\}^{n\times K}\;:\;\sum_{k\in\mathcal{K}} x_{ik}\le 1\quad\forall\, i\in\mathcal{N}\right\}.$$

The constraint ensures that each item, if offered, has one price level. Each row of the constraint matrix corresponds to an item i and includes consecutive ones, corresponding to the different price levels for item i. Thus, the constraint matrix is an interval matrix, which is totally unimodular. If we want to impose a limit of b on the number of products we offer, then we can add the constraint $\sum_{(i,k)\in\mathcal{N}\times\mathcal{K}} x_{ik}\le b$ to the constraints above. The additional constraint amounts to adding a row of ones to the constraint matrix, in which case the constraint matrix remains totally unimodular. In many formulations of the pricing problem under the multinomial logit model, there is a specific relationship between the prices and the preference weights. For example, a common approach is that if the price of the product i is p, then its preference weight is $e^{\alpha_i-\beta_i\, p}$, which arises when the mean utility of the product is linear in its price. In our formulation, the relationship between the prices and the preference weights can be arbitrary.

6.4 Pricing with a Price Ladder Constraint

We can extend the pricing model presented in the previous section to accommodate a price ladder constraint that imposes an ordering of the prices. Suppose there is an inherent ordering $1\succ 2\succ\ldots\succ n$ among the products, where, in some sense, product 1 is the best product and product n is the worst. Such an ordering occurs when the products have a clear ordering in terms of quality, richness of features, or durability. We want to choose the prices of the products in a way that is consistent with their rank; that is, better products have higher prices. We refer to such constraints on the prices as price ladder or quality consistency constraints. Price ladder constraints appear in practice frequently. Rusmevichientong et al. (2006) describe an application in automobile pricing in which a vehicle with more features must have a higher price than a vehicle of the same model with fewer features. Jagabathula and Rusmevichientong (2017) present additional applications of price ladder constraints. We index the items by $\mathcal{N}=\{1,\ldots,n\}$ and the possible price levels by $\mathcal{K}=\{1,\ldots,K\}$. Without loss of generality, we order the prices $\{r_k : k\in\mathcal{K}\}$ corresponding to the different price levels so that $r_1\ge r_2\ge\ldots\ge r_K$. If the price of product i is r_k, then its preference weight is v_{ik}. The price ladder constraint is such that the price of product i should be no larger than the price of product $i-1$. To capture our assortment decisions, we use the vector $\boldsymbol{x} = \{x_{ik\ell} : i \in \mathcal{N},\ k \in \mathcal{K},\ \ell \in \mathcal{K},\ k \geq \ell\} \in \{0,1\}^{n \times K \times (K+1)/2}$, where $x_{ik\ell} = 1$ if and only if we offer item i at price level k and item $i-1$ at price level ℓ. Therefore, we offer item i at price level k if and only if $\sum_{\ell=1}^{k} x_{ik\ell} = 1$. For item $i = 1$, although we do not have an item indexed by zero, we still use the decision variables $\{x_{1k\ell} : k, \ell \in \mathcal{K},\ k \geq \ell\}$ to capture our pricing decisions for item 1. We offer item 1 at price level k if and only if $\sum_{\ell=1}^{k} x_{1k\ell} = 1$.

The expected revenue and the expected utility from our assortment decisions and the set of feasible assortments are given by

$$\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) \;=\; \frac{\sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{K}}\sum_{\ell=1}^{k} r_k\, v_{ik}\, x_{ik\ell}}{v_0 + \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{K}}\sum_{\ell=1}^{k} v_{ik}\, x_{ik\ell}}, \qquad \mathsf{Util}(\boldsymbol{x}) = \log\Big(1 + \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{K}}\sum_{\ell=1}^{k} \frac{v_{ik}}{v_0}\, x_{ik\ell}\Big)$$

$$\mathcal{F} = \left\{ \boldsymbol{x} \in \{0,1\}^{n\times K\times(K+1)/2} \;:\; \sum_{k\in\mathcal{K}}\sum_{\ell=1}^{k} x_{1k\ell} = 1, \quad \sum_{\ell=1}^{k} x_{i-1,k\ell} \;=\; \sum_{\ell=k}^{K} x_{i\ell k} \;\; \forall\, i = 2,\ldots,n,\ k \in \mathcal{K} \right\}.$$

The first constraint ensures that we charge some price for item 1. The second constraint ensures that if we charge the price level k for item $i-1$, then we can charge one of the price levels $k,\ldots,K$ for item i. In particular, the left side of the constraint takes a value of one if we charge price level k for item $i-1$. Noting the definition of the decision variable $x_{ik\ell}$, the right side of the constraint takes a value of one if we charge one of the price levels $k,\ldots,K$ for item i and the price level k for item $i-1$. In the next proposition, we show that the constraints above correspond to flow balance constraints over a certain graph, which implies that the constraint matrix is totally unimodular; see Proposition 3.1 in Chapter III.1 in Nemhauser and Wolsey (1988).

Proposition 6.1 (Connection to Network Flow) *The constraints for the price ladder are flow balance constraints in a directed graph with $(n-1)K + 2$ vertices and $nK(K+1)/2$ edges.*

Proof: Consider a directed graph whose vertices are indexed by $\{(i,k) : i = 2,\ldots,n,\ k \in \mathcal{K}\} \cup \{\mathsf{source}, \mathsf{sink}\}$ and whose edges are indexed by $\{(i,k,\ell) : i \in \mathcal{N},\ k \in \mathcal{K},\ \ell \in \mathcal{K},\ k \geq \ell\}$. For $i = 2,\ldots,n-1$, the edge (i,k,ℓ) leaves the vertex (i,ℓ) and enters the vertex $(i+1,k)$. The edge $(1,k,\ell)$ leaves the vertex source and enters the vertex $(2,k)$. The edge (n,k,ℓ) leaves the vertex (n,ℓ) and enters the vertex sink. The decision variable $x_{ik\ell}$ in our price ladder constraints corresponds to the flow on the edge (i,k,ℓ). The first constraint is the flow balance constraint of the vertex source. The second constraint is the flow balance constraint of the vertex (i,k). The supply at the vertex source is $+1$. The demand at the vertex sink is -1. The flow balance constraint for the vertex sink is redundant, so we do not write these constraints explicitly in our formulation of the price ladder

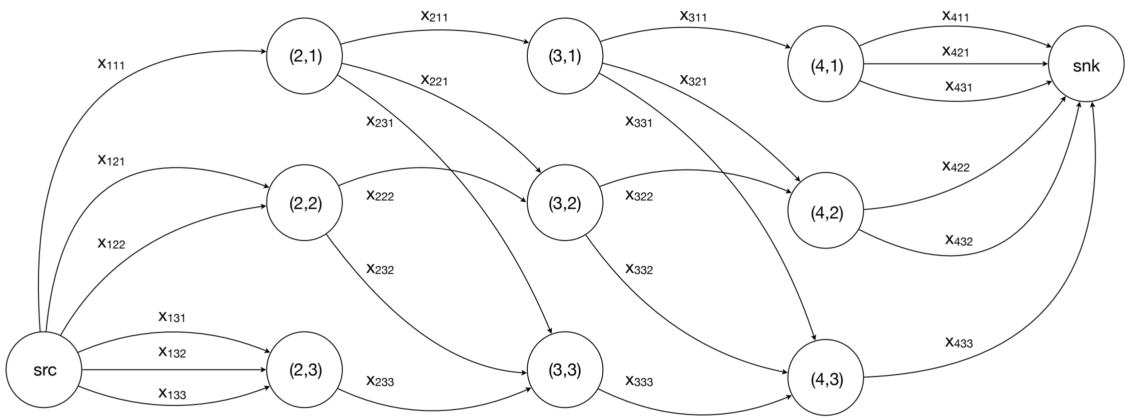

Figure 3 The graph corresponding to the price ladder constraints for a problem instance with $n=4$ products and $K=3$ price levels.

constraints. In Figure 3, we show the construction of the graph for the price ladder constraints for a problem instance with $n=4$ products and $K=3$ price levels. ■

In our price ladder, we use a complete ordering of the products with $1 \succ 2 \succ \ldots \succ n$. By slightly modifying the graph in Figure 3, we can handle a partial ordering. For example, consider the partial ordering $1 \succ \{2,3,4\} \succ 5$, meaning that item 1 must have a price higher than the prices of items 2, 3, and 4, but there is no fixed ordering among the prices of items 2, 3, and 4. Building on the approach described in this section, we can present this partial order using totally unimodular constraints as well. Lastly, in our formulation, the only decision variable is the price to charge for each item. Once again, we can slightly modify the graph in Figure 3 to handle the case in which we choose the products to offer, as well as their prices, while satisfying a price ladder constraint.

6.5 Product Precedence Constraints

We focus on assortment problems in which a particular product cannot be offered unless a certain set of related products is also offered. This kind of a constraint may arise when a company is prohibited, by company policy or law, from offering a more expensive or sophisticated version of the product unless an inexpensive or basic version is also offered. For example, it may not be possible to offer the brand name version of a drug unless the generic version is also offered. To model such product precedence constraints, we use $S_i \subseteq \mathcal{N}$ to denote the set of products that we need to offer in order to be able to offer product i. Thus, the set of feasible assortments is given by $\mathcal{F} = \{\boldsymbol{x} \in \{0,1\}^n : x_i - x_j \leq 0 \ \forall\, i \in \mathcal{N},\ j \in S_i\}$, indicating that we can have $x_i = 1$ only when $x_j = 1$ for all $j \in S_i$. In this constraint matrix, each row includes only a $+1$ and a -1. Such matrices are known to be totally unimodular; see Proposition 2.6 in Chapter III.1 in Nemhauser and Wolsey (1988). Note that the subsets $\{S_i : i \in \mathcal{N}\}$ in the product precedence constraints can be arbitrary. In particular, they can be overlapping and products can have circular dependencies on each other.

7. Computational Experiments

We perform computational experiments on assortment problems with display location effects as well as on pricing problems with discrete price menus. For both problem classes, we work with a large number of randomly generated problem instances. Our goal is to demonstrate that our parametric LP is a viable approach for generating candidate assortments and that the practical performance of our discretization method is substantially better than its theoretical guarantee.

7.1 Computational Results for Display Location Effects

We focus on assortment problems with display location effects, which correspond to the application discussed in Section 6.2. We use the following approach to generate our test problems. In all test problems, we have $n = 60$ items. There are K possible locations indexed by $\mathcal{K} = \{1, \ldots, K\}$, where K is a parameter that we vary. We follow the convention that location 1 is the most desirable and location K is the least desirable. We sample the revenue r_i of item i from the uniform distribution over $[0, 10]$. To come up with the preference weights associated with the item-location pairs, we sample β from the uniform distribution over $[0, 1]$. For each item i, we sample α_i from the uniform distribution over $[0, 2]$ and set the preference weight v_{ik} of item i when offered at location k as $v_{ik} = e^{\alpha_i + (0.1 \times (K-k)) - (\beta \times r_i)}$. Therefore, the items with larger prices tend to be less attractive, and β captures the price sensitivity. To come up with the preference weight of the no-purchase option, letting $\mathcal{N}_K$ be the set of K items with the smallest values for $\{v_{iK} : i \in \mathcal{N}\}$, we set $v_0 = p_0 \sum_{i \in \mathcal{N}_K} v_{iK} / (1 - p_0)$, where p_0 is another parameter that we vary. In this case, ignoring the fact that we can place at most one product at each location, even if we offer all products at the least desirable location, a customer leaves without making a purchase with a probability of p_0.

Parametric LP and Efficient Frontier: Varying $K \in \{15, 30, 45, 60\}$ and $p_0 \in \{0.1, 0.3, 0.5\}$, we obtain 12 parameter configurations. For each parameter configuration, we generate 50 individual test problems by using the approach described in the previous paragraph. For each test problem, we construct a collection of candidate assortments that is guaranteed to include an optimal solution to the Revenue-Utility problem. Constructing the candidate assortments requires obtaining an optimal solution to the Candidate LP for each value of $\gamma \in \mathbb{R}$ by using a parametric LP. Using these candidate assortments, we also construct an efficient frontier that shows all attainable optimal expected revenue-utility pairs in the Revenue-Utility problem as we vary $\lambda \in \mathbb{R}_+$. From

Param. (K, p_0)	# Cand. Assr.	# Opt. Assr.	CPU Secs.
$(15, 0.1)$	399.7	115.0	42.4
$(15, 0.3)$	412.1	115.1	44.1
$(15, 0.5)$	410.3	116.4	45.4
$(30, 0.1)$	970.5	312.3	229.9
$(30, 0.3)$	993.7	319.4	235.1
$(30, 0.5)$	979.9	317.5	224.7

Param. (K, p_0)	# Cand. Assr.	# Opt. Assr.	CPU Secs.
$(45, 0.1)$	1,426.7	393.2	593.7
$(45, 0.3)$	1,369.8	507.1	564.9
$(45, 0.5)$	1,407.4	462.0	588.6
$(60, 0.1)$	1,566.8	468.7	1,098.9
$(60, 0.3)$	1,507.9	643.4	1,089.6
$(60, 0.5)$	1,537.1	577.3	1,090.7

Table 1 The number of candidate assortments, number of optimal assortments, and CPU seconds for assortment problems with display location effects.

the discussion presented immediately before Theorem 4.2, the Collection of Candidate Assortments is given by $\{\boldsymbol{x}^\ell_{\mathsf{Cand}} : \ell = 1, \ldots, L\}$, and the optimal objective value for the Revenue-Utility problem is given by $Z^*_\lambda = \max_{\ell=1,\ldots,L} \mathsf{Rev}(\boldsymbol{x}^\ell_{\mathsf{Cand}}; \boldsymbol{r}) + \lambda\, \mathsf{Util}(\boldsymbol{x}^\ell_{\mathsf{Cand}})$. For each candidate assortment $\boldsymbol{x}^\ell_{\mathsf{Cand}}$, the function $\lambda \mapsto \mathsf{Rev}(\boldsymbol{x}^\ell_{\mathsf{Cand}}; \boldsymbol{r}) + \lambda\, \mathsf{Util}(\boldsymbol{x}^\ell_{\mathsf{Cand}})$ is linear and increasing in λ. Therefore, the function $\lambda \mapsto Z^*_\lambda$ is the pointwise maximum of L linear and increasing functions, so it is continuous, piecewise linear, increasing, and convex in λ. Between successive breakpoints of this function, the optimal solution to the Revenue-Utility problem does not change. Thus, constructing the efficient frontier requires finding the pointwise maximum of L linear functions, which can be done in $O(L \log L)$ operations (Kleinberg and Tardos 2005).

We present our computational results in Table 1. The first column shows the parameter configuration (K, p_0). Recall that we generate 50 test problems in each parameter configuration. The second column shows the average number of candidate assortments that we generate, where the average is computed over the 50 test problems in each parameter configuration. The third column shows the average number of candidate assortments that actually turn out to be an optimal solution to the Revenue-Utility problem for some $\lambda \in \mathbb{R}_+$, corresponding to the average number of assortments that appear on the efficient frontier. The collection of candidate assortments is guaranteed to include an optimal solution to the Revenue-Utility problem. However, some candidate assortments may never be optimal, regardless of what value of λ we consider, and these assortments will not appear on the efficient frontier. The fourth column shows the average CPU seconds needed to generate all the candidate assortments for a test problem. We carried out our computational experiments in Java 1.8.0 on a 2.0 GHz AMD EPYC 7501 32-Core Processor with 4 GB of RAM.

For the smaller test problems with $K = 15$ locations, we can generate all candidate assortments within a minute. For the larger test problems with $K = 60$ locations, this computational effort increases to 18 minutes. After generating the candidate assortments, in each of our test problems, it took less than 0.1 seconds to construct the efficient frontier. The second column in Table 1 shows that the number of candidate assortments L increases approximately linearly with the number of products. In particular, in the assortment problem with display location effects, the

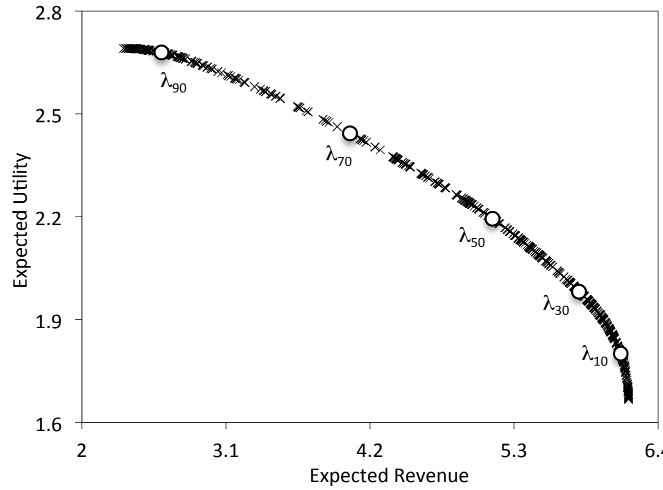

Figure 4 The efficient frontier for a particular instance of the assortment problem with display location effects with $K = 60$ and $p_0 = 0.3$.

number of products corresponds to the number of item-location pairs, which is nK. To normalize the number of candidate assortments, we divide L by nK; we observe that the average value of $L/(nK)$ is roughly constant. Specifically, for the test problems with 15, 30, 45, and 60 locations, the average values of $L/(nK)$ are, respectively, 0.45, 0.55, 0.52, and 0.43. This result is consistent with Theorem 4.2, which shows that the number of candidate assortments is bounded above by an expression that is linear in the number of products. Figure 4 illustrates the efficient frontier for a problem instance with $K = 60$ and $p_0 = 0.3$. Each data point corresponds to a different value of λ that yields a different optimal solution for the Revenue-Utility problem. Between successive data points, the optimal solution to the Revenue-Utility problem does not change.

Discretization Method: Thus far, our discussion has focused on solving the Revenue-Utility problem exactly. We now discuss the performance of our discretization method when we solve the Revenue-Utility problem approximately for some representative values of λ. By the discussion presented earlier in this section, the optimal objective value Z^*_λ of the Revenue-Utility problem is a continuous, piecewise linear, increasing, and convex function of λ. We use $\{\lambda_f : f = 1, \ldots, F\}$ to denote the breakpoints of this function, where F is the number of assortments that appear on the efficient frontier. To choose representative values of λ, we focus on the 10^{th}, 30^{th}, 50^{th}, and 70^{th} percentiles of the data $\{\lambda_f : f = 1, \ldots, F\}$. Letting $\lambda_{10}, \lambda_{30}, \lambda_{50}$, and λ_{70} denote these percentiles, for each of the 50 test problems in a parameter configuration, we use our discretization method to obtain an approximate solution to the Revenue-Utility problem with $\lambda \in \{\lambda_{10}, \lambda_{30}, \lambda_{50}, \lambda_{70}\}$. We use representative values for λ because the Revenue-Utility problem may become simple to solve for extreme values of λ. In particular, if λ is too large, then we put excessive weight on the expected utility, in which case it is near-optimal to maximize the total preference weight of the offered products. Noting that $\mathbf{A}$ is totally unimodular, we can find an assortment that maximizes the total preference weight of the offered products by maximizing the objective function $\sum_{i\in\mathcal{N}} v_i\, x_i$ subject to the constraints that $\mathbf{A}\boldsymbol{x} \leq \boldsymbol{b}$ and $\boldsymbol{x} \in [0,1]^n$. If λ is too small, then we put excessive weight on the expected revenue, in which case it is near-optimal to maximize the expected revenue. By Theorem 5.2, we can maximize the expected revenue by solving a single LP.

We present our computational results in Table 2. The first column shows the parameter configuration (K, p_0). The second column shows the value of $\lambda \in \{\lambda_{10}, \lambda_{30}, \lambda_{50}, \lambda_{70}\}$ that we focus on. The remainder of the table consists of two blocks, each containing three columns. The two blocks, respectively, show the performance of the discretization method when we use our discretization method with the grid sizes of $\rho = 1$ and $\rho = 0.1$. In each block, the first column shows the maximum percent optimality gap of the solutions obtained by the discretization method, where the maximum is computed over the 50 test problems in a parameter configuration. The second column shows the average CPU seconds for the discretization method. The third column shows the average number of candidate assortments generated by the discretization method. Our results indicate that the discretization method can find near-optimal solutions for the Revenue-Utility problem rather quickly. For the larger test problems with $K = 60$ locations, using the smallest grid size of $\rho = 0.1$, the average number of CPU seconds is 7.8 and the discretization method finds solutions with optimality gaps no larger than 0.008% by using an average of 205.2 candidate assortments. Furthermore, the practical performance of the discretization method is substantially better than its theoretical guarantee. In particular, even when we use our discretization method with a grid size of $\rho = 1$, which corresponds to a performance guarantee of 1/2, the optimality gap of the solutions that we found is at most 0.54%.

7.2 Computational Results for Pricing with Discrete Price Menus

In this set of computational experiments, we focus on pricing problems with discrete price menus, as discussed in Section 6.3. We use the following approach to generate our test problems. The number of items is fixed at $n = 100$. There are K possible price levels indexed by $\mathcal{K} = \{1, \ldots, K\}$, where K is a parameter that we vary. The possible prices for an item are $\{r_k : k \in \mathcal{K}\}$, evenly spread over the interval $[1, K]$, where $r_1 = K$ is the largest possible price and $r_K = 1$ is the smallest possible price. To come up with the preference weights associated with item i, we sample α_i from the uniform distribution over $[0,1]$ and β_i from the uniform distribution over $[0, 0.1]$ and set $v_{ik} = e^{\alpha_i - (\beta_i \times r_k)}$ for

Param.		$\rho=1$			$\rho=0.1$		
(K, p_0)	λ	Max. Perc. (%)	CPU Secs.	# Cand. Assr.	Max. Perc. (%)	CPU Secs.	# Cand. Assr.
$(15, 0.1)$	λ_{10}	0.044	0.4	23.4	0.001	1.8	157.4
	λ_{30}	0.289	0.4	23.4	0.003	1.9	157.4
	λ_{50}	0.187	0.4	23.4	0.007	1.9	157.4
	λ_{70}	0.316	0.4	23.4	0.006	1.9	157.4
$(15, 0.3)$	λ_{10}	0.050	0.5	24.6	0.000	2.3	165.4
	λ_{30}	0.157	0.5	24.6	0.002	2.1	165.4
	λ_{50}	0.185	0.4	24.6	0.004	2.0	165.4
	λ_{70}	0.307	0.4	24.6	0.003	2.0	165.4
$(15, 0.5)$	λ_{10}	0.032	0.5	24.8	0.000	2.0	168.4
	λ_{30}	0.168	0.4	24.8	0.003	1.9	168.4
	λ_{50}	0.540	0.4	24.8	0.009	2.1	168.4
	λ_{70}	0.193	0.4	24.8	0.010	2.2	168.4
$(30, 0.1)$	λ_{10}	0.028	1.1	29.0	0.001	4.5	197.4
	λ_{30}	0.175	1.3	29.0	0.004	4.5	197.4
	λ_{50}	0.196	1.1	29.0	0.005	4.2	197.4
	λ_{70}	0.284	1.1	29.0	0.006	4.4	197.4
$(30, 0.3)$	λ_{10}	0.031	1.1	28.4	0.000	3.7	192.0
	λ_{30}	0.227	1.1	28.4	0.003	3.9	192.0
	λ_{50}	0.248	1.0	28.4	0.005	3.8	192.0
	λ_{70}	0.287	1.1	28.4	0.005	4.0	192.0
$(30, 0.5)$	λ_{10}	0.022	0.9	24.8	0.001	3.2	167.8
	λ_{30}	0.165	0.9	24.8	0.004	3.2	167.8
	λ_{50}	0.290	0.9	24.8	0.008	3.2	167.8
	λ_{70}	0.291	0.9	24.8	0.005	3.4	167.8
$(45, 0.1)$	λ_{10}	0.037	1.8	27.8	0.002	5.6	193.0
	λ_{30}	0.199	1.8	27.8	0.004	5.7	193.0
	λ_{50}	0.208	1.8	27.8	0.005	5.6	193.0
	λ_{70}	0.206	1.9	27.8	0.004	6.0	193.0
$(45, 0.3)$	λ_{10}	0.041	2.0	28.4	0.001	6.9	197.4
	λ_{30}	0.193	1.8	28.4	0.004	6.1	197.4
	λ_{50}	0.211	2.1	28.4	0.008	6.9	197.4
	λ_{70}	0.215	2.0	28.4	0.008	6.9	197.4
$(45, 0.5)$	λ_{10}	0.029	1.8	29.8	0.001	6.3	204.2
	λ_{30}	0.163	1.9	29.8	0.002	6.7	204.2
	λ_{50}	0.361	1.9	29.8	0.008	6.6	204.2
	λ_{70}	0.259	2.0	29.8	0.005	6.9	204.2
$(60, 0.1)$	λ_{10}	0.051	2.9	31.6	0.001	7.8	218.4
	λ_{30}	0.193	2.9	31.6	0.004	8.2	218.4
	λ_{50}	0.252	2.7	31.6	0.005	7.8	218.4
	λ_{70}	0.137	2.8	31.6	0.008	8.1	218.4
$(60, 0.3)$	λ_{10}	0.044	2.7	28.4	0.002	7.1	194.0
	λ_{30}	0.313	2.8	28.4	0.005	7.6	194.0
	λ_{50}	0.247	2.7	28.4	0.004	7.3	194.0
	λ_{70}	0.273	2.6	28.4	0.005	7.1	194.0
$(60, 0.5)$	λ_{10}	0.035	2.8	30.0	0.001	8.7	205.2
	λ_{30}	0.204	2.6	30.0	0.004	7.9	205.2
	λ_{50}	0.306	3.0	30.0	0.006	8.4	205.2
	λ_{70}	0.279	2.8	30.0	0.007	8.7	205.2

Table 2 Performance of the discretization method for assortment problems with display location effects.

all $k \in \mathcal{K}$. For the preference weight of the no-purchase option, we set $v_0 = p_0 \sum_{i \in \mathcal{N}} v_{iK} / (1 - p_0)$, where p_0 is another parameter that we vary. In this case, if we offer all items at their lowest possible prices, then a customer leaves without making a purchase with probability p_0. Varying $K \in \{20, 40, 60, 80\}$ and $p_0 \in \{0.1, 0.3, 0.5\}$, we obtain 12 different parameter configurations. In each parameter configuration, we generate 50 test problems by using the approach described in this paragraph. As in our computational experiments for assortment problems with display location effects, for each test problem, we construct a collection of candidate assortments that is guaranteed to include an optimal solution to the Revenue-Utility problem. Using these candidate assortments, we also construct an efficient frontier showing all attainable expected revenue-utility pairs. In addition, we test the performance of our discretization method to check its ability to obtain near-optimal solutions for the Revenue-Utility problem.

Parametric LP and Efficient Frontier: We present our computational results in Table 3. The layout of this table is identical to that of Table 1. For the smaller test problems with $K = 15$ possible price levels, we can generate the efficient frontier in two minutes, whereas the corresponding computational effort for the larger test problems with $K = 80$ possible price levels is about 30 minutes. Given that there are more than 6,000 candidate assortments for the largest test problems, such computational effort corresponds to about 0.3 seconds to generate each candidate assortment. Similar to our results for assortment problems with display location effects, once we generate the candidate assortments, we can quickly construct the efficient frontier. More specifically, for each of our test problems, we can construct the efficient frontier in 0.8 seconds. Lastly, as in the previous section, the number candidate assortments L increases approximately linearly with the number of products, which is nK in this setting, corresponding to the number of item-price combinations, because we have one product for each item-price combination in our formulation of the pricing problem with a discrete price menu. For the test problems with 20, 40, 60, and 80 price levels, the average values of $L/(nK)$ are, respectively, 0.66, 0.77, 0.79, and 0.79. These results are consistent with Theorem 4.2.

Param. (K, p_0)	# Cand. Assr.	# Opt. Assr.	CPU Secs.	Param. (K, p_0)	# Cand. Assr.	# Opt. Assr.	CPU Secs.
$(20, 0.1)$	870.5	757.8	68.3	$(60, 0.1)$	4421.3	2675.1	872.5
$(20, 0.3)$	1516.6	1299.0	112.8	$(60, 0.3)$	4852.1	2844.2	978.6
$(20, 0.5)$	1590.1	1426.9	121.4	$(60, 0.5)$	4874.0	2723.6	988.0
$(40, 0.1)$	2659.3	1912.6	366.0	$(80, 0.1)$	6087.9	3101.4	1,618.8
$(40, 0.3)$	3252.7	2311.3	440.8	$(80, 0.3)$	6436.9	3178.8	1,783.4
$(40, 0.5)$	3279.2	2254.9	453.5	$(80, 0.5)$	6439.6	3073.4	1,810.8

Table 3 The number of candidate assortments, number of optimal assortments, and CPU seconds for pricing problems with discrete price menus.

Discretization Method: Table 4 shows the performance of our discretization method. The layout of this table is identical to that of Table 2. The discretization method continues to obtain solutions with remarkably small optimality gaps by using a substantially smaller number of candidate assortments. For the larger test problems with $K = 80$ price levels, using the grid size of $\rho = 0.1$, the discretization method obtains solutions with optimality gaps of at most 0.008% in about a minute by using 186.4 candidate assortments on average. Using the grid size of $\rho = 1$, the largest optimality gap is 0.328%, with computational effort under four seconds.

8. Conclusion

We examined assortment problems that consider the tradeoff between the expected revenue and the expected utility under the multinomial logit model and totally unimodular constraints. Our characterization of the optimal assortment showed that we can obtain an optimal assortment by shifting the revenues of all products by the same constant and finding an assortment that maximizes only the expected revenue. This characterization enabled us to develop a solution method based on a parametric LP. Furthermore, we proposed a discretization method that approximates the optimal objective value within any prespecified degree of accuracy. The discretization method exploits the fact that we can solve an LP to find an assortment that maximizes only the expected revenue.

Param. (K, p_0)	λ	$\rho = 1$			$\rho = 0.1$		
		Max. Perc. (%)	CPU Secs.	# Cand. Assr.	Max. Perc. (%)	CPU Secs.	# Cand. Assr.
$(20, 0.1)$	λ_{10}	0.093	0.7	17.0	0.005	10.3	111.4
	λ_{30}	0.328	0.7	17.0	0.008	10.2	111.4
	λ_{50}	0.011	0.7	17.0	0.007	10.6	111.4
	λ_{70}	0.013	0.8	17.0	0.004	10.9	111.4
$(20, 0.3)$	λ_{10}	0.038	0.6	17.0	0.001	10.3	111.6
	λ_{30}	0.261	0.6	17.0	0.007	10.2	111.6
	λ_{50}	0.167	0.6	17.0	0.010	10.3	111.6
	λ_{70}	0.018	0.6	17.0	0.003	10.4	111.6
$(20, 0.5)$	λ_{10}	0.003	0.5	17.2	0.000	9.7	110.8
	λ_{30}	0.100	0.6	17.2	0.002	9.7	110.8
	λ_{50}	0.169	0.6	17.2	0.005	9.8	110.8
	λ_{70}	0.029	0.6	17.2	0.003	9.8	110.8
$(40, 0.1)$	λ_{10}	0.042	1.9	21.0	0.001	24.9	138.6
	λ_{30}	0.202	2.0	21.0	0.005	25.5	138.6
	λ_{50}	0.262	2.0	21.0	0.007	26.0	138.6
	λ_{70}	0.013	2.1	21.0	0.006	27.1	138.6
$(40, 0.3)$	λ_{10}	0.014	1.8	21.0	0.000	25.2	137.6
	λ_{30}	0.022	1.8	21.0	0.002	25.3	137.6
	λ_{50}	0.319	1.9	21.0	0.009	25.6	137.6
	λ_{70}	0.012	1.9	21.0	0.006	26.0	137.6
$(40, 0.5)$	λ_{10}	0.007	1.8	20.2	0.000	26.5	138.6
	λ_{30}	0.022	1.9	20.2	0.002	27.2	138.6
	λ_{50}	0.231	2.0	20.2	0.004	27.3	138.6
	λ_{70}	0.030	2.1	20.2	0.003	27.8	138.6

Param. (K, p_0)	λ	$\rho = 1$			$\rho = 0.1$		
		Max. Perc. (%)	CPU Secs.	# Cand. Assr.	Max. Perc. (%)	CPU Secs.	# Cand. Assr.
$(60, 0.1)$	λ_{10}	0.011	3.8	24.0	0.000	46.7	165.0
	λ_{30}	0.064	3.9	24.0	0.002	48.1	165.0
	λ_{50}	0.276	3.9	24.0	0.005	48.1	165.0
	λ_{70}	0.016	4.0	24.0	0.008	48.5	165.0
$(60, 0.3)$	λ_{10}	0.008	3.7	24.0	0.000	48.0	164.0
	λ_{30}	0.054	3.8	24.0	0.002	48.5	164.0
	λ_{50}	0.295	3.9	24.0	0.006	49.5	164.0
	λ_{70}	0.012	4.0	24.0	0.006	49.9	164.0
$(60, 0.5)$	λ_{10}	0.005	3.8	24.0	0.000	47.9	163.2
	λ_{30}	0.008	4.0	24.0	0.002	49.0	163.2
	λ_{50}	0.220	4.1	24.0	0.005	49.9	163.2
	λ_{70}	0.023	4.2	24.0	0.002	50.4	163.2
$(80, 0.1)$	λ_{10}	0.008	3.0	27.6	0.000	56.6	186.6
	λ_{30}	0.079	2.9	27.6	0.002	56.8	186.6
	λ_{50}	0.111	2.9	27.6	0.005	57.4	186.6
	λ_{70}	0.024	2.8	27.6	0.008	58.6	186.6
$(80, 0.3)$	λ_{10}	0.002	2.8	27.8	0.000	59.4	188.0
	λ_{30}	0.090	2.9	27.8	0.002	61.0	188.0
	λ_{50}	0.277	2.8	27.8	0.008	61.1	188.0
	λ_{70}	0.013	2.7	27.8	0.005	60.9	188.0
$(80, 0.5)$	λ_{10}	0.003	2.7	27.2	0.000	60.6	186.4
	λ_{30}	0.010	2.8	27.2	0.001	60.5	186.4
	λ_{50}	0.229	2.7	27.2	0.005	60.7	186.4
	λ_{70}	0.036	2.5	27.2	0.001	60.0	186.4

Table 4 Performance of the discretization method for pricing problems with discrete price menus.

Incorporating customer-centric performance measures into assortment and pricing problems opens up numerous possibilities for future research. In this study, we used the expected utility as the customer-centric performance measure. It would be interesting to investigate other customer-centric objectives that are amenable to efficient optimization. Furthermore, there are other choice models, such as the nested and paired combinatorial logit models, that are based on the random utility maximization principle. It would be interesting to incorporate the expected utility as a performance measure into the assortment and pricing problems under these choice models. Lastly, it is useful to identify additional applications that can be captured by totally unimodular constraints.

References

Abeliuk, A., G. Berbeglia, M. Cebrian, P. Van Hentenryck. 2016. Assortment optimization under a multinomial logit model with position bias and social influence. *4OR* **14**(1) 57–75.

Aouad, A., J. Feldman, D. Segev, D. J. Zhang. 2019. Click-based MNL: Algorithmic frameworks for modeling click data in assortment optimization. Tech. rep., Washington University, St. Louis, MO.

Aouad, A., R. Levi, D. Segev. 2018. Greedy-like algorithms for dynamic assortment planning under multinomial logit preferences. *Operations Research* **66**(5) 1321–1345.

Ashlagi, I., P. Shi. 2016. Optimal allocation without money: An engineering approach. *Management Science* **62**(4) 1078–1097.

Bertsimas, D., J. N. Tsitsiklis. 1997. *Introduction to Linear Optimization*. Athena Scientific, Belmont, MA.

Bront, J. J. M., I. Mendez-Diaz, G. Vulcano. 2009. A column generation algorithm for choice-based network revenue management. *Operations Research* **57**(3) 769–784.

Carstensen, P. J. 1983. Complexity of some parametric integer and network programming problems. *Mathematical Programming* **26**(1) 64–75.

Chen, N., G. Gallego. 2019. Welfare analysis of dynamic pricing. *Management Science* **65**(1) 139–151.

Davis, J. M., G. Gallego, H. Topaloglu. 2013. Assortment planning under the multinomial logit model with totally unimodular constraint structures. Unpublished tech. rep., Cornell University, Ithaca, NY.

Davis, J. M., G. Gallego, H. Topaloglu. 2014. Assortment optimization under variants of the nested logit model. *Operations Research* **62**(2) 250–273.

Desir, A., V. Goyal, J. Zhang. 2016. Near-optimal algorithms for capacity constrained assortment optimization. Tech. rep., Columbia University, New York, NY.

Dong, L., P. Kouvelis, Z. Tian. 2009. Dynamic pricing and inventory control of substitute products. *Manufacturing & Service Operations Management* **11**(2) 317–339.

Feldman, J., D. Zhang, X. Liu, N. Zhang. 2019. Customer choice models versus machine learning: Finding optimal product displays on Alibaba. Tech. rep., Washington University, St. Louis, MO.

Feldman, J. B., H. Topaloglu. 2015. Capacity constraints across nests in assortment optimization under the nested logit model. *Operations Research* **63**(4) 812–822.

Flores, A., G. Berbeglia, P. van Hentenryck. 2019. Assortment optimization under the sequential multinomial logit model. *European Journal of Operational Research* **273**(3) 1052–1064.

Gallego, G., G. Iyengar, R. Phillips, A. Dubey. 2004. Managing flexible products on a network. CORC Technical Report TR-2004-01.

Gallego, G., R. Ratliff, S. Shebalov. 2015. A general attraction model and sales-based linear programming formulation for network revenue management under customer choice. *Operations Research* **63**(1) 212–232.

Gallego, G., R. Wang. 2014. Multi-product price optimization and competition under the nested attraction model. *Operations Research* **62**(2) 450–461.

Jagabathula, S., P. Rusmevichientong. 2017. A nonparametric joint assortment and price choice model. *Management Science* **63**(9) 3128–3145.

Kleinberg, J., E. Tardos. 2005. *Algorithm Design*. Addison Wesley, New York, NY.

Li, G., P. Rusmevichientong, H. Topaloglu. 2015. The d-level nested logit model: Assortment and price optimization problems. *Operations Research* **63**(2) 325–342.

Li, H., W. T. Huh. 2011. Pricing multiple products with the multinomial logit and nested models: Concavity and implications. *Manufacturing & Service Operations Management* **13**(4) 549–563.

Li, H., W. T. Huh. 2015. Pricing under the nested attraction model with a multi-stage choice structure. *Operations Research* **63**(4) 840–850.

Li, H., S. Webster. 2017. Optimal pricing of correlated product options under the paired combinatorial logit model. *Operations Research* **65**(5) 1215–1230.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed., *Frontiers in Economics*. Academic Press, New York, NY, 105–142.

Mendez-Diaz, I., J. J. M. Bront, G. Vulcano, P. Zabala. 2014. A branch-and-cut algorithm for the latent-class logit assortment problem. *Discrete Applied Mathematics* **164**(1) 246–263.

Nemhauser, G. L., L. A. Wolsey. 1988. *Integer and Combinatorial Optimization*. John Wiley & Sons, New York, NY.

Rusmevichientong, P., Z.-J. M. Shen, D. B. Shmoys. 2009. A PTAS for capacitated sum-of-ratios optimization. *Operations Research Letters* **37**(4) 230–238.

Rusmevichientong, P., Z.-J. M. Shen, D. B. Shmoys. 2010. Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations Research* **58**(6) 1666–1680.

Rusmevichientong, P., D. Shmoys, C. Tong, H. Topaloglu. 2014. Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management* **23**(11) 2023–2039.

Rusmevichientong, P., B. Van Roy, P. Glynn. 2006. A nonparametric approach to multiproduct pricing. *Operations Research* **54**(1) 82–98.

Sen, A., A. Atamturk, P. Kaminsky. 2018. A conic integer optimization approach to the constrained assortment problem under the mixed multinomial logit model. *Operations Research* **66**(4) 994–1003.

Song, J.-S., Z. Xue. 2007. Demand management and inventory control for substitutable products. Tech. rep., Duke University, Durham, NC.

Talluri, K., G. J. van Ryzin. 2004. Revenue management under a general discrete choice model of consumer behavior. *Management Science* **50**(1) 15–33.

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, UK.

Truong, V.-A. 2014. Optimal selection of medical formularies. *Journal of Revenue and Pricing Management* **13**(2) 113–132.

Vulcano, G., G. van Ryzin, W. Chaar. 2010. Choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management* **12**(3) 371–392.

Wang, R. 2012. Capacitated assortment and price optimization under the multinomial logit model. *Operations Research Letters* **40**(6) 492–497.

Wang, R., O. Sahin. 2018. The impact of consumer search cost on assortment planning and pricing. *Management Science* **64**(8) 3649–3666.

This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

Online Appendix

Revenue-Utility Tradeoff in Assortment Optimization under the Multinomial Logit Model with Totally Unimodular Constraints

Appendix A: Proof of Theorem 4.2

We establish an upper bound of $(m+n)^{1+m}$ on the number of candidate assortments. To prove this result, we show that as γ varies over $\mathbb{R}$, the number of optimal solutions of the Candidate LP is at most $(1+n)(n+m)^m/m!$ for $m \geq 1$. Using $\mathcal{M} = \{1,\ldots,m\}$ to index the rows of the matrix $\mathbf{A}$, let $a_{\ell i}$ be the $(\ell,i)^{\text{th}}$ entry of the matrix $\mathbf{A}$ and b_ℓ be the ℓ^{th} entry of the vector $\boldsymbol{b}$. Since there exists a nonempty assortment in $\mathcal{F}$, the Candidate LP is feasible and bounded, so the strong duality holds. Using the variables $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_m)$ and $\boldsymbol{\sigma} = (\sigma_1,\ldots,\sigma_n)$, the dual of the Candidate LP is

$$\begin{aligned}
\mathsf{LP}(\gamma) &= \min_{(\boldsymbol{\mu},\boldsymbol{\sigma}) \in \mathbb{R}_+^{m+n}} \left\{ \sum_{\ell \in \mathcal{M}} b_\ell\, \mu_\ell + \sum_{i \in \mathcal{N}} \sigma_i \;\middle|\; \sum_{\ell \in \mathcal{M}} a_{\ell i}\, \mu_\ell + \sigma_i \geq (r_i - \gamma)\, v_i \quad \forall\, i \in \mathcal{N} \right\} \\
&= \min_{(\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\zeta}) \in \mathbb{R}_+^{m+2n}} \left\{ \sum_{\ell \in \mathcal{M}} b_\ell\, \mu_\ell + \sum_{i \in \mathcal{N}} \sigma_i \;\middle|\; \sum_{\ell \in \mathcal{M}} a_{\ell i}\, \mu_\ell + \sigma_i - \zeta_i = (r_i - \gamma)\, v_i \quad \forall\, i \in \mathcal{N} \right\},
\end{aligned}$$

where the last equality follows from introducing the slack variables $\boldsymbol{\zeta} = (\zeta_1,\ldots,\zeta_n)$. To complete the proof, it suffices to show that the function $\mathsf{LP}(\cdot)$ has at most $(1+n)(m+n)^m/m! - 1$ breakpoints. Let $(\boldsymbol{\mu}^*(\gamma), \boldsymbol{\sigma}^*(\gamma), \boldsymbol{\zeta}^*(\gamma))$ denote a basic optimal solution to the dual of the Candidate LP. The constraint coefficients of the decision variables σ_i and ζ_i are linearly dependent, so both of these decision variables cannot be basic at the same time. Therefore, we can partition the set of products $\mathcal{N}$ into three disjoint sets given by

$$\begin{aligned}
\hat{\mathcal{N}}_0(\gamma) &= \{ i \in \mathcal{N} : \sigma_i^*(\gamma) \text{ is nonbasic and } \zeta_i^*(\gamma) \text{ is nonbasic} \} \\
\hat{\mathcal{N}}_1(\gamma) &= \{ i \in \mathcal{N} : \sigma_i^*(\gamma) \text{ is nonbasic and } \zeta_i^*(\gamma) \text{ is basic} \} \\
\hat{\mathcal{N}}_2(\gamma) &= \{ i \in \mathcal{N} : \sigma_i^*(\gamma) \text{ is basic and } \zeta_i^*(\gamma) \text{ is nonbasic} \}.
\end{aligned}$$

Furthermore, we let $\hat{\mathcal{M}}_0(\gamma) = \{ \ell \in \mathcal{M} : \mu_\ell^*(\gamma) \text{ is basic} \}$. We refer to a pair $(\hat{\mathcal{M}}_0(\gamma), \hat{\mathcal{N}}_0(\gamma))$ as a basis, although we also need to fix $\hat{\mathcal{N}}_1(\gamma)$ and $\hat{\mathcal{N}}_2(\gamma)$ to fully specify a basis. Nevertheless, it will shortly become clear that specifying $(\hat{\mathcal{M}}_0(\gamma), \hat{\mathcal{N}}_0(\gamma))$ is enough to compute the values of all of the decision variables $(\boldsymbol{\mu}^*(\gamma), \boldsymbol{\sigma}^*(\gamma), \boldsymbol{\zeta}^*(\gamma))$ in a basic optimal solution to the dual of the Candidate LP. In the next lemma, we show that we can limit the number of bases $\{(\hat{\mathcal{M}}_0(\gamma), \hat{\mathcal{N}}_0(\gamma)) : \gamma \in \mathbb{R}\}$ to $(m+n)^m/m!$ as γ ranges over $\mathbb{R}$.

Lemma A.1 *There exists a collection of bases* $\{(\mathcal{M}_0^k,\mathcal{N}_0^k) : k=1,\ldots,K\}$ *with* $\mathcal{M}_0^k \subseteq \mathcal{M}$, $\mathcal{N}_0^k \subseteq \mathcal{N}$ *and* $K \le (m+n)^m/m!$ *such that* $\{(\hat{\mathcal{M}}_0(\gamma),\hat{\mathcal{N}}_0(\gamma)) : \gamma\in\mathbb{R}\} \subseteq \{(\mathcal{M}_0^k,\mathcal{N}_0^k) : k=1,\ldots,K\}$.

Proof: Because there are n constraints in the dual of the Candidate LP, the number of basic variables satisfies $|\hat{\mathcal{N}}_1(\gamma)|+|\hat{\mathcal{N}}_2(\gamma)|+|\hat{\mathcal{M}}_0(\gamma)|=n$. Since $\mathcal{N}=\hat{\mathcal{N}}_0(\gamma)\cup\hat{\mathcal{N}}_1(\gamma)\cup\hat{\mathcal{N}}_2(\gamma)$, we also have $n=|\hat{\mathcal{N}}_0(\gamma)|+|\hat{\mathcal{N}}_1(\gamma)|+|\hat{\mathcal{N}}_2(\gamma)|$. The last two equalities imply that $|\hat{\mathcal{M}}_0(\gamma)|=|\hat{\mathcal{N}}_0(\gamma)|$. Furthermore, there are n products in the dual of the Candidate LP, so $|\hat{\mathcal{N}}_0(\gamma)|\le n$. Noting that $|\hat{\mathcal{M}}_0(\gamma)|\le m$, we get $|\hat{\mathcal{M}}_0(\gamma)|=|\hat{\mathcal{N}}_0(\gamma)|\le\min\{n,m\}$. Thus, the number of bases $\{(\hat{\mathcal{M}}_0(\gamma),\hat{\mathcal{N}}_0(\gamma)) : \gamma\in\mathbb{R}\}$ as γ ranges over $\mathbb{R}$ is bounded by the number of pairs $(\mathcal{M}_0,\mathcal{N}_0)$ with $\mathcal{M}_0\subseteq\mathcal{M}$, $\mathcal{N}_0\subseteq\mathcal{N}$ and $|\mathcal{M}_0|=|\mathcal{N}_0|\le\min\{m,n\}$. Using the fact that $\binom{x}{y}\le\frac{x^y}{y!}$ for any $1\le y\le x$, the number of possible such pairs is upper bounded by

$$\begin{aligned}\sum_{k=0}^{\min\{m,n\}}\binom{m}{k}\binom{n}{k} &= \sum_{k=0}^{\min\{m,n\}}\binom{m}{m-k}\binom{n}{k} \le \sum_{k=0}^{\min\{m,n\}}\frac{m^{m-k}}{(m-k)!}\frac{n^k}{k!}\\ &\le \sum_{k=0}^{m}\frac{m^{m-k}}{(m-k)!}\frac{n^k}{k!} = \frac{1}{m!}\sum_{k=0}^{m}\binom{m}{k}m^{m-k}n^k = \frac{(m+n)^m}{m!}.\end{aligned}$$

■

In the next lemma, we show that specifying $(\hat{\mathcal{M}}_0(\gamma),\hat{\mathcal{N}}_0(\gamma))$ is enough to compute the values of the decision variables $(\boldsymbol{\mu}^*(\gamma),\boldsymbol{\sigma}^*(\gamma),\boldsymbol{\zeta}^*(\gamma))$ in a basic optimal solution to the dual of the Candidate LP. In this lemma, we use $\mathcal{I}^k$ to denote the set of values of $\gamma\in\mathbb{R}$ such that the optimal basis $(\hat{\mathcal{M}}_0(\gamma),\hat{\mathcal{N}}_0(\gamma))$ in the dual of the Candidate LP is $(\mathcal{M}_0^k,\mathcal{N}_0^k)$; that is, we have

$$\mathcal{I}^k=\Big\{\gamma\in\mathbb{R} \;:\; (\hat{\mathcal{M}}_0(\gamma),\hat{\mathcal{N}}_0(\gamma))=(\mathcal{M}_0^k,\mathcal{N}_0^k)\Big\}.$$

Lemma A.2 *There exist linear functions* $\{f_\ell^k(\cdot) : \ell\in\mathcal{M}\}$ *and* $\{g_i^k(\cdot) : i\in\mathcal{N}\}$ *such that, for each* $\gamma\in\mathcal{I}^k$, *we have* $\mu_\ell^*(\gamma)=f_\ell^k(\gamma)$ *for all* $\ell\in\mathcal{M}$ *and* $\sigma_i^*(\gamma)=[g_i^k(\gamma)]^+$ *for all* $i\in\mathcal{N}$.

Proof: Fix $\gamma\in\mathcal{I}^k$ so that $\hat{\mathcal{M}}_0(\gamma)=\mathcal{M}_0^k$ and $\hat{\mathcal{N}}_0(\gamma)=\mathcal{N}_0^k$. By the same argument at the beginning of the proof of Lemma A.1, we have $|\mathcal{M}_0^k|=|\mathcal{N}_0^k|$. By the definition of $\hat{\mathcal{N}}_0(\gamma)$, we have $\sigma_i^*(\gamma)=\zeta_i^*(\gamma)=0$ for all $i\in\mathcal{N}_0^k$. Moreover, by the definition of $\hat{\mathcal{M}}_0(\gamma)$, we have $\mu_\ell^*(\gamma)=0$ for all $\ell\in\mathcal{M}\setminus\mathcal{M}_0^k$. Thus, the constraints of the dual of the Candidate LP implies that

$$\sum_{\ell\in\mathcal{M}_0^k}a_{\ell i}\,\mu_\ell^*(\gamma)=(r_i-\gamma)\,v_i\quad\forall\, i\in\mathcal{N}_0^k.$$

Because $|\mathcal{M}_0^k|=|\mathcal{N}_0^k|$, the above system of equations has $|\mathcal{N}_0^k|$ unknowns and $|\mathcal{N}_0^k|$ equations. Moreover, since $\{\mu_\ell^*(\gamma) : \ell\in\mathcal{M}_0^k\}$ are basic variables, their constraint coefficients must be linearly independent. Therefore, the values of $\{\mu_\ell^*(\gamma) : \ell\in\mathcal{M}_0^k\}$ are given by the inverse of the matrix with entries $\{a_{\ell i} : \ell \in \mathcal{M}_0^k,\ i \in \mathcal{N}_0^k\}$ multiplied by the vector with entries $\{(r_i - \gamma)\, v_i : i \in \mathcal{N}_0^k\}$. Thus, for all $\ell \in \mathcal{M}_0^k$, $\mu_\ell^*(\gamma)$ is a linear function of γ and this function is completely determined by $(\mathcal{M}_0^k, \mathcal{N}_0^k)$. On the other hand, $\mu_\ell^*(\gamma) = 0$ for all $\ell \in \mathcal{M} \setminus \mathcal{M}_0^k$. Therefore, for all $\ell \in \mathcal{M}$, $\mu_\ell^*(\gamma)$ is a linear function of γ and this function is completely determined by $(\mathcal{M}_0^k, \mathcal{N}_0^k)$. Thus, for all $\ell \in \mathcal{M}$, we have $\mu_\ell^*(\gamma) = f_\ell^k(\gamma)$ for some $f_\ell^k(\cdot)$, where $f_\ell^k(\gamma)$ is linear in γ. Lastly, noting the constraints of the dual of the Candidate LP, in an optimal solution to this problem, we have

$$\sigma_i^*(\gamma) \;=\; \Big[(r_i - \gamma)\, v_i - \sum_{\ell \in \mathcal{M}} a_{\ell i}\, \mu_\ell^*(\gamma)\Big]^+ \;=\; \Big[(r_i - \gamma)\, v_i - \sum_{\ell \in \mathcal{M}} a_{\ell i}\, f_\ell^k(\gamma)\Big]^+.$$

Because $(r_i - \gamma)\, v_i - \sum_{\ell \in \mathcal{M}} a_{\ell i}\, f_\ell^k(\gamma)$ is linear in γ, it follows that $\sigma_i^*(\gamma) = [g_i^k(\gamma)]^+$ for some $g_i^k(\cdot)$, where $g_i^k(\gamma)$ is linear in γ. ■

It is a standard result in parametric LP that $\mathcal{I}^k$ is a finite union of closed intervals except that the first and last of these intervals can be of the form $(-\infty, a]$ and $[a, \infty)$ for some $a \in \mathbb{R}$ (Chapter 5.5, Bertsimas and Tsitsiklis 1997). Let $\overline{\gamma}^k = \max\{\gamma : \gamma \in \mathcal{I}^k\}$ and $\underline{\gamma}^k = \min\{\gamma : \gamma \in \mathcal{I}^k\}$. Using the functions $\{f_\ell^k(\cdot) : \ell \in \mathcal{M}\}$ and $\{g_i^k(\cdot) : i \in \mathcal{N}\}$ in Lemma A.2, we define

$$\mathsf{LP}^k(\gamma) \;=\; \begin{cases} \sum_{\ell \in \mathcal{M}} b_\ell\, f_\ell^k(\gamma) + \sum_{i \in \mathcal{N}} [g_i^k(\gamma)]^+ & \text{if } \gamma \in [\underline{\gamma}^k, \overline{\gamma}^k] \\ +\infty & \text{otherwise.} \end{cases}$$

The expression in the first case above corresponds to the objective function of the dual of the Candidate LP evaluated at the solution $(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\zeta})$ with $\mu_\ell = f_\ell^k(\gamma)$ for all $\ell \in \mathcal{M}$ and $\sigma_i = [g_i^k(\gamma)]^+$ for all $i \in \mathcal{N}$. By Lemma A.2, if $\gamma \in \mathcal{I}^k$, then this solution is optimal to the dual of the Candidate LP, but if $\gamma \in [\underline{\gamma}^k, \overline{\gamma}^k] \setminus \mathcal{I}^k$, then this solution is not necessarily optimal.

In the next lemma, we use the functions $\{\mathsf{LP}^k(\cdot) : k = 1, \ldots, K\}$ to construct the function $\mathsf{LP}(\cdot)$, which corresponds to the optimal objective value of the dual of the Candidate LP.

Lemma A.3 *For each $\gamma \in \mathbb{R}$, we have*

$$\mathsf{LP}(\gamma) = \min\{\mathsf{LP}^k(\gamma) : k = 1, \ldots, K\}.$$

Proof: Fix $\widehat{\gamma} \in \mathbb{R}$ and let $k = 1, \ldots, K$ be such that $\widehat{\gamma} \in \mathcal{I}^k$. By Lemma A.2, the solution $(\boldsymbol{\mu}^*, \boldsymbol{\sigma}^*, \boldsymbol{\zeta}^*)$ with $\mu_\ell^* = f_\ell^k(\widehat{\gamma})$ for all $\ell \in \mathcal{M}$ and $\sigma_i^* = [g_i^k(\widehat{\gamma})]^+$ for all $i \in \mathcal{N}$ is optimal to the dual of the Candidate LP with $\gamma = \widehat{\gamma}$, so we have $\mathsf{LP}(\widehat{\gamma}) = \mathsf{LP}^k(\widehat{\gamma})$. Thus, if we can show that $\mathsf{LP}^t(\widehat{\gamma}) \geq \mathsf{LP}(\widehat{\gamma})$ for all $t \in \{1, \ldots, K\} \setminus \{k\}$, then the desired result follows. Choose an arbitrary $t \neq k$. If $\widehat{\gamma} \notin [\underline{\gamma}^t, \overline{\gamma}^t]$, then $\mathsf{LP}^t(\widehat{\gamma}) = \infty$, so $\mathsf{LP}^t(\widehat{\gamma}) \geq \mathsf{LP}(\widehat{\gamma})$. In the rest of the proof, we assume that $\widehat{\gamma} \in [\underline{\gamma}^t, \overline{\gamma}^t]$.

By the definition of $\underline{\gamma}^t$, we have $\underline{\gamma}^t \in \mathcal{I}^t$. Thus, by Lemma A.2, the solution $(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\zeta})$ with $\mu_\ell = f_\ell^t(\underline{\gamma}^t)$ for all $\ell \in \mathcal{M}$ and $\sigma_i = [g_i^t(\underline{\gamma}^t)]^+$ for all $i \in \mathcal{N}$ is optimal to the dual of the Candidate LP when we solve this problem with $\gamma = \underline{\gamma}^t$. Noting that the decision variable μ_ℓ is nonnegative in the dual of the Candidate LP, it must be the case that $f_\ell^t(\underline{\gamma}^t) \geq 0$ for all $\ell \in \mathcal{M}$. Using the same argument, we also have $f_\ell^t(\overline{\gamma}^t) \geq 0$ for all $\ell \in \mathcal{M}$. In this case, since $f_\ell^t(\underline{\gamma}^t) \geq 0$, $f_\ell^t(\overline{\gamma}^t) \geq 0$ and $\widehat{\gamma} \in [\underline{\gamma}^t, \overline{\gamma}^t]$, using the fact that $f_\ell^t(\gamma)$ is linear in γ, we get $f_\ell^t(\widehat{\gamma}) \geq 0$ for all $\ell \in \mathcal{M}$. Thus, the solution $(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\zeta})$ with $\mu_\ell = f_\ell^t(\widehat{\gamma})$ for all $\ell \in \mathcal{M}$ and

$$\sigma_i \;=\; \Big[(r_i - \widehat{\gamma})\, v_i - \sum_{\ell \in \mathcal{M}} a_{\ell i}\, f_\ell^t(\widehat{\gamma})\Big]^+$$

for all $i \in \mathcal{N}$ satisfies the constraints of the dual of the Candidate LP, along with the nonnegativity constraints, when we solve this problem with $\gamma = \widehat{\gamma}$. By the definition of $g_i^t(\cdot)$ in the proof of Lemma A.2, the right side of the above expression is $[g_i^t(\widehat{\gamma})]^+$. Thus, the solution $(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\zeta})$ with $\mu_\ell = f_\ell^t(\widehat{\gamma})$ for all $\ell \in \mathcal{M}$ and $\sigma_i = [g_i^t(\widehat{\gamma})]^+$ for all $i \in \mathcal{N}$ is feasible to the dual of the Candidate LP when we solve this problem with $\gamma = \widehat{\gamma}$. Furthermore, by the definition of $\mathsf{LP}^t(\cdot)$, this solution provides the objective value of $\mathsf{LP}^t(\widehat{\gamma})$ for the dual of the Candidate LP, so $\mathsf{LP}^t(\widehat{\gamma}) \geq \mathsf{LP}(\widehat{\gamma})$. ■

We need one more lemma to show the final result.

Lemma A.4 *Let $\alpha(\gamma)$ and $\beta(\gamma)$ be piecewise linear and convex in γ and $\theta(\gamma) = \min\{\alpha(\gamma), \beta(\gamma)\}$ be convex in γ. Then, the number of breakpoints of $\theta(\cdot)$ is at most the sum of the number of breakpoints of $\alpha(\cdot)$ and $\beta(\cdot)$.*

Proof: If there exists a breakpoint of $\theta(\cdot)$ that is not a breakpoint of $\alpha(\cdot)$ or $\beta(\cdot)$, then $\alpha(\cdot)$ and $\beta(\cdot)$ must intersect at this point. Since $\theta(\gamma) = \min\{\alpha(\gamma), \beta(\gamma)\}$ and the minimum of two intersecting lines is concave, $\theta(\cdot)$ cannot be convex in a small neighborhood of this intersection point. ■

Here is the proof of Theorem 4.2. Noting the term $\sum_{i \in \mathcal{N}} [g_i^k(\gamma)]^+$ in the definition of $\mathsf{LP}^k(\gamma)$, since $g_i^k(\gamma)$ is linear in γ, $\mathsf{LP}^k(\cdot)$ is piecewise linear and convex and it has $n+2$ breakpoints, including $\underline{\gamma}^k$ and $\overline{\gamma}^k$. Furthermore, $\mathsf{LP}(\gamma)$ is convex in γ by linear programming duality. So, by Lemmas A.3 and A.4, the number of breakpoints of $\mathsf{LP}(\cdot)$ is at most $K\,(n+2)$. If $\overline{\gamma}^k \neq \infty$, then it is a common breakpoint for at least two of the functions $\{\mathsf{LP}^k(\cdot) : k = 1, \ldots, K\}$ and we drop the double-counted breakpoints. Dropping also ∞ and $-\infty$, the number of remaining breakpoints of $\mathsf{LP}(\cdot)$ is at most $K\,(n+2) - (K-1) - 2 \leq (n+1)\,(m+n)^m/m! - 1$, where the inequality uses Lemma A.1. ■