Tractable Models and Algorithms for Assortment Planning with Product Costs

Sumit Kunnumkal* • Victor Martínez-de-Albéniz†

Submitted: January 15, 2016

Abstract

Assortment planning under a logit demand model is a difficult problem when there are product-specific fixed costs. We develop a new continuous relaxation of the problem that is based on the parametrization of the problem on the total assortment attractiveness. This relaxation provides an upper bound and a heuristic integer solution, for which we develop performance guarantees. We analytically prove that it provides a close-to-optimal solution when products are homogeneous in terms of preference weights. Moreover, our formulation can be easily extended to incorporate additional constraints on the assortment, or multiple customer segments. Finally, we provide numerical experiments that show that our method yields tight upper bounds and performs competitively with respect to other approaches found in the literature.

1 Introduction

The optimization of assortment plans is an important problem for most retailers. In the very broad literature on this topic, product fixed costs have been identified as one of the drivers that make optimization difficult. At the same time, these costs may be very significant. They may arise from store or shelf preparation costs and may be especially high for slow movers even if these have high margins. In situations where product assortments change often (e.g., for stores selling new phones or apparel, see Caro and Martínez-de Albéniz 2015), this is even more important, because such fixed costs need to be recovered in a short time.

In this paper, we consider assortment planning under a multinomial logit (MNL) demand model where products involve fixed costs, together with different margins and attractiveness (preference weights). The objective in our approach is to maximize the expected profit, i.e., the expected sales contribution margin minus fixed assortment costs. The resulting optimization problem is known to be NP-hard (Kunnumkal et al. 2009). To circumvent this difficulty, we develop a tractable relaxation of the assortment optimization problem that is based on a parametric continuous knapsack formulation. We use the total attractiveness of the assortment including the attractiveness of the no-purchase option as a parameter in our relaxation. Our

*Indian School of Business, Hyderabad, 500032, India, email: sumit_kunnumkal@isb.edu

†IESE Business School, University of Navarra, Av. Pearson 21, 08034 Barcelona, Spain, email: valbeniz@iese.edu. V. Martínez-de-Albéniz's research was supported in part by the European Research Council - ref. ERC-2011-StG 283300-REACTOPS and by the Spanish Ministry of Economics and Competitiveness (Ministerio de Economía y Competitividad) - ref. ECO2014-59998-P.

relaxation involves (1) solving a continuous knapsack problem for each value of the total attractiveness parameter, and (2) selecting the best possible value of the parameter. This process generates an upper bound on the optimal expected profit.

With this approach, we obtain several useful results. First, we prove that the upper bound can be obtained efficiently. In particular, we show that the best possible value of the total attractiveness parameter, and hence the upper bound can be obtained in polynomial time, namely $O(n^3)$ where n is the total number of products available.

In addition, we provide an analytical characterization of the gap between optimal expected profit and our upper bound. Specifically, we show that the upper bound obtained by our relaxation is never more than twice that of the optimum. In addition, we construct a family of assortment problems which shows that the worst-case gap of 2 is in fact tight. While similar bounds based on the continuous knapsack formulation of the assortment problem under different choice models and assumptions appear in the assortment literature (Kunnumkal et al. 2009, Davis et al. 2014, Gallego and Topaloglu 2014), to our knowledge, the tightness of the worst-case gap is new. We find that the worst-case gap is achieved when one product is significantly different than the rest in terms of margins and attractiveness. This situation may not occur very frequently in practice especially when we think of the products as being potential substitutes. When the characteristics of the products are more similar, we obtain two sharper bounds: 3/2 without making strong assumptions on the parameter space, and a ratio quite close to one when the number of products is large. An appealing feature of the latter bound is that it is a simple function of only the product attractiveness parameters and is independent of the profit margins and the product fixed costs. The sharper bounds are more likely to be applicable in many practical situations and are thus useful in providing a more accurate characterization of the performance of our method.

Our performance bounds are obtained by comparing the upper bound given by our relaxation with the expected profits achieved by certain candidate assortments. Therefore, our continuous relaxation allows us to generate heuristics solutions that perform well, since if a candidate assortment generates a profit that is within a certain factor of our upper bound, then it also within the same factor of the optimal profit. In our computational study, we find that our relaxation generally obtains bounds that are very close to optimal, with average optimality gaps below 1% and worst-case gaps of a few percentage points. Our heuristic thus provides very competitive performance at a reasonable computational cost.

Finally, by taking a novel dual perspective, we extend the relaxation idea to incorporate additional modelling elements. We show that incorporating general linear constraints on the assortment and multiple customer classes (i.e., a mixture of multinomial logit models) still allows us to calculate the upper bound by minimizing a finite number of functions. It turns out that the number of functions to consider is of order $O\left(n^{D+E}\right)$ where D is the number of customer classes and E the number of constraints. So it is only useful when the number of classes and constraints is small. These results can be extended further when there is a single customer class, in which case we show that the complexity of obtaining the upper bound is $O\Big(n^{3(1+E)}\Big)$. When there are constraints on the assortment, we cannot obtain guarantees on the performance of our heuristic in general. However, we are able to recover the performance guarantee of 2 for some classes of constraints that are common in the assortment literature: (1) cardinality constraints which limit the total number of products offered and (2) product precedence constraints which require that a certain set of products be included in the assortment if a given product is part of the assortment.

Our results thus advance the understanding of assortment planning with product fixed costs. We make three main contributions to the literature. First, we build on Kunnumkal et al. (2009) to obtain a new, tractable upper bound on the optimal expected profit. Second, we show that our heuristic has provable performance guarantees, and our computational study further indicates that the heuristic method is efficient and has very competitive performance. Our analytical results explain the good practical performance of the heuristic to a large extent. And third, our approach can be applied to more difficult problems that include constraints on the assortment and multiple customer classes.

The rest of the paper is organized as follows. Section 2 reviews the related literature. Section 3 formulates the problem and Section 4 develops the continuous relaxation and the heuristic. Section 5 adds constraints and customer classes to the problem. Section 6 shows a numerical study of the performance of our heuristic and Section 7 concludes. Proofs of the analytical results are included in the Appendix.

2 Literature Review

We provide a concise review of the assortment planning literature under variants of the MNL choice model. We refer the reader to Kök et al. (2009) for a more detailed review of the assortment planing literature and Anderson et al. (1992) for a background on discrete choice models.

There is a growing literature on assortment optimization under the MNL model and variations of it. Talluri and van Ryzin (2004) show that the assortment optimization problem under the MNL choice model can be solved efficiently and that the optimal assortment is a *revenue-ordered set*, one consisting of a certain number of the products with the highest margins. The assortment problem becomes difficult in general when we move beyond the MNL choice model. Rusmevichientong et al. (2014) and Miranda Bront et al. (2009) show that the assortment optimization problem is NP-hard for the mixture of multinomial logits (MMNL) model. Davis et al. (2014) show that the assortment optimization problem under the nested logit model is tractable under some conditions on the choice parameters, but is intractable in general. They propose approximation schemes to obtain assortments with provable worst-case guarantees in the cases where the assortment optimization problem is intractable. The papers referred to in this paragraph do not model product fixed costs or constraints on the assortment.

The constrained assortment planning problem has been the focus of recent research. The problem becomes hard even under the MNL model when there is a single constraint that limits the total space consumed by the products in the assortment (Rusmevichientong et al. 2009). Tractable models typically include those with cardinality constraints and precedence constraints. Rusmevichientong et al. (2010) show that the cardinality constrained problem under the MNL model can be solved tractably, while Gallego and Topaloglu (2014) and Feldman and Topaloglu (2014) show similar results for the nested logit model. Davis et al. (2013) show that the assortment optimization problem under the MNL model can be solved efficiently when the constraint matrix is totally unimodular. Again, these papers do not model product fixed costs.

Kunnumkal et al. (2009) consider assortment optimization under the MNL model when there are product fixed costs. They show that the resulting assortment optimization problem becomes NP-hard. So the authors focus on approximation schemes to obtain assortments with worst-case guarantees on the expected profit. The authors propose a 2-approximation algorithm and a polynomial time approximation scheme. The first algorithm obtains an assortment which is

guaranteed to obtain at least 50% of the optimal expected profit, while the second one obtains assortments with improved guarantees but at the expense of increased computational effort.

Although there is a somewhat limited literature on the assortment problem with product fixed costs, the model itself is relevant in a number of contexts. As mentioned, constrained assortment optimization is difficult in general and one approach to obtain tractable models is to dualize the difficult constraints by associating Lagrange multipliers with them. The resulting relaxation has precisely the same form as the fixed costs problem if we interpret the Lagrange multipliers as the product fixed costs. Indeed, Feldman and Topaloglu (2015) consider relaxing certain constraints in the assortment optimization problem under the MMNL model by associating Lagrange multipliers with them and the relaxation they obtain is an assortment problem with product fixed costs.

Given the connection to constrained assortment optimization, it is not surprising that the fixed costs model has applications in revenue management as well. The revenue management problem can be viewed as solving a sequence of assortment problems that are linked together by resource constraints. Kunnumkal and Topaloglu (2008) dualize the resource capacity constraints and the sub-problems in their method end up being assortment optimization problems with fixed costs.

The papers closest to ours are Kunnumkal et al. (2009) and Feldman and Topaloglu (2015), but there are some important differences. Kunnumkal et al. (2009) are concerned with approximation schemes that obtain assortments with provable profit guarantees, while our focus is on obtaining tight upper bounds on the optimal expected profit. Now the assortments obtained by Kunnumkal et al. (2009) do provide an implicit bound on the optimal profit, since if a candidate assortment is guaranteed to obtain at least a fraction θ of the optimal profit, then we can conclude that the optimal profit is no more than $1/\theta$ of the profit obtained by that assortment. However, since the guarantees are from a worst-case perspective, they can be quite conservative in practice and may not provide a good indication of the extent of suboptimality. Therefore, it is important to obtain tight upper bounds on the optimal profit to be able to better assess the suboptimality of candidate assortments and to benchmark their performance. In that sense, our work complements Kunnumkal et al. (2009). Moreover, we extend our approach to handle general constraints on the assortment. Our method remains tractable provided the number of constraints is not too large and we recover the performance guarantees from the unconstrained case for certain types of assortment constraints.

Feldman and Topaloglu (2015) propose a Lagrangian relaxation approach to obtain an upper bound on the optimal revenue of the assortment problem under the MMNL model. The relaxed problem ends up being an assortment problem with product fixed costs, which is still difficult to solve. So they obtain an upper bound using a grid-based approximation and the quality of their bound and the computational work involved both depend on the density of the grid. Our method can be viewed as a version of the grid-based approximation of Feldman and Topaloglu (2015) that works with an infinitely dense grid. However, the computational work required to obtain our bound does not depend on the density of the grid and is instead polynomial in the number of products. We also note that the upper bound obtained by the grid-based approximation lacks the theoretical guarantees of our method. Subsequent to a working version of our paper, Kunnumkal (2015) adapted our method to refine the Lagrangian relaxation approach of Feldman and Topaloglu (2015) to the assortment problem under the MMNL model.

Finally we note that there is a body of work on assortment planning with inventory costs; see for example van Ryzin and Mahajan (1999). This line of work is primarily concerned with

inventory levels of the different products that balance the trade-off between stock-outs and inventory carrying costs, and an underlying assumption is that customers make their choice without considering the availability of the products. We do not rely on this assumption and in our model customers choose after observing the assortment.

3 Problem Formulation

We have a set of n products and we have to decide which of them to include in the assortment. We let $\mathcal{J} = \{1, 2, \ldots, n\}$ denote the set of products and for product $j \in \mathcal{J}$, we let p_j denote its profit margin and c_j the fixed cost of including it in the assortment. We let $x_j \in \{0, 1\}$ indicate if product j is included in the assortment. Given an assortment, customers choose among the offered products according to the multinomial logit (MNL) model. The MNL model associates a preference weight v_j with product j and a preference weight v_0 associated with not making a purchase. The probability that a customer purchases product j is given by $v_j x_j/(v_0+\sum_{k\in\mathcal{J}} v_k x_k)$ and the no-purchase probability is given by $v_0/(v_0 + \sum_{k\in\mathcal{J}} v_k x_k)$. Letting

$$Z(x) = \frac{\sum_{j\in\mathcal{J}} p_j v_j x_j}{v_0 + \sum_{j\in\mathcal{J}} v_j x_j} - \sum_{j\in\mathcal{J}} c_j x_j \tag{1}$$

denote the expected profit associated with offering the assortment $x = \{x_j \,|\, \forall j\}$, the optimal expected profits can be obtained by solving the problem

$$(OPT) \qquad Z^{OPT} = \max \quad Z(x) \\ \text{s.t.} \quad x_j \in \{0, 1\}.$$

The optimal assortment can be obtained efficiently in certain special cases. For example, OPT is tractable if the preference weights for all the products are identical, or if the no-purchase preference weight $v_0 = 0$. It is also tractable if the fixed costs are identical for all the products: $c_j = c$ for all j. However Kunnumkal et al. (2009) show that problem OPT is NP-hard in general.

Although OPT is a nonlinear integer program, it can be reformulated as the linear mixed-integer program

$$Z^{OPT} = \max \quad \sum_{j\in\mathcal{J}} p_j u_j - \sum_{j\in\mathcal{J}} c_j x_j \tag{2}$$

$$\text{s.t.} \quad \tfrac{v_0}{v_j} u_j \le u_0 \,\forall j \tag{3}$$

$$u_j \le \tfrac{v_j}{v_0+v_j} x_j \,\forall j \tag{4}$$

$$\textstyle\sum_{j\in\mathcal{J}} u_j + u_0 = 1 \tag{5}$$

$$u_j \ge 0, x_j \in \{0, 1\} \tag{6}$$

by using the transformation $u_j = v_j x_j/(v_0 + \sum_{k\in\mathcal{J}} v_k x_k)$; see for example Topaloglu (2013). While the linear mixed-integer program is still intractable, it is in a form that can be readily handled by most commercial optimization software. The mixed-integer programming formulation tends to be more useful when we benchmark the performance of different approximation methods against the optimal expected profit.

4 An Upper Bound Based on a Parametric Linear Program

In this section, we describe a tractable method to obtain an upper bound on the optimal expected profit. If we let $t = \frac{1}{v_0+\sum_{j\in\mathcal{J}} v_j x_j}$, then $Z(x) = \sum_{j\in\mathcal{J}}(p_j v_j t - c_j)x_j = \sum_{j\in\mathcal{J}} \rho_j(t)x_j$, where

$$\rho_j(t) = p_j v_j t - c_j. \tag{7}$$

Therefore, we can write OPT equivalently as

$$Z^{OPT} = \max_{t\in[t_{min},t_{max}]} \Gamma^b(t) \tag{8}$$

where $V_k = \sum_{j=1}^{k} v_j$, $t_{min} = \frac{1}{V_n+v_0}$, $t_{max} = \frac{1}{\min_j\{v_j\}+v_0}$ and

$$\begin{aligned} \Gamma^b(t) = \quad & \max \sum_{j\in\mathcal{J}} \rho_j(t)x_j \\ & \text{s.t. } \textstyle\sum_{j\in\mathcal{J}} v_j x_j \le \frac{1}{t} - v_0 \\ & x_j \in \{0,1\}. \end{aligned}$$

Here we note that even though we have replaced the constraint $\sum_{j\in\mathcal{J}} v_j x_j = \frac{1}{t} - v_0$ with $\sum_{j\in\mathcal{J}} v_j x_j \le \frac{1}{t} - v_0$, the formulation remains valid since the constraint will be satisfied as an equality at a value of t that maximizes $\Gamma^b(t)$. Computing $\Gamma^b(t)$ involves solving a binary knapsack problem, which is again intractable.

Since we are interested in obtaining a tractable upper bound on Z^{OPT}, we consider the continuous relaxation of the binary knapsack. In doing so, we restrict our attention to the products contained in the set

$$\mathcal{J}(t) = \Big\{ j | v_j \le \frac{1}{t} - v_0 \text{ and } \rho_j(t) > 0 \Big\}. \tag{9}$$

This is because, if $v_j > \frac{1}{t} - v_0$, then product j can never be part of any feasible solution to the binary knapsack. On the other hand, if $\rho_j(t) \le 0$, then product j can be excluded from an optimal solution to the binary knapsack. Therefore, if $j \notin \mathcal{J}(t)$ it cannot be part of an optimal solution to the binary knapsack. Consequently, we can restrict attention to the products in $\mathcal{J}(t)$ when working with the continuous relaxation of the binary knapsack

$$\begin{aligned} \Gamma^f(t) = \quad & \max \sum_{j\in\mathcal{J}(t)} \rho_j(t)x_j && (10) \\ & \text{s.t. } \textstyle\sum_{j\in\mathcal{J}(t)} v_j x_j \le \frac{1}{t} - v_0 && (11) \\ & 0 \le x_j \le 1. && (12) \end{aligned}$$

Since $\Gamma^b(t) \le \Gamma^f(t)$,

$$Z^{UB} = \max_{t\in[t_{min},t_{max}]} \Gamma^f(t) \tag{13}$$

gives us an upper bound on the optimal expected profit.

Lemma 1. $Z^{OPT} \le Z^{UB}$.

While it is easy to see that Z^{UB} is an upper bound on Z^{OPT}, it is not immediately clear whether the maximization in (13) can be carried out in a tractable manner. It is also not clear how well Z^{UB} approximates Z^{OPT}. We explore these questions in the following sections. We note that Kunnumkal et al. (2009) also use the parametric formulation $\Gamma^b(t)$ of the assortment problem. However, as mentioned, their focus is on obtaining candidate assortments with performance guarantees on the expected profit.

4.1 Tractability

Problem (10)-(12) is a continuous knapsack problem and is tractable. However, its optimal solution depends on the parameter t since the objective function coefficients and the knapsack size are functions of t. Therefore, a potential difficulty in obtaining Z^{UB} is that $\Gamma^f(t)$ has to be computed for infinitely many values of t. In this section, we show that it is sufficient to evaluate $\Gamma^f(t)$ at a finite, in fact a polynomial, number of values of t.

We begin with the observation that the optimal solution to a continuous knapsack problem involves filling up the knapsack with items in decreasing order of the profit-to-space ratio until the knapsack is completely filled. In the context of problem (10)-(12), we fill up the knapsack of size $\frac{1}{t} - v_0$ with products in decreasing order of $\frac{\rho_j(t)}{v_j} = p_j t - \frac{c_j}{v_j}$.

Since the profit-to-space ratio depends on the value of t, the order in which the items get placed into the knapsack also depends on the value of t. We bound the number of different orderings that are possible as we vary t. Product k_1 has a higher profit-to-space ratio than product k_2 provided $(p_{k_1} - p_{k_2})t \geq \frac{c_{k_1}}{v_{k_1}} - \frac{c_{k_2}}{v_{k_2}}$. Therefore, we have exactly one critical value

$$\hat{t}_{k_1,k_2} = \frac{c_{k_1}/v_{k_1} - c_{k_2}/v_{k_2}}{p_{k_1} - p_{k_2}}$$

at which the profit-to-space ordering of products k_1 and k_2 changes. Note that if $\hat{t}_{k_1,k_2}$ is smaller than t_{min} or greater than t_{max}, then the profit-to-space ordering of k_1 and k_2 remains the same in the entire range of t of interest. So we find the critical values $\hat{t}_{k_1,k_2}$ for every pair of products k_1 and k_2 and sort these $O(n^2)$ critical values from smallest to largest. This divides the interval $[t_{min}, t_{max}]$ into $O(n^2)$ subintervals. We note that the profit-to-space ordering of the products does not change as t varies within a given subinterval. We conclude that there are $O(n^2)$ possible profit-to-space orderings of the products.

Now consider a particular such subinterval $[\hat{t}_l, \hat{t}_u]$. For simplicity, assume that $\frac{1}{\hat{t}_u} - v_0 \geq v_{max} = \max_j\{v_j\}$ and that $\rho_j(t) > 0$ for all j, so that $\mathcal{J}(t) = \mathcal{J}$ for all $t \in [\hat{t}_l, \hat{t}_u]$. Note that this is not a restrictive assumption since if $\frac{1}{t} - v_0 < v_{max}$, we simply work with a smaller set of products that are admissible given the knapsack size $\frac{1}{t} - v_0$. On the other hand, if $\rho_j(t) \leq 0$ for some j, then we can find the critical value of t at which the profit-to-space ratio of product j becomes equal to zero and analyze the intervals to the left and right of the critical value separately.

Now suppose that $\rho_1(t)/v_1 \geq \ldots \geq \rho_n(t)/v_n > 0$ for all $t \in [\hat{t}_l, \hat{t}_u]$. Since (10)-(12) is a continuous knapsack problem, we simply fill up the knapsack with products starting with product 1 until we use up all the space. Therefore,

$$\Gamma^f(t) = \sum_{j=1}^{\kappa(t)-1} \rho_j(t) + \rho_{\kappa(t)}(t) \left(\frac{\frac{1}{t} - v_0 - V_{\kappa(t)-1}}{v_{\kappa(t)}} \right)$$

where $\kappa(t)$ is the largest index k such that $V_{k-1} = \sum_{j=1}^{k-1} v_j < \frac{1}{t} - v_0$. Note that the index $\kappa(t)$ stays constant as long as $V_{k-1} < \frac{1}{t} - v_0 \leq V_k$. Therefore, the interval $[\hat{t}_l, \hat{t}_u]$ can be further partitioned into $O(n)$ subintervals such that $\kappa(t)$ does not change with t within each subinterval. We note that Kunnumkal et al. (2009) make these observations in developing their approximation algorithms. We build on these observations to next show that problem (13) can be solved in a tractable manner.

Since we have $O(n^2)$ intervals where the profit-to-space ordering of the products does not change and each such interval can be further partitioned into $O(n)$ subintervals where the index $\kappa(t)$ remains constant, the range $[t_{min}, t_{max}]$ can be partitioned into a total of $O(n^3)$ subintervals and problem (13) can be obtained by solving $O(n^3)$ problems of the form $\max_{t\in[l,u]} \Pi_\kappa(t)$ where

$$\Pi_\kappa(t) = \sum_{j=1}^{\kappa-1} \rho_j(t) + \rho_\kappa(t) \left(\frac{\frac{1}{t} - v_0 - V_{\kappa-1}}{v_\kappa} \right) \tag{14}$$

and $V_{\kappa-1} \leq \frac{1}{u} - v_0$ and $\frac{1}{l} - v_0 \leq V_\kappa$. Let

$$\Delta_\kappa = p_\kappa(v_0 + V_{\kappa-1}) - \sum_{j=1}^{\kappa-1} p_j v_j. \tag{15}$$

Lemma 2 below states that the problem $\max_{t\in[l,u]} \Pi_\kappa(t)$ can be solved efficiently, essentially in closed form.

Lemma 2. *Let* $\bar{t}^* = argmax_{t\in[l,u]} \Pi_\kappa(t)$. *If* $\Delta_\kappa \leq 0$, *then* $\bar{t}^* = u$. *Otherwise,* $\bar{t}^* = \max\{l, \min\{t^*, u\}\}$ *where*

$$t^* = \sqrt{\frac{c_\kappa / v_\kappa}{\Delta_\kappa}}. \tag{16}$$

We thus have the following proposition.

Proposition 1. Z^{UB} *can be obtained in a running time of* $O(n^3)$.

To illustrate this result, we describe the intervals and sub-intervals in the following example, see Figure 1. In the example, higher profit margins are associated with higher fixed costs but lower levels of attractiveness (smaller preference weights). It turns out the optimal integer solution is to introduce product 2 (with weight of 3), which results in a profit of 1.8. In contrast, the upper bound is reached at $t = 0.213$ with a value of 1.8238, an optimality gap of 1.32% above the true integer optimum.

To calculate the upper bound, we first compute $\hat{t}_{1,2} = 0.25, \hat{t}_{1,3} = 0.166, \hat{t}_{2,3} = 0.125$. In addition, we note that $\mathcal{J}(t) = \{1, 2, 3\}$ for $t \leq 0.2$, $\mathcal{J}(t) = \{1, 2\}$ for $t \in [0.2, 0.25]$ while $\mathcal{J}(t) = \{1\}$ for $t \in [0.25, 0.333]$. This means, that, given that $t_{max} = \frac{1}{v_0+v_1} = 0.333$ and $t_{min} = \frac{1}{v_0+v_1+v_2+v_3} = 0.1$, we must consider five intervals $[0.1, 0.125], [0.125, 0.166], [0.166, 0.2]$, $[0.2, 0.25]$ and $[0.25, 0.333]$ in computing the upper bound.

1. In the first interval $[0.1, 0.125]$, we have $\mathcal{J}(t) = \{1, 2, 3\}$ and $\rho_3(t)/v_3 \geq \rho_2(t)/v_2 \geq \rho_1(t)/v_1$. In this interval, we have $x_3 = x_2 = 1$ and x_1 varies between 1 and 0 and $\Gamma^f(t)$ is increasing in t.

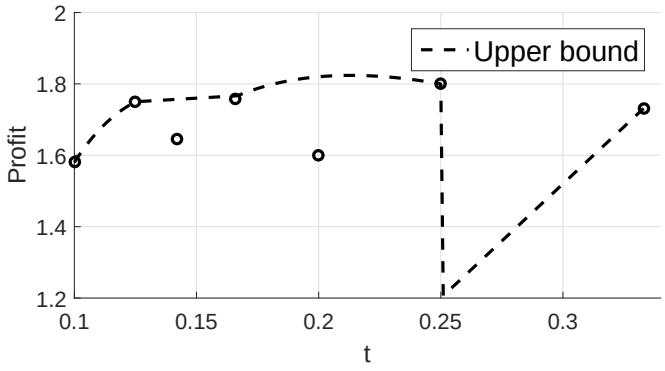

Figure 1: Example with $n = 3$ products. Product characteristics are $v_0 = 1, v_1 = 2, v_2 = 3, v_3 = 4, p_1 = 3.2, p_2 = 2.8, p_3 = 2, c_1 = 0.4, c_2 = 0.3, c_3 = 0$. The curve corresponds to $\Gamma^f(t)$, while the dots correspond to the integer solutions, i.e., all the points $\left(\frac{1}{v_0+\sum_{j=1}^3 v_j x_j}, \frac{\sum_{j=1}^3 p_j v_j x_j}{v_0+\sum_{j=1}^3 v_j x_j} - \sum_{j=1}^3 c_j x_j\right)$ for $x_j \in \{0,1\}$.

2. In the second interval $[0.125, 0.166]$, we still have $\mathcal{J}(t) = \{1,2,3\}$, but $\rho_2(t)/v_2 \geq \rho_3(t)/v_3 \geq \rho_1(t)/v_1$. In this interval $x_2 = 1$, $x_1 = 0$ and x_3 varies between 1 and 1/3 and $\Gamma^f(t)$ is still increasing in t.

3. In the third interval $[0.166, 0.2]$ we have $\mathcal{J}(t) = \{1,2,3\}$ and $\rho_2(t)/v_2 \geq \rho_1(t)/v_1 \geq \rho_3(t)/v_3$. In this interval $x_2 = 1$, $x_3 = 0$ and x_1 varies from 1 to 0.5 and $\Gamma^f(t)$ is increasing in t.

4. In the fourth interval $[0.2, 0.25]$, the profit-to-space ordering of the products remains unchanged but $\mathcal{J}(t) = \{1,2\}$. So we have $x_2 = 1$ in this interval while x_1 varies from 0.5 to 0. $\Gamma^f(t)$ is concave with an interior maximizer at 0.213 (as identified by Lemma 2).

5. Finally, in the last interval $[0.25, 0.333]$, $\mathcal{J}(t) = \{1\}$, which means that in this range the optimal fractional solution stays equal to $x_1 = 1$ and $\Gamma^f(t)$ is increasing.

4.2 Performance Guarantee

In this section, we discuss the tightness of the upper bound Z^{UB}. Kunnumkal et al. (2009) describe an approximation algorithm which obtains an assortment whose expected profit is within a factor of 2 of the optimal value. The same line of analysis here implies that $Z^{UB} \leq 2Z^{OPT}$. We briefly outline the arguments which show the performance bound of 2 and we give an example which shows that the gap of 2 is tight. On the other hand, in our computational experiments, we observe that the gaps between Z^{UB} and Z^{OPT} tend to be much smaller than the theoretical worst-case bound. To explain this, we characterize problem parameter settings where the gaps tend to be small and provide improved performance guarantees in such cases.

4.2.1 A Gap of 2

The analysis in Kunnumkal et al. (2009) implies that $Z^{UB} \leq 2Z^{OPT}$. We summarize the key observation for completeness. By (13) and (8), it suffices to show that $\Gamma^f(t) \leq 2\Gamma^b(t)$. But this follows from the well-known result that the optimal objective function value of the fractional knapsack is within a factor of 2 of that of the binary knapsack; see for example Vazirani (2013).

We next give an example where the gap between Z^{UB} and Z^{OPT} asymptotically approaches 2. We note that it is not a direct extension of the classical knapsack example, since in our setting the objective function coefficients of the products and the knapsack size both depend on the same underlying parameter t.

Consider an assortment problem with two products so that $\mathcal{J} = \{1, 2\}$. Let $v_2 > v_1$ and $p_2 > p_1 > \frac{p_2 v_2}{v_0+v_2}$. Let $c_1 = \frac{v_1}{(v_0+v_2)(v_0+v_1+v_2)}[p_1(v_0+v_2) - p_2 v_2]$ and $c_2 = \frac{v_2}{v_0+v_1+v_2}[p_2 - \frac{p_1 v_1}{v_0+v_1}]$, and note that $c_1, c_2 > 0$. Since $v_2 > v_1$, $t_{min} = \frac{1}{v_0+v_1+v_2}$, $t_{max} = \frac{1}{v_0+v_1}$ and $Z^{UB} = \max_{t\in[t_{min},t_{max}]} \Gamma^f(t)$. It can be verified that $\rho_1(t)/v_1 \geq \rho_2(t)/v_2 > 0$ for all $t \in [t_{min}, t_{max}]$. Therefore when $t = \frac{1}{v_0+v_2}$, the knapsack includes product 1 and a fractional amount of product 2, so that

$$\begin{aligned}\Gamma^f\left(\frac{1}{v_0+v_2}\right) &= \rho_1\left(\frac{1}{v_0+v_2}\right) + \rho_2\left(\frac{1}{v_0+v_2}\right)\left(\frac{v_2-v_1}{v_2}\right) \\ &= Z_{\{1\}} - \frac{p_1 v_1 (v_2-v_1)}{(v_0+v_1)(v_0+v_2)} + \left(1-\frac{v_1}{v_2}\right) Z_{\{2\}},\end{aligned}$$

where we use Z_S to denote the expected profit associated with offering assortment S and the last equality follows from using (7) and rearranging terms. Therefore $Z_{\{1\}} = \rho_1\left(\frac{1}{v_0+v_1}\right)$ denotes the expected profit from offering the assortment consisting of product 1 alone, while $Z_{\{2\}} = \rho_2\left(\frac{1}{v_0+v_2}\right)$ denotes the expected profit from offering the assortment consisting of product 2 alone.

Now set $v_0 = 1$, $v_1 = \epsilon^2$, $v_2 = \epsilon$, $p_1 = 1/\epsilon^2$ and $p_2 = 1/\epsilon^3$, where $0 < \epsilon < 1$. It can be verified that $Z_{\{1\}}$ and $Z_{\{2\}}$ tend to 1 as ϵ approaches 0 and the limit of Z^{OPT} as ϵ approaches 0 is 1. Since $Z^{UB} \leq 2Z^{OPT}$, it follows that the limiting value of Z^{UB} is no more than 2. On the other hand, the limit of $\Gamma^f\left(\frac{1}{v_0+v_2}\right)$ as ϵ approaches 0 is 2. Since $Z^{UB} = \max_{t\in[t_{min},t_{max}]} \Gamma^f(t) \geq \Gamma^f(\frac{1}{v_0+v_2})$, it follows that the limiting value of Z^{UB} as ϵ approaches 0 is 2. Therefore, the gap between Z^{UB} and Z^{OPT} approaches 2 asymptotically.

4.2.2 Performance on Randomly Generated Instances

The example in §4.2.1 requires the preference weights and the profit margins of the products to differ by orders of magnitude and this may not be the case in many situations, especially when we think of the products as being substitutes of each other. So we investigate the performance of the upper bound Z^{UB} on randomly generated test problems.

We generate our test problems in a manner similar to Feldman and Topaloglu (2015). We have $n = 10$ products. We set the preference weight of product j as $v_j = X_j / \sum_{k=1}^{n} X_k$, where X_j is uniformly distributed on $[0, 1]$. We set $v_0 = \frac{\Phi}{1-\Phi}\sum_{j\in\mathcal{J}} v_j$, where $\Phi \in [0, 1]$ is a parameter that we vary in our computational experiments. Note that the no-purchase probability when all the products are offered is Φ. We sample p_j from the uniform distribution on $[0, 2000]$ and sample c_j from the uniform distribution on $[0, \gamma p_j v_j/(v_0 + v_j)]$, where $\gamma \in [0, 1]$ is a second parameter that we vary in our computational experiments. We note that if γ is small, then

the fixed costs are relatively small compared to the profits. On the other hand, if γ is large, then the fixed costs are roughly comparable to the profits. We vary $\Phi \in \{0.75, 0.50, 0.25\}$ and $\gamma \in \{1.00, 0.50, 0.25\}$. For each (Φ, γ) combination, we generate 50 test problems by following the procedure described above.

Table 1 compares the upper bound Z^{UB} with the optimal expected profit Z^{OPT}. In our computational experiments, we obtain Z^{OPT} by solving its linear mixed-integer programming formulation (2)-(6). The first column of Table 1 gives the problem parameters (n, Φ, μ). As mentioned, for each (Φ, γ) pair we generate 50 test problems and the second column of Table 1 gives the average percentage difference between Z^{UB} and Z^{OPT} over the 50 test problems. The third column gives the 5th percentile of the difference, while the fourth column gives the 95th percentile. The last column reports the fraction of instances where Z^{UB} coincides with Z^{OPT}.

We observe that Z^{UB} is remarkably close to Z^{OPT} in our computational experiments. The average percentage difference is at most 0.58% and the 95th percentile of the difference is no more than 3.49%. Moreover, Z^{UB} coincides with Z^{OPT} for at least half of the test problems. We next provide a theoretical basis for these observations.

Problem	% difference between Z^{UB} and Z^{OPT}			% optimal
(n, Φ, γ)	Avg.	5th percentile	95th percentile	
(10, 0.75, 1.00)	0.32	0.00	2.16	64
(10, 0.75, 0.50)	0.16	0.00	0.54	72
(10, 0.75, 0.25)	0.03	0.00	0.18	82
(10, 0.5, 1.00)	0.34	0.00	1.71	60
(10, 0.5, 0.50)	0.18	0.00	1.33	64
(10, 0.5, 0.25)	0.10	0.00	0.59	66
(10, 0.25, 1.00)	0.58	0.00	3.49	60
(10, 0.25, 0.50)	0.16	0.00	1.07	66
(10, 0.25, 0.25)	0.21	0.00	0.87	58

Table 1: Performance gap between Z^{UB} and Z^{OPT} for test problems with 10 products.

4.2.3 A Gap of 3/2

The example in §4.2.1 indicates that the gap between Z^{UB} and Z^{OPT} is essentially 2. On the other hand, our computational experiments in §4.2.2 indicate that the performance of Z^{UB} tends to be much better than the worst-case bound of 2. In this section, we establish conditions for an improved performance guarantee on the upper bound Z^{UB}.

By the discussion in §4.1, it follows that Z^{UB} can be obtained by solving $O(n^3)$ problems of the form $\max_{t\in[l,u]} \Pi_\kappa(t)$ where $V_{\kappa-1} = \sum_{j=1}^{\kappa-1} v_j \leq \frac{1}{u} - v_0$ and $\frac{1}{l} - v_0 \leq V_\kappa = \sum_{j=1}^{\kappa} v_j$. Equivalently, $u \leq \tau_{\kappa-1} = \frac{1}{v_0+V_{\kappa-1}}$ and $l \geq \tau_\kappa = \frac{1}{v_0+V_\kappa}$. So, to bound the gap between Z^{UB} and Z^{OPT}, it suffices to obtain a uniform bound on the gap between $\max_{t\in[l,u]} \Pi_\kappa(t)$ and Z^{OPT}. In the following analysis, we assume that $\frac{1}{u} - v_0 > v_{max}$ and $\mathcal{J}(t) = \mathcal{J}$ for all $t \in [l, u]$. We emphasize that the assumptions are only to reduce the notational burden and that all of our results continue to hold on relaxing them.

Lemma 3. *If* $\Delta_\kappa \leq 0$, *then* $\max_{t\in[l,u]} \Pi_\kappa(t) \leq Z^{OPT}$.

Lemma 4. *Let* $t^* = \ argmax_t \Pi_\kappa(t)$. *If* $\Delta_\kappa > 0$ *and* $t^* \geq \tau_{\kappa-1}$ *or* $t^* \leq \tau_\kappa$, *then* $\max_{t\in[l,u]} \Pi_\kappa(t) \leq Z^{OPT}$.

Lemma 5. *Let* $t^* = \ argmax_t \Pi_\kappa(t)$. *If* $\Delta_\kappa > 0$, $t^* \in (\tau_\kappa, \tau_{\kappa-1})$ *and* $t^* \leq \frac{1}{v_0+v_\kappa}$, *then* $\max_{t\in[l,u]} \Pi_\kappa(t) \leq \frac{3}{2} Z^{OPT}$.

Note that the only case not covered by Lemmas 3-5 is when $\Delta_\kappa > 0$ and $\frac{1}{v_0+v_\kappa} < t^* < \tau_{\kappa-1}$. That is, $V_{\kappa-1} < \frac{1}{t^*} - v_0 < v_\kappa$. We note that for this situation to occur the preference weight of product κ has to be greater than the sum of the preference weights of products $\{1, \ldots, \kappa-1\}$. This is unlikely to be the case if the preference weights of the products are roughly similar and κ is relatively large. That is, we are considering assortments that include a large number of products. In the cases that are covered by Lemmas 3-5, the gap between Z^{UB} and Z^{OPT} is no more than 3/2. More interestingly, in the cases covered by Lemmas 3 and 4, we have $Z^{OPT} = Z^{UB}$ and there is no gap between the optimal expected profit and the upper bound. This explains to a certain degree the good performance of Z^{UB} that we observe in our computational experiments.

4.2.4 A Parametric Bound

The performance guarantees in §4.2.1 and §4.2.3 do not depend on the problem parameters. In this section, we establish a bound that depends only on the preference weights of the products (and is independent of the margins and product costs) and that can be potentially much tighter.

Recall that we can partition the interval $[t_{min}, t_{max}]$ into $O(n^2)$ subintervals where the profit-to-space ordering of the products do not change. Let $[\hat{t}_l, \hat{t}_u]$ be such a subinterval and suppose that we have $\rho_1(t)/v_1 \geq \ldots \geq \rho_n(t)/v_n > 0$ for all $t \in [\hat{t}_l, \hat{t}_u]$. Let κ_u be the largest index such that $\hat{t}_u \leq \tau_k = \frac{1}{v_0+V_k}$ and κ_l be the smallest index such that $\hat{t}_l \geq \tau_k = \frac{1}{v_0+V_k}$ and note that $\kappa_l > \kappa_u$. So we can write $[\hat{t}_l, \hat{t}_u] = \cup_{\kappa\in\{\kappa_l,\ldots,\kappa_u+1\}} \mathcal{I}_\kappa$ where $\mathcal{I}_{\kappa_l} = [\hat{t}_l, \tau_{\kappa_l-1}]$, $\mathcal{I}_\kappa = [\tau_\kappa, \tau_{\kappa-1}]$ for $\kappa \in \{\kappa_l - 1, \ldots, \kappa_u + 2\}$ and $\mathcal{I}_{\kappa_u+1} = [\tau_{\kappa_u+1}, \hat{t}_u]$ and

$$\max_{t\in[\hat{t}_l,\hat{t}_u]} \Gamma^f(t) = \ \max_{\kappa\in\{\kappa_l,\ldots,\kappa_u+1\}} \Big\{ \max_{t\in\mathcal{I}_\kappa} \Pi_\kappa(t) \Big\}.$$

Therefore, in order to bound the gap between $Z^{UB} = \max_{t\in[t_{min},t_{max}]} \Gamma^f(t)$ and Z^{OPT}, it suffices to bound the gap between $\max_{t\in\mathcal{I}_\kappa} \Pi_\kappa(t)$ and Z^{OPT}, which we proceed to do next.

For $\kappa \in \{\kappa_l, \ldots, \kappa_u + 1\}$, let

$$r_\kappa = \min \left\{ \frac{\max_{t\in\mathcal{I}_\kappa} \Pi_\kappa(t)}{Z_{\{1,\ldots,\kappa-1\}}}, \frac{\max_{t\in\mathcal{I}_\kappa} \Pi_\kappa(t)}{Z_{\{1,\ldots,\kappa\}}} \right\} - 1. \tag{17}$$

Recall that $\Pi_\kappa(t)$ gives the expected profit for the assortment comprising of products $\{1, \ldots, \kappa - 1\}$ along with a fractional amount of product κ. Therefore, r_κ can be interpreted as a measure of the local optimality gap between the continuous relaxation and assortments obtained by "rounding down" and "rounding up" the fractional product. We have

$$\frac{\max_{t\in\mathcal{I}_\kappa} \Pi_\kappa(t)}{Z^{OPT}} \leq \min \left\{ \frac{\max_{t\in\mathcal{I}_\kappa} \Pi_\kappa(t)}{Z_{\{1,\ldots,\kappa-1\}}}, \frac{\max_{t\in\mathcal{I}_\kappa} \Pi_\kappa(t)}{Z_{\{1,\ldots,\kappa\}}} \right\} = 1 + r_\kappa$$

where the inequality follows from noting that $Z^{OPT} \geq \max\{Z_{\{1,\ldots,\kappa-1\}}, Z_{\{1,\ldots,\kappa\}}\}$. Therefore, r_κ is an upper bound on the relative gap between $\max_{t\in\mathcal{I}_\kappa} \Pi_\kappa(t)$ and Z^{OPT}.

In the remainder of this section, we establish a bound on r_κ when $\kappa \in \{\kappa_l - 1, \ldots, \kappa_u + 2\}$. The analysis can be adapted to the cases when $\kappa \in \{\kappa_l, \kappa_u + 1\}$; we defer the details to the Online Appendix.

We let $\kappa \in \{\kappa_l - 1, \ldots, \kappa_u + 2\}$ and consider different scenarios. First, if $\Delta_\kappa \leq 0$, then $\Pi_\kappa(t)$ is decreasing (Lemma 2). Therefore, $\max_{t\in\mathcal{I}_\kappa} \Pi_\kappa(t) = \Pi_\kappa(\tau_{\kappa-1}) = Z_{\{1,\ldots,\kappa-1\}}$, where the last equality uses $\tau_{\kappa-1} = \frac{1}{v_0+V_{\kappa-1}}$. As a result, we have $r_\kappa = 0$. Next consider the case that $\Delta_\kappa > 0$ and let t^* denote the unconstrained maximizer of $\Pi_\kappa(t)$ (Lemma 2). If $\Delta_\kappa > 0$, $\Pi_\kappa(t)$ is concave. So if $t^* \geq \tau_{\kappa-1}$, then $\max_{t\in\mathcal{I}_\kappa} \Pi_\kappa(t) = \Pi_\kappa(\tau_{\kappa-1}) = Z_{\{1,\ldots,\kappa-1\}}$ and $r_\kappa = 0$. On the other hand, if $t^* \leq \tau_\kappa$, then $\max_{t\in\mathcal{I}_\kappa} \Pi_\kappa(t) = \Pi_\kappa(\tau_\kappa) = Z_{\{1,\ldots,\kappa\}}$ and again $r_\kappa = 0$. The three cases considered so far result in a trivial bound on r_κ. To obtain a nontrivial bound, we consider the case that $\Delta_\kappa > 0$ and $t^* \in [\tau_\kappa, \tau_{\kappa-1}]$, so that $\max_{t\in\mathcal{I}_\kappa} \Pi_\kappa(t) = \Pi_\kappa(t^*)$. Lemma 6 below shows that r_κ is maximal when the assortments $\{1, \ldots, \kappa-1\}$ and $\{1, \ldots, \kappa\}$ obtain the same expected profits.

Lemma 6. *Let $\kappa \in \{\kappa_l - 1, \ldots, \kappa_u + 2\}$. If $\Delta_\kappa > 0$ and $t^* \in [\tau_\kappa, \tau_{\kappa-1}]$, then r_κ is maximal when $Z_{\{1,\ldots,\kappa-1\}} = Z_{\{1,\ldots,\kappa\}}$.*

Since we are interested in obtaining an upper bound on the gap between Z^{OPT} and $\max_{t\in\mathcal{I}_\kappa} \Pi_\kappa(t) = \Pi_\kappa(t^*)$, we restrict ourselves to the case where r_κ is maximal. If $Z_{\{1,\ldots,\kappa-1\}} = Z_{\{1,\ldots,\kappa\}}$, it implies

$$\frac{c_\kappa}{v_\kappa} = \frac{\Delta_\kappa}{(v_0 + V_{\kappa-1})(v_0 + V_\kappa)}. \tag{18}$$

Using (18) in Equation (35) and simplifying, we have

$$r_\kappa = \frac{\Delta_\kappa \left(\sqrt{v_0 + V_\kappa} - \sqrt{v_0 + V_{\kappa-1}}\right)^2}{(v_0 + V_{\kappa-1})(v_0 + V_\kappa) Z_{\{1,\ldots,\kappa-1\}}}. \tag{19}$$

Lemma 7 below gives a lower bound on $Z_{\{1,\ldots,\kappa-1\}}$ which we use, in turn, to bound r_κ.

Lemma 7. *Let $\kappa \in \{\kappa_l - 1, \ldots, \kappa_u + 2\}$. If $\Delta_\kappa > 0$ and $t^* \in [\tau_\kappa, \tau_{\kappa-1}]$, then*

$$Z_{\{1,\ldots,\kappa-1\}} \geq \frac{\Delta_\kappa V_{\kappa-1}}{v_0(v_0 + V_{\kappa-1})} \left(1 - \sqrt{\frac{v_0 + V_{\kappa-1}}{v_0 + V_\kappa}}\right).$$

Using the lower bound from Lemma 7 in Equation (19), we have the following proposition.

Proposition 2. *Let $\kappa \in \{\kappa_l - 1, \ldots, \kappa_u + 2\}$. If $\Delta_\kappa > 0$ and $t^* \in [\tau_\kappa, \tau_{\kappa-1}]$, then*

$$r_\kappa \leq \frac{v_0 v_\kappa}{V_{\kappa-1}\sqrt{v_0 + V_\kappa}(\sqrt{v_0 + V_\kappa} + \sqrt{v_0 + V_{\kappa-1}})}.$$

Proposition 2 together with the observations preceding Lemma 6 provide a complete characterization of the performance gap for the intervals $\mathcal{I}_\kappa$ with $\kappa \in \{\kappa_l - 1, \ldots, \kappa_u + 2\}$: if $\Delta_\kappa > 0$ and $t^* \in [\tau_\kappa, \tau_{\kappa-1}]$, then the bound in Proposition 2 applies, otherwise $r_\kappa = 0$. As mentioned, it is possible to adapt the analysis to obtain similar bounds for the intervals $\mathcal{I}_{\kappa_l}$ and $\mathcal{I}_{\kappa_u+1}$; we defer the details to the Online Appendix.

Applying the parametric bound to the example in Figure 1, we obtain a bound of $\frac{2}{3\sqrt{6}(\sqrt{6}+\sqrt{3})} = 6.51\%$. This follows from using Proposition 2 to the interval where product 2 is fully included and the marginal product is 1: $v_0 = 1, v_\kappa = 2, V_{\kappa-1} = 3$.

We note that the numerator of the bound in Proposition 2 depends on the preference weight of product κ, while the denominator is a function of the sum of the preference weights of products $\{1,\ldots,\kappa-1\}$, $V_{\kappa-1}$, and the sum of the preference weights of products $\{1,\ldots,\kappa\}$, V_κ. The bound becomes large when v_κ is much larger than $V_{\kappa-1}$ and its worst case value is not better than 2. On the other hand, if the preference weights of the products are not dramatically different and we are considering assortments with a relatively large number of products, then we expect the denominator of the upper bounding term to dominate the numerator and r_κ to be quite small. In such cases, we expect Z^{UB} to be quite close to Z^{OPT} as well. The parametric bound thus provides more insight into why the gap between Z^{UB} and Z^{OPT} is often small in our computational experiments.

5 Assortment Planning with Fixed Costs, Constraints and Multiple Classes

In this section, we consider the assortment problem with fixed costs with additional constraints on the assortment and multiple customer classes, i.e., mixture of logit demands. To solve this more complex problem, we provide a dual formulation for our upper bound. We then study the tractability of the solution method and discuss its performance guarantees.

We now add a total of E constraints that limit the assortments can that be offered:

$$\sum_{j\in\mathcal{J}} \alpha_{e,j} x_j \leq \beta_e \,\forall e \in \mathcal{E} \tag{20}$$

where $\mathcal{E}=\{1,\ldots,E\}$ denotes the set of constraints.

We also consider multiple customer classes $d\in\mathcal{D}=\{1,\ldots,D\}$ and let θ_d denote the fraction of customers belonging to class d so that $\sum_{d\in\mathcal{D}}\theta_d=1$. We let $v_{j,d}$ denote the preference weight associated with class d for product j (we keep $v_{0,d}=v_0$ without loss of generality) and $\breve{p}_{j,d}$ its corresponding margin. Let $Z(x)=\sum_{d\in\mathcal{D}}\theta_d\frac{\sum_{j\in\mathcal{J}}\breve{p}_{j,d}v_{j,d}x_j}{v_0+\sum_{j\in\mathcal{J}}v_{j,d}x_j}-\sum_{j\in\mathcal{J}}c_jx_j=\sum_{d\in\mathcal{D}}\frac{\sum_{j\in\mathcal{J}}p_{j,d}v_{j,d}x_j}{v_0+\sum_{j\in\mathcal{J}}v_{j,d}x_j}-\sum_{j\in\mathcal{J}}c_jx_j$, where $p_{j,d}=\theta_d\breve{p}_{j,d}$ can be interpreted as the expected margin for product j from class d. The optimal expected profits for this extension can be obtained by solving

$$\begin{aligned}(OPT)\qquad Z^{OPT}=\max\quad & Z(x)\\ \text{s.t.}\quad & (20),\quad x_j\in\{0,1\}.\end{aligned}$$

Note that this formulation is generic enough to allow for the same product to be sold at different prices to different customer classes. This may be useful in situations where the different classes are mapped to different retail stores and there is flexibility in terms of setting the store prices.

As in the unconstrained case, we can write the constrained assortment optimization problem equivalently as $Z^{OPT}=\max_{t_1,\ldots,t_D|t_d\in[t_{d,min},t_{d,max}]}\Gamma^b(t_1,\ldots,t_D)$ where

$$\begin{aligned}\Gamma^b(t_1,\ldots,t_D)=\max\quad & \textstyle\sum_{j\in\mathcal{J}}\left(\sum_{d\in\mathcal{D}}p_{j,d}v_{j,d}t_d-c_j\right)x_j\\ \text{s.t.}\quad & x_j\in\{0,1\}\\ & \textstyle\sum_{j\in\mathcal{J}}\alpha_{e,j}x_j\leq\beta_e\,\forall e\in\mathcal{E}\\ & \textstyle\sum_{j\in\mathcal{J}}v_{j,d}x_j\leq\frac{1}{t_d}-v_0\forall d\in\mathcal{D}\end{aligned} \tag{21}$$

and $t_{d,min} = \frac{1}{v_0 + \sum_{j \in \mathcal{J}} v_{j,d}}$ and $t_{d,max} = \frac{1}{v_0 + \min_j \{v_{j,d}\}}$. Even with a single customer class ($D = 1$), $\Gamma^b(t_1, \ldots, t_D)$ is a multidimensional binary knapsack problem and is intractable to solve. We again obtain an upper bound by working with the continuous relaxation of $\Gamma^b(t)$:

$$\begin{aligned} \Gamma^f(t_1, \ldots, t_D) = \max \quad & \textstyle\sum_{j \in \mathcal{J}} \left(\sum_{d \in \mathcal{D}} p_{j,d} v_{j,d} t_d - c_j \right) x_j \\ \text{s.t.} \quad & 0 \le x_j \le 1 \\ & \textstyle\sum_{j \in \mathcal{J}} \alpha_{e,j} x_j \le \beta_e \, \forall e \in \mathcal{E} \\ & \textstyle\sum_{j \in \mathcal{J}} v_{j,d} x_j \le \frac{1}{t_d} - v_0 \forall d \in \mathcal{D}. \end{aligned} \tag{22}$$

We have that $Z^{UB} = \max_{t_1, \ldots, t_D | t_d \in [t_{d,min}, t_{d,max}]} \Gamma^f(t_1, \ldots, t_D)$ is an upper bound on Z^{OPT}. As in the unconstrained single-class case, we can further tighten the continuous relaxation by restricting attention to the products contained in the set $\mathcal{J}(t_1, \ldots, t_D) = \{j | v_{j,d} \le \frac{1}{t_d} - v_0 \, \forall d\}$; we suppress the dependence for ease of notation.

5.1 The Dual

The linear program in (22) can be rewritten through the dual and the strong duality theorem (Bertsimas and Tsitsiklis 1997), as follows. In this formulation, λ_d represents the dual variable associated with the constraint $\sum_{j \in \mathcal{J}} v_{j,d} x_j \le \frac{1}{t_d} - v_0$, μ_e that with the constraint $\sum_{j \in \mathcal{J}} \alpha_{e,j} x_j \le \beta_e$ and z_j the dual variable for $x_j \le 1$.

$$\begin{aligned} \Gamma^f(t_1, \ldots, t_D) = \min \quad & \sum_{d \in \mathcal{D}} \lambda_d \left(\frac{1}{t_d} - v_0 \right) + \sum_{e \in \mathcal{E}} \mu_e \beta_e + \sum_{j \in \mathcal{J}} z_j \\ \text{s.t.} \quad & z_j + \sum_{d \in \mathcal{D}} \lambda_d v_{j,d} + \sum_{e \in \mathcal{E}} \mu_e \alpha_{j,e} \ge \sum_{d \in \mathcal{D}} p_{j,d} v_{j,d} t_d - c_j \\ & \lambda_d, \mu_e, z_j \ge 0 \\ = \min \quad & \sum_{d \in \mathcal{D}} \lambda_d \left(\frac{1}{t_d} - v_0 \right) + \sum_{e \in \mathcal{E}} \mu_e \beta_e \\ & + \textstyle\sum_{j \in \mathcal{J}} \left(\sum_{d \in \mathcal{D}} p_{jd} v_{jd} t_d - c_j - \sum_{d \in \mathcal{D}} \lambda_d v_{jd} - \sum_{e \in \mathcal{E}} \mu_e \alpha_{je} \right)^+ \\ \text{s.t.} \quad & \lambda_d, \mu_e \ge 0 \end{aligned} \tag{23}$$

where $x^+ = \max\{x, 0\}$. As we can see, this dual formulation only requires the optimization of a piecewise-linear objective over $\lambda_d, \mu_e \ge 0$. This suggests that, given $(t_1, \ldots, t_D)$, the upper bound can be computed quickly, by inspecting all the break-points of the piecewise-linear function.

5.2 Alternative View of the Single-class, Unconstrained Case

When we have a single class and no constraints ($D = 1$, $E = 0$), then we recover the continuous knapsack problem described in §3: indeed the minimum of (23) is reached at λ equal to 0 or $[\rho_j(t)/v_j]^+ = [p_j t - c_j / v_j]^+$ for some j, where we drop the customer class index d from the subscripts to simplify the notation. Assuming without loss of generality (as before in §4.1) that $\rho_1(t)/v_1 \ge \ldots \ge \rho_n(t)/v_n \ge 0$, then the primal solution associated with $\lambda = \rho_\kappa(t)/v_\kappa$ is to select $x_j = 1$ for $j \le \kappa - 1$, a fractional value for x_κ, and $x_j = 0$ for $j > \kappa$, which results in an objective

equal to

$$G_\kappa(t) = \sum_{j=1}^{\kappa-1} \rho_j(t) v_j + \frac{\rho_\kappa(t)}{v_\kappa} \left(\frac{1}{t} - v_0 - \sum_{j=1}^{\kappa-1} v_j \right).$$

We define for completeness $\nu_{n+1}(t) = 0$ so $G_{n+1}(t) = \sum_{j \in \mathcal{J}} (\rho_j(t) v_j - c_j)$ and

$$\Gamma^f(t) = \min_{\kappa \in \mathcal{J}} G_\kappa(t). \tag{24}$$

As a result, if we now want to maximize this value by changing t (within the interval such that the order of $\rho_j(t)/v_j$ does not change), then the maximum over t can be either interior, i.e., there is κ such that $t^* = argmax G_\kappa(t)$, or at a breakpoint t such that $G_\kappa(t) = G_{\kappa+1}(t)$. In the case we have an interior solution, the first-order condition is

$$0 = \sum_{j=1}^{\kappa-1} p_j v_j + p_\kappa \left(\frac{1}{t} - v_0 - \sum_{j=1}^{\kappa-1} v_j \right) - \left(p_\kappa t - \frac{c_\kappa}{v_\kappa} \right) \frac{1}{t^2} = \sum_{j=1}^{\kappa-1} (p_j - p_\kappa) v_j - p_\kappa v_0 + \frac{c_\kappa}{v_\kappa} \frac{1}{t^2}$$

which is the same value identified in Lemma 2. In the case we have a solution at a breakpoint, $G_\kappa(t) = G_{\kappa+1}(t)$ yields a quadratic equation in t. One root of the quadratic equation satisfies

$$(p_\kappa - p_{\kappa+1}) t = \frac{c_\kappa}{v_\kappa} - \frac{c_{\kappa+1}}{v_{\kappa+1}},$$

i.e., $t = \hat{t}_{\kappa,\kappa+1}$. The second root satisfies $t = \frac{1}{v_0 + V_\kappa}$. We thus recover all the results presented in §4.1. Specifically, with this alternative view we again see that the complexity to identify the best value of t is $O(n^3)$: for each ordering of $\{\rho_j(t)/v_j | \forall j\}$ (possibly $O(n^2)$ of them), we need to consider $O(n)$ interior candidates and $O(n)$ breakpoint candidates. This is the same structure as the one described for the primal.

5.3 General Case

In the general case with assortment constraints and multiple customer classes, the problem in (23) is a minimization of a piecewise-linear function of $\{\lambda_d | \forall d \in \mathcal{D}\}$ and $\{\mu_e | \forall e \in \mathcal{E}\}$. Compared to §5.2, instead of a search over one dimension (that of λ), we must now search a space of $D + E$ dimensions, so the number of breakpoints to consider for each $(t_1, \ldots, t_D)$ is $\frac{n!}{(D+E)!(n-D-E)!}$, thus $O(n^{D+E})$, polynomial in n but exponential in $D + E$. Still the structure conceptually remains the same as the single-class unconstrained case. Lemma 8 below is the analog of Equation (24) in the unconstrained case.

Lemma 8. *$\Gamma^f(t_1, \ldots, t_D)$ is the minimum of $O(n^{D+E})$ functions of the form*

$$G_\kappa(t_1, \ldots, t_D) = \sum_{d_1, d_2 \in \mathcal{D}} \eta_{\kappa,d_1,d_2} \frac{t_{d_1}}{t_{d_2}} + \sum_{d \in \mathcal{D}} \xi_{\kappa,d} t_d + \sum_{d \in \mathcal{D}} \chi_{\kappa,d} \frac{1}{t_d} + \phi_\kappa, \tag{25}$$

for appropriately defined $\eta_{\kappa,d_1,d_2}, \xi_{\kappa,d}, \chi_{\kappa,d}, \phi_\kappa$.

Hence, in the general case, we find that $\Gamma^f(\cdot)$ can still be computed relatively easily for a given $(t_1, \ldots, t_D)$. However, the existence of multiple classes complicates the functional shape of $G_\kappa(\cdot)$, which are fractional functions of $(t_1, \ldots, t_D)$. The coefficients of $G_\kappa(\cdot)$, $\eta_{\kappa,d_1,d_2}, \xi_{\kappa,d}, \chi_{\kappa,d}$

and ϕ_κ are specified in the proof of the lemma and are more complex to compute since they require the inversion of a square matrix of dimension $D+E$. This can be done in a complexity of $O\Big(\max(D,E)^3\Big)$, by Gauss-Jordan elimination for example.

To generate the upper bound Z^{UB}, we must now search for values of $(t_1,\ldots,t_D)$ that maximize $\Gamma^f(\cdot)$. This is an easy task when there is a single class.

Lemma 9. *When $D=1$, $\Gamma^f(t)$ is either maximized at:*

1. $t^*_\kappa = \sqrt{\frac{\chi_\kappa}{\xi_\kappa}}$ *(when the term inside the square root is positive);*
2. *or, $\hat{t}_{\kappa_1,\kappa_2}$ one of the two solutions of*

$$\eta_{\kappa_1}+\phi_{\kappa_1}+\xi_{\kappa_1}t+\chi_{\kappa_1}\frac{1}{t}=\eta_{\kappa_2}+\phi_{\kappa_2}+\xi_{\kappa_2}t+\chi_{\kappa_2}\frac{1}{t}.$$

Hence, under a single customer class, it is sufficient to inspect a polynomial number of values of t. The number to inspect is dominated by the number of $\hat{t}_{\kappa_1\kappa_2}$: $O(n^{2+2E})$, the order of the square of the number of functions $G_\kappa(t)$ that we consider. For each of these values of t, we then need to compare the $O(n^{1+E})$ values of $G_\kappa(t)$. Taking into account that matrix inversion in this case takes $O(E^3)$, we thus have the following proposition.

Proposition 3. *When $D=1$, Z^{UB} can be obtained in a running time of $O(E^3n^{1+E}+n^{3+3E})$.*

This extends Proposition 1 by incorporating assortment constraints. In the constrained case, we obtain a pseudo-polynomial complexity. Given that typically the number of products is very large but the number of constraints small (constraints on the total number of products and/or space consumed by the products), this means that our heuristic will run reasonably fast in an application.

When we have multiple classes, the problem becomes more complicated. Indeed, the maximizer of $\Gamma^f(t_1,\ldots,t_D)$ can be of two kinds. One possibility is that it is the minimizer of a given $G_\kappa(\cdot)$, in which case we need to first characterize such a minimizer (as in the first case of Lemma 9), and then to guarantee that there is a polynomial number of those. The other possibility is that the solution happens to be at a $(t_1,\ldots,t_D)$ such that multiple $G_\kappa(\cdot)$ attain the same value (although it is not a local minimum for any of the $G_\kappa(\cdot)$), in which case we need to solve equations such as those in the second case of Lemma 9.

There are some special cases where the problem becomes more tractable. We focus here on one such case: the upper bound can be obtained efficiently when the preference weights of the products are identical across the different classes. That is, we have $v_{j,d}=v_j$ for all d and j. Customer heterogeneity can be incorporated by using different $v_{0,d}$, as well as different prices. Without loss of generality, let us assume that for $d\in\{2,\ldots,D\}$, $v_{0,d}=v_{0,1}+w_d$ with $w_d\geq 0$. This implies that we can write $\frac{1}{t_d}=\frac{1}{t_1}+w_d$. Thus, Equation (25) turns into

$$G_\kappa(t_1,\ldots,t_D)=G_\kappa(t_1)=\tilde{\xi}_\kappa t_1+\sum_{d\in\mathcal{D}}\tilde{\chi}_{\kappa,d}\frac{1}{t_1+w_d}+\tilde{\phi}_\kappa, \tag{26}$$

for appropriately defined $\tilde{\xi}_\kappa,\tilde{\chi}_{\kappa,d},\tilde{\phi}_\kappa$.

Lemma 10. *When $D\geq 2$ and $v_{j,d}=v_j$ for all d and j, $\Gamma^f(t_1)$ is either maximized at:*

1. *one of the at most D maxima of $G_\kappa(t_1)$;*

2. *or, one of the at most $D+1$ solutions of $G_{\kappa_1}(t_1) = G_{\kappa_2}(t_1)$.*

Hence, in this case one should find solutions of a polynomial equation of degree $2D$ equal to zero (the first case, in which only D of the $2D$ solutions correspond to a maximum of $G_\kappa(\cdot)$); and of a polynomial equation of degree $D+1$. Approximate solutions of these equations can be quickly obtained via standard numerical methods. Overall, finding the upper bound $\max_{t_1} \Gamma^f(t_1)$ can be done with the following procedure:

- Compute $G_\kappa(t_1)$ for all κ (the number of such functions is $O(n^{D+E})$, the complexity to find one is $O\Big((D+E)^3\Big)$).
- Identify the candidate values of t at which $\Gamma^f(t)$ may be maximized:
  - For each κ, find $t^*_{\kappa,m}$, for $m = 1, \ldots, O(D)$, maxima of $G_\kappa(t)$;
  - For each κ_1, κ_2, find $\hat{t}_{\kappa_1,\kappa_2,m}$, for $m = 1, \ldots, O(D)$, solution to $G_{\kappa_1}(t) = G_{\kappa_2}(t)$.
- Compare for each candidate $t = t^*_{\kappa,m}, \hat{t}_{\kappa_1,\kappa_2,m}$ the values of $G_\kappa(t)$ for all κ and set the minimum to be $\Gamma^f(t)$ (the number of such candidates is $O(n^{2(D+E)})$, the complexity to find one minimum is $O(n^{D+E})$).
- Select the maximum of $\Gamma^f(t)$ for each candidate $t = t^*_{\kappa,m}, \hat{t}_{\kappa_1,\kappa_2,m}$.

In summary, the more general problem is tractable when $D = 1$ and at least in one particular case when $D \geq 2$. However, we need to resort to numerical optimization methods in the general case $D \geq 2$.

5.4 Primal Solutions and Performance Guarantees

The dual approach outlined above provides a way of computing Z^{UB}. However, it is not a priori clear that we can generate primal solutions easily. In the unconstrained case, the upper bound was associated with a fractional solution x_j such that at most one product had a non-integer value. We found that including or excluding this item in the assortment provided a solution with a guaranteed performance (see §4.2). When there are multiple products with non-integer values, then one must consider including or excluding *any* of them, and there may potentially an exponential number of such combinations. This has two implications. First, there is no direct way to obtain a good solution from the tractable computation of the upper bound. Second, we may not be able to guarantee the performance of such a heuristic, as we did before.

Fortunately, we identify in this section some common cases where the "rounding" process can be done in a simple way, without degrading the performance guarantee obtained for the unconstrained case. Throughout we assume $D = 1$, $t \leq \frac{1}{v_0 + \max\{v_j\}}$ and $\mathcal{J}(t) = \mathcal{J}$.

5.4.1 Cardinality Constraints

Consider that we have a constraint on the size of the assortment, so that the total number of items offered is no more than K: $\sum_{j \in \mathcal{J}} x_j \leq K$, where $K \geq 1$. Cardinality constraints arise when there is a need to limit the total number of products in the assortment for operational

reasons; see Rusmevichientong et al. (2010), Gallego and Topaloglu (2014). We show below that $\Gamma^f(t) \leq 2\Gamma^b(t)$ when we have cardinality constraints.

For a given t, let $\hat{x}$ denote an optimal primal solution to $\Gamma^f(t)$. If constraint (20) is nonbinding, then the problem reduces to the unconstrained case and we have $\Gamma^f(t) \leq 2\Gamma^b(t)$ from the earlier arguments. On the other hand, if constraint (11) is nonbinding, then $\Gamma^f(t) = \Gamma^b(t)$ and the result again follows.

So we consider the case where both constraints (11) and (20) are binding. Therefore, we must have at least $n-2$ of the remaining constraints ($0 \leq x_j \leq 1$) binding, which implies that at least $n-2$ of the variables are either at their upper or lower bounds. Equivalently there are at most two variables that take fractional values. Let $\mathcal{F} = \{j \,|\, 0 < \hat{x}_j < 1\}$ so that $\mathcal{F}^c = \{j \,|\, \hat{x}_j \in \{0,1\}\}$. Assume without loss of generality that $\mathcal{F} = \{1,2\}$. We have

$$\begin{aligned} v_1\hat{x}_1 + v_2\hat{x}_2 &= \hat{y} \\ \hat{x}_1 + \hat{x}_2 &= \hat{K}, \end{aligned}$$

where $\hat{y} = \frac{1}{t} - v_0 - \sum_{k\in\mathcal{F}^c} v_k\hat{x}_k$ and $\hat{K} = K - \sum_{k\in\mathcal{F}^c} \hat{x}_k$. Note that $\hat{K} \in \{0,1,2\}$. If $\hat{K} = 0$ or $\hat{K} = 2$, then $\hat{x}_1$ and $\hat{x}_2$ are integer, which contradicts $\mathcal{F} = \{1,2\}$.

So it must be the case that $\hat{K} = 1$. Solving for $\hat{x}_1$ and $\hat{x}_2$, we have $\hat{x}_1 = (\hat{y} - v_2)/(v_1 - v_2)$ and $\hat{x}_2 = (v_1 - \hat{y})/(v_1 - v_2)$. Since $\hat{x}_1 \geq 0$ and $\hat{x}_2 \geq 0$, we must either have $v_1 \geq \hat{y} \geq v_2$ or $v_2 \geq \hat{y} \geq v_1$. Let us assume that $v_1 \geq \hat{y} \geq v_2$ (the other case is symmetric). We show that $\Gamma^f(t) \leq 2\Gamma^b(t)$ by constructing two solutions $\tilde{x}$ and $\bar{x}$ from $\hat{x}$ that are feasible to $\Gamma^b(t)$.

We construct the solution $\tilde{x}$ in the following manner. We set $\tilde{x}_1 = 0, \tilde{x}_2 = 1$ and $\tilde{x}_j = \hat{x}_j$ for $j \in \mathcal{F}^c$. We argue that $\tilde{x}$ is feasible to $\Gamma^b(t)$. We have $\sum_{j\in\mathcal{J}} v_j\tilde{x}_j = v_2 + \sum_{j\in\mathcal{F}^c} v_j\hat{x}_j \leq \hat{y} + \sum_{j\in\mathcal{F}^c} v_j\hat{x}_j = \frac{1}{t} - v_0$, where the inequality holds since $v_2 \leq \hat{y}$. We also have $\sum_{j\in\mathcal{J}} \tilde{x}_j = \tilde{x}_2 + \sum_{j\in\mathcal{F}^c} \tilde{x}_j = \hat{K} + \sum_{j\in\mathcal{F}^c} \hat{x}_j = K$, where we use $\tilde{x}_2 = 1 = \hat{K}$. Therefore $\tilde{x}$ is feasible to $\Gamma^b(t)$ which implies

$$\rho_2(t) + \sum_{j\in\mathcal{F}^c} \rho_j(t)\hat{x}_j = \sum_{j\in\mathcal{J}} \rho_j(t)\tilde{x}_j \leq \Gamma^b(t). \tag{27}$$

We next describe the construction of the solution $\bar{x}$. We set $\bar{x}_1 = 1$, $\bar{x}_2 = 0$ and $\bar{x}_j = 0$ for all $j \in \mathcal{F}^c$. By assumption $v_1 \leq \max_j v_j \leq \frac{1}{t} - v_0$ and so $\sum_{j\in\mathcal{J}} v_j\bar{x}_j \leq \frac{1}{t} - v_0$. The cardinality constraint is trivially satisfied by $\bar{x}$. Therefore the solution $\bar{x}$ is feasible to $\Gamma^b(t)$ and we have

$$\rho_1(t) = \sum_{j\in\mathcal{J}} \rho_j(t)\bar{x}_j \leq \Gamma^b(t). \tag{28}$$

Putting the above inequalities together,

$$\begin{aligned} \Gamma^f(t) &= \rho_1(t)\hat{x}_1 + \rho_2(t)\hat{x}_2 + \sum_{j\in\mathcal{F}^c} \rho_j(t)\hat{x}_j \\ &\leq \rho_1(t) + \rho_2(t) + \sum_{j\in\mathcal{F}^c} \rho_j(t)\hat{x}_j \\ &\leq 2\Gamma^b(t) \end{aligned}$$

where the first inequality uses the facts that $\hat{x}_1, \hat{x}_2 \leq 1$ and $\rho_1(t), \rho_2(t) \geq 0$. The second inequality uses (27) and (28). This implies the following result.

Proposition 4. *For the assortment problem with fixed costs and a cardinality constraint, we have* $Z^{UB} \leq 2Z^{OPT}$.

We note that the bound can be extended to the case with nested cardinality constraints. That is, we have nested subsets of products $S_1 \subset \ldots \subset S_m$ with cardinality restrictions $K_1 \leq \ldots \leq K_m$, respectively. The arguments also apply to the case where the products are partitioned into disjoint subsets with associated restrictions on the cardinality of each subset; see Davis et al. (2013).

5.4.2 Product Precedence Constraints

Consider now that we have constraints which link the offer decisions for the different products. In particular, associated with each product j, we have a subset $\mathcal{O}_j$ of products that have to be offered if product j is offered. That is, we have $x_j - x_i \leq 0$ for all $i \in \mathcal{O}_j$. We follow the convention that $j \in \mathcal{O}_j$. We consider the case where the precedence constraints are nested in the following sense: for two products j and k, we either have $\mathcal{O}_j \subset \mathcal{O}_k$ or $\mathcal{O}_k \subset \mathcal{O}_j$ or $\mathcal{O}_j \cap \mathcal{O}_k = \emptyset$. Product precedence constraints may arise in situations where say a more expensive variant or style cannot be included in the assortment unless a basic version of the product is also part of the assortment; see for example Davis et al. (2013). We show below that $\Gamma^f(t) \leq 2\Gamma^b(t)$ when we have nested product precedence constraints.

For a given t, we let $\mathcal{S}(t) = \{j \mid \sum_{i \in \mathcal{O}_j} v_i \leq \frac{1}{t} - v_0\}$ and note that if $j \notin \mathcal{S}(t)$, then the precedence constraints for product j preclude it from being a part of an optimal solution to $\Gamma^b(t)$. Therefore, we can restrict ourselves to products in $\mathcal{S}(t)$ while solving the continuous relaxation $\Gamma^f(t)$ as well. For ease of notation, we assume that $\mathcal{S}(t) = \mathcal{J}$, but this is without loss of generality and our results continue to hold even on relaxing this assumption.

Fix t and let $\hat{x}$ denote an optimal solution to $\Gamma^f(t)$. To avoid trivial cases, we assume that constraint (11) and at least one of the product precedence constraints are binding at t. Let $\mathcal{F} = \{j \mid 0 < \hat{x}_j < 1\}$ denote the set of variables assuming fractional values and assume that $\mathcal{F}$ is nonempty.

We begin with some preliminary results. The following lemma states that the optimal solution has at most one distinct fractional value.

Lemma 11. *If* $0 < \hat{x}_j, \hat{x}_k < 1$, *then* $\hat{x}_j = \hat{x}_k$.

The following lemma implies that all the variables assuming fractional values are contained in $\mathcal{O}_j$ for some product j. We first define the notion of a maximal element of the set $\mathcal{F}$. We say that $i \in \mathcal{F}$ is a maximal element of $\mathcal{F}$ if $j \notin \mathcal{F}$ for all j such that $i \in \mathcal{O}_j$. That is, if product i is required to be offered if product j is offered, then $j \notin \mathcal{F}$.

Lemma 12. *There exists a unique maximal element of* $\mathcal{F}$.

Corollary 1. *Let* i *denote the unique maximal element of* $\mathcal{F}$.
1. $\mathcal{F} \subset \mathcal{O}_i$.
2. *If* $j \in \mathcal{O}_i^c$, *then* $\hat{x}_j \in \{0, 1\}$.
3. *If* $i \in \mathcal{O}_j$, *then* $\hat{x}_j = 0$.
4. *If* $j \in \mathcal{O}_i \backslash \mathcal{F}$, *then* $\hat{x}_j = 1$.
5. $\sum_{j \in \mathcal{F}} \rho_j(t) \geq 0$.

We are now ready to show that $\Gamma^f(t) \leq 2\Gamma^b(t)$. We show the bound by constructing two solutions $\tilde{x}$ and $\bar{x}$ from $\hat{x}$ that are feasible to $\Gamma^b(t)$.

We construct the solution $\tilde{x}$ in the following manner. We set $\tilde{x}_j = \hat{x}_j$ for $j \in \mathcal{O}_i^c$ and $\tilde{x}_j = 0$ for $j \in \mathcal{O}_i$, where i is the unique maximal element of $\mathcal{F}$. By the second statement of Corollary 1, $\tilde{x}_j \in \{0, 1\}$ for all j. Since $\tilde{x}_j \leq \hat{x}_j$ for all j, we have that $\tilde{x}$ satisfies constraint (11). Next we verify that $\tilde{x}$ satisfies the product precedence constraints. Since $\tilde{x}_j = 0$ for all $j \in \mathcal{O}_i$, the product precedence constraints are satisfied as equalities for all $j \in \mathcal{O}_i$. Next, for j such that $i \in \mathcal{O}_j$, using the third statement of Corollary 1 we have $\tilde{x}_j = \hat{x}_j = 0$ and so that product precedence constraints $\tilde{x}_j - \tilde{x}_k \leq 0$ are satisfied for all $k \in \mathcal{O}_j$. Finally, if $\mathcal{O}_i \cap \mathcal{O}_j = \emptyset$, then $\tilde{x}_k = \hat{x}_k$ for all $k \in \mathcal{O}_j$ and the product precedence constraints are trivially satisfied for all $k \in \mathcal{O}_j$. It follows that $\tilde{x}$ is a feasible solution to $\Gamma^b(t)$ and we have

$$\sum_{j \in \mathcal{O}_i^c} \rho_j(t)\tilde{x}_j = \sum_{j \in \mathcal{J}} \rho_j(t)\tilde{x}_j \leq \Gamma^b(t). \tag{29}$$

Next we describe how we construct the solution $\bar{x}$. We set $\bar{x}_j = 0$ for all $j \in \mathcal{O}_i^c$ and $\bar{x}_j = 1$ for all $j \in \mathcal{O}_i$. Since $i \in \mathcal{S}(t)$ we have $\sum_k v_k \bar{x}_k = \sum_{k \in \mathcal{O}_i} v_k \leq y$ and so $\bar{x}$ satisfies constraint (11). We clearly have $\bar{x}_j - \bar{x}_k = 0$ for all $j \in \mathcal{O}_i$. On the other hand for $j \in \mathcal{O}_i^c$, we have $\tilde{x}_j = 0$ and so the constraint $\tilde{x}_j - \tilde{x}_k \leq 0$ is trivially satisfied for all $k \in \mathcal{O}_j$. Therefore, $\bar{x}$ is also feasible to $\Gamma^b(t)$ and we have

$$\sum_{j \in \mathcal{O}_i} \rho_j(t)\bar{x}_j = \sum_{j \in \mathcal{J}} \rho_j(t)\bar{x}_j \leq \Gamma^b(t).$$

Putting everything together we

$$\begin{aligned}
\Gamma^f(t) = & \quad \textstyle\sum_{j \in \mathcal{J}} \rho_j(t)\hat{x}_j \\
= & \quad \textstyle\sum_{j \in \mathcal{O}_i^c} \rho_j(t)\tilde{x}_j + \sum_{j \in \mathcal{O}_i} \rho_j(t)\hat{x}_j \\
\leq & \quad \Gamma^b(t) + \textstyle\sum_{j \in \mathcal{F}} \rho_j(t)\hat{x}_i + \sum_{j \in \mathcal{O}_i \setminus \mathcal{F}} \rho_j(t) \\
\leq & \quad \Gamma^b(t) + \textstyle\sum_{j \in \mathcal{F}} \rho_j(t)\bar{x}_j + \sum_{j \in \mathcal{O}_i \setminus \mathcal{F}} \rho_j(t)\bar{x}_j \\
\leq & \quad 2\Gamma^b(t)
\end{aligned}$$

where the second equality follows from the fact that $\tilde{x}_j = \hat{x}_j$ for all $j \in \mathcal{O}_i^c$. The first inequality uses (29), the fact that $\hat{x}_j = \hat{x}_i$ for all $j \in \mathcal{F}$ and the fourth statement of Corollary 1 which implies that $\hat{x}_j = 1$ for $j \in \mathcal{O}_i \setminus \mathcal{F}$. The last statement of Corollary 1 implies that $\sum_{j \in \mathcal{F}} \rho_j(t) \geq 0$. This together with the facts that $\hat{x}_i < 1 = \bar{x}_j$ for all $j \in \mathcal{F}$ yields the second inequality. The last inequality follows from (30). The above chain of inequalities shows that $\Gamma^f(t) \leq 2\Gamma^b(t)$ and this implies the following result.

Proposition 5. *For the assortment problem with fixed costs and nested product precedence constraints, we have* $Z^{UB} \leq 2Z^{OPT}$.

6 Computational Experiments

We compare the performance of our tractable approximation method with that of benchmark solution methods on test problems with a larger number of products. We consider the single-class unconstrained assortment optimization problem with fixed costs and generate our test

problems in the same manner as described in §4.2.2. We compare the upper bound obtained by the method proposed in §4, Z^{UB}, with the optimal expected profit, Z^{OPT}. In addition, we include two other benchmark solution methods in our computational experiments that we describe next.

The first benchmark we use is the linear programming relaxation of problem (2)-(6). We let Z^{LP} denote the optimal objective function value of the linear programming relaxation. This gives an upper bound on Z^{OPT} and it can be obtained in a tractable manner.

As the second benchmark, we use the method proposed in Feldman and Topaloglu (2015) which involves solving a relaxation of problem (10)-(12) over a discrete grid. Let $\mathcal{T} = \{t^s | s \in \{1, \ldots, \bar{s}(\sigma)\}\}$ be a set of $\bar{s}(\sigma)$ grid points which cover the interval $[t_{min}, t_{max}]$ where $\sigma > 0$ is a parameter that controls the size of the grid. Feldman and Topaloglu (2015) propose using an exponential grid so that $t^s = (1+\sigma)t^{s-1}$. As σ becomes smaller, the spacing between the grid points decreases and we obtain denser grids. Further, Feldman and Topaloglu (2015) propose solving a relaxation of problem (10)-(12) over the intervals defined by the grid points and take the maximum of these values to obtain an upper bound on the optimal expected profit. We denote the upper bound obtained by this method as $Z^{DG}(\sigma)$ where the argument emphasizes the dependence of the solution on the density of the grid. In our computational experiments, we consider three different grid densities by varying σ over the set $\{0.1, 0.01, 0.001\}$. Note that the grids get progressively denser as σ varies from 0.1 to 0.001.

It is possible to show that the Z^{UB} bound is provably tighter than the linear programming relaxation as well as the discrete grid approximation bounds. In our computational experiments we study the relative tightness of the upper bounds and how much Z^{UB} improves upon Z^{LP} and $Z^{DG}(\sigma)$.

Table 2 gives the tightness of the upper bounds obtained by the different solution methods for assortment problems with $n = 50$ products. The first column describes the problem characteristics using (n, Φ, γ). For each parameter combination, we generate 50 test problems in the manner described in §4.2.2. The second column of the table gives the average percentage difference between Z^{UB} and Z^{OPT}, while the third column gives the 95th percentile of the difference. The fourth column reports the value of the parametric bound described in §4.2.4 averaged over the 50 problem instances. The remaining columns report the averages and the 95th percentiles for Z^{LP}, $Z^{DG}(0.1)$, $Z^{DG}(0.01)$ and $Z^{DG}(0.001)$, respectively.

The percentage difference between Z^{UB} and Z^{OPT} is around 0.04% on average. The gaps tend to decrease as γ gets smaller. Recall that if γ is small, then the fixed costs of including the products are quite small comparable to their profit margins and the problem becomes closer to the assortment problem without fixed costs, where Z^{UB} and Z^{OPT} coincide. While the LP relaxation also displays the same trend, its performance tends to be quite poor in comparison and the gaps are around 25% on average. The quality of the upper bound obtained by the grid-based method, $Z^{DG}(\sigma)$, improves as σ becomes smaller. $Z^{DG}(0.1)$ is somewhat loose, $Z^{DG}(0.01)$ is tighter and $Z^{DG}(0.001)$ is quite close to Z^{UB}. Finally we note that the parametric bound of §4.2.4 gives a more accurate picture of the performance of Z^{UB} compared to the constant factor bounds of 2 and 3/2 (which would imply percentage differences of 100% and 50%, respectively).

Table 3 gives the CPU seconds required to obtain Z^{OPT} as well as the different upper bounds for the test problems with 50 products. All of our computational experiments are carried out on a Pentium Core 2 Duo desktop with 3-GHz CPU and 4-GB RAM. We use CPLEX 11.2 to solve all linear programs. Z^{LP} and $Z^{DG}(0.1)$ can be obtained in less than 1/100 of a second. The other methods take slightly longer but all the upper bounds can be obtained in a fraction

of a second. The solution time of the grid-based method increases as σ becomes smaller (and thus the number of grid points becomes larger).

Table 4 gives the tightness of the upper bounds obtained by the different solution methods for assortment problems with $n = 100$ products. The columns have the same interpretation as in Table 2. Overall, the results display similar trends to the test problems with 50 products. Z^{UB} and $Z^{DG}(0.001)$ are good approximations to Z^{OPT}, while the Z^{LP} and $Z^{DG}(0.1)$ bounds are somewhat loose. The parametric performance guarantee for Z^{UB} is significantly tighter compared to the case with 50 products. This is in line with Proposition 2 which indicates that the gap between Z^{UB} and Z^{OPT} is likely to be small for assortments including a large number of products. To our knowledge, similar theoretical guarantees are not available for the linear programming relaxation and the grid-based approximation.

Feldman and Topaloglu (2015) show that if the fixed costs satisfy a particular scaling property, then no grid, no matter how dense, can improve upon $Z^{DG}(\sigma)$ by more than a factor of $(1 + \sigma)$. However, if the fixed costs do not satisfy the scaling property, then the performance guarantee in Feldman and Topaloglu (2015) does not hold. In our test problems the product fixed costs do not scale in the manner described in Feldman and Topaloglu (2015) and indeed, we observe that their performance guarantee also does not hold. Z^{UB} can be viewed as the limiting value of $Z^{DG}(\sigma)$ as σ tends to zero and we have an infinitely dense grid. In tables 2 and 4 we observe that Z^{UB} can improve upon $Z^{DG}(0.1), Z^{DG}(0.01)$ and $Z^{DG}(0.001)$ by more than $10\%, 1\%$ and 0.1%, respectively.

Table 5 gives the CPU seconds required to obtain Z^{OPT} as well as the different upper bounds for the test problems with 100 products. The time required to obtain the optimal solution is noticeably greater than that for the upper bounds, and we observe instances where it can take several minutes to compute Z^{OPT}. This follows naturally from OPT being an NP-hard problem. On the other hand, Z^{UB} can still be obtained in a fraction of a second. As mentioned, Z^{UB} can be viewed as the limiting value of $Z^{DG}(\sigma)$ as σ tends to zero. The solution time of the grid-based method increases as σ decreases. On the other hand, Z^{UB} can be computed in roughly the same time as $Z^{DG}(0.001)$.

	% difference with Z^{OPT}										
	Z^{UB}			Z^{LP}		$Z^{DG}(0.1)$		$Z^{DG}(0.01)$		$Z^{DG}(0.001)$	
Problem (n, Φ, γ)	Avg.	95%	Param. bound	Avg.	95%	Avg.	95%	Avg.	95%	Avg.	95%
(50, 0.75, 1.00)	0.06	0.15	2.12	30	56	15	18	1.54	1.79	0.21	0.32
(50, 0.75, 0.50)	0.01	0.03	1.01	26	37	14	16	1.38	1.60	0.15	0.19
(50, 0.75, 0.25)	0.00	0.00	0.42	16	30	12	13	1.23	1.33	0.13	0.13
(50, 0.50, 1.00)	0.13	0.32	2.27	31	55	14	17	1.54	1.72	0.27	0.46
(50, 0.50, 0.50)	0.02	0.08	1.57	31	44	13	15	1.34	1.50	0.15	0.22
(50, 0.50, 0.25)	0.01	0.06	0.67	20	31	12	13	1.25	1.34	0.13	0.18
(50, 0.25, 1.00)	0.08	0.35	3.32	33	56	14	16	1.43	1.97	0.21	0.52
(50, 0.25, 0.50)	0.02	0.09	1.72	31	47	13	14	1.28	1.46	0.15	0.23
(50, 0.25, 0.25)	0.02	0.14	1.07	24	33	12	14	1.25	1.37	0.15	0.26

Table 2: Comparison of upper bounds for assortment problems with $n = 50$ products.

Problem (n, Φ, γ)	CPU secs. Z^{OPT}	Z^{UB}	Z^{LP}	$Z^{DG}(0.1)$	$Z^{DG}(0.01)$	$Z^{DG}(0.001)$
(50, 0.75, 1.00)	0.08	0.04	0.00	0.00	0.01	0.10
(50, 0.75, 0.50)	0.38	0.04	0.00	0.00	0.01	0.11
(50, 0.75, 0.25)	0.27	0.03	0.00	0.00	0.01	0.10
(50, 0.50, 1.00)	0.08	0.05	0.00	0.00	0.01	0.12
(50, 0.50, 0.50)	0.59	0.06	0.00	0.00	0.01	0.14
(50, 0.50, 0.25)	0.30	0.03	0.00	0.00	0.01	0.12
(50, 0.25, 1.00)	0.03	0.06	0.00	0.00	0.02	0.17
(50, 0.25, 0.50)	0.14	0.07	0.00	0.00	0.02	0.17
(50, 0.25, 0.25)	0.28	0.05	0.00	0.00	0.02	0.18

Table 3: CPU seconds to obtain Z^{OPT} and the different upper bounds for assortment problems with $n = 50$ products.

	% difference with Z^{OPT}										
	Z^{UB}			Z^{LP}		$Z^{DG}(0.1)$		$Z^{DG}(0.01)$		$Z^{DG}(0.001)$	
Problem (n, Φ, γ)	Avg.	95%	Param. bound	Avg.	95%	Avg.	95%	Avg.	95%	Avg.	95%
(100, 0.75, 1.00)	0.01	0.06	0.60	27	42	15	18	1.53	1.79	0.16	0.19
(100, 0.75, 0.50)	0.01	0.01	0.36	23	39	14	16	1.43	1.62	0.15	0.16
(100, 0.75, 0.25)	0.00	0.01	0.23	14	27	12	13	1.24	1.34	0.13	0.14
(100, 0.50, 1.00)	0.01	0.04	0.60	31	45	14	16	1.44	1.64	0.15	0.18
(100, 0.50, 0.50)	0.03	0.11	0.65	28	37	13	15	1.36	1.54	0.16	0.23
(100, 0.50, 0.25)	0.00	0.02	0.44	19	34	12	13	1.24	1.34	0.13	0.14
(100, 0.25, 1.00)	0.04	0.38	1.73	33	57	13	15	1.35	1.58	0.17	0.49
(100, 0.25, 0.50)	0.03	0.06	0.98	29	40	12	14	1.27	1.44	0.15	0.18
(100, 0.25, 0.25)	0.01	0.02	0.65	25	32	12	13	1.24	1.34	0.13	0.15

Table 4: Comparison of upper bounds for assortment problems with $n = 100$ products.

Problem (n, Φ, γ)	CPU secs. Z^{OPT}	Z^{UB}	Z^{LP}	$Z^{DG}(0.1)$	$Z^{DG}(0.01)$	$Z^{DG}(0.001)$
(100, 0.75, 1.00)	55	0.17	0.01	0.00	0.01	0.09
(100, 0.75, 0.50)	354	0.13	0.01	0.00	0.01	0.09
(100, 0.75, 0.25)	133	0.10	0.01	0.00	0.01	0.10
(100, 0.50, 1.00)	38	0.21	0.01	0.00	0.01	0.12
(100, 0.50, 0.50)	411	0.23	0.01	0.00	0.02	0.15
(100, 0.50, 0.25)	232	0.15	0.01	0.00	0.01	0.15
(100, 0.25, 1.00)	3	0.27	0.00	0.00	0.02	0.21
(100, 0.25, 0.50)	60	0.30	0.00	0.00	0.02	0.22
(100, 0.25, 0.25)	277	0.24	0.01	0.00	0.02	0.20

Table 5: CPU seconds to obtain Z^{OPT} and the different upper bounds for assortment problems with $n = 100$ products.

7 Conclusion

This paper has developed a new continuous relaxation of the assortment planning problem with product fixed costs. This formulation uses a parameter related to the total assortment attractiveness, and simplifies the problem into a parametric continuous knapsack. We prove that this approach requires evaluating a finite and polynomial number of values for the parameter, which means that the upper bound can be calculated in polynomial time, while the original integer program is NP-hard. We also prove that the heuristic based on the relaxation obtains an optimality gap equal to 2 in general, but much tighter when products are all similar. In numerical experiments with random instances, the gap is still tighter, below 1%. Moreover, our methodology can be extended to consider multiple classes and constraints on the assortment.

Our results can be exploited further. One future line of research is to employ our parametric formulation to other assortment demand models, in particular nested demand models. Another direction of work is to develop stronger parametric formulations for the mixed MNL model, where it might be possible to relate through simple relationships the variables $(t_1, \ldots, t_D)$, thereby reducing the number of dimensions of the problem and allowing us to provide stronger heuristics and performance guarantees.

References

Anderson, Simon P., André De Palma, Jacques-François Thisse. 1992. *Discrete choice theory of product differentiation*. MIT Press.

Bertsimas, Dimitris, John N. Tsitsiklis. 1997. *Introduction to linear optimization*, vol. 6. Athena Scientific Belmont, MA.

Caro, Felipe, Victor Martínez-de Albéniz. 2015. Fast fashion: Business model overview and research opportunities. Narendra Agrawal, Stephen A. Smith, eds., *Retail Supply Chain Management: Quantitative Models and Empirical Studies, 2nd Edition*. Springer, New York, 237–264.

Davis, James M., Guillermo Gallego, Huseyin Topaloglu. 2013. Assortment planning under the multinomial logit model with totally unimodular constraint structures. Working paper, Cornell University.

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

Feldman, Jacob B., Huseyin Topaloglu. 2014. Capacity constraints across nests in assortment optimization under the nested logit model. Working paper, Cornell University.

Feldman, Jacob B., Huseyin Topaloglu. 2015. Bounding optimal expected revenues for assortment optimization under mixtures of multinomial logits. *Production and Operations Management* **Forthcoming** N/A.

Gallego, Guillermo, Huseyin Topaloglu. 2014. Constrained assortment optimization for the nested logit model. *Management Science* **60**(10) 2583–2601.

Kök, A. Gürhan, Marshall L. Fisher, Ramnath Vaidyanathan. 2009. Assortment planning: Review of literature and industry practice. *Retail supply chain management*. Springer, 99–153.

Kunnumkal, Sumit. 2015. On upper bounds for assortment optimization under the mixture of multinomial logit models. *Operations Research Letters* **43**(2) 189–194.

Kunnumkal, Sumit, Paat Rusmevichientong, Huseyin Topaloglu. 2009. Assortment optimization under the multinomial logit model with product costs. Working paper, Cornell University.

Kunnumkal, Sumit, Huseyin Topaloglu. 2008. A refined deterministic linear program for the network revenue management problem with customer choice behavior. *Naval Research Logistics Quarterly* **55** 563–580.

Miranda Bront, Juan José, Isabel Méndez-Díaz, Gustavo Vulcano. 2009. A column generation algorithm for choice-based network revenue management. *Operations Research* **57** 769–784.

Rusmevichientong, Paat, Zuo-Jun Max Shen, David B. Shmoys. 2009. A ptas for capacitated sum-of-ratios optimization. *Operations Research Letters* **37**(4) 230–238.

Rusmevichientong, Paat, Zuo-Jun Max Shen, David B. Shmoys. 2010. Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations research* **58**(6) 1666–1680.

Rusmevichientong, Paat, David B. Shmoys, Chaoxu Tong, Huseyin Topaloglu. 2014. Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management* **23**(11) 2023–2039.

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

Topaloglu, Huseyin. 2013. Joint stocking and product offer decisions under the multinomial logit model. *Production and Operations Management* **22**(5) 1182–1199.

van Ryzin, Garrett J., Siddharth Mahajan. 1999. On the relationship between inventory costs and variety benefits in retail assortments. *Management Science* **45**(11) 1496–1509.

Vazirani, Vijay V. 2013. *Approximation algorithms*. Springer Science & Business Media.

Appendix: Proofs

Proof of Lemma 2

Proof. Using (7) and simplifying, the derivative of $\Pi_\kappa(t)$ is such that

$$\Pi'_\kappa(t) = \sum_{j=1}^{\kappa-1} p_j v_j + \frac{c_\kappa/v_\kappa}{t^2} - p_\kappa(v_0 + V_{\kappa-1}) = \frac{c_\kappa/v_\kappa}{t^2} - \Delta_\kappa$$

If $\Delta_\kappa \leq 0$, then $\Pi'_\kappa(t) \geq 0$. So $\Pi_\kappa(t)$ is an increasing function of t and the maximum occurs at the right end-point of the interval, u. On the other hand, if $\Delta_\kappa > 0$, then $\Pi''_\kappa(t) < 0$. Hence $\Pi_\kappa(t)$ is concave and its unconstrained is maximizer is t^*. It follows that the constrained maximizer $\bar{t}^* = \max\{l, \min\{t^*, u\}\}$. □

Proof of Lemma 3

Proof. If $\Delta_\kappa \leq 0$, then $\Pi'_\kappa(t) \geq 0$ for all t. Therefore, $\Pi_\kappa(t)$ is nondecreasing in t. Since $u \leq \frac{1}{v_0+V_{\kappa-1}} = \tau_{\kappa-1}$, we have

$$\max_{t\in[l,u]} \Pi_\kappa(t) \leq \Pi_\kappa(\tau_{\kappa-1}) = Z_{\{1,\ldots,\kappa-1\}} \leq Z^{OPT}$$

where $Z_{\{1,\ldots,\kappa-1\}}$ is the expected profit obtained from the assortment $\{1,\ldots,\kappa-1\}$. □

Proof of Lemma 4

Proof. If $\Delta_\kappa > 0$, then $\Pi''_\kappa(t) < 0$ and $\Pi_\kappa(t)$ is concave. Since t^* is the unconstrained maximizer of $\Pi_\kappa(t)$, the function is increasing to the left of t^* and decreasing to the right of t^*. When $t^* \leq \frac{1}{v_0+V_\kappa} = \tau_\kappa \leq l$, it follows that

$$\max_{t\in[l,u]} \Pi_\kappa(t) \leq \Pi_\kappa(\tau_\kappa) = Z_{\{1,\ldots,\kappa\}} \leq Z^{OPT}.$$

and when $t^* \geq \frac{1}{v_0+V_{\kappa-1}} = \tau_{\kappa-1} \geq u$,

$$\max_{t\in[l,u]} \Pi_\kappa(t) \leq \Pi_\kappa(\tau_{\kappa-1}) = Z_{\{1,\ldots,\kappa\}} \leq Z^{OPT}.$$

□

Proof of Lemma 5

Proof. We begin with some observations. Since t^* is the unconstrained maximizer of $\Pi_\kappa(t)$, we have $\Pi'_\kappa(t^*) = 0$ and so

$$\Delta_\kappa(t^*)^2 = \frac{c_\kappa}{v_\kappa}. \tag{30}$$

Let $Z_{\{1,\ldots,\kappa-1\}}$, $Z_{\{1,\ldots,\kappa\}}$ and $Z_{\{\kappa\}}$, respectively, denote the expected profits associated with the assortments $\{1,\ldots,\kappa-1\}$, $\{1,\ldots,\kappa\}$ and $\{\kappa\}$. We have

$$\begin{aligned}
\Pi_\kappa(t^*) - Z_{\{1,\ldots,\kappa-1\}} &= \textstyle\sum_{j=1}^{\kappa-1} \rho_j(t^*) + \rho_\kappa(t^*)\left(\frac{\frac{1}{t^*}-v_0-V_{\kappa-1}}{v_\kappa}\right) - \sum_{j=1}^{\kappa-1} \rho_j(\tau_{\kappa-1}) \\
&= \left(\tfrac{1}{t^*} - v_0 - V_{\kappa-1}\right)\left(\frac{p_\kappa(v_0+V_{\kappa-1})-\sum_{j=1}^{\kappa-1} p_j v_j}{v_0+V_{\kappa-1}} t^* - \frac{c_\kappa}{v_\kappa}\right) \\
&= \left(\tfrac{1}{t^*} - v_0 - V_{\kappa-1}\right)^2 \tfrac{c_\kappa}{v_\kappa(v_0+V_{\kappa-1})}
\end{aligned} \tag{31}$$

where the first equality uses $\tau_{\kappa-1} = \frac{1}{v_0+V_{\kappa-1}}$, the second equality uses (7) and the last equality uses (15) and (30). In a similar manner, we have

$$\Pi_\kappa(t^*) - Z_{\{1,\ldots,\kappa\}} = \left(V_\kappa + v_0 - \frac{1}{t^*}\right)^2 \frac{c_\kappa}{v_\kappa(v_0+V_\kappa)}. \tag{32}$$

Finally, we note that

$$\begin{aligned}
\Pi_\kappa(t^*) &\leq \textstyle\sum_{j=1}^{\kappa-1} \rho_j(\tau_{\kappa-1}) + \rho_\kappa\left(\frac{1}{v_0+v_\kappa}\right)\left(\frac{\frac{1}{t^*}-v_0-V_{\kappa-1}}{v_\kappa}\right) \\
&= Z_{\{1,\ldots,\kappa-1\}} + Z_{\{\kappa\}}\left(\frac{\frac{1}{t^*}-v_0-V_{\kappa-1}}{v_\kappa}\right)
\end{aligned} \tag{33}$$

where the inequality holds since $\frac{1}{t^*} - v_0 \geq \max\{V_{\kappa-1}, v_\kappa\}$ and $\rho_j(t)$ is an increasing function of t.

We are now ready to prove the lemma. We consider two cases.

Case 1: $\frac{1}{t^*} - v_0 \leq V_{\kappa-1} + \frac{v_\kappa}{2}$. In this case $\frac{1}{t^*} - v_0 - V_{\kappa-1} \leq \frac{v_\kappa}{2}$ and (33) implies that $\Pi_\kappa(t^*) \leq Z_{\{1,\ldots,\kappa-1\}} + Z_{\{\kappa\}} \left(\frac{\frac{1}{t^*} - v_0 - V_{\kappa-1}}{v_\kappa} \right) \leq \frac{3}{2} Z^{OPT}$.

Case 2: $\frac{1}{t^*} - v_0 > V_{\kappa-1} + \frac{v_\kappa}{2}$. Using (31) and (33) we have

$$Z_{\{1,\ldots,\kappa-1\}} + \left(\frac{1}{t^*} - v_0 - V_{\kappa-1} \right)^2 \frac{c_\kappa}{v_\kappa (v_0 + V_{\kappa-1})} = \Pi_\kappa(t^*) \leq Z_{\{1,\ldots,\kappa-1\}} + Z_{\{\kappa\}} \left(\frac{\frac{1}{t^*} - v_0 - V_{\kappa-1}}{v_\kappa} \right),$$

which in turn implies that

$$c_\kappa \leq Z_{\{\kappa\}} \frac{v_0 + V_{\kappa-1}}{\frac{1}{t^*} - v_0 - V_{\kappa-1}}. \tag{34}$$

Using (34) in Equation (32) we have

$$\begin{aligned} \Pi_\kappa(t^*) \leq & \; Z_{\{1,\ldots,\kappa\}} + (V_\kappa + v_0 - \tfrac{1}{t^*})^2 Z_{\{\kappa\}} \tfrac{v_0 + V_{\kappa-1}}{(\frac{1}{t^*} - v_0 - V_{\kappa-1}) v_\kappa (v_0 + V_\kappa)} \\ \leq & \; Z_{\{1,\ldots,\kappa\}} + (\tfrac{v_\kappa}{2})^2 Z_{\{\kappa\}} \tfrac{v_0 + V_{\kappa-1}}{(v_\kappa/2) v_\kappa (v_0 + V_\kappa)} \\ \leq & \; Z_{\{1,\ldots,\kappa\}} + Z_{\{\kappa\}} \tfrac{v_0 + V_{\kappa-1}}{2(v_0 + V_\kappa)} \\ \leq & \; \tfrac{3}{2} Z^{OPT}, \end{aligned}$$

where the second inequality uses the fact that $\frac{1}{t^*} - v_0 > V_{\kappa-1} + \frac{v_\kappa}{2}$. □

Proof of Lemma 6

Proof. Using Equation (16) in (31), we have

$$\Pi_\kappa(t^*) - Z_{\{1,\ldots,\kappa-1\}} = \left(\sqrt{\frac{\Delta_\kappa}{v_0 + V_{\kappa-1}}} - \sqrt{\frac{c_\kappa}{v_\kappa}(v_0 + V_{\kappa-1})} \right)^2. \tag{35}$$

Similarly, using Equation (16) in (32), we have

$$\Pi_\kappa(t^*) - Z_{\{1,\ldots,\kappa\}} = \left(\sqrt{\frac{c_\kappa}{v_\kappa}(v_0 + V_\kappa)} - \sqrt{\frac{\Delta_\kappa}{v_0 + V_\kappa}} \right)^2. \tag{36}$$

Now, if $Z_{\{1,\ldots,\kappa-1\}} > Z_{\{1,\ldots,\kappa\}}$, then $r_\kappa = \left(\sqrt{\frac{\Delta_\kappa}{v_0 + V_{\kappa-1}}} - \sqrt{\frac{c_\kappa}{v_\kappa}(v_0 + V_{\kappa-1})} \right)^2 / Z_{\{1,\ldots,\kappa-1\}}$. Since $\frac{1}{t^*} - v_0 \geq V_{\kappa-1}$, we have using (16) that $\sqrt{\frac{\Delta_\kappa v_\kappa}{c_\kappa}} = \frac{1}{t^*} \geq v_0 + V_{\kappa-1}$. It follows that

$$\left(\sqrt{\frac{\Delta_\kappa}{v_0 + V_{\kappa-1}}} - \sqrt{\frac{c_\kappa}{v_\kappa}(v_0 + V_{\kappa-1})} \right)^2$$

is a decreasing function of c_κ. Since $Z_{\{1,\ldots,\kappa-1\}}$ does not depend on c_κ, r_κ can be increased by decreasing c_κ.

On the other hand, if $Z_{\{1,\ldots,\kappa-1\}} > Z_{\{1,\ldots,\kappa\}}$, then $r_\kappa = \left(\sqrt{\frac{c_\kappa}{v_\kappa}(v_0 + V_\kappa)} - \sqrt{\frac{\Delta_\kappa}{v_0 + V_\kappa}} \right)^2 / Z_{\{1,\ldots,\kappa\}}$. Since $\frac{1}{t^*} - v_0 \leq V_\kappa$, $\sqrt{\frac{\Delta_\kappa v_\kappa}{c_\kappa}} = \frac{1}{t^*} \leq v_0 + V_\kappa$. Therefore, $\left(\sqrt{\frac{c_\kappa}{v_\kappa}(v_0 + V_\kappa)} - \sqrt{\frac{\Delta_\kappa}{v_0 + V_\kappa}} \right)^2$ is an increasing function of c_κ. Since $Z_{\{1,\ldots,\kappa\}}$ is decreasing in c_κ, it follows that r_κ is increasing in c_κ and can be increased by increasing c_κ.

Therefore, r_κ is maximal when $Z_{\{1,\ldots,\kappa-1\}} = Z_{\{1,\ldots,\kappa\}}$. □

Proof of Lemma 7

Proof. Since $t^* \in [\tau_\kappa, \tau_{\kappa-1}] \subset [\hat{t}_l, \hat{t}_u]$, we have $\rho_j(t^*)/v_j \geq \rho_\kappa(t^*)/v_\kappa > 0$ for all $j \in \{1, \ldots, \kappa-1\}$. Multiplying the jth inequality by $v_j/V_{\kappa-1}$, using (7) and adding up the $\kappa - 1$ inequalities have

$$\frac{\sum_{j=1}^{\kappa-1} p_j v_j}{V_{\kappa-1}} t^* - \frac{\sum_{j=1}^{\kappa-1} c_j}{\sum_{j=1}^{\kappa-1} v_j} \geq p_\kappa t^* - \frac{c_\kappa}{v_\kappa}.$$

Dividing both sides of the above inequality by t^*

$$\frac{\sum_{j=1}^{\kappa-1} p_j v_j}{V_{\kappa-1}} - \frac{\sum_{j=1}^{\kappa-1} c_j}{\sum_{j=1}^{\kappa-1} v_j} \frac{1}{t^*} \geq p_\kappa - \frac{c_\kappa}{v_\kappa} \frac{1}{t^*} = \frac{\Delta_\kappa + \sum_{j=1}^{k-1} p_j v_j}{v_0 + V_{\kappa-1}} - \frac{c_\kappa}{v_\kappa} \frac{1}{t^*}, \tag{37}$$

where the equality follows from the definition of Δ_κ. Using (18), we have $t^* = \sqrt{c_k/(\Delta_k v_k)} = \frac{1}{\sqrt{(v_0+V_{\kappa-1})(v_0+V_\kappa)}} = \sqrt{\tau_{\kappa-1}}\sqrt{\tau_\kappa}$. Equivalently $\frac{1}{t^*} = \sqrt{(v_0 + V_{\kappa-1})(v_0 + V_\kappa)}$. Using this in (37) and rearranging, we have

$$-\frac{\sum_{j=1}^{\kappa-1} c_j}{V_{\kappa-1}} \sqrt{(v_0 + V_{\kappa-1})(v_0 + V_\kappa)} \geq \frac{\Delta_\kappa}{v_0 + V_{\kappa-1}} \left(1 - \sqrt{\frac{v_0 + V_{\kappa-1}}{v_0 + V_\kappa}}\right) - \frac{v_0 \sum_{j=1}^{\kappa-1} p_j v_j}{(v_0 + V_{\kappa-1}) V_{\kappa-1}}. \tag{38}$$

We have

$$\begin{aligned} Z_{\{1,\ldots,\kappa-1\}} = \sum_{j=1}^{\kappa-1} \rho_j(V_{\kappa-1}) = \frac{\sum_{j=1}^{\kappa-1} p_j v_j}{v_0 + V_{\kappa-1}} - \sum_{j=1}^{\kappa-1} c_j &\geq \frac{\sum_{j=1}^{\kappa-1} p_j v_j}{v_0 + V_{\kappa-1}} - \sum_{j=1}^{\kappa-1} c_j \frac{\sqrt{(v_0 + V_{\kappa-1})(v_0 + V_\kappa)}}{v_0} \\ &\geq \frac{\Delta_\kappa V_{\kappa-1}}{v_0(v_0 + V_{\kappa-1})} \left(1 - \sqrt{\frac{v_0 + V_{\kappa-1}}{v_0 + V_\kappa}}\right) \end{aligned}$$

where the first equality uses that $V_{\kappa-1}, V_\kappa \geq 0$, and the last inequality follows from (38). □

Proof of Lemma 8

Proof. Given $(t_1, \ldots, t_D)$, we can partition the positive quadrant R_+^{D+E} into $O(n^{D+E})$ regions (polyhedrons) limited by the n hyperplanes $\sum_{d\in\mathcal{D}} p_{j,d} v_{j,d} t_d - \sum_{d\in\mathcal{D}} \lambda_d v_{j,d} - \sum_{e\in\mathcal{E}} \mu_d \alpha_{j,e} = c_j$. Consider one such region P

$$P = \left\{ \begin{pmatrix} \lambda_1, \ldots, \lambda_D, \\ \mu_1, \ldots, \mu_E \end{pmatrix} \geq 0 \;\middle|\; \begin{array}{l} \sum_{d\in\mathcal{D}} p_{j,d} v_{j,d} t_d - \sum_{d\in\mathcal{D}} \lambda_d v_{j,d} - \sum_{e\in\mathcal{E}} \mu_e \alpha_{j,e} \geq c_j \text{ for } j \in \mathcal{J}_P^+; \\ \sum_{d\in\mathcal{D}} p_{j,d} v_{j,d} t_d - \sum_{d\in\mathcal{D}} \lambda_d v_{j,d} - \sum_{e\in\mathcal{E}} \mu_e \alpha_{j,e} \leq c_j \text{ for } j \in \mathcal{J}_P^- \end{array} \right\},$$

where $\mathcal{J}_P^+$ and $\mathcal{J}_P^-$ are disjoint and $\mathcal{J}_P^+ \cup \mathcal{J}_P^- = \mathcal{J}$. Thus the objective in (23) can be written as

$$\sum_{d\in\mathcal{D}} \lambda_d \left(\frac{1}{t_d} - v_0\right) + \sum_{e\in\mathcal{E}} \mu_e \beta_e + \sum_{j\in\mathcal{J}_P^+} \left(\sum_{d\in\mathcal{D}} p_{j,d} v_{j,d} t_d - c_j - \sum_{d\in\mathcal{D}} \lambda_d v_{j,d} - \sum_{e\in\mathcal{E}} \mu_e \alpha_{j,e}\right). \tag{39}$$

Within P, (39) must be minimized at an extreme point, because the objective is linear in $(\lambda_1, \ldots, \lambda_D, \mu_1, \ldots, \mu_E)$. At this extreme point we have a total of $D+E$ constraints that are active and satisfied as equalities. Suppose for instance that a total of $D+E$ constraints in $\mathcal{J}_P^+ \cup \mathcal{J}_P^-$ are active and without loss of generality assume that they are numbered consecutively so that $\{1, \ldots, D+E\}$ represents the set of constraints which define the particular extreme point. As a result we can obtain $(\lambda_1, \ldots, \lambda_D, \mu_1, \ldots, \mu_E)$ by solving

$$\begin{pmatrix} v_{1,1} & \cdots & v_{1,D} & \alpha_{1,1} & \cdots & \alpha_{1,E} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ v_{D+E,1} & \cdots & v_{D+E,D} & \alpha_{D+E,1} & \cdots & \alpha_{D+E,E} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_D \\ \mu_1 \\ \vdots \\ \mu_E \end{pmatrix}$$
$$= \begin{pmatrix} p_{1,1}v_{1,1} & \cdots & p_{1,D}v_{1,D} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ p_{D+E,1}v_{D+E,1} & \cdots & p_{D+E,D}v_{D+E,D} \end{pmatrix} \begin{pmatrix} t_1 \\ \vdots \\ t_D \end{pmatrix} - \begin{pmatrix} c_1 \\ \vdots \\ \vdots \\ c_{D+E} \end{pmatrix}.$$

We can write the above equations in matrix form as $A \begin{pmatrix} \vec{\lambda} \\ \vec{\mu} \end{pmatrix} = B\vec{t} - \vec{c}$, so $\begin{pmatrix} \vec{\lambda} \\ \vec{\mu} \end{pmatrix} = A^{-1}B\vec{t} - A^{-1}\vec{c}$. Therefore, we can express $(\lambda_1, \ldots, \lambda_D, \mu_1, \ldots, \mu_E)$ as a linear function of $(t_1, \ldots, t_D)$.

Hence, at this vertex of P, (39) is equal to

$$G_\kappa(t_1, \ldots, t_D) = \left(\frac{1}{t_1} - \hat{v}_{01} \ldots \frac{1}{t_D} - \hat{v}_{0D}\ \hat{\beta}_1 \ldots \hat{\beta}_E\right)\left(A^{-1}B\vec{t} - A^{-1}\vec{c}\right) + \sum_{j \in \mathcal{J}_P^+}\left(\sum_{d \in \mathcal{D}} p_{j,d}v_{j,d}t_d - c_j\right)$$

where $\hat{v}_{0,d} = v_0 + \sum_{j \in \mathcal{J}_P^+} v_{j,d}$ and $\hat{\beta}_e = \beta_e - \sum_{j \in \mathcal{J}_P^+} \alpha_{j,e}$. This is the form expressed in Equation (25). □

Proof of Lemma 10

Proof. Substituting in Equation (25) $\frac{1}{t_d}$ by $\frac{1}{t_1} + w_d$ yields Equation (26). So $\max_{t_1 \in [t_{1,min}, t_{1,max}]} \Gamma^f(t_1)$ can be obtained either at an interior solution where $\Gamma^f(t_1) = G_\kappa(t_1) < \min_{\kappa' \neq \kappa} G_{\kappa'}(t_1)$, or at a point where $G_{\kappa_1}(t_1) = G_{\kappa_2}(t_1)$.

In the first case, t_1 is a local maximizer of $G_\kappa(\cdot)$: it must satisfy the first-order condition

$$0 = \tilde{\xi}_\kappa - \sum_{d \in \mathcal{D}} \frac{\tilde{\chi}_{\kappa,d}}{(t_1 + w_d)^2}.$$

This can be expressed as

$$P(t_1) = \tilde{\xi}_\kappa \prod_{d \in \mathcal{D}} (t_1 + w_d)^2 - \sum_{d \in \mathcal{D}} \tilde{\chi}_{\kappa,d} \prod_{d' \neq d} (t_1 + w_{d'})^2 = 0$$

where P is a polynomial equation of degree $2D$. From algebra's fundamental theorem, it has exactly $2D$ complex roots c_l, $l \in \{1, \ldots, 2D\}$. Hence P can be written as a constant times

$\prod_{l=1}^{2D}(t_1 - c_l)$. If a root c_l is not real (the complex dimension is non-zero), then its conjugate $\bar{c}_l$ must also be a root, so $(t_1 - c_l)(t_1 - \bar{c}_l) > 0$ for all $t_1 \in R$. Hence there are at most $2D$ real roots to $P(t_1) = 0$. Consider a real root c_l with odd degree: then at $t_1 = c_l$, the sign of P changes, which means that if this is a maximum of $G_\kappa(\cdot)$ (sign changes from positive to negative) then the next root is a minimum. On the other hand if a real root has even degree, then it is a saddle point that is neither a maximum nor a minimum of $G_\kappa(\cdot)$. So half of such roots cannot be maxima and there are at most D maxima to consider. For example if $P(t_1) = (t_1 - 1)(t_1 - 2)(t_1 - 3)^2$, then root 1 is a maximum, root 2 is a minimum and root 3 (double degree) is a saddle point. In this example $D = 2$ and there are at most 2 maxima (there is 1 in reality).

In the second case,

$$(\tilde{\xi}_{\kappa_2} - \tilde{\xi}_{\kappa_1})t_1 + \sum_{d \in \mathcal{D}} (\tilde{\chi}_{\kappa_2,d} - \tilde{\chi}_{\kappa_1,d}) \frac{1}{t_1 + w_d} + (\tilde{\phi}_{\kappa_2} - \tilde{\phi}_{\kappa_1}) = 0$$

or in other words

$$(\tilde{\xi}_{\kappa_2} - \tilde{\xi}_{\kappa_1})t_1 \prod_{d \in \mathcal{D}} (t_1 + w_d) + (\tilde{\phi}_{\kappa_2} - \tilde{\phi}_{\kappa_1}) \prod_{d \in \mathcal{D}} (t_1 + w_d) + \sum_{d \in \mathcal{D}} (\tilde{\chi}_{\kappa_2,d} - \tilde{\chi}_{\kappa_1,d}) \prod_{d' \neq d} (t_1 + w_{d'}) = 0.$$

Again, from algebra's fundamental theorem, there are at most $D + 1$ real roots to this equation. □

Proof of Lemma 11

Proof. Let $\mathcal{F}_j = \{i \,|\, \hat{x}_i = \hat{x}_j\}$, $\bar{v}_j = \sum_{i \in \mathcal{F}_j} v_i$, $\bar{p}_j = \sum_{i \in \mathcal{F}_j} p_i v_i / \bar{v}_j$ and $\bar{c}_j = \sum_{i \in \mathcal{F}_j} c_j$. We define $\mathcal{F}_k, \bar{v}_k, \bar{p}_k$ and $\bar{c}_k$ in an analogous manner.

Suppose $\hat{x}_j \neq \hat{x}_k$. Therefore $\mathcal{F}_j \cap \mathcal{F}_k = \emptyset$. We construct a solution $\tilde{x}$ whose objective function value is strictly larger than that of $\hat{x}$, thus yielding a contradiction. Suppose

$$\bar{p}_j t - \bar{c}_j / \bar{v}_j < \bar{p}_k t - \bar{c}_k / \bar{v}_k. \tag{40}$$

We set $\tilde{x}_i = \hat{x}_i + \epsilon$ for all $i \in \mathcal{F}_k$, $\tilde{x}_i = \hat{x}_i - (\bar{v}_k / \bar{v}_j)\epsilon$ for all $i \in \mathcal{F}_j$ and $\tilde{x}_i = \hat{x}_i$ for all $i \notin \mathcal{F}_j \cup \mathcal{F}_k$. It follows that for sufficiently small $\epsilon > 0$, $\tilde{x}$ is a feasible solution to $\Gamma^f(t)$. In particular, constraint (11) is satisfied by $\tilde{x}$. All the binding precedence constraints at $\hat{x}$ continue to remain binding at $\tilde{x}$. Moreover, since $\mathcal{F}_j \cap \mathcal{F}_k = \emptyset$ none of the nonbinding precedence constraints become binding for sufficiently small ϵ. On the other hand, we have

$$\begin{aligned}
&\sum_{i \in (\mathcal{F}_j \cup \mathcal{F}_k)^c} \Big[p_i v_i t - c_i\Big] \tilde{x}_i + \sum_{i \in \mathcal{F}_j} \Big[p_i v_i t - c_i\Big] \tilde{x}_j + \sum_{i \in \mathcal{F}_k} \Big[p_i v_i t - c_i\Big] \tilde{x}_k \\
> &\sum_{i \in (\mathcal{F}_j \cup \mathcal{F}_k)^c} \Big[p_i v_i t - c_i\Big] \hat{x}_i + \sum_{i \in \mathcal{F}_j} \Big[p_i v_i t - c_i\Big] \hat{x}_j + \sum_{i \in \mathcal{F}_k} \Big[p_i v_i t - c_i\Big] \hat{x}_k
\end{aligned}$$

where the inequality follows from (40). This contradicts $\hat{x}$ being an optimal solution to $\Gamma^f(t)$. The case where $\bar{p}_j t - \bar{c}_j / \bar{v}_j > \bar{p}_k t - \bar{c}_k / \bar{v}_k$ is symmetric and can be handled in a similar manner. By perturbing the profit margins and the costs by infinitesimally small amounts if required, we can assume that the case where the two terms are equal does not occur. □

Proof of Lemma 12

Proof. It can be verified that $\mathcal{F}$ has at least one maximal element. We show that there is exactly one maximal element. By Lemma 11, we have $\hat{x}_j = \hat{x}_k$ for all $j, k \in \mathcal{F}$. Now suppose that we have $j, k \in \mathcal{F}$ such that both are maximal. Since j is maximal and $k \in \mathcal{F}$, we have $j \notin \mathcal{O}_k$. By a similar argument, we have $k \notin \mathcal{O}_j$. Since the precedence constraints are nested, this means that $\mathcal{O}_j \cap \mathcal{O}_k = \emptyset$. Let $\mathcal{F}_j = \{i \in \mathcal{O}_j | \hat{x}_i = \hat{x}_j\}$ and $\mathcal{F}_k = \{i \in \mathcal{O}_k | \hat{x}_i = \hat{x}_k\}$. Since $\mathcal{O}_j \cap \mathcal{O}_k = \emptyset$, we have $\mathcal{F}_j \cap \mathcal{F}_k = \emptyset$. By following the same steps as in the proof of Lemma 11 we can construct a solution $\tilde{x}$ which strictly improves the objective function value, thus yielding a contradiction. □

Proof of Corollary 1

Proof. The first part of the corollary follows by noting that if $j \in \mathcal{F}$ but $j \notin \mathcal{O}_i$, it implies that i is not the unique maximal element of $\mathcal{F}$, contradicting Lemma 12.

Since $\mathcal{F} \subset \mathcal{O}_i$ and $\mathcal{F}$ contains all the products assigned fractional values, we have $\hat{x}_j \in \{0, 1\}$ for $j \in \mathcal{O}_i^c$. This proves the second statement.

If j is such that $i \in \mathcal{O}_j$, then $\mathcal{O}_i \subset \mathcal{O}_j$. The second part of the corollary implies that $\hat{x}_j \in \{0, 1\}$. But the constraint $\hat{x}_j - \hat{x}_i \leq 0$ together with $0 < \hat{x}_i < 1$ implies that $\hat{x}_j = 0$, proving the third statement.

Coming to the next statement, if $j \in \mathcal{O}_i \backslash \mathcal{F}$, then $\hat{x}_i - \hat{x}_j < 0$, which implies that $\hat{x}_j = 1$.

The last statement follows by noting that if $\sum_{j \in \mathcal{F}} \rho_j(t) = \sum_{j \in \mathcal{F}} [p_j v_j t - c_j / v_j] < 0$, then we can construct a solution $\tilde{x}$ which obtains a larger objective function value that $\hat{x}$ in the following manner. We set $\tilde{x}_j = \hat{x}_j - \epsilon$ for all $j \in \mathcal{F}$, where $\epsilon > 0$ is sufficiently small. We set $\tilde{x}_j = \hat{x}_j$ for all $j \in \mathcal{F}^c$. It can be verified that the solution $\tilde{x}$ is feasible and improves on the objective function value compared to $\hat{x}$ yielding a contradiction. □

Online Appendix: Parametric Bounds for the Intervals $\mathcal{I}_{\kappa_l}$ and $\mathcal{I}_{\kappa_u+1}$

We bound the gap between $\max_{t\in\mathcal{I}_\kappa}\Pi_\kappa(t)$ and Z^{OPT} for $\kappa\in\{\kappa_l,\kappa_u+1\}$. All the results from §4.2.4 carry over if $\Delta_\kappa\le 0$. So we focus on the case $\Delta_\kappa>0$. Let

$$\bar{r}_\kappa=\min\left\{\frac{\Pi_\kappa(t^*)}{Z_{\{1,\ldots,\kappa-1\}}},\frac{\Pi_\kappa(t^*)}{Z_{\{1,\ldots,\kappa\}}}\right\}-1 \tag{41}$$

where t^* is the unconstrained maximizer of $\Pi_\kappa(t)$ (Lemma 2). Since $\max_{t\in\mathcal{I}_\kappa}\Pi_\kappa(t)\le\Pi_\kappa(t^*)$, we have $r_\kappa\le\bar{r}_\kappa$. We establish a bound on $\bar{r}_\kappa$, which also bounds r_κ and thus the optimality gap. Moreover, just like in §4.2.4, when bounding $\bar{r}_\kappa$ we restrict attention to the case where $\bar{r}_\kappa$ is maximal. It is possible to show that $\bar{r}_\kappa$ is maximal when $Z_{\{1,\ldots,\kappa_-1\}}=Z_{\{1,\ldots,\kappa\}}$. In this case, we have $t^*=\sqrt{\tau_{\kappa-1}}\sqrt{\tau_\kappa}=\frac{1}{\sqrt{v_0+V_{\kappa-1}}}\frac{1}{\sqrt{v_0+V_\kappa}}$ and

$$\bar{r}_\kappa=\frac{\Delta_\kappa\left(\sqrt{v_0+V_\kappa}-\sqrt{v_0+V_{\kappa-1}}\right)^2}{(v_0+V_{\kappa-1})(v_0+V_\kappa)Z_{\{1,\ldots,\kappa-1\}}}. \tag{42}$$

Analysis for the interval $\mathcal{I}_{\kappa_l}$

We first consider the interval $\mathcal{I}_{\kappa_l}=[\hat{t}_l,\tau_{\kappa_l-1}]$. If $t^*\ge\hat{t}_l$, then it can be verified that the bound described in Proposition 2 applies. So we consider the case when $t^*<\hat{t}_l$. We have the following lower bound on $Z_{\{1,\ldots,\kappa_l-1\}}$.

Lemma 13. *If $\Delta_{\kappa_l}>0$ and $t^*<\hat{t}_l$, then*

$$Z_{\{1,\ldots,\kappa_l-1\}}\ge\frac{\Delta_{\kappa_l}V_{\kappa_l-1}}{v_0(v_0+V_{\kappa_l-1})}\left(1-\frac{t^*}{\hat{t}_l}\sqrt{\frac{v_0+V_{\kappa_l-1}}{v_0+V_{\kappa_l}}}\right).$$

Proof. The proof follows by noting that $\rho_j(\hat{t}_l)/v_j\ge\rho_{\kappa_l}(\hat{t}_l)/v_{\kappa_l}$ for all $j\in\{1,\ldots,\kappa_l-1\}$ and using the same arguments as in the proof of Lemma 7. □

Using the lower bound from Lemma 13 in Equation (42), we have the following bound on $\bar{r}_{\kappa_l}$. Since $r_{\kappa_l}\le\bar{r}_{\kappa_l}$, Proposition 6 implies a bound on r_{κ_l} as well when $\Delta_{\kappa_l}>0$ and $t^*<\hat{t}_l$. Since $t^*<\hat{t}_l$, the bound in Proposition 6 is in fact stronger than that in Proposition 2.

Proposition 6. *If $\Delta_{\kappa_l}>0$ and $t^*<\hat{t}_l$ then*

$$\bar{r}_{\kappa_l}\le\frac{v_0v_{\kappa_l}}{V_{\kappa_l-1}\sqrt{v_0+V_{\kappa_l}}(\sqrt{v_0+V_{\kappa_l}}+\sqrt{v_0+V_{\kappa_l-1}})}\left(\frac{1-\sqrt{\frac{v_0+V_{\kappa_l-1}}{v_0+V_{\kappa_l}}}}{1-\frac{t^*}{\hat{t}_l}\sqrt{\frac{v_0+V_{\kappa_l-1}}{v_0+V_{\kappa_l}}}}\right).$$

To summarize the analysis for the interval $\mathcal{I}_{\kappa_l}$, if $\Delta_{\kappa_l}>0$ and $t^*<\hat{t}_l$, then we bound r_{κ_l} using Proposition 6. Otherwise, the bounds in §4.2.4 apply.

Analysis for the interval $\mathcal{I}_{\kappa_u+1}$

Now we consider the interval $\mathcal{I}_{\kappa_u+1} = [\tau_{\kappa_u+1}, \hat{t}_u]$. It can be verified that the bound described in Proposition 2 continues to apply if $t^* \leq \hat{t}_u$. So we consider the case when $t^* > \hat{t}_u$. We have the following analog of Lemma 13.

Lemma 14. *If $\Delta_{\kappa_u+1} > 0$ and $t^* > \hat{t}_u$, then*

$$Z_{\{1,\ldots,\kappa_u\}} \geq \frac{\Delta_{\kappa_u+1} V_{\kappa_u}}{v_0(v_0+V_{\kappa_u})} \left(1 - \frac{t^*}{\hat{t}_u}\sqrt{\frac{v_0+V_{\kappa_u}}{v_0+V_{\kappa_u+1}}}\right). \tag{43}$$

Lemma 14 together with Equation (42) implies the following proposition, which is the direct counterpart of Proposition 6.

Proposition 7. *If $\Delta_{\kappa_u+1} > 0$ and $t^* > \hat{t}_u$, then*

$$\bar{r}_{\kappa_u+1} \leq \frac{v_0 v_{\kappa_u+1}}{V_{\kappa_u}\sqrt{v_0+V_{\kappa_u+1}}(\sqrt{v_0+V_{\kappa_u+1}}+\sqrt{v_0+V_{\kappa_u}})} \left(\frac{1-\sqrt{\frac{v_0+V_{\kappa_u}}{v_0+V_{\kappa_u+1}}}}{1-\frac{t^*}{\hat{t}_u}\sqrt{\frac{v_0+V_{\kappa_u}}{v_0+V_{\kappa_u+1}}}}\right).$$

Since $r_{\kappa_u+1} \leq \bar{r}_{\kappa_u+1}$, Proposition 7 bounds r_{κ_u+1} as well when $\Delta_{\kappa_u+1} > 0$ and $t^* > \hat{t}_u$. Since $t^* > \hat{t}_u$, $t^*/\hat{t}_u > 1$. As the ratio $t^*/\hat{t}_u$ gets close to 1, the bound in Proposition 7 tends to that in Proposition 2. However, the bound in Proposition 7 can be quite loose if the ratio $t^*/\hat{t}_u$ is large. We provide an alternative bound on r_{κ_u+1} that may be more useful in such situations.

We continue to focus on the case where $\Delta_{\kappa_u+1} > 0$, $\bar{r}_{\kappa_u+1}$ is maximal and $t^* > \hat{t}_u$. Since $\Delta_{\kappa_u+1} > 0$, $\Pi_{\kappa_u+1}(t)$ is concave (Lemma 2). Using the subgradient inequality

$$\Pi_{\kappa_u+1}(\hat{t}_u) \leq Z_{\{1,\ldots,\kappa_u+1\}} + \frac{\Delta_{\kappa_u+1}(V_{\kappa_u+1}+v_0-\frac{1}{\hat{t}_u})v_{\kappa_u+1}}{(v_0+V_{\kappa_u+1})^2(v_0+V_{\kappa_u})}. \tag{44}$$

Since $\Pi_{\kappa_u+1}(t)$ is concave and $t^* > \hat{t}_u$, the function is increasing on $\mathcal{I}_{\kappa_u+1}$ and $\max_{t\in\mathcal{I}_{\kappa_u+1}} \Pi_{\kappa_u+1}(t) = \Pi_{\kappa_u+1}(\hat{t}_u)$. This together with $Z_{\{1,\ldots,\kappa_u\}} = Z_{\{1,\ldots,\kappa_u+1\}}$ (since $\bar{r}_{\kappa_u+1}$ is maximal) implies

$$r_{\kappa_u+1} = \frac{\Pi_{\kappa_u+1}(\hat{t}_u) - Z_{\{1,\ldots,\kappa_u+1\}}}{Z_{\{1,\ldots,\kappa_u\}}}. \tag{45}$$

Using (44) and (43) in Equation (45), we obtain an alternative bound on r_{κ_u+1} given in Proposition 8 below. We expect the bound in Proposition 8 to dominate that in Proposition 7 when $t^*/\hat{t}_u$ is large.

Proposition 8. *If $\Delta_{\kappa_u+1} > 0$ and $t^* > \hat{t}_u$, then*

$$r_{\kappa_u+1} \leq \frac{v_0 v_{\kappa_u+1}}{(v_0+V_{\kappa_u+1})V_{\kappa_u}}.$$

To summarize the analysis for the interval $\mathcal{I}_{\kappa_u+1}$, if $\Delta_{\kappa_u+1} > 0$ and $t^* > \hat{t}_u$, then we use the minimum of the bounds in Propositions 7 and 8 as our bound on r_{κ_u+1}. Otherwise, the bounds in §4.2.4 apply.