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

September 6, 2021

We examine the revenue-utility assortment optimization 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 show that we can solve the revenue-utility assortment optimization problem by finding the assortment that maximizes only the expected revenue, after adjusting the revenue of each product by the same constant. Finding the appropriate revenue adjustment requires solving a nonconvex optimization problem. We give a parametric linear program to generate a collection of candidate assortments that is guaranteed to include an optimal solution to the revenue-utility assortment optimization 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. Moreover, we develop an approximation scheme that limits the number of candidate assortments while ensuring a prespecified solution quality. Lastly, we discuss practical assortment optimization problems that involve totally unimodular constraints. In our computational experiments, we demonstrate that we can obtain significant improvements in the expected utility without incurring a significant loss in the expected revenue.

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. Using discrete choice models, we can develop demand models 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 using discrete choice models to capture the customer choice process can yield better operational decisions (Talluri and van Ryzin 2004, Vulcano et al. 2010, Feldman et al. 2019). Most of this literature focuses on the firm-centric objective of maximizing the expected revenue, leaving customer-centric objectives relatively unexplored. In many application settings, it is important to consider customer-centric objectives. Focusing on customer-centric objectives improves customer satisfaction and loyalty, which is important even for a revenue-maximizing firm, especially when the firm interacts with the customers repeatedly and wants to adopt a far-sighted view. When choosing product assortments in retail (Anupindi et al. 2015), ranking online search results (L'Ecuyer et al. 2017) and making product recommendations (Derakhshan et al. 2019), a profit-maximizing firm may still consider customer-centric objectives to ensure that the customers have access to broader product variety, can find more relevant search results, and derive greater utility from their purchases. Furthermore, in application settings that involve a social component, it is vital to consider the well-being of the customers. When offering schooling options (Shi 2019), setting up the medication menu for an insurer (Truong 2014) and locating healthcare providers (Puig-Junoy et al. 1998), all of which involve a social component, government agencies may need to focus on customer-centric objectives to provide easier access, improved welfare and shorter wait times.

We study an assortment 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. 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 optimization problem. We show that we can solve the revenue-utility assortment optimization problem by finding an assortment that maximizes only the expected revenue, after adjusting the revenue of each product by the same amount. Using this result, 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 problem. We show how to construct an efficient frontier that includes all optimal revenue-utility pairs as the weights on the two objectives vary. We develop an approximation scheme that limits the number of candidate assortments, while maintaining a prespecified solution quality.

Main Contributions: To give an overview of our problem setup, let $\mathcal{N} = \{1, 2, \dots, n\}$ denote the set of available products within which we choose an assortment to offer to the customers. We represent an assortment by a vector $\boldsymbol{x} = (x_1, \dots, x_n) \in \{0,1\}^n$, where we have $x_i = 1$ if and only if we offer product i. The set of feasible assortments is given by $\mathcal{F} = \{\boldsymbol{x} \in \{0,1\}^n : \mathbf{A}\boldsymbol{x} \le \boldsymbol{b}\}$, where

A is a totally unimodular matrix. Customers choose among the offered 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 normalized 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 optimization problem as the $(1, \lambda)$-weighted revenue-utility assortment optimization 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. 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. Under the multinomial logit model, we have a closed-form expression for the expected utility of the alternative that the customer chooses (McFadden 1974). Through totally unimodular constraints, we can impose bounds on the number of offered products, incorporate display location effects that can boost the attractiveness of a product depending on where it is displayed, formulate assortment optimization problems to choose the offered products as well as their prices, and ensure that the price ordering of the products is consistent with their qualities.

Structure of an optimal assortment. We construct a lower bound on the objective function of the revenue-utility assortment optimization problem that requires computing only the expected revenue, but *after* adjusting all product revenues by the same additive constant (Lemma 3.1). We show that this lower bound is tight at the optimal solution, so that we can solve the revenue-utility assortment optimization problem by maximizing only the expected revenue, but after adjusting the revenue associated with each product by the same additive constant (Theorem 3.2). An immediate consequence of this result is that if there is no constraint, then an optimal solution to the revenue-utility assortment optimization problem is revenue-ordered, including a certain number of products with the largest revenues (Corollary 3.3). Computing the appropriate adjustment in the product revenues requires solving a nonconvex optimization problem, but as discussed in the next paragraph, we use this connection to develop a solution method for the revenue-utility problem.

Efficient solution methods. We develop an approach for solving the revenue-utility assortment optimization 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 optimization problem (Theorem 4.1). Using r_i and v_i to denote the revenue and preference weight, respectively, of product i, letting m be the number of constraints, we show that the number of candidate assortments in the collection is at most $\min\{1+n\max_{i\in\mathcal{N}} v_i,\ 2+2n\max_{i\in\mathcal{N}} r_i v_i,\ (m+n)^{1+m}\}$ (Theorem 4.2). When the revenues and preference weights take rational values, we scale them by the same constant so that they take integer values. Such a scaling is innocuous because if we scale all revenues and preference weights, as well as the weight on the expected utility, by the same constant, then the optimal solution to the revenue-utility assortment optimization problem does not change. By the first two terms in the minimum, for fixed revenues and preference weights, the number of candidate assortments grows linearly with the number of products, whereas by the last term, 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 on the expected utility. Once we construct the candidate assortments, we can use the same collection to solve the $(1,\lambda)$-weighted revenue-utility assortment optimization problem for all values of λ. This result allows us to construct an efficient frontier that shows the optimal expected revenue-utility pairs as the weight λ varies.

We also give an approximation scheme that balances solution quality and computational effort. For each fixed $\rho>0$, our approximation scheme rounds the total preference weight of the offered products to integer powers of $1+\rho$. Letting $V_{\min}$ and $V_{\max}$, respectively, be the smallest and largest preference weights, it generates a collection of $O\big(\frac{1}{\rho}\log\big(nV_{\max}/V_{\min}\big)\big)$ candidate assortments and ensures that this collection includes a solution whose objective value is at least $1/(1+\rho)$ fraction of the optimal objective value. Adjusting the value of ρ, we can tradeoff the solution quality with the number of candidate assortments, the latter quantity being a measure of computational effort.

An LP for expected revenue maximization. When $\lambda=0$, the $(1,\lambda)$-weighted revenue-utility assortment optimization problem maximizes only the expected revenue. Even in this simpler setting, we make a 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 6.1). Our result is based on analyzing basic solutions to an LP and can potentially be applied to other assortment optimization problems, providing connections between LP and assortment optimization.

Applications. We discuss five practical problem classes that can be captured by 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 optimization problems with display location effects, in which the attractiveness of a product depends on the location where the product is displayed on a shelf, storefront or 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. 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 follow the same ordering. Fifth, we consider assortment optimization problems with precedence constraints such that a particular product cannot be offered unless certain related products are also offered. To demonstrate the effectiveness of our solution methods, we conduct numerical experiments on assortment optimization problems with display location effects and pricing problems. Using a dataset provided by Expedia, we also demonstrate that we can obtain significant improvements in the expected utility without incurring a significant loss in the expected revenue.

Related Literature: Our paper is related to research on assortment optimization 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 on the offered assortment. Both papers show that an optimal assortment is revenue-ordered. Rusmevichientong et al. (2009) present a polynomial-time approximation scheme when 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 optimization 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. These papers characterize the computational complexity of the problem and provide heuristics, integer programming formulations and approximation methods. Gallego et al. (2015) show that the assortment optimization 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. A constraint on the expected consumption of each resource puts a limit on a linear combination of the purchase probabilities of the different products, whereas a constraint on the offered assortment puts a limit on which specific products are offered. Neither of these constraints is a special case of the other.

This paper is an outgrowth of our earlier study, 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 the present 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, for example, Davis et al. (2014), Feldman and Topaloglu (2015), and Li et al. (2015) for work under other choice models that builds 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 approach, which constructs a lower bound on the objective function of the revenue-utility assortment optimization problem and parametrically maximizes this lower bound, differs substantially from the approaches used in the literature. Other work on assortment optimization under the multinomial logit model includes Wang (2012), Aouad et al. (2018), Sen et al. (2018), Wang and Sahin (2018), Aouad et al. (2019), and Flores et al. (2019).

All of the work on incorporating customer-centric objectives into assortment optimization problems is quite recent. Truong (2014) formulates an assortment optimization problem to find a menu of medication options offered by an insurer. Her objective is to minimize the difference between the expected cost of the purchased medication and the expected utility of the customer. Anupindi et al. (2015) give an integer programming formulation for the assortment optimization problem of a retailer under the assumption that there are multiple customer segments and the customers in each segment associate a deterministic utility with each product. L'Ecuyer et al. (2017) focus on the problem of ranking search results to maximize a linear combination of the expected revenue and the search relevance. The probability of clicking a search result is a separable function of the position of the search result and its other characteristics. Shi (2019) formulates the problem of designing school options as a large-scale optimization problem, in which he maximizes the expected utility subject to a constraint on the expected capacity utilization of the schools. Derakhshan et al. (2019) consider the problem of choosing a sequence of products to recommend in online search results with the goal of maximizing the expected utility of the customer or the market share of the platform, but the authors do not consider the two objectives simultaneously. None of these papers impose constraints on the assortment of offered alternatives.

There is substantial literature on fitting choice models to data and using these choice models to understand the expected utility implications of policy decisions. Duch-Brown et al. (2017) and Quan and Williams (2018) focus on choices for retailers and consumer products, Train (2015) focuses on choices for video streaming services, Puig-Junoy et al. (1998) focus on choices for healthcare providers, Chen and Farias (2012) focus on choices for airlines and fares, Gupta et al. (2006) focus on route choices of commuters, and Kling and Thomson (1996) focus on choices for recreational activity locations. In these papers, the authors fit a multinomial or nested logit model to data and use the fitted choice models to understand the expected utility implications of different policy decisions. In modeling healthcare provider choices, for example, the policy decision may be making certain providers more accessible. In modeling route choices, the policy decision may be changing the tolls on certain roads. The authors use the same closed-form expression that we use to capture the expected utility of the customers. In Chen and Farias (2012), the policy decisions are the fares and flight capacities, and the authors make these decisions by solving a continuous optimization problem. In all other papers, the authors try a small number of hand-picked policy decisions, so there is no optimization problem that automates the policy decision.

We do not explicitly model the repeated interactions between the firm and the customers, though one motivation for customer-centric objectives is certainly to improve long-term outcomes. There is work on repeated interactions. In Calmon et al. (2018), the future budget allocations of the customers depend on their valuations for the past services they received. The authors analyze myopic policies. Aflaki and Popescu (2014) provide a stylized model that focuses on the interactions with one customer. In Adelman and Mersereau (2013), the future demands of the customers depend on their past fill rates and the firm has limited capacity. The authors examine approximate dynamic programming methods. Ho et al. (2006) develop a model in which the purchase rate of a customer depends on the service quality in the last interaction and improving the service quality comes at a cost. In Gans (2002), each customer chooses among suppliers based on her past cumulative service quality. The author analyzes the game between multiple suppliers.

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 price, 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) (Song and Xue 2007, Dong et al. 2009, Li and Huh 2011, Gallego and Wang 2014, Li and Huh 2015, Li and Webster 2017). We allow the preference weight of a product to depend on its price in an arbitrary fashion, without any restriction. Moreover, our discrete price menus 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. Chen and Hausman (2000) examine a joint assortment optimization and pricing problem under the multinomial logit model with a finite price menu. They maximize only the expected revenue by using a continuous relaxation of a fractional programming formulation that yields binary solutions.

Organization: In Section 2, we formulate the revenue-utility assortment optimization problem. In Section 3, we show that the problem can be solved by finding an assortment that maximizes only 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 study the properties of the efficient frontier. In Section 6, we give our approximation scheme. In Section 7, we discuss applications with totally unimodular constraints. In Section 8, we give numerical experiments. In Section 9, we conclude.

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

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

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

We have two goals in mind when choosing the assortment of products 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 $\log v_0 + \theta$. Therefore, if we offer the assortment $\boldsymbol{x}$, then the expected utility of the customer from the alternative she chooses, net of the expected utility she obtains when she must choose the no-purchase option, is $\log\big(v_0 + \sum_{i \in \mathcal{N}} v_i\, x_i\big) + \theta - \log v_0 - \theta = \log\Big(1 + \sum_{i \in \mathcal{N}} \frac{v_i}{v_0}\, x_i\Big)$. 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

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

Throughout the paper, we need to find assortments that maximize the expected revenue under different product revenues, so we make explicit the dependence of the expected revenue $\mathsf{Rev}(\boldsymbol{x}; \boldsymbol{r})$

on the product revenues $\boldsymbol{r}$. We focus on the expected utility of the customer, net of the expected utility she obtains when she must choose the no-purchase option. Intuitively speaking, we use $\mathsf{Util}(\boldsymbol{x})$ to capture the value that the customer derives from the offered assortment. Therefore, the net expected utility $\mathsf{Util}(\boldsymbol{x})$ captures the utility of the customer when we make an assortment $\boldsymbol{x}$ available, as opposed to when the only available alternative is the no-purchase option. Another important reason for focusing on the net expected utility is that if we increase the utility of each alternative by a fixed constant, then the alternative that the customer chooses does not change. Therefore, if we estimate the location parameters of the utilities of the alternatives from the data, then we can estimate them only up to an additive constant (Section 3.5, Train 2003). Noting that $v_i = e^{\mu_i}$ and $v_0 = e^{\mu_0}$, if we shift the location parameters by an additive constant K, then the net expected utility of the customer is $\log(1+\sum_{i\in\mathcal{N}} \frac{e^{\mu_i+K}}{e^{\mu_0+K}}\, x_i) = \log(1+\sum_{i\in\mathcal{N}} \frac{e^{\mu_i}}{e^{\mu_0}}\, x_i)$, which is insensitive to the additive constant that we cannot estimate from the data anyway.

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 constraint matrix, and the vector $\boldsymbol{b}\in\mathbb{R}^m$ on the right side is integral. Recall that the matrix $\mathbf{A}$ is totally unimodular if the determinant of any square submatrix of $\mathbf{A}$ is $+1$, 0 or -1. Thus, the inverse of any square submatrix of a totally unimodular matrix has integer entries, whenever the inverse exists. We express our set of feasible assortments with "less than or equal to" constraints, but negating or duplicating a row of a totally unimodular matrix preserves its total unimodularity (Proposition 2.2 in Chapter III.1, Nemhauser and Wolsey 1988). Therefore, we can accommodate an "equal to" constraint by replacing it with a pair of "less than or equal to" and "greater than or equal to" constraints. 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 net expected utility. We want to solve the problem

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

Throughout the rest of the paper, for conciseness, we refer to $\mathsf{Util}(\boldsymbol{x})$ as expected utility, rather than net expected utility. We close this section with an example to demonstrate that the optimal assortment in the Revenue-Utility problem can change significantly based on the value of λ. Consider a problem instance with four products. The product revenues are $(r_1,r_2,r_3,r_4) = (6,3,2,1)$ and the preference weights are $(v_1,v_2,v_3,v_4) = (2,1,5,8)$. The preference weight of the no-purchase option is $v_0 = 1$. The set of feasible assortments is $\mathcal{F} = \{\boldsymbol{x}\in\{0,1\}^4 : \sum_{i=1}^4 x_i \le 2\}$, so we can offer at most

Inter. for λ	Opt. Assr.	Exp. Rev.	Exp. Util.
$[0.00, 0.87]$	$\{1\}$	4.00	1.10
$[0.87, 1.44]$	$\{1,2\}$	3.75	1.39
$[1.44, 2.62]$	$\{1,3\}$	2.75	2.08
$[2.62, \infty)$	$\{3,4\}$	1.29	2.64

Table 1 Optimal solution to the Revenue-Utility problem for different values of λ.

two products. In Table 1, we give the optimal assortment for all possible values of λ. The first column shows intervals for possible values of λ. The second column shows the optimal assortment to offer when λ is in each interval. The third and fourth columns, respectively, show the expected revenue and the expected utility from the optimal assortment. For $\lambda \in [0, 0.87]$, corresponding to the smallest values for λ, we mainly focus on maximizing the expected revenue. In this case, it is optimal to offer the assortment $\{1\}$, yielding an expected revenue of 4.00. Product 1 has the largest revenue. Offering any other product along with product 1 dilutes the expected revenue from the assortment. For $\lambda \in [0.87, 1.44]$, the weight on the expected utility increases and it is optimal to offer the assortment $\{1, 2\}$. The expected revenue from the assortment $\{1, 2\}$ is 3.75, which is smaller than the expected revenue from the assortment $\{1\}$, but offering product 2 increases the expected utility of the assortment $\{1\}$ from 1.10 to 1.39. For $\lambda \in [1.44, 2.62]$, we put even more weight on the expected utility, in which case, it is optimal to offer the assortment $\{1, 3\}$. To maximize the expected utility, we need to offer products with larger preference weights. When product 2 in the assortment $\{1, 2\}$ is replaced with product 3, the expected revenue of the assortment $\{1, 2\}$ decreases from 3.75 to 2.75 but its expected utility increases from 1.39 to 2.08. For $\lambda \in [2.62, \infty)$, corresponding to the largest values for λ, we mainly focus on maximizing the expected utility. In this case, it is optimal to offer the assortment $\{3, 4\}$. If our focus is on maximizing the expected utility, then it is optimal to offer an assortment that has the largest total preference weight. Products 3 and 4 are the two products with the largest preference weights, so we offer the assortment $\{3, 4\}$ when we focus on maximizing the expected utility. If we do not have any constraints on the offered assortment, then it is optimal to offer all products to maximize the expected utility.

In the next section, we characterize the optimal solution to the Revenue-Utility problem and use this characterization to construct a collection of candidate assortments that is guaranteed to contain an optimal solution to the Revenue-Utility problem.

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. Moreover, we assume that $\mathcal{F}$ includes a nonempty assortment; otherwise, the Revenue-Utility problem is trivial. Since $\mathbf{A}$ is totally unimodular, we can check this assumption by solving the LP $\max_{\boldsymbol{x} \in [0,1]^n} \{\sum_{i \in \mathcal{N}} c_i\, x_i \mid \mathbf{A}\, \boldsymbol{x} \leq \boldsymbol{b}\}$, for any $(c_1, \ldots, c_n)$ with $c_i > 0$ for all $i \in \mathcal{N}$.

Lemma 3.1 (Lower Bound) *Noting that the objective function of the* Revenue-Utility *problem is* $\mathsf{Rev}(\boldsymbol{x}; \boldsymbol{r}) + \lambda \log(1 + V(\boldsymbol{x}))$*, for all* $\boldsymbol{x} \in \mathcal{F}$ *and* $t \geq 0$*, we have*

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

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 $\boldsymbol{x} \in \mathcal{F}$, we have

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

Noting that $\log(1+a)$ 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 $a, b \in \mathbb{R}_+$, 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 $t \geq 0$, 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 by using the identity that we give at the beginning of the proof with $\alpha = \lambda\,(1+t)$ and rearranging the terms. ■

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 $\boldsymbol{x} \in \mathcal{F}$ and $t \geq 0$, we can maximize this function over all $\boldsymbol{x} \in \mathcal{F}$ and $t \geq 0$ to obtain a lower bound on the optimal objective value of the Revenue-Utility problem. In other words, we can solve the problem

$$\max_{t \geq 0} \left\{ \max_{\boldsymbol{x} \in \mathcal{F}} \Big\{ \mathsf{Rev}(\boldsymbol{x}; \boldsymbol{r} + \lambda\,(1+t)\,\boldsymbol{e}) \Big\} + \lambda\,(\log(1+t) - t) \right\}. \qquad \text{(Parametric)}$$

In the next theorem, we show that the lower bound on the optimal objective value of the Revenue-Utility problem obtained by solving the above Parametric problem is tight. Furthermore, we can use 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 (Characterization of 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. Conversely, if $\widehat{\boldsymbol{x}}$ is an optimal solution to the* Revenue-Utility *problem, then $(V(\widehat{\boldsymbol{x}}), \widehat{\boldsymbol{x}})$ is also an optimal solution to the* Parametric *problem.*

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

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

In the above chain of inequalities, (a) holds because $V(\widehat{\boldsymbol{x}}) \geq 0$, in which case, it follows that $(V(\widehat{\boldsymbol{x}}), \widehat{\boldsymbol{x}})$ is a feasible but not necessarily an optimal solution to the Parametric problem. On the other hand, (b) follows by using the identity that we give at the beginning of the proof of Lemma 3.1 with $\alpha = \lambda\,(1 + V(\widehat{\boldsymbol{x}}))$. At the beginning of the proof, we argue that $Z^*_\lambda \geq \mathsf{Rev}(\boldsymbol{x}^*; \boldsymbol{r}) + \lambda \log(1 + V(\boldsymbol{x}^*))$, so all of the above inequalities must hold as equalities. Therefore, 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. Similarly, using the fact that all of the above inequalities must hold as equalities, we also obtain $\mathsf{Rev}(\widehat{\boldsymbol{x}}\,; \boldsymbol{r} + \lambda\,(1 + V(\widehat{\boldsymbol{x}}))\,\boldsymbol{e}) + \lambda\,(\log(1 + V(\widehat{\boldsymbol{x}})) - V(\widehat{\boldsymbol{x}})) = \mathsf{Rev}(\boldsymbol{x}^*; \boldsymbol{r} + \lambda\,(1 + t^*)\,\boldsymbol{e}) + \lambda\,(\log(1 + t^*) - t^*)$, in which case, it follows that $(V(\widehat{\boldsymbol{x}}), \widehat{\boldsymbol{x}})$ is an optimal solution to the Parametric problem as well. ■

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 problem in the Parametric problem to compute t^* is difficult. In Figure 3, we plot the objective function of the outer maximization problem as a function of t for a particular problem instance. This objective function involves multiple local maxima. Instead of trying to find the global maximum for the outer maximization problem, we will generate 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

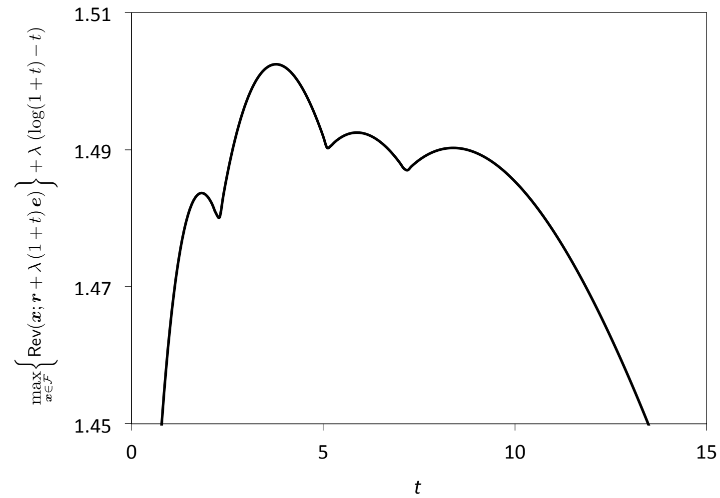

Figure 1 Objective function of the outer maximization in the Parametric problem as a function of $t \ge 0$ for a problem instance with $n = 6$, $\mathcal{F} = \{0,1\}^n$, $\lambda = 0.37$, $(r_1,\dots,r_6) = (1.89, 1.71, 1.65, 0.67, 0.45, 0.34)$, $(v_1,\dots,v_6) = (0.24, 0.54, 1.05, 1.94, 2.11, 2.51)$ and $v_0 = 1$.

$\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda(1+t^*)\boldsymbol{e})$. Thus, the collection of candidate assortments includes an optimal solution to the Revenue-Utility problem as well. By checking the objective value associated with each candidate assortment, we obtain an optimal solution to the Revenue-Utility problem. We discuss the question of generating the candidate assortments in the next section. Lastly, by the identity at the beginning of the proof of Lemma 3.1, note that $\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda(1+t^*)\boldsymbol{e}) = \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r})+\lambda(1+t^*)\frac{V(\boldsymbol{x})}{1+V(\boldsymbol{x})}$. We have $\frac{V(\boldsymbol{x})}{1+V(\boldsymbol{x})} = \frac{\sum_{i\in\mathcal{N}} v_i x_i}{v_0+\sum_{i\in\mathcal{N}} v_i x_i}$, so $\frac{V(\boldsymbol{x})}{1+V(\boldsymbol{x})}$ is the probability that a customer makes a purchase within the assortment $\boldsymbol{x}$. By Theorem 3.2, to align the Revenue-Utility problem with the goal of maximizing only the expected revenue, it turns out that we need to provide an additional reward of $\lambda(1+t^*)$ for each customer who makes a purchase.

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 $r_1 \ge r_2 \ge \ldots \ge r_n$, then there exists an optimal solution $\boldsymbol{x}^*$ to the* Revenue-Utility *problem such that $x_i^* = 1$ for all $i \le j^*$ and $x_i^* = 0$ for all $i > j^*$ for some $j^* \in \mathcal{N}$.*

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 $i \le j^*$ and $x_i^* = 0$ for all $i > j^*$ for some $j^* \in \mathcal{N}$ (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 $t\geq 0$ 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 the existing results on the optimality of revenue-ordered assortments for the unconstrained expected revenue maximization problem. In Section 5, we will show that the optimal revenue-ordered assortment includes more products as the weight λ on the expected utility increases. In the next section, we focus on generating our collection of candidate assortments.

4. Constructing Candidate Assortments

From the discussion that follows Theorem 3.2, we know that we can obtain an optimal solution to the Revenue-Utility problem by solving the problem $\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t^*)\,\boldsymbol{e})$ for some $t^*\geq 0$. Note that $\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t^*)\,\boldsymbol{e}) = \frac{\sum_{i\in\mathcal{N}}(r_i+\lambda\,(1+t^*))\,v_i\,x_i}{v_0+\sum_{i\in\mathcal{N}}v_i\,x_i} \geq \gamma$ if and only if $\sum_{i\in\mathcal{N}}(r_i-\gamma+\lambda\,(1+t^*))\,v_i\,x_i\geq v_0\,\gamma$. Motivated by this observation, we consider the LP

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

Noting that the constraint matrix $\mathbf{A}$ is totally unimodular, it is a standard result that the above Candidate LP has an optimal solution with all decision variables taking binary values (Proposition 2.2 in Chapter III.1, Nemhauser and Wolsey 1988). Therefore, letting $\boldsymbol{x}_{\mathsf{LP}}(\gamma)$ be an optimal solution to the Candidate LP as a function of γ, we have $\boldsymbol{x}_{\mathsf{LP}}(\gamma)\in\mathcal{F}$ for each $\gamma\in\mathbb{R}$. 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 a parametric LP problem, which can be solved by using the homotopy method (Chapter 7.2, Vanderbei 2014). The optimal objective value $\mathsf{LP}(\gamma)$ of the Candidate LP is continuous, piecewise linear, decreasing and convex in γ. The number of breakpoints of the function $\mathsf{LP}(\cdot)$ corresponds to the cardinality of $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma):\gamma\in\mathbb{R}\}$. In the next theorem, we show 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: The result follows by showing that an optimal solution to the Candidate LP with a certain choice of γ is also an optimal solution to the problem $\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t^*)\,\boldsymbol{e})$, and by recalling that an optimal solution to the problem $\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t^*)\,\boldsymbol{e})$ is also an optimal solution to the Revenue-Utility problem. For each $\alpha\in\mathbb{R}$, let $\gamma^\alpha=\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\alpha\,\boldsymbol{e})$. In addition, let $\boldsymbol{x}^\alpha = \boldsymbol{x}_{\mathsf{LP}}(\gamma^\alpha - \alpha)$; that is $\boldsymbol{x}^\alpha$ is an optimal solution to the Candidate LP when we solve this LP with $\gamma = \gamma^\alpha - \alpha$. We claim that $\boldsymbol{x}^\alpha$ is also an optimal solution to the problem $\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\alpha\,\boldsymbol{e})$. To prove the claim, let $\widehat{\boldsymbol{x}}$ be an optimal solution to the problem $\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\alpha\,\boldsymbol{e})$, so $\gamma^\alpha = \mathsf{Rev}(\widehat{\boldsymbol{x}};\boldsymbol{r}+\alpha\,\boldsymbol{e}) = \frac{\sum_{i\in\mathcal{N}}(r_i+\alpha)\,v_i\,\widehat{x}_i}{v_0+\sum_{i\in\mathcal{N}} v_i\,\widehat{x}_i}$. Focusing on the first and last terms in this chain of equalities and rearranging the terms, we have $v_0\,\gamma^\alpha = \sum_{i\in\mathcal{N}}(r_i-\gamma^\alpha+\alpha)\,v_i\,\widehat{x}_i$. Moreover, note that $\widehat{\boldsymbol{x}}$ is a feasible, but not necessarily an optimal, solution to the Candidate LP when we solve this LP with $\gamma=\gamma^\alpha-\alpha$, so $\sum_{i\in\mathcal{N}}(r_i-\gamma^\alpha+\alpha)\,v_i\,x_i^\alpha \geq \sum_{i\in\mathcal{N}}(r_i-\gamma^\alpha+\alpha)\,v_i\,\widehat{x}_i$. Thus, we get $\sum_{i\in\mathcal{N}}(r_i-\gamma^\alpha+\alpha)\,v_i\,x_i^\alpha \geq \sum_{i\in\mathcal{N}}(r_i-\gamma^\alpha+\alpha)\,v_i\,\widehat{x}_i = v_0\,\gamma^\alpha$. Focusing on the first and last terms in this chain of inequalities and collecting the terms that involve γ^α together, we have

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

The left side of the above chain of inequalities corresponds to $\mathsf{Rev}(\boldsymbol{x}^\alpha;\boldsymbol{r}+\alpha\,\boldsymbol{e})$, so $\boldsymbol{x}^\alpha$ must be an optimal solution to the problem $\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\alpha\,\boldsymbol{e})$, establishing the claim.

Let t^* be an optimal solution to the outer maximization problem in the Parametric problem. Using the claim in the previous paragraph with $\alpha = \lambda\,(1+t^*)$, the solution $\boldsymbol{x}_{\mathsf{LP}}(\gamma^{\lambda\,(1+t^*)} - \lambda\,(1+t^*))$ is optimal to the problem $\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t^*)\,\boldsymbol{e})$, in which case, by Theorem 3.2, the solution $\boldsymbol{x}_{\mathsf{LP}}(\gamma^{\lambda(1+t^*)} - \lambda\,(1+t^*))$ is optimal to the Revenue-Utility problem as well. Naturally, we have $\boldsymbol{x}_{\mathsf{LP}}(\gamma^{\lambda\,(1+t^*)} - \lambda\,(1+t^*)) \in \{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma\in\mathbb{R}\}$. Therefore, there exists an optimal solution to the Revenue-Utility problem in the collection of assortments $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma\in\mathbb{R}\}$. ■

We can use a parametric LP to construct the collection of assortments $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma\in\mathbb{R}\}$ such that this collection includes an optimal solution to the Candidate LP for all $\gamma\in\mathbb{R}$. By the above theorem, 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 $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma\in\mathbb{R}\} = \{\boldsymbol{x}^\ell_{\mathsf{Cand}} : \ell = 1,\ldots,L\}$, where L is the number of candidate assortments and $\boldsymbol{x}^\ell_{\mathsf{Cand}}$ is a candidate assortment corresponding to an optimal solution to the Candidate LP for some γ. If there are multiple optimal solutions to the Candidate LP for some γ, then we break ties using a deterministic tie-breaking rule. By expressing the collection of candidate assortments as $\{\boldsymbol{x}^\ell_{\mathsf{Cand}} : \ell = 1,\ldots,L\}$, we highlight the finite nature of the collection of candidate assortments.

In the next theorem, we show that the number of candidate assortments L is upper bounded by the *minimum* of three terms. By the first two terms, for fixed preference weights and revenues, L is linear in the number of products n. By the third term, for a fixed number of constraints m, L is polynomial in n. In the theorem, we assume that the preference weights, as well as the product of the revenues and the preference weights, are integers. Noting that $\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) = \frac{\sum_{i\in\mathcal{N}} r_i\,v_i\,x_i}{v_0+\sum_{i\in\mathcal{N}} v_i\,x_i}$, we can scale the revenues of all products and the parameter λ by the same multiplicative constant without changing the optimal solution to the Revenue-Utility problem. Similarly, noting that $V(\boldsymbol{x}) = 1 + \sum_{i\in\mathcal{N}} \frac{v_i}{v_0}$, we can also scale the preference weights of all products and the no-purchase option by the same multiplicative constant. Therefore, this assumption is innocuous as long as the preference weights and revenues of the products take rational values.

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

Proof: The bound $(m+n)^{1+m}$ follows by counting the number of extreme point solutions of the Candidate 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, by a naive argument, 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 $2 + 2n \max_{i\in\mathcal{N}} r_i\, v_i$ by adapting a technique from Carstensen (1983). Note that $\mathsf{LP}(\gamma)$ 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, using the fact that there exists a binary-valued optimal solution to the Candidate LP taking values in $\mathcal{F}$, we have

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

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 them from the set $\{\ell_{\boldsymbol{x}}(\cdot) : \boldsymbol{x} \in \mathcal{F}\}$ without changing the function $\mathsf{LP}(\cdot)$. Hence, the number of lines necessary to describe the function $\mathsf{LP}(\cdot)$ is equal to the number of different slopes for the lines $\{\ell_{\boldsymbol{x}}(\cdot) : \boldsymbol{x} \in \mathcal{F}\}$. The slope of the line $\ell_{\boldsymbol{x}}(\cdot)$ is $-\sum_{i\in\mathcal{N}} v_i\, x_i$, which is an integer between $-n \max_{i\in\mathcal{N}} v_i$ and zero. Thus, there are at most $1 + n \max_{i\in\mathcal{N}} v_i$ different slopes for the lines $\{\ell_{\boldsymbol{x}}(\cdot) : \boldsymbol{x} \in \mathcal{F}\}$, 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 them from the set $\{\ell_{\boldsymbol{x}}(\cdot) : \boldsymbol{x} \in \mathcal{F}\}$ without changing the function $\mathsf{LP}(\cdot)$. Therefore, the number of lines necessary to describe the function $\mathsf{LP}(\cdot)$ is no larger than twice the number of different intercepts for the lines $\{\ell_{\boldsymbol{x}}(\cdot) : \boldsymbol{x} \in \mathcal{F}\}$. The intercept of the line $\ell_{\boldsymbol{x}}(\cdot)$ is $\sum_{i\in\mathcal{N}} r_i\, v_i\, x_i$, which is an integer between zero and $n \max_{i\in\mathcal{N}} r_i\, v_i$. Thus, there are at most $1 + n \max_{i\in\mathcal{N}} r_i\, v_i$ different intercepts, yielding at most $2 + 2\,n \max_{i\in\mathcal{N}} r_i\, v_i$ breakpoints. ∎

The bound $(m+n)^{1+m}$ is based on counting the number of extreme point solutions of an LP, so it holds even when v_i and $r_i\, v_i$ are not integers. We can construct problem instances where the number of candidate assortments increases exponentially with the number of constraints. In Appendix B, we give a problem instance with n^2 decision variables and $2n$ constraints, where the number of candidate assortments increases exponentially with n. In such a pathological problem instance, the preference weights and revenues must increase exponentially with n; otherwise, by the bounds $1 + n \max_{i\in\mathcal{N}} v_i$ and $2 + 2n \max_{i\in\mathcal{N}} r_i v_i$ in Theorem 4.2, the number of candidate assortments must increase linearly with n. Nevertheless, we emphasize that there is no practical reason for the preference weights or revenues of the products to depend on the number of products, especially when we consider a set of products in the same product category. Thus, we expect the number of candidate assortments in practical applications to increase linearly with n. For the problem instances in our computational experiments, we can construct the collection of candidate assortments in a matter of seconds, even when we have hundreds of products and hundreds of constraints under consideration. We work with noninteger values for v_i and $r_i v_i$ in these problem instances and observe that the number of candidate assortments is roughly linear in n.

5. 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 see how the expected revenue and the expected utility change as a function of λ. In other words, 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 $\lambda \geq 0$. The key observation is that the parameter λ does not play any role in the Candidate LP. Thus, the collection of candidate assortments $\{\boldsymbol{x}^\ell_{\mathsf{Cand}} : \ell = 1,\ldots,L\}$, which is obtained by solving the Candidate LP for all $\gamma \in \mathbb{R}$ through parametric LP, is <u>*independent*</u> of the value of λ. When solving the Revenue-Utility problem for any value of $\lambda \geq 0$, we can use the same candidate assortments; that is, for all $\lambda \geq 0$, $\boldsymbol{x}^*_\lambda = \arg\max \left\{\mathsf{Rev}(\boldsymbol{x}^\ell_{\mathsf{Cand}};\boldsymbol{r}) + \lambda\,\mathsf{Util}(\boldsymbol{x}^\ell_{\mathsf{Cand}}) : \ell = 1,\ldots,L\right\}$. In Figure 2, we show the efficient frontier of attainable expected revenue-utility pairs $\{(\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r}), \mathsf{Util}(\boldsymbol{x}^*_\lambda)) : \lambda \geq 0\}$ for a particular problem instance. For fixed ℓ, $f^\ell(\lambda) = \mathsf{Rev}(\boldsymbol{x}^\ell_{\mathsf{Cand}};\boldsymbol{r}) + \lambda\,\mathsf{Util}(\boldsymbol{x}^\ell_{\mathsf{Cand}})$ is linear in λ, so solving the problem $\max\left\{\mathsf{Rev}(\boldsymbol{x}^\ell_{\mathsf{Cand}};\boldsymbol{r}) + \lambda\,\mathsf{Util}(\boldsymbol{x}^\ell_{\mathsf{Cand}}) : \ell = 1,\ldots,L\right\}$ for all $\lambda \geq 0$ corresponds to finding the pointwise maximum of the L linear functions $\left\{f^\ell(\cdot) : \ell = 1,\ldots,L\right\}$. To find the pointwise maximum of the L linear functions $\left\{f^\ell(\cdot) : \ell = 1,\ldots,L\right\}$, we simply need to find the intersection of each pair of lines and identify the maximum of the lines between successive intersections.

In the next lemma, we give three properties of the efficient frontier. By the first two properties, if we increase λ so that we put more weight on the expected utility, then $\mathsf{Util}(\boldsymbol{x}^*_\lambda)$ increases and $\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r})$ decreases. By the third property, if we connect the successive points on the efficient frontier, then the boundary of the efficient frontier is concave. The proof of the lemma uses standard arguments in multi-objective optimization. We defer it to Appendix C.

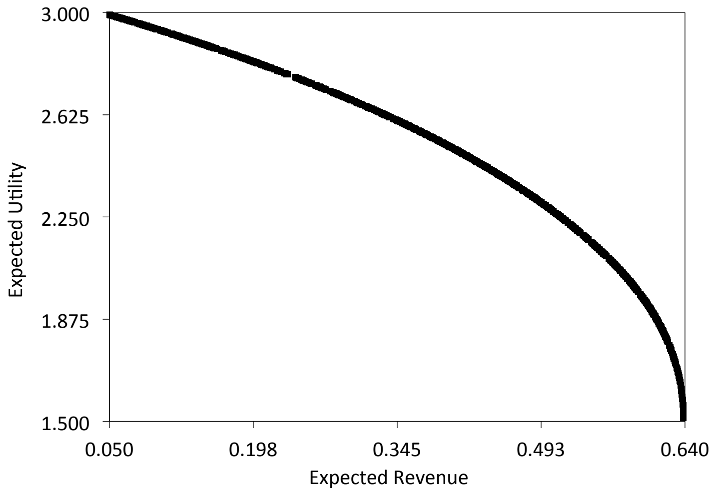

Figure 2 The efficient frontier for a problem instance with $n=1000$, $\mathcal{F}=\{\boldsymbol{x}\in\{0,1\}^n : \sum_{i\in\mathcal{N}} x_i \le 100\}$, $v_0=5$, r_i is sampled from the uniform distribution over $[0,1]$ for each $i\in\mathcal{N}$ and $v_i = 1-r_i$.

Lemma 5.1 (Efficient Frontier) *Using $\boldsymbol{x}^*_\lambda$ to denote an optimal solution to the* Revenue-Utility *problem as a function of λ, for $\lambda>\beta$, we have* $\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r}) \le \mathsf{Rev}(\boldsymbol{x}^*_\beta;\boldsymbol{r})$, $\mathsf{Util}(\boldsymbol{x}^*_\lambda) \ge \mathsf{Util}(\boldsymbol{x}^*_\beta)$ *and*

$$-\frac{1}{\lambda}\Big\{\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r})-\mathsf{Rev}(\boldsymbol{x}^*_\beta;\boldsymbol{r})\Big\} \;\le\; \mathsf{Util}(\boldsymbol{x}^*_\lambda)-\mathsf{Util}(\boldsymbol{x}^*_\beta) \;\le\; -\frac{1}{\beta}\Big\{\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r})-\mathsf{Rev}(\boldsymbol{x}^*_\beta;\boldsymbol{r})\Big\}.$$

By Corollary 3.3, if there is no constraint on the offered assortment, then a revenue-ordered assortment solves the Revenue-Utility problem. Using the second property in Lemma 5.1, we can see that the optimal revenue-ordered assortment includes more products as the weight λ on the expected utility increases. Specifically, if $\lambda>\beta$, then we have $\mathsf{Util}(\boldsymbol{x}^*_\lambda)\ge\mathsf{Util}(\boldsymbol{x}^*_\beta)$ by the second property, but we have $\mathsf{Util}(\boldsymbol{x}^*_\lambda)\ge\mathsf{Util}(\boldsymbol{x}^*_\beta)$ if and only if $V(\boldsymbol{x}^*_\lambda)\ge V(\boldsymbol{x}^*_\beta)$. We can choose $\boldsymbol{x}^*_\lambda$ and $\boldsymbol{x}^*_\beta$ as revenue-ordered assortments when there is no constraint, so we have $V(\boldsymbol{x}^*_\lambda)\ge V(\boldsymbol{x}^*_\beta)$ if and only if the revenue-ordered assortment $\boldsymbol{x}^*_\lambda$ includes more products than the revenue-ordered assortment $\boldsymbol{x}^*_\beta$. Moreover, using the third property in Lemma 5.1, it is evident that if we connect the successive points on the efficient frontier, then the boundary of the efficient frontier is concave. Specifically, let $\{(\mathsf{Rev}^k,\mathsf{Util}^k): k=1,\ldots,K\}$ be the points on the efficient frontier. For $\lambda^1>\lambda^2>\ldots>\lambda^K$, we index these points such that $\mathsf{Rev}^k=\mathsf{Rev}(\boldsymbol{x}^*_{\lambda^k};\boldsymbol{r})$ and $\mathsf{Util}^k=\mathsf{Util}(\boldsymbol{x}^*_{\lambda^k})$. No two points on the efficient frontier can have the same expected revenue, so since $\lambda^1>\lambda^2>\ldots>\lambda^K$, by the first property, we have $\mathsf{Rev}^1<\mathsf{Rev}^2<\ldots<\mathsf{Rev}^K$. Since $\lambda^{k-1}>\lambda^k>\lambda^{k+1}$, by the third property, we get

$$\frac{\mathsf{Util}^k-\mathsf{Util}^{k+1}}{\mathsf{Rev}^k-\mathsf{Rev}^{k+1}} \;\le\; -\frac{1}{\lambda^k} \;\le\; \frac{\mathsf{Util}^{k-1}-\mathsf{Util}^k}{\mathsf{Rev}^{k-1}-\mathsf{Rev}^k},$$

which implies that the slope of the line segment that connects the points $(\mathsf{Rev}^{k-1},\mathsf{Util}^{k-1})$ and $(\mathsf{Rev}^k,\mathsf{Util}^k)$ is greater than the slope of the line segment that connects the points $(\mathsf{Rev}^k,\mathsf{Util}^k)$

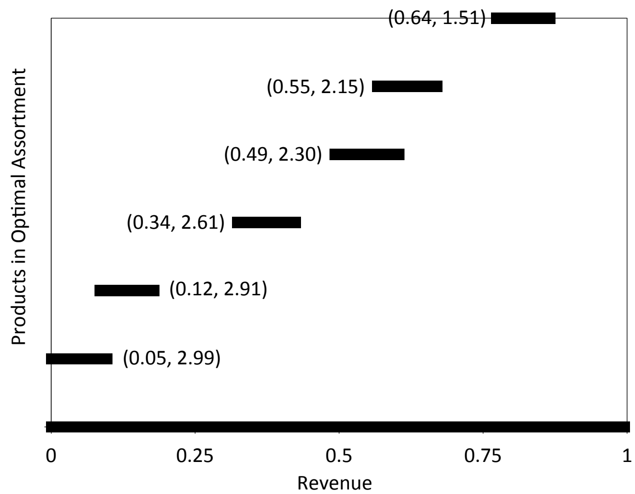

Figure 3 Optimal assortments corresponding to six expected revenue-utility pairs $(0.05, 2.99)$, $(0.12, 2.91)$, $(0.34, 2.61)$, $(0.49, 2.30)$, $(0.55, 2.15)$ and $(0.64, 1.51)$ on the efficient frontier.

and $(\mathsf{Rev}^{k+1}, \mathsf{Util}^{k+1})$. Thus, if we connect the successive points on the efficient frontier, then the boundary of the efficient frontier is concave. Coupled with concavity, the boundary of the efficient frontier on its right-side may be rather steep, in which case, it may be possible to increase the expected utility substantially without incurring a huge loss in the expected revenue. In Figure 2, for example, the expected revenue-utility pairs $(0.638, 1.513)$ and $(0.632, 1.692)$ are on the efficient frontier, so we can increase the expected utility by 11.89% by incurring a loss of no more than 1% in the expected revenue. An increase in $\mathsf{Util}(\boldsymbol{x})$ implies an increase in $V(\boldsymbol{x})$, which, in turn, implies a decrease in the probability that a customer leaves the system without a purchase. In this problem instance, an increase of 11.89% in the expected utility corresponds to a decrease of 16.46% in the no-purchase probability. To make this numerical observation more concrete, by the third property in Lemma 5.1, for $\lambda > 0$ with $\mathsf{Rev}(\boldsymbol{x}^*_\lambda; \boldsymbol{r}) \neq \mathsf{Rev}(\boldsymbol{x}^*_0; \boldsymbol{r})$, we have

$$\frac{\mathsf{Util}(\boldsymbol{x}^*_\lambda) - \mathsf{Util}(\boldsymbol{x}^*_0)}{\mathsf{Rev}(\boldsymbol{x}^*_0; \boldsymbol{r}) - \mathsf{Rev}(\boldsymbol{x}^*_\lambda; \boldsymbol{r})} \;\geq\; \frac{1}{\lambda}.$$

Thus, instead of maximizing only the expected revenue, which corresponds to using a value of zero for λ, if we use a positive value of λ, then the increase in the expected utility relative to the decrease in the expected revenue is at least $1/\lambda$, which can be quite large when λ is small.

The optimal solution to the $\mathsf{Revenue\text{-}Utility}$ problem can be quite different for different values of λ. In Figure 2, the six expected revenue-utility pairs $(0.05, 2.99)$, $(0.12, 2.91)$, $(0.34, 2.61)$, $(0.49, 2.30)$, $(0.55, 2.15)$ and $(0.64, 1.51)$ are on the efficient frontier. In Figure 3, we show the optimal assortments for each one of these six expected revenue-utility pairs. Each cross on the horizontal axis shows the revenue of a product. Each sequence of the horizontal crosses shows the products offered in one of the six assortments. In the problem instance in Figure 2, as a function of its revenue, the preference weight of product i is $v_i = 1 - r_i$, so we can characterize a product only by its revenue. The expected-revenue utility pair $(0.64, 1.51)$ has a relatively large expected revenue. Its corresponding optimal assortment focuses on offering products with large revenues, but products with extremely large revenues have extremely small preference weights, so such products are not offered given that we can offer at most 100 products in this problem instance. The expected revenue-utility pair $(0.05, 2.99)$ has a relatively large expected utility. Its corresponding optimal assortment offers products with small revenues, which have large preference weights.

In our computational experiments, we can construct the collection of candidate assortments for practical problem instances in a matter of seconds. Nevertheless, a natural question is whether we can still use some information on the optimal solution to the Revenue-Utility problem to eliminate some of the candidate assortments from consideration. In the next lemma, we use an upper bound on $V(\boldsymbol{x}^*_\lambda)$ for this purpose. Specifically, we show that we can associate a threshold $\overline{\lambda}(\gamma)$ for each candidate assortment $\boldsymbol{x}_{\mathsf{LP}}(\gamma)$ such that we can eliminate the candidate assortment $\boldsymbol{x}_{\mathsf{LP}}(\gamma)$ from consideration whenever we have $\lambda < \overline{\lambda}(\gamma)$ in the Revenue-Utility problem.

Lemma 5.2 (Thresholds) *Letting $\boldsymbol{x}^*_\lambda$ be an optimal solution to the* Revenue-Utility *problem, $\bar{t}$ be an upper bound on $V(\boldsymbol{x}^*_\lambda)$ and $\overline{\lambda}(\gamma) = \frac{1+V(\boldsymbol{x}_{\mathsf{LP}}(\gamma))}{1+\bar{t}} \times (\mathsf{Rev}(\boldsymbol{x}_{\mathsf{LP}}(\gamma); \boldsymbol{r}) - \gamma)$, the collection of assortments $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \lambda \geq \overline{\lambda}(\gamma)\}$ includes an optimal solution to the* Revenue-Utility *problem.*

The proof of the lemma uses the fact that if $\boldsymbol{x}^*_\lambda$ is an optimal solution to the Revenue-Utility problem, then $V(\boldsymbol{x}^*_\lambda)$ is an optimal solution to the outer maximization problem in the Parametric problem. We give the proof in Appendix D. To get the efficient frontier, we solve the Revenue-Utility problem simultaneously for all $\lambda \geq 0$. Each candidate assortment $\boldsymbol{x}_{\mathsf{LP}}(\gamma)$ corresponds to a specific value for $\gamma \in \mathbb{R}$, so we can compute $\overline{\lambda}(\gamma)$ for each candidate assortment $\boldsymbol{x}_{\mathsf{LP}}(\gamma)$ and eliminate the candidate assortment $\boldsymbol{x}_{\mathsf{LP}}(\gamma)$ from consideration when solving the Revenue-Utility problem for any $\lambda < \overline{\lambda}(\gamma)$. Since $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma \in \mathbb{R}\}$ includes an optimal solution to the Revenue-Utility problem, a simple upper bound on $V(\boldsymbol{x}^*_\lambda)$ is $\bar{t} = \max\{V(\boldsymbol{x}_{\mathsf{LP}}(\gamma)) : \gamma \in \mathbb{R}\}$. In the problem instance in Figure 2, we have 861 candidate assortments in the collection $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma \in \mathbb{R}\}$ and considering the values of λ in $[0, 0.95]$ is sufficient to construct the efficient frontier. In Figure 4, for each $\lambda \in [0, 0.95]$, we show the number of candidate assortments left after eliminating the unnecessary assortments using Lemma 5.2. When solving the Revenue-Utility problem with $\lambda = 0.2$, for example, we need to consider 455 candidate assortments, eliminating the other 406. In Appendix D, we also give a stronger result that uses both upper and lower bounds on $V(\boldsymbol{x}^*_\lambda)$ to eliminate candidate assortments, but using this result to eliminate candidate assortments is more computationally intensive. Lastly, note that not all candidate assortments appear on the efficient frontier. In Figure 2, we have 682 points on the efficient frontier, but 861 candidate assortments in the collection $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma \in \mathbb{R}\}$.

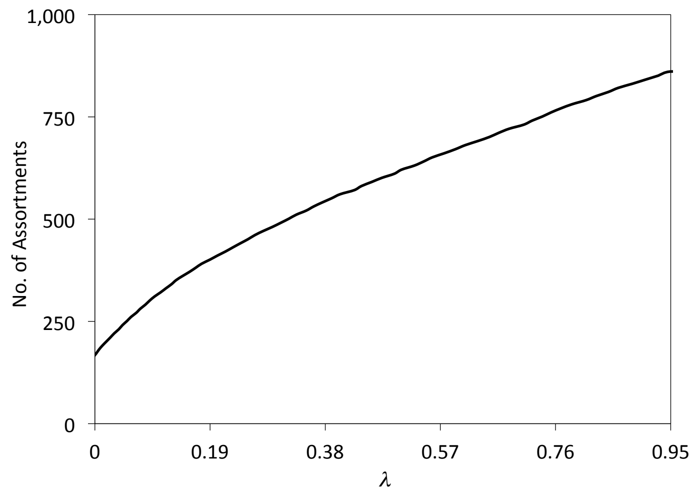

Figure 4 Number of candidate assortments left after eliminating the unnecessary assortments using Lemma 5.2.

6. Fully Polynomial-Time Approximation Scheme

In this section, we give a fully polynomial-time approximation scheme that allows us to use a smaller collection of candidate assortments to obtain a near-optimal solution to the Revenue-Utility problem, while controlling the quality of the solutions that we obtain.

Maximizing Expected Revenue Through an LP:

Our approximation scheme requires repeatedly finding an assortment that maximizes only the expected revenue under totally unimodular constraints on the offered assortment. Thus, we begin by formulating an LP to find an assortment that maximizes only the expected revenue and use this LP as a subroutine in our approximation scheme. Furthermore, there can be settings where the goal is to maximize only the expected revenue without paying attention to the expected utility. In this case, the LP that we formulate indicates that we can simply solve a single LP to find an assortment that maximizes the expected revenue under totally unimodular constraints. After formulating the LP, we give our approximation scheme. Specifically, we consider the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\ \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r})$, which finds an assortment that maximizes the expected revenue. Using the decision variables $\boldsymbol{y}=(y_1,\ldots,y_n)\in\mathbb{R}^n_+$ and $y_0\in\mathbb{R}_+$, we consider the LP

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

The Charnes-Cooper transformation is useful for converting a fractional program into an LP (Charnes and Cooper 1962). We can use this transformation to immediately show that the above LP has the same optimal objective value as the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r})$, but we will ultimately need extra care to convert an optimal solution to the above LP into a binary solution that characterizes an assortment. Since $\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) = \frac{\sum_{i\in\mathcal{N}} r_i\, v_i\, x_i}{v_0+\sum_{i\in\mathcal{N}} v_i\, x_i}$, the problem $\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r})$ is equivalent to

$$\max_{\boldsymbol{x}\in\mathbb{Z}_+^n}\left\{\frac{\sum_{i\in\mathcal{N}} r_i\, v_i\, x_i}{v_0+\sum_{i\in\mathcal{N}} v_i\, x_i}\;\middle|\; \mathbf{A}\boldsymbol{x}\le \boldsymbol{b},\; x_i \le 1\;\forall\, i\in\mathcal{N}\right\}. \qquad (\mathsf{Max\text{-}Revenue})$$

By the Charnes-Cooper transformation, for some fixed $\boldsymbol{f},\boldsymbol{c}\in\mathbb{R}_+^n$, $d\in\mathbb{R}_+$, $\mathbf{A}\in\mathbb{R}^{m\times n}$ and $\boldsymbol{b}\in\mathbb{R}^m$, the optimal objective value of the fractional program

$$\max_{\boldsymbol{x}\in\mathbb{R}_+^n}\left\{\frac{\boldsymbol{f}^\top\boldsymbol{x}}{d+\boldsymbol{c}^\top\boldsymbol{x}}\;\middle|\;\mathbf{A}\,\boldsymbol{x}\le\boldsymbol{b},\quad \boldsymbol{x}\le\boldsymbol{e}\right\}$$

is equal to that of the LP $\max_{(\boldsymbol{y},y_0)\in\mathbb{R}_+^{n+1}}\{\boldsymbol{f}^\top\boldsymbol{y}\,|\,\mathbf{A}\,\boldsymbol{y}\le y_0\,\boldsymbol{b},\;\boldsymbol{y}\le y_0\,\boldsymbol{e},\;\boldsymbol{c}^\top\boldsymbol{y}+d\,y_0=1\}$. In addition, letting $(\boldsymbol{y}^*,y_0^*)$ be an optimal solution to the LP, $(y_1^*/y_0^*,\ldots,y_n^*/y_0^*)$ is an optimal solution to the fractional program. Note that the objective function of the Max-Revenue problem is quasiconvex, so an optimal solution is attained at an extreme point. Since the matrix $\mathbf{A}$ is totally unimodular, the polyhedron $\{\boldsymbol{x}\in[0,1]^n:\mathbf{A}\,\boldsymbol{x}\le\boldsymbol{b}\}$ has integral extreme points. Therefore, we can relax the integrality constraints in the Max-Revenue problem, in which case, we can apply the Charnes-Cooper transformation to this problem. Setting $\boldsymbol{f}=(r_1\,v_1,\ldots,r_n\,v_n)$, $\boldsymbol{c}=(v_1,\ldots,v_n)$ and $d=v_0$ in the Charnes-Cooper transformation, it immediately follows that the optimal objective value of the Revenue LP is equal to that of the Max-Revenue problem, which is in turn equal to the optimal objective value of the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r})$. In the next theorem, we go one step beyond the equality of the optimal objective values by showing that we can convert an optimal solution to the Revenue LP into a binary solution that characterizes an assortment.

Theorem 6.1 (Revenue LP) *In a basic optimal solution* $(\boldsymbol{y}^*,y_0^*)$ *to the* Revenue LP*, we have* $y_i^*/y_0^*\in\{0,1\}$ *for all* $i\in\mathcal{N}$*, so the solution* $\left(\frac{y_1^*}{y_0^*},\ldots,\frac{y_n^*}{y_0^*}\right)$ *is optimal to the problem* $\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r})$.

Proof: For the solution $(\boldsymbol{y}^*,y_0^*)$, defining the slack variables $\boldsymbol{\zeta}^*=(\zeta_1^*,\ldots,\zeta_m^*)$ and $\boldsymbol{\gamma}^*=(\gamma_1^*,\ldots,\gamma_n^*)$ for the first and second sets of constraints, by the constraints of the Revenue LP, we have

$$\mathbf{A}\,\boldsymbol{y}^*+\boldsymbol{\zeta}^*=y_0^*\,\boldsymbol{b},\qquad y_i^*+\gamma_i^*=y_0^*\;\;\forall\, i\in\mathcal{N},\qquad v_0\,y_0^*+\sum_{i\in\mathcal{N}}v_i\,y_i^*=1.$$

Note that $y_0^*>0$; otherwise, we have $y_i^*=0$ for all $i\in\mathcal{N}$ by the second set of constraints, in which case, the left side of the third constraint must be zero. Also, since $y_0^*>0$, we cannot have $y_i^*=0$ and $\gamma_i^*=0$ for some $i\in\mathcal{N}$. So, defining the sets $\mathcal{N}^0=\{i\in\mathcal{N}:y_i^*$ is basic and γ_i^* is basic$\}$, $\mathcal{N}^1=\{i\in\mathcal{N}:y_i^*$ is basic and γ_i^* is nonbasic$\}$ and $\mathcal{N}^2=\{i\in\mathcal{N}:y_i^*$ is nonbasic and γ_i^* is basic$\}$, the three sets partition $\mathcal{N}$, satisfying $|\mathcal{N}^0|+|\mathcal{N}^1|+|\mathcal{N}^2|=n$. We claim that $y_i^*\in\{0,y_0^*\}$ for all $i\in\mathcal{N}^0$. To prove the claim, using $\mathcal{M}=\{1,\ldots,m\}$ to index the rows of the matrix $\mathbf{A}$, let

$\mathcal{M}^0 = \{\ell \in \mathcal{M} : \zeta^*_\ell$ is nonbasic$\}$. Observe that the four sets $\mathcal{N}^0$, $\mathcal{N}^1$, $\mathcal{N}^2$ and $\mathcal{M}^0$ completely characterize the basis. Since $y^*_0 > 0$, the number of basic variables in the solution $(\boldsymbol{y}^*, y^*_0, \boldsymbol{\zeta}^*, \boldsymbol{\gamma}^*)$ is $1 + 2|\mathcal{N}^0| + |\mathcal{N}^1| + |\mathcal{N}^2| + m - |\mathcal{M}^0| = 1 + |\mathcal{N}^0| + n + m - |\mathcal{M}^0|$. Since there are $m+n+1$ constraints in the Revenue LP, we must have $1 + |\mathcal{N}^0| + n + m - |\mathcal{M}^0| = m + n + 1$, so $|\mathcal{M}^0| = |\mathcal{N}^0|$. On the other hand, we have $\zeta^*_\ell = 0$ for all $\ell \in \mathcal{M}^0$ and $y^*_i = 0$ for all $i \in \mathcal{N}^2$. Moreover, for all $i \in \mathcal{N}^1$, since $\gamma^*_i = 0$, we have $y^*_i = y^*_0$. Therefore, letting $a_{\ell i}$ be the $(\ell, i)^{\text{th}}$ entry of the matrix $\mathbf{A}$ and b_ℓ be the ℓ^{th} entry of the vector $\boldsymbol{b}$, by the first set of constraints in the Revenue LP, for each $\ell \in \mathcal{M}_0$, we have $\sum_{i\in\mathcal{N}^0} a_{\ell i}\, y^*_i + \sum_{i\in\mathcal{N}^1} a_{\ell i}\, y^*_0 = b_\ell\, y^*_0$, which is equivalent to

$$\sum_{i\in\mathcal{N}^0} a_{\ell i}\, y^*_i = \left(b_\ell - \sum_{i\in\mathcal{N}^1} a_{\ell i} \right) y^*_0 \quad \forall\, \ell \in \mathcal{M}^0.$$

Viewing $\{y^*_i : i \in \mathcal{N}^0\}$ as unknowns, there are $|\mathcal{N}^0|$ unknowns and $|\mathcal{M}^0|$ equations above. Since $|\mathcal{M}^0| = |\mathcal{N}^0|$, the number of equations is equal to the number of unknowns.

In the above system of equations, there must be a unique solution for $\{y^*_i : i \in \mathcal{N}^0\}$; otherwise, we have two different basic solutions corresponding to the same basis. Therefore, $\{y^*_i : i \in \mathcal{N}^0\}$ is given by the inverse of the matrix with entries $\{a_{\ell i} : \ell \in \mathcal{M}^0,\ i \in \mathcal{N}^0\}$ multiplying the vector with entries $\{(b_\ell - \sum_{i\in\mathcal{N}^1} a_{\ell i})\, y^*_0 : \ell \in \mathcal{M}^0\}$. Since the matrix $\mathbf{A}$ is totally unimodular, all of its entries take values in $\{-1, 0, +1\}$ and the inverses of its submatrices have integer entries (Proposition 2.1 in Chapter III.1, Nemhauser and Wolsey 1988). Thus, for each $i \in \mathcal{N}^0$, y^*_i must be an integer multiple of y^*_0. Also, by the second set of constraints in the Revenue LP, we have $y^*_i \le y^*_0$, which implies that $y^*_i \in \{0, y^*_0\}$ for all $i \in \mathcal{N}^0$, establishing the claim.

If $i \in \mathcal{N}^0$, then $y^*_i \in \{0, y^*_0\}$ by the claim in the previous paragraph. If $i \in \mathcal{N}^1$, then $\gamma^*_i = 0$, which implies that $y^*_i = y^*_0$. If $i \in \mathcal{N}^2$, then $y^*_i = 0$. Thus, we have $y^*_i / y^*_0 \in \{0, 1\}$ for all $i \in \mathcal{N}$. ■

Approximation Scheme for Revenue-Utility Maximization:

We return to the problem of solving the Revenue-Utility problem for a fixed value of λ. As discussed at the end of Section 4, it takes a matter of seconds to generate the candidate assortments for the problem instances in our computational experiments, but one can come up with pathological cases in which the number of candidate assortments increases exponentially with the input size. We give an approximation scheme that keeps the number of candidate assortments polynomial in the input size while incurring a bounded loss in optimality. By Theorem 3.2, letting $\boldsymbol{x}^*$ be an optimal solution to the Revenue-Utility problem, we can recover an optimal solution to the Revenue-Utility problem by solving the inner maximization problem in the Parametric problem with $t = V(\boldsymbol{x}^*)$. Letting $V_{\min} = \min_{i\in\mathcal{N}} v_i$ and $V_{\max} = \max_{i\in\mathcal{N}} v_i$, since $\mathcal{F}$ includes a nonempty assortment, we have $V(\boldsymbol{x}^*) \in [V_{\min}, nV_{\max}]$. In our approximation scheme, for some fixed accuracy parameter $\rho > 0$, we focus on integer powers of $1+\rho$ over the interval $[V_{\min}, nV_{\max}]$. Instead of solving the problem $\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t)\,\boldsymbol{e})$ with $t=V(\boldsymbol{x}^*)$, we solve this problem for each value of t that is an integer power of $1+\rho$. For each value of t, we get a candidate assortment by solving the problem $\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t)\,\boldsymbol{e})$. We check the objective value that each candidate assortment provides for the Revenue-Utility problem and pick the best one. Here is our approximation scheme.

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

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

Approximation Scheme: 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, $\widehat{\boldsymbol{x}}=\arg\max_{t\in\mathsf{Grid}}\ \mathsf{Rev}(\widehat{\boldsymbol{x}}_t,\boldsymbol{r})+\lambda\,\log(1+V(\widehat{\boldsymbol{x}}_t))$.

In the next theorem, we show that the output $\widehat{\boldsymbol{x}}$ of the approximation scheme provides a $1/(1+\rho)$-approximate solution to the Revenue-Utility problem.

Theorem 6.2 (Approximation Scheme) *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

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

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 approximation scheme, $\widehat{\boldsymbol{x}}_{\widehat{t}}=\arg\max_{\boldsymbol{x}\in\mathcal{F}}\mathsf{Rev}(\boldsymbol{x}\,;\,\boldsymbol{r}+\lambda\,(\,1+\widehat{t}\,)\,\boldsymbol{e})$.

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, so we get

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

where (c) holds because $\frac{a}{1+a}$ is increasing in a and $V(\boldsymbol{x}^*)\ge\widehat{t}$, so we have $\frac{V(\boldsymbol{x}^*)}{1+V(\boldsymbol{x}^*)}\ge\frac{\widehat{t}}{1+\widehat{t}}$, 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

$(1+a)^{1+\rho} \geq 1+(1+\rho)\,a$. Using this inequality with $a = V(\boldsymbol{x}^*)/(1+\rho)$, we get $\left(1+\frac{V(\boldsymbol{x}^*)}{1+\rho}\right)^{1+\rho} \geq 1+V(\boldsymbol{x}^*)$. Therefore, we have

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

where (e) holds because $\mathsf{Rev}(\boldsymbol{x}^*;\boldsymbol{r}) \geq 0$. By the three chains of inequalities displayed above, we have $\mathsf{Rev}\big(\widehat{\boldsymbol{x}};\boldsymbol{r}\big) + \lambda \log(1+V(\widehat{\boldsymbol{x}})) \geq Z^*_\lambda/(1+\rho)$, as desired. ■

Our approximation scheme constructs one candidate assortment for each $t \in \mathsf{Grid}$ 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, which implies that we can find a $1/(1+\rho)$-approximate solution to the $\mathsf{Revenue\text{-}Utility}$ problem by checking the objective values of $O\left(\frac{1}{\rho}\log(nV_{\max}/V_{\min})\right)$ candidate assortments. Thus, the accuracy parameter ρ balances solution quality with computational effort. The running time of our approximation scheme is polynomial in the input size and $\frac{1}{\rho}$, making it a fully polynomial-time approximation scheme. Our approximation scheme solves the problem $\max_{\boldsymbol{x}\in\mathcal{F}}\ \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t)\,\boldsymbol{e})$ for each $t \in \mathsf{Grid}$. In this problem, we maximize only the expected revenue and the revenue of product i is $r_i+\lambda(1+t)$. To solve this problem, we can replace the revenue of product i with $r_i+\lambda(1+t)$ in the $\mathsf{Revenue\ LP}$.

7. Applications of Totally Unimodular Constraints

In this section, we give examples of assortment optimization settings that fit our formulation with totally unimodular constraints. In each of these settings, we can use our approach to solve the $\mathsf{Revenue\text{-}Utility}$ problem, yielding an assortment that maximizes a linear combination of the expected revenue and the expected utility.

7.1 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 $\mathcal{F} = \{\boldsymbol{x}\in\{0,1\}^n : \sum_{i\in\mathcal{N}} x_i \leq b\}$. Here, the constraint matrix $\mathbf{A} = (1,\ldots,1)$ has a single row consisting of all ones. Therefore, the matrix $\mathbf{A}$ 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. 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 (Corollary 2.10 in Chapter III.1, 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.

7.2 Display Location Effects

When a product is displayed at a prominent location in a retail store or at the top of a list of search results in online retail, customers may be more likely to choose it. We consider the case where the preference weight of a product depends on the location at which the product is displayed (Abeliuk et al. 2016, Chen and Jiang 2017, L'Ecuyer et al. 2017, Gallego et al. 2019). 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 locations at which we can display the items. If we display item i at location k, then its preference weight is v_{ik}. We use the vector $\boldsymbol{x} = \{x_{ik} : i \in \mathcal{N},\ k \in \mathcal{K}\} \in \{0,1\}^{n \times K}$ to capture our assortment decisions, where $x_{ik} = 1$ if and only if we offer item i at location k. Thus, the set of products is $\mathcal{N} \times \mathcal{K}$ and offering the product (i,k) corresponds to displaying item i at location k.

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

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

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

where the first set of constraints ensures that each item is displayed at no more than one location and the second set of constraints ensures that each location is used by at most one item. The above constraint matrix is the constraint matrix of an assignment problem, which is totally unimodular (Corollary 2.9 in Chapter III.1, Nemhauser and Wolsey 1988). If we convert the first set of constraints into equalities, then all products must be offered and we choose only the product locations. If the locations have a natural sequence, as in online search results, then our formulation allows skipping a location, but if $v_{i1} \geq v_{i2} \geq \ldots \geq v_{iK}$ for all $i \in \mathcal{N}$, so that locations with smaller indices are more preferable, then we can show that it is not optimal to skip a location.

7.3 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

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

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

The constraints ensure 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 offered products, then we can add the constraint $\sum_{(i,k) \in \mathcal{N} \times \mathcal{K}} x_{ik} \leq b$. The additional constraint amounts to adding a row of ones to the constraint matrix and the constraint matrix is still an interval matrix. 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, 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.

7.4 Pricing with a Price Ladder Constraint

We can extend the pricing model 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 ranking 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 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) give additional applications. We index the items by $\mathcal{N} = \{1, 2, \ldots, n\}$ and the possible price levels by $\mathcal{K} = \{1, 2, \ldots, K\}$. In our formulation, all products must be offered and we need to choose their prices. Without loss of generality, we order the prices $\{r_k : k \in \mathcal{K}\}$ corresponding to the different price levels so that $r_1 \geq r_2 \geq \ldots \geq r_K$. 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 $i-1$. We capture our assortment decisions using 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. Thus, 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 $k, \ldots, K$

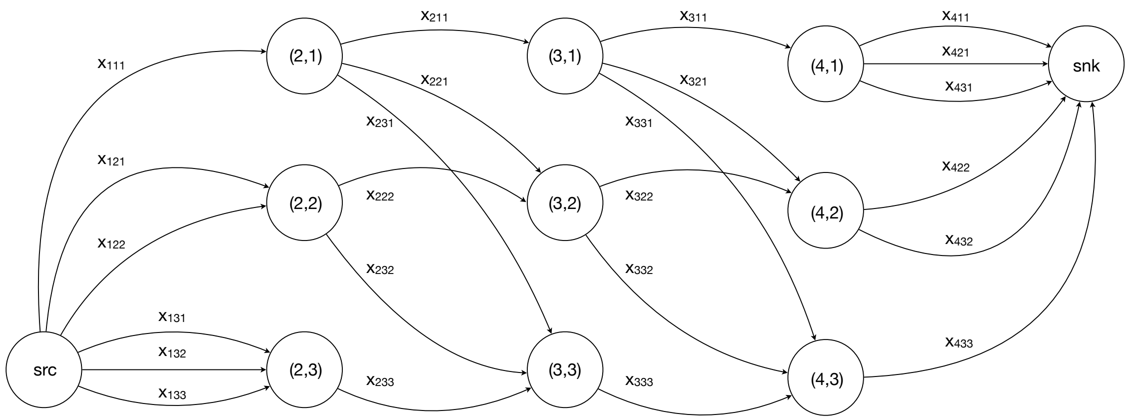

Figure 5 The graph corresponding to the price ladder constraints for a problem instance with $n = 4$ items and $K = 3$ price levels.

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 (Proposition 3.1 in Chapter III.1, Nemhauser and Wolsey 1988).

Proposition 7.1 (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 this constraint explicitly in our formulation of the price ladder constraints. In Figure 5, we show the construction of the graph for the price ladder constraints for a problem instance with $n = 4$ items and $K = 3$ price levels. ■

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

7.5 Product Precedence Constraints

We focus on assortment optimization 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 must be offered 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 totally unimodular (Proposition 2.6 in Chapter III.1, 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.

8. Computational Experiments

In our first set of computational experiments, we work with a large number of randomly generated instances of the assortment optimization problem with display location effects. Our goal is to demonstrate that generating the candidate assortments and applying our approximation scheme requires reasonable computational effort. In Appendix E, we also give similar computational experiments on a large number of randomly generated instances of the pricing problem with discrete price menus. In our second set of computational experiments, we use a publicly available dataset that is provided by Expedia and work with pricing problems with discrete price menus. Our goal is to demonstrate the possible tradeoffs between the expected revenue and the expected utility on a dataset coming from a real-world setting.

8.1 Computational Results for Display Location Effects

We focus on assortment optimization problems with display location effects, which correspond to the application discussed in Section 7.2. We use the following approach to generate our test problems. In all of our 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 higher prices tend to be less attractive and the parameter β captures the price sensitivity shared by all items. To determine the preference weight of the no-purchase option, letting $\mathcal{N}_K$ be the set of K items with the smallest values for the preference weights $\{v_{iK} : i \in \mathcal{N}\}$, 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 item at each location, even if we offer all items 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 test problems as 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, which requires obtaining an optimal solution to the Candidate LP for all $\gamma \in \mathbb{R}$ through parametric LP. Using the 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 $\lambda \in \mathbb{R}_+$. Letting $\{\boldsymbol{x}^\ell_{\mathsf{Cand}} : \ell = 1, \ldots, L\}$ to be the collection of candidate assortments, the optimal objective value of the Revenue-Utility problem is $Z^*_\lambda = \max_{\ell=1,\ldots,L} \mathsf{Rev}(\boldsymbol{x}^\ell_{\mathsf{Cand}}; \boldsymbol{r}) + \lambda\, \mathsf{Util}(\boldsymbol{x}^\ell_{\mathsf{Cand}})$. For fixed $\boldsymbol{x}^\ell_{\mathsf{Cand}}$, the function $\lambda \mapsto \mathsf{Rev}(\boldsymbol{x}^\ell_{\mathsf{Cand}}; \boldsymbol{r}) + \lambda\, \mathsf{Util}(\boldsymbol{x}^\ell_{\mathsf{Cand}})$ is linear in λ. Therefore, the function $\lambda \mapsto Z^*_\lambda$ is the pointwise maximum of L linear functions, so it is continuous, piecewise linear and convex in λ. When λ takes values between the 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 (Chapter 5, Kleinberg and Tardos 2005).

We give our computational results in Table 2. In this table, 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

Param. (K, p_0)	# Cand. Assr.	# Opt. Assr.	CPU Secs.
$(15, 0.1)$	476.41	84.46	1.89
$(15, 0.3)$	474.30	97.42	1.85
$(15, 0.5)$	476.50	102.72	1.88
$(30, 0.1)$	1095.02	187.18	9.28
$(30, 0.3)$	1084.58	212.20	9.35
$(30, 0.5)$	1112.58	244.66	9.59

Param. (K, p_0)	# Cand. Assr.	# Opt. Assr.	CPU Secs.
$(45, 0.1)$	1514.96	290.4	21.77
$(45, 0.3)$	1476.48	323.02	21.33
$(45, 0.5)$	1533.12	379.16	22.14
$(60, 0.1)$	1650.78	356.28	37.48
$(60, 0.3)$	1628.40	391.08	36.56
$(60, 0.5)$	1527.14	360.62	34.23

Table 2 The number of candidate assortments, number of optimal assortments and CPU seconds for assortment optimization problems with display location effects.

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. We emphasize that the collection of candidate assortments is guaranteed to include an optimal solution to the Revenue-Utility problem, but some candidate assortments may never be optimal to the Revenue-Utility problem, regardless of the value of λ. Also, using Lemma 5.2 to eliminate candidate assortments did not improve our computation times. The fourth column shows the average CPU seconds to generate all candidate assortments for a test problem. We carried out our computational experiments in Java 1.7.0 on a 2.8 GHz Intel Core i7 CPU running MacOS with 16 GB of RAM. We used Gurobi 6.5.0 as our LP solver.

For the smaller test problems with $K = 15$ locations, we can generate all candidate assortments in less than two seconds. For the larger test problems with $K = 60$ locations, this computational effort increases to just over half a minute. After generating the candidate assortments, in each of our test problems, it took less than a millisecond to construct the efficient frontier. The number of candidate assortments L increases approximately linearly with the number of products. Specifically, the number of products corresponds to the number of item-location pairs, which is nK. Normalizing the number of candidate assortments by dividing L by nK, the average value of $L/(nK)$ is roughly constant. For the test problems with 15, 30, 45, and 60 locations, the average values of $L/(nK)$ are, respectively, 0.53 0.61, 0.56, and 0.45. This result is consistent with Theorem 4.2, which shows that L is bounded above by an expression that is linear in the number of products.

Figure 6 gives the efficient frontier for a problem instance with $K = 45$ and $p_0 = 0.5$. Each data point corresponds to a different value of λ with a different optimal solution for the Revenue-Utility problem. Between successive data points, the optimal solution does not change.

Approximation Scheme: Thus far, our discussion has focused on exact solutions of the Revenue-Utility problem. We now discuss the performance of our approximation scheme when we find approximate solutions to the Revenue-Utility problem for some representative values of λ. Based

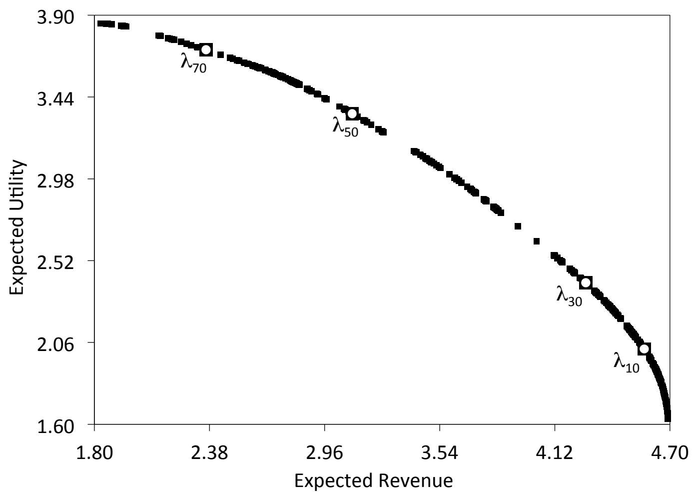

Figure 6 Efficient frontier for a particular instance of the assortment optimization problem with display location effects with $K = 45$ and $p_0 = 0.5$.

on the discussion presented earlier in this section, the optimal objective value Z^*_λ of the Revenue-Utility problem is a continuous, piecewise linear 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 breakpoints. 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 approximation scheme 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 $\mathbf{A}\boldsymbol{x} \le \boldsymbol{b}$ 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 6.1, we can maximize the expected revenue by solving a single LP. In Figure 6, we show the expected revenue-utility pairs corresponding to λ_{10}, λ_{30}, λ_{50} and λ_{70} for a particular problem instance.

We give our computational results in Table 3. In this table, the first column shows the parameter configuration (K, p_0). The second column shows the value of $\lambda \in \{\lambda_{10}, \lambda_{30}, \lambda_{50}, \lambda_{70}\}$ on which we focus. The remainder of the table has two blocks, each containing three columns. The two blocks, respectively, show the performance of our approximation scheme when we use this scheme with the accuracy parameters of $\rho = 1$ and $\rho = 0.1$. In each block, the first column shows the maximum

		$\rho=1$			$\rho=0.1$		
Param. (K,p_0)	λ	Max. % Gap	CPU Secs.	# Cand. Assr.	Max. % Gap	CPU Secs.	# Cand. Assr.
$(15,0.1)$	λ_{10}	0.044	0.35	23.4	0.001	1.77	157.4
	λ_{30}	0.289	0.40	23.4	0.003	1.88	157.4
	λ_{50}	0.187	0.45	23.4	0.007	1.88	157.4
	λ_{70}	0.316	0.42	23.4	0.006	1.94	157.4
$(15,0.3)$	λ_{10}	0.050	0.46	24.6	0.000	2.29	165.4
	λ_{30}	0.157	0.49	24.6	0.002	2.14	165.4
	λ_{50}	0.185	0.42	24.6	0.004	2.03	165.4
	λ_{70}	0.307	0.44	24.6	0.003	1.99	165.4
$(15,0.5)$	λ_{10}	0.032	0.45	24.8	0.000	2.02	168.4
	λ_{30}	0.168	0.43	24.8	0.003	1.93	168.4
	λ_{50}	0.540	0.44	24.8	0.009	2.08	168.4
	λ_{70}	0.193	0.43	24.8	0.010	2.19	168.4
$(30,0.1)$	λ_{10}	0.028	1.08	29.0	0.001	4.51	197.4
	λ_{30}	0.175	1.28	29.0	0.004	4.46	197.4
	λ_{50}	0.196	1.06	29.0	0.005	4.20	197.4
	λ_{70}	0.284	1.13	29.0	0.006	4.35	197.4
$(30,0.3)$	λ_{10}	0.031	1.06	28.4	0.000	3.72	192.0
	λ_{30}	0.227	1.15	28.4	0.003	3.88	192.0
	λ_{50}	0.248	0.99	28.4	0.005	3.82	192.0
	λ_{70}	0.287	1.09	28.4	0.005	3.96	192.0
$(30,0.5)$	λ_{10}	0.022	0.88	24.8	0.001	3.23	167.8
	λ_{30}	0.165	0.87	24.8	0.004	3.20	167.8
	λ_{50}	0.290	0.91	24.8	0.008	3.25	167.8
	λ_{70}	0.291	0.90	24.8	0.005	3.44	167.8

		$\rho=1$			$\rho=0.1$		
Param. (K,p_0)	λ	Max. % Gap	CPU Secs.	# Cand. Assr.	Max. % Gap	CPU Secs.	# Cand. Assr.
$(45,0.1)$	λ_{10}	0.037	1.83	27.8	0.002	5.56	193.0
	λ_{30}	0.199	1.84	27.8	0.004	5.71	193.0
	λ_{50}	0.208	1.83	27.8	0.005	5.59	193.0
	λ_{70}	0.206	1.90	27.8	0.004	6.00	193.0
$(45,0.3)$	λ_{10}	0.041	2.02	28.4	0.001	6.86	197.4
	λ_{30}	0.193	1.84	28.4	0.004	6.09	197.4
	λ_{50}	0.211	2.08	28.4	0.008	6.87	197.4
	λ_{70}	0.215	1.99	28.4	0.008	6.93	197.4
$(45,0.5)$	λ_{10}	0.029	1.84	29.8	0.001	6.31	204.2
	λ_{30}	0.163	1.90	29.8	0.002	6.75	204.2
	λ_{50}	0.361	1.94	29.8	0.008	6.57	204.2
	λ_{70}	0.259	1.97	29.8	0.005	6.86	204.2
$(60,0.1)$	λ_{10}	0.051	2.93	31.6	0.001	7.84	218.4
	λ_{30}	0.193	2.93	31.6	0.004	8.16	218.4
	λ_{50}	0.252	2.69	31.6	0.005	7.79	218.4
	λ_{70}	0.137	2.78	31.6	0.008	8.10	218.4
$(60,0.3)$	λ_{10}	0.044	2.67	28.4	0.002	7.11	194.0
	λ_{30}	0.313	2.77	28.4	0.005	7.55	194.0
	λ_{50}	0.247	2.66	28.4	0.004	7.32	194.0
	λ_{70}	0.273	2.58	28.4	0.005	7.11	194.0
$(60,0.5)$	λ_{10}	0.035	2.83	30.0	0.001	8.66	205.2
	λ_{30}	0.204	2.56	30.0	0.004	7.90	205.2
	λ_{50}	0.306	2.97	30.0	0.006	8.38	205.2
	λ_{70}	0.279	2.76	30.0	0.007	8.74	205.2

Table 3 Performance of the approximation scheme for assortment optimization problems with display location effects.

percent optimality gap of the solutions obtained by the approximation scheme, where the maximum is computed over the 50 test problems in a parameter configuration. The second column shows the average CPU seconds for the approximation scheme. The third column shows the average number of candidate assortments generated by the approximation scheme. Our results indicate that our approximation scheme can find near-optimal solutions for the Revenue-Utility problem rather quickly. For the larger test problems with $K=60$ locations, using the smallest accuracy parameter of $\rho=0.1$, the average CPU seconds is 7.88 and the approximation scheme finds solutions with optimality gaps no larger than 0.008% by using an average of 205.2 candidate assortments. The practical performance of the approximation scheme is substantially better than its theoretical guarantee. Using an accuracy parameter of $\rho=1$, which corresponds to a performance guarantee of 1/2, the optimality gap of the solutions that we get is at most 0.54%. In Appendix E, we give additional computational experiments on a large number of randomly generated instances of the pricing problem with discrete price menus.

8.2 Computational Results on the Expedia Dataset

We give computational experiments on a pricing problem by using a publicly available dataset provided by Expedia (Kaggle 2016). These experiments have two goals. First, we want to demonstrate that we may significantly improve customer-centric objectives, such as expected utility and purchase probability, by incurring a relatively small degradation in the firm-centric objective of expected revenue. Second, we want to demonstrate that shifting the weight from expected revenue to expected utility may significantly change the optimal decisions in the Revenue-Utility problem.

Dataset and Experimental Setup: We have the results of search queries for hotels on Expedia. A search query is akin to an assortment, corresponding to a set of hotels displayed to a customer. In the dataset, each row corresponds to a hotel displayed in a specific search query. Each column corresponds to the attributes of the hotel displayed in the search query and the purchase decision of the customer making the search query. We preprocessed the dataset to remove the columns and rows with missing entries. After preprocessing, we end up with 14 columns and 595,965 rows, corresponding to 34,461 search queries. In the first column, we have a unique user identifier giving the customer making the search query. Using this column, we can access the assortment of hotels displayed to a customer in a search query. In the second column, we have an indicator for whether the customer making the search query booked the hotel. The remaining 12 columns give the attributes of the hotel, including its star rating, review score and displayed price. In the dataset, the customer booked a hotel in 16.92% of the search queries. The average number of hotels displayed in a search query is 17.24, ranging between 4 and 37. In Appendix F, we discuss the details of our preprocessing approach and the 14 columns that we use.

Based on the single dataset provided by Expedia, to enrich our experimental setup, we use bootstrapping to generate multiple datasets. In each dataset, we vary the fraction of search queries that result in a booking. In particular, each dataset that we generate includes 10,000 search queries. Using φ to denote the fraction of search queries that result in a booking, we sample $\varphi \times 10,000$ search queries within the original Expedia search queries that resulted in a booking. Also, among the original Expedia search queries that did not result in a booking, we sample $(1-\varphi) \times 10,000$ search queries. Putting the two samples together, we get 10,000 search queries in our dataset. We vary φ over $\{0.1, 0.3, 0.5\}$ to obtain three datasets. In this way, in each of the three datasets, we have 10,000 search queries. For each search query, we have access to the attributes of the hotels displayed to the customer, along with the hotel, if any, the customer booked. Note that one of the attributes of each hotel is its displayed price. For economy of space, we report our results on one bootstrapped dataset for each value for $\varphi \in \{0.1, 0.3, 0.5\}$, but we repeated our computational experiments on 10 bootstrapped datasets. Our results do not change significantly from one bootstrapped dataset to another. In Appendix G, we give the results on all bootstrapped datasets.

We split each dataset into testing and training portions, containing, respectively, 64% and 36% of the search queries. We use maximum likelihood estimation on the training data to fit a multinomial logit model to each dataset. The preference weight of hotel i is of the form $e^{\beta^0 + \sum_{\ell=1}^{12} \beta^\ell z_i^\ell}$, where the vector $(z_i^1,\ldots,z_i^{12})$ captures the values of the last 12 columns giving the attributes of hotel i and the vector $(\beta^0,\beta^1,\ldots,\beta^{12})$ corresponds to the coefficients that we estimate. We normalize the preference weight of the no-purchase option as $v_0=1$. Since one of the attributes correspond to the displayed price of the hotel, once we fit a multinomial logit model, we can use the expression $e^{\beta^0+\sum_{\ell=1}^{12}\beta^\ell z_i^\ell}$ to compute the preference weight of hotel i at any price. Using $\mathcal{K}=\{1,\ldots,K\}$ with $K=100$ to index the possible price levels that we can charge for a hotel and $\{r_k : k\in\mathcal{K}\}$ to denote the set of 100 possible prices equally spaced over the interval $[0,R]$, let v_{ik} be the preference weight for hotel i when we charge price level k. To choose R, assuming that the prices take values over a continuum, we use the approach in Li and Huh (2011) to compute the prices that maximize only the expected revenue from a customer making a search query. We choose R as four times the maximum price charged for any hotel. Given the assortment of hotels displayed in a search query, building on the discussion in Section 7.3, we can formulate an LP to compute the prices to charge for these hotels to maximize a linear combination of the expected revenue and the expected utility. Letting $\mathcal{N}$ be the assortment of hotels displayed in a search query, we use $\boldsymbol{x}=\{x_{ik} : i\in\mathcal{N},\ k\in\mathcal{K}\}\in\{0,1\}^{|\mathcal{N}|\times K}$ to capture the prices charged for the hotels, where $x_{ik}=1$ if and only if we charge price level k for hotel i. To find the prices to charge for the hotels displayed in a search query without changing the assortment of hotels displayed, we simply modify the definition of the set of feasible solutions in Section 7.3 as $\mathcal{F}=\left\{\boldsymbol{x}\in\{0,1\}^{|\mathcal{N}|\times K} : \sum_{k\in\mathcal{K}} x_{ik}=1 \quad \forall\, i\in\mathcal{N}\right\}$.

Results: We compute two sets of prices for each search query in the testing data. First, we compute the prices that maximize the expected revenue from the assortment of hotels displayed in a search query. We call these prices revenue-maximizing prices. To compute such prices, we set $\lambda=0$ in the Revenue-Utility problem. In this way, we get the largest achievable expected revenue. Second, we compute the prices that maximize the expected utility, while losing at most 1% of the expected revenue when compared with the largest achievable one. We call these prices utility-focused prices. By Lemma 5.1, the expected utility in an optimal solution to the Revenue-Utility problem is increasing in λ, so we compute the utility-focused prices by using bisection over λ.

We give our computational results in Table 4. In the top portion of the table, we compare the performance of the revenue-maximizing and utility-focused prices. Each row corresponds to a bootstrapped dataset with a different value for φ. The first column gives the value of φ. The remainder of the table has three blocks, each containing four columns. In the first block, we compare the probability that a customer makes a purchase under the revenue-maximizing and utility-focused prices. Letting $\boldsymbol{x}^t_{\mathsf{Rev}}$ capture the revenue-maximizing prices for the assortment of hotels displayed in search query t of the testing data, under these prices, the customer makes a purchase with probability $\phi^t_{\mathsf{Rev}}=V(\boldsymbol{x}^t_{\mathsf{Rev}})/(1+V(\boldsymbol{x}^t_{\mathsf{Rev}}))$. Similarly, letting $\boldsymbol{x}^t_{\mathsf{Util}}$ capture the utility-focused prices

1% Allowed Decrease in Expected Revenue

φ	% Increase in Purch. Prb. Avg.	Min.	Max.	Std.	% Increase in Exp. Utility Avg.	Min.	Max.	Std.	% Change in Price Avg.	Min.	Max.	Std.
0.1	12.49	8.73	14.33	0.85	14.45	14.30	15.05	0.08	11.98	8.94	13.41	0.66
0.3	13.36	10.48	14.33	0.55	14.39	14.29	14.69	0.04	12.65	10.37	13.40	0.42
0.5	14.05	11.90	14.34	0.21	14.35	14.28	14.53	0.03	13.18	11.51	13.39	0.16

2% Allowed Decrease in Expected Revenue

φ	% Increase in Purch. Prb. Avg.	Min.	Max.	Std.	% Increase in Exp. Utility Avg.	Min.	Max.	Std.	% Change in Price Avg.	Min.	Max.	Std.
0.1	17.65	12.12	20.39	1.25	20.56	20.34	21.48	0.12	16.65	12.55	18.57	0.89
0.3	18.93	14.73	20.40	0.81	20.46	20.33	20.92	0.06	17.55	14.54	18.57	0.56
0.5	19.95	16.71	20.40	0.31	20.41	20.33	20.61	0.04	18.26	15.98	18.57	0.21

3% Allowed Decrease in Expected Revenue

φ	% Increase in Purch. Prb. Avg.	Min.	Max.	Std.	% Increase in Exp. Utility Avg.	Min.	Max.	Std.	% Change in Price Avg.	Min.	Max.	Std.
0.1	21.57	14.64	24.98	1.57	25.27	24.98	26.50	0.17	20.12	15.29	22.32	1.04
0.3	23.18	17.93	25.00	1.02	25.13	24.97	25.79	0.08	21.17	17.67	22.33	0.66
0.5	24.47	20.42	25.03	0.39	25.05	24.97	25.40	0.04	21.99	19.37	22.34	0.25

Table 4 Percent changes in the purchase probabilities, expected utilities and prices as a function of the maximum allowed percent change in the expected revenue.

for the assortment of hotels displayed in search query t of the testing data, under these prices a customer makes a purchase with probability $\phi^t_{\mathsf{Util}} = V(\boldsymbol{x}^t_{\mathsf{Util}})/(1+V(\boldsymbol{x}^t_{\mathsf{Util}}))$. Thus, $100 \times \frac{\phi^t_{\mathsf{Util}} - \phi^t_{\mathsf{Rev}}}{\phi^t_{\mathsf{Util}}}$ captures the percent improvement in the purchase probability if we are willing to incur 1% loss in the expected revenue in favor of increasing the expected utility. For the 3,600 search queries in the testing data, the four columns in the first block, respectively, give the average, minimum, maximum and standard deviation of the data $\{100 \times \frac{\phi^t_{\mathsf{Util}} - \phi^t_{\mathsf{Rev}}}{\phi^t_{\mathsf{Util}}} : t = 1, \ldots, 3{,}600\}$. In the second block, we compare the expected utility of a customer under the revenue-maximizing and utility-focused prices. In particular, $100 \times \frac{\mathsf{Util}(\boldsymbol{x}^t_{\mathsf{Util}}) - \mathsf{Util}(\boldsymbol{x}^t_{\mathsf{Rev}})}{\mathsf{Util}(\boldsymbol{x}^t_{\mathsf{Util}})}$ gives the percent improvement in the expected utility of the customer in search query t if we are willing to incur 1% loss in the expected revenue. The second block focuses on the data $\{100 \times \frac{\mathsf{Util}(\boldsymbol{x}^t_{\mathsf{Util}}) - \mathsf{Util}(\boldsymbol{x}^t_{\mathsf{Rev}})}{\mathsf{Util}(\boldsymbol{x}^t_{\mathsf{Util}})} : t = 1, \ldots, 3{,}600\}$ and gives for this data the same four statistics provided in the first block. In the third block, we compare the revenue-maximizing and utility-focused prices. Letting $p^t_{i,\mathsf{Rev}}$ and $p^t_{i,\mathsf{Util}}$, respectively, be the revenue-maximizing and utility-focused prices for hotel i in search query t, $100 \times \frac{|p^t_{i,\mathsf{Rev}} - p^t_{i,\mathsf{Util}}|}{p^t_{i,\mathsf{Util}}}$ corresponds to the percent change in the price charged for the hotel if we are willing to incur 1% loss in the expected revenue in favor of increasing the expected utility. Letting $\mathcal{N}^t$ be the assortment of hotels displayed in search query t, the third block focuses on the data $\{100 \times \frac{|p^t_{i,\mathsf{Rev}} - p^t_{i,\mathsf{Util}}|}{p^t_{i,\mathsf{Util}}} : t = 1, \ldots, 3{,}600,\ i \in \mathcal{N}^t\}$ and gives for this data the same four statistics provided in the first block.

Our results indicate that we can significantly improve the purchase probability and the expected utility of a customer by incurring a rather small loss in the expected revenue. If we are willing to incur 1% loss in the expected revenue, then we can increase the purchase probability of a customer by 13.30% on average. As the value of φ increases, so that a larger fraction of the customers in

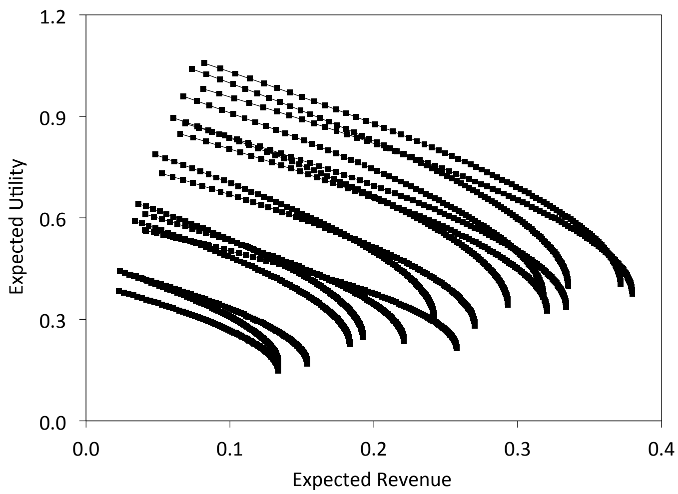

Figure 7 Efficient frontiers for the assortments of hotels displayed to customers in 16 different search queries.

the training data make a booking, the improvements in the purchase probabilities become more noticeable. Intuitively speaking, as a larger fraction of the customers make a booking, the estimated mean utilities of the hotels, relative to the mean utility of the no-purchase option, end up being larger, in which case, we can obtain more noticeable improvements in the purchase probabilities by adjusting the prices charged for the hotels. If we are willing to incur 1% loss in the expected revenue, then we can increase the expected utility of a customer by 14.39% on average. These observations are consistent with the discussion presented in Section 5. Specifically, the boundary of the efficient frontier on its right-side may be rather steep, in which case, it may be possible to increase the expected utility substantially without incurring a huge loss in the expected revenue. In Figure 7, we show the efficient frontier for the assortments of hotels displayed to customers in 16 different search queries in the testing data. The revenue-maximizing prices correspond to the right-most point on each efficient frontier. Lastly, the average percent deviation between the revenue-maximizing and utility-focused prices is 12.60%. Therefore, if we shift our focus from maximizing only the expected revenue to increasing the expected utility with a relatively small loss in the expected revenue, then the prices that we need to charge can change significantly.

We repeated our computational experiments by allowing at most 2% and 3% loss in the expected revenue when we compute the utility-focused prices. Our results are shown in the middle and bottom portions of Table 4. Our observations remain qualitatively unchanged.

9. Conclusion

We examined assortment optimization problems that consider the tradeoff between the expected revenue and the expected utility under the multinomial logit model and totally unimodular constraints. In future work, one can investigate customer-centric objectives, other than the expected utility, that are amenable to efficient optimization. We gave a problem instance where the number of candidate assortments increases exponentially with the number of constraints, but this instance does not establish the computational complexity of our assortment optimization problem. One can focus on characterizing the computational complexity of the problem. Alternatively, one can explore polynomial-time algorithms to solve our assortment optimization problem. Our work on the Expedia dataset involved a pricing application. One can carry out empirical work on other datasets to check the extent of improvements in the expected utility, especially when we have more stringent constraints on the feasible assortments that we can offer. Lastly, one can work with other choice models, such as the nested and paired combinatorial logit models.

Acknowledgements: We thank the Area Editor, Professor Omar Besbes, the associate editor, and three anonymous referees whose comments substantially improved both our technical results and computational experiments. This work was supported in part by NSF grant CMMI 1825406. The first author is supported by the NSF Graduate Research Fellowship.

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.

Adelman, D., A. J. Mersereau. 2013. Dynamic capacity allocation to customers who remember past service. *Management Science* **59**(3) 592–612.

Aflaki, S., I. Popescu. 2014. Managing retention in service relationships. *Management Science* **60**(2) 415–433.

Anupindi, R., S. Gupta, M. A. Venkataramanan. 2015. Managing variety on the retail shelf: Using household scanner panel data to rationalize assortments. N. Agrawal, S. A. Smith, eds., *Retail Supply Chain Management: Quantitative Models and Empirical Studies*. Springer US, Boston, MA, 265–291.

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.

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.

Calmon, A. P., F. D. Ciocan, G. Romero. 2018. Revenue management with repeated customer interactions. Tech. rep., University of Toronto, Toronto, ON.

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

Charnes, A., W. W. Cooper. 1962. Programming with linear fractional functionals. *Naval Research Logistics Quarterly* **9**(3–4) 181–186.

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

Chen, R., H. Jiang. 2017. Assortment optimization with position effects under the nested logit model. Tech. rep., Tsinghua University, Beijing, China.

Chen, Y., V. F. Farias. 2012. What's on the table: Revenue management and the welfare gap in the US airline industry. Tech. rep., Massachusetts Institute of Technology, Cambridge, MA.

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.

Derakhshan, M., N. Golrezai, V. Manshadi, V. Mirrokni. 2019. Product ranking on online platforms. Tech. rep., Massachusetts Institution of Technology, Cambridge, MA.

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.

Duch-Brown, N., L. Grzybowski, A. Romahn, F. Verboven. 2017. The impact of online sales on consumers and firms: Evidence from consumer electronics. *International Journal of Industrial Organization* **52** 30–62.

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., A. Li, V. A. Truong, X. Wang. 2019. Approximation algorithms for product framing and pricing. *Operations Research* (to appear).

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.

Gans, N. 2002. Customer loyalty and supplier quality competition. *Management Science* **48**(2) 207–221.

Gupta, S., S. Kalmanje, K. M. Kockelman. 2006. Road pricing simulations: Traffic, land use and welfare impacts for Austin, Texas. *Transportation Planning and Technology* **29**(1) 1–23.

Ho, T.-H., Y.-H. Park, Y.-P. Zhou. 2006. Incorporating satisfaction into customer value analysis: Optimal investment in lifetime value. *Marketing Science* **25**(3) 260–277.

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

Kaggle. 2016. Kaggle: Expedia hotel recommendations. URL `https://www.kaggle.com/c/expedia-hotel-recommendations`.

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

Kling, C. L., C. J. Thomson. 1996. The implications of model specification for welfare estimation in nested logit models. *American Journal of Agricultural Economics* **78**(1) 103–114.

L'Ecuyer, P., P. Maille, N. E. Stier-Moses, B. Tuffin. 2017. Revenue-maximizing rankings for online platforms with quality-sensitive consumers. *Operations Research* **65**(2) 408–423.

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.

Puig-Junoy, J., M. Saez, E. Martinez-Garcia. 1998. Why do patients prefer hospital emergency visits? A nested multinomial logit analysis for patient-initiated contacts. *Health Care Management Science* **1**(1) 39–52.

Quan, T. W., K. R. Williams. 2018. Product variety, across-market demand heterogeneity, and the value of online retail. *The RAND Journal of Economics* **49**(4) 877–913.

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.

Shi, P. 2019. Optimal priority-based allocation mechanisms. Tech. rep., University of Southern California, Los Angeles, CA.

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.

Train, Kenneth. 2015. Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples. *Journal of Choice Modelling* **16** 15–22.

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

Vanderbei, R. J. 2014. *Linear Programming: Foundations and Extensions*. Springer, New York, NY.

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: Bounding the Number of Candidate Assortments

We establish an upper bound of $(1+n)(n+m)^m/m! \le (m+n)^{1+m}$ 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 $(1+n)(n+m)^m/m!$. 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 $\mathcal{F}$ is nonempty, 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

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

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. As a function of γ, 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 $\mathsf{Candidate\ LP}$, 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 $\mathsf{Candidate\ LP}$ 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 $\mathsf{Candidate\ LP}$, 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 $\mathsf{Candidate\ LP}$.

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 $\mathsf{Candidate\ LP}$ 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 $\mathsf{Candidate\ LP}$

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 LP 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. ∎

Appendix B: Exponentially Many Candidate Assortments

Recall that as a function of γ, we use $\mathsf{LP}(\gamma)$ to denote the optimal objective value of the Candidate LP. In this section, we give an instance of a variant of the Candidate LP for which the number of breakpoints of $\mathsf{LP}(\cdot)$ increases exponentially with the number of constraints. The instance that we give is adapted from Zadeh (1973), where the author gives an instance of the min-cost network flow problem for which the number of iterations of the simplex method to reach the optimal solution increases exponentially with the number of nodes. His instance immediately implies that if we view the optimal cost in the min-cost network flow problem as a function of the *total supply* in the network, then the number of breakpoints of this function increases exponentially with the number of nodes. With some additional work, we will be able to convert the problem instance in Zadeh (1973) into an instance of a variant of the Candidate LP for which the number of breakpoints of LP$(\cdot)$ increases exponentially with the number of constraints. Specifically, for some generic index set $\mathcal{N}=\{1,\ldots,n\}$ and fixed parameters $\{r_{ij}:i,j\in\mathcal{N}\}$, $\{s_i:i\in\mathcal{N}\}$, $\{d_j:j\in\mathcal{N}\}$ and $\{u_{ij}:i,j\in\mathcal{N}\}$, using the decision variables $\boldsymbol{x}=\{x_{ij}:i\in\mathcal{N},\ j\in\mathcal{N}\}$, we consider the LP

$$f(\gamma) \;=\; \max_{\boldsymbol{x}\in\mathbb{R}_+^{n^2}} \left\{ \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}(r_{ij}-\gamma)\,x_{ij} \;\middle|\; \sum_{j\in\mathcal{N}} x_{ij} \;\geq\; s_i \;\forall\, i\in\mathcal{N},\;\; \sum_{i\in\mathcal{N}} x_{ij} \;\geq\; d_j \;\forall\, j\in\mathcal{N},\; x_{ij}\leq u_{ij}\;\forall\, i,j\in\mathcal{N}\right\}. \qquad \text{(Parametric Cost LP)}$$

The above constraint matrix is the constraint matrix of an assignment problem, which is totally unimodular (Corollary 2.9 in Chapter III.1, Nemhauser and Wolsey 1988).

The Parametric Cost LP is a variant of the Candidate LP, where the products are indexed by $(i,j)\in\mathcal{N}^2$ and the decision variable for product (i,j) has an upper bound of u_{ij} instead of one. An upper bound other than one on a decision variable in the Candidate LP simply captures the case with multiple products sharing the same revenues, preference weights and constraint matrix coefficients. In this case, the Candidate LP focuses on how many of these products to offer. Noting that $f(\gamma)$ is the optimal objective value of the Parametric Cost LP as a function of γ, using the fact that there are n^2 decision variables and $2n$ constraints in this LP, by precisely the same argument in the proof of Theorem 4.2, we can show that the number of breakpoints of $f(\cdot)$ is bounded by $\min\{1+n^2\max_{(i,j)\in\mathcal{N}}u_{ij},\ 2+2\,n^2\max_{(i,j)\in\mathcal{N}^2}r_{ij}\,u_{ij},(2n+n^2)^{2n+1}\}$. In the rest of our discussion, we will use the two sequences of integers $\{\overline{p}_i:i\in\mathcal{N}\}$ and $\{\overline{q}_i:i\in\mathcal{N}\}$ defined as

$$\overline{p}_i=\begin{cases}1 & \text{if } i=1\\ 4 & \text{if } i=2\\ 2^{i-3}+2^{i-1} & \text{if } i\geq 3,\end{cases} \qquad \overline{q}_i=\begin{cases}2 & \text{if } i=1\\ 2 & \text{if } i=2\\ 2^{i-3}+2^{i-1} & \text{if } i\geq 3.\end{cases}$$

Using these sequences of integers, we give specific values for the parameters in the Parametric Cost LP that ensure that the number of breakpoints of $f(\cdot)$ increases exponentially with n.

Note that the parameters in the Parametric Cost LP are $\{r_{ij}:i,j\in\mathcal{N}\}$, $\{s_i:i\in\mathcal{N}\}$, $\{d_j:j\in\mathcal{N}\}$ and $\{u_{ij}:i,j\in\mathcal{N}\}$. Using the sequences of integers $\{\overline{p}_i:i\in\mathcal{N}\}$ and $\{\overline{q}_i:i\in\mathcal{N}\}$ defined in the previous paragraph, we set the values of these parameters as follows. For all $i,j\in\mathcal{N}$, we set $r_{ij}=2^{\max\{i,j\}-1}-1$. Moreover, for all $i,j\in\mathcal{N}$, we set

$$u_{ij}=\begin{cases}1 & \text{if } i=j=1\\ \min\{\overline{p}_i,\overline{q}_j\} & \text{if } i\neq j\\ 0 & \text{otherwise.}\end{cases}$$

Lastly, using $\{u_{ij}:i,j\in\mathcal{N}\}$ as defined above, we set $s_i=\sum_{j\in\mathcal{N}}u_{ij}-\overline{p}_i$ for all $i\in\mathcal{N}$ and $d_j=\sum_{i\in\mathcal{N}}u_{ij}-\overline{q}_j$ for all $j\in\mathcal{N}$.

To argue that the number of breakpoints of $f(\cdot)$ increases exponentially with n, we first relate the Parametric Cost LP to the problem instance in Zadeh (1973). Consider the LP

$$g(\delta)\;=\;\max_{\boldsymbol{x}\in\mathbb{R}_+^{n^2}}\left\{\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}r_{ij}\,x_{ij}\;\middle|\;\sum_{j\in\mathcal{N}}x_{ij}\;\geq\;s_i\;\forall\,i\in\mathcal{N},\;\;\sum_{i\in\mathcal{N}}x_{ij}\;\geq\;d_j\;\forall\,j\in\mathcal{N},\right.$$
$$\left.x_{ij}\leq u_{ij}\;\forall\,i,j\in\mathcal{N},\;\;\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}x_{ij}=\delta\right\}.\qquad\text{(Parametric Flow LP)}$$

The sum of all decision variables in the above LP is restricted to be δ. Naturally, we must have $\delta\geq\sum_{i\in\mathcal{N}}s_i$ and $\delta\geq\sum_{j\in\mathcal{N}}d_j$ for the LP to be feasible. With the way we set the parameters $\{r_{ij}:i,j\in\mathcal{N}\}$, $\{s_i:i\in\mathcal{N}\}$, $\{d_j:j\in\mathcal{N}\}$ and $\{u_{ij}:i,j\in\mathcal{N}\}$ in the previous paragraph, it is straightforward to check that the Parametric Flow LP is equivalent to the instance of the min-cost network flow problem provided by Zadeh (1973). Noting that $g(\delta)$ is the optimal objective value of the Parametric Flow LP as a function of δ, Zadeh (1973) shows that the number of breakpoints of $g(\cdot)$ increases exponentially with n. Building on this result, we adapt an approach from Carstensen (1983) to show that the number of breakpoints of $f(\cdot)$ also increases exponentially with n. Our approach is based on arguing that $f(\cdot)$ is the convex conjugate of $g(\cdot)$. In particular, using $\mathcal{X}\subseteq\mathbb{R}_+^{n^2}$ to denote the set of feasible solutions to the Parametric Cost LP, we have

$$\begin{aligned}f(\gamma)\;&=\;\max_{\boldsymbol{x}\in\mathcal{X}}\left\{\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}(r_{ij}-\gamma)\,x_{ij}\right\}\;=\;\max_{\delta\in\mathbb{R}}\max_{\boldsymbol{x}\in\mathcal{X}}\left\{\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}(r_{ij}-\gamma)\,x_{ij}\;:\;\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}x_{ij}=\delta\right\}\\&=\;\max_{\delta\in\mathbb{R}}\left\{\max_{\boldsymbol{x}\in\mathcal{X}}\left\{\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}r_{ij}\,x_{ij}\;:\;\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}x_{ij}=\delta\right\}-\gamma\,\delta\right\}\;=\;\max_{\delta\in\mathbb{R}}\Big\{g(\delta)-\gamma\,\delta\Big\},\end{aligned}$$

where the third equality follows from the fact that any $\boldsymbol{x}\in\mathcal{X}$ satisfying $\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}x_{ij}=\delta$ is a feasible solution to the Parametric Flow LP.

By the above chain of equalities, $f(\cdot)$ is the convex conjugate of $g(\cdot)$, so by Lemma 1 in Carstensen (1983), the numbers of breakpoints of $f(\cdot)$ and $g(\cdot)$ differ by one. Thus, since the number of breakpoints of $g(\cdot)$ increases exponentially with n, the number of breakpoints of $f(\cdot)$ must also increase exponentially with n, as long as we set the parameters $\{r_{ij}:i,j\in\mathcal{N}\}$, $\{s_i:i\in\mathcal{N}\}$, $\{d_j:j\in\mathcal{N}\}$ and $\{u_{ij}:i,j\in\mathcal{N}\}$ as discussed earlier in this section. In Figure EC.1, we plot the number of breakpoints of $f(\cdot)$ as a function of n, which indeed increases exponentially with n.

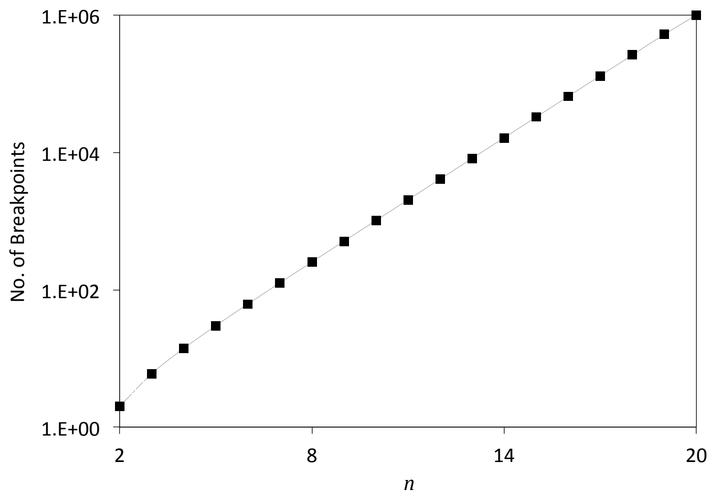

Figure EC.1 Number of breakpoints of $f(\cdot)$ as a function of n; note that *the vertical axis is in* log-*scale*.

Appendix C: Proof of Lemma 5.1

Using the fact that $\boldsymbol{x}^*_\lambda$ is an optimal solution to the Revenue-Utility problem as a function of λ, we have the inequalities

$$\begin{aligned}\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r})+\lambda\,\mathsf{Util}(\boldsymbol{x}^*_\lambda) &\geq \mathsf{Rev}(\boldsymbol{x}^*_\beta;\boldsymbol{r})+\lambda\,\mathsf{Util}(\boldsymbol{x}^*_\beta)\\ \mathsf{Rev}(\boldsymbol{x}^*_\beta;\boldsymbol{r})+\beta\,\mathsf{Util}(\boldsymbol{x}^*_\beta) &\geq \mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r})+\beta\,\mathsf{Util}(\boldsymbol{x}^*_\lambda).\end{aligned}$$

Subtracting the two inequalities above, we have $(\lambda-\beta)\,\mathsf{Util}(\boldsymbol{x}^*_\lambda)\geq(\lambda-\beta)\,\mathsf{Util}(\boldsymbol{x}^*_\beta)$. Noting that $\lambda>\beta$, the last inequality yields $\mathsf{Util}(\boldsymbol{x}^*_\lambda)\geq\mathsf{Util}(\boldsymbol{x}^*_\beta)$, establishing the second property in the lemma. In this case, using the fact that $\mathsf{Util}(\boldsymbol{x}^*_\lambda)\geq\mathsf{Util}(\boldsymbol{x}^*_\beta)$, the second inequality above yields $\mathsf{Rev}(\boldsymbol{x}^*_\beta;\boldsymbol{r})\geq\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r})+\beta\,(\mathsf{Util}(\boldsymbol{x}^*_\lambda)-\mathsf{Util}(\boldsymbol{x}^*_\beta))\geq\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r})$, establishing the first property in the lemma. Lastly, write the two inequalities above as $\lambda\,(\mathsf{Util}(\boldsymbol{x}^*_\lambda)-\mathsf{Util}(\boldsymbol{x}^*_\beta))\geq\mathsf{Rev}(\boldsymbol{x}^*_\beta;\boldsymbol{r})-\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r})$ and $\mathsf{Rev}(\boldsymbol{x}^*_\beta;\boldsymbol{r})-\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r})\geq\beta\,(\mathsf{Util}(\boldsymbol{x}^*_\lambda)-\mathsf{Util}(\boldsymbol{x}^*_\beta))$. Multiplying the first inequality by $1/\lambda$ and the second inequality by $1/\beta$, we obtain

$$-\frac{1}{\lambda}\Big\{\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r})-\mathsf{Rev}(\boldsymbol{x}^*_\beta;\boldsymbol{r})\Big\}\;\leq\;\mathsf{Util}(\boldsymbol{x}^*_\lambda)-\mathsf{Util}(\boldsymbol{x}^*_\beta)\;\leq\;-\frac{1}{\beta}\Big\{\mathsf{Rev}(\boldsymbol{x}^*_\lambda;\boldsymbol{r})-\mathsf{Rev}(\boldsymbol{x}^*_\beta;\boldsymbol{r})\Big\},$$

yielding the third property in the lemma. Therefore, all of the three properties that are given in the lemma hold. ■

Appendix D: Eliminating Candidate Assortments

In the next lemma, we use both upper and lower bounds on $V(\boldsymbol{x}^*_\lambda)$ to eliminate candidate assortments, in which case, Lemma 5.2 follows as a corollary to this result.

Lemma D.1 *Letting $\boldsymbol{x}^*_\lambda$ be an optimal solution to the* Revenue-Utility *problem, $\bar{t}$ be an upper bound on $V(\boldsymbol{x}^*_\lambda)$, $\underline{t}$ be a lower bound on $V(\boldsymbol{x}^*_\lambda)$,*

$$\overline{\gamma}_\lambda = \max_{\boldsymbol{x}\in\mathcal{F}} \left\{ \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) - \frac{\lambda\,(1+\bar{t})}{1+V(\boldsymbol{x})} \right\} \quad \text{and} \quad \underline{\gamma}_\lambda = \max_{\boldsymbol{x}\in\mathcal{F}} \left\{ \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) - \frac{\lambda\,(1+\underline{t})}{1+V(\boldsymbol{x})} \right\},$$

the collection of assortments $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \overline{\gamma}_\lambda \le \gamma \le \underline{\gamma}_\lambda\}$ includes an optimal solution to the Revenue-Utility *problem.*

Proof: Fix $\lambda \ge 0$. Let t^* be an optimal solution to the outer maximization problem in the Parametric problem. By the discussion in the proof of Theorem 4.1, letting $\gamma^* = \max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t^*)\,\boldsymbol{e})$, the solution $\boldsymbol{x}_{\mathsf{LP}}(\gamma^* - \lambda\,(1+t^*))$ is optimal to the problem $\max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t^*)\,\boldsymbol{e})$, in which case, by Theorem 3.2, the solution $\boldsymbol{x}_{\mathsf{LP}}(\gamma^* - \lambda\,(1+t^*))$ is optimal to the Revenue-Utility problem as well. Therefore, if we can show that $\overline{\gamma}_\lambda \le \gamma^* - \lambda\,(1+t^*) \le \underline{\gamma}_\lambda$, then it follows that $\boldsymbol{x}_{\mathsf{LP}}(\gamma^* - \lambda\,(1+t^*)) \in \{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \overline{\gamma}_\lambda \le \gamma \le \underline{\gamma}_\lambda\}$, so the desired result holds. By the definition of γ^*, we have the chain of equalities

$$\begin{aligned}
\gamma^* \;&=\; \max_{\boldsymbol{x}\in\mathcal{F}} \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}+\lambda\,(1+t^*)\,\boldsymbol{e}) \;\stackrel{(a)}{=}\; \max_{\boldsymbol{x}\in\mathcal{F}} \left\{ \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) + \lambda\,(1+t^*)\frac{V(\boldsymbol{x})}{1+V(\boldsymbol{x})} \right\} \\
&=\; \max_{\boldsymbol{x}\in\mathcal{F}} \left\{ \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) + \lambda\,(1+t^*)\left(1 - \frac{1}{1+V(\boldsymbol{x})}\right) \right\},
\end{aligned}$$

where (a) follows from the identity that we give at the beginning of the proof of Lemma 3.1. Arranging the terms, the above chain of equalities yields

$$\gamma^* - \lambda\,(1+t^*) \;=\; \max_{\boldsymbol{x}\in\mathcal{F}} \left\{ \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) - \frac{\lambda\,(1+t^*)}{1+V(\boldsymbol{x})} \right\}.$$

By Theorem 3.2, we can choose t^* as $V(\boldsymbol{x}^*_\lambda)$, in which case, we have $\underline{t} \le V(\boldsymbol{x}^*_\lambda) = t^* \le \bar{t}$. Therefore, by the above equality, it follows that

$$\overline{\gamma}_\lambda \;=\; \max_{\boldsymbol{x}\in\mathcal{F}} \left\{ \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) - \frac{\lambda\,(1+\bar{t})}{1+V(\boldsymbol{x})} \right\} \le\; \gamma^* - \lambda\,(1+t^*) \;\le \max_{\boldsymbol{x}\in\mathcal{F}} \left\{ \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) - \frac{\lambda\,(1+\underline{t})}{1+V(\boldsymbol{x})} \right\} \;=\; \underline{\gamma}_\lambda.$$

■

Lemma 5.2 follows as a corollary to the above lemma.

Proof of Lemma 5.2:

We claim that if $\gamma \ge \overline{\gamma}_\lambda$, then we have $\lambda \ge \overline{\lambda}(\gamma)$. If we can show this claim, then it follows that $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma \ge \overline{\gamma}_\lambda\} \subseteq \{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \lambda \ge \overline{\lambda}(\gamma)\}$. In this case, noting that the collection of assortments $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \gamma \ge \overline{\gamma}_\lambda\}$ includes an optimal solution to the Revenue-Utility problem by Lemma D.1, it must be the case that the collection of assortments $\{\boldsymbol{x}_{\mathsf{LP}}(\gamma) : \lambda \ge \overline{\lambda}(\gamma)\}$ includes an optimal solution

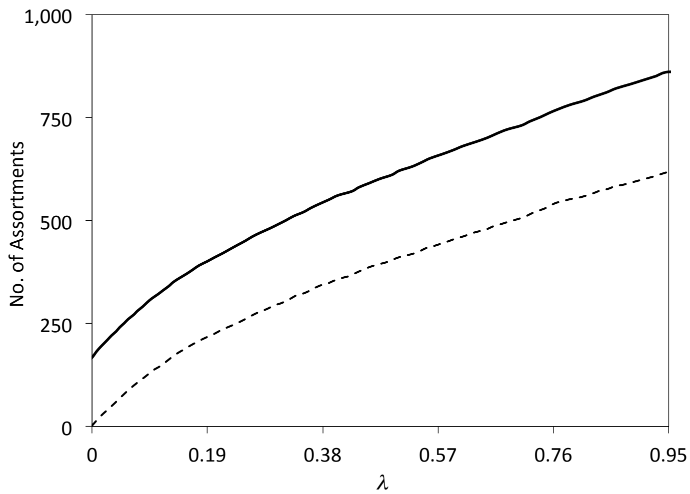

Figure EC.2 Number of candidate assortments left after eliminating the unnecessary assortments using Lemmas 5.2 and D.1.

to the $\mathsf{Revenue\text{-}Utility}$ problem as well. Thus, we proceed to showing that the claim holds. If $\gamma \geq \overline{\gamma}_\lambda$, then we have

$$\gamma \;\geq\; \overline{\gamma}_\lambda \;=\; \max_{\boldsymbol{x}\in\mathcal{F}} \left\{ \mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) - \frac{\lambda\,(1+\bar{t})}{1+V(\boldsymbol{x})} \right\} \;\geq\; \mathsf{Rev}(\boldsymbol{x}_{\mathsf{LP}}(\gamma);\boldsymbol{r}) - \frac{\lambda\,(1+\bar{t})}{1+V(\boldsymbol{x}_{\mathsf{LP}}(\gamma))}.$$

Focusing on the first and last terms in the above chain of inequalities and arranging the terms, we get $\lambda \geq \frac{1+V(\boldsymbol{x}_{\mathsf{LP}}(\gamma))}{1+\bar{t}} \times \Big(\mathsf{Rev}(\boldsymbol{x}_{\mathsf{LP}}(\gamma);\boldsymbol{r}) - \gamma\Big) \;=\; \overline{\lambda}(\gamma)$, which establishes the claim. ∎

Computing $\overline{\gamma}_\lambda$ and $\underline{\gamma}_\lambda$ in Lemma D.1 requires solving the problems $\max_{\boldsymbol{x}\in\mathcal{F}}\{\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) - \frac{\lambda\,(1+\bar{t})}{1+V(\boldsymbol{x})}\}$ and $\max_{\boldsymbol{x}\in\mathcal{F}}\{\mathsf{Rev}(\boldsymbol{x};\boldsymbol{r}) - \frac{\lambda\,(1+\underline{t})}{1+V(\boldsymbol{x})}\}$. In contrast, computing $\overline{\lambda}(\gamma)$ in Lemma 5.2 does not require solving an optimization problem. Thus, eliminating candidate assortments by using Lemma D.1 is computationally more intensive. In Figure EC.2, we focus on the problem instance in Figure 2. For each $\lambda \in [0, 0.95]$, we compare the number of candidate assortments left after eliminating the unnecessary ones by using Lemmas 5.2 and D.1. The solid data series show the number of candidate assortments left when we use Lemma 5.2 to eliminate the unnecessary ones. This data series is identical to the one in Figure 4. The dotted data series show the number of candidate assortments left when we use Lemma D.1 to eliminate the unnecessary ones. Tracing the proof of Lemma 5.2, any candidate assortment that is eliminated by using Lemma 5.2 must be eliminated by using Lemma D.1. In that sense, Lemma D.1 is stronger than Lemma 5.2. Therefore, we indeed expect the dotted data series to lie below the solid one. In Figure EC.2, we observe that we can eliminate a significantly larger number of candidate assortments by using Lemma D.1, though at the expense of additional computational effort. When λ is close to zero, we can use Lemma D.1 to eliminate all but a handful of the candidate assortments.

Appendix E: Computational Results for Pricing with Discrete Price Menus

We focus on pricing problems with discrete price menus, as discussed in Section 7.3. 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 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 get 12 parameter configurations. For each parameter configuration, we generate 50 test problems by using the approach described above. 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. Lastly, we test the performance of our approximation scheme.

Parametric LP and Efficient Frontier: We present our computational results in Table EC.1. The layout of this table is identical to that of Table 2. For the smaller test problems with $K = 20$ possible price levels, we can generate the candidate assortments in about ten seconds, whereas the corresponding computational effort for the larger test problems with $K = 80$ possible price levels is less than four minutes. Given that there are more than 6,000 candidate assortments for the largest test problems, such computational effort corresponds to about 0.03 seconds to generate each candidate assortment. Similar to our results for assortment optimization problems with display location effects, once we generate the candidate assortments, we can quickly construct the efficient frontier. For each of our test problems, we can construct the efficient frontier in 0.003 seconds. Lastly, as in the test problems with display location effects, the number of 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 discrete price menus. For the test problems with 20, 40, 60, and 80 price levels, the average values of $L/(nK)$ are, respectively, 0.93, 0.90, 0.89, and 0.85. These results are consistent with Theorem 4.2.

Approximation Scheme: Table EC.2 shows the performance of our approximation scheme. The layout of this table is identical to that of Table 3. Our approximation scheme obtains solutions

Param. (K, p_0)	# Cand. Assr.	# Opt. Assr.	CPU Secs.	Param. (K, p_0)	# Cand. Assr.	# Opt. Assr.	CPU Secs.
$(20, 0.1)$	1877.38	1271.66	12.22	$(60, 0.1)$	5398.14	2876.58	92.65
$(20, 0.3)$	1838.96	1456.40	12.15	$(60, 0.3)$	5373.28	2915.42	93.93
$(20, 0.5)$	1886.98	1568.26	12.53	$(60, 0.5)$	5375.42	2746.22	93.58
$(40, 0.1)$	3726.78	2276.22	47.99	$(80, 0.1)$	6974.04	3324.90	228.38
$(40, 0.3)$	3496.50	2291.74	45.17	$(80, 0.3)$	6690.46	3149.60	221.75
$(40, 0.5)$	3574.12	2271.92	46.27	$(80, 0.5)$	6838.40	2956.66	225.42

Table EC.1 The number of candidate assortments, number of optimal assortments and CPU seconds for pricing problems with discrete price menus.

		$\rho = 1$			$\rho = 0.1$		
Param. (K, p_0)	λ	Max. % Gap	CPU Secs.	# Cand. Assr.	Max. % Gap	CPU Secs.	# Cand. Assr.
$(20, 0.1)$	λ_{10}	0.093	0.70	17.0	0.005	10.26	111.4
	λ_{30}	0.328	0.69	17.0	0.008	10.19	111.4
	λ_{50}	0.011	0.70	17.0	0.007	10.57	111.4
	λ_{70}	0.013	0.76	17.0	0.004	10.92	111.4
$(20, 0.3)$	λ_{10}	0.038	0.58	17.0	0.001	10.27	111.6
	λ_{30}	0.261	0.59	17.0	0.007	10.22	111.6
	λ_{50}	0.167	0.61	17.0	0.010	10.34	111.6
	λ_{70}	0.018	0.63	17.0	0.003	10.35	111.6
$(20, 0.5)$	λ_{10}	0.003	0.54	17.2	0.000	9.73	110.8
	λ_{30}	0.100	0.58	17.2	0.002	9.66	110.8
	λ_{50}	0.169	0.58	17.2	0.005	9.80	110.8
	λ_{70}	0.029	0.59	17.2	0.003	9.84	110.8
$(40, 0.1)$	λ_{10}	0.042	1.92	21.0	0.001	24.88	138.6
	λ_{30}	0.202	1.95	21.0	0.005	25.49	138.6
	λ_{50}	0.262	2.03	21.0	0.007	25.95	138.6
	λ_{70}	0.013	2.13	21.0	0.006	27.08	138.6
$(40, 0.3)$	λ_{10}	0.014	1.81	21.0	0.000	25.19	137.6
	λ_{30}	0.022	1.84	21.0	0.002	25.31	137.6
	λ_{50}	0.319	1.86	21.0	0.009	25.64	137.6
	λ_{70}	0.012	1.90	21.0	0.006	26.00	137.6
$(40, 0.5)$	λ_{10}	0.007	1.81	20.2	0.000	26.47	138.6
	λ_{30}	0.022	1.94	20.2	0.002	27.19	138.6
	λ_{50}	0.231	1.98	20.2	0.004	27.26	138.6
	λ_{70}	0.030	2.05	20.2	0.003	27.75	138.6

		$\rho = 1$			$\rho = 0.1$		
Param. (K, p_0)	λ	Max. % Gap	CPU Secs.	# Cand. Assr.	Max. % Gap	CPU Secs.	# Cand. Assr.
$(60, 0.1)$	λ_{10}	0.011	3.80	24.0	0.000	46.73	165.0
	λ_{30}	0.064	3.95	24.0	0.002	48.08	165.0
	λ_{50}	0.276	3.91	24.0	0.005	48.14	165.0
	λ_{70}	0.016	3.96	24.0	0.008	48.49	165.0
$(60, 0.3)$	λ_{10}	0.008	3.72	24.0	0.000	48.03	164.0
	λ_{30}	0.054	3.81	24.0	0.002	48.54	164.0
	λ_{50}	0.295	3.89	24.0	0.006	49.48	164.0
	λ_{70}	0.012	4.01	24.0	0.006	49.86	164.0
$(60, 0.5)$	λ_{10}	0.005	3.81	24.0	0.000	47.92	163.2
	λ_{30}	0.008	3.99	24.0	0.002	48.97	163.2
	λ_{50}	0.220	4.08	24.0	0.005	49.89	163.2
	λ_{70}	0.023	4.21	24.0	0.002	50.40	163.2
$(80, 0.1)$	λ_{10}	0.008	3.03	27.6	0.000	56.56	186.6
	λ_{30}	0.079	2.88	27.6	0.002	56.78	186.6
	λ_{50}	0.111	2.92	27.6	0.005	57.36	186.6
	λ_{70}	0.024	2.82	27.6	0.008	58.58	186.6
$(80, 0.3)$	λ_{10}	0.002	2.80	27.8	0.000	59.40	188.0
	λ_{30}	0.090	2.86	27.8	0.002	60.98	188.0
	λ_{50}	0.277	2.78	27.8	0.008	61.11	188.0
	λ_{70}	0.013	2.70	27.8	0.005	60.88	188.0
$(80, 0.5)$	λ_{10}	0.003	2.66	27.2	0.000	60.59	186.4
	λ_{30}	0.010	2.78	27.2	0.001	60.47	186.4
	λ_{50}	0.229	2.71	27.2	0.005	60.69	186.4
	λ_{70}	0.036	2.54	27.2	0.001	60.01	186.4

Table EC.2 Performance of the approximation scheme for pricing problems with discrete price menus.

with remarkably small optimality gaps by using substantially smaller numbers of candidate assortments. For the larger test problems with $K = 80$ price levels, using an accuracy parameter of $\rho = 0.1$, the approximation scheme obtains solutions with optimality gaps of at most 0.008% in about a minute by using 186.4 candidate assortments on average. With an accuracy parameter of $\rho = 1$, the largest optimality gap is 0.328%, with computational effort under four seconds.

Appendix F: Preprocessing the Expedia Dataset

We preprocessed the dataset from Expedia as follows. The full dataset provided by Expedia has about ten million rows and 54 columns. We used three criteria to drop some of the rows and columns in the dataset. First, in some search queries, the price is displayed as the total amount over the whole length of the stay, whereas in some others, the price is displayed as the amount per night. We are not able to tell with certainty which approach is used in each search query. Thus, we focused only on the search queries for a single night stay and dropped the others. Second, we dropped the columns for which the entries were missing for more than 25% of the rows. For the remaining columns, we dropped the search queries for which the entries were missing in one of the remaining columns. Third, some of the rows in the dataset included exorbitantly high or low values. Specifically, we dropped all search queries which had a column entry that falls outside the 0.5^{th} and 99.5^{th} percentile band of all entries in the corresponding column.

Once we preprocess the dataset, we have 595,965 rows corresponding to 34,561 search queries and 14 columns. We discuss the first two columns in the main text. The remaining 12 columns give the star rating and the average review score for the hotel, an indicator for whether the hotel is a chain, two location scores, the average past price of the hotel, the displayed price, an indicator for whether the price is a promotion, the number of days until the day of stay, the number of adults and children for the stay, and an indicator for whether the stay is over a weekend.

Appendix G: Computational Results for All Bootstrapped Datasets

In Table EC.3, we give a summary of our results for all of the 10 bootstrapped datasets. In the top, middle and bottom portions of the table, we compute the utility-focused prices by allowing, respectively, at most 1%, 2% and 3% loss in the expected revenue. Each of the top, middle and bottom portions of the table has three blocks, each containing six columns. Each block corresponds to a value of $\varphi \in \{0.1, 0.3, 0.5\}$, capturing the fraction of search queries that result in a booking in the bootstrapped dataset. In each block, there are six columns. The first two columns give the average and standard deviation of the percent improvement in the purchase probabilities when we use the utility-focused prices instead of revenue-maximizing prices. The next two columns give the average and standard deviation of the percent improvement in the expected utilities when we, once again, use the utility-focused prices instead of revenue-maximizing prices. The last two columns give the average and standard deviation of the percent gap between the revenue-maximizing and utility-focused prices. A row of the table corresponds to a different set of three bootstrapped datasets. Each of these three bootstrapped datasets is obtained by using a value of $\varphi \in \{0.1, 0.3, 0.5\}$. The first rows in the top, middle and bottom portions of the table correspond to the three datasets that we work with in Section 8.2. Therefore, the statistics that we report in these rows match the ones in Table 4. The results in Table EC.3 indicate that our observations do not change significantly from one bootstrapped dataset to the next.

Online Appendix References

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

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

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

Zadeh, N. 1973. A bad network problem for the simplex method and other minimum cost flow algorithms. *Mathematical Programming* **5**(1) 255–266.

1% Allowed Decrease in Expected Revenue

$\varphi = 0.1$						$\varphi = 0.3$						$\varphi = 0.5$					
% Inc. in Prch. Prob.		% Inc. in Exp. Util.		% Chg. in Price		% Inc. in Prch. Prob.		% Inc. in Exp. Util.		% Chg. in Price		% Inc. Prch. Prob.		% Inc. Exp. Util.		% Chg. in Price	
Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
12.49	0.85	14.45	0.08	11.98	0.66	13.36	0.55	14.39	0.04	12.65	0.42	14.05	0.21	14.35	0.03	13.18	0.16
12.48	0.84	14.45	0.08	11.96	0.66	13.33	0.55	14.39	0.04	12.63	0.43	14.05	0.19	14.35	0.03	13.18	0.15
12.48	0.84	14.45	0.08	11.96	0.66	13.35	0.55	14.38	0.04	12.64	0.43	14.05	0.20	14.35	0.03	13.18	0.15
12.47	0.86	14.45	0.08	11.96	0.68	13.35	0.51	14.38	0.04	12.65	0.40	14.05	0.19	14.35	0.03	13.18	0.15
12.46	0.83	14.45	0.07	11.95	0.65	13.36	0.56	14.39	0.04	12.65	0.43	14.05	0.19	14.35	0.03	13.18	0.15
12.47	0.83	14.45	0.07	11.96	0.65	13.34	0.54	14.39	0.04	12.64	0.42	14.05	0.18	14.35	0.03	13.18	0.14
12.50	0.85	14.45	0.08	11.98	0.67	13.36	0.55	14.39	0.04	12.65	0.43	14.03	0.18	14.35	0.03	13.16	0.14
12.50	0.83	14.45	0.08	11.98	0.65	13.36	0.55	14.38	0.04	12.65	0.43	14.05	0.20	14.35	0.03	13.18	0.15
12.45	0.85	14.45	0.08	11.94	0.67	13.34	0.55	14.39	0.04	12.63	0.43	14.06	0.18	14.35	0.03	13.19	0.14
12.42	0.85	14.45	0.08	11.92	0.67	13.33	0.54	14.39	0.04	12.63	0.42	14.05	0.19	14.35	0.03	13.18	0.15

2% Allowed Decrease in Expected Revenue

$\varphi = 0.1$						$\varphi = 0.3$						$\varphi = 0.5$					
% Inc. in Prch. Prob.		% Inc. in Exp. Util.		% Chg. in Price		% Inc. in Prch. Prob.		% Inc. in Exp. Util.		% Chg. in Price		% Inc. Prch. Prob.		% Inc. Exp. Util.		% Chg. in Price	
Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
17.65	1.25	20.56	0.12	16.65	0.89	18.93	0.81	20.46	0.06	17.55	0.56	19.95	0.31	20.41	0.04	18.26	0.21
17.63	1.23	20.57	0.12	16.64	0.88	18.89	0.81	20.46	0.06	17.53	0.57	19.96	0.28	20.41	0.04	18.26	0.19
17.63	1.25	20.57	0.12	16.64	0.88	18.92	0.82	20.46	0.06	17.55	0.57	19.95	0.30	20.41	0.04	18.26	0.20
17.61	1.27	20.57	0.12	16.63	0.90	18.92	0.76	20.46	0.06	17.55	0.53	19.96	0.29	20.41	0.04	18.26	0.20
17.61	1.22	20.57	0.12	16.62	0.87	18.93	0.83	20.46	0.06	17.56	0.58	19.95	0.28	20.41	0.04	18.26	0.19
17.62	1.23	20.57	0.12	16.63	0.87	18.91	0.79	20.46	0.06	17.54	0.55	19.96	0.27	20.41	0.03	18.26	0.18
17.66	1.26	20.56	0.12	16.66	0.89	18.93	0.82	20.46	0.06	17.56	0.57	19.92	0.27	20.41	0.04	18.24	0.19
17.66	1.22	20.56	0.12	16.66	0.87	18.93	0.82	20.46	0.06	17.55	0.57	19.95	0.29	20.41	0.04	18.26	0.20
17.59	1.25	20.57	0.12	16.61	0.89	18.90	0.81	20.46	0.06	17.53	0.57	19.97	0.27	20.41	0.03	18.27	0.18
17.55	1.26	20.57	0.13	16.58	0.89	18.89	0.80	20.46	0.06	17.53	0.56	19.95	0.28	20.41	0.04	18.26	0.20

3% Allowed Decrease in Expected Revenue

$\varphi = 0.1$						$\varphi = 0.3$						$\varphi = 0.5$					
% Inc. in Prch. Prob.		% Inc. in Exp. Util.		% Chg. in Price		% Inc. in Prch. Prob.		% Inc. in Exp. Util.		% Chg. in Price		% Inc. Prch. Prob.		% Inc. Exp. Util.		% Chg. in Price	
Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
21.57	1.57	25.27	0.17	20.12	1.04	23.18	1.02	25.13	0.08	21.17	0.66	24.47	0.39	25.05	0.04	21.99	0.25
21.55	1.56	25.28	0.16	20.10	1.02	23.13	1.03	25.14	0.08	21.14	0.66	24.48	0.36	25.05	0.04	22.00	0.22
21.54	1.57	25.28	0.16	20.10	1.03	23.17	1.03	25.13	0.08	21.16	0.67	24.47	0.37	25.05	0.04	21.99	0.24
21.52	1.60	25.28	0.17	20.08	1.06	23.17	0.96	25.13	0.07	21.16	0.62	24.48	0.37	25.05	0.04	22.00	0.23
21.52	1.54	25.28	0.16	20.08	1.02	23.19	1.05	25.13	0.08	21.17	0.67	24.47	0.36	25.05	0.04	21.99	0.22
21.53	1.55	25.28	0.16	20.09	1.02	23.15	1.00	25.13	0.08	21.15	0.65	24.48	0.34	25.05	0.04	22.00	0.21
21.58	1.59	25.27	0.16	20.12	1.05	23.19	1.03	25.13	0.08	21.17	0.66	24.44	0.34	25.06	0.04	21.97	0.22
21.58	1.54	25.27	0.16	20.12	1.01	23.18	1.03	25.13	0.08	21.16	0.67	24.47	0.37	25.05	0.04	21.99	0.23
21.49	1.58	25.28	0.17	20.06	1.04	23.14	1.03	25.14	0.08	21.14	0.66	24.49	0.33	25.05	0.04	22.00	0.21
21.44	1.59	25.29	0.17	20.03	1.05	23.13	1.01	25.14	0.08	21.14	0.65	24.47	0.36	25.05	0.04	21.99	0.23

Table EC.3 Computational results on 10 bootstrapped datasets.