Assortment Optimization under the Paired Combinatorial Logit Model

Heng Zhang, Paat Rusmevichientong
Marshall School of Business, University of Southern California, Los Angeles, CA 90089
hengz@usc.edu, rusmevic@marshall.usc.edu

Huseyin Topaloglu
School of Operations Research and Information Engineering, Cornell Tech, New York, NY 10044
topaloglu@orie.cornell.edu

August 14, 2019

We consider uncapacitated and capacitated assortment problems under the paired combinatorial logit model, where the goal is to find a set of products to offer to maximize the expected revenue obtained from a customer. In the uncapacitated setting, we can offer any set of products, whereas in the capacitated setting, there is an upper bound on the number of products that we can offer. We establish that even the uncapacitated assortment problem is strongly NP-hard. To develop an approximation framework for our assortment problems, we transform the assortment problem into an equivalent problem of finding the fixed point of a function, but computing the value of this function at any point requires solving a nonlinear integer program. Using a suitable linear programming relaxation of the nonlinear integer program and randomized rounding, we obtain a 0.6-approximation algorithm for the uncapacitated assortment problem. Using randomized rounding on a semidefinite programming relaxation, we obtain an improved 0.79-approximation algorithm, but the semidefinite programming relaxation can get difficult to solve in practice for large problem instances. Finally, using iterative rounding, we obtain a 0.25-approximation algorithm for the capacitated assortment problem. Our computational experiments on randomly generated problem instances demonstrate that our approximation algorithms, on average, yield expected revenues that are within 1.1% of an efficiently computable upper bound on the optimal expected revenue.

Key words: Customer choice modeling, paired combinatorial logit model, assortment optimization.

---

1. Introduction

Traditional revenue management models commonly assume that each customer arrives into the system with the intention to purchase a particular product. If this product is available for purchase, then the customer purchases it; otherwise, the customer leaves the system without a purchase. In reality, however, customers observe the set of available alternatives and make a choice among the available alternatives. Under such a customer choice process, the demand for a particular product depends on the availability of other products. In this case, discrete choice models provide a useful representation of demand since discrete choice models capture the demand for each product as a function of the entire set of products in the offer set. A growing body of literature indicates that capturing the choice process of customers using discrete choice models can significantly improve the quality of operational decisions; see, for example, Talluri and van Ryzin (2004), Gallego et al. (2004) and Vulcano et al. (2010). While more refined choice models yield a more accurate representation of the customer choice process, the assortment and other operational problems under more refined choice models become more challenging. Thus, it is useful to identify realistic choice models, where the corresponding operational problems remain efficiently solvable.

In this paper, we study assortment problems under the paired combinatorial logit (PCL) model. There is a fixed revenue for each product. Customers choose among the offered products according to the PCL model, which is discussed in Section 2. The goal is to find an offer set that maximizes the expected revenue obtained from a customer. We consider both the uncapacitated version, where we can offer any subset of products, as well as the capacitated version, where there is an upper bound on the number of products that we can offer. We show that even the uncapacitated assortment problem is strongly NP-hard. We give a framework for constructing approximation algorithms for the assortment problem. We use this framework to develop approximation algorithms for the uncapacitated and capacitated versions. Our computational experiments on randomly generated problem instances demonstrate that our approximation algorithms perform quite well, yielding solutions with optimality gaps under 1.1% on average.

The PCL model is compatible with random utility maximization, where each customer associates random utilities with the alternatives. The utilities are sampled from a certain distribution. The customer knows the utilities and chooses the alternative that provides the largest utility. Other choice models, such as the multinomial logit, nested logit and a mixture of multinomial logit models, are also compatible with random utility maximization. The PCL model allows certain correlations between the utilities of any pair of alternatives. In contrast, the multinomial logit model assumes that the utilities are independent. In the nested logit model, the alternatives are grouped into nests. There is a single parameter governing the correlation between the utilities of the alternatives in the same nest. The utilities of the alternatives in different nests are independent. A mixture of multinomial logit models allows general correlations, but it presents difficulties in solving the corresponding assortment problem, as discussed in our literature review.

By allowing correlations between the utilities of any pair of alternatives, the PCL model captures the situation where the preference of a customer for one product offers insight into their inclination towards another product. The work in the existing literature shows that there can be significant correlations between the utilities of the alternatives when passengers choose, for example, among travel modes and paths. When such correlations between the utilities of the alternatives are present, it turns out that the PCL model can provide better predictions of the demand process of the passengers; see, for example, Prashker and Bekhor (1998), Koppleman and Wen (2000) and Chen et al. (2003). However, although the PCL model can provide better predictions of the demand process when there are correlations between the utilities of different alternatives, as far as we are aware, there is little research on understanding the complexity of making operational decisions under the PCL model and providing efficient algorithms for making such decisions. Our work in this paper is directed towards filling this gap.

1.1. Main Contributions

We make four main contributions. First, we show that the uncapacitated assortment problem under the PCL model is strongly NP-hard. Our proof uses a reduction from the max-cut problem. This result is in contrast with the assortment problem under the closely-related multinomial logit and nested logit models. In particular, as we discuss in our literature review in Section 1.2, there exist polynomial-time algorithms to solve even the capacitated assortment problem under the multinomial logit model. Similarly, there exist polynomial-time algorithms to solve the capacitated assortment problem under the nested logit model, as long as a customer deciding to make a purchase in one of the nests cannot leave without making a purchase and the so-called dissimilarity parameters in the nested logit model do not exceed one. The latter condition ensures the compatibility of the nested logit model with random utility maximization.

Second, we give a framework to develop approximation algorithms for assortment problems under the PCL model. In particular, we show that the assortment problem is equivalent to finding the fixed point of a function $f:\mathbb{R}\rightarrow\mathbb{R}$, whose evaluation at a certain point requires solving a nonlinear integer program. We design an upper bound $f^R:\mathbb{R}\rightarrow\mathbb{R}$ to $f(\cdot)$ and compute the fixed point $\hat{z}$ of $f^R(\cdot)$. In this case, we develop an algorithm to find an α-approximate solution for the nonlinear integer program that computes the value of $f(\cdot)$ at $\hat{z}$. This α-approximate solution turns out to be an α-approximate solution to the assortment problem as well. Davis et al. (2014), Gallego and Topaloglu (2014), and Feldman and Topaloglu (2015) use a connection in this spirit between the assortment problem under the nested logit model and the fixed point of a function, but the functions they work with are significantly simpler as they are separable by the nests.

Third, we use our approximation framework to give an approximation algorithm for the uncapacitated assortment problem. We construct the upper bound $f^R(\cdot)$ through a linear programming (LP) relaxation of the nonlinear integer program that computes $f(\cdot)$. We show that we can compute the fixed point of $f^R(\cdot)$ by solving an LP. After we compute the fixed point $\hat{z}$ of $f^R(\cdot)$, we get a solution to the LP that computes $f^R(\cdot)$ at $\hat{z}$. We use randomized rounding on this solution to get a 0.6-approximate solution to the nonlinear integer program that computes $f(\cdot)$ at

$\hat{z}$. Lastly, to obtain a deterministic algorithm, we de-randomize this approach by using the standard method of conditional expectations; see Section 15.1 in Alon and Spencer (2000) and Section 5.2 in Williamson and Shmoys (2011). Our framework can allow other approximation algorithms. For example, if we construct $f^R(\cdot)$ by using a semidefinite programming (SDP) relaxation of the nonlinear integer program, then we can use the spherical rounding method of Goemans and Willamson (1995) to obtain a 0.79-approximation algorithm for the uncapacitated assortment problem. This approximation algorithm requires solving an SDP. We can theoretically solve an SDP in polynomial time, but solving a large-scale SDP in practice can be computationally difficult. Thus, the SDP relaxation can be less appealing than the LP relaxation.

Fourth, we give an approximation algorithm for the capacitated assortment problem, also by using our approximation framework. Here, we exploit the structural properties of the extreme points of the LP relaxation and use an iterative rounding method, followed by coupled randomized rounding, to develop a 0.25-approximation algorithm. In this algorithm, if there are n products that can be offered to the customers, then we solve at most n successive LP relaxations, fixing the value of one decision variable after solving each LP relaxation. Once we solve the LP relaxations, we perform coupled randomized rounding on the solution of the last LP relaxation to obtain a solution to the assortment problem. Using the method of conditional expectations, we can de-randomize this solution to obtain a deterministic algorithm with the same performance guarantee.

In our computational experiments, the practical performance of our approximation algorithms is substantially better than their theoretical guarantees, yielding, on average, expected revenues within 1.1% of an upper bound on the optimal expected revenue. We also test the ability of the PCL model to predict the choices of the customers. On average, compared with the multinomial logit benchmark, we get a 6.38% reduction in the errors of the predicted choice probabilities.

1.2. Literature Review

Considering operational decisions under the PCL model, prior to ours, the only work we are aware of is Li and Webster (2017), where the authors study pricing problems under the PCL model. The authors give sufficient conditions for the price sensitivities of the products to ensure that the pricing problem can be solved efficiently. Subsequent to our work, Ghuge et al. (2019) give approximation algorithms for constrained assortment problems under the PCL model. Despite limited work on solving operational problems under the PCL model, there is considerable work, especially in the transportation literature, on using the PCL model to capture travel mode and path choices. Koppleman and Wen (2000) estimate the parameters of the PCL model by using real data on the travel mode choices of the passengers. Their empirical results indicate that there are statistically significant correlations between the utilities that a passenger associates with different travel modes. Prashker and Bekhor (1998) and Chen et al. (2003) use numerical examples to demonstrate that the PCL model can provide improvements over the multinomial logit model in predicting path choices. The authors argue that different paths overlap with each other to varying extents, creating complex correlations between the utilities provided by different paths. Chen et al. (2014) and Karoonsoontawong and Lin (2015) study various traffic equilibrium problems under the PCL model and discuss the benefits from using this choice model. The numerical work in the transportation literature demonstrates that the PCL model can provide improvements over the multinomial logit and nested logit models in predicting travel mode and path choices, especially when the utilities provided by different alternatives exhibit complex correlation structures.

There is considerable work on assortment problems under the multinomial logit and nested logit models. In the multinomial logit model, the utilities of the products are independent of each other. In the nested logit model, the products are grouped into disjoint nests. Associated with each nest, there is a dissimilarity parameter characterizing the correlation between the utilities of the products in the same nest, but the utilities of the products in different nests are independent of each other. As shown by Daganzo and Kusnic (1993), the utilities of the products in the same nest have the same correlation coefficient. In the PCL model, there exists one nest for each pair of products, so the nests are overlapping. Associated with each nest, there is a separate dissimilarity parameter characterizing the correlation between the utilities of each pair of products. Therefore, when compared with the multinomial logit and nested logit models, we can use the PCL model to specify a significantly more general correlation structure between the utilities of the products. Train (2002) provides a thorough discussion of the multinomial logit, nested logit and PCL models. Koppleman and Wen (2000) discuss the correlation structure of the utilities under the PCL model, including the joint distributions and the correlation coefficients of the utilities.

Talluri and van Ryzin (2004) and Gallego et al. (2004) give an efficient algorithm for the uncapacitated assortment problem under the multinomial logit model, whereas Rusmevichientong et al. (2010) give an efficient algorithm for the capacitated version. Davis et al. (2013) formulate an LP to solve the assortment problem under the multinomial logit model when there are constraints on the offered assortment that can be represented through a totally unimodular constraint structure. In a mixture of multinomial logit models, there are multiple customer types and customers of different types choose according to different multinomial logit models. McFadden and Train (2000) show that a mixture of a multinomial logit models can allow arbitrary correlations between the utilities of the alternatives. Bront et al. (2009) show that the assortment problem under this choice model is NP-hard. In their proof, the authors use a reduction from the minimum vertex cover problem, in which case, they end up with an assortment problem where the numbers of products and customer types are as large as the number of vertices in the minimum vertex cover problem. This result raises the question of whether the problem is still NP-hard when the number of customer types is fixed. Rusmevichientong et al. (2014) show that the problem is NP-hard even when the number of customer types is fixed at two. Letting m be the number of customer types, Desire and Goyal (2016) show that there is no polynomial-time algorithm with an approximation guarantee of $O(1/m^{1-\delta})$ for any constant $\delta > 0$ unless $NP \subseteq BPP$.

Davis et al. (2014) categorize the assortment problem under the nested logit model along two dimensions, which are based on whether the dissimilarity parameter for a nest exceeds one and whether the no purchase option is available within the nests. Considering the case where none of the dissimilarity parameters exceeds one and the no purchase option is not available within the nests, the authors develop an efficient algorithm to find the optimal assortment for the unconstrained version of the problem. In other cases, the authors show that the assortment problem is NP-hard. Most of the other literature on the assortment problem under the nested logit model focuses on the case where none of the dissimilarity parameters exceeds one and the no purchase option is not available within the nests. Under this case, Gallego and Topaloglu (2014) study the assortment problem when a capacity constraint limits the total number of products offered in each nest and give an efficient algorithm to compute the optimal assortment. They also give a 0.5-approximation algorithm when each product occupies a certain amount of space and there is a constraint that limits the total space consumption of the products offered in each nest. In contrast, Feldman and Topaloglu (2015) give an efficient algorithm to compute the optimal assortment when a capacity constraint limits the total number of products offered in all nests. They also give a 0.25-approximation algorithm when there is a constraint that limits the total space consumption of the products offered in all nests. Li et al. (2015) give an efficient algorithm to compute the optimal assortment under the multi-level nested logit model, where the products are hierarchically organized into nests and subnests. Our paper complements all this work on assortment problems under the multinomial logit and nested logit models by working with the PCL model. Vulcano et al. (2010) and Dai et al. (2014) use the multinomial and nested logit models in airline applications.

The paper is organized as follows. In Section 2, we formulate the uncapacitated and capacitated assortment problems and show that even the uncapacitated version is strongly NP-hard. In Section 3, we give our approximation framework. In Section 4, we use our approximation framework to give a 0.6-approximation algorithm for the uncapacitated problem. Also, we discuss how to obtain a 0.79-approximation algorithm by using an SDP relaxation. In Section 5, we use our approximation framework to give a 0.25-approximation algorithm for the capacitated problem. In Section 6, we give our computational experiments. In Section 7, we conclude.

2. Assortment Problem

In this section, we formulate the assortment problem under the PCL model, characterize its complexity and discuss the use of the PCL model in assortment problems.

2.1. Problem Formulation and Complexity

The set of products is indexed by $N = \{1,\dots,n\}$. We use the vector $\boldsymbol{x} = (x_1,\dots,x_n) \in \{0,1\}^n$ to capture the subset of products that we offer to the customers, where $x_i = 1$ if and only if we offer product i. We refer to the vector $\boldsymbol{x}$ simply as the assortment or the subset of products that we offer. Throughout the paper, we denote the vectors and matrices in bold font. We denote the collection of nests by $M = \{(i,j) \in N^2 : i \neq j\}$. For each nest $(i,j) \in M$, we let $\gamma_{ij} \in [0,1]$ be the dissimilarity parameter of the nest. For each product i, we let v_i be the preference weight of product i. Under the PCL model, we can view the choice process of a customer as taking place in two stages. First, the customer decides to make a purchase in one of the nests or to leave without a purchase. In particular, letting $V_{ij}(\boldsymbol{x}) = v_i^{1/\gamma_{ij}}\, x_i + v_j^{1/\gamma_{ij}}\, x_j$ and using $v_0 > 0$ to denote the preference weight of the no purchase option, if we offer the subset of products $\boldsymbol{x}$, then a customer decides to make a purchase in nest (i,j) with probability $P_{ij}(\boldsymbol{x}) = V_{ij}(\boldsymbol{x})^{\gamma_{ij}} / (v_0 + \sum_{(k,\ell)\in M} V_{k\ell}(\boldsymbol{x})^{\gamma_{k\ell}})$. Second, if the customer decides to make a purchase in nest (i,j), then she chooses product i with probability $q^i_{ij}(\boldsymbol{x}) = v_i^{1/\gamma_{ij}}\, x_i \,/\, V_{ij}(\boldsymbol{x})$, whereas she chooses product j with probability $q^j_{ij}(\boldsymbol{x}) = v_j^{1/\gamma_{ij}}\, x_j \,/\, V_{ij}(\boldsymbol{x})$. Thus, if we offer the subset of products $\boldsymbol{x}$, then a customer chooses product i with probability $\sum_{k\in N: k\neq i}(P_{ik}(\boldsymbol{x})\, q^i_{ik}(\boldsymbol{x}) + P_{ki}(\boldsymbol{x})\, q^i_{ki}(\boldsymbol{x}))$, fully specifying the choice probabilities under the PCL model. To formulate our assortment problem, we use $p_i \geq 0$ to denote the revenue of product i. If we offer the subset of products $\boldsymbol{x}$ and a customer has already decided to make a purchase in nest (i,j), then the expected revenue that we obtain from the customer is

$$R_{ij}(\boldsymbol{x}) \;=\; q^i_{ij}(\boldsymbol{x})\, p_i + q^j_{ij}(\boldsymbol{x})\, p_j \;=\; \frac{p_i\, v_i^{1/\gamma_{ij}}\, x_i + p_j\, v_j^{1/\gamma_{ij}}\, x_j}{V_{ij}(\boldsymbol{x})}.$$

We use $\pi(\boldsymbol{x})$ to denote the expected revenue that we obtain from a customer when we offer the subset of products $\boldsymbol{x}$. In this case, we have

$$\pi(\boldsymbol{x}) \;=\; \sum_{(i,j)\in M} P_{ij}(\boldsymbol{x})\, R_{ij}(\boldsymbol{x}) \;=\; \sum_{(i,j)\in M} \frac{V_{ij}(\boldsymbol{x})^{\gamma_{ij}}}{v_0 + \sum_{(k,\ell)\in M} V_{k\ell}(\boldsymbol{x})^{\gamma_{k\ell}}}\, R_{ij}(\boldsymbol{x}) \;=\; \frac{\sum_{(i,j)\in M} V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\, R_{ij}(\boldsymbol{x})}{v_0 + \sum_{(i,j)\in M} V_{ij}(\boldsymbol{x})^{\gamma_{ij}}}.$$

Throughout the paper, we consider both uncapacitated and capacitated assortment problems. In the uncapacitated assortment problem, we can offer any subset of products to the customers. In the capacitated assortment problem, we have an upper bound on the number of products that we can offer to the customers. To capture both the uncapacitated and capacitated assortment problems succinctly, for some $c \in \mathbb{Z}_+$, we use $\mathcal{F} = \{\boldsymbol{x} \in \{0,1\}^n : \sum_{i \in N} x_i \le c\}$ to denote the feasible subsets of products that we can offer to the customers. Since there are n products, the constraint $\sum_{i \in N} x_i \le c$ is not binding when we have $c \ge n$. Thus, we obtain the uncapacitated assortment problem by choosing a value of c that is no smaller than n, whereas we obtain the capacitated assortment problem with other values of c. In the assortment problem, our goal is to find a feasible subset of products to offer that maximizes the expected revenue obtained from a customer, corresponding to the combinatorial optimization problem

$$z^* \;=\; \max_{\boldsymbol{x} \in \mathcal{F}} \; \pi(\boldsymbol{x}) \;=\; \max_{\boldsymbol{x} \in \mathcal{F}} \left\{ \frac{\sum_{(i,j) \in M} V_{ij}(\boldsymbol{x})^{\gamma_{ij}} \, R_{ij}(\boldsymbol{x})}{v_0 + \sum_{(i,j) \in M} V_{ij}(\boldsymbol{x})^{\gamma_{ij}}} \right\}. \qquad \text{(Assortment)}$$

Our formulation of the PCL model is slightly different from the one that often appears in the literature. In the existing literature, the collection of nests is often $\{(i,j) \in N^2 : i < j\}$, whereas in our formulation, the collection of nests is $\{(i,j) \in N^2 : i \ne j\}$. If we let $\gamma_{ij} = \gamma_{ji}$ for all $(i,j) \in N^2$ with $i > j$ and double the preference weight of the no purchase option in our formulation, then it is simple to check that the two formulations of the PCL model are consistent. Thus, our approximation algorithms remain applicable to the formulation of the PCL model that often appears in the literature. Our formulation of the PCL model will reduce the notational burden.

If $\gamma_{ij} \in [0,1]$ and $\gamma_{ij} = \gamma_{ji}$ for all $(i,j) \in M$, then the PCL model is compatible with the random utility maximization principle, where each customer associates a random utility with the products and the no purchase option, choosing the alternative with the largest utility; see Koppleman and Wen (2000). The parameter γ_{ij} characterizes the degree of correlation between the utilities of products i and j. The approximation algorithms that we give require that $\gamma_{ij} \in [0,1]$, but not necessarily $\gamma_{ij} = \gamma_{ji}$, but the latter condition is needed if we want to ensure that the PCL model that we work with is compatible with the random utility maximization principle.

In the next theorem, we show that the Assortment problem is strongly NP-hard even when $\mathcal{F} = \{0,1\}^n$ so that the problem is uncapacitated and $\gamma_{ij} = \gamma_{ji}$ for all $(i,j) \in M$.

Theorem 2.1 (Computational Complexity) *The* Assortment *problem is strongly NP-hard, even when we have $\mathcal{F} = \{0,1\}^n$ and $\gamma_{ij} = \gamma_{ji}$ for all $(i,j) \in M$.*

The proof of Theorem 2.1 is in Appendix A. It uses a reduction from the max-cut problem, which is a well-known NP-hard problem; see Section A.2.2 in Garey and Johnson (1979). Motivated by this complexity result, in the rest of the paper, we focus on developing approximation algorithms for the Assortment problem. For $\alpha \in [0,1]$, an α-approximation algorithm is a polynomial-time algorithm that, for any problem instance, computes an assortment $\hat{\boldsymbol{x}} \in \mathcal{F}$, whose expected revenue is at least α times the optimal expected revenue; that is, noting that the optimal expected revenue in the Assortment problem is z^*, the assortment $\hat{\boldsymbol{x}}$ satisfies $\pi(\hat{\boldsymbol{x}}) \ge \alpha \, z^*$.

2.2. Paired Combinatorial Logit Model in Assortment Problems

There is work on assortment problems under the multinomial logit and nested logit models. The multinomial logit, nested logit and PCL models are all compatible with the random utility maximization principle. In the random utility maximization principle, a customer associates random utilities with the products and the no purchase option, which are sampled from a certain joint utility distribution. The customer knows the sampled values of the utilities and she chooses the alternative with the largest utility. A natural question is whether the PCL model captures a phenomenon in the choice process of the customers that cannot be captured by the multinomial logit or nested logit models. We point out two phenomena that can be captured by the PCL model, but not by the multinomial logit and nested logit models. Also, we give an example where these two phenomena naturally occur, potentially making the PCL model a viable candidate.

First, there can be different levels of correlation between the utilities of different pairs of products. In particular, if there are two products that are similar to each other, then knowing that a customer associates a high utility with one may indicate that the customer also associates a high utility with the other. Under the multinomial logit model, the utilities of the products are uncorrelated, yielding a correlation coefficient matrix of zero for the utilities. Under the nested logit model, the products are partitioned into nests. The correlation coefficient for the utilities of any pair of products in the same nest is a constant determined by the dissimilarity parameter of the nest. The correlation coefficient for the utilities of any pair of products in different nests is zero; see Daganzo and Kusnic (1993). Therefore, the correlation coefficient matrix for the utilities under the nested logit model is block diagonal and all (non-diagonal) entries in each block are the same.

Under the PCL model, we can have a different correlation coefficient for the utilities of each pair of products. Recalling that $\gamma_{ij} = \gamma_{ji}$ when the PCL model is compatible with the random utility maximization principle, under the PCL model, the correlation coefficient between the utilities of products i and j depends only on the dissimilarity parameter γ_{ij} and the number of products n; see Koppleman and Wen (2000). Thus, each entry in the correlation coefficient matrix for the utilities can be different, capturing different degrees of correlation between the utilities. In Figure 1, we consider the case with three products and show the correlation coefficient between the utilities of products i and j as a function of γ_{ij}. By choosing a different value for γ_{ij} for each pair of products i and j, we can have a different correlation coefficient for the utilities of each pair of products. Such flexibility is not available under the multinomial logit and nested logit models.

Second, we may have a case where the correlation structure of the utilities does not satisfy, what we call, the "transitivity" property. In particular, we may have a case where the utility of product

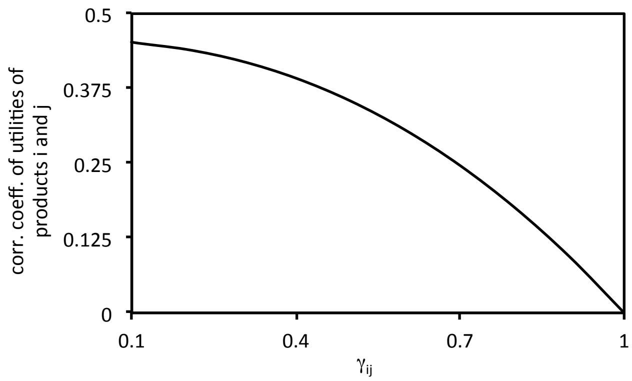

Figure 1 Correlation coefficient of the utilities of products i and j under the PCL model as a function of γ_{ij}.

1 is correlated with the utility of product 2 and the utility of product 2 is correlated with the utility of product 3, but the utility of product 1 is not correlated with the utility of product 3. We cannot capture such a correlation structure under the multinomial logit or nested logit models. Under the multinomial logit model, the utilities of the products are uncorrelated. Under the nested logit model, if the utilities of products 1 and 2 are to be correlated, then they have to be in the same nest. Similarly, if the utilities of products 2 and 3 are to be correlated, then they have to be in the same nest as well. Therefore, products 1 and 3 end up being in the same nest, in which case, the utilities of products 1 and 3 must also be correlated.

We can use the PCL model to capture the case where the correlation structure of the utilities does not satisfy the "transitivity" property. Noting Figure 1, if $\gamma_{ij} = 1$, then the utilities of products i and j under the PCL model are uncorrelated, where as if $\gamma_{ij} < 1$, then the utilities of products i and j are correlated. Therefore, if we use the PCL model with $\gamma_{12} < 1$, $\gamma_{23} < 1$ and $\gamma_{13} = 1$, then the utility of product 1 is correlated with the utility of product 2, the utility of product 2 is correlated with the utility of product 3 and the utility of product 1 is not correlated with the utility of product 3. Thus, such choices of γ_{12}, γ_{23} and γ_{13} allow the PCL model to handle correlation structures that do not satisfy the "transitivity" property.

It is not difficult to give examples where there are different levels of correlation between the utilities of different pairs of alternatives and the correlation structure of the utilities does not satisfy the "transitivity" property. In Appendix B, we give an example that focuses on the path choices of commuters traveling from an origin node to a destination node in a network. In this example, there are different paths that connect the origin node to the destination node. The disutility of a path is the sum of the travel times on the edges that are included in the path. The travel time on the edges are random but the commuter knows the travel times before deciding which path to take. The commuter chooses the path that provides the largest utility. Since two different paths may use a common edge, the utilities provided by two different paths can be correlated. We make two observations in our example. First, we can have different levels of correlation between the utilities of different pairs of paths. In particular, paths P_1 and P_2 may share three common edges, but paths P_1 and P_3 may share only one common edge. In this case, we expect the correlation between the utilities of paths P_1 and P_2 to be relatively high, but the correlation between the utilities of paths P_1 and P_3 to be relatively low. Second, the correlation structure between the utilities of different paths may not satisfy the "transitivity" property. In particular, path R_1 may share common edges with path R_2, path R_2 may share common edges with path R_3, but path R_1 may not share any common edges with path R_3. In this case, we expect that the utility of path R_1 is correlated with the utility of path R_2, the utility of path R_2 is correlated with the utility of path R_3, but the utility of path R_1 is not correlated with the utility of path R_3. In Appendix B, we give a specific network where it is indeed the case that there can be different levels of correlation between the utilities of different pairs of paths and the correlation structure of the utilities may violate the "transitivity" property. In addition, we give a numerical study to check the ability of the PCL model to predict the path choices of commuters. Although our example focuses on the path choices of commuters, similar situations occur when customers choose among products that share different numbers of common features.

The generalized nested logit model subsumes the multinomial logit, nested logit and PCL models, as well as the cross nested logit, ordered extreme value and product differentiation models, as special cases; see Section 2.2 in Wen and Koppelman (2001). As discussed in Section 4.4.2 in Train (2002), we can view the choice process of a customer under any generalized nested logit model as taking place in two stages. In the first stage, the customer chooses a nest. In the second stage, the customer chooses a product within the nest. Often times, the two stages in the choice process occur as a result of the correlation structure between the utilities of the products. For the nested logit model, for example, using γ_m to denote the dissimilarity parameter of nest m and $m(i)$ to denote the nest that includes product i, the utility of product i is given by $\mathsf{Utility}_i = \mu_i + \zeta_{m(i)} + \gamma_{m(i)}\,\eta_i$, where μ_i is a deterministic constant, the random variable $\zeta_{m(i)}$ has the so-called C distribution with parameter γ_m and the random variable η_i has the Gumbel distribution with location and shape parameters $(0,1)$; see Section 3 in Cardell (1997). Letting M be the set of nests and N be the set of products, the random variables $\{\zeta_m : m \in M\} \cup \{\eta_i : i \in N\}$ are independent of each other. We can interpret ζ_m as the utility extracted from a product in nest m based on the common characteristics of the products in this nest and η_i as the utility specific to product i. Due to this structure of the utilities, we can view the choice process of the customer under the nested logit model as taking place in two stages. Since each product appears in only one nest in the nested logit model with the utilities of the products in different nests being uncorrelated, the two stages may correspond to the thought process that a customer goes through when making a purchase. In the first stage, the customer chooses a group of products that are similar to each other, which is captured by a nest. In the second stage, the customer chooses a product within this group. However, the use of the nested logit model does not require that the customers actually go through a two stage thought process when making a purchase. The two stages occur simply as a result of the correlation structure of the utilities. In Appendix C, we give a similar description of the utilities of the products under the PCL model. We can interpret the choice process of a customer under the PCL model also as taking place in two stages, but the two stages, once again, do not necessarily reflect the actual thought process of the customers. The two stages occur simply as a result of the correlation structure of the utilities and we only use the two stages to intuitively describe the PCL model.

In the Assortment problem, the assortment that we offer is a decision variable, but we use the PCL model with the same parameters to capture choices within different assortments, simply dropping the products that are not offered. In Appendix D, we carefully justify this approach by using the fact that the PCL model is based on the random utility maximization principle. In Appendix E, we also give a numerical study to check the ability of the PCL model to predict choices within different assortments. Compared with the multinomial logit benchmark, fitting a PCL model provides larger out of sample log-likelihoods and smaller errors in the predicted choice probabilities.

3. A General Framework for Approximation Algorithms

In this section, we provide a general framework that is useful for developing approximation algorithms for the Assortment problem. Our framework is applicable to both the uncapacitated and capacitated problems simultaneously.

3.1. Connection to a Fixed Point Problem

Note that $\pi(\boldsymbol{x}) = \frac{\sum_{(i,j)\in M} V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\, R_{ij}(\boldsymbol{x})}{v_0 + \sum_{(i,j)\in M} V_{ij}(\boldsymbol{x})^{\gamma_{ij}}} \geq z$ if and only if $\sum_{(i,j)\in M} V_{ij}(\boldsymbol{x})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}) - z) \geq v_0\, z$. Thus, there exists an assortment with a revenue of z or more if and only if $\sum_{(i,j)\in M} V_{ij}(\boldsymbol{x})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}) - z) \geq v_0\, z$ for some $\boldsymbol{x} \in \{0,1\}^n$. To use this observation, we define the function $f : \mathbb{R} \to \mathbb{R}$ as

$$f(z) \;=\; \max_{\boldsymbol{x}\in\mathcal{F}} \left\{ \sum_{(i,j)\in M} V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\,(R_{ij}(\boldsymbol{x}) - z) \right\}. \qquad \text{(Function Evaluation)}$$

For each $\boldsymbol{x} \in \mathcal{F}$, the objective function of the Function Evaluation problem on the right side above is decreasing in z, which implies that $f(z)$ is also decreasing in z. A simple lemma, given as Lemma F.1 in Appendix F, also shows that $f(\cdot)$ is continuous and it satisfies $f(0) \geq 0$. Therefore, $f(\cdot)$ is decreasing and continuous, satisfying $f(0)\geq 0$. Also, defining the function $g:\mathbb{R}\rightarrow\mathbb{R}$ as $g(z)=v_0\,z$, $g(\cdot)$ is strictly increasing and continuous, satisfying $g(0)=0$ and $\lim_{z\rightarrow\infty} g(z)=\infty$. In this case, there exists a unique $\hat{z}\geq 0$ satisfying $f(\hat{z})=v_0\,\hat{z}$. Note that the value of $\hat{z}$ that satisfies $f(\hat{z})=v_0\,\hat{z}$ is the fixed point of the function $f(\cdot)/v_0$. It is possible to show that this value of $\hat{z}$ corresponds to the optimal objective value of the Assortment problem and we can solve the Function Evaluation problem with $z=\hat{z}$ to obtain an optimal solution to the Assortment problem. We do not give a proof for this result, since this result immediately follows as a corollary to a more general result that we shortly give in Theorem 3.1.

Since the Function Evaluation problem is a nonlinear integer program, finding the fixed point of $f(\cdot)/v_0$ can be difficult. Instead, we will use an LP or SDP relaxation of the Function Evaluation problem to construct an upper bound $f^R(\cdot)$ on $f(\cdot)$ so that $f^R(z)\geq f(z)$ for all $z\in\mathbb{R}$. The upper bound $f^R(\cdot)$ will be decreasing and continuous, satisfying $f^R(0)\geq 0$. Thus, by the same argument in the previous paragraph, there exists a unique $\hat{z}\geq 0$ satisfying $f^R(\hat{z})=v_0\,\hat{z}$. In the next theorem, we show that this value of $\hat{z}$ upper bounds the optimal objective value of the Assortment problem and we can use this value of $\hat{z}$ to obtain an approximate solution.

Theorem 3.1 (Approximation Framework) *Assume that $f^R(\cdot)$ satisfies $f^R(z)\geq f(z)$ for all $z\in\mathbb{R}$. Let $\hat{z}\geq 0$ satisfy $f^R(\hat{z}) = v_0\,\hat{z}$ and $\hat{\boldsymbol{x}}\in\mathcal{F}$ be such that*

$$\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}\,(R_{ij}(\hat{\boldsymbol{x}})-\hat{z})\geq \alpha\, f^R(\hat{z}) \qquad \text{(Sufficient Condition)}$$

for some $\alpha\in[0,1]$. Then, we have $\pi(\hat{\boldsymbol{x}})\geq \alpha\,\hat{z}\geq\alpha\,z^*$; so, $\hat{z}$ upper bounds the optimal objective value of the Assortment *problem and $\hat{\boldsymbol{x}}$ is an α-approximate solution to the* Assortment *problem.*

Proof: Noting that $v_0\,\hat{z}=f^R(\hat{z})$, we have $\alpha\,v_0\,\hat{z}=\alpha\,f^R(\hat{z})\leq\sum_{(i,j)\in M}V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}\,(R_{ij}(\hat{\boldsymbol{x}})-\hat{z})\leq \sum_{(i,j)\in M}V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}\,(R_{ij}(\hat{\boldsymbol{x}})-\alpha\,\hat{z})$, where the first inequality uses the Sufficient Condition. Thus, we have $\alpha\,v_0\,\hat{z}\leq\sum_{(i,j)\in M}V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}\,(R_{ij}(\hat{\boldsymbol{x}})-\alpha\,\hat{z})$, in which case, solving for $\hat{z}$ in the last inequality, we get $\alpha\,\hat{z}\leq\sum_{(i,j)\in M}V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}\,R_{ij}(\hat{\boldsymbol{x}})/(v_0+\sum_{(i,j)\in M}V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}})=\pi(\hat{\boldsymbol{x}})$. Next, we show that $\hat{z}\geq z^*$. We let $\boldsymbol{x}^*$ be an optimal solution to the Assortment problem. Since $\boldsymbol{x}^*$ is a feasible but not necessarily an optimal solution to the Function Evaluation problem when we solve this problem with $z=\hat{z}$, we have $f(\hat{z})\geq\sum_{(i,j)\in M}V_{ij}(\boldsymbol{x}^*)^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^*)-\hat{z})$. Noting that $v_0\,\hat{z}=f^R(\hat{z})\geq f(\hat{z})$, we obtain $v_0\,\hat{z}\geq\sum_{(i,j)\in M}V_{ij}(\boldsymbol{x}^*)^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^*)-\hat{z})$. Solving for $\hat{z}$ in this inequality, we get $\hat{z}\geq \sum_{(i,j)\in M}V_{ij}(\boldsymbol{x}^*)^{\gamma_{ij}}\,R_{ij}(\boldsymbol{x}^*)/(v_0+\sum_{(i,j)\in M}V_{ij}(\boldsymbol{x}^*)^{\gamma_{ij}})=\pi(\boldsymbol{x}^*)=z^*$. □

As a corollary, if we apply Theorem 3.1 with $f^R(\cdot)=f(\cdot)$, then the value of $\hat{z}$ that satisfies $f(\hat{z})=v_0\,\hat{z}$ is the optimal objective value of the Assortment problem, but we will use the theorem above only to obtain an approximate solution. By Theorem 3.1, to obtain an α-approximate solution to the Assortment problem, it is enough to execute the following three steps.

Approximation Framework

Step 1: Construct a decreasing and continuous upper bound $f^R(\cdot)$ on $f(\cdot)$ with $f^R(0) \geq 0$.
Step 2: Find the fixed point $\hat{z}$ of $f^R(\cdot)/v_0$; that is, find the value of $\hat{z}$ such that $f^R(\hat{z}) = v_0\, \hat{z}$.
Step 3: Find an assortment $\hat{\boldsymbol{x}} \in \mathcal{F}$ such that $\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}\,(R_{ij}(\hat{\boldsymbol{x}}) - \hat{z}) \geq \alpha\, f^R(\hat{z})$.

In Section 3.2, we show how to construct the upper bound $f^R(\cdot)$ on $f(\cdot)$ by using an LP relaxation of the Function Evaluation problem. In Section 3.3, we show how to compute the fixed point of $f^R(\cdot)/v_0$ by solving an LP. Using these results, we can execute Steps 1 and 2 in our approximation framework. In Section 4, we tackle the uncapacitated problem and use randomized rounding on an optimal solution of an LP relaxation of the Function Evaluation problem to construct an assortment $\hat{\boldsymbol{x}}$ that satisfies the Sufficient Condition with $\alpha = 0.6$, yielding a 0.6-approximation algorithm. We also discuss an SDP relaxation to satisfy the Sufficient Condition with $\alpha = 0.79$. In Section 5, we tackle the capacitated problem and use iterative rounding to construct an assortment that satisfies the Sufficient Condition with $\alpha = 0.25$, yielding a 0.25-approximation algorithm.

Davis et al. (2014), Gallego and Topaloglu (2014), and Feldman and Topaloglu (2015) use analogues of the Function Evaluation problem and Theorem 3.1 to approximately solve assortment problems under the nested logit model, but these authors face different challenges. First, under the nested logit model, since each product appears in one nest, the analogue of the Function Evaluation problem decomposes by the nests. Second, the authors characterize α-approximate solutions for the subproblem for each nest for all possible values of $z \in \mathbb{R}$. In this way, they construct a collection of candidate assortments for each nest that includes an α-approximate assortment to offer in the nest. Third, the authors solve an LP to stitch together an α-approximate solution to the assortment problem by picking one assortment from the candidate collection for each nest.

The steps described above are not possible under the PCL model. First, since a product appears in multiple nests under the PCL model, the Function Evaluation problem does not decompose by the nests. Second, the idea of constructing a collection of candidate assortments for each nest does not yield a useful algorithmic approach under the PCL model, since there are two products in each nest, yielding four possible assortments in each nest anyway. Third, there is a stronger interaction between the nests under the PCL model. If we offer an assortment that includes product i in nest (i,j), then we must offer assortments that include product i in all nests $\{(i,\ell) : \ell \in N \setminus \{i\}\}$. Due to this interaction, we cannot solve an LP to stitch together an approximate solution to the assortment problem by picking one assortment from a candidate collection for each nest.

In our approach, instead of characterizing α-approximate solutions to the Function Evaluation problem for all values of $z \in \mathbb{R}$, we find the fixed point $\hat{z}$ of the upper bound $f^R(\cdot)/v_0$. In this case, we find an α-approximate solution to the Function Evaluation problem only at $z = \hat{z}$.

3.2. Constructing an Upper Bound

We construct an upper bound $f^R(\cdot)$ on $f(\cdot)$ by using an LP relaxation of the Function Evaluation problem. Noting the definition of $V_{ij}(\boldsymbol{x})$ and $R_{ij}(\boldsymbol{x})$, we have

$$V_{ij}(\boldsymbol{x})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x})-z) \;=\; (v_i^{1/\gamma_{ij}}\,x_i+v_j^{1/\gamma_{ij}}\,x_j)^{\gamma_{ij}}\,\frac{(p_i-z)\,v_i^{1/\gamma_{ij}}\,x_i+(p_j-z)\,v_j^{1/\gamma_{ij}}\,x_j}{v_i^{1/\gamma_{ij}}\,x_i+v_j^{1/\gamma_{ij}}\,x_j}.$$

We let $\rho_{ij}(z)$ be the expression on the right side above when $x_i=1$ and $x_j=1$ and $\theta_i(z)$ be the expression on the right side above when $x_i=1$ and $x_j=0$. In other words, we have

$$\rho_{ij}(z) \;=\; (v_i^{1/\gamma_{ij}}+v_j^{1/\gamma_{ij}})^{\gamma_{ij}}\,\frac{(p_i-z)\,v_i^{1/\gamma_{ij}}+(p_j-z)\,v_j^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}}+v_j^{1/\gamma_{ij}}} \quad \text{and} \quad \theta_i(z) \;=\; v_i\,(p_i-z).$$

There are only four possible values of (x_i,x_j). In this case, letting $\mu_{ij}(z)=\rho_{ij}(z)-\theta_i(z)-\theta_j(z)$ for notational brevity, we can express $V_{ij}(\boldsymbol{x})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x})-z)$ succinctly as

$$V_{ij}(\boldsymbol{x})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x})-z)=\rho_{ij}(z)\,x_i\,x_j+\theta_i(z)\,x_i\,(1-x_j)+\theta_j(z)\,(1-x_i)\,x_j=\mu_{ij}(z)\,x_i\,x_j+\theta_i(z)\,x_i+\theta_j(z)\,x_j.$$

Writing its objective function as $\sum_{(i,j)\in M}(\mu_{ij}(z)\,x_i\,x_j+\theta_i(z)\,x_i+\theta_j(z)\,x_j)$, the Function Evaluation problem is equivalent to

$$f(z)=\max\left\{\sum_{(i,j)\in M}(\mu_{ij}(z)\,x_i\,x_j+\theta_i(z)\,x_i+\theta_j(z)\,x_j) \;:\; \sum_{i\in N}x_i\le c,\quad x_i\in\{0,1\}\quad\forall\, i\in N\right\}.$$

In general, the sign of $\mu_{ij}(z)$ can be positive or negative, but we shortly show that $\mu_{ij}(z)\le 0$ whenever $p_i\ge z$ and $p_j\ge z$. To construct an upper bound $f^R(\cdot)$ on $f(\cdot)$, we use a standard approach to linearize the term $x_i\,x_j$ in the objective function above. Define the decision variable $y_{ij}\in\{0,1\}$ with the interpretation that $y_{ij}=x_i\,x_j$. To ensure that y_{ij} takes the value $x_i\,x_j$, we impose the constraints $y_{ij}\ge x_i+x_j-1$, $y_{ij}\le x_i$ and $y_{ij}\le x_j$. If $x_i=0$ or $x_j=0$, then the constraints $y_{ij}\le x_i$ and $y_{ij}\le x_j$ ensure that $y_{ij}=0$. If $x_i=1$ and $x_j=1$, then the constraint $y_{ij}\ge x_i+x_j-1$ ensures that $y_{ij}=1$. We define our upper bound $f^R(\cdot)$ on $f(\cdot)$ by using the LP relaxation

$$\begin{aligned} f^R(z) \;=\; \max \quad & \sum_{(i,j)\in M}(\mu_{ij}(z)\,y_{ij}+\theta_i(z)\,x_i+\theta_j(z)\,x_j) && \text{(Upper Bound)}\\ \text{s.t.} \quad & y_{ij}\ge x_i+x_j-1 \quad \forall\,(i,j)\in M\\ & y_{ij}\le x_i,\quad y_{ij}\le x_j \quad \forall\,(i,j)\in M\\ & \sum_{i\in N}x_i\le c\\ & 0\le x_i\le 1 \quad \forall\, i\in N,\quad y_{ij}\ge 0 \quad \forall\,(i,j)\in M. \end{aligned}$$

Since the Upper Bound problem is an LP relaxation of the Function Evaluation problem, we have $f^R(z)\ge f(z)$ for all $z\in\mathbb{R}$. Setting $x_i=0$ for all $i\in N$ and $y_{ij}=0$ for all $(i,j)\in M$ gives a feasible solution to the Upper Bound problem, so $f^R(\cdot) \geq 0$. Since $\theta_i(z)$ and $\mu_{ij}(z)$ are continuous in z and the optimal objective value of a bounded LP is continuous in its objective function coefficients, $f^R(z)$ is continuous in z. It is not immediately clear that $f^R(z)$ is decreasing in z since it is not immediately clear that the objective function coefficient $\mu_{ij}(z)$ in the Upper Bound problem is decreasing in z. In the next lemma, we show that $f^R(z)$ is decreasing in z. Since $f^R(\cdot)$ is decreasing and continuous with $f^R(0) \geq 0$, there exists a unique $\hat{z} \geq 0$ satisfying $f^R(\hat{z}) = v_0\,\hat{z}$.

Lemma 3.2 (Monotonicity of Upper Bound) *The optimal objective value $f^R(z)$ of the* Upper Bound *problem is decreasing in z.*

Proof: Consider $z^+ \geq z^-$ and let $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ with $\boldsymbol{y}^* = \{y^*_{ij} : (i,j) \in M\}$ be an optimal solution to the Upper Bound problem when we solve this problem with $z = z^+$. Since $\mu_{ij}(z) = \rho_{ij}(z) - \theta_i(z) - \theta_j(z)$ by the definition of $\mu_{ij}(z)$, we obtain

$$\begin{aligned}
f^R(z^+) &= \sum_{(i,j)\in M} \left(\mu_{ij}(z^+)\, y^*_{ij} + \theta_i(z^+)\, x^*_i + \theta_j(z^+)\, x^*_j\right) \\
&= \sum_{(i,j)\in M} \left(\rho_{ij}(z^+)\, y^*_{ij} + \theta_i(z^+)\,(x^*_i - y^*_{ij}) + \theta_j(z^+)\,(x^*_j - y^*_{ij})\right) \\
&\leq \sum_{(i,j)\in M} \left(\rho_{ij}(z^-)\, y^*_{ij} + \theta_i(z^-)\,(x^*_i - y^*_{ij}) + \theta_j(z^-)\,(x^*_j - y^*_{ij})\right) \\
&= \sum_{(i,j)\in M} \left(\mu_{ij}(z^-)\, y^*_{ij} + \theta_i(z^-)\, x^*_i + \theta_j(z^-)\, x^*_j\right) \\
&\leq f^R(z^-),
\end{aligned}$$

where the first inequality is by the fact that $\rho_{ij}(z)$ and $\theta_i(z)$ are decreasing in z, along with the fact that $y^*_{ij} \leq x^*_i$ and $y^*_{ij} \leq x^*_j$, whereas the second inequality is by the fact that $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is a feasible but not necessarily an optimal solution to the Upper Bound problem with $z = z^-$. □

One useful property of the Upper Bound problem is that there exists an optimal solution to this problem where the decision variable x_i takes a non-zero value only when $p_i \geq z$ and the decision variable y_{ij} takes a non-zero value only when $p_i \geq z$ and $p_j \geq z$. Thus, we need to keep the decision variable x_i only when $p_i \geq z$ and we need to keep the decision variable y_{ij} only when $p_i \geq z$ and $p_j \geq z$. This property allows us to significantly simplify the Upper Bound problem. In particular, let $N(z) = \{i \in N : p_i \geq z\}$ and $M(z) = \{(i,j) \in N(z)^2 : i \neq j\}$. In Lemma G.1 in Appendix G, we show that there exists an optimal solution $\boldsymbol{x}^* = \{x_i : i \in N\}$ and $\boldsymbol{y}^* = \{y^*_{ij} : (i,j) \in N\}$ to the Upper Bound problem such that $x^*_i = 0$ for all $i \notin N(z)$ and $y^*_{ij} = 0$ for all $(i,j) \notin M(z)$. The proof of this result follows by showing that if $\hat{\boldsymbol{x}} = \{\hat{x}_i : i \in N\}$ and $\hat{\boldsymbol{y}} = \{\hat{y}_{ij} : (i,j) \in N\}$ is a feasible solution to the Upper Bound problem, then we can set $\hat{x}_i = 0$ for all $i \notin N(z)$ and $\hat{y}_{ij} = 0$ for all $(i,j) \notin M(z)$ while making sure that the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ remains feasible to the Upper Bound problem and we do not degrade the objective value provided by this solution. In this case, letting $\mathbf{1}(\cdot)$ be the indicator function and dropping the decision variable x_i for all $i \notin N(z)$ and the decision variable y_{ij} for all $(i,j) \notin M(z)$, we write the objective function of the Upper Bound problem as

$$\sum_{(i,j)\in M} \mathbf{1}(i \in N(z),\ j \in N(z))\,(\mu_{ij}(z)\,y_{ij} + \theta_i(z)\,x_i + \theta_j(z)\,x_j) \\ + \sum_{(i,j)\in M} \mathbf{1}(i \in N(z),\ j \notin N(z))\,\theta_i(z)\,x_i \ + \sum_{(i,j)\in M} \mathbf{1}(i \notin N(z),\ j \in N(z))\,\theta_j(z)\,x_j.$$

For the last two sums, we have $\sum_{(i,j)\in M} \mathbf{1}(i \in N(z),\ j \notin N(z))\,\theta_i(z)\,x_i = |N \setminus N(z)| \sum_{i\in N(z)} \theta_i(z)\,x_i$ and $\sum_{(i,j)\in M} \mathbf{1}(i \notin N(z),\ j \in N(z))\,\theta_j(z)\,x_j = |N \setminus N(z)| \sum_{j\in N(z)} \theta_j(z)\,x_j$. Thus, the objective function of the Upper Bound problem takes the form $\sum_{(i,j)\in M(z)} (\mu_{ij}(z)\,y_{ij} + \theta_i(z)\,x_i + \theta_j(z)\,x_j) + 2\,|N \setminus N(z)| \sum_{i\in N(z)} \theta_i(z)\,x_i$. A simple lemma, given as Lemma G.2 in Appendix G, shows that $\mu_{ij}(z) \le 0$ for all $(i,j) \in M(z)$. So, the decision variable y_{ij} takes its smallest possible value in the Upper Bound problem, which implies that the constraints $y_{ij} \le x_i$ and $y_{ij} \le x_j$ are redundant. In this case, the Upper Bound problem is equivalent to the problem

$$\begin{aligned} f^R(z) \ = \ \max \quad & \sum_{(i,j)\in M(z)} (\mu_{ij}(z)\,y_{ij} + \theta_i(z)\,x_i + \theta_j(z)\,x_j) + 2\,|N \setminus N(z)| \sum_{i\in N(z)} \theta_i(z)\,x_i \\ \text{s.t.} \quad & y_{ij} \ge x_i + x_j - 1 \quad \forall\,(i,j) \in M(z) \qquad \text{(Compact Upper Bound)} \\ & \sum_{i\in N(z)} x_i \le c \\ & 0 \le x_i \le 1 \ \ \forall\, i \in N(z), \quad y_{ij} \ge 0 \ \ \forall\,(i,j) \in M(z). \end{aligned}$$

Both the Upper Bound and Compact Upper Bound problems will be useful. We will use the Upper Bound problem to find the fixed point of $f^R(\cdot)/v_0$. We will use the Compact Upper Bound problem above to find an assortment $\hat{\boldsymbol{x}}$ satisfying the Sufficient Condition.

Noting the objective function $\sum_{(i,j)\in M}(\mu_{ij}(z)\,x_i\,x_j + \theta_i(z)\,x_i + \theta_j(z)\,x_j)$ in the equivalent formulation of the Function Evaluation problem, this problem does not decompose by the nests even when we focus on the uncapacitated assortment problem. Therefore, unlike the approach used by Davis et al. (2014), Gallego and Topaloglu (2014), and Feldman and Topaloglu (2015) for tackling assortment problems under the nested logit model, since the Function Evaluation problem does not decompose by the nests under the PCL model, it is difficult to characterize the optimal or approximate solutions to the Function Evaluation problem for all values of $z \in \mathbb{R}$. Instead, we approximate the Function Evaluation problem without decomposing it. Rather than characterizing approximate solutions for all values of $z \in \mathbb{R}$, we construct an approximate solution to the Function Evaluation problem with $z = \hat{z}$, where $\hat{z}$ is the fixed point of $f^R(\cdot)/v_0$.

3.3. Computing the Fixed Point

To compute the fixed point of $f^R(\cdot)/v_0$, we use the dual of the Upper Bound problem. For each $(i,j)\in M$, let α_{ij}, β_{ij}, and γ_{ij} be the dual variables of the constraints $y_{ij}\geq x_i+x_j-1$, $y_{ij}\leq x_i$ and $y_{ij}\leq x_j$, respectively. For each $i\in N$, let λ_i be the dual variable of the constraint $x_i\leq 1$. Let δ be the dual variable of the constraint $\sum_{i\in N} x_i\leq c$. The dual of the Upper Bound problem is

$$\begin{aligned} f^R(z) \;=\; \min\;\; & c\,\delta+\sum_{i\in N}\lambda_i+\sum_{(i,j)\in M}\alpha_{ij} && && \text{(Dual)}\\ \text{s.t.}\;\; & -\alpha_{ij}+\beta_{ij}+\gamma_{ij} && \geq\; \mu_{ij}(z) && \forall\,(i,j)\in M\\ & \delta+\lambda_i+\sum_{j\in N}\mathbf{1}(j\neq i)\,(\alpha_{ij}+\alpha_{ji}-\beta_{ij}-\gamma_{ji}) && \geq\; 2\,(n-1)\,\theta_i(z) && \forall\, i\in N\\ & \alpha_{ij}\geq 0,\;\beta_{ij}\geq 0,\;\gamma_{ij}\geq 0\;\;\forall\,(i,j)\in M,\;\;\lambda_i\geq 0\;\;\forall\, i\in N,\;\;\delta\geq 0. \end{aligned}$$

In the Dual problem above, the decision variables are $\boldsymbol{\alpha}=\{\alpha_{ij}:(i,j)\in M\}$, $\boldsymbol{\beta}=\{\beta_{ij}:(i,j)\in M\}$, $\boldsymbol{\gamma}=\{\gamma_{ij}:(i,j)\in M\}$, $\boldsymbol{\lambda}=\{\lambda_i:i\in N\}$ and δ. We treat z as fixed. In the next theorem, we show that if we treat z as a decision variable and add one constraint to the Dual problem that involves the decision variable z, then we can recover the fixed point of $f^R(\cdot)/v_0$.

Theorem 3.3 (Fixed Point Computation by Using an LP) *Let* $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\lambda}},\hat{\delta},\hat{z})$ *be an optimal solution to the problem*

$$\begin{aligned} \min\;\; & c\,\delta+\sum_{i\in N}\lambda_i+\sum_{(i,j)\in M}\alpha_{ij} && && \text{(Fixed Point)}\\ \text{s.t.}\;\; & -\alpha_{ij}+\beta_{ij}+\gamma_{ij} && \geq\; \mu_{ij}(z) && \forall\,(i,j)\in M\\ & \delta+\lambda_i+\sum_{j\in N}\mathbf{1}(j\neq i)\,(\alpha_{ij}+\alpha_{ji}-\beta_{ij}-\gamma_{ji}) && \geq\; 2\,(n-1)\,\theta_i(z) && \forall\, i\in N\\ & c\,\delta+\sum_{i\in N}\lambda_i+\sum_{(i,j)\in M}\alpha_{ij} && =\; v_0\,z\\ & \alpha_{ij}\geq 0,\;\beta_{ij}\geq 0,\;\gamma_{ij}\geq 0\;\;\forall\,(i,j)\in M,\;\;\lambda_i\geq 0\;\;\forall\, i\in N,\;\;\delta\geq 0,\;\;z \textit{ is free.} \end{aligned}$$

Then, we have $f^R(\hat{z})=v_0\,\hat{z}$*; so,* $\hat{z}$ *is the fixed point of* $f^R(\cdot)/v_0$*.*

Proof: Let $\bar{z}$ be the fixed point of $f^R(\cdot)/v_0$ so that $f^R(\bar{z})=v_0\,\bar{z}$. We will show that $\hat{z}=\bar{z}$. Let $(\bar{\boldsymbol{\alpha}},\bar{\boldsymbol{\beta}},\bar{\boldsymbol{\gamma}},\bar{\boldsymbol{\lambda}},\bar{\delta})$ be an optimal solution to the Dual problem when we solve this problem with $z=\bar{z}$. Thus, we have $c\,\bar{\delta}+\sum_{i\in N}\bar{\lambda}_i+\sum_{(i,j)\in M}\bar{\alpha}_{ij}=f^R(\bar{z})=v_0\,\bar{z}$, which implies that the solution $(\bar{\boldsymbol{\alpha}},\bar{\boldsymbol{\beta}},\bar{\boldsymbol{\gamma}},\bar{\boldsymbol{\lambda}},\bar{\delta},\bar{z})$ satisfies the last constraint in the Fixed Point problem in the theorem. Furthermore, since the solution $(\bar{\boldsymbol{\alpha}},\bar{\boldsymbol{\beta}},\bar{\boldsymbol{\gamma}},\bar{\boldsymbol{\lambda}},\bar{\delta})$ is feasible to the Dual problem, it satisfies the first two constraints in the Fixed Point problem as well. Thus, $(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\gamma}}, \bar{\boldsymbol{\lambda}}, \bar{\delta}, \bar{z})$ is a feasible but not necessarily an optimal solution to the the Fixed Point problem, which implies that

$$v_0\, \bar{z} = f^R(\bar{z}) = c\,\bar{\delta} + \sum_{i\in N} \bar{\lambda}_i + \sum_{(i,j)\in M} \bar{\alpha}_{ij} \;\geq\; c\,\hat{\delta} + \sum_{i\in N} \hat{\lambda}_i + \sum_{(i,j)\in M} \hat{\alpha}_{ij} = v_0\, \hat{z},$$

where the last equality uses the fact that $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}}, \hat{\delta}, \hat{z})$ satisfies the last constraint in the Fixed Point problem. The chain of inequalities above implies that $\bar{z} \geq \hat{z}$. Next, we show that $\bar{z} \leq \hat{z}$. Note that $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}}, \hat{\delta})$ is a feasible solution to the Dual problem with $z = \hat{z}$, so that the objective value of the Dual problem at the solution $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}}, \hat{\delta})$ is no smaller than its optimal objective value. Therefore, we have $f^R(\hat{z}) \;\leq\; c\,\hat{\delta} + \sum_{i\in N} \hat{\lambda}_i + \sum_{(i,j)\in M} \hat{\alpha}_{ij} = v_0\, \hat{z}$, where the equality uses the fact that $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}}, \hat{\delta}, \hat{z})$ satisfies the last constraint in the Fixed Point problem. Since $f^R(\cdot)$ is a decreasing function, having $f^R(\bar{z}) = v_0\, \bar{z}$ and $f^R(\hat{z}) \leq v_0\, \hat{z}$ implies that $\bar{z} \leq \hat{z}$. □

Since $\mu_{ij}(z)$ and $\theta_i(z)$ are linear functions of z, the Fixed Point problem is an LP. Thus, we can compute the fixed point of $f^R(\cdot)/v_0$ by solving an LP.

4. Applying the Approximation Framework to the Uncapacitated Problem

In Sections 3.2 and 3.3, we show how to construct an upper bound $f^R(\cdot)$ on $f(\cdot)$ by using an LP relaxation of the Function Evaluation problem and how to find the fixed point of $f^R(\cdot)/v_0$ by using the Fixed Point problem. This discussion allows us to execute Steps 1 and 2 in our approximation framework that we give in Section 3.1. In this section, we focus on Step 3 in our approximation framework for the uncapacitated problem. In particular, letting $\hat{z}$ be such that $f^R(\hat{z}) = v_0\, \hat{z}$, we give an efficient approach to find a subset of products $\hat{\boldsymbol{x}}$ that satisfies $\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{x}}) - \hat{z}) \geq 0.6\, f^R(\hat{z})$. In this case, Theorem 3.1 implies that $\hat{\boldsymbol{x}}$ is a 0.6-approximate solution to the uncapacitated assortment problem. Letting $\hat{z}$ satisfy $f^R(\hat{z}) = v_0\, \hat{z}$, since the value of $\hat{z}$ is fixed, to simplify our notation, we exclude the reference to $\hat{z}$ in this section. In particular, we let $\mu_{ij} = \mu_{ij}(\hat{z})$, $\theta_i = \theta_i(\hat{z})$, $\rho_{ij} = \rho_{ij}(\hat{z})$, $f^R = f^R(\hat{z})$, $\hat{N} = N(\hat{z})$ and $\hat{M} = M(\hat{z})$. Also, since we consider the uncapacitated assortment problem, we omit the constraint $\sum_{i\in N(z)} x_i \leq c$. Thus, we write the Compact Upper Bound problem as

$$\begin{aligned} f^R \;=\; \max \quad & \sum_{(i,j)\in \hat{M}} (\mu_{ij}\, y_{ij} + \theta_i\, x_i + \theta_j\, x_j) + 2\,|N \setminus \hat{N}| \sum_{i\in \hat{N}} \theta_i\, x_i \\ \text{s.t.} \quad & y_{ij} \geq x_i + x_j - 1 \quad \forall\, (i,j) \in \hat{M} \\ & 0 \leq x_i \leq 1 \quad \forall\, i \in \hat{N}, \quad y_{ij} \geq 0 \quad \forall\, (i,j) \in \hat{M}. \end{aligned} \tag{1}$$

Our goal is to find $\hat{\boldsymbol{x}}$ that satisfies $\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{x}}) - \hat{z}) \geq 0.6\, f^R$, where f^R is the optimal objective value of the problem above. We use randomized rounding for this purpose. Let $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ be an optimal solution to the problem above. We define a random subset of products $\hat{\boldsymbol{X}} = \{\hat{X}_i : i \in N\}$ by independently rounding each coordinate of $\boldsymbol{x}^*$ as follows. For each $i \in \hat{N}$, we set

$$\hat{X}_i = \begin{cases} 1 & \text{with probability } x_i^* \\ 0 & \text{with probability } 1 - x_i^*. \end{cases} \tag{2}$$

For each $i \in N \setminus \hat{N}$, we set $\hat{X}_i = 0$. Note that the subset of products $\hat{\boldsymbol{X}}$ is a random variable with $\mathbb{E}\{\hat{X}_i\} = x_i^*$ for all $i \in \hat{N}$. In the next theorem, we give the main result of this section.

Theorem 4.1 (0.6-Approximation) *Let $\hat{\boldsymbol{X}}$ be obtained by using the randomized rounding approach. Then, we have*

$$\mathbb{E}\left\{ \sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{X}}) - \hat{z}) \right\} \ge 0.6\, f^R.$$

Proof: Here, we will show that $\mathbb{E}\{\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \ge 0.5\, f^R$ and briefly discuss how to refine the analysis to get the approximation guarantee of 0.6. The details of the refined analysis are in Appendix H. As discussed at the beginning of Section 3.2, we have $V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{X}}) - \hat{z}) = \mu_{ij}\, \hat{X}_i\, \hat{X}_j + \theta_i\, \hat{X}_i + \theta_j\, \hat{X}_j$. So, since $\{\hat{X}_i : i \in N\}$ are independent and $\mathbb{E}\{\hat{X}_i\} = x_i^*$, we have

$$\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} = \begin{cases} \mu_{ij}\, x_i^*\, x_j^* + \theta_i\, x_i^* + \theta_j\, x_j^* & \text{if } i \in \hat{N},\ j \in \hat{N},\ i \neq j \\ \theta_i\, x_i^* & \text{if } i \in \hat{N},\ j \notin \hat{N} \\ \theta_j\, x_j^* & \text{if } i \notin \hat{N},\ j \in \hat{N} \\ 0 & \text{if } i \notin \hat{N},\ j \notin \hat{N}. \end{cases}$$

Letting $[a]^+ = \max\{a, 0\}$ and considering the four cases above through the indicator function, we can write $\sum_{i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\}$ equivalently as

$$\begin{aligned}
\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} &= \sum_{(i,j)\in M} \mathbf{1}(i \in \hat{N},\ j \in \hat{N})\, (\mu_{ij}\, x_i^*\, x_j^* + \theta_i\, x_i^* + \theta_j\, x_j^*) \\
&\quad + \sum_{(i,j)\in M} \mathbf{1}(i \in \hat{N},\ j \notin \hat{N})\, \theta_i\, x_i^* + \sum_{(i,j)\in M} \mathbf{1}(i \notin \hat{N},\ j \in \hat{N})\, \theta_j\, x_j^* \\
&= \sum_{(i,j)\in \hat{M}} (\mu_{ij}\, x_i^*\, x_j^* + \theta_i\, x_i^* + \theta_j\, x_j^*) + 2\, |N \setminus \hat{N}| \sum_{i \in \hat{N}} \theta_i\, x_i^* \\
&= \sum_{(i,j)\in \hat{M}} (\mu_{ij}\, [x_i^* + x_j^* - 1]^+ + \theta_i\, x_i^* + \theta_j\, x_j^*) + 2\, |N \setminus \hat{N}| \sum_{i \in \hat{N}} \theta_i\, x_i^* + \sum_{(i,j)\in \hat{M}} \mu_{ij}\, (x_i^*\, x_j^* - [x_i^* + x_j^* - 1]^+) \\
&= f^R + \sum_{(i,j)\in \hat{M}} \mu_{ij}\, (x_i^*\, x_j^* - [x_i^* + x_j^* - 1]^+) \\
&\ge f^R + \frac{1}{4} \sum_{(i,j)\in \hat{M}} \mu_{ij}.
\end{aligned}$$

In the chain of inequalities above, the fourth equality follows because we have $\mu_{ij} \le 0$ for all $(i,j) \in \hat{M}$ by Lemma G.2 so that the decision variable y_{ij}^* takes its smallest possible value in an optimal solution to the Compact Upper Bound problem, which implies that $y^*_{ij} = [x^*_i + x^*_j - 1]^+$. The last inequality follows from the fact that we have $0 \leq ab - [a+b-1]^+ \leq 1/4$ for any $a, b \in [0,1]$. To complete the proof, we proceed to giving a lower bound on f^R. Let $\hat{x}_i = \frac{1}{2}$ for all $i \in \hat{N}$ and $\hat{y}_{ij} = 0$ for all $(i,j) \in \hat{M}$. In this case, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is a feasible solution to the LP that computes f^R at the beginning of this section, which implies that the objective value of this LP at $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ provides a lower bound on f^R. Therefore, we can lower bound f^R as

$$f^R \geq \sum_{(i,j)\in\hat{M}} \frac{\theta_i + \theta_j}{2} + |N \setminus \hat{N}| \sum_{i\in\hat{N}} \theta_i \;\geq\; \sum_{(i,j)\in\hat{M}} \frac{\theta_i + \theta_j}{2} \;\geq\; \sum_{(i,j)\in\hat{M}} \frac{\theta_i + \theta_j - \rho_{ij}}{2} \;=\; -\sum_{(i,j)\in\hat{M}} \frac{\mu_{ij}}{2},$$

where the second inequality holds since $\theta_i \geq 0$ for all $i \in \hat{N}$, the third inequality holds since $\rho_{ij} \geq 0$ for all $(i,j) \in \hat{M}$ and the equality follows from the definition of μ_{ij}. Using the lower bound above on f^R in the earlier chain of inequalities, we have $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \geq f^R + \frac{1}{4}\sum_{(i,j)\in\hat{M}} \mu_{ij} \geq \frac{1}{2} f^R$, which is the desired result. Next, we briefly discuss how to refine the analysis to improve the approximation guarantee to 0.6. The refined analysis is lengthy and we defer the details of the refined analysis to Appendix H.

The discussion above uses a lower bound on f^R that is based on a feasible solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ with $\hat{x}_i = \frac{1}{2}$ for all $i \in \hat{N}$ and $\hat{y}_{ij} = 0$ for all $(i,j) \in \hat{M}$. This lower bound may not be tight. In our refined analysis, we discuss that if $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is an extreme point of the feasible region in the LP that computes f^R at the beginning of this section, then we have $\hat{x}_i \in \{0, \frac{1}{2}, 1\}$ for all $i \in \hat{N}$. Motivated by this observation, we enumerate over a relatively small collection of feasible solutions $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ to the LP that computes f^R, where we have $\hat{x}_i \in \{0, \frac{1}{2}, 1\}$ for all $i \in \hat{N}$ and $\hat{y}_{ij} = [x_i + x_j - 1]^+$ for all $(i,j) \in \hat{M}$. We pick the best one of these solutions to obtain a tighter lower bound on f^R. In this case, we can show that $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \geq 0.6\, f^R$. □

The subset of products $\hat{\boldsymbol{X}}$ is a random variable, but in Theorem 3.1, we need a deterministic subset of products $\hat{\boldsymbol{x}}$ that satisfies the Sufficient Condition. In particular, even if the subset of products $\hat{\boldsymbol{X}}$ satisfies $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \geq 0.6\, f^R$, Theorem 3.1 does not necessarily imply that $\mathbb{E}\{\pi(\hat{\boldsymbol{X}})\} \geq 0.6\, z^*$. Nevertheless, we can use the method of conditional expectations to de-randomize the subset of products $\hat{\boldsymbol{X}}$ so that we obtain a deterministic subset of products $\hat{\boldsymbol{x}}$ that satisfies $\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{x}}) - \hat{z}) \geq 0.6\, f^R$, in which case, we obtain $\pi(\hat{\boldsymbol{x}}) \geq 0.6\, z^*$ by Theorem 3.1. Therefore, the subset of products $\hat{\boldsymbol{x}}$ that we obtain by de-randomizing the subset of products $\hat{\boldsymbol{X}}$ is a 0.6-approximate solution to the Assortment problem.

The method of conditional expectations is standard in the discrete optimization literature; see Section 15.1 in Alon and Spencer (2000) and Section 5.2 in Williamson and Shmoys (2011). We give an overview of our use of the method of conditional expectations but defer the detailed discussion of this method to Appendix I. In the method of conditional expectations, we inductively construct a subset of products $\boldsymbol{x}^{(k)} = (\hat{x}_1, \ldots, \hat{x}_k, \hat{X}_{k+1}, \ldots, \hat{X}_n)$ for all $k \in N$, where the first k products in this subset are deterministic and the last $n-k$ products are random variables. In particular, we start with $\boldsymbol{x}^{(0)} = \hat{\boldsymbol{X}}$ and construct the subset of products $\boldsymbol{x}^{(k)} = (\hat{x}_1, \ldots, \hat{x}_{k-1}, \hat{x}_k, \hat{X}_{k+1}, \ldots, \hat{X}_n)$ by using the subset of products $\boldsymbol{x}^{(k-1)} = (\hat{x}_1, \ldots, \hat{x}_{k-1}, \hat{X}_k, \hat{X}_{k+1}, \ldots, \hat{X}_n)$. These subsets of products are constructed in such a way that they satisfy $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{x}^{(k)})^{\gamma_{ij}} (R_{ij}(\boldsymbol{x}^{(k)}) - \hat{z})\} \geq 0.6\, f^R$ for all $k \in N$. In this case, the subset of products $\boldsymbol{x}^{(n)} = (\hat{x}_1, \ldots, \hat{x}_n)$ corresponds to a deterministic subset of products and it satisfies $\sum_{(i,j)\in M} V_{ij}(\boldsymbol{x}^{(n)})^{\gamma_{ij}} (R_{ij}(\boldsymbol{x}^{(n)}) - \hat{z}) \geq 0.6\, f^R$. In other words, the deterministic subset of products $\boldsymbol{x}^{(n)}$ satisfies the Sufficient Condition with $\alpha = 0.6$. Constructing the subset of products $\boldsymbol{x}^{(k)}$ by using the subset of products $\boldsymbol{x}^{(k-1)}$ takes $O(n)$ operations, in which case, the method of conditional expectations takes $O(n^2)$ operations.

Thus, for the uncapacitated problem, we execute the approximation framework given in Section 3.1 as follows, yielding a 0.6-approximation algorithm. **(a)** Solve the LP given in the Fixed Point problem in Theorem 3.3 to compute the fixed point $\hat{z}$ of $f^R(\cdot)/v_0$. **(b)** Recalling that $\mu_{ij} = \mu_{ij}(\hat{z})$, $\theta_i = \theta_i(\hat{z})$, $\hat{M} = M(\hat{z})$ and $\hat{N} = N(\hat{z})$, solve the LP in (1) that computes f^R at the beginning of this section to obtain the optimal solution $(\boldsymbol{x}^*, \boldsymbol{y}^*)$. The solution $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ characterizes the random subset of products $\hat{\boldsymbol{X}}$ through our randomized rounding approach given in (2) in this section. **(c)** De-randomize the subset of products $\hat{\boldsymbol{X}}$ by using the method of conditional expectations to obtain a deterministic subset of products $\hat{\boldsymbol{x}}$ satisfying $\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{x}}) - \hat{z}) \geq 0.6\, f^R$. By Theorem 4.1, the subset of products $\hat{\boldsymbol{x}}$ is a 0.6-approximate solution to the uncapacitated problem. In terms of computational effort, we solve two LP formulations, each with $O(n^2)$ decision variables and $O(n^2)$ constraints. The method of conditional expectations takes $O(n^2)$ operations.

Approximation Framework with a Semidefinite Programming Relaxation

We can also use an SDP relaxation of the Function Evaluation problem to construct an upper bound $f^R(\cdot)$ on $f(\cdot)$. This SDP has $O(n^2)$ decision variables and $O(n^2)$ constraints. We can compute the fixed point $\hat{z}$ of $f^R(\cdot)/v_0$ by solving an SDP of the same size. In this case, we can construct a subset of products $\hat{\boldsymbol{X}}$ that satisfies $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \geq 0.79\, f^R(\hat{z})$, where $f^R(\cdot)$ refers to the upper bound constructed by using the SDP relaxation. This approach provides a stronger approximation guarantee, but comes at the expense of solving an SDP. We summarize this approach in the next theorem and defer the details to Appendix J.

Theorem 4.2 (SDP Relaxation) *There exists an algorithm to find the fixed point $\hat{z}$ of a function $f^R(\cdot)/v_0$ that satisfies $f^R(z) \geq f(z)$ for all $z \in \mathbb{R}$ and to construct a random subset of products $\hat{\boldsymbol{X}}$ that satisfies $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \geq 0.79\, f^R(\hat{z})$. This algorithm requires solving two SDP formulations, each with $O(n^2)$ decision variables and $O(n^2)$ constraints.*

5. Applying the Approximation Framework to the Capacitated Problem

We consider Step 3 in our approximation framework for the capacitated problem. Letting $\hat{z}$ satisfy $f^R(\hat{z}) = v_0\,\hat{z}$ for $f^R(\cdot)$ given by the Compact Upper Bound problem, we focus on finding a subset of products $\hat{\boldsymbol{x}}$ such that $\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{x}}) - \hat{z}) \geq \frac{1}{4}\, f^R(\hat{z})$ and $\sum_{i\in N} \hat{x}_i \leq c$. Then, by Theorem 3.1, the subset of products $\hat{\boldsymbol{x}}$ is a 0.25-approximate solution to the Assortment problem.

5.1. Half-Integral Solutions Through Iterative Rounding

As in the previous section, since the value of $\hat{z}$ is fixed, to simplify our notation, we drop the argument $\hat{z}$ in the Compact Upper Bound problem. In particular, we let $\mu_{ij} = \mu_{ij}(\hat{z})$, $\theta_i = \theta_i(\hat{z})$, $\rho_{ij} = \rho_{ij}(\hat{z})$, $f^R = f^R(\hat{z})$, $\hat{N} = N(\hat{z})$ and $\hat{M} = M(\hat{z})$. Noting that the optimal objective value of the Compact Upper Bound problem is f^R, we write the Compact Upper Bound problem as

$$
\begin{aligned}
f^R \;=\; \max \quad & \sum_{(i,j)\in\hat{M}} (\mu_{ij}\, y_{ij} + \theta_i\, x_i + \theta_j\, x_j) + 2\,|N\setminus\hat{N}| \sum_{i\in\hat{N}} \theta_i\, x_i \\
\text{s.t.} \quad & y_{ij} \geq x_i + x_j - 1 \quad \forall\,(i,j)\in\hat{M} \\
& \sum_{i\in\hat{N}} x_i \leq c \\
& 0 \leq x_i \leq 1 \;\;\forall\, i\in\hat{N}, \quad y_{ij} \geq 0 \;\;\forall\,(i,j)\in\hat{M}.
\end{aligned}
$$

We can construct counterexamples to show that all of the decision variables $\{x_i : i\in\hat{N}\}$ can take strictly positive values in an extreme point of the set of feasible solutions to the LP above. In Example K.1 in Appendix K, we give one such counterexample. In this case, letting $\{x_i^* : i\in\hat{N}\}$ be an optimal solution to the LP above, if we construct a random subset of products by using a naive randomized rounding approach that sets $x_i = 1$ with probability x_i^* and $x_i = 0$ with probability $1 - x_i^*$, then the random subset of products may violate the capacity constraint. To address this difficulty, we will use an iterative rounding algorithm, where we iteratively solve a modified version of the LP above after fixing some of the decision variables $\{x_i : i\in\hat{N}\}$ at $\frac{1}{2}$. In this way, we construct a feasible solution to the LP above, where the decision variables $\{x_i : i\in\hat{N}\}$ ultimately all take values in $\{0, \frac{1}{2}, 1\}$. In the feasible solution that we construct, since the smallest strictly positive value for the decision variables $\{x_i : i\in\hat{N}\}$ is $\frac{1}{2}$, noting the constraint $\sum_{i\in\hat{N}} x_i \leq c$, no more than $2\,c$ of the decision variables $\{x_i : i\in\hat{N}\}$ can take strictly positive values. In this case, we will be able to use a coupled randomized rounding algorithm on the feasible solution to obtain a random subset of products that includes no more than c products.

Since we will solve the LP above after fixing some of the decision variables $\{x_i : i\in\hat{N}\}$ at $\frac{1}{2}$, we study the extreme points of the set of feasible solutions to this LP with some of the decision variables $\{x_i : i \in \hat{N}\}$ fixed at $\frac{1}{2}$. In particular, if we fix the decision variables $\{x_i : i \in H\}$ at $\frac{1}{2}$, then the set of feasible solutions to the LP is given by the polyhedron

$$\mathcal{P}(H) = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in [0,1]^{|\hat{N}|} \times \mathbb{R}_+^{|\hat{M}|} \ : \ y_{ij} \geq x_i + x_j - 1 \ \ \forall (i,j) \in \hat{M}, \ \sum_{i \in \hat{N}} x_i \leq c, \quad x_i = \frac{1}{2} \ \ \forall i \in H \right\}.$$

In the next lemma, we show a useful structural property of the extreme points of $\mathcal{P}(H)$. In particular, it turns out that if none of the decision variables $\{x_i : i \in \hat{N}\}$ take a fractional value larger than $\frac{1}{2}$ in an extreme point of $\mathcal{P}(H)$, then all of the decision variables $\{x_i : i \in \hat{N}\}$ must take values in $\{0, \frac{1}{2}, 1\}$. This result shortly becomes useful for arguing that our iterative rounding algorithm terminates with a feasible solution to the LP that computes f^R at the beginning of this section, where all of the decision variables $\{x_i : i \in \hat{N}\}$ are guaranteed to take values in $\{0, \frac{1}{2}, 1\}$. We defer the proof of the next lemma to Appendix K.

Lemma 5.1 (Extreme Points) *For any $H \subseteq \hat{N}$, let $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ be an extreme point of $\mathcal{P}(H)$. If there is no product $i \in \hat{N}$ such that $\frac{1}{2} < \hat{x}_i < 1$, then we have $\hat{x}_i \in \{0, \frac{1}{2}, 1\}$ for all $i \in \hat{N}$.*

We consider the following iterative rounding algorithm to construct a feasible solution to the LP that computes f^R at the beginning of this section.

Iterative Rounding

Step 1: Set the iteration counter to $k = 1$ and initialize the set of fixed variables $H^k = \varnothing$.

Step 2: Let $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ be an optimal solution to the LP

$$f^k = \max \left\{ \sum_{(i,j) \in \hat{M}} (\mu_{ij}\, y_{ij} + \theta_i\, x_i + \theta_j\, x_j) + 2\,|N \setminus \hat{N}| \sum_{i \in \hat{N}} \theta_i\, x_i : (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}(H^k) \right\}. \quad \text{(Variable Fixing)}$$

Step 3: If there exists some $i_k \in \hat{N}$ such that $\frac{1}{2} < x^k_{i_k} < 1$, then set $H^{k+1} = H^k \cup \{i_k\}$, increase k by one and go to Step 2. Otherwise, stop.

Without loss of generality, we assume that the optimal solution $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ to the Variable Fixing problem in Step 2 is an extreme point of $\mathcal{P}(H^k)$. In Step 3 of the iterative rounding algorithm, if there does not exist some $i \in \hat{N}$ such that $\frac{1}{2} < x^k_i < 1$, then we stop. By Lemma 5.1, if there does not exist some $i \in \hat{N}$ such that $\frac{1}{2} < x^k_i < 1$, then we have $x^k_i \in \{0, \frac{1}{2}, 1\}$ for all $i \in \hat{N}$. Therefore, if the iterative rounding algorithm stops at iteration k with the solution $(\boldsymbol{x}^k, \boldsymbol{y}^k)$, then we must have $x^k_i \in \{0, \frac{1}{2}, 1\}$ for all $i \in \hat{N}$. Similar iterative rounding approaches are often used to design approximation algorithms for discrete optimization problems; see Lau et al. (2011).

We use $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ to denote an optimal solution to the Variable Fixing problem at the last iteration of the iterative rounding algorithm. By the discussion in the previous paragraph, we have

$x_i^* \in \{0, \frac{1}{2}, 1\}$ for all $i \in \hat{N}$. Also, since this solution is feasible to the Variable Fixing problem, we have $\sum_{i \in \hat{N}} x_i^* \leq c$. Therefore, noting that the smallest strictly positive value of $\{x_i^* : i \in \hat{N}\}$ is $\frac{1}{2}$, no more than $2c$ of the decision variables $\{x_i^* : i \in \hat{N}\}$ can take strictly positive values. Nevertheless, including each product $i \in \hat{N}$ in a subset with probability x_i^* may not provide a solution that satisfies the capacity constraint, because this subset can include as many as $2c$ products. Instead, we use a coupled randomized rounding approach to obtain a random subset of products that satisfies the capacity constraint with probability one.

5.2. Feasible Subsets Through Coupled Randomized Rounding

Letting $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ be an optimal solution to the Variable Fixing problem at the last iteration of the iterative rounding algorithm, we use the following coupled randomized rounding approach to obtain a random subset of products $\hat{\boldsymbol{X}} = \{\hat{X}_i : i \in N\}$.

Coupled Randomized Rounding

Recall that we have $x_i^* \in \{0, \frac{1}{2}, 1\}$ for all $i \in \hat{N}$. We assume that the number of the decision variables in $\{x_i^* : i \in \hat{N}\}$ that take the value $\frac{1}{2}$ is even. The idea of coupled randomized rounding is similar under the odd case. Letting 2ℓ be the number of the decision variables in $\{x_i^* : i \in \hat{N}\}$ that take the value $\frac{1}{2}$, we use $\{x_{i_1}^*, x_{j_1}^*, x_{i_2}^*, x_{j_2}^*, \ldots, x_{i_\ell}^*, x_{j_\ell}^*\}$ to denote these decision variables. We view each of $(i_1, j_1), (i_2, j_2), \ldots, (i_\ell, j_\ell)$ as a pair. Using the solution $(\boldsymbol{x}^*, \boldsymbol{y}^*)$, we define the random subset of products $\hat{\boldsymbol{X}} = \{\hat{X}_i : i \in N\}$ as follows. For each pair (i_m, j_m), we set $(\hat{X}_{i_m}, \hat{X}_{j_m}) = (1, 0)$ with probability 0.5, whereas we set $(\hat{X}_{i_m}, \hat{X}_{j_m}) = (0, 1)$ with probability 0.5. Note that the decision variables $\{x_i^* : i \in \hat{N}\}$ that are not in the set $\{x_{i_1}^*, x_{j_1}^*, x_{i_2}^*, x_{j_2}^*, \ldots, x_{i_\ell}^*, x_{j_\ell}^*\}$ take the value zero or one. Thus, for all $i \in \hat{N} \setminus \{i_1, j_1, i_2, j_2, \ldots, i_\ell, j_\ell\}$, we set $\hat{X}_i = 1$ if $x_i^* = 1$, whereas we set $\hat{X}_i = 0$ if $x_i^* = 0$. Finally, we set $\hat{X}_i = 0$ for all $i \in N \setminus \hat{N}$.

In our construction, $\hat{X}_{i_m}$ and $\hat{X}_{j_m}$ for the pair (i_m, j_m) are dependent, but the components of $\hat{\boldsymbol{X}}$ corresponding to different pairs are independent. Also, we have $\mathbb{E}\{\hat{X}_i\} = x_i^*$ for all $i \in \hat{N}$. Lastly, the subset of products $\hat{\boldsymbol{X}}$ always satisfies the capacity constraint $\sum_{i \in N} \hat{X}_i \leq c$. In particular, we let $S = \{i \in \hat{N} : x_i^* = \frac{1}{2}\}$ and $L = \{i \in \hat{N} : x_i^* = 1\}$. We have $|S| = 2\ell$. By the definition of $\hat{\boldsymbol{X}}$, there are exactly $\ell + |L|$ products in $\hat{\boldsymbol{X}}$, so $\sum_{i \in N} \hat{X}_i = \ell + |L|$. Since $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is a feasible solution to the Variable Fixing problem, we have $\sum_{i \in \hat{N}} x_i^* \leq c$, but we can write the last sum as $\sum_{i \in \hat{N}} x_i^* = 2\ell \cdot \frac{1}{2} + |L| = \ell + |L|$, indicating $\hat{\boldsymbol{X}}$ satisfies the capacity constraint $\sum_{i \in N} \hat{X}_i = \ell + |L| \leq c$.

The main result of this section is stated in the following theorem.

Theorem 5.2 (0.25-Approximation) *Let $\hat{\boldsymbol{X}}$ be obtained by using the coupled randomized rounding approach. Then, we have*

$$\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}})-\hat{z})\} \geq \frac{1}{4}\, f^R.$$

We give the proof of this theorem in Section 5.3. The subset of products $\hat{\boldsymbol{X}}$ is random, but we can use the method of conditional expectations discussed in Section 4 and Appendix I to obtain a deterministic subset of products $\hat{\boldsymbol{x}}$ that satisfies $\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{x}})-\hat{z}) \geq \frac{1}{4} f^R$ and $\sum_{i\in N} \hat{x}_i \leq c$. The only difference is that we condition on the pair $(\hat{X}_{i_m}, \hat{X}_{j_m})$ when we consider the products (i_m, j_m) that are paired in the coupled randomized rounding approach. For product i that is not paired, we condition on $\hat{X}_i$. Using the same argument in Appendix I, for the capacitated problem, the method of conditional expectations takes $O(n^2)$ operations.

Thus, for the capacitated problem, we can execute the approximation framework given in Section 3.1 as follows, yielding a 0.25-approximation algorithm. **(a)** Solve the LP given in the Fixed Point problem in Theorem 3.3 to compute the fixed point $\hat{z}$ of $f^R(\cdot)/v_0$. **(b)** Recalling that $\mu_{ij} = \mu_{ij}(\hat{z})$, $\theta_i = \theta_i(\hat{z})$, $\hat{M} = M(\hat{z})$ and $\hat{N} = N(\hat{z})$ in the Variable Fixing problem, we execute the iterative rounding algorithm. Let $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ be an optimal solution to the Variable Fixing problem at the last iteration of the iterative rounding algorithm. Through the coupled randomized rounding approach, the solution $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ characterizes the random subset of products $\hat{\boldsymbol{X}}$. **(c)** De-randomize the subset of products $\hat{\boldsymbol{X}}$ by using the method of conditional expectations to obtain a deterministic subset of products $\hat{\boldsymbol{x}}$ satisfying $\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{x}})-\hat{z}) \geq \frac{1}{4} f^R$ and $\sum_{i\in N} \hat{x}_i \leq c$. In terms of computational effort, to compute the fixed point of $f^R(\cdot)/v_0$, we solve an LP with $O(n^2)$ decision variables and $O(n^2)$ constraints. The iterative rounding algorithm terminates in $O(n)$ iterations. At each iteration of the iterative rounding algorithm, we solve an LP with $O(n^2)$ decision variables and $O(n^2)$ constraints. The method of conditional expectations takes $O(n^2)$ operations.

5.3. Proof of Theorem 5.2

We devote this section in its entirety to the proof of Theorem 5.2. The proof relies on the next two lemmas. As the iterations of the iterative rounding algorithm progress, we fix additional decision variables at the value $\frac{1}{2}$ in the Variable Fixing problem. Therefore, noting that the optimal objective value of the Variable Fixing problem at iteration k is f^k, since the Variable Fixing problem at iteration $k+1$ has one more decision variable fixed at $\frac{1}{2}$, we have $f^{k+1} \leq f^k$ for all $k = 1, 2, \ldots$. In the next lemma, we give an upper bound on the degradation in the optimal objective value of the Variable Fixing problem at the successive iterations of the iterative rounding algorithm.

Lemma 5.3 (Reduction in Objective) *For all $k = 1, 2, \ldots$, we have $f^k - f^{k+1} \le (|N| - 1)\,\theta_{i_k}$.*

Proof: We have $\sum_{(i,j)\in\hat{M}} \theta_i\, x_i = \sum_{i\in\hat{N}} \sum_{j\in\hat{N}} \mathbf{1}(i \ne j)\,\theta_i\, x_i = (|\hat{N}| - 1) \sum_{i\in\hat{N}} \theta_i\, x_i$. Similarly, we have $\sum_{(i,j)\in\hat{M}} \theta_j\, x_j = (|\hat{N}| - 1) \sum_{i\in\hat{N}} \theta_i\, x_i$. In this case, since $|\hat{N}| - 1 + |N \setminus \hat{N}| = |N| - 1$, we can write the objective function of the Variable Fixing problem as

$$\sum_{(i,j)\in\hat{M}} (\mu_{ij}\, y_{ij} + \theta_i\, x_i + \theta_j\, x_j) + 2\,|N \setminus \hat{N}| \sum_{i\in\hat{N}} \theta_i\, x_i$$
$$= \sum_{(i,j)\in\hat{M}} \mu_{ij}\, y_{ij} + 2\,(|\hat{N}| - 1) \sum_{i\in\hat{N}} \theta_i\, x_i + 2\,|N \setminus \hat{N}| \sum_{i\in\hat{N}} \theta_i\, x_i = \sum_{(i,j)\in\hat{M}} \mu_{ij}\, y_{ij} + 2\,(|N| - 1) \sum_{i\in\hat{N}} \theta_i\, x_i.$$

Since the iterative rounding algorithm did not stop at iteration k, we have $x^k_{i_k} \in (\frac{1}{2}, 1)$. We define the solution $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ to the Variable Fixing problem as follows. We set $\tilde{x}_i = x^k_i$ for all $i \in \hat{N} \setminus \{i_k\}$ and $\tilde{x}_{i_k} = \frac{1}{2}$. Also, we set $\tilde{y}_{ij} = [\tilde{x}_i + \tilde{x}_j - 1]^+$ for all $(i,j) \in \hat{M}$.

We claim that $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ is a feasible solution to the Variable Fixing problem at iteration $k+1$. In particular, since $x^k_{i_k} \in (\frac{1}{2}, 1)$, but $\tilde{x}_{i_k} = \frac{1}{2}$, we have $\tilde{x}_i \le x^k_i$ for all $i \in \hat{N}$. Therefore, $\sum_{i\in\hat{N}} \tilde{x}_i \le \sum_{i\in\hat{N}} x^k_i \le c$, where the last inequality uses the fact that $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ is a feasible solution to the Variable Fixing problem at iteration k so that it satisfies the capacity constraint. Also, we have $x^k_i = \frac{1}{2}$ for all $i \in H^k$. Since $H^{k+1} = H^k \cup \{i_k\}$ and $\tilde{x}_{i_k} = \frac{1}{2}$, we have $\tilde{x}_i = \frac{1}{2}$ for all $i \in H^{k+1}$. Thus, it follows that $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ is a feasible solution to the Variable Fixing problem at iteration $k+1$. Since $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ is not necessarily an optimal to the same problem at iteration $k+1$, we have

$$\begin{aligned}
f^k - f^{k+1} &\le f^k - \sum_{(i,j)\in\hat{M}} \mu_{ij}\, \tilde{y}_{ij} - 2\,(|N| - 1) \sum_{i\in\hat{N}} \theta_i\, \tilde{x}_i \\
&= \sum_{(i,j)\in\hat{M}} \mu_{ij}\,(y^k_{ij} - \tilde{y}_{ij}) + 2\,(|N| - 1)\,\theta_{i_k} \left(x^k_{i_k} - \frac{1}{2} \right) + 2\,(|N| - 1) \sum_{i\in\hat{N}\setminus\{i_k\}} \theta_i\,(x^k_i - \tilde{x}_i) \\
&\le \sum_{(i,j)\in\hat{M}} \mu_{ij}\,(y^k_{ij} - \tilde{y}_{ij}) + (|N| - 1)\,\theta_{i_k},
\end{aligned}$$

where the equality uses the fact that $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ is an optimal solution to the Variable Fixing problem at iteration k and the second inequality uses the fact that $x^k_{i_k} \le 1$ and $\tilde{x}_i = x^k_i$ for all $i \in \hat{N} \setminus \{i_k\}$. Assume for the moment that $y^k_{ij} = [x^k_i + x^k_j - 1]^+$ for all $(i,j) \in \hat{M}$. By our construction, we have $\tilde{y}_{ij} = [\tilde{x}_i + \tilde{x}_j - 1]^+$ as well. Earlier in this paragraph, we showed that $x^k_i \ge \tilde{x}_i$ for all $i \in \hat{N}$. Since $[\cdot]^+$ is an increasing function, we get $y^k_{ij} \ge \tilde{y}_{ij}$ for all $(i,j) \in \hat{M}$. By Lemma G.2, we have $\mu_{ij} \le 0$ for all $(i,j) \in \hat{M}$. In this case, we obtain $\sum_{(i,j)\in\hat{M}} \mu_{ij}\,(y^k_{ij} - \tilde{y}_{ij}) + (|N| - 1)\,\theta_{i_k} \le (|N| - 1)\,\theta_{i_k}$ and the desired result follows from the chain of inequalities above.

To complete the proof, we argue that $y^k_{ij} = [x^k_i + x^k_j - 1]^+$ for all $(i,j) \in \hat{M}$. Note that $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ is an extreme point of $\mathcal{P}(H^k)$. If $y^k_{ij} > [x^k_i + x^k_j - 1]^+$ for some $(i,j) \in \hat{M}$, then we can perturb only this component of $\boldsymbol{y}^k$ by $+\epsilon$ and $-\epsilon$ for a small enough $\epsilon > 0$, while keeping the other components of $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ constant. The two points that we obtain are in $\mathcal{P}(H^k)$ and $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ is a convex combination of the two points, which contradicts the fact that $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ is an extreme point of $\mathcal{P}(H^k)$. □

In the next lemma, we accumulate the degradation in the optimal objective value of the Variable Fixing problem over the iterations of the iterative rounding algorithm to compare the optimal objective value of the Variable Fixing problem at the last iteration with f^R.

Lemma 5.4 (Objective at Last Iteration) *If* $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ *is an optimal solution to the* Variable Fixing *problem at the last iteration of the iterative rounding algorithm, then we have*

$$\sum_{(i,j)\in \hat{M}} (\mu_{ij}\, y^*_{ij} + \theta_i\, x^*_i + \theta_j\, x^*_j) + 2\,|N\setminus \hat{N}| \sum_{i\in\hat{N}} \theta_i\, x^*_i \;\geq\; \frac{1}{2}\, f^R.$$

Proof: By the discussion at the beginning of the proof of Lemma 5.3, we can write the objective function of the Variable Fixing problem as $\sum_{(i,j)\in\hat{M}} \mu_{ij}\, y_{ij} + 2\,(|N|-1) \sum_{i\in\hat{N}} \theta_i\, x_i$. We let K be the last iteration of the iterative rounding algorithm. Consider the Variable Fixing problem at iteration K. In this problem, we fix the decision variables in $\{x_i : i \in H^K\}$ at $\frac{1}{2}$ and have $H^K = \{i_1, \ldots, i_{K-1}\}$ by the construction of the iterative rounding algorithm. If we set $x_i = \frac{1}{2}$ for all $i \in H^K$, $x_i = 0$ for all $i \in \hat{N}\setminus\{H^K\}$ and $y_{ij} = 0$ for all $(i,j)\in\hat{M}$, then we obtain a feasible solution to the Variable Fixing problem at iteration K and this solution provides the objective value of $2\,(|N|-1)\sum_{i\in H^K} \frac{\theta_i}{2} = (|N|-1)\sum_{i\in H^K}\theta_i$. Since the optimal objective value of the Variable Fixing problem at iteration K is f^K, we get $f^K \geq (|N|-1)\sum_{i\in H^K}\theta_i$. By Lemma 5.3, we also have $f^k - f^{k+1} \leq (|N|-1)\,\theta_{i_k}$ for all $k = 1,\ldots,K-1$. In this case, we obtain

$$\begin{aligned}
\sum_{(i,j)\in \hat{M}} (\mu_{ij}\, y^*_{ij} + \theta_i\, x^*_i + \theta_j\, x^*_j) + 2\,|N\setminus \hat{N}| \sum_{i\in\hat{N}} \theta_i\, x^*_i \;&=\; f^K - \frac{1}{2}\,(|N|-1)\sum_{i\in H^K}\theta_i + \frac{1}{2}\,(|N|-1)\sum_{i\in H^K}\theta_i \\
&\geq\; f^K - \frac{1}{2}\,(|N|-1)\sum_{i\in H^K}\theta_i + \frac{1}{2}\sum_{k=1}^{K-1}(f^k - f^{k+1}) \;\geq\; \frac{1}{2}\, f^K + \frac{1}{2}\sum_{k=1}^{K-1}(f^k - f^{k+1}) \;=\; \frac{1}{2}\, f^1.
\end{aligned}$$

In the chain of inequalities above, the first equality is from the fact that $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to the Variable Fixing problem at iteration K. The first inequality is by the fact that $f^k - f^{k+1} \leq (|N|-1)\,\theta_{i_k}$ for all $k = 1,\ldots,K-1$ and $H^K = \{i_1,\ldots,i_{K-1}\}$. The second inequality holds since $f^K \geq (|N|-1)\sum_{i\in H^K}\theta_i$. Since $H^1 = \varnothing$, the Variable Fixing problem at the first iteration is identical to the LP that computes f^R at the beginning of Section 5.1. Therefore, we get $f^1 = f^R$, in which case, the desired result follows from the chain of inequalities above. □

Finally, here is the proof of Theorem 5.2.

Proof of Theorem 5.2: Since $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an extreme point solution to the Variable Fixing problem, by the same discussion at the end of the proof of Lemma 5.3, we have $y^*_{ij} = [x^*_i + x^*_j - 1]^+$

for all $(i,j) \in \hat{M}$. Also, by the same discussion at the beginning of Section 3.2, we have $V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z}) \;=\; \mu_{ij}\,\hat{X}_i\,\hat{X}_j + \theta_i\,\hat{X}_i + \theta_j\,\hat{X}_j$. There are four cases to consider.

Case 1: Suppose $i \in \hat{N}$ and $j \in \hat{N}$ with $i \neq j$. We claim that $\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \geq \frac{1}{2}\,(\mu_{ij}\,y^*_{ij} + \theta_i\,x^*_i + \theta_j\,x^*_j)$ in this case.

First, assume that the products i and j are paired in the coupled randomized rounding approach. Thus, we must have $x^*_i = \frac{1}{2}$ and $x^*_j = \frac{1}{2}$, so that $y^*_{ij} = [x^*_i + x^*_j - 1]^+ = 0$. Also, since products i and j are paired, we have $(\hat{X}_i, \hat{X}_j) = (1,0)$ or $(\hat{X}_i, \hat{X}_j) = (0,1)$, so that $\hat{X}_i\,\hat{X}_j = 0$. So, we get

$$\begin{aligned}\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \;&=\; \mu_{ij}\mathbb{E}\{\hat{X}_i\,\hat{X}_j\} + \theta_i\,\mathbb{E}\{\hat{X}_i\} + \theta_j\,\mathbb{E}\{\hat{X}_j\} \\ &=\; \frac{\theta_i}{2} + \frac{\theta_j}{2} \;=\; \frac{1}{2}\,(\mu_{ij}\,y^*_{ij} + \theta_i\,x^*_i + \theta_j\,x^*_j).\end{aligned}$$

Second, assume that the products i and j are not paired. Thus, $\hat{X}_i$ and $\hat{X}_j$ are independent, in which case, $\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} = \mu_{ij}\,\mathbb{E}\{\hat{X}_i\}\,\mathbb{E}\{\hat{X}_j\} + \theta_i\,\mathbb{E}\{\hat{X}_i\} + \theta_j\,\mathbb{E}\{\hat{X}_j\} = \mu_{ij}\,x^*_i\,x^*_j + \theta_i\,x^*_i + \theta_j\,x^*_j$. If $x^*_i \in \{0,1\}$ or $x^*_j \in \{0,1\}$, then we have $[x^*_i + x^*_j - 1]^+ = x^*_i\,x^*_j$. Thus, we get

$$\begin{aligned}\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \;&=\; \mu_{ij}\,x^*_i\,x^*_j + \theta_i\,x^*_i + \theta_j\,x^*_j \;\;\geq\;\; \frac{1}{2}\,(\mu_{ij}\,x^*_i\,x^*_j + \theta_i\,x^*_i + \theta_j\,x^*_j) \\ &=\; \frac{1}{2}\,(\mu_{ij}\,[x^*_i + x^*_j - 1]^+ + \theta_i\,x^*_i + \theta_j\,x^*_j) \;=\; \frac{1}{2}\,(\mu_{ij}\,y^*_{ij} + \theta_i\,x^*_i + \theta_j\,x^*_j),\end{aligned}$$

where the inequality uses the fact that $\mu_{ij}\,x^*_i\,x^*_j + \theta_i\,x^*_i + \theta_j\,x^*_j = \rho_{ij}\,x^*_i\,x^*_j + \theta_i\,x^*_i\,(1 - x^*_j) + \theta_j\,(1 - x^*_i)\,x^*_j$ and $\rho_{ij} \geq 0$, $\theta_i \geq 0$ and $\theta_j \geq 0$ for all $(i,j) \in \hat{M}$. If $x^*_i = \frac{1}{2}$ and $x^*_j = \frac{1}{2}$, then $y^*_{ij} = [x^*_i + x^*_j - 1]^+ = 0$, so since $\rho_{ij} \geq 0$, we obtain

$$\begin{aligned}\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \;&=\; \mu_{ij}\,x^*_i\,x^*_j + \theta_i\,x^*_i + \theta_j\,x^*_j \;=\; \frac{\rho_{ij} - \theta_i - \theta_j}{4} + \frac{\theta_i}{2} + \frac{\theta_j}{2} \\ &\geq\; \frac{\theta_i}{4} + \frac{\theta_j}{4} \;=\; \frac{1}{2}\,(\mu^*_{ij}\,y^*_{ij} + \theta_i\,x^*_i + \theta_j\,x^*_j),\end{aligned}$$

where the last equality uses the fact that $y^*_{ij} = 0$. In all cases, we get $\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \geq \frac{1}{2}(\mu_{ij}\,y^*_{ij} + \theta_i\,x^*_i + \theta_j\,x^*_j)$ for all $i \in \hat{N}$ and $j \in \hat{N}$ with $i \neq j$, which is the desired the claim.

Case 2: Suppose $i \in \hat{N}$ and $j \notin \hat{N}$. Since $\hat{X}_j = 0$, we get $V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z}) = \mu_{ij}\,\hat{X}_i\,\hat{X}_j + \theta_i\,\hat{X}_i + \theta_j\,\hat{X}_j = \theta_i\,\hat{X}_i$. Thus, we have $\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} = \theta_i\,\mathbb{E}\{\hat{X}_i\} = \theta_i\,x^*_i \geq \frac{1}{2}\,\theta_i\,x^*_i$.

Case 3: Suppose $i \notin \hat{N}$ and $j \in \hat{N}$. Using the same argument as in Case 2, it follows that $\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \geq \frac{1}{2}\,\theta_j\,x^*_j$.

Case 4: Suppose $i \notin \hat{N}$ and $j \notin \hat{N}$ with $i \neq j$. In this case, noting that $\hat{X}_i = 0$ and $\hat{X}_j = 0$, we get $\mathbb{E}\{V_{ij}(\hat{X})^{\gamma_{ij}}(R_{ij}(\hat{X}) - \hat{z})\} = 0$.

Collecting the four cases above, if $i \in \hat{N}$ and $j \in \hat{N}$, then we have $\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \geq \frac{1}{2}\,(\mu_{ij}\,y^*_{ij} + \theta_i\,x^*_i + \theta_j\,x^*_j)$. Also, if $i \in \hat{N}$ and $j \notin \hat{N}$, then we have $\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \geq \frac{1}{2}\,\theta_i\,x^*_i$.

Similarly, if $i \notin \hat{N}$ and $j \in \hat{N}$, then we have $\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \geq \frac{1}{2}\,\theta_j\, x_j^*$. Finally, if $i \notin \hat{N}$ and $j \notin \hat{N}$, then we have $\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} = 0$. Therefore, we get

$$\begin{aligned}
\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} &\geq \frac{1}{2}\Bigg\{ \sum_{(i,j)\in M} \mathbf{1}(i \in \hat{N},\ j \in \hat{N})\,(\mu_{ij}\, y_{ij}^* + \theta_i\, x_i^* + \theta_j\, x_j^*) \\
&\qquad + \sum_{(i,j)\in M} \mathbf{1}(i \in \hat{N},\ j \notin \hat{N})\,\theta_i\, x_i^* + \sum_{(i,j)\in M} \mathbf{1}(i \notin \hat{N},\ j \in \hat{N})\,\theta_j\, x_j^* \Bigg\} \\
&= \frac{1}{2}\Bigg\{ \sum_{(i,j)\in \hat{M}} (\mu_{ij}\, y_{ij}^* + \theta_i\, x_i^* + \theta_j\, x_j^*) + 2\,|N \setminus \hat{N}| \sum_{i\in\hat{N}} \theta_i\, x_i^* \Bigg\} \\
&\geq \frac{1}{4} f^R,
\end{aligned}$$

where the equality follows because $\sum_{(i,j)\in M} \mathbf{1}(i \in \hat{N},\ j \notin \hat{N})\,\theta_i\, x_i^* = |N \setminus \hat{N}| \sum_{i\in\hat{N}} \theta_i\, x_i^* = \sum_{(i,j)\in M} \mathbf{1}(i \notin \hat{N},\ j \in \hat{N})\,\theta_j\, x_j^*$ and the last inequality follows from Lemma 5.4. $\square$

As an alternative to the coupled randomized rounding approach, letting $S = \{i \in \hat{N} : x_i^* = \frac{1}{2}\}$ and recalling that we have $|S| = 2\ell$ in our discussion of the coupled randomized rounding approach, we can simply sample ℓ elements of S without replacement. Using $\hat{S}$ to denote these elements, we can define the random subset of products $\hat{\boldsymbol{X}} = \{\hat{X}_i : i \in N\}$ as follows. For each $i \in \hat{N}$, we set $\hat{X}_i = 1$ if $i \in \hat{S}$ and $\hat{X}_i = 0$ if $i \notin \hat{S}$. For each $i \in N \setminus \hat{N}$, we set $\hat{X}_i = 0$. In this case, the subset of products $\hat{\boldsymbol{X}}$ still satisfies Theorem 5.2. The only difference in the proof of Theorem 5.2 is that since the subset $\hat{S}$ is obtained by sampling ℓ elements of S without replacement, for $i, j \in \hat{N}$ with $i \neq j$, we have $\hat{X}_i = 1$ and $\hat{X}_j = 1$ with probability $\binom{2\ell-2}{\ell-2}/\binom{2\ell}{\ell}$. Furthermore, for any $i, j \in \hat{N}$ with $i \neq j$, $\hat{X}_i$ and $\hat{X}_j$ are never independent, so computing the conditional expectations involved in the method of conditional expectations gets slightly more complicated.

6. Computational Experiments

In this section, we present computational experiments to test the performance of our approximation algorithms on a large number of randomly generated test problems.

6.1. Computational Setup

We work with both uncapacitated and capacitated problems. To obtain a 0.6-approximate solution for the uncapacitated assortment problem, we execute Steps (a)-(c) discussed at the end of Section 4. To obtain a 0.25-approximate solution for the capacitated assortment problem, we execute Steps (a)-(c) discussed at the end of Section 5.2. We do not test the performance of the approximation algorithm based on an SDP relaxation, because the approximation algorithm based on an LP relaxation already performs quite well. Also, although we can solve an SDP in polynomial time in theory, the size of our test problems prevents us from solving SDP relaxations for the large number of problem instances that we consider. We carry out our computational experiments using Matlab with 3.1 GHz Intel Core i7 CPU and 16 GB RAM. We use Gurobi 6.5.0 as our LP solver.

By Theorem 3.1, if $\hat{z}$ satisfies $f^R(\hat{z}) = v_0\,\hat{z}$, then we have $\hat{z} \geq z^*$, so $\hat{z}$ is an upper bound on the optimal expected revenue z^* for the Assortment problem. Recalling that $\pi(\hat{\boldsymbol{x}})$ is the expected revenue from the subset of products $\hat{\boldsymbol{x}}$, to evaluate the quality of the subset of products $\hat{\boldsymbol{x}}$ obtained by our approximation algorithms, we report the quantity $100 \times \pi(\hat{\boldsymbol{x}})/\hat{z}$, which corresponds to the percentage of the upper bound captured by the subset $\hat{\boldsymbol{x}}$. This quantity provides a conservative estimate of the optimality gaps of the solutions obtained by our approximation algorithms, because $\hat{z}$ is an upper bound on the optimal expected revenue, rather than the optimal expected revenue itself. To compute the upper bound $\hat{z}$, we solve the Fixed Point problem in Theorem 3.3, which is an LP. Letting $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}}, \hat{\delta}, \hat{z})$ be an optimal solution to the Fixed Point problem, by Theorem 3.3, $\hat{z}$ satisfies $f^R(\hat{z}) = v_0\,\hat{z}$, in which case, by Theorem 3.1, $\hat{z}$ is an upper bound on the optimal expected revenue z^*. In our test problems, we have $n = 50$ or $n = 100$ products. Finding the optimal subset of products through enumeration requires checking the expected revenues from $O(2^n)$ assortments, which is not computationally feasible for the sizes of our problem instances. Thus, we provide comparisons with an upper bound on the optimal expected revenue.

6.2. Uncapacitated Problems

We randomly generate a large number of test problems and check the performance of our approximation algorithm on each test problem. To generate the dissimilarity parameters of the nests, we sample γ_{ij} from the uniform distribution over $[0, \bar{\gamma}]$ for all $(i,j) \in M$, where $\bar{\gamma}$ is a parameter that we vary in our computational experiments. To generate the preference weights of the products, we sample v_i from the uniform distribution over $[0,1]$ for all $i \in N$. Using $\mathbf{1} \in \mathbb{R}^n$ to denote the vector of all ones, if we offer all products, then a customer leaves without a purchase with probability $v_0/(v_0 + \sum_{(i,j)\in M} V_{ij}(\mathbf{1})^{\gamma_{ij}})$. To generate the preference weight of the no purchase option, we set $v_0 = \phi_0 \sum_{(i,j)\in M} V_{ij}(\mathbf{1})^{\gamma_{ij}}/(1-\phi_0)$, where ϕ_0 is a parameter that we vary. In this case, if we offer all products, then a customer leaves without a purchase with probability ϕ_0. We work with two classes of test problems when generating the revenues of the products. In the first class, we sample the revenue p_i of each product i from the uniform distribution over $[0,1]$. We refer to these problem instances as independent instances since the preference weights and the revenues are independent. In the second class, we set the revenue p_i of each product i as $p_i = 1 - v_i$. We refer to these problem instances as correlated instances since the preference weights and the revenues are correlated. In the correlated instances, more expensive products have smaller preference weights, making them less desirable. As stated earlier, we use $n = 50$ or $n = 100$ products.

We vary $\bar{\gamma}$ over $\{0.1, 0.5, 1.0\}$ and ϕ_0 over $\{0.25, 0.50, 0.75\}$. Using I and C to, respectively, refer to the independent and correlated instances, we label our test problems as $(T, n, \bar{\gamma}, \phi_0) \in \{I, C\} \times \{50, 100\} \times \{0.1, 0.5, 1.0\} \times \{0.25, 0.50, 0.75\}$, where T is the class of the test problem and n, $\bar{\gamma}$ and ϕ_0 are as discussed above. In this way, we obtain 36 parameter combinations. In each parameter combination, we randomly generate 1000 individual test problems by using the approach discussed in the previous paragraph. We use our approximation algorithm to obtain an approximate solution for each test problem. Also, for each test problem, we solve the Fixed Point problem in Theorem 3.3 to find the value of $\hat{z}$ satisfying $f^R(\hat{z}) = v_0\, \hat{z}$. For test problem s, we use $\hat{\boldsymbol{x}}^s$ to denote the solution obtained by our approximation algorithm and $\hat{z}^s$ to denote the value of $\hat{z}$ satisfying $f^R(\hat{z}) = v_0\, \hat{z}$. In this case, the data $\{100 \times \pi(\hat{\boldsymbol{x}}^s)/\hat{z}^s : s = 1, \ldots, 1000\}$ characterizes the quality of the solutions obtained for the 1000 test problems in a parameter combination.

We give our computational results in Table 1. The first column in this table shows the parameter combination. The next five columns, respectively, show the average, minimum, 5th percentile, 95th percentile and standard deviation of the data $\{100 \times \pi(\hat{\boldsymbol{x}}^s)/\hat{z}^s : s = 1, \ldots, 1000\}$. The last column shows the average CPU seconds to run our approximation algorithm over the 1000 test problems in a parameter configuration. The results in Table 1 indicate that our approximation algorithm performs remarkably well. Over all test problems, on average, our approximation algorithm obtains 99.5% of the upper bound on the optimal expected revenue. In the worst case, our approximation algorithm obtains 95.5% of the upper bound on the optimal expected revenue. In Appendix L, we give confidence intervals for some of the performance measures in Table 1. For the largest test problems with $n = 100$, on average, our approximation algorithm runs in 0.23 seconds. The CPU seconds varied no more than 10% from one test problem to another.

6.3. Capacitated Problems

The approach that we use to generate the capacitated test problems is the same as the one that we use to generate the uncapacitated ones, but we also need to choose the available capacity in the capacitated test problems. We set the capacity c as $c = \lceil \delta\, n \rceil$, where δ is a parameter that we vary. We label our test problems by $(T, n, \bar{\gamma}, \phi_0, \delta) \in \{I, C\} \times \{50, 100\} \times \{0.1, 0.5, 1.0\} \times \{0.25, 0.75\} \times \{0.2, 0.5, 0.8\}$, which yields 72 parameter combinations. We randomly generate 1000 individual test problems in each parameter combination. Using $\hat{\boldsymbol{x}}^s$ and $\hat{z}^s$ with the same interpretation that we have for the uncapacitated test problems, the data $\{100 \times \pi(\hat{\boldsymbol{x}}^s)/\hat{z}^s : s = 1, \ldots, 1000\}$ continues to characterize the quality of the solutions obtained

Param. Conf. $(T, n, \bar{\gamma}, \phi_0)$	$\pi(\hat{\boldsymbol{x}})/\hat{z}$ Avg.	Min	5th	95th	Std.	CPU Secs.
$(I,\ 50, 0.1, 0.25)$	98.6	95.5	97.3	99.7	0.7	0.06
$(I,\ 50, 0.1, 0.50)$	99.7	98.8	99.2	100.0	0.2	0.05
$(I,\ 50, 0.1, 0.75)$	99.9	99.6	99.8	100.0	0.1	0.05
$(I,\ 50, 0.5, 0.25)$	98.8	96.2	97.7	99.7	0.6	0.04
$(I,\ 50, 0.5, 0.50)$	99.7	98.8	99.3	100.0	0.2	0.05
$(I,\ 50, 0.5, 0.75)$	99.9	99.7	99.8	100.0	0.1	0.05
$(I,\ 50, 1.0, 0.25)$	98.8	96.0	97.7	99.7	0.6	0.05
$(I,\ 50, 1.0, 0.50)$	99.8	98.6	99.4	100.0	0.2	0.05
$(I,\ 50, 1.0, 0.75)$	99.9	99.7	99.9	100.0	0.1	0.05
$(I, 100, 0.1, 0.25)$	98.6	96.4	97.7	99.3	0.5	0.23
$(I, 100, 0.1, 0.50)$	99.7	99.0	99.3	99.9	0.2	0.25
$(I, 100, 0.1, 0.75)$	99.9	99.8	99.9	100.0	0.1	0.25
$(I, 100, 0.5, 0.25)$	98.8	97.0	98.0	99.5	0.5	0.22
$(I, 100, 0.5, 0.50)$	99.7	99.2	99.4	100.0	0.2	0.24
$(I, 100, 0.5, 0.75)$	99.9	99.8	99.9	100.0	0.1	0.26
$(I, 100, 1.0, 0.25)$	98.8	97.1	98.1	99.5	0.4	0.22
$(I, 100, 1.0, 0.50)$	99.8	99.1	99.5	100.0	0.1	0.24
$(I, 100, 1.0, 0.75)$	99.9	99.8	99.9	100.0	0.1	0.26
Average	99.5	98.3	99.0	99.8	0.3	0.15

Param. Conf. $(T, n, \bar{\gamma}, \phi_0)$	$\pi(\hat{\boldsymbol{x}})/\hat{z}$ Avg.	Min	5th	95th	Std.	CPU Secs.
$(C,\ 50, 0.1, 0.25)$	98.6	96.2	97.3	99.7	0.7	0.04
$(C,\ 50, 0.1, 0.50)$	99.7	98.6	99.2	100.0	0.3	0.05
$(C,\ 50, 0.1, 0.75)$	99.9	99.7	99.8	100.0	0.1	0.05
$(C,\ 50, 0.5, 0.25)$	98.8	96.4	97.6	99.7	0.6	0.04
$(C,\ 50, 0.5, 0.50)$	99.7	98.9	99.3	100.0	0.2	0.05
$(C,\ 50, 0.5, 0.75)$	99.9	99.6	99.8	100.0	0.1	0.05
$(C,\ 50, 1.0, 0.25)$	98.8	96.4	97.6	99.8	0.6	0.04
$(C,\ 50, 1.0, 0.50)$	99.7	99.0	99.3	100.0	0.2	0.05
$(C,\ 50, 1.0, 0.75)$	99.9	99.7	99.9	100.0	0.0	0.06
$(C, 100, 0.1, 0.25)$	98.6	96.9	97.7	99.4	0.5	0.21
$(C, 100, 0.1, 0.50)$	99.7	98.9	99.4	99.9	0.2	0.23
$(C, 100, 0.1, 0.75)$	99.9	99.8	99.9	100.0	0.0	0.26
$(C, 100, 0.5, 0.25)$	98.8	97.0	97.9	99.5	0.5	0.21
$(C, 100, 0.5, 0.50)$	99.7	99.2	99.5	100.0	0.2	0.24
$(C, 100, 0.5, 0.75)$	99.9	99.8	99.9	100.0	0.1	0.27
$(C, 100, 1.0, 0.25)$	98.8	97.1	98.1	99.5	0.4	0.21
$(C, 100, 1.0, 0.50)$	99.8	99.1	99.5	100.0	0.1	0.25
$(C, 100, 1.0, 0.75)$	99.9	99.8	99.9	100.0	0.1	0.28
Average	99.5	98.4	99.0	99.9	0.3	0.14

Table 1 Computational results for the uncapacitated test problems.

for the 1000 test problems in a parameter combination. We give our computational results in Table 2. The layout of this table is identical to that of Table 1. The optimality gaps reported in Table 2 are slightly larger than those in Table 1, but our approximation algorithm for the capacitated problems also performs remarkably well. On average, our approximation algorithm obtains 98.9% of the upper bound on the optimal expected revenue, corresponding to an average optimality gap of no larger than 1.1%. In Appendix L, we give confidence intervals for some of the performance measures in Table 2. The slightly larger optimality gaps in Table 2 can be attributed to the performance of the approximation algorithm being inferior or the upper bounds being looser, but it is not possible to say which one of these factors plays a dominant role without knowing the optimal expected revenues. Over the largest test problems with $n = 100$, on average, our approximation algorithm runs in 0.25 seconds. To put this running time in perspective, if we partition the 100 products into two nests and assume that the customers choose under the nested logit model, then the average CPU seconds for the approach proposed by Feldman and Topaloglu (2015) is 0.86 seconds. We solve multiple LP relaxations in the iterative rounding algorithm. Over all of our test problems, the iterative rounding algorithm terminated after solving at most five LP relaxations, with only 1.32 LP relaxations on average.

7. Conclusions

In this paper, we developed approximation algorithms for the uncapacitated and capacitated assortment problems under the PCL model. We can extend our work to a slightly more general version of the PCL model. In particular, the generalized nested logit model is a more general version of both the nested logit and PCL models. Under the generalized nested logit model, each product

Param. Conf. $(T, n, \bar{\gamma}, \phi_0, \delta)$	$\pi(\hat{\boldsymbol{x}})/\hat{z}$ Avg.	Min	5th	95th	Std.	CPU Secs.
$(I, 50, 0.1, 0.25, 0.8)$	95.8	90.2	93.2	98.1	1.5	0.05
$(I, 50, 0.1, 0.25, 0.5)$	98.1	94.4	96.1	99.6	1.0	0.05
$(I, 50, 0.1, 0.25, 0.2)$	99.5	96.1	98.4	100.0	0.6	0.05
$(I, 50, 0.1, 0.75, 0.8)$	99.9	99.2	99.6	100.0	0.1	0.06
$(I, 50, 0.1, 0.75, 0.5)$	99.9	99.5	99.8	100.0	0.1	0.06
$(I, 50, 0.1, 0.75, 0.2)$	99.9	99.7	99.9	100.0	0.0	0.06
$(I, 50, 0.5, 0.25, 0.8)$	96.3	90.1	93.8	98.5	1.5	0.05
$(I, 50, 0.5, 0.25, 0.5)$	98.1	94.1	96.3	99.5	1.0	0.05
$(I, 50, 0.5, 0.25, 0.2)$	99.4	96.6	98.2	100.0	0.6	0.06
$(I, 50, 0.5, 0.75, 0.8)$	99.9	99.3	99.6	100.0	0.1	0.06
$(I, 50, 0.5, 0.75, 0.5)$	99.9	99.5	99.7	100.0	0.1	0.06
$(I, 50, 0.5, 0.75, 0.2)$	99.9	99.6	99.9	100.0	0.1	0.06
$(I, 50, 1.0, 0.25, 0.8)$	96.6	92.8	94.4	98.5	1.2	0.05
$(I, 50, 1.0, 0.25, 0.5)$	98.0	93.2	96.2	99.5	1.0	0.05
$(I, 50, 1.0, 0.25, 0.2)$	99.4	96.5	98.2	100.0	0.6	0.06
$(I, 50, 1.0, 0.75, 0.8)$	99.9	99.4	99.7	100.0	0.1	0.06
$(I, 50, 1.0, 0.75, 0.5)$	99.9	99.1	99.7	100.0	0.1	0.06
$(I, 50, 1.0, 0.75, 0.2)$	99.9	98.9	99.9	100.0	0.1	0.19
$(I, 100, 0.1, 0.25, 0.8)$	95.8	90.9	94.0	97.5	1.1	0.23
$(I, 100, 0.1, 0.25, 0.5)$	98.1	95.8	96.9	99.1	0.7	0.25
$(I, 100, 0.1, 0.25, 0.2)$	99.5	97.7	98.8	100.0	0.4	0.26
$(I, 100, 0.1, 0.75, 0.8)$	99.9	99.6	99.7	100.0	0.1	0.32
$(I, 100, 0.1, 0.75, 0.5)$	99.9	99.5	99.8	100.0	0.1	0.30
$(I, 100, 0.1, 0.75, 0.2)$	99.9	99.9	99.9	100.0	0.1	0.28
$(I, 100, 0.5, 0.25, 0.8)$	96.3	92.6	94.6	97.8	1.0	0.24
$(I, 100, 0.5, 0.25, 0.5)$	98.0	94.6	96.6	99.1	0.7	0.27
$(I, 100, 0.5, 0.25, 0.2)$	99.4	97.5	98.7	100.0	0.4	0.26
$(I, 100, 0.5, 0.75, 0.8)$	99.9	99.4	99.7	100.0	0.1	0.33
$(I, 100, 0.5, 0.75, 0.5)$	99.9	99.7	99.8	100.0	0.1	0.33
$(I, 100, 0.5, 0.75, 0.2)$	99.9	99.7	99.9	100.0	0.0	0.30
$(I, 100, 1.0, 0.25, 0.8)$	96.5	92.8	94.8	98.0	1.0	0.24
$(I, 100, 1.0, 0.25, 0.5)$	98.0	95.3	96.7	99.1	0.7	0.28
$(I, 100, 1.0, 0.25, 0.2)$	99.4	97.1	98.6	100.0	0.4	0.27
$(I, 100, 1.0, 0.75, 0.8)$	99.9	99.6	99.7	100.0	0.1	0.20
$(I, 100, 1.0, 0.75, 0.5)$	99.9	99.5	99.8	100.0	0.1	0.20
$(I, 100, 1.0, 0.75, 0.2)$	99.9	99.7	99.9	100.0	0.1	0.20
Average	98.9	96.9	98.1	99.6	0.5	0.16

Param. Conf. $(T, n, \bar{\gamma}, \phi_0, \delta)$	$\pi(\hat{\boldsymbol{x}})/\hat{z}$ Avg.	Min	5th	95th	Std.	CPU Secs.
$(C, 50, 0.1, 0.25, 0.8)$	96.0	89.3	93.4	98.2	1.5	0.06
$(C, 50, 0.1, 0.25, 0.5)$	98.1	93.9	96.2	99.6	1.0	0.06
$(C, 50, 0.1, 0.25, 0.2)$	99.5	97.5	98.5	100.0	0.5	0.06
$(C, 50, 0.1, 0.75, 0.8)$	99.9	99.3	99.6	100.0	0.1	0.07
$(C, 50, 0.1, 0.75, 0.5)$	99.9	99.3	99.8	100.0	0.1	0.07
$(C, 50, 0.1, 0.75, 0.2)$	99.9	99.6	99.9	100.0	0.0	0.07
$(C, 50, 0.5, 0.25, 0.8)$	96.3	91.9	94.0	98.4	1.3	0.06
$(C, 50, 0.5, 0.25, 0.5)$	98.0	94.3	96.0	99.5	1.1	0.06
$(C, 50, 0.5, 0.25, 0.2)$	99.4	96.9	98.2	100.0	0.6	0.06
$(C, 50, 0.5, 0.75, 0.8)$	99.9	99.3	99.6	100.0	0.1	0.07
$(C, 50, 0.5, 0.75, 0.5)$	99.9	99.3	99.7	100.0	0.1	0.07
$(C, 50, 0.5, 0.75, 0.2)$	99.9	99.2	99.9	100.0	0.1	0.07
$(C, 50, 1.0, 0.25, 0.8)$	96.5	91.6	94.2	98.6	1.3	0.07
$(C, 50, 1.0, 0.25, 0.5)$	98.0	94.3	96.3	99.5	1.0	0.07
$(C, 50, 1.0, 0.25, 0.2)$	99.4	97.2	98.2	100.0	0.6	0.06
$(C, 50, 1.0, 0.75, 0.8)$	99.9	99.4	99.6	100.0	0.1	0.07
$(C, 50, 1.0, 0.75, 0.5)$	99.9	99.2	99.7	100.0	0.1	0.08
$(C, 50, 1.0, 0.75, 0.2)$	99.9	98.7	99.9	100.0	0.1	0.07
$(C, 100, 0.1, 0.25, 0.8)$	95.8	92.5	94.1	97.5	1.1	0.21
$(C, 100, 0.1, 0.25, 0.5)$	98.1	95.6	96.8	99.2	0.7	0.21
$(C, 100, 0.1, 0.25, 0.2)$	99.5	98.1	98.8	100.0	0.4	0.21
$(C, 100, 0.1, 0.75, 0.8)$	99.9	99.3	99.7	100.0	0.1	0.25
$(C, 100, 0.1, 0.75, 0.5)$	99.9	99.6	99.8	100.0	0.1	0.25
$(C, 100, 0.1, 0.75, 0.2)$	99.9	99.9	99.9	100.0	0.0	0.25
$(C, 100, 0.5, 0.25, 0.8)$	96.3	93.0	94.6	97.7	1.0	0.22
$(C, 100, 0.5, 0.25, 0.5)$	98.0	94.9	96.8	99.1	0.7	0.22
$(C, 100, 0.5, 0.25, 0.2)$	99.4	97.7	98.7	100.0	0.4	0.23
$(C, 100, 0.5, 0.75, 0.8)$	99.9	99.4	99.7	100.0	0.1	0.26
$(C, 100, 0.5, 0.75, 0.5)$	99.9	99.6	99.8	100.0	0.1	0.26
$(C, 100, 0.5, 0.75, 0.2)$	99.9	99.7	99.9	100.0	0.0	0.26
$(C, 100, 1.0, 0.25, 0.8)$	96.5	92.9	95.0	98.0	0.9	0.23
$(C, 100, 1.0, 0.25, 0.5)$	98.1	95.9	96.8	99.1	0.7	0.24
$(C, 100, 1.0, 0.25, 0.2)$	99.4	97.8	98.7	100.0	0.4	0.24
$(C, 100, 1.0, 0.75, 0.8)$	99.9	99.5	99.7	100.0	0.1	0.27
$(C, 100, 1.0, 0.75, 0.5)$	99.9	99.0	99.8	100.0	0.1	0.27
$(C, 100, 1.0, 0.75, 0.2)$	99.9	99.7	99.9	100.0	0.0	0.28
Average	98.9	97.1	98.1	99.6	0.5	0.15

Table 2 Computational results for the capacitated test problems.

can be in multiple nests and each nest can include an arbitrary number of products. For each product and nest combination, there is a membership parameter that characterizes the extent to which the product is a member of the nest. Considering the generalized nested logit model with at most two products in each nest, we can generalize our results to tackle the uncapacitated and capacitated assortment problems under this choice model. We discuss this extension in Appendix M. There are several future research directions to pursue. First, our approximation algorithms exploit the fact that we can formulate the Function Evaluation problem as an integer program by linearizing the quadratic terms in the objective function. Our performance guarantees are based on the fact that we can choose the values of the decision variables within $\{0, \frac{1}{2}, 1\}$ to construct a provably good feasible solution to the LP relaxation of the integer program. This observation does not hold when each nest includes more than two products. Solving assortment problems under variants of the generalized nested logit model that have more than two products in each nest is a worthwhile and highly non-trivial extension. Second, we can consider a variant of the multinomial logit model with synergies between the pairs of products, where the deterministic component of the utility of product i increases by Δ_i when product i is offered along with some other product i_p. Using $\boldsymbol{x}$ to denote the subset of offered products, if we formulate an assortment problem under such a choice model, then the objective function can be written as a ratio of two quadratic functions of $\boldsymbol{x}$. An interesting question is whether we can use an approach similar to ours to develop approximation algorithms under this variant of the multinomial logit model. A straightforward extension of our approach does not work when the increase Δ_i in the deterministic component of the utility of product i has no relationship with the base deterministic component of the utility of product i when this product is not offered along with product i_p. More research is needed in this direction. Third, although there is work on pricing problems under the PCL model, this work assumes that the price sensitivities of the products satisfy certain conditions. We can formulate a pricing problem as a variant of our assortment problem. In particular, we can create multiple copies of a product, corresponding to offering a product at different price levels. In this case, we need to impose the constraint that we offer at most one copy of a particular product, meaning that each product should have one price level, if offered. Our efforts to extend our approximation algorithms to this type of a pricing problem showed that the pricing problem is considerably more difficult and more work is also needed in this direction. Fourth, the PCL model is flexible as it allows a rather general correlation structure among the utilities of the products. Empirical studies in the path choice domain demonstrate that the flexibility of the PCL model may be beneficial. It would be useful to conduct additional empirical studies in the operations management domain to understand the benefits of the PCL model in predicting the customer purchase behavior.

Acknowledgements: The authors thank the area editor Anton Kleywegt, the associate editor and three anonymous referees whose comments substantially improved the paper. The second and third authors were supported in part by National Science Foundation grants CMMI-1824860 and CMMI-1825406. The third author was also supported in part by Schmidt Sciences. We provide the computer code and data used in our computational experiments as supplemental material.

Author Biographies: Heng Zhang finished his PhD at the USC Marshall School of Business. In fall 2019, he will start as an assistant professor in W.P. Carey School of Business in Arizona State University. His research interests are revenue management, choice modeling, and optimization.

Paat Rusmevichientong is a professor of data sciences and operations at the USC Marshall School of Business. His research interests include revenue management, choice modeling, pricing, assortment optimization, and approximate dynamic programming.

Huseyin Topaloglu is a professor in the School of Operations Research and Information Engineering at Cornell University. His research focuses on revenue management, assortment optimization, transportation logistics, and approximate dynamic programming.

References

Alon, N., J. H. Spencer. 2000. *The Probabilistic Methods*. John Wiley & Sons, Inc., New York, NY.

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

Cardell, N. S. 1997. Variance components structure for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory* **13**(2) 185–213.

Chen, A., P. Kasikitwiwat, Z. Ji. 2003. Solving the overlapping problem in route choice with paired combinatorial logit model. *Transportation Research Record* **1857**(1) 65–73.

Chen, A., S. Ryu, X. Xu, K. Choi. 2014. Computation and application of the paired combinatorial logit stochastic user equilibrium problem. *Computers & Operations Research* **43**(1) 68–77.

Daganzo, C. F., M. Kusnic. 1993. Two properties of the nested logit model. *Transportation Science* **27**(4) 395–400.

Dai, J., W. Ding, A. J. Kleywegt, X. Wang, Y. Zhang. 2014. Choice based revenue management for parallel flights. Tech. rep., Georgia Tech, Atlanta, GA.

Davis, J. M., G. Gallego, H. Topaloglu. 2013. Assortment planning under the multinomial logit model with totally unimodular constraint structures. Tech. rep., Cornell University, School of Operations Research and Information Engineering.

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

Desire, Antoine, Vineet Goyal. 2016. Near-optimal algorithms for capacity constrained assortment optimization. Working Paper, Columbia University.

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

Gallego, G., G. Iyengar, R. Phillips, A. Dubey. 2004. Managing flexible products on a network. Working Paper, Columbia University.

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

Garey, M. R., D. S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, New York, NY.

Ghuge, R., J. Kwon, V. Nagarajan, A. Sharma. 2019. Constrained assortment optimization under the paired combinatorial logit model. Tech. rep., University of Michigan, Ann Arbor, MI.

Goemans, M. X., D. P. Willamson. 1995. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM* **42**(6) 1115–1145.

Karoonsoontawong, A., D.-Y. Lin. 2015. Combined gravity model trip distribution and paired combinatorial logit stochastic user equilibrium problem. *Networks and Spatial Economics* **15**(4) 1011–1048.

Koppleman, F. S., C.-H. Wen. 2000. The paired combinatorial logit model: Properties, estimation and application. *Transportation Research Part B: Methodological* **34**(2) 75–89.

Lau, L. C., R. Ravi, M. Singh. 2011. *Iterative Methods in Combinatorial Optimization*. Cambridge University Press, New York, NY.

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

Li, H., S. Webster. 2017. Optimal pricing of correlated product options under the pairwise combinatorial logit model. *Operations Research* **65**(5) 125–130.

McFadden, Daniel, Kenneth Train. 2000. Mixed MNL models for discrete response. *Journal of Applied Economics* **15** 447–470.

Prashker, J., S. Bekhor. 1998. Investigation of stochastic network loading procedures. *Transportation Research Record* **1645**(1) 94–102.

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

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

Train, K. 2002. *Discrete Choice Methods with Simulation*. Cambridge University Press.

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

Wen, C.-H., F. S. Koppelman. 2001. The generalized nested logit model. *Transportation Research Part B* **35**(7) 627–641.

Williamson, D. P., D. B. Shmoys. 2011. *The Design of Approximation Algorithms*. Cambridge University Press.

Electronic Companion: Assortment Optimization under the Paired Combinatorial Logit Model

Heng Zhang, Paat Rusmevichientong
Marshall School of Business, University of Southern California, Los Angeles, CA 90089
hengz@usc.edu, rusmevic@marshall.usc.edu

Huseyin Topaloglu
School of Operations Research and Information Engineering, Cornell Tech, New York, NY 10044
topaloglu@orie.cornell.edu

August 14, 2019

Appendix A: Computational Complexity

In this section, we give a proof for Theorem 2.1 by using a polynomial-time reduction from the max-cut problem. In the max-cut problem, we have an undirected graph $G=(V,E)$, where V is the set of vertices and E is the set of edges. We denote the edge between vertex i and j as (i,j). The goal is to find a subset of vertices S such that the number of edges in the set $\{(i,j)\in E : i\in S,\ j\in V\setminus S\}$ is maximized. In other words, the goal is to find a subset of vertices S such that the number of edges that connect a vertex in S and a vertex in $V\setminus S$ is maximized. We refer to a subset of vertices S as a cut and $|\{(i,j)\in E : i\in S,\ j\in V\setminus S\}|$ as the objective value provided by the cut S. The max-cut problem is strongly NP-hard; see Section A.2.2 in Garey and Johnson (1979). Here, we focus on graphs where the degrees of all vertices are even. We show that the max-cut problem over graphs with even vertex degrees continues to be strongly NP-hard. For any $\delta>0$, de la Vega and Karpinski (2006) show that the max-cut problem over the graphs $G=(E,V)$ with $|E|=\Omega(|V|^{2-\delta})$ is hard to approximate within a constant factor. Therefore, defining the class of graphs $\mathcal{G}_1(\beta)=\{G=(V,E) : |E|\geq\beta\,|V|^{1.5}\}$, there exist constants $\alpha\in(0,1)$ and $\beta>0$ such that it is hard to approximate the max-cut problem over the graphs in $\mathcal{G}_1(\beta)$ within a factor of $1-\alpha$. Defining the class of graphs $\mathcal{G}_2(\beta)=\{G=(V,E) : |E|\geq\beta\,|V|^{1.5}$ and all vertices in G have even degrees$\}$, in the next lemma, we show that the same result holds over the graphs in $\mathcal{G}_2(\beta)$.

Lemma A.1 *There exist constants $\alpha\in(0,1)$ and $\beta>0$ such that it is hard to approximate the max-cut problem over the graphs in $\mathcal{G}_2(\beta)$ within a factor of $1-\alpha$.*

Proof: Let $\alpha_1\in(0,1)$ and $\beta_1>0$ be such that it is hard to approximate the max-cut problem over the graphs in $\mathcal{G}_1(\beta_1)$ within a factor of $1-\alpha_1$. Note that the existence of α_1 and β_1 is guaranteed by the discussion right before the lemma. Fix any $\epsilon<2\,\alpha_1/3$. Set $\alpha=\alpha_1-3\,\epsilon/2$ and $\beta=\beta_1/2^{1.5}$. Assume that there exists a $(1-\alpha)$-approximation algorithm for the max-cut problem over the graphs in $\mathcal{G}_2(\beta)$. We will show that the existence of this approximation algorithm directly implies the existence of a $(1-\alpha_1)$-approximation algorithm for the max-cut problem over the graphs in $\mathcal{G}_1(\beta_1)$, which is a contradiction.

Choose any graph $G_1=(V_1,E_1)\in\mathcal{G}_1(\beta_1)$ and let $n=|V_1|$. If we have $n\leq(2/(\beta_1\,\epsilon))^2$, then we can enumerate all of the cuts in G_1 to solve the max-cut problem over G_1 in polynomial time, since the number of vertices is bounded by a constant. Also, if all of the vertices in G_1 have even degrees, noting that $G_1\in\mathcal{G}_1(\beta_1)$ and $\beta_1\geq\beta$, we have $|E_1|\geq\beta_1\,|V_1|^{1.5}\geq\beta\,|V_1|^{1.5}$, which implies that $G_1\in\mathcal{G}_2(\beta)$. Therefore, we can use the $(1-\alpha)$-approximation algorithm for the max-cut problem over the graphs in $\mathcal{G}_2(\beta)$ to find a $(1-\alpha)$-approximate cut in G_1. Since $\alpha_1\geq\alpha$, a $(1-\alpha)$-approximate cut is also a $(1-\alpha_1)$-approximate cut. Therefore, we can obtain a $(1-\alpha_1)$-approximate cut in G_1 in polynomial time, which is a contradiction. In the rest of the proof, we assume that $n > (2/(\beta_1\,\epsilon))^2$ and some of the vertices in G_1 have odd degrees.

Let k be the number of vertices in G_1 with odd degrees. Note that k must be an even number, otherwise the sum of the degrees of all vertices in G_1 would be an odd number, but we know that this sum is equal to twice the number of edges. We add k auxiliary vertices to G_1. Using additional k edges, we connect each one of these auxiliary vertices to one of the k vertices with odd degrees. Since k is an even number, we also use $k/2$ additional edges to form a perfect matching between the auxiliary vertices. We denote this new graph by $\bar{G} = (\bar{V}, \bar{E})$. By our construction, all of the vertices in $\bar{G}$ have even degrees. Furthermore, we have $|\bar{E}| \geq |E_1| \geq \beta_1\,|V_1|^{1.5} = \frac{\beta_1}{2^{1.5}}\,(2\,|V_1|)^{1.5} \geq \frac{\beta_1}{2^{1.5}}\,(|V_1| + k)^{1.5} = \frac{\beta_1}{2^{1.5}}\,|\bar{V}|^{1.5} = \beta\,|\bar{V}|^{1.5}$, which implies that $\bar{G} \in \mathcal{G}_2(\beta)$.

We use OPT_1 and $\overline{\text{OPT}}$ to, respectively, denote the optimal objective values of the max-cut problems in G_1 and $\bar{G}$. We have $\overline{\text{OPT}} \geq \text{OPT}_1$, because $\bar{V} \supseteq V_1$ and $\bar{E} \supseteq E_1$. Since $\bar{G} \in \mathcal{G}_2(\beta)$, we use the $(1-\alpha)$-approximation algorithm for the max-cut problem over the graphs in $\mathcal{G}_2(\beta)$ to find a $(1-\alpha)$-approximate cut in $\bar{G}$. That is, letting $\overline{\text{CUT}}$ be the objective value provided by the cut, we have $\overline{\text{CUT}} \geq (1-\alpha)\overline{\text{OPT}}$. By removing the auxiliary vertices in the $(1-\alpha)$-approximate cut in $\bar{G}$, we obtain the corresponding cut in G_1. We use CUT_1 to denote the objective value provided by the cut in G_1. Since the graphs G_1 and $\bar{G}$ differ in $3k/2$ edges, the objective values of the two cuts cannot differ by more than $3k/2$, so $\text{CUT}_1 \geq \overline{\text{CUT}} - 3k/2$. Therefore, we obtain

$$\text{CUT}_1 \;\geq\; \overline{\text{CUT}} - \frac{3}{2}\,k \;\geq\; (1-\alpha)\,\overline{\text{OPT}} - \frac{3}{2}\,k \;\geq\; (1-\alpha)\,\text{OPT}_1 - \frac{3}{2}\,k \;\geq\; (1-\alpha)\,\text{OPT}_1 - \frac{3}{2}\,|V_1|.$$

It is well-known that the optimal objective value of the max-cut problem over any graph is at least half of the number of edges; see Section 12.4 in Kleinberg and Tardos (2005). Thus, noting that $n > (2/(\beta_1\,\epsilon))^2$, we have $|V_1| = n \leq \frac{1}{2}\,\epsilon\,\beta_1\,n^{1.5} = \frac{1}{2}\,\epsilon\,\beta_1\,|V_1|^{1.5} \leq \frac{1}{2}\,\epsilon\,|E_1| \leq \epsilon\,\text{OPT}_1$, where the second inequality uses the fact that $G_1 \in \mathcal{G}_1(\beta_1)$. In this case, we get $(1-\alpha)\,\text{OPT}_1 - \frac{3}{2}\,|V_1| \geq (1 - \alpha - \frac{3}{2}\epsilon)\,\text{OPT}_1 = (1-\alpha_1)\,\text{OPT}_1$, so using the chain of inequalities above, it follows that $\text{CUT}_1 \geq (1-\alpha_1)\,\text{OPT}_1$. Therefore, if we use the $(1-\alpha)$-approximate approximation algorithm to find a $(1-\alpha)$-approximate cut in $\bar{G}$ and drop the auxiliary vertices, then we obtain a $(1-\alpha_1)$-approximate cut in G_1 in polynomial time, which is a contradiction. □

By the lemma above, the max-cut problem is strongly NP-hard when the degrees of all vertices are even. In the proof Theorem 2.1, we will need the fact that the max-cut problem is strongly NP-hard when the degrees of all vertices are divisible by four. We define the class of graphs $\mathcal{G}_4(\beta) = \{G = (V,E) : |E| \geq \beta\,|V|^{1.5}$ and all vertices in G have degrees divisible by four$\}$. In the next lemma, we repeat an argument similar to the one in the proof of Lemma A.1 to show that an analogue of the result in Lemma A.1 holds for the graphs in $\mathcal{G}_4(\beta)$.

Lemma A.2 *There exist constants $\alpha \in (0,1)$ and $\beta > 0$ such that it is hard to approximate the max-cut problem over the graphs in $\mathcal{G}_4(\beta)$ within a factor of $1-\alpha$.*

Proof: Let $\alpha_2 \in (0,1)$ and $\beta_2 > 0$ be such that it is hard to approximate the max-cut problem over the graphs in $\mathcal{G}_2(\beta_2)$ within a factor of $1-\alpha_2$. The existence of α_2 and β_2 is guaranteed by Lemma A.1. Fix any $\epsilon < \alpha_2/11$. Set $\alpha = \alpha_2 - 11\,\epsilon$ and $\beta = \beta_2/6^{1.5}$. To get a contradiction, we assume that there exists a $(1-\alpha)$-approximation algorithm for the max-cut problem over the graphs in $\mathcal{G}_4(\beta)$. We will show that the existence of this approximation algorithm implies the existence of a $(1-\alpha_2)$-approximation algorithm for the max-cut problem over the graphs in $\mathcal{G}_2(\beta_2)$. Choose any graph $G_2 = (V_2, E_2) \in \mathcal{G}_2(\beta_2)$ and let $n = |V_2|$. If $n \le (2/(\beta_2\,\epsilon))^2$ or all of the vertices in G_2 have degrees divisible by four, then we reach a contradiction by the same argument in the second paragraph of the proof of Lemma A.1. Therefore, we assume that $n > (2/(\beta_2\,\epsilon))^2$ and the degrees of some of the vertices in G_2 are not divisible by four.

Let k be the number of vertices in G_2 with degrees not divisible by four. Since $G_2 \in \mathcal{G}_2(\beta_2)$, these vertices must have even degrees. If $k \ge 3$, then we can add k vertices and $3k$ edges to G_2 to obtain a graph $\bar{G} = (\bar{V}, \bar{E})$ with all the vertices having degrees divisible by four. In particular, let $\{i_1, \ldots, i_k\}$ be the vertices with even degree but not divisible by four. We add k auxiliary vertices $\{j_1, \ldots, j_k\}$ to the graph G_2. Using additional edges, for $s = 1, \ldots, k-1$, we connect i_s to j_s and j_{s+1}. Also, we connect i_k to j_k and j_1. Finally, we add the edges $(j_1, j_2), (j_2, j_3), \ldots, (j_{k-1}, j_k), (j_k, j_1)$. We denote this new graph by $\bar{G} = (\bar{V}, \bar{E})$. In Figure 2.a, we show the k vertices $\{i_1, \ldots, i_k\}$ in G_2 with even degrees but not divisible by four, along with the k auxiliary vertices $\{j_1, \ldots, j_k\}$. The solid edges are the ones that we add to G_2 to get $\bar{G}$. The dotted edges are already in G_2. By our construction, all of the vertices in $\bar{G}$ have degrees divisible by four.

If $k = 1$, then we can add 5 vertices and 11 edges to G_2 get a graph $\bar{G} = (\bar{V}, \bar{E})$ with all vertices having degrees divisible by four. In Figure 2.b, we show the only vertex in G_2 with even degree but not divisible by four, along with the 5 vertices and 11 edges that we add to get the graph $\bar{G}$. If $k = 2$, then we can add 5 vertices and 12 edges to G_2 to get a graph $\bar{G} = (\bar{V}, \bar{E})$ with all vertices having degrees divisible by four. In Figure 2.b, we show the two vertices in G_2 with even degrees but not divisible by four, along with the 5 vertices and 12 edges that we add to get the graph $\bar{G}$. Collecting the three cases discussed in the previous and this paragraph together, if k is the number of vertices in G_2 with even degrees but not divisible by four, then we can add at most $5k$ vertices and $11k$ edges to G_2 to obtain $\bar{G}$. Note that $|\bar{E}| \ge |E_2| \ge \beta_2\,|V_2|^{1.5} = \frac{\beta_2}{6^{1.5}}\,(6\,|V_2|)^{1.5} \ge \frac{\beta_2}{6^{1.5}}\,(|V_2| + 5k)^{1.5} \ge \frac{\beta_2}{6^{1.5}}\,|\bar{V}|^{1.5} = \beta\,|\bar{V}|^{1.5}$, which implies that $\bar{G} \in \mathcal{G}_4(\beta)$.

In this case, we can use precisely the same argument in the last two paragraphs of the proof of Lemma A.1 to show that if we have a $(1-\alpha)$-approximate cut in $\bar{G}$, then we can obtain a $(1-\alpha_2)$

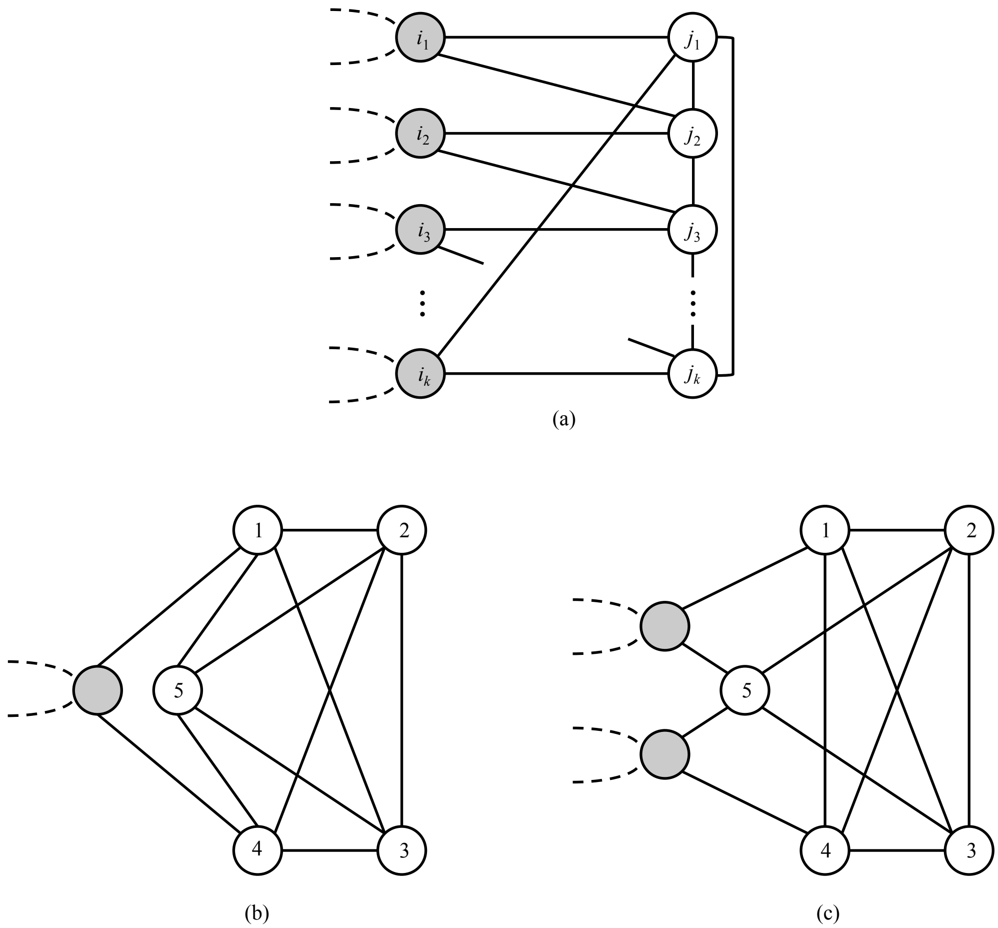

Figure 2 Constructing a graph with all vertices having degrees divisible by four.

approximate cut in G_2 in polynomial time. Since our choice of the graph $G_2 \in \mathcal{G}(\beta_2)$ is arbitrary, we obtain a contradiction Lemma A.1. □

Therefore, by Lemma A.2, the max-cut problem is strongly NP-hard when the degrees of the vertices are divisible by four. We use the decision variable $y_i \in \{-1, +1\}$ to capture whether vertex i is included in the cut, where $y_i = +1$ if and only if the vertex is included. If vertex i is included in the cut and vertex j is not, then we have $y_i\, y_j = -1$. Thus, we can formulate the max-cut problem over the graph $G = (V, E)$ as

$$\max_{\boldsymbol{y} \in \{-1,+1\}^{|V|}} \left\{ \frac{1}{2} \sum_{(i,j)\in E} (1 - y_i\, y_j) \right\},$$

where we use the decision variables $\boldsymbol{y} = \{y_i : i \in V\}$. Using the change of variables $y_i = 2\, x_i - 1$ with $x_i \in \{0, 1\}$ and letting d_i be the degree of vertex i, the objective function of the problem above is $\frac{1}{2}\sum_{(i,j)\in E}(1-(2x_i-1)\,(2\,x_j-1)) = \frac{1}{2}\sum_{(i,j)\in E}(2x_i+2x_j-4x_i\,x_j) = \sum_{i\in V} d_i\,x_i - 2\sum_{(i,j)\in E} x_i\,x_j$, which implies that the problem $\max_{\boldsymbol{x}\in\{0,1\}^{|V|}}\{\sum_{i\in V} d_i\,x_i - 2\sum_{(i,j)\in E} x_i\,x_j\}$ is strongly NP-hard when d_i is divisible by four for all $i\in V$.

To show Theorem 2.1, we use a feasibility version of the max-cut problem. Given an undirected graph $G=(V,E)$, we assume that d_i is divisible by four for all $i\in V$. For a fixed target objective value K, we consider the problem of whether there exists $\boldsymbol{x}=\{x_i : i\in V\}\in\{0,1\}^{|V|}$ such that $\sum_{i\in V} d_i\,x_i - 2\sum_{(i,j)\in E} x_i\,x_j \geq K$. We refer to this problem as the max-cut feasibility problem. By the discussion in the previous paragraph, the max-cut feasibility problem is strongly NP-complete. Below is the proof of Theorem 2.1.

Proof of Theorem 2.1: Throughout the proof, we use the formulation of the PCL model, where the set of nests is given by $M=\{(i,j)\in N^2 : i<j\}$. We observe that the formulation that we use in the paper is equivalent to the formulation with the set of nests $M=\{(i,j)\in N^2 : i<j\}$. In particular, if we are given an assortment problem with the set of nests $M=\{(i,j)\in N^2 : i<j\}$, the dissimilarity parameters $\{\gamma_{ij} : (i,j)\in M\}$ and the no purchase preference weight v_0, then we can define an assortment problem with the set of nests $M'=\{(i,j)\in N^2 : i\neq j\}$, the dissimilarity parameters $\{\gamma'_{ij} : (i,j)\in M'\}$ and the no purchase preference weight v'_0, where $\gamma'_{ij}=\gamma'_{ji}=\gamma_{ij}$ and $v'_0 = 2\,v_0$. In this case, the expected revenues obtained by any subset of products are identical in the two problems.

Assume that we have an instance of the max-cut feasibility problem over the graph $G=(V,E)$ with target objective value K. Letting d_i be the degree of vertex i, we assume that d_i is divisible by four for all $i\in V$. We construct an instance of our assortment problem in such a way that there exists $\boldsymbol{x}=\{x_i : i\in V\}\in\{0,1\}^{|V|}$ that satisfies $\sum_{i\in V} d_i\,x_i - 2\sum_{(i,j)\in E} x_i\,x_j\geq K$ if and only if there exists a subset of products in our assortment problem that provides an expected revenue of $K+8\,(|V|-1)^2$ or more. Thus, an instance of the max-cut feasibility problem can be reduced to an instance of the feasibility version of our assortment problem, in which case, the desired result follows. We construct the instance of our assortment problem as follows.

Let $n=|V|$. In the instance of our assortment problem, there are $2n-1$ products. We partition the products into two subsets V and W so that the set of products is $N=V\cup W$. Since $|N|=2n-1$ and $|V|=n$, we have $|W|=n-1$. We index the products in V by $\{1,\ldots,n\}$ and the products in W by $\{n+1,\ldots,2n-1\}$. The set of nests is $M=\{(i,j)\in N^2 : i<j\}$. Letting $T=K+8\,(n-1)^2$, the revenues of the products are given by $p_i=1+T$ for all $i\in V$ and $p_i=4+T$ for all $i\in W$. The preference weights of the products are given by $v_i=2$ for all $i\in V$ and $v_i=1$ for all $i\in W$. The preference weight of the no purchase option is $v_0=1$. Since E is the set of edges in an

undirected graph, we follow the convention that $i<j$ for all $(i,j)\in E$. The dissimilarity parameters of the nests are

$$\gamma_{ij}=\begin{cases}0 & \text{if } i\in V,\ j\in V,\ (i,j)\in E\\ 1 & \text{if } i\in V,\ j\in V,\ (i,j)\notin E\\ 1 & \text{if } i\in V,\ j\in W,\ j\in\{n+1,\ldots,n+d_i/4\}\\ 0 & \text{if } i\in V,\ j\in W,\ j\in\{n+1+d_i/4\ldots,2n-1\}\\ 1 & \text{if } i\in W,\ j\in W.\end{cases}$$

In the feasibility version of the assortment problem, we are interested in whether there exists a vector $\boldsymbol{x}\in\{0,1\}^{2n-1}$ such that $\sum_{(i,j)\in M}V_{ij}(\boldsymbol{x})^{\gamma_{ij}}R_{ij}(\boldsymbol{x})/(v_0+\sum_{(i,j)\in M}V_{ij}(\boldsymbol{x})^{\gamma_{ij}})\geq T$. In other words, arranging the terms in this inequality, we are interested in whether there exists a vector $\boldsymbol{x}\in\{0,1\}^{2n-1}$ such that $\sum_{(i,j)\in M}V_{ij}(\boldsymbol{x})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x})-T)\geq v_0T$. By the definition of $R_{ij}(\boldsymbol{x})$ and $V_{ij}(\boldsymbol{x})$, we observe that we have

$$R_{ij}(\boldsymbol{x})-T=\frac{p_i\,v_i^{1/\gamma_{ij}}\,x_i+p_j\,v_j^{1/\gamma_{ij}}\,x_j}{V_{ij}(\boldsymbol{x})}-T=\frac{(p_i-T)\,v_i^{1/\gamma_{ij}}\,x_i+(p_j-T)\,v_j^{1/\gamma_{ij}}\,x_j}{V_{ij}(\boldsymbol{x})}.$$

Note that $p_i-T=1$ or $p_i-T=4$ for all $i\in N$. For notational brevity, let $\hat{R}_{ij}(\boldsymbol{x})=R_{ij}(\boldsymbol{x})-T$. In this case, we are interested in whether there exists a vector $\boldsymbol{x}\in\{0,1\}^{2n-1}$ such that $\sum_{(i,j)\in M}V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\hat{R}_{ij}(\boldsymbol{x})\geq v_0T$. It is simple to check that we can offer all of the products with the largest revenue without degrading the expected revenue from a subset of products. Noting that the products in the set W have the largest revenue, we can set $x_i=1$ for all $i\in W$ in the feasibility version of the assortment problem. Therefore, the only question is which of the products in V to include in the subset of products $\boldsymbol{x}$ such that $\sum_{(i,j)\in M}V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\hat{R}_{ij}(\boldsymbol{x})\geq v_0T$. We proceed to computing $V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\hat{R}_{ij}(\boldsymbol{x})$ for all $(i,j)\in M$. Since we offer the products in the set W, if $i\in W$ and $j\in W$, then noting that $\gamma_{ij}=1$, $p_i=p_j=4+T$ and $v_i=v_j=1$, we have $V_{ij}(\boldsymbol{x})^{\gamma_{ij}}=2$ and $\hat{R}_{ij}=4$. Therefore, we have $\sum_{(i,j)\in M}\mathbf{1}(i\in W,\ j\in W)\,V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\hat{R}_{ij}(\boldsymbol{x})=8\,(n-1)\,(n-2)/2$, where we use the fact that $|W|=n-1$ and we have $i<j$ for all $(i,j)\in M$. Similarly, considering the dissimilarity parameters of the other nests, along with the revenues and the preference weights of the products in these nests, if $i\in V$ and $j\in W$, then we have

$$V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\hat{R}_{ij}(\boldsymbol{x})=\begin{cases}6 & \text{if } x_i=1,\ j\in\{n+1,\ldots,n+d_i/4\}\\ 2 & \text{if } x_i=1,\ j\in\{n+1+d_i/4,\ldots,2n-1\}\\ 4 & \text{if } x_i=0,\end{cases}$$

where, once again, we use the fact that we offer all of the products in the set W. We can write the expression above succinctly as $4+2\,x_i$ for all $j\in\{n+1,\ldots,n+d_i/4\}$ and $4-2\,x_i$ for all $j\in\{n+1+d_i/4,\ldots,2n-1\}$. Therefore, we have $\sum_{(i,j)\in M}\mathbf{1}(i\in V,\ j\in W)\,V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\hat{R}_{ij}(\boldsymbol{x})=$

$\sum_{i\in V}\sum_{j\in W}\big[\mathbf{1}(j\in\{n+1,\ldots,n+d_i/4\})\,(4+2\,x_i)+\mathbf{1}(j\in\{n+1+d_i/4,\ldots,2n-1\})\,(4-2\,x_i)\big]=\sum_{i\in V}\big[\frac{d_i}{4}\,(4+2\,x_i)+(n-1-\frac{d_i}{4})\,(4-2\,x_i)\big]$. Next, if $i\in V$ and $j\in V$, then we have

$$V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\,\hat{R}_{ij}(\boldsymbol{x})=\begin{cases}2 & \text{if } x_i=1,\ x_j=1,\ (i,j)\in E\\ 2 & \text{if } x_i=1,\ x_j=0,\ (i,j)\in E\\ 2 & \text{if } x_i=0,\ x_j=1,\ (i,j)\in E\\ 0 & \text{if } x_i=0,\ x_j=0,\ (i,j)\in E\\ 4 & \text{if } x_i=1,\ x_j=1,\ (i,j)\notin E\\ 2 & \text{if } x_i=1,\ x_j=0,\ (i,j)\notin E\\ 2 & \text{if } x_i=0,\ x_j=1,\ (i,j)\notin E\\ 0 & \text{if } x_i=0,\ x_j=0,\ (i,j)\notin E.\end{cases}$$

If $(i,j)\in E$, then we can write the expression above succinctly as $2\,x_i+2\,x_j-2\,x_i\,x_j$. So, we have $\sum_{(i,j)\in M}\mathbf{1}(i\in V,\ \ j\in V,\ \ (i,j)\in E)\,V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\,\hat{R}_{ij}(\boldsymbol{x})=\sum_{(i,j)\in E}(2\,x_i+2\,x_j-2\,x_i\,x_j)$, where we use the fact that having $(i,j)\in E$ implies that $i\in V$, $j\in V$ and $i<j$, which, in turn, implies that $(i,j)\in M$. On the other hand, if $(i,j)\notin E$, then we can write the expression above succinctly as $2\,x_i+2\,x_j$. Note that $\sum_{i\in V}\sum_{j\in V}\mathbf{1}(i<j,\ \ (i,j)\notin E)\,(x_i+x_j)=\sum_{i\in V}x_i\sum_{j\in V}\big[\mathbf{1}(i<j,\ (i,j)\notin E)+\mathbf{1}(j<i,\ (j,i)\notin E)\big]=\sum_{i\in V}x_i\,(n-1-d_i)$, where the last equality uses the fact that $\sum_{j\in V}\big[\mathbf{1}(i<j,\ (i,j)\notin E)+\mathbf{1}(j<i,\ (j,i)\notin E)\big]$ corresponds to the number of vertices that are not connected to vertex i with an edge, which is given by $n-1-d_i$. Thus, we obtain $\sum_{(i,j)\in M}\mathbf{1}(i\in V,\ j\in V,\ (i,j)\notin E)\,V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\,\hat{R}_{ij}(\boldsymbol{x})=\sum_{i\in V}\sum_{j\in V}\mathbf{1}(i<j,\ (i,j)\notin E)\times(2\,x_i+2\,x_j)=\sum_{i\in V}2\,(n-1-d_i)\,x_i$. Putting the discussion so far together, we have

$$\begin{aligned}\sum_{(i,j)\in M}V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\,\hat{R}_{ij}(\boldsymbol{x})&=\sum_{(i,j)\in M}\mathbf{1}(i\in W,\ j\in W)\,V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\,\hat{R}_{ij}(\boldsymbol{x})\\&\quad+\sum_{(i,j)\in M}\mathbf{1}(i\in V,\ j\in W)\,V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\,\hat{R}_{ij}(\boldsymbol{x})+\sum_{(i,j)\in M}\mathbf{1}(i\in V,\ j\in V,\ (i,j)\in E)\,V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\,\hat{R}_{ij}(\boldsymbol{x})\\&\quad+\sum_{(i,j)\in M}\mathbf{1}(i\in V,\ j\in V,\ (i,j)\notin E)\,V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\,\hat{R}_{ij}(\boldsymbol{x})\\&=4\,(n-1)\,(n-2)+\sum_{i\in V}\left\{\frac{d_i}{4}\,(4+2\,x_i)+\left(n-1-\frac{d_i}{4}\right)(4-2\,x_i)\right\}\\&\quad+\sum_{(i,j)\in E}(2\,x_i+2\,x_j-2\,x_i\,x_j)+\sum_{i\in V}2\,(n-1-d_i)\,x_i\\&=4\,(n-1)\,(n-2)+\sum_{i\in V}d_i+\sum_{i\in V}\frac{d_i}{2}\,x_i+4\,n\,(n-1)-\sum_{i\in V}2\,(n-1)\,x_i-\sum_{i\in V}d_i+\sum_{i\in V}\frac{d_i}{2}\,x_i\\&\quad+\sum_{i\in V}2\,d_i\,x_i-\sum_{(i,j)\in E}2\,x_i\,x_j+\sum_{i\in V}2\,(n-1)\,x_i-\sum_{i\in V}2\,d_i\,x_i\\&=8\,(n-1)^2+\sum_{i\in V}d_i\,x_i-2\sum_{(i,j)\in E}x_i\,x_j.\end{aligned}$$

Therefore, there exists $\boldsymbol{x}\in\{0,1\}^{|V|}$ that satisfies $\sum_{i\in V}d_i\,x_i-2\sum_{(i,j)\in E}x_i\,x_j\geq K$ if and only if there exists $\boldsymbol{x}=\{0,1\}^{2n-1}$ that satisfies $\sum_{(i,j)\in M}V_{ij}(\boldsymbol{x})^{\gamma_{ij}}\,\hat{R}_{ij}(\boldsymbol{x})\geq K+8\,(n-1)^2$. □

Appendix B: An Example with Different Levels of Correlation Between the Utilities

Consider the path choices of commuters traveling from the origin node to the destination node in the network shown in Figure 3. The edges in the network are labeled as $\{e_1, e_2, e_3, e_4, f_1, f_2, f_3, f_4\}$. The disutility of a path is the sum of the travel times on the edges that are included in the path. Possibly due to traffic conditions, the travel time on the edges are random but each commuter knows the travel times (or generates predictions) before deciding which path to take. Each commuter chooses the path that provides the largest utility. The modeler is interested in estimating the frequency with which each path is utilized over different days. We assume that the travel times on different edges are independent, but since two different paths may use a common edge, the utilities provided by two different paths can be correlated. Only for the brevity of discussion, we limit the paths that a commuter can choose to five possible paths, although our argument holds when there are other paths in consideration. In particular, a commuter chooses among the five paths given by

$$P_1 \;:\; e_1 \to e_2 \to e_3 \to e_4 \qquad P_2 \;:\; e_1 \to e_2 \to e_3 \to f_4 \qquad P_3 \;:\; e_1 \to e_2 \to f_3 \to f_4$$
$$P_4 \;:\; e_1 \to f_2 \to f_3 \to f_4 \qquad P_5 \;:\; f_1 \to f_2 \to f_3 \to f_4.$$

In this example, the utilities of different pairs of paths may have different levels of correlation and the correlation structure of the utilities may not satisfy the "transitivity" property. In particular, since different pairs of paths have different numbers of common edges, we expect their utilities to have different levels of correlation. Paths P_1 and P_2 have three common edges, so we expect the correlation between their utilities to be relatively high. Paths P_1 and P_4 have only one common edge, so we expect the correlation between their utilities to be relatively low. Also, the correlation structure of the utilities does not satisfy the "transitivity" property. The utilities of paths P_1 and P_3 are correlated since these paths have common edges. Similarly, the utilities of paths P_3 and P_5 are correlated as well. However, the utilities of paths P_1 and P_5 are uncorrelated, since these paths do not have any common edges. As the PCL model allows different levels of correlation between the utilities of different pairs of alternatives and accommodates utilities that do not satisfy the "transitivity" property, it can be a viable option to capture the path choices.

Numerical Study: We give a brief numerical study to check the ability of the PCL model to predict the path choices of commuters in the specific setting in Figure 3. In Appendix E, we give a more detailed numerical study to check the prediction ability of the PCL model. Consider the case where the utility provided by each edge is a normal random variable with mean 3 and standard deviation 1. The utility of a path is the sum of the utilities of the edges included in the path. A commuter chooses the path that provides the largest utility. Assuming that the commuters choose among the paths according to such a ground choice model, we generate the choice history from τ

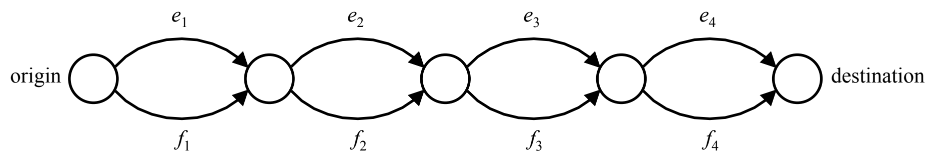

Figure 3 Network for a commuter that needs to travel from the origin node to the destination node.

commuters. We capture this choice history by $\{(S_t, i_t) : t = 1, \ldots, \tau\}$, where S_t is the subset of paths offered to commuter τ and i_t is the path chosen by commuter t. To generate the subset S_t, we include each path in the subset with probability 0.5. Given that we offer the subset S_t, we sample the path i_t according to the ground choice model. We refer to the choice history $\{(S_t, i_t) : t = 1, \ldots, \tau\}$ as the training data. We vary τ to work with different levels of of data availability. Note that the ground choice model that governs the choices in the training data is not the PCL model. We fit a PCL model to the training data by using standard maximum likelihood estimation; see, for example, Vulcano et al. (2012). We use the fmincon routine in Matlab to maximize the likelihood functions. For comparison purpose, we also fit a multinomial logit model. Using the same approach that we use for generating the training data, we also generate the choice history for another 10,000 commuters from the ground choice model. We refer to this choice history as the testing data and use it to test the performance of the fitted choice models.

Using the testing data, we compute the out of sample log-likelihoods of the fitted PCL and multinomial logit models. A larger out of sample log-likelihood indicates that the fitted choice model does a better job of predicting the choices of the commuters that are not in the training data. Furthermore, we compute the mean absolute errors in the choice probabilities of the fitted choice models. In particular, using $\phi_i^{\mathrm{GR}}(S)$ and $\phi_i^{\mathrm{PCL}}(S)$ to, respectively, denote the choice probability of path i out of the subset of paths S under the ground choice model and the fitted PCL model, letting $N = \{P_1, \ldots, P_5\}$ be the set of all paths, the mean absolute error in the choice probabilities of the fitted PCL model is given by $\frac{1}{\sum_{S \subseteq N} |S|} \sum_{S \subseteq N} \sum_{i \in S} |\phi_i^{\mathrm{PCL}}(S) - \phi_i^{\mathrm{GR}}(S)|$. We focus on mean absolute, rather than mean percent, errors because calculating mean percent errors may require divisions by small choice probabilities, putting disproportionate weight on misestimating small choice probabilities. We compute the mean absolute error in the choice probabilities of the fitted multinomial logit model similarly.

We give our numerical results in Figure 4. On the left side of the figure, we focus on the out of sample log-likelihoods of the fitted choice models. The horizontal axis shows the number of commuters τ in the training data set, capturing different levels of data availability to fit the

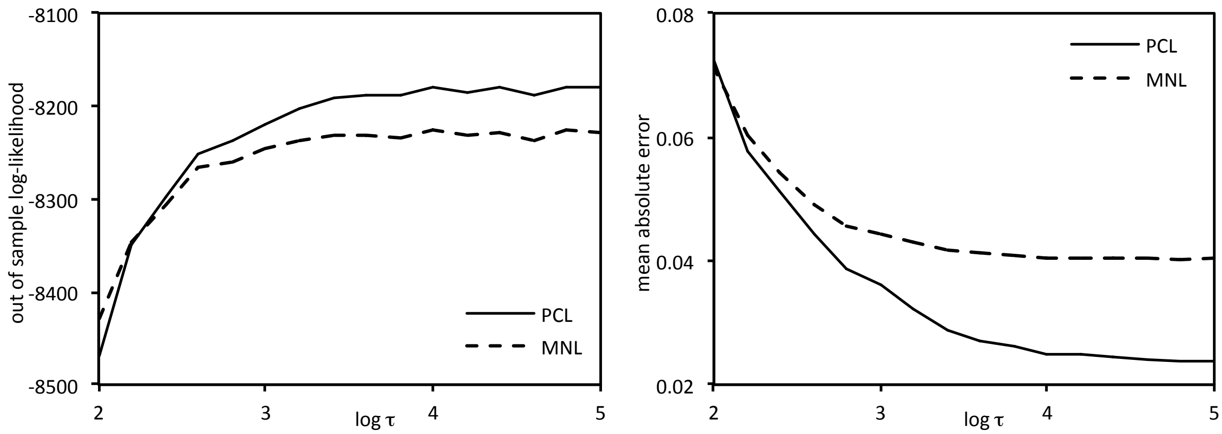

Figure 4 Out of sample log-likelihoods and mean absolute errors in the choice probabilities of the fitted PCL and multinomial logit models in the commuter path choice setting.

choice models. The two data series show the out of sample log-likelihoods of the fitted PCL and multinomial logit models when we have different levels of data availability to fit the two choice models. On the right side of the figure, we focus on the mean absolute errors in the choice probabilities of the fitted choice models. The two data series show the mean absolute errors of the fitted PCL and multinomial logit models when we have different levels of data availability to fit the two choice models. To eliminate the effect of noise, we replicated our numerical experiments 100 times. The data series in Figure 4 show the averages of our results over 100 replications. Using multiple replications smoothed out the data series, but in each replication, the data series had qualitatively the same nature, except that they were less smooth.

The results in Figure 4 indicate that if τ is reasonably large so that we have a reasonably large amount of data for fitting the choice models, then the out of sample log-likelihoods and the mean absolute errors of the fitted PCL model are significantly better than those of the fitted multinomial logit model. The performance of the two fitted choice models is comparable when we have too little data. The number of parameters for the PCL model is $O(|N|^2)$, whereas the number of parameters for the multinomial logit model is $O(|N|)$. If we have too little data, then it may be difficult to estimate the larger number of parameters for the PCL model, but as the data availability increases, the fitted PCL model yields significantly better out of sample log-likelihoods and mean absolute errors when compared with the fitted multinomial logit model.

Appendix C: Motivation for the Utilities under the Paired Combinatorial Logit Model

As we discuss at the beginning of Section 2.2, the PCL model is compatible with the random utility maximization principle. In this section, we give a motivation for the specific form of the utilities under the PCL model. Consider a process where a customer makes a choice by comparing the pairs of products. When the customer compares the pair of products $\{i,j\}$, she assigns the random value $X^i_{\{i,j\}}$ to product i and the random value $X^j_{\{i,j\}}$ to product j. The pair $\{i,j\}$ is unordered in the sense that we treat the pairs $\{i,j\}$ and $\{j,i\}$ as the same. After comparing every pair of products, the customer has the collection of values $\{X^i_{\{i,j\}} : j \in N,\ j \neq i\}$ associated with product i. The customer assigns a utility to product i by aggregating these values. One way to aggregate the values is to average them. Another way to aggregate the values is by taking the maximum. In the latter case, the utility of product i is given by

$$\mathsf{Utility}_i = \max\left\{X^i_{\{i,j\}} : j \in N,\ j \neq i\right\}. \tag{3}$$

Intuitively, in (3), the customer assigns the utility to product i based on her "best experience" among the pairwise comparisons of product i with all other products. If $X^i_{\{i,j\}}$ and $X^j_{\{i,j\}}$ follow a specific form, then the pairwise comparison process described above yields the PCL model.

In particular, let $\eta^i_{\{i,j\}}$ and $\eta^j_{\{i,j\}}$ each have the Gumbel distribution with location and scale parameters $(0,1)$. For a fixed $\gamma_{\{i,j\}} \in [0,1]$, let $Y_{\{i,j\}}$ have the C distribution with parameter $\gamma_{\{i,j\}}$; see Cardell (1997). Consider the case where $X^i_{\{i,j\}}$ and $X^j_{\{i,j\}}$ have the form

$$X^i_{\{i,j\}} = \mu_i + Y_{\{i,j\}} + \gamma_{\{i,j\}}\,\eta^i_{\{i,j\}} \quad \text{and} \quad X^j_{\{i,j\}} = \mu_j + Y_{\{i,j\}} + \gamma_{\{i,j\}}\,\eta^j_{\{i,j\}}, \tag{4}$$

where μ_i and μ_j are deterministic constants. Lastly, to capture the utility of the no purchase option, let $\mathsf{Utility}_0$ have the Gumbel distribution with location and scale parameters $(0,1)$. We assume that $\{Y_{\{i,j\}} : (i,j) \in N^2,\ i<j\} \cup \{\eta^k_{\{i,j\}} : (i,j) \in N^2,\ i<j,\ k \in \{i,j\}\} \cup \{\mathsf{Utility}_0\}$ are all independent of each other. In this case, by Theorem 2.1 in Cardell (1997), if the random variables $X^i_{\{i,j\}}$ and $X^j_{\{i,j\}}$ have the form in (4), then they have the Gumbel distribution with location and scale parameters, respectively, $(\mu_i,1)$ and $(\mu_j,1)$. Also, if the utility of product i has the form in (3)-(4) above and the customer chooses the alternative that provides the largest utility, then we can follow the same line of reasoning in Section 3 of Cardell (1997) to show that the probability that a customer chooses each product is precisely the same as the choice probability under the PCL model.

We can interpret the form of $X^i_{\{i,j\}}$ and $X^j_{\{i,j\}}$ in (4) as follows. The deterministic constants μ_i and μ_j, respectively, reflect the intrinsic values of products i and j. The random variable $Y_{\{i,j\}}$ captures the contribution to the value based on the common characteristics of products

i and j when the customer compares the two products. The random variables $\gamma_{\{i,j\}}\,\eta^i_{\{i,j\}}$ and $\gamma_{\{i,j\}}\,\eta^j_{\{i,j\}}$ represent idiosyncratic noises. Using the fact that the maximum of independent Gumbel random variables with the same shape parameter is also a Gumbel random variable with the same shape parameter, it follows that $\mathsf{Utility}_i$ has the Gumbel distribution with location and scale parameters $(\mu_i, 1)$. However, since $X^i_{\{i,j\}}$ and $X^j_{\{i,j\}}$ both depend on $Y_{\{i,j\}}$, $\mathsf{Utility}_i$ and $\mathsf{Utility}_j$ can be correlated. Also, it is instructive to consider special cases. If $\gamma_{\{i,j\}} = 0$, then $X^i_{\{i,j\}} = \mu_i + Y_{\{i,j\}}$ and $X^j_{\{i,j\}} = \mu_i + Y_{\{i,j\}}$, making them perfectly correlated, corresponding to the case where products i and j are perfectly similar to each other. Since the C distribution with parameter 1 is a degenerate distribution with a point mass at zero, if $\gamma_{\{i,j\}} = 1$, then $X^i_{\{i,j\}} = \mu_i + \eta^i_{\{i,j\}}$ and $X^j_{\{i,j\}} = \mu_j + \eta^j_{\{i,j\}}$, making $X^i_{\{i,j\}}$ and $X^j_{\{i,j\}}$ independent of each other, in which case, products i and j do not share any common characteristics.

The discussion in this section provides some motivation for the utilities of the products under the PCL model and can shed light into the aspects of the choice process that are captured by the PCL model, but not by other choice models such as the multinomial logit or nested logit models. Nevertheless, we emphasize that using the PCL model does not imply that the customers necessarily make a choice through pairwise comparisons. The important point is that the PCL model is a choice model that is based on the random utility maximization principle, where the utility of each product has the Gumbel distribution and there can be different levels of correlation between the utilities of different pairs of products.

Appendix D: Using the Same Paired Combinatorial Logit Model for Different Assortments

In the Assortment problem, we use the PCL model with the same parameters to capture the choices of the customers within different assortments. We simply drop the products that are not available in the offered assortment from the choice model. In this section, we justify this approach by using the fact that the PCL model is based on the random utility maximization principle. In particular, the PCL model is a generalized extreme value (GEV) model, which is a broad class of choice models based on the random utility maximization principle. In a GEV model, if we offer the assortment $\boldsymbol{x}$, then a customer associates the random utilities $\{\mu_i(x_i) + \epsilon_i : i \in N\}$ with the products, where $\mu_i(x_i)$ is the deterministic component and ϵ_i is the random shock for the utility of product i. For some fixed $\beta_i \in \mathbb{R}$, the deterministic component is given by $\mu_i(x_i) = \beta_i$ if $x_i = 1$, whereas $\mu_i(x_i) = -\infty$ if $x_i = 0$. Therefore, if a product is not offered, then its utility is negative infinity. Similarly, a customer associates the random utility $\mu_0 + \epsilon_0$ with the no purchase option. The no purchase option is always available, but for notational uniformity, we use $\mu_0(x_0)$ to denote the deterministic component of its utility. For some fixed $\beta_0 \in \mathbb{R}$, we have

$\mu_0(x_0)=\beta_0$. The random shocks $(\epsilon_0,\epsilon_1,\ldots,\epsilon_n)$ have a cumulative distribution function of the form $\mathbb{P}\{\epsilon_0\le u_0,\epsilon_1\le u_1,\ldots,\epsilon_n\le u_n\}=\exp(-G(e^{-u_0},e^{-u_1},\ldots,e^{-u_n}))$ for some function $G:\mathbb{R}_+^{n+1}\to\mathbb{R}_+$. Different choices of $G(\cdot,\cdot,\ldots,\cdot)$ yield different GEV models. A customer chooses the alternative with the largest utility. Therefore, if we offer the assortment $\boldsymbol{x}$, then a customer chooses product i with probability $\mathbb{P}\{\mu_i(x_i)+\epsilon_i=\max_{j\in N\cup\{0\}}\mu_j(x_j)+\epsilon_j\}$.

Theorem 1 in McFadden (1978) shows that if the function $G(\cdot,\cdot,\ldots,\cdot)$ satisfies a number of properties that ensure that $F(u_0,u_1,\ldots,u_n)=\exp(-G(e^{-u_0},e^{-u_1},\ldots,e^{-u_n}))$ is a cumulative distribution function, then the purchase probability of product i under a GEV model is given by $\mathbb{P}\{\mu_i(x_i)+\epsilon_i=\max_{j\in N\cup\{0\}}\mu_j(x_j)+\epsilon_j\}=\frac{e^{\mu_i(x_i)}\,\partial_i G(e^{\mu_0(x_0)},e^{\mu_1(x_1)},\ldots,e^{\mu_n(x_n)})}{G(e^{\mu_0(x_0)},e^{\mu_1(x_1)},\ldots,e^{\mu_n(x_n)})}$, where $\partial_i G(y_0,y_1,\ldots,y_n)$ is the partial derivative of $G(\cdot,\cdot,\ldots,\cdot)$ with respect to the i-th coordinate evaluated at $(y_0,y_1,\ldots,y_n)$. The PCL model is a GEV model with the choice of $G(\cdot,\cdot,\ldots,\cdot)$ given by $G(y_0,y_1,\ldots,y_n)=y_0+\sum_{(i,j)\in M}(y_i^{1/\gamma_{ij}}+y_j^{1/\gamma_{ij}})^{\gamma_{ij}}$ with $\gamma_{ij}=\gamma_{ji}$. The preference weight v_i of product i in Section 2.1 is related to the deterministic component of the utility of this product through the relationship $v_i=e^{\beta_i}$. Therefore, noting that $\mu_i(x_i)=\beta_i=\log v_i$ if $x_i=1$, whereas $\mu_i(x_i)=-\infty$ if $x_i=0$, we get $e^{\mu_i(x_i)}=v_i\,x_i$. In this case, if we offer the assortment $\boldsymbol{x}$, then the purchase probability of product i under the PCL model is

$$\mathbb{P}\Big\{\mu_i(x_i)+\epsilon_i=\max_{j\in N\cup\{0\}}\mu_j(x_j)+\epsilon_j\Big\}=\frac{e^{\mu_i(x_i)}\,\partial_i G(e^{\mu_0(x_0)},e^{\mu_1(x_1)},\ldots,e^{\mu_n(x_n)})}{G(e^{\mu_0(x_0)},e^{\mu_1(x_1)},\ldots,e^{\mu_n(x_n)})}$$

$$=\frac{e^{\mu_i(x_i)}\sum_{j\in N:(i,j)\in M}2\,e^{\mu_i(x_i)\,(1/\gamma_{ij}-1)}\big(e^{\mu_i(x_i)/\gamma_{ij}}+e^{\mu_j(x_j)/\gamma_{ij}}\big)^{\gamma_{ij}-1}}{e^{\mu_0(x_0)}+\sum_{(k,\ell)\in M}\big(e^{\mu_k(x_k)/\gamma_{k\ell}}+e^{\mu_\ell(x_\ell)/\gamma_{k\ell}}\big)^{\gamma_{k\ell}}}=\frac{\sum_{j\in N:(i,j)\in M}2\,\frac{v_i^{1/\gamma_{ij}}\,x_i}{V_{ij}(\boldsymbol{x})}\,V_{ij}(\boldsymbol{x})^{\gamma_{ij}}}{v_0+\sum_{(k,\ell)\in M}V_{k\ell}(\boldsymbol{x})^{\gamma_{k\ell}}},\qquad(5)$$

where the second equality follows directly by differentiating $G(\cdot,\cdot,\ldots,\cdot)$ and noting that $\gamma_{ij}=\gamma_{ji}$, whereas the third equality holds since $e^{\mu_i(x_i)/\gamma_{ij}}+e^{\mu_j(x_j)/\gamma_{ij}}=v_i^{1/\gamma_{ij}}\,x_i+v_j^{1/\gamma_{ij}}\,x_j=V_{ij}(\boldsymbol{x})$. In the choice process discussed in Section 2.1, if we offer the assortment $\boldsymbol{x}$, then a customer chooses nest (i,j) with probability $P_{ij}(\boldsymbol{x})=V_{ij}(\boldsymbol{x})^{\gamma_{ij}}/(v_0+\sum_{(k,\ell)\in M}V_{k\ell}(\boldsymbol{x})^{\gamma_{k\ell}})$. When $\gamma_{ij}=\gamma_{ji}$, if the customer chooses nest (i,j) or nest (j,i), then she purchases product i with probability $q^i_{ij}(\boldsymbol{x})=v_i^{1/\gamma_{ij}}\,x_i/V_{ij}(\boldsymbol{x})$. Thus, the purchase probability of product i under any assortment $\boldsymbol{x}$ in the choice process discussed in Section 2.1, which is given by $\sum_{j\in N:j\neq i}(P_{ij}(\boldsymbol{x})\,q^i_{ij}(\boldsymbol{x})+P_{ji}(\boldsymbol{x})\,q^i_{ji}(\boldsymbol{x}))$, is the same as the purchase probability in (5), which is obtained by using the random utility maximization principle, justifying the use of the PCL model with the same parameters to capture choices within different assortments.

Appendix E: Prediction Ability of the Paired Combinatorial Logit Model

We give a numerical study to check the ability of the PCL model to capture the choices of the customers within different assortments. We work with both synthetically generated data sets and data sets that are based on a real-world hotel revenue management application.

E.1. Synthetically Generated Data

In this set of numerical experiments, we synthetically generate the past purchase history of the customers under the assumption that the customers choose according to a complicated ground choice model. We fit a PCL model to this past purchase history. For comparison purpose, we also fit a multinomial logit model. Our goal is to understand the benefits from the PCL model from the perspective of predicting the choices of the customers within different assortments. In the ground choice model, there are m customer types. A customer of a particular type ranks a subset of the products according to a certain preference order and purchases the highest ranking available product. If none of the products in her preference list is available for purchase, then the customer leaves the system without a purchase. Thus, customers of different types are differentiated only by their preference lists. We use β^g to denote the probability that a customer of type g arrives into the system. Letting k^g be the number of products in the preference list of customers of type g, we use the tuple $(i^g(1),\ldots,i^g(k^g))$ to denote the preference list of customers of type g, where $i^g(\ell)\in N$ is the product with the ℓ-th highest ranking in the preference list. Therefore, the parameters of the ground choice model are $(\beta^1,\ldots,\beta^m)$ and $\{(i^g(1),\ldots,i^g(k^g)) : g=1,\ldots,m\}$.

Generating the Ground Choice Model and Estimation: We generate the parameters of the ground choice model as follows. We assume that the products $\{1,\ldots,n\}$ are indexed in the order of decreasing quality, where product 1 has the highest quality and product n has the lowest quality. A product with a higher quality also has a higher price. A customer of a particular type has a certain minimum acceptable quality and a maximum acceptable price. She drops products not satisfying these criteria from consideration. Also, she generally prefers a product with higher quality, but we add some noise to enrich the choice behavior. In particular, to generate the preference list for customers of type g, we sample ℓ^g from the uniform distribution over $\{1,\ldots,n\}$ and u^g from the uniform distribution over $\{\ell^g,\ldots,n\}$. Focusing on the products $\{\ell^g,\ldots,u^g\}$, we drop each one of these products with probability 0.1. We denote the remaining products by $\{j^g(1),\ldots,j^g(k^g)\}$, where product $j^g(1)$ has the highest quality and product $j^g(k^g)$ has the lowest quality. With probability 0.5, we randomly pick one of the products in $\{j^g(1),\ldots,j^g(k^g-1)\}$ and flip its ordering in the tuple $(j^g(1),\ldots,j^g(k^g))$ with its successor to obtain the tuple $(i^g(1),\ldots,i^g(k^g))$, yielding the preference list for customers of type g. With the remaining probability of 0.5, we keep the tuple $(j^g(1),\ldots,j^g(k^g))$ unchanged to obtain the tuple $(i^g(1),\ldots,i^g(k^g))$, once again, yielding the preference list for customers of type g. The approach that we use to generate the ground choice model follows the one in Feldman and Topaloglu (2017). A customer of each type g arrives with probability $\beta^g = 1/m$. In our numerical study, we set $m=50$ and $n=10$.

The design of our numerical study is similar to the one in Appendix B. Once we generate the ground choice model as in the previous paragraph, we generate the purchase history of τ customers under the assumption that the customers choose according to the ground choice model. We denote this purchase history by $\{(S_t,i_t): t=1,\ldots,\tau\}$, where S_t is the subset of products offered to customer t and i_t is the product purchased by customer t. To generate the subset of products S_t, we include each product in the subset with probability 0.5. Given that we offer the subset S_t, we sample product i_t according to the ground choice model. If the customer does not make a purchase, then we set $i_t = 0$. We use $\tau \in \{1000, 1750, 2500\}$ to capture three levels of data availability in the purchase history. We refer to these purchase histories as training data. We fit a PCL model and a multinomial logit model to the training data by using standard maximum likelihood estimation; see, for example, Vulcano et al. (2012). We use the fmincon routine in Matlab to maximize the likelihood functions. To use as testing data, using the same approach for generating the training data, we generate the purchase history of another 2500 customers under the assumption that these customers also choose according to the same ground choice model. We use the testing data to test the performance of the fitted choice models.

Out of Sample Log-Likelihoods: One approach to compare the fitted PCL and multinomial logit models is to check their out of sample log-likelihoods on the testing data. A larger out of sample log-likelihood indicates that the corresponding fitted choice model does a better job of predicting the choices of the customers that are not in the training data. To compare the out of sample log-likelihoods, we generate 10 different ground choice models. For each ground choice model, we generate three past purchase histories with $\tau \in \{1000, 1750, 2500\}$ customers. We fit a PCL model and a multinomial logit model to the three past purchase histories. When estimating the parameters of the PCL model, the fmincon routine in Matlab terminated with convergence to a local maximizer for all of our test problems and took 121 to 356 seconds depending on the number customers in the past purchase history. Once we fit a PCL and a multinomial logit model to a past purchase history, we check the out of sample log-likelihoods of the two fitted choice models on the testing data. We give our results in Table 3. Each row in the table shows the results for one of the 10 ground choice models. There are three blocks of three columns in the table. Each block corresponds to a different value of τ, capturing a different level of data availability. The first and second columns in each block show the out of sample log-likelihoods of the fitted PCL and

	$\tau = 1000$			$\tau = 1750$			$\tau = 2500$		
Grnd. Mod.	Log-Li. PCL	Log-Li. MNL	% Gap	Log-Li. PCL	Log-Li. MNL	% Gap	Log-Li. PCL	Log-Li. MNL	% Gap
1	−4171	−4189	0.42%	−4166	−4182	0.37%	−4164	−4181	0.40%
2	−4268	−4291	0.53%	−4266	−4294	0.64%	−4264	−4291	0.62%
3	−4189	−4212	0.56%	−4177	−4204	0.64%	−4179	−4207	0.68%
4	−4099	−4114	0.37%	−4093	−4112	0.47%	−4090	−4110	0.50%
5	−4228	−4237	0.21%	−4208	−4224	0.37%	−4203	−4220	0.42%
6	−4228	−4249	0.48%	−4220	−4243	0.53%	−4217	−4240	0.56%
7	−4242	−4261	0.45%	−4234	−4255	0.49%	−4231	−4254	0.54%
8	−4293	−4306	0.30%	−4289	−4304	0.35%	−4290	−4305	0.35%
9	−4213	−4223	0.22%	−4202	−4221	0.46%	−4200	−4220	0.47%
10	−4162	−4174	0.30%	−4162	−4173	0.28%	−4162	−4176	0.33%
Avg.			0.38%			0.46%			0.49%

Table 3 Out of sample log-likelihoods of the fitted PCL and multinomial logit models.

multinomial logit models. The third column shows the percent gap between the two log-likelihoods. Our results indicate that the fitted PCL model provides consistent improvements over the fitted multinomial logit model. The average gap between the out of sample log-likelihoods is 0.44%. Shortly, we also comment on the statistical significance of these gaps.

The PCL model has $O(n^2)$ parameters, corresponding to the preference weights of the products and the dissimilarity parameters of the nests, whereas the multinomial logit model has $O(n)$ parameters, corresponding to the mean utilities of the products. Due to its larger number of parameters, we expect that the PCL model provides more flexibility in modeling the choice behavior of the customers. However, due to its larger number of parameters, the PCL model may also over-fit to the training data, especially when we have too few customers in the training data. In this case, the PCL model may not provide satisfactory performance when we check its out of sample log-likelihoods on the testing data. Therefore, we cannot claim that the out of sample log-likelihoods of the fitted PCL model will always exceed those of the fitted multinomial logit model. In our numerical study, nevertheless, the possibility of over-fitting does not appear to be problematic for the PCL model, even when we have as few as 1000 customers in the training data and we estimate more than 100 parameters. As discussed in Section 1.1 in Bishop (2006), the possibility of over-fitting is a concern when working with a model with a large number of parameters, but this concern goes away when the amount of training data increases. Thus, the multinomial logit model may be preferable to the PCL model when the training data is scarce. As the amount of training data increases, the PCL model may be preferable to the multinomial logit model.

When over-fitting is a concern, one approach is to regularize the log-likelihood function by subtracting the penalty term $\lambda \sum_{(i,j)\in M}(1-\gamma_{ij})$ from the log-likelihood function, where λ is an adjustable penalty multiplier. As discussed in Section 2.2, if $\gamma_{ij} = 1$ for all $(i,j) \in M$, then the utilities are uncorrelated, so the PCL model becomes equivalent to the multinomial logit model.

	$\tau = 1000$			$\tau = 1750$			$\tau = 2500$		
Grnd. Mod.	Err. PCL	Err. MNL	% Gap	Err. PCL	Err. MNL	% Gap	Err. PCL	Err. MNL	% Gap
1	0.049	0.051	5.42%	0.048	0.051	5.60%	0.047	0.050	6.29%
2	0.047	0.050	5.79%	0.045	0.050	9.34%	0.045	0.049	8.94%
3	0.049	0.051	5.65%	0.047	0.050	7.29%	0.047	0.051	7.57%
4	0.049	0.051	5.65%	0.047	0.050	7.02%	0.046	0.050	7.55%
5	0.047	0.048	3.40%	0.044	0.047	5.60%	0.044	0.047	6.89%
6	0.048	0.051	6.22%	0.046	0.050	7.61%	0.046	0.050	7.96%
7	0.051	0.054	5.82%	0.050	0.053	6.43%	0.049	0.053	7.45%
8	0.048	0.050	3.18%	0.047	0.049	3.98%	0.047	0.049	4.08%
9	0.051	0.054	6.10%	0.050	0.054	7.59%	0.050	0.054	7.62%
10	0.044	0.047	6.18%	0.044	0.046	6.36%	0.044	0.047	6.70%
Avg.			5.34%			6.68%			7.10%

Table 4 Mean absolute errors in the choice probabilities of the fitted PCL and multinomial logit models.

Thus, the penalty term discourages moving away from the specification of the multinomial logit model unless it is clearly beneficial to do so. One can experiment with different values for the penalty multiplier λ to obtain a different fitted PCL model for each value. By checking the out of sample log-likelihoods of the fitted PCL model, one can pick the best one. This approach closely follows the standard regularization idea in model selection; see Section 3.5 in Hastie et al. (2017). Another approach to address over-fitting is to fit two versions of the PCL model after setting some or all of the dissimilarity parameters $\{\gamma_{ij} : (i,j) \in M\}$ to one. In this case, one can use the likelihood ratio test to compare the fit of the two models. Likelihood ratio test is a common approach in the choice modeling literature to avoid over-specifying the model; see, for example, Horowitz (1982), Gunn and Bates (1982), and Hausman and McFadden (1984).

Errors in Choice Probabilities: We also compare the errors in the purchase probabilities of the fitted choice models. In particular, we use $\phi_i^{\mathrm{GR}}(S)$ and $\phi_i^{\mathrm{PCL}}(S)$ to, respectively, denote the purchase probability of product i out of assortment S under the ground choice model and the fitted PCL model. For the testing data $\{(S_t, i_t) : t = 1, \ldots, 2500\}$, the mean absolute error in the purchase probabilities of the fitted PCL model is $\frac{1}{2500}\sum_{t=1}^{2500}\frac{1}{|S_t|}\sum_{i\in S_t}|\phi_i^{\mathrm{PCL}}(S_t) - \phi_i^{\mathrm{GR}}(S_t)|$. We can compute the mean absolute error of the fitted multinomial logit model similarly. We give our results in Table 4. The layout of this table is identical to that of Table 3. The only difference is that smaller mean absolute errors indicate better performance. The fitted PCL model has consistent improvements over the fitted multinomial logit model. Over all of our test problems, the average percent gap between the mean absolute errors of the two fitted choice models is 6.38%.

The results in Tables 3 and 4 are based on one training data set for each ground choice model and for each value of τ. To check the statistical significance of the results in these tables, we use bootstrapping. In particular, for each ground choice model and for each value of τ, we generate 10 past purchase histories to use as the training data. For each of the 10 bootstrapped past purchase

	$\tau = 1000$				$\tau = 1750$				$\tau = 2500$			
Grnd. Mod.	Mean Diff.	Std. Dev.	t-Sta.	p-Val.	Mean Diff.	Std. Dev.	t-Sta.	p-Val.	Mean Diff.	Std. Dev.	t-Sta.	p-Val.
1	0.38	0.08	14.92	5.9×10^{-8}	0.40	0.08	16.51	2.5×10^{-8}	0.42	0.08	17.60	1.4×10^{-8}
2	0.43	0.07	18.76	8.0×10^{-9}	0.49	0.07	20.71	3.3×10^{-9}	0.51	0.07	23.89	9.4×10^{-10}
3	0.48	0.06	23.22	1.2×10^{-9}	0.53	0.09	19.17	6.6×10^{-9}	0.54	0.08	20.64	3.4×10^{-9}
4	0.40	0.14	9.04	4.1×10^{-6}	0.49	0.09	16.54	2.4×10^{-8}	0.55	0.07	23.70	1.0×10^{-9}
5	0.31	0.08	12.75	2.3×10^{-7}	0.36	0.08	14.40	8.1×10^{-8}	0.39	0.07	17.42	1.5×10^{-8}
6	0.42	0.06	21.70	2.2×10^{-9}	0.46	0.06	22.35	1.7×10^{-9}	0.48	0.07	21.25	2.7×10^{-9}
7	0.53	0.13	12.88	2.1×10^{-7}	0.58	0.11	17.05	1.8×10^{-8}	0.59	0.09	21.43	2.5×10^{-9}
8	0.24	0.11	6.86	3.7×10^{-5}	0.29	0.06	15.49	4.3×10^{-8}	0.31	0.08	13.06	1.9×10^{-7}
9	0.46	0.17	8.89	4.7×10^{-6}	0.53	0.08	21.33	2.6×10^{-9}	0.58	0.06	28.35	2.1×10^{-10}
10	0.34	0.07	15.66	3.9×10^{-8}	0.37	0.06	19.27	6.3×10^{-9}	0.39	0.05	25.15	6.0×10^{-10}

Table 5 Statistical comparison of the out of sample log-likelihoods of the two fitted choice models.

histories, we replicate the results in Tables 3 and 4. In this case, we can check whether the gaps in the performance of the fitted PCL and multinomial logit models are statistically significant. It turns out that a sample size of 10 is enough to ensure that our results are statistically significant at a high significance level. In Table 5, we check the statistical significance of the gaps between the out of sample log-likelihoods of the fitted PCL and multinomial logit models. Each row in this table corresponds to a different ground choice model, whereas each block of four columns corresponds to a different value of τ. In each block, the first column shows the average percent gap between the out of sample log-likelihoods of the fitted PCL and multinomial logit models, where the average is computed over the 10 bootstrapped training data sets. The second column shows the standard deviation of the percent gaps between the out of sample log-likelihoods of the fitted PCL and multinomial logit models. In other words, using $\mathrm{LL}_s^{\mathrm{PCL}}$ and $\mathrm{LL}_s^{\mathrm{MNL}}$ to, respectively, denote the out of sample log-likelihoods of the fitted PCL and multinomial logit models from the s-th bootstrapped training data set, the first and second columns show the average and standard deviation of the data $\{100 \times (\mathrm{LL}_s^{\mathrm{PCL}} - \mathrm{LL}_s^{\mathrm{MNL}})/|\mathrm{LL}_s^{\mathrm{PCL}}| : s = 1, \ldots, 10\}$. The third column shows the t-statistic in the paired t-test to check the statistical significance of the gap between the out of sample log-likelihoods of the fitted PCL and multinomial logit models. The fourth column shows the p-value of the same test. The results in Table 5 indicate that the p-values are essentially zero for all ground choice models and all values of τ. Therefore, for all ground choice models and all levels of data availability, with extremely high statistical significance, the out of sample log-likelihoods of the fitted PCL model are better than those of the fitted multinomial logit model.

In Table 6, we use a similar approach to check the statistical significance of the gaps between the mean absolute errors of the choice probabilities from the fitted PCL and multinomial logit models. The layout of this table is identical to that of Table 5. The only difference is that if we use $\mathrm{MAE}_s^{\mathrm{PCL}}$ and $\mathrm{MAE}_s^{\mathrm{MNL}}$ to, respectively, denote the mean absolute errors of the fitted PCL and multinomial logit models that are obtained by using the s-th bootstrapped training

	$\tau = 1000$				$\tau = 1750$				$\tau = 2500$			
Grnd. Mod.	Mean Diff.	Std. Dev.	t-Sta.	p-Val.	Mean Diff.	Std. Dev.	t-Sta.	p-Val.	Mean Diff.	Std. Dev.	t-Sta.	p-Val.
1	4.96	0.56	27.91	2.4×10^{-10}	5.65	0.71	25.26	5.7×10^{-10}	5.88	0.58	31.80	7.4×10^{-11}
2	6.23	1.62	12.20	3.3×10^{-7}	8.14	1.31	19.67	5.3×10^{-9}	8.59	0.89	30.42	1.1×10^{-10}
3	5.96	0.49	38.29	1.4×10^{-11}	6.58	0.79	26.33	4.0×10^{-10}	7.17	0.45	49.90	1.3×10^{-12}
4	5.28	1.17	14.31	8.5×10^{-8}	6.43	1.04	19.56	5.5×10^{-9}	7.25	0.53	43.66	4.3×10^{-12}
5	5.19	1.17	14.07	9.8×10^{-8}	5.95	0.88	21.36	2.5×10^{-9}	6.49	0.88	23.40	1.1×10^{-9}
6	6.47	0.72	28.42	2.0×10^{-10}	7.49	0.64	36.80	2.0×10^{-11}	8.18	0.65	39.62	1.0×10^{-11}
7	6.83	1.18	18.37	9.6×10^{-9}	7.78	1.16	21.14	2.8×10^{-9}	8.19	0.81	31.82	7.4×10^{-11}
8	3.76	0.81	14.76	6.5×10^{-8}	4.52	0.62	23.01	1.3×10^{-9}	4.75	0.60	24.97	6.4×10^{-10}
9	6.06	1.37	13.99	1.0×10^{-7}	6.76	0.65	32.83	5.6×10^{-11}	7.48	0.84	28.18	2.2×10^{-10}
10	5.56	0.84	20.99	3.0×10^{-9}	6.42	1.03	19.78	5.0×10^{-9}	6.82	0.73	29.59	1.4×10^{-10}

Table 6 Statistical comparison of the mean absolute errors of the two fitted choice models.

data set, the first and second columns show the average and standard deviation of the data $\{100 \times (\text{MAE}_s^{\text{MNL}} - \text{MAE}_s^{\text{PCL}})/\text{MAE}_s^{\text{PCL}} : s = 1, \ldots, 10\}$. Once again, for all ground choice models and all levels of data availability, the p-values in Table 6 are essentially zero, indicating that the mean absolute errors of the choice probabilities predicted by the fitted PCL model are, with very high statistical significance, smaller than those predicted by the fitted multinomial logit model.

E.2. Hotel Revenue Management Data

In this set of numerical experiments, we use a data set from Bodea et al. (2009), which is based on a real-world hotel revenue management application. In this data set, we have customer purchase records for five different hotels from March 12, 2007 to April 15, 2007. Each purchase record gives the room type availability at the time of the booking and the room type that was booked. Room types take values such as king non-smoking and queen smoking. Each room type corresponds to a different product. We consider the choice behavior of the customers between the different products. In van Ryzin and Vulcano (2015), the authors also use this data set in their numerical experiments. We focus on the purchase records for each hotel separately. For one of the hotels, letting D be the number of days over which we have purchase records and R_ℓ be the number of purchase records on day ℓ, we use $\{(S^{\text{rec}}_{\ell,r}, i^{\text{rec}}_{\ell,r}) : \ell = 1, \ldots, D,\ r = 1, \ldots R_\ell\}$ to denote the purchase records in the data set, where $S^{\text{rec}}_{\ell,r}$ is the subset of products that is available when the purchase record r on day ℓ occurred and $i^{\text{rec}}_{\ell,r}$ is the product that was purchased within this subset. The purchase records in the data set do not include no purchase events by arriving customers.

Data Set to Fit the Choice Models: To obtain a data set that we can use to fit a PCL and a multinomial logit model, we build on the purchase records as follows. In the purchase records, the subsets of products that are offered on day ℓ are $\{S^{\text{rec}}_{\ell,r} : \ell = 1, \ldots, R_\ell\}$. There are R_ℓ subsets that are offered on day ℓ, but some of these subsets can be identical to each other. We assume that each subset $S^{\text{rec}}_{\ell,r}$ was offered for a duration of $24/R_\ell$ hours on day ℓ. During this duration of time, there was one purchase for product $i^{\text{rec}}_{r,\ell}$ and no other purchases. Timing of this purchase does not affect our fitting results, so we place this purchase at the beginning of the duration of $24/R_\ell$ hours. In the duration of $24/R_\ell$ hours, there may or may not have been other customer arrivals. The fact that there are no other purchases during this duration may be due to the fact that there were no customer arrivals or the arriving customers did not purchase anything. Thus, during the D days in the purchase records, we have access to the subsets of products that were offered at each time instant and the timing of the purchases. To obtain a data set that we can use to fit a PCL and a multinomial logit model, we divide each day into K time slots, each time slot representing a small enough duration of time that there is at most one customer arrival in each time slot. Since each time slot corresponds to a particular time interval on a particular day, we look up the subset of products offered in each time slot. Furthermore, on day ℓ, out of the time slots in which the subset $S^{\text{rec}}_{\ell,r}$ was offered, we assume that product $i^{\text{rec}}_{\ell,r}$ was purchased in one of the time slots and no products were purchased in the other time slots. This approach yields the subset that was offered and the product, if any, that was purchased in each time slot and on each day. We capture this data by $\{(S_{\ell,k}, i_{\ell,k}) : \ell = 1,\ldots,D,\ k = 1,\ldots,K\}$, where $S_{\ell,k}$ is the subset offered in time slot k on day ℓ and $i_{\ell,k}$ is the product purchased in time slot k on day ℓ. If there is no purchase in time slot k on day ℓ, then we set $i_{\ell,k} = 0$. In this way, the frequency with which each subset is offered in the data set $\{(S_{\ell,k}, i_{\ell,k}) : \ell = 1,\ldots,D,\ k = 1,\ldots,K\}$ is roughly equal to the frequency with which each subset is offered in the purchase records $\{(S^{\text{rec}}_{\ell,r}, i^{\text{rec}}_{\ell,r}) : \ell = 1,\ldots,D,\ r = 1,\ldots R_\ell\}$, but there are small rounding errors due to the fact that we divide each day into K discrete time slots. Furthermore, the number of purchases of each product out of each subset in the data set $\{(S_{\ell,k}, i_{\ell,k}) : \ell = 1,\ldots,D,\ k = 1,\ldots,K\}$ is exactly equal to the number of purchases of each product out of each subset in the purchase records $\{(S^{\text{rec}}_{\ell,r}, i^{\text{rec}}_{\ell,r}) : \ell = 1,\ldots,D,\ r = 1,\ldots R_\ell\}$.

Estimation: We use the data set $\{(S_{\ell,k}, i_{\ell,k}) : \ell = 1,\ldots,D,\ k = 1,\ldots,K\}$ to fit a PCL and a multinomial logit model. We split the DK offered subset-purchased product pairs in this data set into training and testing data sets. The training data set includes 9/10 fraction of the randomly chosen offered subset-purchased product pairs, whereas the testing data set includes the remaining pairs. We use $\{(S_t, i_t) : t = 1,\ldots,\tau\}$ to denote the offered subset-purchased product pairs in the training data set. Using $\boldsymbol{v} = \{v_i : i \in N\}$ and $\boldsymbol{\gamma} = \{\gamma_{ij} : (i,j) \in M\}$ to capture the parameters of the PCL model, we use $\phi^{\text{PCL}}_i(S \mid \boldsymbol{v}, \boldsymbol{\gamma})$ to denote purchase probability of product i out of the subset S under the PCL model with parameters $\boldsymbol{v}$ and $\boldsymbol{\gamma}$. Similarly, we use $\phi^{\text{PCL}}_0(S \mid \boldsymbol{v}, \boldsymbol{\gamma})$ to denote the no purchase probability under the PCL model with parameters $\boldsymbol{v}$ and $\boldsymbol{\gamma}$ when we offer the subset S. We assume that customers arrive according to a Poisson process with a stationary arrival rate. Our results qualitatively remained the same when we worked with different arrival rates in different

Problem instance characteristics	Hotel 1	Hotel 2	Hotel 3	Hotel 4	Hotel 5
No. Prd.	10	14	9	9	8
$\sum_{\ell=1}^{D} R_\ell$	1341	487	1271	309	260
DK	3596	1456	3477	1020	987

Out of sample log-likelihoods	Hotel 1	Hotel 2	Hotel 3	Hotel 4	Hotel 5
PCL	−434.77	−196.96	−390.19	−109.14	−97.24
MNL	−436.62	−198.17	−390.47	−109.29	−97.66
% Gap	0.42%	0.61%	0.07%	0.14%	0.43%

Table 7 Numerical results for the hotel revenue management data.

weeks. Using α to denote the probability that there is a customer arrival in each time slot, the log-likelihood function to fit a PCL model is

$$L(\boldsymbol{v},\boldsymbol{\gamma},\alpha) = \sum_{t=1}^{\tau} \log \Big(\mathbf{1}(i_t \neq 0)\, \alpha\, \phi_{i_t}^{\text{PCL}}(S_t \,|\, \boldsymbol{v},\boldsymbol{\gamma}) + \mathbf{1}(i_t = 0) \Big\{ (1-\alpha) + \alpha\, \phi_{i_t}^{\text{PCL}}(S_t \,|\, \boldsymbol{v},\boldsymbol{\gamma}) \Big\} \Big).$$

In the log-likelihood function above, we use the fact that if there is a purchase for product i in a time slot, then there was a customer arrival in this time slot and the customer purchased product i, whereas if there is no purchase in a time slot, then either there was no customer arrival or there was a customer arrival and the arriving customer did not purchase anything. We fit a PCL model to the training data by maximizing the log-likelihood function above subject to the constraints that $v_i \geq 0$ for all $i \in N$, $\gamma_{ij} \in [0,1]$ for all $(i,j) \in M$ and $\alpha \in [0,1]$. We use the fmincon routine in Matlab to maximize the log-likelihood function. Once we fit a PCL model, we compute the out of sample log-likelihood of the fitted PCL model by using the testing data. The log-likelihood function that we use for this purpose is the same as the one above, but we use the offered subset-purchased product pairs in the testing data instead of those in the training data. By using a similar approach, we fit a multinomial logit model to the training data and compute the out of sample log-likelihood of the fitted multinomial logit model by using the testing data.

Out of Sample Log-Likelihoods: We give our results in Table 7. On the left side of the table, we show statistics for the data sets for each hotel. The first row shows the number of products. The second row shows the number of purchase records $\{(S_{\ell,r}^{\text{rec}}, i_{\ell,r}^{\text{rec}}) : \ell = 1,\ldots,D,\ r = 1,\ldots,R_\ell\}$, which is given by $\sum_{\ell=1}^{D} R_\ell$. The third row shows the total number of data points in the training and testing data sets $\{(S_{\ell,k}, i_{\ell,k}) : \ell = 1,\ldots,D,\ k = 1,\ldots,K\}$, which is given by DK. The value of DK is different for different hotels because the booking requests arrive over different numbers of days for different hotels and we divide a day into different numbers of time slots for different hotels depending on the number of booking arrivals per day. On the right side of the table, we show the out of sample log-likelihoods of the fitted PCL and multinomial logit models, along with the percent gap between the out of sample log-likelihoods of the two fitted choice models. The gaps range between 0.07% and 0.61%, favoring the fitted PCL model. The results in this table are based on one partition of the data points between the training and testing data sets.

We use bootstrapping to check the statistical significance of the gaps between the out of sample log-likelihoods of the two fitted choice models. We generate 100 randomly chosen partitions of

	Hotel 1	Hotel 2	Hotel 3	Hotel 4	Hotel 5
Mean Diff.	0.51%	0.27%	0.18%	0.45%	0.50%
Std. Dev.	0.22%	0.73%	0.25%	1.41%	0.77%
t-Sta.	23.01	3.74	7.02	3.22	6.52
p-Val.	7.97×10^{-42}	1.54×10^{-4}	1.41×10^{-10}	8.79×10^{-4}	1.48×10^{-9}

Table 8 Statistical comparison of the out of sample log-likelihoods for the hotel revenue management data.

the data points between the training and testing data sets, where, for each partition, the training data set includes 9/10 fraction of the data points and the testing data set includes the remaining data points. For each bootstrapped partition, we replicate the results in Table 7. In this way, we have the out of sample log-likelihoods for the two fitted choice models from each bootstrapped partition. In Table 8, we check the statistical significance of the gaps between the out of sample log-likelihoods of the two fitted choice models. Each column in this table focuses on a different hotel. The first row shows the average percent gap between the out of sample log-likelihoods of the fitted PCL and multinomial logit models, where the average is computed over the 100 bootstrapped partitions. The second row shows the standard deviation of the percent gaps between the out of sample log-likelihoods of the fitted PCL and multinomial logit models. In particular, using $\mathrm{LL}_s^{\mathrm{PCL}}$ and $\mathrm{LL}_s^{\mathrm{MNL}}$ to, respectively, denote the out of sample log-likelihoods of the fitted PCL and multinomial logit models from the s-th bootstrapped partition, the first and second rows show the average and standard deviation of the data $\{100 \times (\mathrm{LL}_s^{\mathrm{PCL}} - \mathrm{LL}_s^{\mathrm{MNL}})/|\mathrm{LL}_s^{\mathrm{PCL}}| : s = 1, \ldots, 100\}$. The third row shows the t-statistic in the paired t-test to check the statistical significance between the out of sample log-likelihoods of the fitted PCL and multinomial logit models. The fourth row shows the p-value of the same test. In Table 8, the p-values are very close to zero. Therefore, for the data sets originating from each of the hotels, with very high statistical significance, the out of sample log-likelihoods of the fitted PCL model are better than those of the fitted multinomial logit model. Note that since we do not have access to the actual ground choice model that governs the choices of the customers in the data sets, we do not compare the mean absolute errors in the predicted purchase probabilities of the two fitted choice models.

Appendix F: Existence and Uniqueness of the Fixed Point

In the next lemma, we show that $f(z)$ given by the optimal objective value of the Function Evaluation problem is continuous in z and satisfies $f(0) \geq 0$.

Lemma F.1 *Letting $f(z)$ be the optimal objective value of the* Function Evaluation *problem, $f(z)$ is continuous in z and satisfies $f(0) \geq 0$.*

Proof: Setting $\boldsymbol{x} = \mathbf{0} \in \mathbb{R}^n$ is a feasible solution to the Function Evaluation problem. Noting that $V_{ij}(\mathbf{0}) = 0$, this solution provides an objective value of zero for the Function Evaluation problem. Therefore, we have $f(z) \geq 0$ for all $z \in \mathbb{R}$. Next, we show that $f(z)$ is continuous in z. For fixed $\boldsymbol{x} \in \mathcal{F}$, $\sum_{(i,j)\in M} V_{ij}(\boldsymbol{x})^{\gamma_{ij}} (R_{ij}(\boldsymbol{x}) - z)$ is linear in z. Furthermore, there are at most 2^n possible values of $\boldsymbol{x} \in \mathcal{F}$. So, noting the Function Evaluation problem, $f(z)$ is the pointwise maximum of at most 2^n linear functions of z. The pointwise maximum of linear functions is continuous. □

Appendix G: Simplifying the Upper Bound

In the next lemma, we show that we can drop some of the decision variables from the Upper Bound problem without changing the optimal objective value of this problem. This lemma becomes useful to obtain the Compact Upper Bound problem.

Lemma G.1 *There exists an optimal solution $\boldsymbol{x}^* = \{x_i^* : i \in N\}$ and $\boldsymbol{y}^* = \{y_{ij} : (i,j) \in M\}$ to the* Upper Bound *problem such that $x_i^* = 0$ for all $i \notin N(z)$ and $y_{ij}^* = 0$ for all $(i,j) \notin M(z)$.*

Proof: Letting $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ be an optimal solution to the Upper Bound problem, we define the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ as follows. We set $\hat{x}_i = x_i^*$ for all $i \in N(z)$, $\hat{x}_i = 0$ for all $i \notin N(z)$, $\hat{y}_{ij} = y_{ij}^*$ for all $(i,j) \in N(z)^2$ with $i \neq j$ and $\hat{y}_{ij} = 0$ for all $(i,j) \notin N(z)^2$ with $i \neq j$. Observe that $\hat{x}_i \leq x_i^*$ and $\hat{y}_{ij} \leq y_{ij}^*$. We claim that the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is feasible to the Upper Bound problem. To establish the claim, we note that if $(i,j) \in N(z)^2$, then we have $\hat{y}_{ij} = y_{ij}^* \geq x_i^* + x_j^* - 1 = \hat{x}_i + \hat{x}_j - 1$, where the two equalities follow from the definition of $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ and the inequality follows from the fact that $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is a feasible solution to the Upper Bound problem, so it satisfies the first constraint in this problem. Also, if $(i,j) \notin N(z)^2$, then we have $\hat{y}_{ij} = 0 \geq \hat{x}_i + \hat{x}_j - 1$, where the equality follows from the definition of $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ and the inequality follows from the fact that if $(i,j) \notin N(z)^2$, then we have $\hat{x}_i = 0$ or $\hat{x}_j = 0$, along with $\hat{x}_i \leq x_i^* \leq 1$ and $\hat{x}_j \leq x_j^* \leq 1$. Therefore, the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ satisfies the first constraint in the Upper Bound problem. On the other than, if $(i,j) \in N(z)^2$, then we have $\hat{y}_{ij} = y_{ij}^* \leq x_i^* = \hat{x}_i$, where the two equalities follow from the definition of $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ and the inequality follows from the fact that $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is a feasible solution to the Upper Bound problem. If $(i,j) \notin N(z)^2$, then we have $\hat{y}_{ij} = 0 \leq \hat{x}_i$, where the equality is by the definition of $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ and the inequality is simply by the fact that $\hat{x}_i \geq 0$. Therefore, the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ satisfies the second constraint in the Upper Bound problem. We can use the same approach to show that the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ satisfies the third constraint in the Upper Bound problem. Finally, since $\hat{x}_i \leq x_i^*$ for all $i \in N$, we have $\sum_{i\in N} \hat{x}_i \leq \sum_{i\in N} x_i^* \leq c$. Thus, the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is feasible to the Upper Bound problem.

Next, we claim that the objective value provided by the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ for the Upper Bound problem is at least as large as the objective value provided by the solution $(\boldsymbol{x}^*, \boldsymbol{y}^*)$. Nest (i,j) contributes the quantity $\mu_{ij}(z)\, y_{ij} + \theta_i(z)\, x_i + \theta_j(z)\, x_j$ to the objective function of the Upper Bound problem. To establish the claim, we show that the the contribution of each nest under the solution

$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is at least as large as the contribution under the solution $(\boldsymbol{x}^*, \boldsymbol{y}^*)$. If $i \in N(z)$ and $j \in N(z)$, then we have $\mu_{ij}(z)\, y^*_{ij} + \theta_i(z)\, x^*_i + \theta_j(z)\, x^*_j = \mu_{ij}(z)\, \hat{y}_{ij} + \theta_i(z)\, \hat{x}_i + \theta_j(z)\, \hat{x}_j$. If $i \notin N(z)$ and $j \notin N(z)$, then by the definition of $N(z)$, we have $\rho_{ij}(z) \le 0$, $\theta_i(z) \le 0$ and $\theta_j(z) \le 0$. In this case, we obtain $\mu_{ij}(z)\, y^*_{ij} + \theta_i(z)\, x^*_i + \theta_j(z)\, x^*_j = \rho_{ij}(z)\, y^*_{ij} + \theta_i(z)\,(x^*_i - y^*_{ij}) + \theta_j(z)\,(x^*_i - y^*_{ij}) \le 0 = \mu_{ij}(z)\, \hat{y}_{ij} + \theta_i(z)\, \hat{x}_i + \theta_j(z)\, \hat{x}_j$, where the inequality follows by noting that $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is a feasible solution to the Upper Bound problem so that $y^*_{ij} \le x^*_i$ and $y^*_{ij} \le x^*_j$, whereas the equality is by the fact that $\hat{y}_{ij} = 0$, $\hat{x}_i = 0$ and $\hat{x}_j = 0$ whenever $i \notin N(z)$ and $j \notin N(z)$. If $i \in N(z)$ and $j \notin N(z)$, then we have

$$\left(\frac{v_i^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}} + v_j^{1/\gamma_{ij}}} \right)^{\gamma_{ij}} (p_i - z) \ge \frac{v_i^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}} + v_j^{1/\gamma_{ij}}} (p_i - z) \ge \frac{(p_i - z)\, v_i^{1/\gamma_{ij}} + (p_j - z)\, v_j^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}} + v_j^{1/\gamma_{ij}}},$$

where the first inequality follows by noting that $a^\gamma \ge a$ for $a \in [0,1]$ and $\gamma \in [0,1]$, along with the fact that $i \in N(z)$ so that $p_i \ge z$, whereas the second inequality follows from the fact that $j \notin N(z)$ so that $p_j < z$. Focusing on the first and last expressions in the chain of inequalities above and noting the definitions of $\rho_{ij}(z)$ and $\theta_i(z)$, we obtain $\theta_i(z) \ge \rho_{ij}(z)$. In this case, we have $\mu_{ij}(z)\, y^*_{ij} + \theta_i(z)\, x^*_i + \theta_j(z)\, x^*_j = (\rho_{ij}(z) - \theta_i(z))\, y^*_{ij} + \theta_i(z)\, x^*_i + \theta_j(z)\,(x^*_i - y^*_{ij}) \le \theta_i(z)\, x^*_i = \mu_{ij}(z)\, \hat{y}_{ij} + \theta_i(z)\, \hat{x}_i + \theta_j(z)\, \hat{x}_j$, where the inequality follows from the fact that $\rho_{ij}(z) - \theta_i(z) \le 0$ and $j \notin N(z)$ so that $\theta_j(z) < 0$, whereas the second equality follows from the fact that $i \in N(z)$ and $j \notin N(z)$, in which case, we have $\hat{y}_{ij} = 0$, $\hat{x}_i = x^*_i$ and $\hat{x}_j = 0$. If $i \notin N(z)$ and $j \in N(z)$, then we can use the same approach to show that $\mu_{ij}(z)\, y^*_{ij} + \theta_i(z)\, x^*_i + \theta_j(z)\, x^*_j \le \mu_{ij}(z)\, \hat{y}_{ij} + \theta_i(z)\, \hat{x}_i + \theta_j(z)\, \hat{x}_j$. The only difference is that we need to interchange the roles of the decision variables x_i and x_j. Therefore, in all of the four cases considered, the contribution of nest (i,j) under the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is at least as large as the contribution of nest (i,j) under the solution $(\boldsymbol{x}^*, \boldsymbol{y}^*)$. □

In the next lemma, we show that $\mu_{ij}(z) \le 0$ for all $(i,j) \in M(z)$. This lemma is used at multiple places throughout the paper.

Lemma G.2 *For all $z \in \mathbb{R}$ and $(i,j) \in M(z)$, we have $\mu_{ij}(z) \le 0$.*

Proof: Since $(i,j) \in M(z)$, we have $i \in N(z)$ and $j \in N(z)$, which implies that $p_i - z \ge 0$ and $p_j - z \ge 0$. Using the fact that $a \le a^\gamma$ for $a \in [0,1]$ and $\gamma \in [0,1]$, we obtain

$$\frac{(p_i - z)\, v_i^{1/\gamma_{ij}} + (p_j - z)\, v_j^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}} + v_j^{1/\gamma_{ij}}} \le (p_i - z) \left(\frac{v_i^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}} + v_j^{1/\gamma_{ij}}} \right)^{\gamma_{ij}} + (p_j - z) \left(\frac{v_j^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}} + v_j^{1/\gamma_{ij}}} \right)^{\gamma_{ij}}.$$

Multiplying both sides of this inequality with $(v_i^{1/\gamma_{ij}} + v_j^{1/\gamma_{ij}})^{\gamma_{ij}}$ and using the definitions of $\rho_{ij}(z)$ and $\theta_i(z)$, we get $\rho_{ij}(z) \le \theta_i(z) + \theta_j(z)$. Thus, we have $\mu_{ij}(z) = \rho_{ij}(z) - \theta_i(z) - \theta_j(z) \le 0$. □

Appendix H: Improving the Performance Guarantee

We give the full proof for Theorem 4.1 with the 0.6-approximation guarantee. The proof is lengthy, so we begin with an outline. In particular, the proof uses the following four steps.

Step 1: Using the vector $\boldsymbol{\theta} = \{\theta_i : i \in \hat{N}\}$, we will construct a function $F : \mathbb{R}_+^{|\hat{N}|} \to \mathbb{R}_+$ that satisfies $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \geq f^R - \frac{1}{2}F(\boldsymbol{\theta})$.

Step 2: We will also construct a function $G : \mathbb{R}_+^{|\hat{N}|} \times \mathbb{Z}_+^3 \to \mathbb{R}_+$ that satisfies $f^R \geq G(\boldsymbol{\theta}, k_1, k_2, |\hat{N}|)$ for any $1 \leq k_1 \leq k_2 \leq |\hat{N}|$.

To establish Step 1, we observe that $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \geq f^R + \frac{1}{4}\sum_{(i,j)\in \hat{M}} \mu_{ij}$ by the discussion in the proof of Theorem 4.1 in the main body of the paper. Thus, it will be enough to give a lower bound on μ_{ij} as a function of θ_i and θ_j. To establish Step 2, we construct a feasible solution to the problem that computes f^R at the beginning of Section 4. In particular, recalling that $\theta_i \geq 0$ for all $i \in \hat{N}$, we index the products so that $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_{|\hat{N}|} \geq 0$. The solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ obtained by setting $\hat{x}_i = 1$ for all $i \in \{1, \ldots, k_1\}$, $\hat{x}_i = \frac{1}{2}$ for all $i \in \{k_1 + 1, \ldots, k_2\}$ and $\hat{x}_i = 0$ for all $i \in \{k_2 + 1, \ldots, |\hat{N}|\}$ along with $\hat{y}_{ij} = [\hat{x}_i + \hat{x}_j - 1]^+$ for all $(i,j) \in \hat{M}$ is feasible to the problem that computes f^R at the beginning of Section 4. In this case, f^R will be lower bounded by the objective function of the problem evaluated at this solution. Using Steps 1 and 2, we get

$$\begin{aligned}
\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \;&\geq\; f^R - \frac{1}{2}F(\boldsymbol{\theta}) \;=\; \left(1 - \frac{F(\boldsymbol{\theta})}{2\, f^R}\right) f^R \\
&\geq \left(1 - \frac{F(\boldsymbol{\theta})}{2 \times \max_{\substack{(k_1,k_2):\\ 1\leq k_1\leq k_2\leq |\hat{N}|}} G(\boldsymbol{\theta}, k_1, k_2, |\hat{N}|)}\right) f^R \;=\; \left(1 - \frac{1}{2} \times \min_{\substack{(k_1,k_2):\\ 1\leq k_1\leq k_2\leq |\hat{N}|}} \left\{\frac{F(\boldsymbol{\theta})}{G(\boldsymbol{\theta}, k_1, k_2, |\hat{N}|)}\right\}\right) f^R \\
&\geq \left(1 - \frac{1}{2} \times \min_{\substack{(k_1,k_2):\\ 1\leq k_1\leq k_2\leq |\hat{N}|}} \left\{ \max_{\substack{(\theta_1,\ldots,\theta_{|\hat{N}|}):\\ \theta_1\geq\ldots\geq\theta_{|\hat{N}|}\geq 0}} \left\{\frac{F(\boldsymbol{\theta})}{G(\boldsymbol{\theta}, k_1, k_2, |\hat{N}|)}\right\}\right\}\right) f^R. \qquad (6)
\end{aligned}$$

Step 3: Letting $a \vee b = \max\{a, b\}$, we will construct functions $\Gamma_1 : \mathbb{Z}_+^3 \to \mathbb{R}_+$, $\Gamma_2 : \mathbb{Z}_+^3 \to \mathbb{R}_+$ and $\Gamma_3 : \mathbb{Z}_+^3 \to \mathbb{R}_+$ that, for any $1 \leq k_1 \leq k_2 \leq |\hat{N}|$, satisfy

$$\max_{\substack{(\theta_1,\ldots,\theta_{|\hat{N}|}):\\ \theta_1\geq\ldots\geq\theta_{|\hat{N}|}\geq 0}} \left\{\frac{F(\boldsymbol{\theta})}{G(\boldsymbol{\theta}, k_1, k_2, |\hat{N}|)}\right\} \;\leq\; \Gamma_1(k_1, k_2, |\hat{N}|) \vee \Gamma_2(k_1, k_2, |\hat{N}|) \vee \Gamma_3(k_1, k_2, |\hat{N}|).$$

Step 4: For any $|\hat{N}|$, we will show that there exist k_1 and k_2 such that $1 \leq k_1 \leq k_2 \leq |\hat{N}|$, $\Gamma_1(k_1, k_2, |\hat{N}|) \leq 0.8$, $\Gamma_2(k_1, k_2, |\hat{N}|) \leq 0.8$ and $\Gamma_3(k_1, k_2, |\hat{N}|) \leq 0.8$.

To establish Step 3, we note that $F(\boldsymbol{\theta})$ and $G(\boldsymbol{\theta}, k_1, k_2, |\hat{N}|)$ are linear in $\boldsymbol{\theta}$ in our construction. Thus, the objective function of the problem $\max_{(\theta_1,\ldots,\theta_{|\hat{N}|}):\theta_1\geq\ldots\geq\theta_{|\hat{N}|}\geq 0}\left\{\frac{F(\boldsymbol{\theta})}{G(\boldsymbol{\theta},k_1,k_2,|\hat{N}|)}\right\}$ is quasi-linear in $\boldsymbol{\theta}$, so an optimal solution occurs at an extreme point of the set of feasible solutions. In this case, we construct the functions $\Gamma_1(\cdot,\cdot,\cdot)$, $\Gamma_2(\cdot,\cdot,\cdot)$ and $\Gamma_3(\cdot,\cdot,\cdot)$ by checking the objective value of the last maximization problem at the possible extreme points of the set of feasible solutions. To establish Step 4, we show that if $|\hat{N}|$ is large enough, then we can choose k_1 and k_2 as fixed fractions of $|\hat{N}|$ to obtain $\Gamma_1(k_1,k_2,|\hat{N}|) \le 0.8$, $\Gamma_2(k_1,k_2,|\hat{N}|) \le 0.8$ and $\Gamma_3(k_1,k_2,|\hat{N}|) \le 0.8$. In particular, using $\lceil\cdot\rceil$ to denote the round up function and fixing $\hat{\beta}_1 = 0.088302$ and $\hat{\beta}_2 = 0.614542$ we show that if $|\hat{N}| \ge 786$, then we have $\Gamma_1(\lceil\hat{\beta}_1\,|\hat{N}|\rceil, \lceil\hat{\beta}_2\,|\hat{N}|\rceil, |\hat{N}|) \le 0.8$, $\Gamma_2(\lceil\hat{\beta}_1\,|\hat{N}|\rceil, \lceil\hat{\beta}_2\,|\hat{N}|\rceil, |\hat{N}|) \le 0.8$ and $\Gamma_3(\lceil\hat{\beta}_1\,|\hat{N}|\rceil, \lceil\hat{\beta}_2\,|\hat{N}|\rceil, |\hat{N}|) \le 0.8$. On the other hand, if $|\hat{N}| < 786$, then we enumerate all values of $(k_1,k_2) \in \mathbb{Z}^2$ with $1 \le k_1 \le k_2 \le |\hat{N}|$ to numerically check that $\Gamma_1(k_1,k_2,|\hat{N}|) \le 0.8$, $\Gamma_2(k_1,k_2,|\hat{N}|) \le 0.8$ and $\Gamma_3(k_1,k_2,|\hat{N}|) \le 0.8$. Using Steps 3 and 4, we get

$$\min_{\substack{(k_1,k_2):\\ 1\le k_1\le k_2\le |\hat{N}|}} \left\{ \max_{\substack{(\theta_1,\ldots,\theta_{|\hat{N}|}):\\ \theta_1\ge\ldots\ge\theta_{|\hat{N}|}\ge 0}} \left\{ \frac{F(\boldsymbol{\theta})}{G(\boldsymbol{\theta},k_1,k_2,|\hat{N}|)} \right\} \right\}$$
$$\le \min_{\substack{(k_1,k_2):\\ 1\le k_1\le k_2\le |\hat{N}|}} \left\{ \Gamma_1(k_1,k_2,|\hat{N}|) \vee \Gamma_2(k_1,k_2,|\hat{N}|) \vee \Gamma_3(k_1,k_2,|\hat{N}|) \right\} \le 0.8. \quad (7)$$

Therefore, by (6) and (7), we get $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\} \ge 0.6\, f^R$, which is the desired result. In Appendix H.1, we establish Steps 1 and 2. In Appendix H.2, we establish Step 3. In Appendix H.3, we establish Step 4.

H.1. Preliminary Bounds

In this section, we establish Steps 1 and 2 in our outline of the proof of Theorem 4.1. Throughout this section, we let $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ be an optimal solution to the LP that computes f^R at the beginning of Section 4. Also, we recall that the random subset of products $\hat{\boldsymbol{X}} = \{\hat{X}_i : i \in N\}$ is defined as follows. For all $i \in \hat{N}$, we have $\hat{X}_i = 1$ with probability x_i^* and $\hat{X}_i = 0$ with probability $1 - x_i^*$. Lastly, we have $\hat{X}_i = 0$ for all $i \in N \setminus \hat{N}$. Different components of the vector $\hat{\boldsymbol{X}}$ are independent of each other. For notational brevity, we let $m = |\hat{N}|$. In this case, since $n = |N|$, we write the LP that computes f^R at the beginning of Section 4 as

$$f^R = \max\left\{ \sum_{(i,j)\in\hat{M}} (\mu_{ij}\, y_{ij} + \theta_i\, x_i + \theta_j\, x_j) + 2\,(n-m) \sum_{i\in\hat{N}} \theta_i\, x_i \,:\, y_{ij} \ge x_i + x_j - 1 \quad \forall\,(i,j)\in\hat{M},\right.$$
$$\left. 0 \le x_i \le 1 \quad \forall\, i \in \hat{N}, \quad y_{ij} \ge 0 \quad \forall\,(i,j)\in\hat{M} \right\}. \quad (8)$$

We index the elements of $\hat{N}$ as $\{1,\ldots,m\}$ and the elements of $N \setminus \hat{N}$ as $\{m+1,\ldots,n\}$. Without loss of generality, we assume that the products in $\hat{N}$ are indexed such that $\theta_1 \ge \theta_2 \ge \ldots \ge \theta_m$. In the next lemma, we give a lower bound on $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - \hat{z})\}$.

Lemma H.1 *We have* $\sum_{(i,j)\in M}\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}})-\hat{z})\}\geq\Big(1-\frac{\sum_{i=1}^m(m-i)\,\theta_i}{2\,f^R}\Big)f^R$.

Proof: Noting the discussion in the proof of Theorem 4.1 in the main body of the paper, we have $\sum_{(i,j)\in M}\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}})-\hat{z})\}\geq f^R+\frac{1}{4}\sum_{(i,j)\in\hat{M}}\mu_{ij}$. Lemma H.2 given below shows that $\mu_{ij}\geq-\max\{\theta_i,\theta_j\}$ for all $(i,j)\in\hat{M}$. In this case, we obtain

$$\begin{aligned}\sum_{(i,j)\in M}\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}})-\hat{z})\} &\geq f^R+\frac{1}{4}\sum_{(i,j)\in\hat{M}}\mu_{ij} \geq f^R-\frac{1}{4}\sum_{(i,j)\in\hat{M}}\max\{\theta_i,\theta_j\}\\ &= f^R-\frac{1}{4}\sum_{i\in\hat{N}}\sum_{j\in\hat{N}}\mathbf{1}(i<j)\max\{\theta_i,\theta_j\}-\frac{1}{4}\sum_{i\in\hat{N}}\sum_{j\in\hat{N}}\mathbf{1}(i>j)\max\{\theta_i,\theta_j\}\\ &= f^R-\frac{1}{4}\sum_{i\in\hat{N}}\sum_{j\in\hat{N}}\mathbf{1}(i<j)\,\theta_i-\frac{1}{4}\sum_{i\in\hat{N}}\sum_{j\in\hat{N}}\mathbf{1}(i>j)\,\theta_j\\ &= f^R-\frac{1}{4}\sum_{i\in\hat{N}}\Big\{\sum_{j\in\hat{N}}\mathbf{1}(i<j)\Big\}\theta_i-\frac{1}{4}\sum_{j\in\hat{N}}\Big\{\sum_{i\in\hat{N}}\mathbf{1}(i>j)\Big\}\theta_j\\ &= f^R-\frac{1}{4}\sum_{i\in\hat{N}}(m-i)\,\theta_i-\frac{1}{4}\sum_{j\in\hat{N}}(m-j)\,\theta_j = \left(1-\frac{\sum_{i\in\hat{N}}(m-i)\,\theta_i}{2\,f^R}\right)f^R,\end{aligned}$$

where the second equality uses the fact that $\theta_1\geq\theta_2\geq\ldots\geq\theta_m$ and the fourth equality uses the fact that $|\hat{N}|=m$. $\square$

We use the next lemma in the proof of Lemma H.1.

Lemma H.2 *For all* $z\in\mathbb{R}$ *and* $(i,j)\in M(z)$*, we have* $\mu_{ij}(z)\geq-\max\{\theta_i(z),\theta_j(z)\}$.

Proof: For $(i,j)\in M(z)$, we have $i\in N(z)$ and $j\in N(z)$, which implies that $p_i\geq z$ and $p_j\geq z$. Using the definitions of $\rho_{ij}(z)$ and $\theta_i(z)$, we get

$$\begin{aligned}\rho_{ij}(z) &= (v_i^{1/\gamma_{ij}}+v_j^{1/\gamma_{ij}})^{\gamma_{ij}}\,\frac{(p_i-z)\,v_i^{1/\gamma_{ij}}+(p_j-z)\,v_j^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}}+v_j^{1/\gamma_{ij}}} = \frac{(p_i-z)\,v_i^{1/\gamma_{ij}}+(p_j-z)\,v_j^{1/\gamma_{ij}}}{(v_i^{1/\gamma_{ij}}+v_j^{1/\gamma_{ij}})^{1-\gamma_{ij}}}\\ &= (p_i-z)\,v_i\left(\frac{v_i^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}}+v_j^{1/\gamma_{ij}}}\right)^{1-\gamma_{ij}}+(p_j-z)\,v_j\left(\frac{v_j^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}}+v_j^{1/\gamma_{ij}}}\right)^{1-\gamma_{ij}}\\ &= \theta_i(z)\left(\frac{v_i^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}}+v_j^{1/\gamma_{ij}}}\right)^{1-\gamma_{ij}}+\theta_j(z)\left(\frac{v_j^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}}+v_j^{1/\gamma_{ij}}}\right)^{1-\gamma_{ij}}\\ &\geq \theta_i(z)\left(\frac{v_i^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}}+v_j^{1/\gamma_{ij}}}\right)+\theta_j(z)\left(\frac{v_j^{1/\gamma_{ij}}}{v_i^{1/\gamma_{ij}}+v_j^{1/\gamma_{ij}}}\right) \geq \min\{\theta_i(z),\theta_j(z)\},\end{aligned}$$

where the first inequality uses the fact that $a^{1-\gamma}\geq a$ for $a\in[0,1]$ and $\gamma\in[0,1]$. In this case, we get $\mu_{ij}(z)=\rho_{ij}(z)-\theta_i(z)-\theta_j(z)\geq\min\{\theta_i(z),\theta_j(z)\}-\theta_i(z)-\theta_j(z)=-\max\{\theta_i(z),\theta_j(z)\}$. $\square$

Using the vector $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_m)$, the function $\sum_{i=1}^m(m-i)\,\theta_i$ in Lemma H.1 corresponds to the function $F(\boldsymbol{\theta})$ in Step 1 in our outline of the proof of Theorem 4.1. Therefore, Lemma H.1

establishes Step 1. Next, we focus on establishing Step 2. In particular, we construct a function $G : \mathbb{R}_+^m \times \mathbb{Z}_+^3 \to \mathbb{R}_+$ that satisfies $f^R \geq G(\boldsymbol{\theta}, k_1, k_2, m)$ for any $1 \leq k_1 \leq k_2 \leq m$.

Lemma H.3 *We have*

$$f^R \;\geq \max_{\substack{(k_1,k_2):\\ 1\le k_1\le k_2\le m}} \left\{ \sum_{i=1}^{k_1} (2\,n - k_1 - k_2 + 2\,i - 2)\,\theta_i + \sum_{i=k_1+1}^{k_2} (n-1)\,\theta_i \right\}.$$

Proof: Consider the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \in \mathbb{R}_+^{|\hat{N}|} \times \mathbb{R}_+^{|\hat{M}|}$ to problem (8) that is obtained by letting $\hat{x}_i = 1$ for all $i \in \{1,\dots,k_1\}$, $x_i = \frac{1}{2}$ for all $i \in \{k_1+1,\dots,k_2\}$ and $\hat{x}_i = 0$ for all $i \in \{k_2+1,\dots,m\}$ and $\hat{y}_{ij} = [\hat{x}_i + \hat{x}_j - 1]^+$. The solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is feasible but not necessarily optimal to problem (8), in which case, noting that the optimal objective value of problem (8) is f^R, we get

$$\begin{aligned} f^R \;&\geq \sum_{(i,j)\in\hat{M}} (\mu_{ij}\,\hat{y}_{ij} + \theta_i\,\hat{x}_i + \theta_j\,\hat{x}_j) + 2\,(n-m) \sum_{i\in\hat{N}} \theta_i\,\hat{x}_i \\ &= \sum_{(i,j)\in\hat{M}} \mu_{ij}\,\hat{y}_{ij} + \sum_{i\in\hat{N}}\sum_{j\in\hat{N}} \mathbf{1}(i\neq j)\,(\theta_i\,\hat{x}_i + \theta_j\,\hat{x}_j) + 2\,(n-m)\sum_{i\in\hat{N}}\theta_i\,\hat{x}_i. \end{aligned}$$

Since $|\hat{N}| = m$, we have $\sum_{i\in\hat{N}}\sum_{j\in\hat{N}} \mathbf{1}(i\neq j)\,\theta_i\,\hat{x}_i = (m-1)\sum_{i\in\hat{N}}\theta_i\,\hat{x}_i$. Similarly, we have $\sum_{j\in\hat{N}}\sum_{i\in\hat{N}} \mathbf{1}(i\neq j)\,\theta_j\,\hat{x}_j = (m-1)\sum_{j\in\hat{N}}\theta_j\,\hat{x}_j$. Thus, the chain of inequalities above yields

$$\begin{aligned} f^R \;&\geq \sum_{(i,j)\in\hat{M}} \mu_{ij}\,\hat{y}_{ij} + 2\,(m-1)\sum_{i\in\hat{N}}\theta_i\,\hat{x}_i + 2\,(n-m)\sum_{i\in\hat{N}}\theta_i\,\hat{x}_i \\ &= \sum_{(i,j)\in\hat{M}} \mu_{ij}\,\hat{y}_{ij} + 2\,(n-1)\sum_{i\in\hat{N}}\theta_i\,\hat{x}_i \\ &\geq -\sum_{(i,j)\in\hat{M}} \max\{\theta_i,\theta_j\}\,\hat{y}_{ij} + 2\,(n-1)\sum_{i\in\hat{N}}\theta_i\,\hat{x}_i, \end{aligned}$$

where the last inequality follows from the fact that we have $\mu_{ij} \geq -\max\{\theta_i,\theta_j\}$ for all $(i,j)\in\hat{M}$ by Lemma H.2. We compute each one of the two sums on the right side of the inequality above separately. Considering the sum $\sum_{i\in\hat{N}}\theta_i\,\hat{x}_i$, the definition of $\hat{x}_i$ implies that

$$\sum_{i\in\hat{N}}\theta_i\,\hat{x}_i = \sum_{i=1}^{k_1}\theta_i + \sum_{i=k_1+1}^{k_2}\frac{\theta_i}{2}.$$

On the other hand, considering the sum $\sum_{(i,j)\in\hat{M}}\max\{\theta_i,\theta_j\}\,\hat{y}_{ij}$, we have $\sum_{(i,j)\in\hat{M}}\max\{\theta_i,\theta_j\}\,\hat{y}_{ij} = \sum_{(i,j)\in\hat{M}}\mathbf{1}(i<j)\,\theta_i\,\hat{y}_{ij} + \sum_{(i,j)\in\hat{M}}\mathbf{1}(i>j)\,\theta_j\,\hat{y}_{ij}$, where we use the fact that $\theta_1\geq\theta_2\geq\ldots\geq\theta_m$. Noting the definition of $\hat{\boldsymbol{x}}$ and using the fact that $\hat{y}_{ij} = [\hat{x}_i + \hat{x}_j - 1]^+$ at the beginning of the proof, we have $\hat{y}_{ij} = 1$ for all $i\in\{1,\dots,k_1\}$ and $j\in\{1,\dots,k_1\}$. Similarly, we have $\hat{y}_{ij} = \frac{1}{2}$ for all $i\in\{1,\dots,k_1\}$ and

$j\in\{k_1+1,\ldots,k_2\}$. Lastly, we have $\hat{y}_{ij}=\frac{1}{2}$ for all $i\in\{k_1+1,\ldots,k_2\}$ and $j\in\{1,\ldots,k_1\}$. For the other cases not considered by the preceding three conditions, we have $\hat{y}_{ij}=0$. Thus, we get

$$\begin{aligned}\sum_{(i,j)\in\hat{M}}\mathbf{1}(i<j)\,\theta_i\,\hat{y}_{ij} &= \sum_{(i,j)\in\hat{M}}\mathbf{1}(i<j\le k_1)\,\theta_i+\sum_{(i,j)\in\hat{M}}\mathbf{1}(i\le k_1<j\le k_2)\,\theta_i\,\frac{1}{2}\\ &=\sum_{i\in\hat{N}}\theta_i\sum_{j\in\hat{N}}\mathbf{1}(i<j\le k_1)+\frac{1}{2}\sum_{i\in\hat{N}}\theta_i\sum_{j\in\hat{N}}\mathbf{1}(i\le k_1<j\le k_2)\\ &=\sum_{i\in\hat{N}}\theta_i\,\mathbf{1}(i\le k_1)\,(k_1-i)+\frac{1}{2}\sum_{i\in\hat{N}}\theta_i\,\mathbf{1}(i\le k_1)\,(k_2-k_1)\\ &=\sum_{i\in\hat{N}}\mathbf{1}(i\le k_1)\left\{\frac{1}{2}\,k_1+\frac{1}{2}\,k_2-i\right\}\theta_i=\sum_{i=1}^{k_1}\left\{\frac{1}{2}\,k_1+\frac{1}{2}\,k_2-i\right\}\theta_i.\end{aligned}$$

By the same computation in the chain of equalities above, we also have $\sum_{(i,j)\in\hat{M}}\mathbf{1}(i>j)\,\theta_j\,\hat{y}_{ij}=\sum_{j=1}^{k_1}\left(\frac{k_1}{2}+\frac{k_2}{2}-j\right)\theta_j$. Therefore, we obtain

$$\begin{aligned}f^R &\ge -\sum_{(i,j)\in\hat{M}}\max\{\theta_i,\theta_j\}\,\hat{y}_{ij}+2\,(n-1)\sum_{i\in\hat{N}}\theta_i\,\hat{x}_i\\ &=-\sum_{i=1}^{k_1}\left\{\frac{1}{2}\,k_1+\frac{1}{2}\,k_2-i\right\}\theta_i-\sum_{j=1}^{k_1}\left\{\frac{1}{2}\,k_1+\frac{1}{2}\,k_2-j\right\}\theta_j+2\,(n-1)\sum_{i=1}^{k_1}\theta_i+(n-1)\sum_{i=k_1+1}^{k_2}\theta_i\\ &=\sum_{i=1}^{k_1}(2\,n-k_1-k_2+2\,i-2)\,\theta_i+\sum_{i=k_1+1}^{k_2}(n-1)\,\theta_i.\end{aligned}$$

The inequality above holds for all choices of k_1 and k_2 such that $1\le k_1\le k_2\le m$. In this case, the desired follows by taking the maximum of the expression on the right side above over all (k_1,k_2) that satisfies $1\le k_1\le k_2\le m$. □

Viewing the objective function of the maximization problem in Lemma H.3 as a function of $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_m)$, k_1, k_2 and m, this objective function corresponds to the function $G(\boldsymbol{\theta},k_1,k_2,m)$ in Step 2. Thus, Lemma H.3 establishes Step 2. In the proof of Lemma H.3, we construct a feasible solution $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})\in\mathbb{R}_+^{|\hat{N}|}\times\mathbb{R}_+^{|\hat{M}|}$ to problem (8) by setting $\hat{x}_i=1$ for all $i\in\{1,\ldots,k_1\}$, $x_i=\frac{1}{2}$ for all $i\in\{k_1+1,\ldots,k_2\}$ and $\hat{x}_i=0$ for all $i\in\{k_2+1,\ldots,m\}$ and $\hat{y}_{ij}=[\hat{x}_i+\hat{x}_j-1]^+$ for all $(i,j)\in\hat{M}$. Our choice of this solution is motivated by the fact that if we maximize the function $-\sum_{(i,j)\in\hat{M}}\max\{\theta_i,\theta_j\}\,\hat{y}_{ij}+2\,(n-1)\sum_{i\in\hat{N}}\theta_i\,\hat{x}_i$ over the feasible set of problem (8), then there exists an optimal solution to problem (8) of this form for some choices of k_1 and k_2. Our development does not require showing this result explicitly, so we do not dwell on it further.

H.2. Removing Dependence on Product Revenues and Preference Weights

In this section, we establish Step 3. In Lemma H.1, $\sum_{i=1}^m(m-i)\,\theta_i$ is a function of $(\theta_1,\ldots,\theta_m)$. Similarly, the optimal objective value of the maximization problem on the right side of the inequality in Lemma H.3 also depends on $(\theta_1,\ldots,\theta_m)$. Next, we remove the dependence of these bounds on $(\theta_1,\ldots,\theta_m)$. In particular, by Lemma H.3, we have

$$\begin{aligned}
\frac{\sum_{i=1}^m (m-i)\,\theta_i}{f^R} \;&\leq\; \frac{\sum_{i=1}^m (m-i)\,\theta_i}{\max\limits_{\substack{(k_1,k_2):\\ 1\le k_1\le k_2\le m}} \left\{ \sum_{i=1}^{k_1} (2\,n-k_1-k_2+2\,i-2)\,\theta_i + \sum_{i=k_1+1}^{k_2} (n-1)\,\theta_i \right\}} \\
&= \min_{\substack{(k_1,k_2):\\ 1\le k_1\le k_2\le m}} \left\{ \frac{\sum_{i=1}^m (m-i)\,\theta_i}{\sum_{i=1}^{k_1} (2\,n-k_1-k_2+2\,i-2)\,\theta_i + \sum_{i=k_1+1}^{k_2} (n-1)\,\theta_i} \right\} \\
&\leq \min_{\substack{(k_1,k_2):\\ 1\le k_1\le k_2\le m}} \left\{ \frac{\sum_{i=1}^m (m-i)\,\theta_i}{\sum_{i=1}^{k_1} (2\,m-k_1-k_2+2\,i-2)\,\theta_i + \sum_{i=k_1+1}^{k_2} (m-1)\,\theta_i} \right\} \\
&\leq \min_{\substack{(k_1,k_2):\\ 1\le k_1\le k_2\le m}} \left\{ \max_{\substack{(\theta_1,\ldots,\theta_m):\\ \theta_1\ge\ldots\ge\theta_m\ge 0}} \left\{ \frac{\sum_{i=1}^m (m-i)\,\theta_i}{\sum_{i=1}^{k_1} (2\,m-k_1-k_2+2\,i-2)\,\theta_i + \sum_{i=k_1+1}^{k_2} (m-1)\,\theta_i} \right\} \right\}, \qquad (9)
\end{aligned}$$

where the second inequality is by the fact that $n\ge m$ and $2\,m-k_1-k_2+2\,i-2\ge 0$ whenever $k_1\le k_2\le m$. There are two features of the maximization problem on the right side of (9). First, if $(\theta_1^*,\ldots,\theta_m^*)$ is an optimal solution to this problem, then $(\alpha\,\theta_1^*,\ldots,\alpha\,\theta_m^*)$ is also an optimal solution for any $\alpha>0$. Thus, we can assume that $\theta_1\le 1$. Second, the objective function of the maximization problem on the right side above is quasi-linear. Thus, an optimal solution occurs at an extreme point of the polyhedron $\{(\theta_1,\ldots,\theta_m)\in\mathbb{R}^m : 1\ge\theta_1\ge\theta_2\ge\ldots\ge\theta_m\ge 0\}$. It is simple to check that an extreme point $(\hat\theta_1,\ldots,\hat\theta_m)$ of this polyhedron is of the form $\hat\theta_i=1$ for all $i=1,\ldots,\ell$ and $\hat\theta_i=0$ for all $i=\ell+1,\ldots,m$ for some $\ell\in\{0,\ldots,m\}$. In particular, if we have $0<\theta_i<1$ for some $i\in\{0,\ldots,m\}$, then we can express $(\hat\theta_1,\ldots,\hat\theta_m)$ as a convex combination of two points in the polyhedron. This argument shows that an optimal solution $(\hat\theta_1,\ldots,\hat\theta_m)$ to the maximization problem is of the form $\hat\theta_i=1$ for all $i=1,\ldots,\ell$ and $\hat\theta_i=0$ for all $i=\ell+1,\ldots,m$ for some $\ell\in\{0,\ldots,m\}$. Building on these observations, we give an upper bound on the optimal objective value of the maximization problem on the right side of (9). Throughout the rest of our discussion, we will use the functions

$$\begin{aligned}
\Gamma_1(k_1,k_2,m) \;&=\; \frac{m\,(m-1)/2}{m\,k_1+m\,k_2-k_1\,k_2-k_2}, \qquad (10)\\
\Gamma_2(k_1,k_2,m) \;&=\; \frac{m-1}{2\,m-k_1-k_2},\\
\Gamma_3(k_1,k_2,m) \;&=\; \max_{q\in\mathbb{R}_+} \left\{ \frac{m\,q-q\,(q+1)/2}{m\,k_1-k_1\,k_2+m\,q-q} \right\}.
\end{aligned}$$

In the next lemma, we use $\Gamma_1(k_1,k_2,m)$, $\Gamma_2(k_1,k_2,m)$ and $\Gamma_3(k_1,k_2,m)$ to give an upper bound on the optimal objective value of the maximization problem on the right side of (9).

Lemma H.4 *We have*

$$\max_{\substack{(\theta_1,\ldots,\theta_m):\\ \theta_1\geq\ldots\geq\theta_m\geq 0}} \left\{ \frac{\sum_{i=1}^m (m-i)\,\theta_i}{\sum_{i=1}^{k_1}(2\,m-k_1-k_2+2\,i-2)\,\theta_i + \sum_{i=k_1+1}^{k_2}(m-1)\,\theta_i} \right\} \\ \leq \; \Gamma_1(k_1,k_2,m) \vee \Gamma_2(k_1,k_2,m) \vee \Gamma_3(k_1,k_2,m). \quad (11)$$

Proof: By the discussion right before the lemma, there exists an optimal solution $(\hat{\theta}_1,\ldots,\hat{\theta}_m)$ to the maximization problem in (11) such that $\hat{\theta}_i = 1$ for all $i = 1,\ldots,\ell$ and $\hat{\theta}_i = 0$ for all $i = \ell+1,\ldots,m$ for some $\ell \in \{0,\ldots,m\}$. The denominator of the objective function of the maximization problem in (11) does not depend on $\{\theta_i : i = k_2+1,\ldots,m\}$, which implies that if $\ell > k_2$, then we can assume that $\ell = m$. In particular, if $\ell > k_2$, then setting $\ell = m$ increases the nominator without changing the denominator. So, we assume that $\ell \leq k_2$ or $\ell = m$. If $\ell = m$ so that $\hat{\theta}_1 = \ldots = \hat{\theta}_m = 1$, then the objective value of the maximization problem in (11) at the optimal solution $(\hat{\theta}_1,\ldots,\hat{\theta}_m)$ is

$$\frac{\sum_{i=1}^m (m-i)}{\sum_{i=1}^{k_1}(2\,m-k_1-k_2+2\,i-2) + \sum_{i=k_1+1}^{k_2}(m-1)} = \frac{m\,(m-1)/2}{k_1\,(2\,m-k_1-k_2)+k_1\,(k_1-1)+(k_2-k_1)\,(m-1)} \\ = \frac{m\,(m-1)/2}{m\,k_1+m\,k_2-k_1\,k_2-k_2} \;=\; \Gamma_1(k_1,k_2,m). \quad (12)$$

On the other hand, if $\ell \leq k_1$ so that $\hat{\theta}_1 = \ldots = \hat{\theta}_\ell = 1$ and $\hat{\theta}_{\ell+1} = \ldots \hat{\theta}_m = 0$, then the objective value of the maximization problem in (11) is

$$\frac{\sum_{i=1}^{\ell}(m-i)}{\sum_{i=1}^{\ell}(2\,m-k_1-k_2+2\,i-2)\,\theta_i} \;=\; \frac{m\,\ell-\ell\,(\ell+1)/2}{\ell\,(2\,m-k_1-k_2)+\ell\,(\ell-1)} \;=\; \frac{m-(\ell+1)/2}{2\,m-k_1-k_2+\ell-1},$$

which is decreasing in ℓ. Therefore, if we maximize the expression above over all ℓ satisfying $1 \leq \ell \leq k_1$, then the maximizer occurs at $\ell = 1$. Thus, we get

$$\frac{\sum_{i=1}^{\ell}(m-i)}{\sum_{i=1}^{\ell}(2\,m-k_1-k_2+2\,i-2)\,\theta_i} \;\leq\; \frac{m-1}{2\,m-k_1-k_2} \;=\; \Gamma_2(k_1,k_2,m). \quad (13)$$

Finally, if $k_1+1 \leq \ell \leq k_2$ so that $\hat{\theta}_1 = \ldots = \hat{\theta}_{k_1} = \hat{\theta}_{k_1+1} = \ldots = \hat{\theta}_\ell = 1$ and $\hat{\theta}_{\ell+1} = \ldots \hat{\theta}_m = 0$, then the objective value of the maximization problem in (11) is

$$\frac{\sum_{i=1}^{\ell}(m-i)}{\sum_{i=1}^{k_1}(2\,m-k_1-k_2+2\,i-2) + \sum_{i=k_1+1}^{\ell}(m-1)} = \frac{m\,\ell-\ell\,(\ell+1)/2}{k_1\,(2\,m-k_1-k_2)+k_1\,(k_1-1)+(\ell-k_1)\,(m-1)} \\ = \frac{m\,\ell-\ell\,(\ell+1)/2}{m\,k_1-k_1\,k_2+m\,\ell-\ell} \;\leq\; \Gamma_3(k_1,k_2,m). \quad (14)$$

Putting (12), (13) and (14) together, the optimal objective value of the maximization problem in (11) is no larger than $\Gamma_1(k_1,k_2,m) \vee \Gamma_2(k_1,k_2,m) \vee \Gamma_3(k_1,k_2,m)$. □

In the objective function of the maximization problem in (11), recalling that $F(\boldsymbol{\theta})$ corresponds to the function in the numerator and $G(\boldsymbol{\theta},k_1,k_2,m)$ corresponds to the function in the denominator, Lemma H.4 establishes Step 3. Note that $\Gamma_1(k_1,k_2,m)$ and $\Gamma_2(k_1,k_2,m)$ in (10) have explicit expressions. Next, we give an explicit expression for $\Gamma_3(k_1,k_2,m)$ as well. Since we are interested in the values of k_2 satisfying $k_2 \leq m$, the denominator of the fraction in the definition of $\Gamma_3(k_1,k_2,m)$ is non-negative. Furthermore, the numerator in this fraction is concave in q. Therefore, the objective function of the maximization problem in the definition of $\Gamma_3(k_1,k_2,m)$ is quasi-concave, which implies that we can use the first order condition to characterize the optimal solution to this problem. In particular, differentiating the fraction in the maximization problem in the definition of $\Gamma_3(k_1,k_2,m)$ with respect to q, the first order condition is

$$\frac{(m-q-\frac{1}{2})\,(m\,k_1-k_1\,k_2+m\,q-q)-(m\,q-q\,(q+1)/2)\,(m-1)}{(m\,k_1-k_1\,k_2+m\,q-q)^2} = \frac{-(m-1)\,q^2/2-k_1\,(m-k_2)\,q+k_1\,(m-k_2)\,(m-\frac{1}{2})}{(m\,k_1-k_1\,k_2+m\,q-q)^2}=0.$$

There is only one positive solution to the second equality above. Using $q(k_1,k_2,m)$ to denote this positive solution, we have

$$q(k_1,k_2,m)=\frac{\sqrt{k_1^2\,(m-k_2)^2+2\,k_1\,(m-k_2)\,(m-1)\,(m-\frac{1}{2})}-k_1\,(m-k_2)}{m-1}. \tag{15}$$

To obtain an explicit expression for $\Gamma_3(k_1,k_2,m)$, consider the function $h(q)=f(q)/g(q)$. Assume that the derivative of $h(q)$ at $\hat{q}$ is zero. In other words, using $f'(\hat{q})$ and $q'(\hat{q})$ to, respectively, denote the derivatives of $f(\cdot)$ and $g(\cdot)$ evaluated at $\hat{q}$, we have $f'(\hat{q})\,g(\hat{q})-f(\hat{q})\,g'(\hat{q})=0$. In this case, we obtain $f(\hat{q})/g(\hat{q})=f'(\hat{q})/g'(\hat{q})$, which implies that $h(\hat{q})=f(\hat{q})/g(\hat{q})=f'(\hat{q})/g'(\hat{q})$. To use this observation, we note that $\Gamma_3(k_1,k_2,m)$ is given by the value of the fraction in the definition of $\Gamma_3(k_1,k_2,m)$ evaluated at $q(k_1,k_2,m)$. Furthermore, the derivative of this fraction with respect to q evaluated at $q(k_1,k_2,m)$ is zero. Since the derivative of the numerator and denominator of this fraction with respect to q are, respectively, $m-q-\frac{1}{2}$ and $m-1$, it follows that

$$\Gamma_3(k_1,k_2,m)=\frac{m-q(k_1,k_2,m)-\frac{1}{2}}{m-1}=1+\frac{1}{2\,(m-1)}-\frac{q(k_1,k_2,m)}{m-1}, \tag{16}$$

which, noting (15), yields an explicit expression for $\Gamma_3(k_1,k_2,m)$. Therefore, we have explicit expressions for $\Gamma_1(k_1,k_2,m)$, $\Gamma_2(k_1,k_2,m)$ and $\Gamma_3(k_1,k_2,m)$.

H.3. Uniform Bounds

In this section, we establish Step 4. In particular, for any value of m, we show that there exist k_1 and k_2 satisfying $1 \leq k_1 \leq k_2 \leq m$, $\Gamma_1(k_1,k_2,m) \leq 0.8$, $\Gamma_2(k_1,k_2,m) \leq 0.8$ and $\Gamma_3(k_1,k_2,m) \leq 0.8$. Since we have explicit expressions for $\Gamma_1(k_1,k_2,m)$, $\Gamma_2(k_1,k_2,m)$ and $\Gamma_3(k_1,k_2,m)$, if m is small, then we can enumerate over all possible values of k_1 and k_2 that satisfy $1 \leq k_1 \leq k_2 \leq m$ to ensure that there exist k_1 and k_2 such that $\Gamma_1(k_1,k_2,m) \le 0.8$, $\Gamma_2(k_1,k_2,m) \le 0.8$ and $\Gamma_3(k_1,k_2,m) \le 0.8$. In particular, through complete enumeration, it is simple to numerically verify that if $m < 786$, then there exist k_1 and k_2 satisfying $1 \le k_1 \le k_2 \le m$, $\Gamma_1(k_1,k_2,m) \le 0.8$, $\Gamma_2(k_1,k_2,m) \le 0.8$ and $\Gamma_3(k_1,k_2,m) \le 0.8$. Thus, we only need to consider the case where $m \ge 786$. We begin with some intuition for our approach. Assume that we always choose k_1 and k_2 as a fixed fraction of m. In particular, we always choose k_1 and k_2 as $k_1 = \hat\beta_1 m$ and $k_2 = \hat\beta_2 m$ for some $\hat\beta_1 \in (0,1]$, $\hat\beta_2 \in (0,1]$ and $\hat\beta_1 \le \hat\beta_2$. Recall that we want to find some k_1 and k_2 satisfying $1 \le k_1 \le k_2 \le m$, $\Gamma_1(k_1,k_2,m) \le 0.8$, $\Gamma_2(k_1,k_2,m) \le 0.8$ and $\Gamma_3(k_1,k_2,m) \le 0.8$. Thus, there is no harm in trying to choose $k_1 = \hat\beta_1 m$ and $k_2 = \hat\beta_2 m$. Naturally, k_1 and k_2 need to be integers and we shortly address this issue. By (10) and (16), if we choose k_1 and k_2 as $k_1 = \hat\beta_1 m$ and $k_2 = \hat\beta_2 m$, then we have

$$\Gamma_1(\hat\beta_1 m, \hat\beta_2 m, m) = \frac{m\,(m-1)/2}{\hat\beta_1 m^2 + \hat\beta_2 m^2 - \hat\beta_1 \hat\beta_2 m^2 - \hat\beta_2 m}, \qquad \Gamma_2(\hat\beta_1 m, \hat\beta_2 m, m) = \frac{m-1}{2m - \hat\beta_1 m - \hat\beta_2 m},$$

$$\Gamma_3(\hat\beta_1 m, \hat\beta_2 m, m) = 1 + \frac{1}{2\,(m-1)} + \frac{\hat\beta_1 m\,(m - \hat\beta_2 m)}{(m-1)^2} - \frac{\sqrt{\hat\beta_1^2 m^2 (m - \hat\beta_2 m)^2 + 2\,\hat\beta_1 m\,(m - \hat\beta_2 m)\,(m-1)\,(m - \frac{1}{2})}}{(m-1)^2}.$$

We let $\gamma_1(\hat\beta_1,\hat\beta_2) = \lim_{m\to\infty} \Gamma_1(\hat\beta_1 m, \hat\beta_2 m, m)$, $\gamma_2(\hat\beta_1,\hat\beta_2) = \lim_{m\to\infty} \Gamma_2(\hat\beta_1 m, \hat\beta_2 m, m)$ and $\gamma_3(\hat\beta_1,\hat\beta_2) = \lim_{m\to\infty} \Gamma_3(\hat\beta_1 m, \hat\beta_2 m, m)$. Thus, taking limits in the expressions above, we get

$$\gamma_1(\hat\beta_1,\hat\beta_2) = \frac{1}{2\,(\hat\beta_1 + \hat\beta_2 - \hat\beta_1 \hat\beta_2)}, \qquad \gamma_2(\hat\beta_1,\hat\beta_2) = \frac{1}{2 - \hat\beta_1 - \hat\beta_2},$$

$$\gamma_3(\hat\beta_1,\hat\beta_2) = 1 + \hat\beta_1\,(1 - \hat\beta_2) - \sqrt{\hat\beta_1^2\,(1 - \hat\beta_2^2) + 2\,\hat\beta_1\,(1 - \hat\beta_2)}.$$

Roughly speaking, if m is large and we choose k_1 and k_2 as $k_1 = \hat\beta_1 m$ and $k_2 = \hat\beta_2 m$, then $\Gamma_1(k_1,k_2,m) \vee \Gamma_2(k_1,k_2,m) \vee \Gamma_3(k_1,k_2,m)$ behaves similarly to $\gamma_1(\hat\beta_1,\hat\beta_2) \vee \gamma_2(\hat\beta_1,\hat\beta_2) \vee \gamma_3(\hat\beta_1,\hat\beta_2)$. We want to find some k_1 and k_2 with $1 \le k_1 \le k_2 \le m$ such that $\Gamma_1(k_1,k_2,m) \vee \Gamma_2(k_1,k_2,m) \vee \Gamma_3(k_1,k_2,m) \le 0.8$. Using $\gamma_1(\hat\beta_1,\hat\beta_2) \vee \gamma_2(\hat\beta_1,\hat\beta_2) \vee \gamma_3(\hat\beta_1,\hat\beta_2)$ as an approximation to $\Gamma_1(k_1,k_2,m) \vee \Gamma_2(k_1,k_2,m) \vee \Gamma_3(k_1,k_2,m)$, we choose $\hat\beta_1$ and $\hat\beta_2$ to ensure that $\gamma_1(\hat\beta_1,\hat\beta_2) \vee \gamma_2(\hat\beta_1,\hat\beta_2) \vee \gamma_3(\hat\beta_1,\hat\beta_2)$ is as small as possible. In particular, we choose $\hat\beta_1$ and $\hat\beta_2$ as the solution to the system of equations $\gamma_1(\hat\beta_1,\hat\beta_2) = \gamma_2(\hat\beta_1,\hat\beta_2)$ and $\gamma_2(\hat\beta_1,\hat\beta_2) = \gamma_3(\hat\beta_1,\hat\beta_2)$. Solving this system of equations numerically, we obtain $\hat\beta_1 \approx 0.088302$ and $\hat\beta_2 \approx 0.614542$, yielding $\gamma_1(\hat\beta_1,\hat\beta_2) \vee \gamma_2(\hat\beta_1,\hat\beta_2) \vee \gamma_3(\hat\beta_1,\hat\beta_2) \approx 0.770917$. Since we need to ensure that $k_1 \le k_2$, we also need to ensure that $\hat\beta_1 \le \hat\beta_2$. Fortunately, the solution to the system of equations $\gamma_1(\hat\beta_1,\hat\beta_2) = \gamma_2(\hat\beta_1,\hat\beta_2)$ and $\gamma_2(\hat\beta_1,\hat\beta_2) = \gamma_3(\hat\beta_1,\hat\beta_2)$ already satisfies this requirement. Also, we do not need the precise solution to the last system of equations, since our goal is to find some upper bound on $\gamma_1(\hat\beta_1,\hat\beta_2) \vee \gamma_2(\hat\beta_1,\hat\beta_2) \vee \gamma_3(\hat\beta_1,\hat\beta_2)$. An imprecise solution simply yields a slightly looser upper bound. Lastly, since the best upper bound we can find on

$\gamma_1(\hat\beta_1,\hat\beta_2)\vee\gamma_2(\hat\beta_1,\hat\beta_2)\vee\gamma_3(\hat\beta_1,\hat\beta_2)$ is roughly equal to 0.8, we will be able to show that there exist k_1 and k_2 with $1\le k_1\le k_2\le m$ and $\Gamma_1(k_1,k_2,m)\vee\Gamma_2(k_1,k_2,m)\vee\Gamma_3(k_1,k_2,m)\le 0.8$.

The preceding discussion provides some intuition, but it is not precise. We need an upper bound on $\Gamma_1(k_1,k_2,m)\vee\Gamma_2(k_1,k_2,m)\vee\Gamma_3(k_1,k_2,m)$, not on $\gamma_1(\hat\beta_1,\hat\beta_2)\vee\gamma_2(\hat\beta_1,\hat\beta_2)\vee\gamma_3(\hat\beta_1,\hat\beta_2)$. These two quantities are different for finite values of m. Also, k_1 and k_2 need to be integers, but choosing $k_1=\hat\beta_1\,m$ and $k_2=\hat\beta_2\,m$ does not necessarily provide integer values for k_1 and k_2. To address these issues, we choose k_1 and k_2 as $k_1=\lceil\hat\beta_1\,m\rceil$ and $k_2=\lceil\hat\beta_1\,m\rceil$. In this case, setting $\hat\beta_1=0.088302$ and $\hat\beta_2=0.614542$, we proceed to showing that $\Gamma_1(\lceil\hat\beta_1\,m\rceil,\lceil\hat\beta_2\,m\rceil,m)\le 0.8$, $\Gamma_2(\lceil\hat\beta_1\,m\rceil,\lceil\hat\beta_2\,m\rceil,m)\le 0.8$ and $\Gamma_3(\lceil\hat\beta_1\,m\rceil,\lceil\hat\beta_2\,m\rceil,m)\le 0.8$, as long as $m\ge 786$. Therefore, for any value of $m\ge 786$, there exist values of k_1 and k_2 satisfying $1\le k_1\le k_2\le m$, $\Gamma_1(k_1,k_2,m)\le 0.8$, $\Gamma_2(k_1,k_2,m)\le 0.8$ and $\Gamma_3(k_1,k_2,m)\le 0.8$, which establishes Step 4. Throughout this section, we fix $\hat\beta_1=0.088302$ and $\hat\beta_2=0.614542$. In the next lemma, we give a bound on $\Gamma_1(\lceil\hat\beta_1\,m\rceil,\lceil\hat\beta_2\,m\rceil,m)$.

Lemma H.5 *If $m\ge 786$, then we have $\Gamma_1(\lceil\hat\beta_1\,m\rceil,\lceil\hat\beta_2\,m\rceil,m)\le 0.8$.*

Proof: We have $\hat\beta_1\,m\le\lceil\hat\beta_1\,m\rceil\le\hat\beta_1\,m+1$ and $\hat\beta_2\,m\le\lceil\hat\beta_2\,m\rceil\le\hat\beta_2\,m+1$. In this case, noting the definition of $\Gamma_1(k_1,k_2,m)$, it follows that

$$\begin{aligned}\Gamma_1(\lceil\hat\beta_1\,m\rceil,\lceil\hat\beta_2\,m\rceil,m)&=\frac{m\,(m-1)/2}{m\,\lceil\hat\beta_1\,m\rceil+m\,\lceil\hat\beta_2\,m\rceil-\lceil\hat\beta_1\,m\rceil\,\lceil\hat\beta_2\,m\rceil-\lceil\hat\beta_2\,m\rceil}\\&\le\frac{m^2/2}{\hat\beta_1\,m^2+\hat\beta_2\,m^2-(\hat\beta_1\,m+1)(\hat\beta_2\,m+1)-\hat\beta_2\,m-1}\le\frac{\frac{1}{2}}{\hat\beta_1+\hat\beta_2-\hat\beta_1\,\hat\beta_2-(\hat\beta_1+2\,\hat\beta_2+2)/m}.\end{aligned}$$

The expression on the right side above is decreasing in m. Computing this expression with $\hat\beta_1=0.088302$, $\hat\beta_2=0.614542$ and $m=786$, we get a value that does not exceed 0.78. Therefore, we have $\Gamma_1(\lceil\hat\beta_1\,m\rceil,\lceil\hat\beta_2\,m\rceil,m)\le 0.8$ for all $m\ge 786$. □

In the proof of Lemma H.5, we can check that $\Gamma_1(\lceil\hat\beta_1\,m\rceil,\lceil\hat\beta_2\,m\rceil,m)\le 0.8$ for all $m\ge 141$, but we need to impose a lower bound of 786 on m anyway when dealing with $\Gamma_3(\lceil\hat\beta_1\,m\rceil,\lceil\hat\beta_2\,m\rceil,m)$ shortly. In the next lemma, we give an upper bound on $\Gamma_2(\lceil\hat\beta_1\,m\rceil,\lceil\hat\beta_2\,m\rceil,m)$.

Lemma H.6 *If $m\ge 786$, then we have $\Gamma_2(\lceil\hat\beta_1\,m\rceil,\lceil\hat\beta_2\,m\rceil,m)\le 0.8$.*

Proof: Similar to our approach in the proof of Lemma H.5, we have $\hat\beta_1\,m\le\lceil\hat\beta_1\,m\rceil\le\hat\beta_1\,m+1$ and $\hat\beta_2\,m\le\lceil\hat\beta_2\,m\rceil\le\hat\beta_2\,m+1$. Noting the definition of $\Gamma_2(k_1,k_2,m)$, it follows that

$$\Gamma_2(\lceil\hat\beta_1\,m\rceil,\lceil\hat\beta_2\,m\rceil,m)=\frac{m-1}{2\,m-\lceil\hat\beta_1\,m\rceil-\lceil\hat\beta_2\,m\rceil}\le\frac{m}{2\,m-(\hat\beta_1\,m+1)-(\hat\beta_2\,m+1)}=\frac{1}{2-\hat\beta_1-\hat\beta_2-2/m}.$$

The expression on the right side above is decreasing in m. If we compute this expression with $\hat\beta_1=0.088302$, $\hat\beta_2=0.614542$ and $m=786$, then we get a value that does not exceed 0.78. □

In the next lemma, we come up with an upper bound on $\Gamma_3(\lceil\hat\beta_1\,m\rceil,\lceil\hat\beta_2\,m\rceil,m)$.

Lemma H.7 *If $m \geq 786$, then we have $\Gamma_3(\lceil \hat{\beta}_1 m \rceil, \lceil \hat{\beta}_2 m \rceil, m) \leq 0.8$.*

Proof: We begin by providing bounds for several quantities. These bounds become useful later in the proof. For $m \geq 786$, we bound $\lceil \hat{\beta}_1 m \rceil/(m-1)$ and $(m - \lceil \hat{\beta}_2 m \rceil)/(m-1)$ as

$$\hat{\beta}_1 \leq \frac{\lceil \hat{\beta}_1 m \rceil}{m-1} \leq \hat{\beta}_1 + \frac{2}{m} \tag{17}$$

$$1 - \hat{\beta}_2 - \frac{1}{m} \leq \frac{m - \lceil \hat{\beta}_2 m \rceil}{m-1} \leq 1 - \hat{\beta}_2 + \frac{1}{m}. \tag{18}$$

In particular, we have $\lceil \hat{\beta}_1 m \rceil/(m-1) \leq (\hat{\beta}_1 m + 1)/(m-1) = \hat{\beta}_1 + (\hat{\beta}_1 + 1)/(m-1) \leq \hat{\beta}_1 + 2/m$, where the last inequality uses the fact that $\hat{\beta}_1 = 0.088302$ so that we have $(\hat{\beta}_1 + 1)/(m-1) \leq 2/m$ for all $m \geq 3$, but we already assume that $m \geq 786$. Also, we have $\lceil \hat{\beta}_1 m \rceil/(m-1) \geq \hat{\beta}_1 m/(m-1) \geq \hat{\beta}_1$. Therefore, the chain of inequalities in (17) holds. On the other hand, we have $(m - \lceil \hat{\beta}_2 m \rceil)/(m-1) \leq (m - \hat{\beta}_2 m)/(m-1) = 1 - \hat{\beta}_2 + (1 - \hat{\beta}_2)/(m-1) \leq 1 - \hat{\beta}_2 + 1/m$, where the last inequality uses the fact that $\hat{\beta}_2 = 0.614542$ so that we have $(1 - \hat{\beta}_2)/(m-1) \leq 1/m$ for all $m \geq 2$, but once again, we already assume that $m \geq 786$. Also, we have $(m - \lceil \hat{\beta}_2 m \rceil)/(m-1) \geq (m - \hat{\beta}_2 m - 1)/(m-1) = 1 - \hat{\beta}_2 - \hat{\beta}_2/(m-1) \geq 1 - \hat{\beta}_2 - 1/m$, where the last inequality uses the fact that $\hat{\beta}_2/(m-1) \leq 1/m$ for all $m \geq 3$. Therefore, the chain of inequalities in (18) holds as well. Next, we define the function $\Lambda(k_1, k_2, m)$ and the constant $\lambda(\hat{\beta}_1, \hat{\beta}_2)$ as

$$\Lambda(k_1, k_2, m) = \frac{k_1^2 (m - k_2)^2 + 2 k_1 (m - k_2)(m-1)(m - \frac{1}{2})}{(m-1)^4}$$

$$\lambda(\hat{\beta}_1, \hat{\beta}_2) = \hat{\beta}_1^2 (1 - \hat{\beta}_2)^2 + 2 \hat{\beta}_1 (1 - \hat{\beta}_2).$$

Relating $\Lambda(\lceil \hat{\beta}_1 m \rceil, \lceil \hat{\beta}_2 m \rceil, m)$ to $\lambda(\hat{\beta}_1, \hat{\beta}_2)$, we will relate $\Gamma_3(\lceil \hat{\beta}_1 m \rceil, \lceil \hat{\beta}_2 m \rceil, m)$ to $\gamma_3(\hat{\beta}_1, \hat{\beta}_2)$. We claim that $\sqrt{\Lambda(\lceil \hat{\beta}_1 m \rceil, \lceil \hat{\beta}_2 m \rceil, m)} \geq \sqrt{\lambda(\hat{\beta}_1, \hat{\beta}_2)} - 8/m$. To see the claim, note that

$$\frac{\lceil \hat{\beta}_1 m \rceil^2 (m - \lceil \hat{\beta}_2 m \rceil)^2}{(m-1)^4} \geq \hat{\beta}_1^2 \left(1 - \hat{\beta}_2 - \frac{1}{m}\right)^2 = \hat{\beta}_1^2 \left((1 - \hat{\beta}_2)^2 - \frac{2}{m}(1 - \hat{\beta}_2) + \frac{1}{m^2}\right)$$
$$\geq \hat{\beta}_1^2 (1 - \hat{\beta}_2)^2 - \frac{2}{m} \hat{\beta}_1^2 (1 - \hat{\beta}_2) \geq \hat{\beta}_1^2 (1 - \hat{\beta}_2)^2 - \frac{1}{m},$$

where the first inequality uses (17) and (18), whereas the third inequality uses the fact that $\hat{\beta}_1^2 (1 - \hat{\beta}_2)^2 \leq \frac{1}{2}$. Furthermore, we have

$$\frac{\lceil \hat{\beta}_1 m \rceil (m - \lceil \hat{\beta}_2 m \rceil)(m-1)(m - \frac{1}{2})}{(m-1)^4} \geq \hat{\beta}_1 \left(1 - \hat{\beta}_2 - \frac{1}{m}\right) \left(\frac{m - \frac{1}{2}}{m-1}\right)$$
$$\geq \hat{\beta}_1 \left(1 - \hat{\beta}_2 - \frac{1}{m}\right) \geq \hat{\beta}_1 (1 - \hat{\beta}_2) - \frac{1}{m},$$

where the first inequality uses (17) and (18) and the third inequality uses the fact that $\hat{\beta}_1 \leq 1$. Multiplying the inequality above by two, adding the last two inequalities and noting the definition of

$\Lambda(k_1,k_2,m)$, we get $\Lambda(\lceil\hat\beta_1 m\rceil,\lceil\hat\beta_2 m\rceil,m)\ge\lambda(\hat\beta_1,\hat\beta_2)-3/m$. Since $\hat\beta_1=0.088302$ and $\hat\beta_2=0.614542$, we can compute the value of $\lambda(\hat\beta_1,\hat\beta_2)$ to check that $\sqrt{\lambda(\hat\beta_1,\hat\beta_2)}\ge 1/4$. Therefore, for all $m\ge 64$, we have $16\sqrt{\lambda(\hat\beta_1,\hat\beta_2)}-64/m\ge 4-64/m\ge 3$, in which case, we obtain

$$\Lambda(\lceil\hat\beta_1 m\rceil,\lceil\hat\beta_2 m\rceil,m)\ \ge\ \lambda(\hat\beta_1,\hat\beta_2)-\frac{3}{m}\ \ge\ \lambda(\hat\beta_1,\hat\beta_2)-\frac{16}{m}\sqrt{\lambda(\hat\beta_1,\hat\beta_2)}+\frac{64}{m^2}\ =\ \left(\sqrt{\lambda(\hat\beta_1,\hat\beta_2)}-\frac{8}{m}\right)^2.$$

Taking the square root above, we obtain $\sqrt{\Lambda(\lceil\hat\beta_1 m\rceil,\lceil\hat\beta_2 m\rceil,m)}\ge\sqrt{\lambda(\hat\beta_1,\hat\beta_2)}-8/m$, which establishes the desired claim. Noting (15) and the definition of $\Lambda(k_1,k_2,m)$, we have

$$\begin{aligned}
\frac{q(\lceil\hat\beta_1 m\rceil,\lceil\hat\beta_2 m\rceil,m)}{m-1} &= \sqrt{\Lambda(\lceil\hat\beta_1 m\rceil,\lceil\hat\beta_2 m\rceil,m)}-\frac{\lceil\hat\beta_1 m\rceil\,(m-\lceil\hat\beta_2 m\rceil)}{(m-1)^2}\\
&\ge \sqrt{\Lambda(\lceil\hat\beta_1 m\rceil,\lceil\hat\beta_2 m\rceil,m)}-\left(\hat\beta_1+\frac{2}{m}\right)\left(1-\hat\beta_2+\frac{1}{m}\right)\\
&\ge \sqrt{\lambda(\hat\beta_1,\hat\beta_2)}-\frac{8}{m}-\left(\hat\beta_1+\frac{2}{m}\right)\left(1-\hat\beta_2+\frac{1}{m}\right)\\
&= \sqrt{\lambda(\hat\beta_1,\hat\beta_2)}-\hat\beta_1\,(1-\hat\beta_2)-\frac{1}{m}\,(8+\hat\beta_1+2\,(1-\hat\beta_2))-\frac{2}{m^2}\\
&\ge \sqrt{\lambda(\hat\beta_1,\hat\beta_2)}-\hat\beta_1\,(1-\hat\beta_2)-\frac{10}{m},
\end{aligned}$$

where the first inequality uses (17) and (18), whereas the third inequality holds since $8+\hat\beta_1+2\,(1-\hat\beta_2)\le 9$ and $2/m^2\le 1/m$ for all $m\ge 2$. Using (16) and the inequality above, we get

$$\begin{aligned}
\Gamma_3(\lceil\hat\beta_1 m\rceil,\lceil\hat\beta_2 m\rceil,m) &= 1+\frac{1}{2\,(m-1)}-\frac{q(\lceil\hat\beta_1 m\rceil,\lceil\hat\beta_2 m\rceil,m)}{m-1}\\
&\le 1+\frac{1}{2\,(m-1)}-\sqrt{\lambda(\hat\beta_1,\hat\beta_2)}+\hat\beta_1\,(1-\hat\beta_2)+\frac{10}{m}\\
&\le 1-\sqrt{\lambda(\hat\beta_1,\hat\beta_2)}+\hat\beta_1\,(1-\hat\beta_2)+\frac{11}{m}\le 0.786+\frac{11}{m},
\end{aligned}$$

where the second inequality uses the fact that $1/(2\,(m-1))\le 1/m$ for all $m\ge 2$ and the third inequality follows from the fact that $\sqrt{\lambda(\hat\beta_1,\hat\beta_2)}\ge 1/4$ and $\hat\beta_1\le 0.09$ and $\hat\beta_2\ge 0.6$. The result follows by noting that $0.786+11/m\le 0.8$ for all $m\ge 786$. □

Putting Lemmas H.5, H.6 and H.7 together establishes Step 4.

Appendix I: Method of Conditional Expectations for the Uncapacitated Problem

Assume that we have a random subset of products $\hat{\boldsymbol{X}}=\{\hat X_i : i\in N\}$ that satisfies the inequality $\sum_{(i,j)\in M}\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}})-\hat z)\}\ge 0.6\,f^R$. In the method of condition expectations, we inductively construct a subset of products $\boldsymbol{x}^{(k)}=(\hat x_1,\ldots,\hat x_k,\hat X_{k+1},\ldots,\hat X_n)$ for all $k\in N$, where the first k products in this subset are deterministic and the last $n-k$ products are random variables. Each one of these subsets of products is constructed to ensure that we have $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{x}^{(k)})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^{(k)}) - \hat{z})\} \geq 0.6\, f^R$ for all $k \in N$. In this case, the subset of products $\boldsymbol{x}^{(n)} = (\hat{x}_1,\ldots,\hat{x}_n)$ is a deterministic subset of products that satisfies $\sum_{(i,j)\in M} V_{ij}(\boldsymbol{x}^{(n)})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^{(n)}) - \hat{z}) \geq 0.6\, f^R$, as desired. To inductively construct the subset of products $\boldsymbol{x}^{(k)} = (\hat{x}_1,\ldots,\hat{x}_k,\hat{X}_{k+1},\ldots,\hat{X}_n)$ for all $k\in N$, we start with $\boldsymbol{x}^{(0)} = \hat{\boldsymbol{X}}$. By Theorem 4.1, we have $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{x}^{(0)})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^{(0)}) - \hat{z})\} \geq 0.6\, f^R$. Assuming that we have a subset of products $\boldsymbol{x}^{(k)}$ that satisfies $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{x}^{(k)})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^{(k)}) - \hat{z})\} \geq 0.6\, f^R$, we show how to construct a subset of products $\boldsymbol{x}^{(k+1)}$ that satisfies $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{x}^{(k+1)})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^{(k+1)}) - \hat{z})\} \geq 0.6\, f^R$. By the induction assumption, we have $0.6\, f^R \leq \sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{x}^{(k)})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^{(k)}) - \hat{z})\}$. Conditioning on $\hat{X}_{k+1}$, we write the last inequality as

$$0.6\, f^R \;\leq\; \mathbb{P}\{\hat{X}_{k+1}=1\} \sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{x}^{(k)})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^{(k)}) - \hat{z}) \,|\, \hat{X}_{k+1}=1\} + \mathbb{P}\{\hat{X}_{k+1}=0\} \sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{x}^{(k)})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^{(k)}) - \hat{z}) \,|\, \hat{X}_{k+1}=0\}.$$

We define the two subsets of products as $\tilde{\boldsymbol{x}}^{(k)} = (\hat{x}_1,\ldots,\hat{x}_k,1,\hat{X}_{k+2},\ldots,\hat{X}_n)$ and as $\bar{\boldsymbol{x}}^{(k)} = (\hat{x}_1,\ldots,\hat{x}_k,0,\hat{X}_{k+2},\ldots,\hat{X}_n)$. By the definition of $\boldsymbol{x}^{(k)}$, given that $\hat{X}_{k+1} = 1$, we have $\boldsymbol{x}^{(k)} = \tilde{\boldsymbol{x}}^{(k)}$. Given that $\hat{X}_{k+1} = 0$, we have $\boldsymbol{x}^{(k)} = \bar{\boldsymbol{x}}^{(k)}$. So, we write the inequality above as

$$\begin{aligned} 0.6\, f^R &\leq \mathbb{P}\{\hat{X}_{k+1}=1\} \sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\tilde{\boldsymbol{x}}^{(k)})^{\gamma_{ij}}(R_{ij}(\tilde{\boldsymbol{x}}^{(k)}) - \hat{z})\} \\ &\qquad + \mathbb{P}\{\hat{X}_{k+1}=0\} \sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\bar{\boldsymbol{x}}^{(k)})^{\gamma_{ij}}(R_{ij}(\bar{\boldsymbol{x}}^{(k)}) - \hat{z})\} \\ &\leq \max\Big\{ \sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\tilde{\boldsymbol{x}}^{(k)})^{\gamma_{ij}}(R_{ij}(\tilde{\boldsymbol{x}}^{(k)}) - \hat{z})\},\; \sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\bar{\boldsymbol{x}}^{(k)})^{\gamma_{ij}}(R_{ij}(\bar{\boldsymbol{x}}^{(k)}) - \hat{z})\}\Big\}. \end{aligned}$$

Thus, either $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\tilde{\boldsymbol{x}}^{(k)})^{\gamma_{ij}}(R_{ij}(\tilde{\boldsymbol{x}}^{(k)}) - \hat{z})\}$ or $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\bar{\boldsymbol{x}}^{(k)})^{\gamma_{ij}}(R_{ij}(\bar{\boldsymbol{x}}^{(k)}) - \hat{z})\}$ is at least $0.6\, f^R$, indicating that we can use $\tilde{\boldsymbol{x}}^{(k)}$ or $\bar{\boldsymbol{x}}^{(k)}$ as $\boldsymbol{x}^{(k+1)}$. In both $\tilde{\boldsymbol{x}}^{(k)}$ and $\bar{\boldsymbol{x}}^{(k)}$, the first $k+1$ products are deterministic and the last $n-k-1$ products are random variables, as desired.

Considering the computational effort for the method of conditional expectations, we can compute $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{x}^{(0)})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^{(0)}) - \hat{z})\}$ in $O(n^2)$ operations. The subset of products $\tilde{\boldsymbol{x}}^{(k)}$ differs from the subset of products $\boldsymbol{x}^{(k)}$ only in product $k+1$, which implies that the quantity $\mathbb{E}\{V_{ij}(\tilde{\boldsymbol{x}}^{(k)})^{\gamma_{ij}}(R_{ij}(\tilde{\boldsymbol{x}}^{(k)}) - \hat{z})\}$ differs from $\mathbb{E}\{V_{ij}(\boldsymbol{x}^{(k)})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^{(k)}) - \hat{z})\}$ only for the nests that include product $k+1$. There are $O(n)$ such nests. Therefore, if we know the value of $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{x}^{(k)})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^{(k)}) - \hat{z})\}$, then we can compute $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\tilde{\boldsymbol{x}}^{(k)})^{\gamma_{ij}}(R_{ij}(\tilde{\boldsymbol{x}}^{(k)}) - \hat{z})\}$ in $O(n)$ operations. Similarly, if we know the value of $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{x}^{(k)})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^{(k)}) - \hat{z})\}$, then we can compute $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\bar{\boldsymbol{x}}^{(k)})^{\gamma_{ij}}(R_{ij}(\bar{\boldsymbol{x}}^{(k)}) - \hat{z})\}$ in $O(n)$ operations. Therefore, given the subset of products $\boldsymbol{x}^{(k)}$ and the value of $\sum_{(i,j)\in M}\mathbb{E}\{V_{ij}(\boldsymbol{x}^{(k)})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x}^{(k)})-\hat{z})\}$, we can construct the subset of products $\boldsymbol{x}^{(k+1)}$ in $O(n)$ operations. In the method of conditional expectations, we construct $O(n)$ subsets of products of the form $\boldsymbol{x}^{(k)}=(\hat{x}_1,\ldots,\hat{x}_k,\hat{X}_{k+1},\ldots,\hat{X}_n)$. Thus, the method of conditional expectations takes $O(n^2)$ operations.

Appendix J: Semidefinite Programming Relaxation

We describe an approximation algorithm for the uncapacitated problem that provides an α-approximate solution with $\alpha=\frac{2}{\pi}\min_{\theta\in[0,\arccos(-1/3)]}(2\pi-3\theta)/(1+3\cos\theta)\geq 0.79$. Our development generally follows the one in the main text. We develop an upper bound $f^R(\cdot)$ on $f(\cdot)$. This upper bound is based on an SDP relaxation of the Function Evaluation problem. Next, we show how to obtain a random subset of products $\hat{\boldsymbol{X}}$ such that $\sum_{(i,j)\in M}\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}})-\hat{z})\}\geq 0.79\,f^R(\hat{z})$, where $\hat{z}$ satisfies $f^R(\hat{z})=v_0\,\hat{z}$. We also show how to find the value of $\hat{z}$ that satisfies $f^R(\hat{z})=v_0\,\hat{z}$. Lastly, we discuss how to de-randomize the random subset of products $\hat{\boldsymbol{X}}$.

J.1. Constructing an Upper Bound

We build on an approximation algorithm for quadratic optimization problems given in Goemans and Willamson (1995). Recall that we can represent $V_{ij}(\boldsymbol{x})^{\gamma_{ij}}(R_{ij}(\boldsymbol{x})-z)$ in the objective function of the Function Evaluation problem by $\rho_{ij}(z)\,x_i\,x_j+\theta_i(z)\,x_i\,(1-x_j)+\theta_j(z)\,x_j\,(1-x_i)$. Instead of using the decision variables $\boldsymbol{x}=(x_1,\ldots,x_n)\in\{0,1\}^n$ to capture the subset of offered products, we use $\boldsymbol{y}=(y_0,y_1,\ldots,y_n)\in\{-1,1\}^{n+1}$, where we have $y_0\,y_i=1$ if we offer product i, whereas $y_0\,y_i=-1$ if we do not offer product i. In this case, the decision variable x_i is captured by $(1+y_0\,y_i)/2$. Thus, the expression $\rho_{ij}(z)\,x_i\,x_j+\theta_i(z)\,x_i\,(1-x_j)+\theta_j(z)\,x_j\,(1-x_i)$ is equivalent to

$$\begin{aligned}\rho_{ij}(z)\,\frac{1+y_0\,y_i}{2}\,\frac{1+y_0\,y_j}{2}+\theta_i(z)\,\frac{1+y_0\,y_i}{2}\,\frac{1-y_0\,y_j}{2}+\theta_j(z)\,\frac{1-y_0\,y_i}{2}\,\frac{1+y_0\,y_j}{2}\\ =\frac{\rho_{ij}(z)}{4}\,(1+y_0\,y_i+y_0\,y_j+y_i\,y_j)+\frac{\theta_i(z)}{4}\,(1+y_0\,y_i-y_0\,y_j-y_i\,y_j)\\ +\frac{\theta_i(z)}{4}\,(1-y_0\,y_i+y_0\,y_j-y_i\,y_j).\end{aligned}$$

We define the function $q(y_0,y_i,y_j)=1+y_0\,y_i+y_0\,y_j+y_i\,y_j$ so that the expression above can be written as $\rho_{ij}(z)\,q(y_0,y_i,y_j)/4+\theta_i(z)\,q(y_0,y_i,-y_j)/4+\theta_j(z)\,q(y_0,-y_i,y_j)/4$. In this case, if there is no capacity constraint, then the Function Evaluation problem is equivalent to

$$f(z)=\max_{\substack{\boldsymbol{y}\in\{-1,1\}^{n+1}:\\ y_i=-y_0\;\forall\,i\in N\setminus N(z)}}\left\{\frac{1}{4}\sum_{(i,j)\in M}\big(\rho_{ij}(z)\,q(y_0,y_i,y_j)+\theta_i(z)\,q(y_0,y_i,-y_j)+\theta_j(z)\,q(y_0,-y_i,y_j)\big)\right\},\quad(19)$$

where the constraint $y_i=-y_0$ for all $i\in N\setminus N(z)$ follows from the fact that we can use an argument similar to the one in the proof of Lemma G.1 to show that if $i\notin N(z)$, then there exists an optimal solution to the Function Evaluation problem that does not offer product i. To construct an upper bound on $f(\cdot)$, letting $\boldsymbol{a}\cdot\boldsymbol{b}$ denote the scalar product of the two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, for $(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\in\mathbb{R}^{n+1}\times\mathbb{R}^{n+1}\times\mathbb{R}^{n+1}$, we define the function $p:\mathbb{R}^{n+1}\times\mathbb{R}^{n+1}\times\mathbb{R}^{n+1}\to\mathbb{R}$ as $p(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})=1+\boldsymbol{u}\cdot\boldsymbol{v}+\boldsymbol{u}\cdot\boldsymbol{w}+\boldsymbol{v}\cdot\boldsymbol{w}$. In this case, using the decision variables $\boldsymbol{v}=(\boldsymbol{v}^0,\boldsymbol{v}^1,\ldots,\boldsymbol{v}^n)$ with $\boldsymbol{v}^i\in\mathbb{R}^{n+1}$ for all $i=0,1,\ldots,n$, we define $f^R(z)$

$$\begin{aligned} f^R(z) \;=\max\;\; & \frac{1}{4}\sum_{(i,j)\in M}\left(\rho_{ij}(z)\,p(\boldsymbol{v}^0,\boldsymbol{v}^i,\boldsymbol{v}^j)+\theta_i(z)\,p(\boldsymbol{v}^0,\boldsymbol{v}^i,-\boldsymbol{v}^j)+\theta_j(z)\,p(\boldsymbol{v}^0,-\boldsymbol{v}^i,\boldsymbol{v}^j)\right) \qquad (20)\\ \text{s.t.}\;\; & \boldsymbol{v}^i\cdot\boldsymbol{v}^i=1\;\;\forall i\in N\cup\{0\},\qquad \boldsymbol{v}^i=-\boldsymbol{v}^0\;\;\forall i\in N\setminus N(z)\\ & p(\boldsymbol{v}^0,\boldsymbol{v}^i,\boldsymbol{v}^j)\ge 0\;\;\forall (i,j)\in M,\qquad p(\boldsymbol{v}^0,\boldsymbol{v}^i,-\boldsymbol{v}^j)\ge 0\;\;\forall (i,j)\in M\\ & p(\boldsymbol{v}^0,-\boldsymbol{v}^i,\boldsymbol{v}^j)\ge 0\qquad \forall (i,j)\in M. \end{aligned}$$

Using a feasible solution $\boldsymbol{y}\in\{-1,1\}^{n+1}$ to problem (19), we can come up with a feasible solution $\boldsymbol{v}=(\boldsymbol{v}^0,\boldsymbol{v}^1,\ldots,\boldsymbol{v}^n)$ to problem (20) such that the two solutions provide the same objective values. In particular, we can set $v^i_k=y_i/\sqrt{n+1}$ for all $i,k\in N\cup\{0\}$. Thus, we have $f^R(z)\ge f(z)$.

Next, we formulate problem (20) as an SDP. We define the $(n+1)$-by-$(n+1)$ symmetric matrix $\boldsymbol{\Lambda}(z)=\{\Lambda_{ij}(z):(i,j)\in(N\cup\{0\})\times(N\cup\{0\})\}$ as

$$\Lambda_{ij}(z)=\begin{cases} 0 & \text{if } i=j\\ \dfrac{1}{4}\left(\rho_{ij}(z)-\theta_i(z)+\theta_j(z)\right) & \text{if } (i,j)\in N^2 \text{ and } i<j\\ \displaystyle\sum_{k\in N\setminus\{j\}}\frac{1}{4}\left(\rho_{kj}(z)+\theta_j(z)-\theta_k(z)\right) & \text{if } i=0 \text{ and } j\in N. \end{cases}$$

Since $\boldsymbol{\Lambda}(z)$ is symmetric, we give only the entries that are above the diagonal. We use $\mathbb{S}^{n+1}_+$ to denote the set of $(n+1)$-by-$(n+1)$ symmetric positive semidefinite matrices. In this case, using the decision variables $\boldsymbol{X}=\{X_{ij}:(i,j)\in(N\cup\{0\})\times(N\cup\{0\})\}\in\mathbb{R}^{(n+1)\times(n+1)}$, we can equivalently formulate problem (20) as the SDP given by

$$\begin{aligned} f^R(z)=\max_{\boldsymbol{X}\in\mathbb{S}^{n+1}_+}\Big\{\;&\mathrm{tr}(\boldsymbol{\Lambda}(z)\,\boldsymbol{X})+\frac{1}{4}\sum_{(i,j)\in M}\left(\rho_{ij}(z)+\theta_i(z)+\theta_j(z)\right)\;:\;X_{ii}=1\;\;\forall i\in N\cup\{0\},\\ & X_{i0}=X_{0i}=-1\;\;\forall i\in N\setminus N(z),\quad X_{0i}+X_{0j}+X_{ij}\ge -1\;\;\forall (i,j)\in N^2 \text{ with } i<j,\\ & X_{0i}-X_{0j}-X_{ij}\ge -1\;\;\forall (i,j)\in N^2 \text{ with } i<j,\quad -X_{0i}+X_{0j}-X_{ij}\ge -1\;\;\forall (i,j)\in N^2 \text{ with } i<j,\\ & X_{i0}+X_{j0}+X_{ij}\ge -1\;\;\forall (i,j)\in N^2 \text{ with } i>j,\quad X_{i0}-X_{j0}-X_{ij}\ge -1\;\;\forall (i,j)\in N^2 \text{ with } i>j,\\ -&X_{i0}+X_{j0}-X_{ij}\ge -1\;\;\forall (i,j)\in N^2 \text{ with } i>j\;\Big\}. \qquad (21) \end{aligned}$$

Problem (21) is useful to demonstrate that we can compute the upper bound $f^R(z)$ at any point z by solving an SDP, but to show the performance guarantee for the approximation algorithm we propose, we primarily work with problem (20). Later in our discussion, we use the dual of problem (21) to find the value of $\hat{z}$ that satisfies $f^R(\hat{z})=v_0\,\hat{z}$.

J.2. Randomized Rounding and Performance Guarantee

Fix any $z \in \mathbb{R}_+$. To obtain a random assortment $\hat{\boldsymbol{X}}$ such that $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - z)\} \geq 0.79\, f^R(z)$, we study the following randomized rounding algorithm. Using $\|\cdot\|$ to denote the Euclidean norm, the inputs of the algorithm are $\boldsymbol{v} = (\boldsymbol{v}^0, \boldsymbol{v}^1, \ldots, \boldsymbol{v}^n)$ and $\boldsymbol{u} = (u_0, u_1, \ldots, u_n) \in \mathbb{R}^{n+1}$, where we have $\boldsymbol{v}^i \in \mathbb{R}^{n+1}$ and $\|\boldsymbol{v}^i\| = 1$ for all $i = 0, 1, \ldots, n$.

Randomized Rounding

Step 1: If $\boldsymbol{v}^0 \cdot \boldsymbol{u} \geq 0$, then set $y_0 = 1$. Otherwise, set $y_0 = -1$.

Step 2: For all $i \in N \setminus N(z)$, set $y_i = -y_0$.

Step 3: For all $i \in N(z)$, if $\boldsymbol{v}^i \cdot \boldsymbol{u} \geq 0$, then set $y_i = 1$; otherwise, set $y_i = -1$.

Step 4: Let $\boldsymbol{X} = (X_1, \ldots, X_n) \in \{0,1\}^n$ be such that $X_i = 1$ if $y_0\, y_i = 1$; otherwise, $X_i = 0$.

As a function of its input $(\boldsymbol{v}, \boldsymbol{u})$, we let $\boldsymbol{X}^{\mathrm{RR}}(\boldsymbol{v}, \boldsymbol{u})$ be the output of the randomized rounding algorithm. To get a random subset of products $\hat{\boldsymbol{X}}$ satisfying $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}}) - z)\} \geq 0.79\, f^R(z)$, we will use the input $(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})$, where $\hat{\boldsymbol{v}}$ is an optimal solution to problem (20) and the components of $\hat{\boldsymbol{u}}$ are independent and have the standard normal distribution.

Considering the input $(\boldsymbol{v}, \boldsymbol{u})$ for the randomized rounding algorithm, we will write $(\boldsymbol{v}, \boldsymbol{u}) \in \mathcal{I}$ if and only if for each $i \in \{0, 1, \ldots, n\}$, there exists some $k \in \{0, 1, \ldots, n\}$ such that $v^i_k \neq 0$, u_k has a normal distribution with non-zero variance and u_k is independent of $\{u_j : j \in (N \cup \{0\}) \setminus \{k\}\}$. Observe that if we have $(\boldsymbol{v}, \boldsymbol{u}) \in \mathcal{I}$, then for each $i \in \{0, 1, \ldots, n\}$, there exists some $k \in \{0, 1, \ldots, n\}$ such that $v^i_k\, u_k$ has a normal distribution with non-zero variance and $v^i_k\, u_k$ is independent of $\{v^i_j\, u_j : j \in (N \cup \{0\}) \setminus \{k\}\}$, in which case, it follows that $\boldsymbol{v}^i \cdot \boldsymbol{u} = \sum_{j \in N \cup \{0\}} v^i_j\, u_j$ is non-zero with probability one.

Letting $\hat{\boldsymbol{v}}$ be an optimal solution to problem (20), by the first constraint in this problem, we have $\|\boldsymbol{v}^i\| = 1$ for each $i = 0, 1, \ldots, n$. Therefore, for each $i \in \{0, 1, \ldots, n\}$, there exists some $k \in \{0, 1, \ldots, n\}$ such that $\hat{v}^i_k \neq 0$. In this case, letting $\hat{\boldsymbol{u}}$ be a vector with all components being independent and having the standard normal distribution, we have $(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}}) \in \mathcal{I}$. In this section, we show that if we execute the randomized rounding algorithm with the input $(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})$, then its output $\boldsymbol{X}^{\mathrm{RR}}(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})$ satisfies $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{X}^{\mathrm{RR}}(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}}))^{\gamma_{ij}}(R_{ij}(\boldsymbol{X}^{\mathrm{RR}}(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})) - z)\} \geq 0.79\, f^R(z)$. We use the following two lemmas from Goemans and Willamson (1995).

Lemma J.1 *For all $y \in [-1, 1]$, we have $\frac{1}{\pi} \arccos(y) \geq \chi\,(1 - y)/2$ for some fixed $\chi \in [0.87, \infty)$.*

The lemma above is from Lemmas 3.4 and 3.5 in Goemans and Willamson (1995). For any $(\boldsymbol{v}, \boldsymbol{u}) \in \mathcal{I}$, we define $S_{\boldsymbol{u}}(\boldsymbol{v}^i, \boldsymbol{v}^j)$ as $S_{\boldsymbol{u}}(\boldsymbol{v}^i, \boldsymbol{v}^j) = \mathbb{P}\{\mathrm{sign}(\boldsymbol{v}^i \cdot \boldsymbol{u}) = \mathrm{sign}(\boldsymbol{v}^j \cdot \boldsymbol{u})\}$, where $\mathrm{sign}(x) = 1$

if $x>0$, whereas $\text{sign}(x)=-1$ if $x<0$. Since $(\boldsymbol{v},\boldsymbol{u})\in\mathcal{I}$, for all $i=0,1,\ldots,n$, $\boldsymbol{v}^i\cdot\boldsymbol{u}$ is non-zero with probability one. Therefore, we do not specify $\text{sign}(x)$ for $x=0$. For $(\boldsymbol{v},\boldsymbol{u})\in\mathcal{I}$, an elementary computation in probability yields the identity

$$\mathbb{P}\{\text{sign}(\boldsymbol{v}^0\cdot\boldsymbol{u})=\text{sign}(\boldsymbol{v}^i\cdot\boldsymbol{u})=\text{sign}(\boldsymbol{v}^j\cdot\boldsymbol{u})\}=\frac{1}{2}\Big[S_{\boldsymbol{u}}(\boldsymbol{v}^0,\boldsymbol{v}^i)+S_{\boldsymbol{u}}(\boldsymbol{v}^0,\boldsymbol{v}^j)+S_{\boldsymbol{u}}(\boldsymbol{v}^i,\boldsymbol{v}^j)-1\Big]. \qquad (22)$$

Goemans and Willamson (1995) show this identity in the proof of Lemma 7.3.1 in their paper. In the next lemma, letting $p(\cdot,\cdot,\cdot)$ be as defined right before problem (20), we give a lower bound on the probability above when the components of the vector $\boldsymbol{u}$ are standard normal.

Lemma J.2 *Assume that the components of the vector $\boldsymbol{u}$ are independent and have the standard normal distribution and $\|\boldsymbol{v}^0\|=1$, $\|\boldsymbol{v}^i\|=1$ and $\|\boldsymbol{v}^j\|=1$. For some fixed $\alpha\in[0.79,0.87]$, we have*

$$\frac{1}{2}\Big[S_{\boldsymbol{u}}(\boldsymbol{v}^0,\boldsymbol{v}^i)+S_{\boldsymbol{u}}(\boldsymbol{v}^0,\boldsymbol{v}^j)+S_{\boldsymbol{u}}(\boldsymbol{v}^i,\boldsymbol{v}^j)-1\Big]$$
$$=1-\frac{1}{2\pi}(\arccos(\boldsymbol{v}^0\cdot\boldsymbol{v}^i)+\arccos(\boldsymbol{v}^0\cdot\boldsymbol{v}^j)+\arccos(\boldsymbol{v}^i\cdot\boldsymbol{v}^j))\geq\frac{\alpha}{4}\,p(\boldsymbol{v}^0,\boldsymbol{v}^i,\boldsymbol{v}^j).$$

The lemma above is from Lemmas 7.3.1 and 7.3.2 in Goemans and Willamson (1995). In the next lemma, we give an equivalent expression for $\sum_{(i,j)\in M}\mathbb{E}\{V_{ij}(X^{\text{RR}}(\boldsymbol{v},\boldsymbol{u}))^{\gamma_{ij}}(R_{ij}(\boldsymbol{X}^{\text{RR}}(\boldsymbol{v},\boldsymbol{u}))-z)\}$.

Lemma J.3 *For any input $(\boldsymbol{v},\boldsymbol{u})\in\mathcal{I}$ of the randomized rounding algorithm, the output of the algorithm $X^{\text{RR}}(\boldsymbol{v},\boldsymbol{u})$ satisfies*

$$\begin{aligned}\sum_{(i,j)\in M}&\mathbb{E}\{V_{ij}(X^{\text{RR}}(\boldsymbol{v},\boldsymbol{u}))^{\gamma_{ij}}(R_{ij}(\boldsymbol{X}^{\text{RR}}(\boldsymbol{v},\boldsymbol{u}))-z)\}\\
=\;&\frac{1}{2}\sum_{(i,j)\in M(z)}\Bigg\{\rho_{ij}(z)\Big[S_{\boldsymbol{u}}(\boldsymbol{v}^0,\boldsymbol{v}^i)+S_{\boldsymbol{u}}(\boldsymbol{v}^0,\boldsymbol{v}^j)+S_{\boldsymbol{u}}(\boldsymbol{v}^i,\boldsymbol{v}^j)-1\Big]\\
&+\;\theta_i(z)\Big[S_{\boldsymbol{u}}(\boldsymbol{v}^0,\boldsymbol{v}^i)+S_{\boldsymbol{u}}(\boldsymbol{v}^0,-\boldsymbol{v}^j)+S_{\boldsymbol{u}}(\boldsymbol{v}^i,-\boldsymbol{v}^j)-1\Big]\\
&+\;\theta_j(z)\Big[S_{\boldsymbol{u}}(\boldsymbol{v}^0,-\boldsymbol{v}^i)+S_{\boldsymbol{u}}(\boldsymbol{v}^0,\boldsymbol{v}^j)+S_{\boldsymbol{u}}(-\boldsymbol{v}^i,\boldsymbol{v}^j)-1\Big]\Bigg\}\\
&+\;2\,|N\setminus N(z)|\sum_{i\in N(z)}\theta_i(z)\,S_{\boldsymbol{u}}(\boldsymbol{v}^0,\boldsymbol{v}^i).\end{aligned} \qquad (23)$$

Proof: Fixing the input $(\boldsymbol{v},\boldsymbol{u})$, for notational brevity, we use $\tilde{\boldsymbol{X}}$ to denote the output of the randomized rounding algorithm for the fixed input. By the discussion at the beginning of Section 3.2, we have $\sum_{(i,j)\in M}\mathbb{E}\{V_{ij}(\tilde{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\tilde{\boldsymbol{X}})-z)\}=\sum_{(i,j)\in M}(\rho_{ij}(z)\,\mathbb{P}\{\tilde{X}_i=1,\ \tilde{X}_j=1\}+\theta_i(z)\,\mathbb{P}\{\tilde{X}_i=1,\ \tilde{X}_j=0\}+\theta_j(z)\,\mathbb{P}\{\tilde{X}_i=0,\ \tilde{X}_j=1\})$. We consider four cases.

Case 1: Suppose $i\in N(z)$ and $j\in N(z)$ with $i\neq j$. By Steps 3 and 4 of the randomized rounding algorithm, to have $\tilde{X}_i=1$, we need to have $y_0\,y_i=1$, which, in turn, requires that we have $\text{sign}(y_0) = \text{sign}(y_i)$. The last equality holds if and only if $\text{sign}(\boldsymbol{v}^0 \cdot \boldsymbol{u}) = \text{sign}(\boldsymbol{v}^i \cdot \boldsymbol{u})$. In this case, by (22), it follows that $\mathbb{P}\{\tilde{X}_i = 1,\ \tilde{X}_j = 1\} = \mathbb{P}\{\text{sign}(\boldsymbol{v}^0 \cdot \boldsymbol{u}) = \text{sign}(\boldsymbol{v}^i \cdot \boldsymbol{u}) = \text{sign}(\boldsymbol{v}^j \cdot \boldsymbol{u})\} = \frac{1}{2}(S_{\boldsymbol{u}}(\boldsymbol{v}^0, \boldsymbol{v}^i) + S_{\boldsymbol{u}}(\boldsymbol{v}^0, \boldsymbol{v}^j) + S_{\boldsymbol{u}}(\boldsymbol{v}^i, \boldsymbol{v}^j) - 1)$. Similarly, to have $\tilde{X}_j = 0$, we need to have $y_0\, y_i = -1$, which, in turn, requires that we have $\text{sign}(y_0) = -\text{sign}(y_i)$. The last equality holds if and only if $\text{sign}(\boldsymbol{v}^0 \cdot \boldsymbol{u}) = -\text{sign}(\boldsymbol{v}^j \cdot \boldsymbol{u})$, which we can equivalently write as $\text{sign}(\boldsymbol{v}^0 \cdot \boldsymbol{u}) = \text{sign}(-\boldsymbol{v}^j \cdot \boldsymbol{u})$. Thus, by (22), we get $\mathbb{P}\{\tilde{X}_i = 1,\ \tilde{X}_j = 0\} = \mathbb{P}\{\text{sign}(\boldsymbol{v}^0 \cdot \boldsymbol{u}) = \text{sign}(\boldsymbol{v}^i \cdot \boldsymbol{u}) = \text{sign}(-\boldsymbol{v}^j \cdot \boldsymbol{u})\} = \frac{1}{2}(S_{\boldsymbol{u}}(\boldsymbol{v}^0, \boldsymbol{v}^i) + S_{\boldsymbol{u}}(\boldsymbol{v}^0, -\boldsymbol{v}^j) + S_{\boldsymbol{u}}(\boldsymbol{v}^i, -\boldsymbol{v}^j) - 1)$. Interchanging the roles of $\tilde{X}_i$ and $\tilde{X}_j$ in the last chain of equalities, we also have $\mathbb{P}\{\tilde{X}_i = 0,\ \tilde{X}_j = 1\} = \mathbb{P}\{\text{sign}(\boldsymbol{v}^0 \cdot \boldsymbol{u}) = \text{sign}(-\boldsymbol{v}^i \cdot \boldsymbol{u}) = \text{sign}(\boldsymbol{v}^j \cdot \boldsymbol{u})\} = \frac{1}{2}(S_{\boldsymbol{u}}(\boldsymbol{v}^0, -\boldsymbol{v}^i) + S_{\boldsymbol{u}}(\boldsymbol{v}^0, \boldsymbol{v}^j) + S_{\boldsymbol{u}}(-\boldsymbol{v}^i, \boldsymbol{v}^j) - 1)$.

Case 2: Suppose $i \in N(z)$ and $j \notin N(z)$. By Steps 2 and 4 of the randomized rounding algorithm, we have $\tilde{X}_j = 0$. By an argument similar to the one in Case 1, we also have $\mathbb{P}\{\tilde{X}_i = 1\} = \mathbb{P}\{\text{sign}(\boldsymbol{v}^0 \cdot \boldsymbol{u}) = \text{sign}(\boldsymbol{v}^i \cdot \boldsymbol{u})\} = S_{\boldsymbol{u}}(\boldsymbol{v}^0, \boldsymbol{v}^i)$. So, we get $\mathbb{P}\{\tilde{X}_i = 1,\ \tilde{X}_j = 0\} = S_{\boldsymbol{u}}(\boldsymbol{v}^0, \boldsymbol{v}^i)$.

Case 3: Suppose $i \notin N(z)$ and $j \in N$. By the same argument in Case 2, we have $\mathbb{P}\{\tilde{X}_i = 0,\ \tilde{X}_j = 1\} = S_{\boldsymbol{u}}(\boldsymbol{v}^0, \boldsymbol{v}^j)$.

Case 4: Suppose $i \notin N(z)$ and $j \notin N(z)$ with $i \neq j$. In this case, we have $\tilde{X}_i = 0$ and $\tilde{X}_j = 0$. Putting all of the cases together, under Case 1, if $\tilde{X}_i$ or $\tilde{X}_j$ is non-zero, then we may have $\tilde{X}_i = 1$, $\tilde{X}_j = 1$, or $\tilde{X}_i = 1$, $\tilde{X}_j = 0$, or $\tilde{X}_i = 0$, $\tilde{X}_j = 1$. Under Case 2, if $\tilde{X}_i$ or $\tilde{X}_j$ is non-zero, then we must have $\tilde{X}_i = 1$ and $\tilde{X}_j = 0$. Under Case 3, if $\tilde{X}_i$ or $\tilde{X}_j$ is non-zero, then we must have $\tilde{X}_i = 0$ and $\tilde{X}_j = 1$. Collecting these observations, we obtain

$$
\begin{aligned}
&\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\tilde{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\tilde{\boldsymbol{X}}) - z)\} \\
&\quad = \sum_{(i,j)\in M} \Big\{\rho_{ij}(z)\,\mathbb{P}\{\tilde{X}_i = 1,\ \tilde{X}_j = 1\} + \theta_i(z)\,\mathbb{P}\{\tilde{X}_i = 1,\ \tilde{X}_j = 0\} + \theta_j(z)\,\mathbb{P}\{\tilde{X}_i = 0,\ \tilde{X}_j = 1\}\Big\} \\
&\quad = \sum_{(i,j)\in M} \mathbf{1}(i \in N(z),\ j \in N(z))\Big\{\rho_{ij}(z)\,\mathbb{P}\{\tilde{X}_i = 1,\ \tilde{X}_j = 1\} \\
&\qquad\qquad + \theta_i(z)\,\mathbb{P}\{\tilde{X}_i = 1,\ \tilde{X}_j = 0\} + \theta_j(z)\,\mathbb{P}\{\tilde{X}_i = 0,\ \tilde{X}_j = 1\}\Big\} \\
&\qquad + \sum_{(i,j)\in M} \mathbf{1}(i \in N(z),\ j \notin N(z))\,\theta_i(z)\,\mathbb{P}\{\tilde{X}_i = 1,\ \tilde{X}_j = 0\} \\
&\qquad + \sum_{(i,j)\in M} \mathbf{1}(i \notin N(z),\ j \in N(z))\,\theta_j(z)\,\mathbb{P}\{\tilde{X}_i = 0,\ \tilde{X}_i = 1\},
\end{aligned}
$$

in which case, plugging the expressions for the probabilities $\mathbb{P}\{\tilde{X}_i = 1,\ \tilde{X}_j = 1\}$, $\mathbb{P}\{\tilde{X}_i = 1,\ \tilde{X}_j = 0\}$ and $\mathbb{P}\{\tilde{X}_i = 0,\ \tilde{X}_j = 1\}$ that we have under Cases 1, 2 and 3 above yields the desired result. □

In the next theorem, we give a performance guarantee for the subset of products obtained by the randomized rounding algorithm. Throughout our discussion, α is as given in Lemma J.2.

Theorem J.4 *For a fixed value of $z \in \mathbb{R}_+$, let the subset of products $\boldsymbol{X}^{\text{RR}}(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})$ be the output of the randomized rounding algorithm with the input $(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})$, where we have $\hat{\boldsymbol{v}} = (\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^1, \ldots, \hat{\boldsymbol{v}}^n)$ with $\hat{\boldsymbol{v}}^i \in \mathbb{R}^{n+1}$ and $\|\hat{\boldsymbol{v}}^i\| = 1$ for all $i = 0, 1, \ldots, n$ and the components of the vector $\hat{\boldsymbol{u}}$ are independent and have the standard normal distribution. If $\hat{\boldsymbol{v}}^i = -\hat{\boldsymbol{v}}^0$ for all $i \in N \setminus N(z)$, then we have*

$$\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{X}^{\text{RR}}(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}}))^{\gamma_{ij}} (R_{ij}(\boldsymbol{X}^{\text{RR}}(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})) - z)\}$$
$$\geq \frac{\alpha}{4} \sum_{(i,j)\in M} \Big\{ \rho_{ij}(z)\, p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) + \theta_i(z)\, p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j) + \theta_j(z)\, p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) \Big\}.$$

In particular, if we choose $\hat{\boldsymbol{v}}$ in the input $(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})$ as an optimal solution to problem (20)*, then we have $\sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{X}^{\text{RR}}(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}}))^{\gamma_{ij}} (R_{ij}(\boldsymbol{X}^{\text{RR}}(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})) - z)\} \geq 0.79\, f^R(z)$.*

Proof: For notational brevity, we use $\hat{\boldsymbol{X}}$ to denote the output of the randomized rounding algorithm with the input $(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})$. Note that $(\hat{\boldsymbol{v}}, \hat{\boldsymbol{u}}) \in \mathcal{I}$. We consider four cases.

Case 1: Suppose $i \in N(z)$ and $j \in N(z)$ with $i \neq j$. By Lemma J.2, we have the inequality $\frac{1}{2}(S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^j) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) - 1) \geq \frac{\alpha}{4}\, p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j)$, in which case, since $\rho_{ij}(z) \geq 0$ whenever $i \in N(z)$ and $j \in N(z)$, we obtain $\frac{1}{2}\, \rho_{ij}(z)\, (S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^j) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) - 1) \geq \frac{\alpha}{4}\, \rho_{ij}(z)\, p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j)$. We also get $\frac{1}{2}\, \theta_i(z)\, (S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^j) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j) - 1) \geq \frac{\alpha}{4}\, \theta_i(z)\, p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j)$ and $\frac{1}{2}\, \theta_j(z)\, (S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^j) + S_{\hat{\boldsymbol{u}}}(-\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) - 1) \geq \frac{\alpha}{4}\, \theta_j(z)\, p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j)$ by following the same reasoning. Adding the last three inequalities, we have

$$\begin{aligned}
\frac{1}{2} \Big\{ \rho_{ij}(z) \Big[& S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^j) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) - 1 \Big] \\
& + \theta_i(z) \Big[S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^j) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j) - 1 \Big] \\
& \quad + \theta_j(z) \Big[S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^j) + S_{\hat{\boldsymbol{u}}}(-\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) - 1 \Big] \Big\} \\
& \qquad \geq \frac{\alpha}{4} \Big\{ \rho_{ij}(z)\, p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) + \theta_i(z)\, p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j) + \theta_j(z)\, p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) \Big\}. \qquad (24)
\end{aligned}$$

Case 2: Suppose $i \in N(z)$ and $j \notin N(z)$. By Lemma J.1, for all $y \in [-1, 1]$, we have $1 - \frac{1}{\pi} \arccos(y) = \frac{1}{\pi}(\pi - \arccos(y)) = \arccos(-y)/\pi \geq \chi\,(1 + y)/2$. The definition of $S_{\boldsymbol{u}}(\boldsymbol{v}^i, \boldsymbol{v}^j)$ implies that $S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^i) = 1$. In this case, by Lemma J.2, we obtain

$$\begin{aligned}
S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) &= \frac{1}{2} \Big[S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i) + S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^i) - 1 \Big] \\
&= 1 - \frac{1}{2\pi}(2 \arccos(\hat{\boldsymbol{v}}^0 \cdot \hat{\boldsymbol{v}}^i) + \arccos(\hat{\boldsymbol{v}}^i \cdot \hat{\boldsymbol{v}}^i)) = 1 - \frac{1}{\pi} \arccos(\hat{\boldsymbol{v}}^0 \cdot \hat{\boldsymbol{v}}^i) \geq \frac{\chi}{2}\,(1 + \hat{\boldsymbol{v}}^0 \cdot \hat{\boldsymbol{v}}^i),
\end{aligned}$$

where the third equality uses the fact that $\|\hat{\boldsymbol{v}}^i\| = 1$. On the other hand, since $j \in N \setminus N(z)$, we have $\hat{\boldsymbol{v}}^j = -\hat{\boldsymbol{v}}^0$. Therefore, we obtain $p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) = p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^0) = 1 - \hat{\boldsymbol{v}}^0 \cdot \hat{\boldsymbol{v}}^0 = 0$, where the second equality uses the definition of $p(\cdot, \cdot, \cdot)$ and the last equality uses the fact that $\|\hat{\boldsymbol{v}}^0\| = 1$. By the same argument, we have $p(\hat{\boldsymbol{v}}^0,-\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)=0$. Also, we have $p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,-\hat{\boldsymbol{v}}^j)=p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^0)=2+2\,\hat{\boldsymbol{v}}^0\cdot\hat{\boldsymbol{v}}^i\geq 0$, where the inequality uses the fact that $\|\hat{\boldsymbol{v}}^0\|=\|\hat{\boldsymbol{v}}^i\|=1$ so that $\hat{\boldsymbol{v}}^0\cdot\hat{\boldsymbol{v}}^i\geq -1$. Since $i\in N(z)$, we have $\theta_i(z)\geq 0$. In this case, multiplying the chain of inequalities above by $\theta_i(z)$, we get

$$\begin{aligned}\theta_i(z)\,S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j) \;\geq\; \theta_i(z)\,\frac{\chi}{2}\,(1+\hat{\boldsymbol{v}}^0\cdot\hat{\boldsymbol{v}}^i) \;\geq\; \theta_i(z)\,\frac{\alpha}{2}\,(1+\hat{\boldsymbol{v}}^0\cdot\hat{\boldsymbol{v}}^i)\\ =\frac{\alpha}{4}\Big\{\rho_{ij}(z)\,p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)+\theta_i(z)\,p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,-\hat{\boldsymbol{v}}^j)+\theta_j(z)\,p(\hat{\boldsymbol{v}}^0,-\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)\Big\},\end{aligned}\tag{25}$$

where the second inequality holds since $\chi\in[0.87,\infty)$, $\alpha\in[0.79,0.87]$ and $\hat{\boldsymbol{v}}^0\cdot\hat{\boldsymbol{v}}^i\geq -1$, whereas the equality uses the fact that $p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)=0=p(\hat{\boldsymbol{v}}^0,-\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)$ and $p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,-\hat{\boldsymbol{v}}^j)=2+2\,\hat{\boldsymbol{v}}^0\cdot\hat{\boldsymbol{v}}^i$.

Case 3: Suppose $i\notin N(z)$ and $j\in N$. Interchanging the roles of products i and j in Case 2, the same reasoning in Case 2 yields

$$\theta_j(z)\,S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)\geq\frac{\alpha}{4}\Big\{\rho_{ij}(z)\,p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)+\theta_i(z)\,p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,-\hat{\boldsymbol{v}}^j)+\theta_j(z)\,p(\hat{\boldsymbol{v}}^0,-\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)\Big\}.\tag{26}$$

Case 4: Suppose $i\notin N(z)$ and $j\notin N(z)$ with $i\neq j$. Since $i\notin N(z)$ and $j\notin N(z)$, we have $\hat{\boldsymbol{v}}^i=\hat{\boldsymbol{v}}^j=-\hat{\boldsymbol{v}}^0$. In this case, we obtain $p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)=p(\hat{\boldsymbol{v}}^0,-\hat{\boldsymbol{v}}^0,-\hat{\boldsymbol{v}}^0)=1-\hat{\boldsymbol{v}}^0\cdot\hat{\boldsymbol{v}}^0=0$, where the second equality uses the definition of $p(\cdot,\cdot,\cdot)$ and the third equality uses the fact that $\|\hat{\boldsymbol{v}}^0\|=1$. Similarly, we have $p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,-\hat{\boldsymbol{v}}^j)=0$ and $p(\hat{\boldsymbol{v}}^0,-\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)=0$ as well. Therefore, it follows that $\frac{\alpha}{4}\,(\rho_{ij}(z)\,p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)+\theta_i(z)\,p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,-\hat{\boldsymbol{v}}^j)+\theta_j(z)\,p(\hat{\boldsymbol{v}}^0,-\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j))=0$. To put the four cases considered above together, recalling that we use $\hat{\boldsymbol{X}}$ to denote the output of the randomized rounding algorithm with the input $(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}})$, by Lemma J.3, we have

$$\begin{aligned}&\sum_{(i,j)\in M}\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}})-z)\}\\
&=\;\frac{1}{2}\sum_{(i,j)\in M}\mathbf{1}(i\in N(z),\;j\in N(z))\Big\{\rho_{ij}(z)\Big[S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i)+S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^j)+S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)-1\Big]\\
&\qquad\qquad+\;\theta_i(z)\Big[S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i)+S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0,-\hat{\boldsymbol{v}}^j)+S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^i,-\hat{\boldsymbol{v}}^j)-1\Big]\\
&\qquad\qquad+\;\theta_j(z)\Big[S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0,-\hat{\boldsymbol{v}}^i)+S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^j)+S_{\hat{\boldsymbol{u}}}(-\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)-1\Big]\Big\}\\
&\quad+\sum_{(i,j)\in M}\mathbf{1}(i\in N(z),\;j\notin N(z))\,\theta_i(z)\,S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i)+\sum_{(i,j)\in M}\mathbf{1}(i\notin N(z),\;j\in N(z))\,\theta_j(z)\,S_{\hat{\boldsymbol{u}}}(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^j)\\
&\geq\;\frac{\alpha}{4}\sum_{(i,j)\in M}\Big\{\rho_{ij}(z)\,p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)+\theta_i(z)\,p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,-\hat{\boldsymbol{v}}^j)+\theta_j(z)\,p(\hat{\boldsymbol{v}}^0,-\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)\Big\},\end{aligned}$$

where the inequality follows from (24), (25) and (26), along with the fact that $p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)=p(\hat{\boldsymbol{v}}^0,\hat{\boldsymbol{v}}^i,-\hat{\boldsymbol{v}}^j)=p(\hat{\boldsymbol{v}}^0,-\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{v}}^j)=0$ when $i\notin N(z)$ and $j\notin N(z)$. The chain of inequalities above establishes the first inequality in the lemma. To see the second inequality in the lemma, choosing $\hat{\boldsymbol{v}}$ as an optimal solution to problem (20) in the chain of inequalities above and noting the objective function of problem (20), we obtain $\sum_{(i,j)\in M}\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}})-z)\}\geq\alpha\, f^R(z)$. The optimal objective value of problem (20) is non-negative since setting $\hat{\boldsymbol{v}}^i=-\hat{\boldsymbol{v}}^0$ for all $i\in N$ provides a feasible solution to this problem with an objective value of zero. Therefore, noting that $f^R(z)\geq 0$ and $\alpha\geq 0.79$, the last inequality yields $\sum_{(i,j)\in M}\mathbb{E}\{V_{ij}(\hat{\boldsymbol{X}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{X}})-z)\}\geq 0.79\, f^R(z)$. □

Thus, by Theorem J.4, letting $\hat{\boldsymbol{v}}$ be an optimal solution to problem (20) and $\hat{\boldsymbol{u}}$ be a vector whose components are independent and have the standard normal distribution, if we use the randomized rounding algorithm with the input $(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}})$, then the output of the algorithm $\boldsymbol{X}^{\rm RR}(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}})$ satisfies $\sum_{(i,j)\in M}\mathbb{E}\{V_{ij}(\boldsymbol{X}^{\rm RR}(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}}))^{\gamma_{ij}}(R_{ij}(\boldsymbol{X}^{\rm RR}(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}}))-z)\}\geq 0.79\, f^R(z)$.

The vector $\hat{\boldsymbol{u}}$ is a random variable, so the subset of products $\boldsymbol{X}^{\rm RR}(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}})$ is a random variable as well, but to use Theorem 3.1 to get an approximate solution, we need a deterministic subset of products $\hat{\boldsymbol{x}}$ that satisfies the Sufficient Condition. Since we construct the upper bound $f^R(\cdot)$ by using an SDP relaxation, the method of conditional expectations discussed in Section 4 does not work. Nevertheless, Mahajan and Ramesh (1999) give a procedure to de-randomize the solutions that are obtained through SDP relaxations. We shortly adopt their de-randomization procedure to de-randomize the subset of products $\boldsymbol{X}^{\rm RR}(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}})$. This de-randomization procedure is rather involved. As an alternative, we can simply simulate many realizations of the random variable $\boldsymbol{X}^{\rm RR}(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}})$. In particular, since we know the distribution of $\hat{\boldsymbol{u}}$, we can simulate many realizations of the random variable $\hat{\boldsymbol{u}}$ and compute $\boldsymbol{X}^{\rm RR}(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}})$ for each realization. Therefore, simulating many realizations of $\boldsymbol{X}^{\rm RR}(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}})$ is straightforward. Since we have $\sum_{(i,j)\in M}\mathbb{E}\{V_{ij}(\boldsymbol{X}^{\rm RR}(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}}))^{\gamma_{ij}}(R_{ij}(\boldsymbol{X}^{\rm RR}(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}}))-z)\}\geq 0.79\, f^R(z)$, there must be realizations $\hat{\boldsymbol{x}}$ of the random variable $\boldsymbol{X}^{\rm RR}(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}})$ with strictly positive probability that satisfy $\sum_{(i,j)\in M}V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{x}})-z)\geq 0.79\, f^R(z)$. Also, since we know the value of $f^R(z)$, if we find a realization $\hat{\boldsymbol{x}}$ that satisfies $\sum_{(i,j)\in M}V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{x}})-z)\geq\alpha\, f^R(z)$ for some α other than 0.79, then we can be sure that this subset is an α-approximate solution. Therefore, it is entirely possible that simulating many realizations of the random variable $\boldsymbol{X}^{\rm RR}(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}})$ may provide a deterministic subset of products $\hat{\boldsymbol{x}}$ that satisfies $\sum_{(i,j)\in M}V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{x}})-z)\geq\alpha\, f^R(z)$ for a value of α that is larger than 0.79. Furthermore, since we know the values of both $\sum_{(i,j)\in M}V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}(R_{ij}(\hat{\boldsymbol{x}})-z)$ and $f^R(z)$, we can compute the value of α. In the next section, we show how to compute the fixed point of $f^R(\cdot)/v_0$ by solving an SDP. After this discussion, we show how to de-randomize the output $\boldsymbol{X}^{\rm RR}(\hat{\boldsymbol{v}},\hat{\boldsymbol{u}})$ of the randomized rounding algorithm, if desired.

J.3. Computing the Fixed Point

In Section 3.3, we use the dual of the Upper Bound problem to find the value of $\hat{z}$ satisfying $f^R(\hat{z})=v_0\,\hat{z}$, where $f^R(z)$ is given by the optimal objective value of the Upper Bound problem.

In this section, we use the dual of problem (21) to find the value of $\hat{z}$ satisfying $f^R(\hat{z}) = v_0\,\hat{z}$, where the $f^R(z)$ is given by the optimal objective value of the SDP in (21). To formulate the dual of problem (21), we let $\boldsymbol{\beta} = \{\beta_i : i \in N \cup \{0\}\}$ be the dual variables associated with the first constraint in problem (21). Writing the second and third constraints as $X_{i0} = -1$ and $X_{0i} = -1$, we let $\{\psi_{i0} : i \in N \setminus N(z)\}$ and $\{\psi_{0i} : i \in N \setminus N(z)\}$ be the dual variables associated with the second and third constraints in problem (21). Also, we let $\{\gamma^1_{ij} : (i,j) \in N^2 \text{ with } i < j\}$, $\{\gamma^2_{ij} : (i,j) \in N^2 \text{ with } i < j\}$ and $\{\gamma^3_{ij} : (i,j) \in N^2 \text{ with } i < j\}$ be the dual variables associated with the fourth, fifth and sixth constraints. Similarly, we let $\{\gamma^1_{ij} : (i,j) \in N^2 \text{ with } i > j\}$, $\{\gamma^2_{ij} : (i,j) \in N^2 \text{ with } i > j\}$ and $\{\gamma^3_{ij} : (i,j) \in N^2 \text{ with } i > j\}$ be the dual variables associated with the last three constraints. We define the $(n+1)$-by-$(n+1)$ symmetric matrix of decision variables $\boldsymbol{\Gamma} = \{\Gamma_{ij} : (i,j) \in (N \cup \{0\}) \times (N \cup \{0\})\}$ as

$$\Gamma_{ij} = \begin{cases} 0 & \text{if } i = j \\ \gamma^1_{ij} - \gamma^2_{ij} - \gamma^3_{ij} & \text{if } (i,j) \in N^2 \text{ and } i < j \\ \displaystyle\sum_{k \in N \setminus \{j\}} \gamma^1_{ij} + \gamma^2_{ij} - \gamma^3_{ij} & \text{if } i = 0 \text{ and } j \in N. \end{cases}$$

Shortly, we restrict $\boldsymbol{\Gamma}$ to be a symmetric matrix. Therefore, we give only the entries that are above the diagonal. Also, we define the $(n+1)$-by-$(n+1)$ matrix of decision variables $\boldsymbol{\Psi} = \{\psi_{ij} : (i,j) \in (N \cup \{0\}) \times (N \cup \{0\})\}\}$, where all entries other than $\{\psi_{i0} : i \in N \setminus N(z)\}$ and $\{\psi_{0i} : i \in N \setminus N(z)\}$ are set to zero. For fixed value of z, the dual of problem (21) is given by

$$\begin{aligned} \min \quad & \sum_{i \in N \cup \{0\}} \beta_i - \sum_{i \in N \setminus N(z)} (\psi_{i0} + \psi_{0i}) + \sum_{(i,j) \in M} (\gamma^1_{ij} + \gamma^2_{ij} + \gamma^3_{ij}) + \frac{1}{4} \sum_{(i,j) \in M} (\rho_{ij}(z) + \theta_i(z) + \theta_j(z)) \\ \text{s.t.} \quad & \text{diag}(\boldsymbol{\beta}) + \boldsymbol{\Psi} - \boldsymbol{\Gamma} - \boldsymbol{\Lambda}(z) \in \mathbb{S}^{n+1}_+ \\ & \boldsymbol{\beta} \in \mathbb{R}^{n+1}, \quad \boldsymbol{\Psi} \in \mathbb{R}^{(n+1)\times(n+1)}, \quad \boldsymbol{\Gamma} \in \mathbb{R}^{(n+1)\times(n+1)}_+, \end{aligned}$$

where we use $\text{diag}(\boldsymbol{\beta})$ to denote the diagonal matrix with diagonal entries $\{\beta_i : i \in N \cup \{0\}\}$. Similar to our approach in Section 3.3, to find the value of $\hat{z}$ that satisfies $f^R(\hat{z}) = v_0\,\hat{z}$, we solve

$$\begin{aligned} \min \quad & \sum_{i \in N \cup \{0\}} \beta_i - \sum_{i \in N \setminus N(z)} (\psi_{i0} + \psi_{0i}) + \sum_{(i,j) \in M} (\gamma^1_{ij} + \gamma^2_{ij} + \gamma^3_{ij}) + \frac{1}{4} \sum_{(i,j) \in M} (\rho_{ij}(z) + \theta_i(z) + \theta_j(z)) \\ \text{s.t.} \quad & \text{diag}(\boldsymbol{\beta}) + \boldsymbol{\Psi} - \boldsymbol{\Gamma} - \boldsymbol{\Lambda}(z) \in \mathbb{S}^{n+1}_+ \\ & \sum_{i \in N \cup \{0\}} \beta_i - \sum_{i \in N \setminus N(z)} (\psi_{i0} + \psi_{0i}) + \sum_{(i,j) \in M} (\gamma^1_{ij} + \gamma^2_{ij} + \gamma^3_{ij}) + \frac{1}{4} \sum_{(i,j) \in M} (\rho_{ij}(z) + \theta_i(z) + \theta_j(z)) = v_0\, z \\ & \boldsymbol{\beta} \in \mathbb{R}^{n+1}, \quad \boldsymbol{\Psi} \in \mathbb{R}^{(n+1)\times(n+1)}, \quad \boldsymbol{\Gamma} \in \mathbb{R}^{(n+1)\times(n+1)}_+, \quad z \in \mathbb{R}. \end{aligned}$$

By using precisely the same argument in the proof of Theorem 3.3, we can show that if $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Gamma}}, \hat{z})$ is an optimal solution to the SDP above, then $f^R(\hat{z}) = v_0\,\hat{z}$.

J.4. Preliminary Bounds for De-Randomizing the Subset of Products

In this section and the next, we discuss how to de-randomize the output of our randomized rounding algorithm. In this section, we provide preliminary bounds that will be useful in the analysis of the de-randomization approach. In the next section, we give the de-randomization approach and its analysis. In our de-randomization approach, we follow Mahajan and Ramesh (1999), where the authors de-randomize an SDP relaxation-based approximation algorithm for the 3-vertex coloring problem. We adapt the approach in Mahajan and Ramesh (1999) to our assortment optimization setting. Letting $\hat{\boldsymbol{v}} = (\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^1, \ldots, \hat{\boldsymbol{v}}^n)$ be an optimal solution to problem (20), the starting point in the de-randomization approach is to compute a so-called discretized version of $\hat{\boldsymbol{v}}$. The discretized version is discussed in Section 3.1 and Appendix 1 in Mahajan and Ramesh (1999). The next lemma summarizes this discussion. Here, for any vector $\boldsymbol{v} = (v_0, v_1, \ldots, v_n) \in \mathbb{R}^{n+1}$, we use $\boldsymbol{v}[k, \ldots, \ell] \in \mathbb{R}^{\ell-k+1}$ to denote the vector $(v_k, v_{k+1}, \ldots, v_\ell)$.

Lemma J.5 *Letting $\hat{\boldsymbol{v}} = (\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^1, \ldots, \hat{\boldsymbol{v}}^n)$ be an optimal solution to problem* (20)*, in polynomial time, we can obtain the solution $\bar{\boldsymbol{v}} = (\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^1, \ldots, \bar{\boldsymbol{v}}^n)$ that satisfies the following properties.*

(*a*) *We have $||\bar{\boldsymbol{v}}^i|| = 1$ for all $i \in N \cup \{0\}$ and $\bar{\boldsymbol{v}}^i = -\bar{\boldsymbol{v}}^0$ for all $i \in N \setminus N(z)$.*
(*b*) *We have $|\bar{\boldsymbol{v}}^i \cdot \bar{\boldsymbol{v}}^j - \hat{\boldsymbol{v}}^i \cdot \hat{\boldsymbol{v}}^j| = O(\frac{1}{n})$ for all $i, j \in N \cup \{0\}$; that is, the scalar product of any pair of vectors changes by $O(\frac{1}{n})$.*
(*c*) *Letting $\bar{\boldsymbol{v}}^i = (\bar{v}^i_0, \bar{v}^i_1, \ldots, \bar{v}^i_n)$, we have $|\bar{v}^i_j| = \Omega(\frac{1}{n^2})$ for all $i, j \in N \cup \{0\}$.*
(*d*) *For all $i, j \in N(z) \cup \{0\}$ and $h \in N \cup \{0\}$, if we rotate the coordinate system so that $\bar{\boldsymbol{v}}^i[h \ldots n] = (b_1, 0, \ldots, 0)$ and $\bar{\boldsymbol{v}}^j[h \ldots n] = (b'_1, b'_2, \ldots, 0)$, then we have $|b_1| = \Omega(\frac{1}{n^2})$ and $|b'_2| = \Omega(\frac{1}{n^4})$.*

Throughout our discussion, we use $\bar{\boldsymbol{v}}$ to denote the discretized version of $\hat{\boldsymbol{v}}$ as discussed in the lemma above, where $\hat{\boldsymbol{v}}$ is an optimal solution to problem (20). We define $C(z)$ as

$$C(z) = \frac{1}{4} \left\{ \sum_{(i,j) \in M(z)} (\rho_{ij}(z) + \theta_i(z) + \theta_j(z)) + 2\,|N \setminus N(z)| \sum_{i \in N(z)} \theta_i(z) \right\}.$$

In the next lemma, we give a simple bound on $f^R(z)$.

Lemma J.6 *We have $\frac{1}{4} f^R(z) \leq C(z) \leq f^R(z)$.*

Proof: Using $\boldsymbol{e}_i \in \mathbb{R}^{n+1}$ to denote the unit vector with a one in the i-th component, we define the solution $\tilde{\boldsymbol{v}} = (\tilde{\boldsymbol{v}}^0, \tilde{\boldsymbol{v}}^1, \ldots, \tilde{\boldsymbol{v}}^n)$ to problem (20) as follows. For all $i \in N(z) \cup \{0\}$, we set $\tilde{\boldsymbol{v}}^i = \boldsymbol{e}_i$. For all $i \in N \setminus N(z)$, we set $\tilde{\boldsymbol{v}}^i = -\tilde{\boldsymbol{v}}^0$. If $i \in N(z)$ and $j \in N(z)$, then we have $\tilde{\boldsymbol{v}}^0 \cdot \tilde{\boldsymbol{v}}^i = 0$, $\tilde{\boldsymbol{v}}^0 \cdot \tilde{\boldsymbol{v}}^j = 0$ and $\tilde{\boldsymbol{v}}^i \cdot \tilde{\boldsymbol{v}}^j = 0$, so that $p(\tilde{\boldsymbol{v}}^0, \tilde{\boldsymbol{v}}^i, \tilde{\boldsymbol{v}}^j) = p(\tilde{\boldsymbol{v}}^0, \tilde{\boldsymbol{v}}^i, -\tilde{\boldsymbol{v}}^j) = p(\tilde{\boldsymbol{v}}^0, -\tilde{\boldsymbol{v}}^i, \tilde{\boldsymbol{v}}^j) = 1$. On the other hand, if $i \in N(z)$ and $j \in N \setminus N(z)$, then we have $\tilde{\boldsymbol{v}}^0 \cdot \tilde{\boldsymbol{v}}^i = 0$, $\tilde{\boldsymbol{v}}^0 \cdot \tilde{\boldsymbol{v}}^j = -1$ and $\tilde{\boldsymbol{v}}^i \cdot \tilde{\boldsymbol{v}}^j = 0$, so we get

$p(\tilde{\boldsymbol{v}}^0, \tilde{\boldsymbol{v}}^i, \tilde{\boldsymbol{v}}^j) = 0$, $p(\tilde{\boldsymbol{v}}^0, \tilde{\boldsymbol{v}}^i, -\tilde{\boldsymbol{v}}^j) = 2$ and $p(\tilde{\boldsymbol{v}}^0, -\tilde{\boldsymbol{v}}^i, \tilde{\boldsymbol{v}}^j) = 0$. Similarly, if $i \in N \setminus N(z)$ and $j \in N(z)$, then we have $p(\tilde{\boldsymbol{v}}^0, \tilde{\boldsymbol{v}}^i, \tilde{\boldsymbol{v}}^j) = 0$, $p(\tilde{\boldsymbol{v}}^0, \tilde{\boldsymbol{v}}^i, -\tilde{\boldsymbol{v}}^j) = 0$ and $p(\tilde{\boldsymbol{v}}^0, -\tilde{\boldsymbol{v}}^i, \tilde{\boldsymbol{v}}^j) = 2$. Lastly, if $i \in N \setminus N(z)$ and $j \in N \setminus N(z)$, then we have $p(\tilde{\boldsymbol{v}}^0, \tilde{\boldsymbol{v}}^i, \tilde{\boldsymbol{v}}^j) = p(\tilde{\boldsymbol{v}}^0, \tilde{\boldsymbol{v}}^i, -\tilde{\boldsymbol{v}}^j) = p(\tilde{\boldsymbol{v}}^0, -\tilde{\boldsymbol{v}}^i, \tilde{\boldsymbol{v}}^j) = 0$. Thus, the solution $\tilde{\boldsymbol{v}}$ is feasible to problem (20). Also, it is simple to check that this solution provides an objective value of $\frac{1}{4}\{\sum_{(i,j)\in M(z)}(\rho_{ij}(z) + \theta_i(z) + \theta_j(z)) + 4\,|N \setminus N(z)| \sum_{i\in N(z)} \theta_i(z)\}$ for problem (20). Since $\tilde{\boldsymbol{v}}$ is a feasible but not necessarily an optimal solution to problem (20), we obtain

$$f^R(z) \;\geq\; \frac{1}{4}\left\{ \sum_{(i,j)\in M(z)} (\rho_{ij}(z) + \theta_i(z) + \theta_j(z)) + 4\,|N \setminus N(z)| \sum_{i\in N(z)} \theta_i(z) \right\} \;\geq\; C(z).$$

Let $\hat{\boldsymbol{v}} = (\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^1, \ldots, \hat{\boldsymbol{v}}^n)$ be an optimal solution to problem (20). Since $\|\hat{\boldsymbol{v}}^i\| = 1$ for all $i \in N \cup \{0\}$, we have $\hat{\boldsymbol{v}}^i \cdot \hat{\boldsymbol{v}}^j \leq 1$ for all $i, j \in N \cup \{0\}$, so we get $p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) \leq 4$, $p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j) \leq 4$ and $p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) \leq 4$. Also, if $i \in N(z)$ and $j \in N \setminus N(z)$, then $\hat{\boldsymbol{v}}^0 \cdot \hat{\boldsymbol{v}}^0 = 1$ and $\hat{\boldsymbol{v}}^j = -\hat{\boldsymbol{v}}^0$ by the first two constraints in problem (20), which imply that $p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) = p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^0) = 0$ and $p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) = p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^0) = 0$. Similarly, if $i \in N \setminus N(z)$ and $j \in N(z)$, then we have $p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) = p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^j) = 0$ and $p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j) = p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^j) = 0$. Lastly, if $i \in N \setminus N(z)$ and $j \in N \setminus N(z)$, then we have $p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) = p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j) = p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) = 0$. So, we get

$$\begin{aligned}
f^R(z) \;&=\; \frac{1}{4} \sum_{(i,j)\in M} (\rho_{ij}(z)\, p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) + \theta_i(z)\, p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j) + \theta_j(z)\, p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j)) \\
&=\; \frac{1}{4} \sum_{(i,j)\in M} \mathbf{1}(i \in N(z),\; j \in N(z)) \left\{ \rho_{ij}(z)\, p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) + \theta_i(z)\, p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j) + \theta_j(z)\, p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) \right\} \\
&\qquad + \frac{1}{4} \sum_{(i,j)\in M} \mathbf{1}(i \in N(z),\; j \notin N(z))\, \theta_i(z)\, p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j) \\
&\qquad\qquad + \frac{1}{4} \sum_{(i,j)\in M} \mathbf{1}(i \notin N(z),\; j \in N(z))\, \theta_j(z)\, p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) \\
&\leq\; \sum_{(i,j)\in M} \mathbf{1}(i \in N(z),\; j \in N(z))\, (\rho_{ij}(z) + \theta_i(z) + \theta_j(z)) \\
&\qquad + \sum_{(i,j)\in M} \mathbf{1}(i \in N(z),\; j \notin N(z))\, \theta_i(z) + \sum_{(i,j)\in M} \mathbf{1}(i \notin N(z),\; j \in N(z))\, \theta_j(z) \;=\; 4\, C(z),
\end{aligned} \tag{27}$$

where the inequality holds since $p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) \leq 4$, $p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j) \leq 4$ and $p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) \leq 4$ for all $(i,j) \in M$, along with $\rho_{ij}(z) \geq 0$ for all $(i,j) \in M(z)$ and $\theta_i(z) \geq 0$ for all $i \in N(z)$. $\square$

In the next lemma, we use $C(z)$ to bound the loss in the objective value of problem (20) when we use the discretized solution $\bar{\boldsymbol{v}}$ instead of the optimal solution $\hat{\boldsymbol{v}}$. We define $g^R(z)$ as

$$g^R(z) \;=\; \frac{1}{4} \sum_{(i,j)\in M} (\rho_{ij}(z)\, p(\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j) + \theta_i(z)\, p(\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i, -\bar{\boldsymbol{v}}^j) + \theta_j(z)\, p(\bar{\boldsymbol{v}}^0, -\bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j)),$$

which is the objective value of problem (20) evaluated at $\bar{\boldsymbol{v}}$.

Lemma J.7 *We have* $g^R(z) \geq f^R(z) - O(\frac{1}{n})\, C(z)$.

Proof: By the second part of Lemma J.5, we have $\bar{\boldsymbol{v}}^i \cdot \bar{\boldsymbol{v}}^j \geq \hat{\boldsymbol{v}}^i \cdot \hat{\boldsymbol{v}}^j - O(\frac{1}{n})$, which implies that $p(\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j) \geq p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) - O(\frac{1}{n})$, $p(\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i, -\bar{\boldsymbol{v}}^j) \geq p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j) - O(\frac{1}{n})$ and $p(\bar{\boldsymbol{v}}^0, -\bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j) \geq p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) - O(\frac{1}{n})$. Furthermore, noting that $||\bar{\boldsymbol{v}}^i|| = 1$ for all $i \in N \cup \{0\}$ and $\bar{\boldsymbol{v}}^i = -\bar{\boldsymbol{v}}^0$ for all $i \in N \setminus N(z)$ by the first part of Lemma J.5, using the same argument that we use to obtain the second equality in (27), we obtain

$$
\begin{aligned}
g^R(z) = \frac{1}{4} \sum_{(i,j)\in M} \mathbf{1}(i \in N(z),\ j \in N(z)) \Big\{ \rho_{ij}(z)\, p(\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j) + \theta_i(z)\, p(\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i, -\bar{\boldsymbol{v}}^j) + \theta_j(z)\, p(\bar{\boldsymbol{v}}^0, -\bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j) \Big\} \\
+ \frac{1}{4} \sum_{(i,j)\in M} \mathbf{1}(i \in N(z),\ j \notin N(z))\, \theta_i(z)\, p(\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i, -\bar{\boldsymbol{v}}^j) \\
+ \frac{1}{4} \sum_{(i,j)\in M} \mathbf{1}(i \notin N(z),\ j \in N(z))\, \theta_j(z)\, p(\bar{\boldsymbol{v}}^0, -\bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j).
\end{aligned}
$$

Since $\rho_{ij}(z) \geq 0$ for all $(i,j) \in M(z)$ and $\theta_i(z) \geq 0$ for all $i \in N(z)$, the desired result follows by noting that $p(\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j) \geq p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) - O(\frac{1}{n})$, $p(\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i, -\bar{\boldsymbol{v}}^j) \geq p(\hat{\boldsymbol{v}}^0, \hat{\boldsymbol{v}}^i, -\hat{\boldsymbol{v}}^j) - O(\frac{1}{n})$ and $p(\bar{\boldsymbol{v}}^0, -\bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j) \geq p(\hat{\boldsymbol{v}}^0, -\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j) - O(\frac{1}{n})$, along with the definition of $C(z)$ and the equivalent definition of $f^R(z)$ given by the second equality in (27). □

J.5. De-Randomization Algorithm and Analysis

In this section, we give the de-randomization algorithm and show that we can use this algorithm to de-randomize the output of our randomized rounding algorithm. For any vector $\boldsymbol{w} = (w^0, w^1, \ldots, w^n, w^{n+1}, \ldots, w^{2n+1}) \in \mathbb{R}^{2n+2}$, we define $\boldsymbol{w}(1) \in \mathbb{R}^{n+1}$ and $\boldsymbol{w}(2) \in \mathbb{R}^{n+1}$ as $\boldsymbol{w}(1) = (w^0, w^1, \ldots, w^n)$ and $\boldsymbol{w}(2) = (w^{n+1}, w^{n+2}, \ldots, w^{2n+1})$. Thus, the vectors $\boldsymbol{w}(1)$ and $\boldsymbol{w}(2)$, respectively, correspond to the first and last $n+1$ components of $\boldsymbol{w}$. For a vector $\boldsymbol{w} \in \mathbb{R}^{2n+2}$, note that we can express $\boldsymbol{w}(1) - \boldsymbol{w}(2)$ as $\mathbf{D}\boldsymbol{w}$ for an appropriate matrix $\mathbf{D} \in \mathbb{R}^{(n+1)\times(2n+2)}$. In particular, indexing the elements of $\mathbf{D}$ by $\{d_{ij} : i = 0, 1, \ldots, n,\ j = 0, 1, \ldots, 2n+1\}$, it is enough to set $d_{ij} = 1$ when $i = j$, $d_{ij} = -1$ when $i + n + 1 = j$ and $d_{ij} = 0$ otherwise. Throughout this section, for notational brevity, we will write $\mathbf{D}\boldsymbol{w}$ instead of $\boldsymbol{w}(1) - \boldsymbol{w}(2)$.

Consider using the input $(\bar{\boldsymbol{v}}, \hat{\boldsymbol{u}})$ in the randomized rounding algorithm, where $\bar{\boldsymbol{v}}$ is the discretized version of the optimal solution $\hat{\boldsymbol{v}}$ to problem (20) and the components of the vector $\hat{\boldsymbol{u}}$ are independent and have the standard normal distribution. If we multiply $\hat{\boldsymbol{u}}$ by a positive constant, then the output of the randomized rounding algorithm does not change. Let $\boldsymbol{W}$ be a vector taking values in $\mathbb{R}^{2n+2}$ such that its components are independent and have the standard normal distribution. In this case, the components of the vector $\boldsymbol{W}(1) - \boldsymbol{W}(2) = \mathbf{D}\boldsymbol{W}$ are independent and have normal distribution with mean zero and variance 2. Therefore, using the input $(\bar{\boldsymbol{v}}, \mathbf{D}\boldsymbol{W})$ in the randomized rounding algorithm is equivalent to using the input $(\bar{\boldsymbol{v}}, \hat{\boldsymbol{u}})$.

In our de-randomization approach, we start with the vector $\boldsymbol{W}$ taking values in $\mathbb{R}^{2n+2}$, where the components of $\boldsymbol{W}$ are independent and have the standard normal distribution. Iteratively, we fix one additional component of this vector. Therefore, after $2n+2$ iterations, we obtain a deterministic vector. Using $\bar{\boldsymbol{w}} \in \mathbb{R}^{2n+2}$ to denote the deterministic vector obtained after $2n+2$ iterations, we use the input $(\bar{\boldsymbol{v}}, \mathbf{D}\bar{\boldsymbol{w}})$ in the randomized rounding algorithm. Since $\mathbf{D}\bar{\boldsymbol{w}}$ is deterministic, the output of the randomized rounding algorithm is also deterministic. We will show that $\sum_{(i,j)\in M} V_{ij}(\boldsymbol{X}^{\mathrm{RR}}(\bar{\boldsymbol{v}}, \mathbf{D}\bar{\boldsymbol{w}}))^{\gamma_{ij}} \big(R_{ij}(\boldsymbol{X}^{\mathrm{RR}}(\bar{\boldsymbol{v}}, \mathbf{D}\bar{\boldsymbol{w}})) - z\big) \ge (0.79 - O(\frac{1}{n}))\, f^R(z)$, so the subset of products $\boldsymbol{X}^{\mathrm{RR}}(\bar{\boldsymbol{v}}, \mathbf{D}\bar{\boldsymbol{w}})$ satisfies the Sufficient Condition with $\alpha = 0.79 - O(\frac{1}{n})$. We give the de-randomization algorithm below. In this algorithm, for any random vector $\boldsymbol{W}$ taking values in $\mathbb{R}^{2n+2}$, we let $\boldsymbol{W}(\ell, \delta)$ be the vector also taking values in $\mathbb{R}^{2n+2}$ that is obtained by fixing the ℓ-th component of $\boldsymbol{W}$ at δ. Also, for any random vector $\boldsymbol{W}$ taking values in $\mathbb{R}^{2n+2}$, we define

$$\Phi(\boldsymbol{W}) = \sum_{(i,j)\in M} \mathbb{E}\Big\{ V_{ij}(\boldsymbol{X}^{\mathrm{RR}}(\bar{\boldsymbol{v}}, \mathbf{D}\boldsymbol{W}))^{\gamma_{ij}} \big(R_{ij}(\boldsymbol{X}^{\mathrm{RR}}(\bar{\boldsymbol{v}}, \mathbf{D}\boldsymbol{W})) - z\big) \Big\},$$

where $\bar{\boldsymbol{v}}$ is the discretized version of the optimal solution to problem (20). Lastly, we use the operator precedence $\mathbf{D}\boldsymbol{W}(\ell,\delta) = \mathbf{D}(\boldsymbol{W}(\ell,\delta))$, not $\mathbf{D}\boldsymbol{W}(\ell,\delta) = (\mathbf{D}\boldsymbol{W})(\ell,\delta)$.

<u>**De-Randomization**</u>

Step 1: Set $\ell = 0$. Define the random vector $\boldsymbol{W}^{(0)} = (W^0, W^1, \ldots, W^n, W^{n+1}, \ldots, W^{2n+1})$, where W^i has the standard normal distribution for all $i = 0, 1, \ldots, 2n+1$ and $\{W^i : i = 0, 1, \ldots, 2n+1\}$ are independent.

Step 2: If $\ell < 2n+1$, then define the set S as

$$S = \Big\{ \delta \in [-3\sqrt{\ln n}, 3\sqrt{\ln n}] : \ \delta \text{ is a multiple of } \frac{1}{n^9} \Big\} \bigcup \Big\{ \delta \in [-3\sqrt{\ln n}, 3\sqrt{\ln n}] : \ \ \Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta)) \text{ is not differentiable in } \delta \Big\}.$$

If $\ell = 2n+1$, then letting $\Delta = \{\delta \in \mathbb{R} \ : \ \bar{\boldsymbol{v}}^i \cdot \mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta) = 0 \text{ for some } i \in N \cup \{0\}\}$, $\delta_{\max} = \max_{\delta \in \Delta} \delta$ and $\delta_{\min} = \min_{\delta\in\Delta}$, define the set S as $S = \Delta \cup \{\delta_{\min} - \epsilon, \delta_{\max} + \epsilon\}$ for any $\epsilon > 0$.

Step 3: For each $\delta \in S$, find $f(\delta)$ such that $|f(\delta) - \Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))| = O\left(\frac{1}{n^5}\right) C(z)$. Set $\bar{w}^\ell = \arg\max_{\delta \in S} f(\delta)$.

Step 4: Define the random vector $\boldsymbol{W}^{(\ell+1)} = (\bar{w}^0, \bar{w}^1, \ldots, \bar{w}^\ell, W^{\ell+1}, \ldots, W^{2n+1})$. Increase ℓ by one. If $\ell \le 2n+1$, then go to Step 2; otherwise, return $\bar{\boldsymbol{w}} = (\bar{w}^0, \bar{w}^1, \ldots, \bar{w}^{2n+1})$.

If we have $\ell = 2n+1$, then the vector $\boldsymbol{W}^{(\ell)}(\ell,\delta)$ is of the form $(\bar{w}^0, \bar{w}^1, \ldots, \bar{w}^{2n}, \delta)$, which is deterministic. In this case, we can compute the elements of Δ in Step 2 by solving the linear equation $\bar{\boldsymbol{v}}^i \cdot \mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta) = 0$ for δ for all $i = N \cup \{0\}$. Thus, we can obtain the elements of Δ explicitly and $|\Delta| = O(n)$. Therefore, we can execute Step 2 in the de-randomization algorithm efficiently when $\ell = 2n+1$. On the other hand, if $\ell < 2n+1$, then the vector $\boldsymbol{W}^{(\ell)}(\ell,\delta)$ is of the form $(\bar{w}^0, \bar{w}^1, \ldots, \bar{w}^{\ell-1}, \delta, W^{\ell+1}, \ldots, W^{2n+1})$ and the last $2n+1-\ell$ components of this vector are independent and have the standard normal distribution. By the third part of Lemma J.5, we have $\bar{\boldsymbol{v}}^i_k \neq 0$ for all $i,k \in N \cup \{0\}$. Therefore, we have $(\bar{\boldsymbol{v}}, \mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)) \in \mathcal{I}$. In Section 5 in Mahajan and Ramesh (1999), the authors discuss how to compute the points of non-differentiability for $\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))$ efficiently. At the end of this section, we argue that the number of points of non-differentiability for $\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))$ is polynomial in n. Thus, we can execute Step 2 in the de-randomization algorithm efficiently when $\ell < 2n+1$ as well. At the end of this section, we also discuss how to construct $f(\delta)$ for each $\delta \in S$. In this case, we can find $\bar{w}^\ell$ in Step 3 of the de-randomization algorithm by checking the value of $f(\delta)$ for each $\delta \in S$.

Next, establish the performance guarantee for the output of the de-randomization algorithm. In the random vector $\boldsymbol{W}^{(\ell)}(\ell,\delta)$, the ℓ-th component is fixed at δ, whereas in the random vector $\boldsymbol{W}^{(\ell)}$, the ℓ-th component has the standard normal distribution. In the next lemma, we show that $\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))$ is not too much smaller $\Phi(\boldsymbol{W}^{(\ell)})$, as long as we choose some $\delta \in [-3\sqrt{\ln n}, 3\sqrt{\ln n}]$ to maximize the former quantity. After the next lemma, we build on this result to show that $\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))$ is not too much smaller $\Phi(\boldsymbol{W}^{(\ell)})$, as long as we choose some $\delta \in S$.

Lemma J.8 *For any $\ell < 2n+1$, we have*

$$\max_{\delta \in [-3\sqrt{\ln n}, 3\sqrt{\ln n}]} \left\{ \Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta)) \right\} \;\geq\; \Phi(\boldsymbol{W}^{(\ell)}) - O\Big(\frac{1}{n^{4.5}}\Big)\, C(z).$$

Proof: Letting $\bar{\boldsymbol{v}}$ be the discretized version of the optimal solution to problem (20), for any random vector $\boldsymbol{W}$ taking values in $\mathbb{R}^{2n+2}$ such that $(\bar{\boldsymbol{v}}, \mathbf{D}\boldsymbol{W}) \in \mathcal{I}$, we define $T(\boldsymbol{W})$ as

$$\begin{aligned} T(\boldsymbol{W}) \;=\; \frac{1}{2} \sum_{(i,j)\in M(z)} \Big\{ & \rho_{ij}(z) \Big[S_{\mathbf{D}\boldsymbol{W}}(-\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i) + S_{\mathbf{D}\boldsymbol{W}}(-\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^j) + S_{\mathbf{D}\boldsymbol{W}}(-\bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j) \Big] \\ & + \theta_i(z) \Big[S_{\mathbf{D}\boldsymbol{W}}(-\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i) + S_{\mathbf{D}\boldsymbol{W}}(-\bar{\boldsymbol{v}}^0, -\bar{\boldsymbol{v}}^j) + S_{\mathbf{D}\boldsymbol{W}}(-\bar{\boldsymbol{v}}^i, -\bar{\boldsymbol{v}}^j) \Big] \\ & + \theta_j(z) \Big[S_{\mathbf{D}\boldsymbol{W}}(-\bar{\boldsymbol{v}}^0, -\bar{\boldsymbol{v}}^i) + S_{\mathbf{D}\boldsymbol{W}}(-\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^j) + S_{\mathbf{D}\boldsymbol{W}}(\bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j) \Big] \Big\} \\ & + 2\,|N \setminus N(z)| \sum_{i \in N(z)} \theta_i(z)\, S_{\mathbf{D}\boldsymbol{W}}(-\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i). \end{aligned} \tag{28}$$

Note that $\Phi(\boldsymbol{W})$ can be obtained by setting $\boldsymbol{u} = \mathbf{D}\boldsymbol{W}$ and $\boldsymbol{v}^i = \bar{\boldsymbol{v}}^i$ for all $i \in N \cup \{0\}$ in the expression on the right side of (23). Furthermore, since $(\bar{\boldsymbol{v}}, \mathbf{D}\boldsymbol{W}) \in \mathcal{I}$, the definition of $S_{\boldsymbol{u}}(\boldsymbol{v}^i, \boldsymbol{v}^j)$ implies that $S_{\mathbf{D}\boldsymbol{W}}(\bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j) + S_{\mathbf{D}\boldsymbol{W}}(-\bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j) = 1$ for all $i,j \in N \cup \{0\}$ with $i \neq j$. In this case, noting the definition of $C(z)$, it follows that $T(\boldsymbol{W})+\Phi(\boldsymbol{W})=4\,C(z)$. Also, since $\rho_{ij}(z)\geq 0$ for all $(i,j)\in M(z)$ and $\theta_i(z)\geq 0$ for all $i\in N(z)$, we have $T(\boldsymbol{W})\geq 0$. In this case, noting that $(\bar{\boldsymbol{v}},\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta))\in\mathcal{I}$ by the discussion right after the de-randomization algorithm, we obtain

$$\min_{\delta\in[-3\sqrt{\ln n},3\sqrt{\ln n}]} T(\boldsymbol{W}^{(\ell)}(\ell,\delta)) \;\leq\; \frac{\int_{-3\sqrt{\ln n}}^{3\sqrt{\ln n}} T(\boldsymbol{W}^{(\ell)}(\ell,\delta))\,e^{-\frac{\delta^2}{2}}\,d\delta}{\int_{-3\sqrt{\ln n}}^{3\sqrt{\ln n}} e^{-\frac{\delta^2}{2}}\,d\delta} \;\leq\; \frac{\int_{-\infty}^{\infty} T(\boldsymbol{W}^{(\ell)}(\ell,\delta))\,e^{-\frac{\delta^2}{2}}\,d\delta}{\int_{-3\sqrt{\ln n}}^{3\sqrt{\ln n}} e^{-\frac{\delta^2}{2}}\,d\delta}$$
$$= \frac{T(\boldsymbol{W}^{(\ell)})}{\int_{-3\sqrt{\ln n}}^{3\sqrt{\ln n}} e^{-\frac{\delta^2}{2}}\,d\delta} \;\leq\; \frac{T(\boldsymbol{W}^{(\ell)})}{1-O\left(\frac{1}{n^{4.5}}\right)}. \tag{29}$$

In the chain of inequalities above, the second inequality holds since $T(\boldsymbol{W})\geq 0$. To see the equality, consider each term in the sum on the right side of (28) when we compute $T(\boldsymbol{W}^{(\ell)})$. By the definition of $S_{\boldsymbol{u}}(\boldsymbol{v}^i,\boldsymbol{v}^j)$, we have $S_{\mathbf{D}\boldsymbol{W}}(-\bar{\boldsymbol{v}}^0,\bar{\boldsymbol{v}}^i)=\mathbb{P}\{\mathrm{sign}(-\mathbf{D}\boldsymbol{W}^{(\ell)}\cdot\bar{\boldsymbol{v}}^0)=\mathrm{sign}(\mathbf{D}\boldsymbol{W}^{(\ell)}\cdot\bar{\boldsymbol{v}}^i)\}$. The corresponding term is given by $\mathbb{P}\{\mathrm{sign}(-\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)\cdot\bar{\boldsymbol{v}}^0)=\mathrm{sign}(\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)\cdot\bar{\boldsymbol{v}}^i)\}$, when we compute $T(\boldsymbol{W}^{(\ell)}(\ell,\delta))$. The vectors $\boldsymbol{W}^{(\ell)}$ and $\boldsymbol{W}^{(\ell)}(\ell,\delta)$ agree in all components except for the ℓ-th component. The ℓ-th component of $\boldsymbol{W}^{(\ell)}$ has the standard normal distribution, whereas the ℓ-th component of $\boldsymbol{W}^{(\ell)}(\ell,\delta)$ is fixed at δ. Therefore, by conditioning, we get

$$\mathbb{P}\{\mathrm{sign}(-\mathbf{D}\boldsymbol{W}^{(\ell)}\cdot\bar{\boldsymbol{v}}^0)=\mathrm{sign}(\mathbf{D}\boldsymbol{W}^{(\ell)}\cdot\bar{\boldsymbol{v}}^i)\}$$
$$=\int_{-\infty}^{\infty}\mathbb{P}\{\mathrm{sign}(-\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)\cdot\bar{\boldsymbol{v}}^0)=\mathrm{sign}(\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)\cdot\bar{\boldsymbol{v}}^i)\}\,e^{-\frac{\delta^2}{2}}\,d\delta.$$

Using the same argument for each term in the sum on the right side of (28), we get $T(\boldsymbol{W}^{(\ell)})=\int_{-\infty}^{\infty}T(\boldsymbol{W}^{(\ell)}(\ell,\delta))\,e^{-\frac{\delta^2}{2}}\,d\delta$. The last inequality in (29) holds since $1-\int_{-3\sqrt{\ln n}}^{3\sqrt{\ln n}}e^{-\frac{\delta^2}{2}}\,d\delta=O(\frac{1}{n^{4.5}})$, which is shown in the proof of Lemma 4.2 in Mahajan and Ramesh (1999).

We can compute $\Phi(\boldsymbol{W}^{(\ell)})$ by replacing $\boldsymbol{u}$ with $\mathbf{D}\boldsymbol{W}^{(\ell)}$ and $\boldsymbol{v}$ with $\bar{\boldsymbol{v}}$ on the right side of (23). By (22), each expression delineated with square brackets on the right side of (23) corresponds to a probability. Thus, $\Phi(\boldsymbol{W}^{(\ell)})\geq 0$. Since $T(\boldsymbol{W}^{(\ell)}(\ell,\delta))+\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))=4C(z)$, we get

$$\max_{\delta\in[-3\sqrt{\ln n},3\sqrt{\ln n}]}\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta)) \;=\; 4\,C(z) \;-\; \min_{\delta\in[-3\sqrt{\ln n},3\sqrt{\ln n}]} T(\boldsymbol{W}^{(\ell)}(\ell,\delta))$$
$$\geq\; 4\,C(z) \;-\; \frac{T(\boldsymbol{W}^{(\ell)})}{1-O\left(\frac{1}{n^{4.5}}\right)} \;=\; \frac{4\,C(z)-T(\boldsymbol{W}^{(\ell)})-O\left(\frac{1}{n^{4.5}}\right)C(z)}{1-O\left(\frac{1}{n^{4.5}}\right)}$$
$$=\; \frac{\Phi(\boldsymbol{W}^{(\ell)})-O\left(\frac{1}{n^{4.5}}\right)C(z)}{1-O\left(\frac{1}{n^{4.5}}\right)} \;\geq\; \Phi(\boldsymbol{W}^{(\ell)})-O\Big(\frac{1}{n^{4.5}}\Big)\,C(z),$$

where the first inequality uses (29) and the second inequality holds since $\Phi(\boldsymbol{W}^{(\ell)})\geq 0$. $\square$

Next, we will show that the lemma above holds when $\delta\in S$ instead of $\delta\in[-3\sqrt{\ln n},3\sqrt{\ln n}]$. We need the following lemma from Appendix 2 in Mahajan and Ramesh (1999).

Lemma J.9 *Letting* $h_\delta^{(\ell)}(\bar{\boldsymbol{v}}^i,\bar{\boldsymbol{v}}^j) = S_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\bar{\boldsymbol{v}}^i,\bar{\boldsymbol{v}}^j)$ *for* $i,j \in N\cup\{0\}$ *with* $i\neq j$, *we have* $\frac{dh_\delta^{(\ell)}(\bar{\boldsymbol{v}}^i,\bar{\boldsymbol{v}}^j)}{d\delta} = O(n^4)$ *whenever the derivative exists.*

In the next lemma, we build on Lemmas J.8 and J.9 to show that the inequality in Lemma J.8 continues to hold when we choose $\delta\in S$ and consider any $\ell\le 2n+1$.

Lemma J.10 *For any* $\ell\le 2n+1$, *we have*

$$\max_{\delta\in S}\Big\{\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))\Big\} \;\ge\; \Phi(\boldsymbol{W}^{(\ell)}) - O\Big(\frac{1}{n^{4.5}}\Big)\,C(z).$$

Proof: First, fix $\ell<2n+1$. Let $\delta^* = \arg\max_{\delta\in[-3\sqrt{\ln n},3\sqrt{\ln n}]}\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))$. By the definition of S in the de-randomization algorithm, there exists $\hat{\delta}\in S$ such that $|\delta^*-\hat{\delta}| = O(\frac{1}{n^9})$ and $S_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\bar{\boldsymbol{v}}^i,\bar{\boldsymbol{v}}^j)$ is differentiable in δ over the interval $(\min\{\delta^*,\hat{\delta}\},\max\{\delta^*,\hat{\delta}\})$. So, by Lemma J.9, we get

$$|S_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta^*)}(\bar{\boldsymbol{v}}^i,\bar{\boldsymbol{v}}^j) - S_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\hat{\delta})}(\bar{\boldsymbol{v}}^i,\bar{\boldsymbol{v}}^j)| \;=\; O\Big(\frac{1}{n^5}\Big).$$

Since $(\bar{\boldsymbol{v}},\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta^*))\in\mathcal{I}$, we can compute $\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta^*))$ by replacing $\boldsymbol{u}$ with $\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta^*)$ and $\boldsymbol{v}$ with $\bar{\boldsymbol{v}}$ on the right side of (23). Similarly, we can compute $\Phi(\boldsymbol{W}^{(\ell)}(\ell,\hat{\delta}))$ by replacing $\boldsymbol{u}$ with $\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\hat{\delta})$ and $\boldsymbol{v}$ with $\bar{\boldsymbol{v}}$ on the right side of (23). So, using the equality above, we get

$$\begin{aligned}
&|\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta^*)) - \Phi(\boldsymbol{W}^{(\ell)}(\ell,\hat{\delta}))| \\
&\qquad\le\; \frac{1}{2}\sum_{(i,j)\in M(z)} 3\Big\{\rho_{ij}(z)+\theta_i(z)+\theta_j(z)\Big\}\,O\Big(\frac{1}{n^5}\Big) \;+\; 2\,|N\setminus N(z)|\sum_{i\in N(z)}\theta_i(z)\,O\Big(\frac{1}{n^5}\Big) \\
&\qquad=\; O\Big(\frac{1}{n^5}\Big)\,C(z), \qquad\qquad (30)
\end{aligned}$$

where the last equality uses the fact that $6\,C(z) \ge \frac{3}{2}\sum_{(i,j)\in M(z)}(\rho_{ij}(z)+\theta_i(z)+\theta_j(z)) + 2\,|N\setminus N(z)|\sum_{i\in N(z)}\theta_i(z)$. Thus, it follows that

$$\begin{aligned}
\max_{\delta\in S}\Big\{\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))\Big\} \;&\ge\; \Phi(\boldsymbol{W}^{(\ell)}(\ell,\hat{\delta})) \;\ge\; \Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta^*)) - O\Big(\frac{1}{n^5}\Big)\,C(z) \\
&=\; \max_{\delta\in[-3\sqrt{\ln n},3\sqrt{\ln n}]}\Big\{\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))\Big\} - O\Big(\frac{1}{n^5}\Big)\,C(z) \;\ge\; \Phi(\boldsymbol{W}^{(\ell)}) - O\Big(\frac{1}{n^{4.5}}\Big)\,C(z),
\end{aligned}$$

where the second inequality uses (30) and the third inequality uses Lemma J.8 along with the fact that $O(\frac{1}{n^5}) + O(\frac{1}{n^{4.5}}) = O(\frac{1}{n^{4.5}})$.

Second, fix $\ell = 2n+1$. Noting the discussion right after the de-randomization algorithm, $\boldsymbol{W}^{(\ell)}(\ell,\delta)$ is a deterministic vector of the form $(\bar{w}^0,\bar{w}^1,\ldots,\bar{w}^{2n},\delta)$. Furthermore, by the definition of S, for any $i\in N\cup\{0\}$, the sign of $\bar{\boldsymbol{v}}^i\cdot\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)$ does not change when δ takes values between two consecutive elements of S. Observe that if we execute the randomized rounding algorithm with the input $(\bar{\boldsymbol{v}},\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta))$, the output of the algorithm depends only on the signs of $\{\bar{\boldsymbol{v}}^i \cdot \mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta) : i \in N \cup \{0\}\}$. Therefore, as δ takes values between two consecutive elements of S, the value of $\boldsymbol{X}^{\mathrm{RR}}(\bar{\boldsymbol{v}}, \mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta))$ does not change. Noting the definition of $\Phi(\boldsymbol{W})$, it follows that the value of $\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))$ does not change either as δ takes values between two consecutive elements of S. Furthermore, as δ takes values smaller than $\delta_{\min} - \epsilon$ or larger than $\delta_{\max} + \epsilon$, the value of $\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))$ does not change. Therefore, we obtain

$$\max_{\delta \in S} \left\{ \Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta)) \right\} \;=\; \max_{\delta \in \mathbb{R}} \left\{ \Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta)) \right\} \;\geq\; \int_{-\infty}^{\infty} \Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))\, e^{-\frac{\delta^2}{2}}\, d\delta \;=\; \Phi(\boldsymbol{W}^{(\ell)}),$$

where the last equality follows from the same argument that we use to show that $\int_{-\infty}^{\infty} T(\boldsymbol{W}^{(\ell)}(\ell,\delta))\, e^{-\frac{\delta^2}{2}}\, d\delta = T(\boldsymbol{W}^{(\ell)})$ in the proof of Lemma J.8. □

In the next theorem, we give the performance guarantee for the output of the de-randomization algorithm.

Theorem J.11 *Letting $\bar{\boldsymbol{w}}$ be the output of the de-randomization algorithm and $\hat{\boldsymbol{x}} = \boldsymbol{X}^{\mathrm{RR}}(\bar{\boldsymbol{v}}, \mathbf{D}\bar{\boldsymbol{w}})$, $\hat{\boldsymbol{x}}$ is a deterministic subset of products that satisfies*

$$\sum_{(i,j) \in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{x}}) - z) \;\geq\; \left(0.79 - O\!\left(\frac{1}{n}\right)\right) f^R(z).$$

Proof: Since the vector $\bar{\boldsymbol{w}}$ in Step 4 of the de-randomization algorithm is deterministic, it follows that $\hat{\boldsymbol{x}}$ is a deterministic subset of products. Noting Step 4 of the de-randomization algorithm, we have $\boldsymbol{W}^{(\ell+1)} = \boldsymbol{W}^{(\ell)}(\ell, \bar{w}^\ell)$. In this case, since $\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta)) \geq f(\delta) - O\left(\frac{1}{n^5}\right) C(z)$ for each $\delta \in S$ in Step 3 of the de-randomization algorithm, we get $\Phi(\boldsymbol{W}^{(\ell+1)}) = \Phi(\boldsymbol{W}^{(\ell)}(\ell,\bar{w}^\ell)) \geq f(\bar{w}^\ell) - O\left(\frac{1}{n^5}\right) C(z)$. Furthermore, since, we also have $f(\delta) \geq \Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta)) - O\left(\frac{1}{n^5}\right) C(z)$ for each $\delta \in S$ in Step 3 of the de-randomization algorithm, we get $\max_{\delta \in S} f(\delta) \geq \max_{\delta \in S} \Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta)) - O\left(\frac{1}{n^5}\right) C(z)$. In this case, since $\bar{w}^\ell = \arg\max_{\delta \in S} f(\delta)$ in Step 3, we obtain

$$\begin{aligned}\Phi(\boldsymbol{W}^{(\ell+1)}) \;&=\; \Phi(\boldsymbol{W}^{(\ell)}(\ell,\bar{w}^\ell)) \;\geq\; f(\bar{w}^\ell) - O\left(\frac{1}{n^5}\right) C(z) \\ &=\; \max_{\delta \in S} \{f(\delta)\} - O\left(\frac{1}{n^5}\right) C(z) \;\geq\; \max_{\delta \in S} \left\{ \Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta)) \right\} - O\left(\frac{1}{n^5}\right) C(z).\end{aligned}$$

Noting Lemma J.10 and the fact that $O(\frac{1}{n^5}) + O(\frac{1}{n^{4.5}}) = O(\frac{1}{n^{4.5}})$, the chain of inequalities above implies that $\Phi(\boldsymbol{W}^{(\ell+1)}) \geq \Phi(\boldsymbol{W}^{(\ell)}) - O(\frac{1}{n^{4.5}})\, C(z)$. There are $2n+2$ iterations in the de-randomization algorithm. Adding the last inequality over $\ell = 0, 1, \ldots, 2n+1$, we obtain $\Phi(\boldsymbol{W}^{(2n+2)}) \geq \Phi(\boldsymbol{W}^{(0)}) - O(\frac{1}{n^{3.5}})\, C(z)$.

We have $\boldsymbol{W}^{(2n+2)} = \bar{\boldsymbol{w}}$ at the last iteration of the de-randomization algorithm. Also, $\boldsymbol{W}^{(0)}$ is the random vector taking values in $\mathbb{R}^{2n+2}$, where the components are independent and have the standard normal distribution. As discussed at the beginning of this section, using the input

$(\bar{\boldsymbol{v}}, \mathbf{D}\boldsymbol{W}^{(0)})$ in the randomized rounding algorithm is equivalent to using the input $(\bar{\boldsymbol{v}}, \hat{\boldsymbol{u}})$, where $\hat{\boldsymbol{u}}$ is a vector taking values in $\mathbb{R}^{n+1}$ with the components being independent and having the standard normal distribution. Therefore, we obtain

$$\begin{aligned}
\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{x}}) - z) \;&=\; \sum_{(i,j)\in M} V_{ij}(\boldsymbol{X}^{\mathrm{RR}}(\bar{\boldsymbol{v}}, \mathbf{D}\bar{\boldsymbol{w}}))^{\gamma_{ij}} (R_{ij}(\boldsymbol{X}^{\mathrm{RR}}(\bar{\boldsymbol{v}}, \mathbf{D}\bar{\boldsymbol{w}})) - z) \\
&= \Phi(\bar{\boldsymbol{w}}) \;=\; \Phi(\boldsymbol{W}^{(2n+2)}) \;\geq\; \Phi(\boldsymbol{W}^{(0)}) - O\Big(\frac{1}{n^{3.5}}\Big)\, C(z) \\
&= \sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{X}^{\mathrm{RR}}(\bar{\boldsymbol{v}}, \mathbf{D}\boldsymbol{W}^{(0)}))^{\gamma_{ij}} (R_{ij}(\boldsymbol{X}^{\mathrm{RR}}(\bar{\boldsymbol{v}}, \mathbf{D}\boldsymbol{W}^{(0)})) - z)\} - O\Big(\frac{1}{n^{3.5}}\Big)\, C(z) \\
&= \sum_{(i,j)\in M} \mathbb{E}\{V_{ij}(\boldsymbol{X}^{\mathrm{RR}}(\bar{\boldsymbol{v}}, \hat{\boldsymbol{u}}))^{\gamma_{ij}} (R_{ij}(\boldsymbol{X}^{\mathrm{RR}}(\bar{\boldsymbol{v}}, \hat{\boldsymbol{u}})) - z)\} - O\Big(\frac{1}{n^{3.5}}\Big)\, C(z) \\
&\geq \frac{\alpha}{4} \sum_{(i,j)\in M} \Big\{\rho_{ij}(z)\, p(\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j) + \theta_i(z)\, p(\bar{\boldsymbol{v}}^0, \bar{\boldsymbol{v}}^i, -\bar{\boldsymbol{v}}^j) + \theta_j(z)\, p(\bar{\boldsymbol{v}}^0, -\bar{\boldsymbol{v}}^i, \bar{\boldsymbol{v}}^j)\Big\} - O\Big(\frac{1}{n^{3.5}}\Big)\, C(z) \\
&= \alpha\, g^R(z) - O\Big(\frac{1}{n^{3.5}}\Big)\, C(z),
\end{aligned}$$

where the second and fourth equalities follow from the definition of $\Phi(\boldsymbol{W})$ and the second inequality is by Theorem J.4. Thus, we have $\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{x}}) - z) \geq \alpha\, g^R(z) - O(\frac{1}{n^{3.5}})$, in which case, noting Lemma J.7 and the fact that $O(\frac{1}{n}) + O(\frac{1}{n^{3.5}}) = O(\frac{1}{n})$, we get

$$\begin{aligned}
\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{x}}) - z) \;&\geq\; \alpha\, g^R(z) - O\Big(\frac{1}{n^{3.5}}\Big)\, C(z) \;\geq\; \alpha\, f^R(z) - O\Big(\frac{1}{n}\Big)\, C(z) \\
&\geq\; 0.79 f^R(z) - O\Big(\frac{1}{n}\Big)\, C(z) \;\geq\; \Big(0.79 - O\Big(\frac{1}{n}\Big)\Big) f^R(z).
\end{aligned}$$

Here, the third inequality uses the fact that $\alpha \geq 0.79$ and the fact that the optimal objective value of problem (20) is non-negative as discussed at the end of the proof of Theorem J.4. The last inequality holds since $f^R(z) \geq C(z)$ by Lemma J.6. □

By the theorem above, for any $\epsilon > 0$, these exists a constant K such that if $n \geq K/\epsilon$, then we can use the de-randomization algorithm to obtain a deterministic subset of products that satisfies $V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{x}}) - z) \geq (0.79 - \epsilon)\, f^R(z)$. If $n < K/\epsilon$, then we can enumerate all possible subsets in constant time. Thus, we have a $0.79 - \epsilon$ approximation algorithm for any $\epsilon > 0$.

Closing this section, we consider any iteration $\ell < 2n + 1$ in the de-randomization algorithm and argue that the set S in Step 2 includes a polynomial number of elements and discuss how to construct $f(\delta)$ for each $\delta \in S$ in Step 3. By Lemma 3.4 in Mahajan and Ramesh (1999), $S_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}$ is non-differentiable in δ at no more than two values of δ. Furthermore, in Section 5 in Mahajan and Ramesh (1999), the authors show how to compute the points of non-differentiability for $S_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}$ efficiently. Noting (23), we can express $\Phi(\boldsymbol{W}^{(\ell)}(\ell, \delta))$ as a linear combination of $\{S_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\mp\bar{\boldsymbol{v}}^i,\mp\bar{\boldsymbol{v}}^j): i,j\in N\cup\{0\}$ with $i\neq j\}$. Since there are $O(n^2)$ elements in the set $\{(i,j): i,j\in N$ with $i\neq j\}$, $\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))$ is non-differentiable in δ at no more than $O(n^2)$ values of δ. Also, the set $\{\delta\in[-3\sqrt{\ln n},3\sqrt{\ln n}]:\delta$ is a multiple of $\frac{1}{n^9}\}$ has $O(n^9\sqrt{\ln n})$ elements. Next, we focus on constructing $f(\delta)$ such that $|f(\delta)-\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))| = O\left(\frac{1}{n^5}\right)C(z)$ for each $\delta\in S$. In Section 7 in Mahajan and Ramesh (1999), the authors give an algorithm to construct an approximation to $S_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\bar{\boldsymbol{v}}^i,\bar{\boldsymbol{v}}^j)$ with an error of $O(\frac{1}{n^5})$. Using $\tilde{S}_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\bar{\boldsymbol{v}}^i,\bar{\boldsymbol{v}}^j)$ to denote the approximation, we construct $f(\delta)$ in Step 3 of the de-randomization algorithm as

$$\begin{aligned} f(\delta) \;=\; \frac{1}{2}\sum_{(i,j)\in M(z)} \Bigg\{ & \rho_{ij}(z)\Big[\tilde{S}_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\bar{\boldsymbol{v}}^0,\bar{\boldsymbol{v}}^i)+\tilde{S}_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\bar{\boldsymbol{v}}^0,\bar{\boldsymbol{v}}^j)+\tilde{S}_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\bar{\boldsymbol{v}}^i,\bar{\boldsymbol{v}}^j)-1\Big] \\ &+\;\theta_i(z)\Big[\tilde{S}_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\bar{\boldsymbol{v}}^0,\bar{\boldsymbol{v}}^i)+\tilde{S}_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\bar{\boldsymbol{v}}^0,-\bar{\boldsymbol{v}}^j)+\tilde{S}_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\bar{\boldsymbol{v}}^i,-\bar{\boldsymbol{v}}^j)-1\Big] \\ &+\;\theta_j(z)\Big[\tilde{S}_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\bar{\boldsymbol{v}}^0,-\bar{\boldsymbol{v}}^i)+\tilde{S}_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\bar{\boldsymbol{v}}^0,\bar{\boldsymbol{v}}^j)+\tilde{S}_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(-\bar{\boldsymbol{v}}^i,\bar{\boldsymbol{v}}^j)-1\Big]\Bigg\} \\ &+\;2\,|N\setminus N(z)|\sum_{i\in N(z)}\theta_i(z)\,\tilde{S}_{\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)}(\bar{\boldsymbol{v}}^0,\bar{\boldsymbol{v}}^i). \end{aligned}$$

We can compute $\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))$ by replacing $\boldsymbol{u}$ with $\mathbf{D}\boldsymbol{W}^{(\ell)}(\ell,\delta)$ and $\boldsymbol{v}$ with $\bar{\boldsymbol{v}}$ on the right side of (23). Therefore, we obtain

$$|f(\delta)-\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))| \;\le\; \frac{1}{2}\sum_{(i,j)\in M(z)} 3\Big\{\rho_{ij}(z)+\theta_i(z)+\theta_j(z)\Big\}\,O\Big(\frac{1}{n^5}\Big) \;+\; 2\,|N\setminus N(z)|\sum_{i\in N(z)}\theta_i(z)\,O\Big(\frac{1}{n^5}\Big).$$

By the same reasoning that we use to obtain the equality in (30), the right side of the inequality above is $O(\frac{1}{n^5})\,C(z)$. Lastly, in the proof of Lemma J.10, we show that we can compute $\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))$ exactly when $\ell=2n+1$. Thus, when $\ell=2n+1$, we can use $f(\delta)=\Phi(\boldsymbol{W}^{(\ell)}(\ell,\delta))$.

Appendix K: Structural Properties of the Extreme Points

We focus on the extreme points of the polyhedron given by the set of feasible solutions to the LP that computes f^R at the beginning of Section 5.1. This polyhedron is given by

$$\mathcal{P} \;=\; \left\{(\boldsymbol{x},\boldsymbol{y})\in[0,1]^{|\hat{N}|}\times\mathbb{R}_+^{|\hat{M}|} \;:\; y_{ij}\ge x_i+x_j-1 \;\;\forall\,(i,j)\in\hat{M}, \quad \sum_{i\in\hat{N}} x_i\le c\right\}.$$

If we have $c\ge n$ so that there is no capacity constraint, then $\mathcal{P}$ is the boolean quadric polytope studied by Padberg (1989). By Theorem 7 in Padberg (1989), all components of any extreme point of the boolean quadric polytope take values in $\{0,\frac{1}{2},1\}$. Also, Hochbaum (1998) studies optimization problems over the feasible set $\mathcal{P}\cap\in\{0,1\}^{|\hat{N}|}\times\mathbb{R}_+^{|\hat{M}|}$ with $c\ge n$ and constructs half-integral solutions with objective values exceeding the optimal, in which case, she can obtain 0.5-approximate solutions when the objective function coefficients are all positive. By Theorem 7 in Padberg (1989), if $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$ is an extreme point of $\mathcal{P}$ with $c\ge n$, then $\hat{x}_i\in\{0,\frac{1}{2},1\}$ for all $i\in\hat{N}$. In the next counterexample, we show that this property does not hold when there is a capacity constraint.

Example K.1 (Dense Extreme Points in Capacitated Problem) Consider the polyhedron $\mathcal{P}$ for the case where we have $c=3$ and $|\hat{N}|=7$ with $\hat{N}=\{1,\ldots,7\}$. Let $\hat{\boldsymbol{x}}=(\frac{2}{5},\frac{2}{5},\frac{2}{5},\frac{2}{5},\frac{2}{5},\frac{2}{5},\frac{3}{5})$ and $\hat{\boldsymbol{y}}=\mathbf{0}\in\mathbb{R}_+^{|\hat{M}|}$. Note that we have $\sum_{i\in\hat{N}}\hat{x}_i=c$ and $\hat{x}_i+\hat{x}_j-1\le 0=\hat{y}_{ij}$ for all $(i,j)\in\hat{M}$, which implies that $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})\in\mathcal{P}$. We claim that $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$ is an extreme point of $\mathcal{P}$. Assume on the contrary that there exist $(\hat{\boldsymbol{x}}+\boldsymbol{\epsilon},\hat{\boldsymbol{y}}+\boldsymbol{\delta})\in\mathcal{P}$ and $(\hat{\boldsymbol{x}}-\boldsymbol{\epsilon},\hat{\boldsymbol{y}}-\boldsymbol{\delta})\in\mathcal{P}$ with $(\boldsymbol{\epsilon},\boldsymbol{\delta})$ non-zero so that we have $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})=\frac{1}{2}(\hat{\boldsymbol{x}}+\boldsymbol{\epsilon},\hat{\boldsymbol{y}}+\boldsymbol{\delta})+\frac{1}{2}(\hat{\boldsymbol{x}}-\boldsymbol{\epsilon},\hat{\boldsymbol{y}}-\boldsymbol{\delta})$. Since we have $\hat{\boldsymbol{y}}=\mathbf{0}$, $\hat{\boldsymbol{y}}+\boldsymbol{\delta}\in\mathbb{R}_+^{|\hat{M}|}$ and $\hat{\boldsymbol{y}}-\boldsymbol{\delta}\in\mathbb{R}_+^{|\hat{M}|}$, it must be the case that $\boldsymbol{\delta}=\mathbf{0}$. Noting that $(\hat{\boldsymbol{x}}+\boldsymbol{\epsilon},\hat{\boldsymbol{y}}+\boldsymbol{\delta})\in\mathcal{P}$ and $(\hat{\boldsymbol{x}}-\boldsymbol{\epsilon},\hat{\boldsymbol{y}}-\boldsymbol{\delta})\in\mathcal{P}$, for each $i\in\{1,\ldots,6\}$, the constraint $y_{i7}\ge x_i+x_7-1$ that defines $\mathcal{P}$ yields

$$0=\hat{y}_{i7}+\delta_{i7}\ge\hat{x}_i+\epsilon_i+\hat{x}_7+\epsilon_7-1\qquad\text{and}\qquad 0=\hat{y}_{i7}-\delta_{i7}\ge\hat{x}_i-\epsilon_i+\hat{x}_7-\epsilon_7-1.$$

In the case, since $\hat{x}_i+\hat{x}_7=1$ by the definition of $\hat{\boldsymbol{x}}$, the inequalities above imply that $\epsilon_7=-\epsilon_i$ for all $i\in\{1,\ldots,6\}$. Also, the constraint $\sum_{i\in\hat{N}}x_i\le c$ yields $\sum_{i=1}^7\hat{x}_i+\sum_{i=1}^7\epsilon_i\le c$ and $\sum_{i=1}^7\hat{x}_i-\sum_{i=1}^7\epsilon_i\le c$, in which case, noting that $\sum_{i=1}^7\hat{x}_i=3=c$ by the definition of $\hat{\boldsymbol{x}}$, we obtain $\sum_{i=1}^7\epsilon_i=0$. Combining the last equality with the fact that $\epsilon_i=-\epsilon_7$ for all $i\in\{1,\ldots,6\}$, it follows that $\epsilon_i=0$ for all $i\in\{1,\ldots,7\}$. So, $(\boldsymbol{\epsilon},\boldsymbol{\delta})$ is the zero vector, which is a contradiction.

Next, we give the proof of Lemma 5.1. Throughout the proof, we use the fact that if $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$ is an extreme point of $\mathcal{P}(H)$, then we must have $\hat{y}_{ij}=[\hat{x}_i+\hat{x}_j-1]^+$ for all $(i,j)\in\hat{M}$. This result holds because if $\hat{y}_{ij}>[\hat{x}_i+\hat{x}_j-1]^+$ for some $(i,j)\in\hat{M}$, we can perturb only this component of $\hat{\boldsymbol{y}}$ by $+\epsilon$ and $-\epsilon$ for a small enough $\epsilon>0$ while keeping the other components of $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$ constant. In this case, the two points that we obtain in this fashion are in $\mathcal{P}(H)$ and $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$ can be written as a convex combination of the two points, which contradicts the fact that $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$ is an extreme point of $\mathcal{P}(H)$. Therefore, it indeed holds that $\hat{y}_{ij}=[\hat{x}_i+\hat{x}_j-1]^+$ for all $(i,j)\in\hat{M}$ for any extreme point $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$ of $\mathcal{P}(H)$. Below is the proof of Lemma 5.1.

Proof of Lemma 5.1: To get a contradiction, assume that we have an extreme point $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$ such that $\hat{x}_i\notin(\frac{1}{2},1)$ for all $i\in\hat{N}$ and $\hat{x}_k\notin\{0,\frac{1}{2},1\}$ for some $k\in\hat{N}$. We define $F=\{k\in\hat{N}:\hat{x}_k\notin\{0,\frac{1}{2},1\}\}$. Consider some $k\in F$. Since we assume that $\hat{x}_i\notin(\frac{1}{2},1)$ for all $i\in\hat{N}$ and we have $\hat{x}_k\notin\{0,\frac{1}{2},1\}$, it follows that $\hat{x}_k\in(0,\frac{1}{2})$. Therefore, we have $\hat{x}_k\in(0,\frac{1}{2})$ for all $k\in F$. Also, by the definition of F, we have $\hat{x}_i\in\{0,\frac{1}{2},1\}$ for all $i\notin F$. Summing up the discussion so far, we obtain $\hat{x}_k\in(0,\frac{1}{2})$ for all $k\in F$ and $\hat{x}_i\in\{0,\frac{1}{2},1\}$ for all $i\notin F$. Thus, we can partition $\hat{N}$ into three subsets $F=\{k\in\hat{N}:\hat{x}_k\in(0,\frac{1}{2})\}$, $S=\{i\in\hat{N}:\hat{x}_i\in\{0,\frac{1}{2}\}\}$ and $L=\{i\in\hat{N}:\hat{x}_i=1\}$ so that $\hat{N}=F\cup S\cup L$. Since we assume that $\hat{x}_k\notin\{0,\frac{1}{2},1\}$ for some $k\in\hat{N}$, $|F|=|\hat{N}\setminus(S\cup L)|\ge 1$.

First, we consider the case $|F|=1$. We use k to denote the single element of F. Given the extreme point $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$, we define the solution $(\boldsymbol{x}^1,\boldsymbol{y}^1)$ as follows. For a small enough $\epsilon>0$, we set

$x_k^1 = \hat{x}_k + \epsilon$, $x_i^1 = \hat{x}_i$ for all $i \in S \cup L$ and $y_{ij}^1 = [x_i^1 + x_j^1 - 1]^+$ for all $(i,j) \in \hat{M}$. Since we have $\hat{x}_k \in (0, \frac{1}{2})$ and $\hat{x}_i \in \{0, \frac{1}{2}, 1\}$ for all $i \in \hat{N} \setminus \{k\}$, the sum $\sum_{i \in \hat{N}} \hat{x}_i$ cannot be an integer. So, the constraint $\sum_{i \in \hat{N}} x_i \le c$ is not tight at the extreme point $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$, which implies that $\sum_{i \in \hat{N}} x_i^1 = \sum_{i \in \hat{N}} \hat{x}_i + \epsilon \le c$ for a small enough ϵ. Also, since $\hat{x}_k \in (0, \frac{1}{2})$, we obtain $x_k^1 = \hat{x}_k + \epsilon \le 1$ for a small enough ϵ. In this case, $(\boldsymbol{x}^1, \boldsymbol{y}^1) \in \mathcal{P}(H)$. Similarly, we define the solution $(\boldsymbol{x}^2, \boldsymbol{y}^2)$ as follows. We set $x_k^2 = \hat{x}_k - \epsilon$, $x_i^2 = \hat{x}_i$ for all $i \in S \cup L$ and $y_{ij}^2 = [x_i^2 + x_j^2 - 1]^+$ for all $(i,j) \in \hat{M}$. By using an argument similar to the one earlier in this paragraph, we can verify that $(\boldsymbol{x}^2, \boldsymbol{y}^2) \in \mathcal{P}(H)$.

Consider $(k,j) \in \hat{M}$ with $j \in S$. We have $y_{kj}^1 = [x_k^1 + x_j^1 - 1]^+ = [\hat{x}_k + \epsilon + \hat{x}_j - 1]^+ = 0 = [\hat{x}_k + \hat{x}_j - 1]^+ = \hat{y}_{kj}$, where the third and fourth equalities follow from the fact that $\hat{x}_k \in (0, \frac{1}{2})$ and $\hat{x}_j \in \{0, \frac{1}{2}\}$ for all $j \in S$, whereas the last equality follows from the fact that $\hat{y}_{ij} = [\hat{x}_i + \hat{x}_j - 1]^+$ in the extreme point $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$. Similarly, we have $y_{jk}^1 = \hat{y}_{jk}$ for all $(j,k) \in \hat{M}$ with $j \in S$. Consider $(k,j) \in \hat{M}$ with $j \in L$. We have $y_{kj}^1 = [x_k^1 + x_j^1 - 1]^+ = [\hat{x}_k + \epsilon + \hat{x}_j - 1]^+ = [\hat{x}_k + \hat{x}_j - 1]^+ + \epsilon = \hat{y}_{kj} + \epsilon$, where the third equality follows from the fact that $\hat{x}_j = 1$ for all $j \in L$. Similarly, we have $y_{jk}^1 = \hat{y}_{jk} + \epsilon$ for all $(j,k) \in \hat{M}$ with $j \in L$. For the other cases not considered by the preceding four conditions, we have $y_{ij}^1 = \hat{y}_{ij}$. By following precisely the same line of reasoning used in this paragraph, we can also show that $y_{kj}^2 = \hat{y}_{kj}$ for all $(k,j) \in \hat{M}$ with $j \in S$, $y_{jk}^2 = \hat{y}_{jk}$ for all $(j,k) \in \hat{M}$ with $j \in S$, $y_{kj}^2 = \hat{y}_{kj} - \epsilon$ for all $(k,j) \in \hat{M}$ with $j \in L$ and $y_{jk}^2 = \hat{y}_{jk} - \epsilon$ for all $(j,k) \in \hat{M}$ with $j \in L$. Also, we have $y_{ij}^2 = \hat{y}_{ij}$ for the other cases not considered by the preceding four conditions. In this case, we get $y_{ij}^1 + y_{ij}^2 = 2\,\hat{y}_{ij}$ for all $(i,j) \in \hat{M}$. The discussion in this and the previous paragraph shows that $\hat{\boldsymbol{x}} = \frac{1}{2}\,\boldsymbol{x}^1 + \frac{1}{2}\,\boldsymbol{x}^2$ and $\hat{\boldsymbol{y}} = \frac{1}{2}\,\boldsymbol{y}^1 + \frac{1}{2}\,\boldsymbol{y}^2$ for $(\boldsymbol{x}^1, \boldsymbol{y}^1) \in \mathcal{P}(H)$ and $(\boldsymbol{x}^2, \boldsymbol{y}^2) \in \mathcal{P}(H)$, so $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ cannot be an extreme point of $\mathcal{P}(H)$. Thus, we get a contradiction and the desired result follows.

Second, we consider the case $|F| \ge 2$. We use k and k' to denote any two elements of F. Given the extreme point $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$, we define the solution $(\boldsymbol{x}^1, \boldsymbol{y}^1)$ as follows. For a small enough $\epsilon > 0$, we set $x_k^1 = \hat{x}_k + \epsilon$, $x_{k'}^1 = \hat{x}_{k'} - \epsilon$, $x_i^1 = \hat{x}_i$ for all $i \in (F \setminus \{k, k'\}) \cup S \cup L$ and $y_{ij}^1 = [x_i^1 + x_j^1 - 1]^+$ for all $(i,j) \in \hat{M}$. Since $\sum_{i \in \hat{N}} x_i^1 = \sum_{i \in \hat{N}} \hat{x}_i$, it follows that $(\boldsymbol{x}^1, \boldsymbol{y}^1)$ satisfies the constraint $\sum_{i \in \hat{N}} x_i^1 \le c$. Also, since $\hat{x}_k \in (0, \frac{1}{2})$ and $\hat{x}_{k'} \in (0, \frac{1}{2})$, we have $x_k^1 \le 1$ and $x_{k'}^1 \ge 0$. In this case, it follows that $(\boldsymbol{x}^1, \boldsymbol{y}^1) \in \mathcal{P}(H)$. Similarly, we define the solution $(\boldsymbol{x}^2, \boldsymbol{y}^2)$ as follows. We set $x_k^2 = \hat{x}_k - \epsilon$, $x_{k'}^2 = \hat{x}_{k'} + \epsilon$, $x_i^2 = \hat{x}_i$ for all $i \in (F \setminus \{k, k'\}) \cup S \cup L$ and $y_{ij}^2 = [x_i^2 + x_j^2 - 1]^+$ for all $(i,j) \in \hat{M}$. Using an argument similar to the one earlier in this paragraph, we have $(\boldsymbol{x}^2, \boldsymbol{y}^2) \in \mathcal{P}(H)$. Lastly, following an argument similar to the one in the previous paragraph, we can show that $\hat{\boldsymbol{x}} = \frac{1}{2}\,\boldsymbol{x}^1 + \frac{1}{2}\,\boldsymbol{x}^2$ and $\hat{\boldsymbol{y}} = \frac{1}{2}\,\boldsymbol{y}^1 + \frac{1}{2}\,\boldsymbol{y}^2$, in which case, we, once more, reach a contradiction. □

Appendix L: Confidence Intervals for the Performance Measures

In this section, we provide confidence intervals for the performance measures in Tables 1 and 2. For economy of space, we focus on some of the performance measures. In particular, noting that our

Param. Conf. $(T, n, \bar{\gamma}, \phi_0)$	Confidence Interval 1st	5th	50th
$(I,\ 50, 0.1, 0.25)$	[96.0, 96.8]	[97.2, 97.5]	[98.6, 98.7]
$(I,\ 50, 0.1, 0.50)$	[98.9, 99.1]	[99.2, 99.3]	[99.6, 99.7]
$(I,\ 50, 0.1, 0.75)$	[99.7, 99.8]	[99.8, 99.8]	[99.9, 100]
$(I,\ 50, 0.5, 0.25)$	[96.6, 97.2]	[97.5, 97.8]	[98.8, 98.9]
$(I,\ 50, 0.5, 0.50)$	[99.0, 99.2]	[99.3, 99.3]	[99.8, 99.9]
$(I,\ 50, 0.5, 0.75)$	[99.8, 99.8]	[99.8, 99.8]	[99.9, 100]
$(I,\ 50, 1.0, 0.25)$	[96.4, 97.2]	[97.6, 97.8]	[98.8, 98.9]
$(I,\ 50, 1.0, 0.50)$	[99.0, 99.3]	[99.4, 99.4]	[99.8, 99.8]
$(I,\ 50, 1.0, 0.75)$	[99.8, 99.8]	[99.8, 99.9]	[99.9, 100]
$(I, 100, 0.1, 0.25)$	[96.9, 97.4]	[97.6, 97.8]	[98.6, 98.6]
$(I, 100, 0.1, 0.50)$	[99.1, 99.3]	[99.3, 99.4]	[99.7, 99.7]
$(I, 100, 0.1, 0.75)$	[99.8, 99.8]	[99.8, 99.9]	[99.9, 100]
$(I, 100, 0.5, 0.25)$	[97.4, 97.7]	[97.9, 98.0]	[98.8, 98.8]
$(I, 100, 0.5, 0.50)$	[99.2, 99.3]	[99.4, 99.5]	[99.6, 99.8]
$(I, 100, 0.5, 0.75)$	[99.8, 99.9]	[99.9, 99.9]	[99.9, 100]
$(I, 100, 1.0, 0.25)$	[97.4, 97.8]	[98.0, 98.1]	[98.8, 98.9]
$(I, 100, 1.0, 0.50)$	[99.2, 99.4]	[99.5, 99.5]	[99.8, 99.9]
$(I, 100, 1.0, 0.75)$	[99.8, 99.9]	[99.9, 99.9]	[99.9, 100]

Param. Conf. $(T, n, \bar{\gamma}, \phi_0)$	Confidence Interval 1st	5th	50th
$(C,\ 50, 0.1, 0.25)$	[96.4, 96.8]	[97.1, 97.5]	[98.6, 98.7]
$(C,\ 50, 0.1, 0.50)$	[98.7, 99.0]	[99.2, 99.3]	[99.7, 99.7]
$(C,\ 50, 0.1, 0.75)$	[99.7, 99.8]	[99.8, 99.8]	[99.9, 100]
$(C,\ 50, 0.5, 0.25)$	[97.0, 97.3]	[97.6, 97.7]	[98.8, 98.9]
$(C,\ 50, 0.5, 0.50)$	[98.9, 99.1]	[99.3, 99.4]	[99.8, 99.8]
$(C,\ 50, 0.5, 0.75)$	[99.7, 99.8]	[99.8, 99.8]	[99.9, 100]
$(C,\ 50, 1.0, 0.25)$	[96.7, 97.3]	[97.6, 97.7]	[98.9, 99.0]
$(C,\ 50, 1.0, 0.50)$	[99.1, 99.2]	[99.3, 99.4]	[99.8, 99.8]
$(C,\ 50, 1.0, 0.75)$	[99.8, 99.8]	[99.9, 99.9]	[99.9, 100]
$(C, 100, 0.1, 0.25)$	[97.1, 97.4]	[97.6, 97.8]	[98.6, 98.7]
$(C, 100, 0.1, 0.50)$	[99.0, 99.2]	[99.3, 99.4]	[99.7, 99.7]
$(C, 100, 0.1, 0.75)$	[99.8, 99.8]	[99.9, 99.9]	[99.9 , 100]
$(C, 100, 0.5, 0.25)$	[97.3, 97.6]	[97.8, 98.0]	[98.8, 98.8]
$(C, 100, 0.5, 0.50)$	[99.2, 99.4]	[99.4, 99.5]	[99.7, 99.8]
$(C, 100, 0.5, 0.75)$	[99.8, 99.9]	[99.9, 99.9]	[99.9, 100]
$(C, 100, 1.0, 0.25)$	[97.3, 97.8]	[98.0, 98.1]	[98.8, 98.9]
$(C, 100, 1.0, 0.50)$	[99.2, 99.4]	[99.5, 99.5]	[99.8, 99.8]
$(C, 100, 1.0, 0.75)$	[99.9, 99.9]	[99.9, 99.9]	[99.9, 100]

Table 9 Confidence intervals for the uncapacitated test problems.

test problems are random, we provide 95% confidence intervals for the 1st, 5th and 50th percentiles of the ratio between the expected revenue of the solution from our approximation algorithm and the upper bound on the optimal expected revenue. See, for example, Chapter 5 in Meeker et al. (2017) for computing confidence intervals for percentiles. In Tables 9 and 10, we give the confidence intervals for the performance measures in, respectively, Tables 1 and 2.

Appendix M: Generalized Nested Logit Model with at Most Two Products per Nest

In this section, we give extensions of our results to the generalized nested logit model with at most two products in each nest.

M.1. Assortment Problem

We index the set of products by $N = \{1, \ldots, n\}$. We use $\boldsymbol{x} = (x_1, \ldots, x_n) \in \{0, 1\}^n$ to capture the subset of products that we offer to the customers, where $x_i = 1$ if and only if we offer product i. We denote the collection of nests by M. For each nest $q \in M$, we let $\gamma_q \in [0, 1]$ be the dissimilarity parameter of the nest. For each product i and nest q, we let $\alpha_{iq} \in [0, 1]$ be the membership parameter of product i for nest q. We have $\sum_{q \in M} \alpha_{iq} = 1$ for all $i \in N$, so the membership parameters for a particular product adds up to one. Letting $B_q = \{i \in N : \alpha_{iq} > 0\}$, B_q corresponds to the set of products with strictly positive membership parameters for nest q. We use v_i to denote the preference weight of product i and v_0 to denote the preference weight of the no purchase option. Letting $V_q(\boldsymbol{x}) = \sum_{i \in B_q} (\alpha_{iq} v_i)^{1/\gamma_q} x_i$, under the generalized nested logit model, if we offer the subset of products $\boldsymbol{x}$, then a customer decides to make a purchase in nest q with probability $V_q(\boldsymbol{x})^{\gamma_q} / (v_0 + \sum_{\ell \in M} V_\ell(\boldsymbol{x})^{\gamma_q})$. If the customer decides to make a purchase in nest q, then she chooses product $i \in B_q$ with probability $(\alpha_{iq} v_i)^{1/\gamma_q} x_i / V_q(\boldsymbol{x})$. In this case, if we offer the subset

Param. Conf. $(T, n, \bar{\gamma}, \phi_0, \delta)$	Confidence Interval 1st	5th	50th
$(I,\ 50, 0.1, 0.25, 0.8)$	[90.9, 92.3]	[92.9, 93.5]	[95.9, 96.1]
$(I,\ 50, 0.1, 0.25, 0.5)$	[94.7, 95.5]	[95.9, 96.2]	[98.1, 98.2]
$(I,\ 50, 0.1, 0.25, 0.2)$	[96.9, 98.0]	[98.3, 98.5]	[99.6, 99.7]
$(I,\ 50, 0.1, 0.75, 0.8)$	[99.3, 99.5]	[99.6, 99.6]	[99.9, 99.9]
$(I,\ 50, 0.1, 0.75, 0.5)$	[99.6, 99.7]	[99.8, 99.8]	[99.9, 100]
$(I,\ 50, 0.1, 0.75, 0.2)$	[99.7, 99.8]	[99.9, 99.9]	[99.9, 100]
$(I,\ 50, 0.5, 0.25, 0.8)$	[90.8, 92.7]	[93.4, 94.0]	[96.3, 96.5]
$(I,\ 50, 0.5, 0.25, 0.5)$	[94.2, 95.6]	[96.1, 96.4]	[98.1, 98.2]
$(I,\ 50, 0.5, 0.25, 0.2)$	[96.8, 97.7]	[98.1, 98.4]	[99.5, 99.7]
$(I,\ 50, 0.5, 0.75, 0.8)$	[99.3, 99.5]	[99.6, 99.6]	[99.9, 99.9]
$(I,\ 50, 0.5, 0.75, 0.5)$	[99.5, 99.7]	[99.7, 99.8]	[99.9, 100]
$(I,\ 50, 0.5, 0.75, 0.2)$	[99.7, 99.8]	[99.8, 99.9]	[99.9, 100]
$(I,\ 50, 1.0, 0.25, 0.8)$	[93.1, 93.7]	[94.2, 94.5]	[96.6, 96.7]
$(I,\ 50, 1.0, 0.25, 0.5)$	[93.6, 95.3]	[96.0, 96.4]	[98.1, 98.3]
$(I,\ 50, 1.0, 0.25, 0.2)$	[97.0, 97.8]	[98.2, 98.4]	[99.5, 99.6]
$(I,\ 50, 1.0, 0.75, 0.8)$	[99.5, 99.6]	[99.6, 99.7]	[99.9, 99.9]
$(I,\ 50, 1.0, 0.75, 0.5)$	[99.4, 99.6]	[99.7, 99.7]	[99.9, 100]
$(I,\ 50, 1.0, 0.75, 0.2)$	[99.0, 99.7]	[99.8, 99.9]	[99.9, 100]
$(I, 100, 0.1, 0.25, 0.8)$	[92.2, 93.5]	[94.0, 94.2]	[95.8, 95.9]
$(I, 100, 0.1, 0.25, 0.5)$	[95.9, 96.5]	[96.7, 97.0]	[98.1, 98.2]
$(I, 100, 0.1, 0.25, 0.2)$	[98.0, 98.5]	[98.8, 98.9]	[99.5, 99.6]
$(I, 100, 0.1, 0.75, 0.8)$	[99.6, 99.6]	[99.7, 99.7]	[99.9, 99.9]
$(I, 100, 0.1, 0.75, 0.5)$	[99.7, 99.8]	[99.8, 99.8]	[99.9, 100]
$(I, 100, 0.1, 0.75, 0.2)$	[99.9, 99.9]	[99.9, 99.9]	[99.9, 100]
$(I, 100, 0.5, 0.25, 0.8)$	[92.9, 94.0]	[94.4, 94.7]	[96.3, 96.4]
$(I, 100, 0.5, 0.25, 0.5)$	[95.5, 96.3]	[96.6, 96.9]	[98.0, 98.1]
$(I, 100, 0.5, 0.25, 0.2)$	[97.9, 98.4]	[98.6, 98.8]	[99.5, 99.5]
$(I, 100, 0.5, 0.75, 0.8)$	[99.6, 99.6]	[99.7, 99.7]	[99.9, 99.9]
$(I, 100, 0.5, 0.75, 0.5)$	[99.7, 99.8]	[99.8, 99.8]	[99.9, 99.9]
$(I, 100, 0.5, 0.75, 0.2)$	[99.8, 99.9]	[99.9, 99.9]	[99.9, 100]
$(I, 100, 1.0, 0.25, 0.8)$	[93.4, 94.3]	[94.7, 95.0]	[96.5, 96.7]
$(I, 100, 1.0, 0.25, 0.5)$	[95.6, 96.3]	[96.6, 96.8]	[98.0, 98.2]
$(I, 100, 1.0, 0.25, 0.2)$	[97.9, 98.2]	[98.5, 98.6]	[99.4, 99.5]
$(I, 100, 1.0, 0.75, 0.8)$	[99.6, 99.7]	[99.7, 99.7]	[99.9, 99.9]
$(I, 100, 1.0, 0.75, 0.5)$	[99.7, 99.8]	[99.8, 99.8]	[99.9, 100]
$(I, 100, 1.0, 0.75, 0.2)$	[99.8, 99.9]	[99.9, 99.9]	[99.9, 100]

Param. Conf. $(T, n, \bar{\gamma}, \phi_0, \delta)$	Confidence Interval 1st	5th	50th
$(C,\ 50, 0.1, 0.25, 0.8)$	[90.7, 92.7]	[93.1, 93.6]	[95.8, 96.1]
$(C,\ 50, 0.1, 0.25, 0.5)$	[95.1, 95.8]	[96.1, 96.4]	[98.1, 98.3]
$(C,\ 50, 0.1, 0.25, 0.2)$	[97.7, 98.2]	[98.4, 98.6]	[99.5, 99.7]
$(C,\ 50, 0.1, 0.75, 0.8)$	[99.3, 99.5]	[99.6, 99.6]	[99.9, 99.9]
$(C,\ 50, 0.1, 0.75, 0.5)$	[99.5, 99.7]	[99.7, 99.8]	[99.9, 100]
$(C,\ 50, 0.1, 0.75, 0.2)$	[99.7, 99.8]	[99.9, 99.9]	[99.9, 100]
$(C,\ 50, 0.5, 0.25, 0.8)$	[92.7, 93.1]	[93.8, 94.2]	[96.3, 96.5]
$(C,\ 50, 0.5, 0.25, 0.5)$	[94.6, 95.3]	[95.8, 96.1]	[98.0, 98.2]
$(C,\ 50, 0.5, 0.25, 0.2)$	[97.2, 97.8]	[98.1, 98.3]	[99.4, 99.6]
$(C,\ 50, 0.5, 0.75, 0.8)$	[99.4, 99.5]	[99.6, 99.7]	[99.9, 99.9]
$(C,\ 50, 0.5, 0.75, 0.5)$	[99.4, 99.7]	[99.7, 99.7]	[99.9, 100]
$(C,\ 50, 0.5, 0.75, 0.2)$	[99.6, 99.8]	[99.8, 99.9]	[99.9, 100]
$(C,\ 50, 1.0, 0.25, 0.8)$	[92.5, 93.5]	[94.1, 94.4]	[96.5, 96.8]
$(C,\ 50, 1.0, 0.25, 0.5)$	[94.8, 95.5]	[96.1, 96.4]	[98.0, 98.3]
$(C,\ 50, 1.0, 0.25, 0.2)$	[97.5, 97.9]	[98.1, 98.3]	[99.5, 99.6]
$(C,\ 50, 1.0, 0.75, 0.8)$	[99.4, 99.6]	[99.6, 99.7]	[99.9, 99.9]
$(C,\ 50, 1.0, 0.75, 0.5)$	[99.2, 99.6]	[99.7, 99.7]	[99.9, 100]
$(C,\ 50, 1.0, 0.75, 0.2)$	[99.4, 99.8]	[99.8, 99.9]	[99.9, 100]
$(C, 100, 0.1, 0.25, 0.8)$	[92.7, 93.5]	[93.9, 94.2]	[95.8, 96.0]
$(C, 100, 0.1, 0.25, 0.5)$	[95.8, 96.4]	[96.7, 97.0]	[98.2, 98.2]
$(C, 100, 0.1, 0.25, 0.2)$	[98.3, 98.5]	[98.7, 98.9]	[99.5, 99.6]
$(C, 100, 0.1, 0.75, 0.8)$	[99.5, 99.6]	[99.7, 99.7]	[99.9, 99.9]
$(C, 100, 0.1, 0.75, 0.5)$	[99.7, 99.8]	[99.8, 99.8]	[100, 100]
$(C, 100, 0.1, 0.75, 0.2)$	[99.9, 99.9]	[99.9, 99.9]	[99.9, 100]
$(C, 100, 0.5, 0.25, 0.8)$	[93.3, 94.0]	[94.4, 94.7]	[96.2, 96.4]
$(C, 100, 0.5, 0.25, 0.5)$	[95.9, 96.5]	[96.7, 96.9]	[98.0, 98.1]
$(C, 100, 0.5, 0.25, 0.2)$	[98.0, 98.4]	[98.6, 98.8]	[99.5, 99.5]
$(C, 100, 0.5, 0.75, 0.8)$	[99.5, 99.6]	[99.7, 99.7]	[99.9, 99.9]
$(C, 100, 0.5, 0.75, 0.5)$	[99.7, 99.8]	[99.8, 99.8]	[99.9, 100]
$(C, 100, 0.5, 0.75, 0.2)$	[99.9, 99.9]	[99.9, 99.9]	[99.9, 100]
$(C, 100, 1.0, 0.25, 0.8)$	[93.9, 94.3]	[94.9, 95.1]	[96.5, 96.6]
$(C, 100, 1.0, 0.25, 0.5)$	[96.0, 96.4]	[96.7, 97.0]	[98.1, 98.1]
$(C, 100, 1.0, 0.25, 0.2)$	[97.9, 98.4]	[98.6, 98.8]	[99.4, 99.5]
$(C, 100, 1.0, 0.75, 0.8)$	[99.6, 99.7]	[99.7, 99.7]	[99.9, 100]
$(C, 100, 1.0, 0.75, 0.5)$	[99.5, 99.8]	[99.8, 99.8]	[99.9, 100]
$(C, 100, 1.0, 0.75, 0.2)$	[99.8, 99.9]	[99.9, 99.9]	[99.9, 100]

Table 10 Confidence intervals for the capacitated test problems.

of products $\boldsymbol{x}$ and a customer has already decided to make a purchase in nest q, then the expected revenue that we obtain from the customer is $R_q(\boldsymbol{x}) = \sum_{i \in B_q} p_i \, (\alpha_{iq} \, v_i)^{1/\gamma_q} \, x_i / V_q(\boldsymbol{x})$, where p_i is the revenue of product i. Letting $\pi(\boldsymbol{x})$ be the expected revenue that we obtain from a customer when we offer the subset of products $\boldsymbol{x}$, we have

$$\pi(\boldsymbol{x}) \ = \ \sum_{q \in M} \frac{V_q(\boldsymbol{x})^{\gamma_q}}{v_0 + \sum_{\ell \in M} V_\ell(\boldsymbol{x})^{\gamma_\ell}} \, R_q(\boldsymbol{x}) \ = \ \frac{\sum_{q \in M} V_q(\boldsymbol{x})^{\gamma_q} \, R_q(\boldsymbol{x})}{v_0 + \sum_{q \in M} V_q(\boldsymbol{x})^{\gamma_q}}.$$

Defining the set of feasible subsets of products $\mathcal{F}$ that we can offer in the same way that we define in the main body of the paper, we formulate our assortment problem as

$$z^* \ = \ \max_{\boldsymbol{x} \in \mathcal{F}} \ \pi(\boldsymbol{x}) \ = \ \max_{x \in \mathcal{F}} \left\{ \frac{\sum_{q \in M} V_q(\boldsymbol{x})^{\gamma_q} \, R_q(\boldsymbol{x})}{v_0 + \sum_{q \in M} V_q(\boldsymbol{x})^{\gamma_q}} \right\}. \tag{31}$$

The PCL model is a special case of the generalized nested logit model. Since the assortment problem under the PCL model is strongly NP-hard, the problem above is strongly NP-hard as well.

M.2. A Framework for Approximation Algorithms

To relate problem (31) to the problem of computing the fixed point of a function, we define the function $f:\mathbb{R}\rightarrow\mathbb{R}$ as

$$f(z) \;=\; \max_{\boldsymbol{x}\in\mathcal{F}} \left\{ \sum_{q\in M} V_q(\boldsymbol{x})^{\gamma_q}\,(R_q(\boldsymbol{x})-z)\right\}. \tag{32}$$

We can show that $f(\cdot)$ is decreasing and continuous with $f(0)\geq 0$. Therefore, there exists a unique $\hat{z}\geq 0$ that satisfies $f(\hat{z})=v_0\,\hat{z}$. In our approximation framework, we will construct an upper bound $f^R(\cdot)$ on $f(\cdot)$ so that $f^R(z)\geq f(z)$ for all $z\in\mathbb{R}$. This upper bound will be decreasing and continuous with $f^R(0)\geq 0$, so that there also exists a unique $\hat{z}\geq 0$ that satisfies $f^R(\hat{z})=v_0\,\hat{z}$. Theorem 3.1 continues to hold, as long as we replace $\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}\,(R_{ij}(\hat{\boldsymbol{x}})-\hat{z})\geq \alpha\, f^R(\hat{z})$ in the Sufficient Condition with $\sum_{q\in M} V_q(\hat{\boldsymbol{x}})^{\gamma_q}\,(R_q(\hat{\boldsymbol{x}})-\hat{z})\geq \alpha\, f^R(\hat{z})$. The approximation framework given in Section 3.1 continues to hold as well, as long as we replace $\sum_{(i,j)\in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}}\,(R_{ij}(\hat{\boldsymbol{x}})-\hat{z})\geq \alpha\, f^R(\hat{z})$ in Step 3 with $\sum_{q\in M} V_q(\hat{\boldsymbol{x}})^{\gamma_q}\,(R_q(\hat{\boldsymbol{x}})-\hat{z})\geq \alpha\, f^R(\hat{z})$. The preceding discussion holds with an arbitrary number of products in each nest. In the rest of the discussion, we focus on the case where there are at most two products in each nest so that $|B_q|\leq 2$ for all $q\in M$.

We proceed to constructing an upper bound $f^R(\cdot)$ on $f(\cdot)$. To construct this upper bound, we use an LP relaxation of problem (32). In particular, we define $\rho_q(z)$ and $\theta_{iq}(z)$ as

$$\rho_q(z) = \left(\sum_{i\in B_q} (\alpha_{iq}\,v_i)^{1/\gamma_q}\right)^{\gamma_q} \frac{\sum_{i\in B_q}(p_i-z)\,(\alpha_{iq}\,v_i)^{1/\gamma_q}}{\sum_{i\in B_q}(\alpha_{iq}\,v_i)^{1/\gamma_q}} \quad \text{and} \quad \theta_{iq}(z)=\alpha_{iq}\,v_i\,(p_i-z).$$

As in the beginning of Section 3.2, if $x_i=1$ for all $i\in B_q$, then $V_q(\boldsymbol{x})^{\gamma_q}\,(R_q(\boldsymbol{x})-z)=\rho_q(z)$, whereas if $x_i=1$ for exactly one $i\in B_q$, then $V_q(\boldsymbol{x})^{\gamma_q}\,(R_q(\boldsymbol{x})-z)=\theta_{iq}(z)$. Note that since $|B_q|=2$, if we offer some product in nest q, then we must have either $x_i=1$ for all $i\in B_q$ or $x_i=1$ for exactly one $i\in B_q$. In this case, letting $\mu_q(z)=\rho_q(z)-\sum_{i\in B_q}\theta_{iq}(z)$ for notational brevity, we have

$$V_q(\boldsymbol{x})^{\gamma_q}\,(R_q(\boldsymbol{x})-z) \;=\; \rho_q(z)\prod_{j\in B_q}x_j+\sum_{i\in B_q}\theta_{iq}(z)\left(1-\prod_{j\in B_q\setminus\{i\}}x_j\right)x_i \;=\; \mu_q(z)\prod_{i\in B_q}x_i+\sum_{i\in B_q}\theta_{iq}(z)\,x_i. \tag{33}$$

Thus, we can use the expression $\sum_{q\in M}(\mu_q(z)\prod_{i\in B_q}x_i+\sum_{i\in B_q}\theta_{iq}(z)\,x_i)$ to equivalently write the objective function of problem (32), in which case, this problem becomes

$$f(z) \;=\; \max\left\{\sum_{q\in M}\left(\mu_q(z)\prod_{i\in B_q}x_i+\sum_{i\in B_q}\theta_{iq}(z)\,x_i\right) \;:\; \sum_{i\in N}x_i\leq c,\quad x_i\in\{0,1\}\quad\forall\, i\in N\right\}.$$

To linearize the term $\prod_{i\in B_q}x_i$ in the objective function above, we define the decision variable $y_q\in\{0,1\}$ with the interpretation that $y_q=\prod_{i\in B_q}x_i$. To ensure that y_q takes the value $\prod_{i\in B_q}x_i$, we impose the constraints $y_q \geq \sum_{i\in B_q} x_i - |B_q| + 1$ and $y_q \leq x_i$ for all $i \in B_q$. In this case, noting that $|B_q| \leq 2$, if $x_i = 1$ for all $i \in B_q$, then the constraints $y_q \geq \sum_{i\in B_q} x_i - |B_q| + 1$ and $y_q \leq x_i$ for all $i \in B_q$ ensures that $y_q = 1$. If $x_i = 0$ for some $i \in B_q$, then the constraint $y_q \leq x_i$ ensures that $y_q = 0$. Using the decision variables $\{y_q : q \in M\}$, we can formulate the problem above as an integer program. We use the LP relaxation of this integer program to construct the upper bound $f^R(\cdot)$ on $f(\cdot)$. In particular, we define the upper bound $f^R(\cdot)$ as

$$\begin{aligned} f^R(z) \;=\; \max \quad & \sum_{q\in M}\left(\mu_q(z)\,y_q + \sum_{i\in B_q}\theta_{iq}(z)\,x_i\right) \\ \text{s.t.} \quad & y_q \geq \sum_{i\in B_q} x_i - |B_q| + 1 \quad \forall\, q\in M \\ & y_q \leq x_i \quad \forall\, i \in B_q,\; q\in M \\ & \sum_{i\in N} x_i \leq c \\ & 0 \leq x_i \leq 1 \quad \forall\, i\in N, \quad y_q \geq 0 \quad \forall\, q \in M. \end{aligned} \tag{34}$$

In the formulation of the LP above, we use the fact that if we offer some product in a nest, then we offer either all or one of the products in this nest, which holds when the number of products in this nest is at most two. Lemma 3.2 continues to hold so that $f^R(z)$ is decreasing in z. In this case, by the same argument right before Lemma 3.2, $f^R(\cdot)$ is decreasing and continuous with $f^R(0) \geq 0$. So, there exists a unique $\hat{z} \geq 0$ satisfying $f^R(\hat{z}) = v_0\,\hat{z}$. Using the same approach in the proof of Theorem 3.3, we can show that we can solve an LP to compute the fixed point of $f^R(\cdot)/v_0$.

We define $N(z) = \{i \in N : p_i \geq z\}$ and $M(z) = \{q \in M : p_i \geq z \;\; \forall\, i \in B_q\}$. As in the proof of Lemma G.1, we can show that there exists an optimal solution $\boldsymbol{x}^* = \{x_i^* : i \in N\}$ and $\boldsymbol{y}^* = \{y_q^* : q \in M\}$ to problem (34) with $x_i^* = 0$ for all $i \notin N(z)$ and $y_q^* = 0$ for all $q \notin M(z)$. So, we can assume that $x_i^* = 0$ for all $i \notin N(z)$ and $y_q^* = 0$ for all $q \notin M(z)$. Also, as in the proof of Lemma G.2, we can show that $\mu_q(z) \leq 0$ for all $q \in M(z)$, in which case, the decision variables $\{y_q : q \in M(z)\}$ take their smallest possible value in an optimal solution to problem (34). Thus, the constraint $y_q \leq x_i$ is redundant. In this case, problem (34) is equivalent to the problem

$$\begin{aligned} f^R(z) \;=\; \max \quad & \sum_{q\in M}\left(\mu_q(z)\,y_q + \sum_{i\in B_q}\theta_{iq}(z)\,x_i\right) \\ \text{s.t.} \quad & y_q \geq \sum_{i\in B_q} x_i - |B_q| + 1 \quad \forall\, q\in M \\ & \sum_{i\in N} x_i \leq c \\ & 0 \leq x_i \leq 1 \quad \forall\, i\in N, \quad y_q \geq 0 \quad \forall\, q \in M. \end{aligned} \tag{35}$$

Working with problem (35), rather than problem (34), will be more convenient. In the next two sections, we focus on the uncapacitated and capacitated problems separately.

M.3. Uncapacitated Problem

We consider the case where $c \geq n$ so that there is no capacity constraint. We let $\hat{z}$ be such that $f^R(\hat{z}) = v_0\,\hat{z}$. Throughout this section, since the value of $\hat{z}$ is fixed, as is done in the main body of the paper, we exclude the reference to $\hat{z}$. In particular, we let $\mu_q = \mu_q(\hat{z})$, $\theta_{iq} = \theta_{iq}(\hat{z})$, $\rho_{iq} = \rho_{iq}(z)$, $\hat{N} = N(z)$, $\hat{M} = M(z)$ and $f^R = f^R(\hat{z})$. We let $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ be an optimal solution to problem (35) with $z = \hat{z}$. As discussed at the end of the previous section, without loss of generality, we assume that $x_i^* = 0$ for all $i \notin \hat{N}$ and $y_q^* = 0$ for all $i \notin \hat{M}$. We define the random subset of products $\hat{\boldsymbol{X}} = \{\hat{X}_i : i \in N\}$ as follows. For each $i \in N$, we have $\hat{X}_i = 1$ with probability x_i^*, whereas $\hat{X}_i = 0$ with probability $1 - x_i^*$. Different components of the vector $\hat{\boldsymbol{X}}$ are independent of each other. Through minor modifications in the proof of Theorem 4.1, we can show that

$$\mathbb{E}\Bigg\{ \sum_{q\in M} V_q(\hat{\boldsymbol{X}})^{\gamma_q} \left(R_q(\hat{\boldsymbol{X}}) - \hat{z}\right) \Bigg\} \geq 0.5\, f^R. \tag{36}$$

In particular, using the same argument in the proof of Lemma G.2, we can show that $\mu_q \leq 0$ for all $q \in \hat{M}$. In this case, for $q \in \hat{M}$, the decision variable y_q takes its smallest possible value in an optimal solution to problem (35). Thus, without loss of generality, we can assume that the optimal solution $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ to problem (35) satisfies $y_q^* = \left[\sum_{i\in B_q} x_i^* - |B_q| + 1\right]^+$ for all $q \in \hat{M}$. Furthermore, by the definition of $\hat{\boldsymbol{X}}$, we have $\mathbb{E}\{\hat{X}_i\} = x_i^*$ for all $i \in N$ and the different components of $\hat{\boldsymbol{X}}$ are independent of each other. In this case, noting (33), we get

$$\begin{aligned}
\sum_{q\in M} \mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q} (R_q(\hat{\boldsymbol{X}}) - \hat{z})\} \;&=\; \sum_{q\in M} \left(\mu_q \prod_{i\in B_q} \mathbb{E}\{\hat{X}_i\} + \sum_{i\in B_q} \theta_{iq}\, \mathbb{E}\{\hat{X}_i\} \right) \\
&= \sum_{q\in M} \left(\mu_q \prod_{i\in B_q} x_i^* + \sum_{i\in B_q} \theta_{iq}\, x_i^* \right) \\
&= \sum_{q\in M} \mathbf{1}(q \in \hat{M}) \left(\mu_q \prod_{i\in B_q} x_i^* + \sum_{i\in B_q} \theta_{iq}\, x_i^* \right) + \sum_{q\in M} \mathbf{1}(q \notin \hat{M}) \left(\sum_{i\in B_q} \theta_{iq}\, x_i^* \right) \\
&= \sum_{q\in M} \mathbf{1}(q \in \hat{M}) \left(\mu_q \left[\sum_{i\in B_q} x_i^* - |B_q| + 1 \right]^+ + \sum_{i\in B_q} \theta_{iq}\, x_i^* \right) + \sum_{q\in M} \mathbf{1}(q \notin \hat{M}) \left(\sum_{i\in B_q} \theta_{iq}\, x_i^* \right) \\
&\qquad + \sum_{q\in M} \mathbf{1}(q \in \hat{M}) \mu_q \left(\prod_{i\in B_q} x_i^* - \left[\sum_{i\in B_q} x_i^* - |B_q| + 1 \right]^+ \right) \\
&= f^R + \sum_{q\in \hat{M}} \mu_q \left(\prod_{i\in B_q} x_i^* - \left[\sum_{i\in B_q} x_i^* - |B_q| + 1 \right]^+ \right) \\
&\geq f^R + \frac{1}{4} \sum_{q\in \hat{M}} \mu_q.
\end{aligned} \tag{37}$$

In the chain of inequalities above, the third equality uses the fact that if $q \notin \hat{M}$, then there exists some $j \in B_q$ such that $p_j < \hat{z}$, in which case, we get $j \notin \hat{N}$. Having $j \notin \hat{N}$ implies that $x_j^* = 0$. Thus, there exists some $j \in B_q$ such that $x_j^* = 0$, which yields $\prod_{i \in B_q} x_i^* = 0$. The fifth equality is by the fact that $y_q^* = \big[\sum_{i \in B_q} x_i^* - |B_q| + 1\big]^+$ for all $q \in \hat{M}$ and $y_q^* = 0$ for all $q \notin \hat{M}$. To see that the inequality holds, if $|B_q| = 2$ for nest q, then we use the fact that $0 \le ab - [a+b-1]^+ \le 1/4$ for all $a, b \in [0,1]$, whereas if $|B_q| = 1$ for nest q, then we use the fact that $a - [a]^+ = 0$ for all $a \in [0,1]$, along with the fact that $\mu_q \le 0$ for all $q \in \hat{M}$.

Next, we give a feasible solution to problem (35), which, in turn, allows us to construct a lower bound on f^R. We define the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ to problem (35) as

$$\hat{x}_i = \begin{cases} \frac{1}{2} & \text{if } i \in \hat{N} \\ 0 & \text{if } i \notin \hat{N}, \end{cases} \qquad \hat{y}_q = \begin{cases} 0 & \text{if } |B_q| = 2 \\ 1 & \text{if } |B_q| = 1. \end{cases}$$

It is straightforward to check that the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is feasible to problem (35) when we do not have a capacity constraint. Therefore, the objective value of problem (35) evaluated at $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ provides a lower bound on f^R, so we have

$$\begin{aligned} f^R \;&\ge\; \sum_{q \in M} \left(\mu_q\, \hat{y}_q + \sum_{i \in B_q} \theta_{iq}\, \hat{x}_i \right) \;=\; \sum_{q \in M} \mathbf{1}(q \in \hat{M}) \sum_{i \in B_q} \theta_{iq}\, \hat{x}_i + \sum_{q \in M} \mathbf{1}(q \notin \hat{M}) \sum_{i \in B_q} \theta_{iq}\, \hat{x}_i \\ &=\; \frac{1}{2} \sum_{q \in M} \mathbf{1}(q \in \hat{M}) \sum_{i \in B_q} \theta_{iq} + \frac{1}{2} \sum_{q \in M} \mathbf{1}(q \notin \hat{M}) \sum_{i \in B_q} \mathbf{1}(i \in \hat{N})\, \theta_{iq} \;\ge\; \frac{1}{2} \sum_{q \in M} \mathbf{1}(q \in \hat{M}) \sum_{i \in B_q} \theta_{iq} \\ &\ge\; \frac{1}{2} \sum_{q \in M} \mathbf{1}(q \in \hat{M}) \left(\sum_{i \in B_q} \theta_{iq} - \rho_q \right) \;=\; -\frac{1}{2} \sum_{q \in \hat{M}} \mu_q, \end{aligned}$$

In the chain of inequalities above, the first equality holds because the definition of μ_q immediately implies that $\mu_q = 0$ when $|B_q| = 1$. Also, we have $\hat{y}_q = 0$ when $|B_q| = 2$. Therefore, we have $\mu_q\, \hat{y}_q = 0$ for all $q \in M$. The second equality holds since having $q \in \hat{M}$ implies having $i \in \hat{N}$ for all $i \in B_q$, in which case, we have $\hat{x}_i = \frac{1}{2}$ for all $i \in B_q$. The second inequality holds because if we have $i \in \hat{N}$, then $\theta_{iq} \ge 0$ by the definition of $\hat{N}$ and θ_{iq}. The last inequality follows from the fact that if we have $q \in \hat{M}$, then the definition of ρ_q implies that $\rho_q \ge 0$. The chain of inequalities above yields the inequality $f^R + \frac{1}{2} \sum_{q \in \hat{M}} \mu_q \ge 0$, in which case, by the chain of inequalities in (37), we get $\sum_{q \in M} \mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q} (R_q(\hat{\boldsymbol{X}}) - \hat{z})\} \ge \frac{1}{2} f^R + \frac{1}{2}(f^R + \frac{1}{2} \sum_{q \in \hat{M}} \mu_q) \ge \frac{1}{2} f^R$, establishing (36).

The subset of products $\hat{\boldsymbol{X}}$ is a random variable but we can use the method of conditional expectations to de-randomize the subset of products $\hat{\boldsymbol{X}}$ so that we obtain a deterministic subset of products $\hat{\boldsymbol{x}}$ that satisfies $\sum_{(i,j) \in M} V_{ij}(\hat{\boldsymbol{x}})^{\gamma_{ij}} (R_{ij}(\hat{\boldsymbol{x}}) - \hat{z}) \ge 0.5\, f^R$. In this case, $\hat{\boldsymbol{x}}$ is a 0.5-approximate solution to the uncapacitated problem under the generalized nested logit model with at most two products in each nest. Unfortunately, our approach in Appendix H to obtain a 0.6-approximate solution does not extend to the generalized nested logit model. In Appendix H, letting $\hat{N} = \{1, \ldots, m\}$, we index the products such that $\theta_1 \ge \theta_2 \ge \ldots \ge \theta_m$. Under the generalized nested logit model, it is not possible to ensure that we have $\theta_{1q} \ge \theta_{2q} \ge \ldots \ge \theta_{mq}$ for all $q \in M$.

M.4. Capacitated Problem

Considering the assortment problem with the capacity constraint, we let $\hat{z}$ be such that $f^R(\hat{z}) = v_0\,\hat{z}$, where $f^R(z)$ is the optimal objective value of problem (35). Our goal is to find a subset of products $\hat{\boldsymbol{x}}$ such that $\sum_{q\in M} V_q(\hat{\boldsymbol{x}})^{\gamma_q}\,(R_q(\hat{\boldsymbol{x}}) - \hat{z}) \geq 0.25\, f^R(\hat{z})$ and $\sum_{i\in N}\hat{x}_i \leq c$.

M.4.1. Half-Integral Solutions Through Iterative Rounding Similar to our approach for the capacitated problem, since $\hat{z}$ is fixed, we exclude the reference to $\hat{z}$. In particular, we let $\mu_q = \mu_q(\hat{z})$, $\theta_{iq} = \theta_{iq}(\hat{z})$, $\rho_{iq} = \rho_{iq}(\hat{z})$, $\hat{N} = N(\hat{z})$, $\hat{M} = M(\hat{z})$ and $f^R = f^R(\hat{z})$. For any $H \subseteq \hat{N}$, we use the polyhedron $\mathcal{P}(H)$ to denote the set of feasible solutions to problem (35) after we fix the values of the decision variables $\{x_i : i \in H\}$ at $\frac{1}{2}$ and the values of the decision variables $\{x_i : i \notin \hat{N}\}$ and $\{y_q : q \notin \hat{M}\}$ at zero. Therefore, the polyhedron $\mathcal{P}(H)$ is given by

$$\mathcal{P}(H) \;=\; \Bigg\{ (\boldsymbol{x},\boldsymbol{y}) \in [0,1]^{|N|} \times \mathbb{R}_+^{|M|} \;:\; y_q \geq \sum_{i\in B_q} x_i - |B_q| + 1 \quad \forall\, q\in M, \quad \sum_{i\in N} x_i \leq c,$$
$$x_i = \frac{1}{2} \quad \forall\, i \in H, \quad x_i = 0 \quad \forall\, i \notin \hat{N}, \quad y_q = 0 \quad \forall\, q \notin \hat{M} \Bigg\}. \tag{38}$$

An analogue of Lemma 5.1 holds for the polyhedron $\mathcal{P}(H)$ given above. In particular, for any $H \subseteq \hat{N}$, letting $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$ be an extreme point of $\mathcal{P}(H)$, we can use the same approach in the proof of Lemma 5.1 to show that if there is no product $i \in \hat{N}$ such that $\frac{1}{2} < \hat{x}_i < 1$, then we have $\hat{x}_i \in \{0, \frac{1}{2}, 1\}$ for all $i \in \hat{N}$. We use the same iterative rounding algorithm given in Section 5.1, as long as we modify the objective function of the Variable Fixing problem to reflect the objective function of problem (35). Thus, we replace the Variable Fixing problem in Step 2 with

$$f^k \;=\; \max \Bigg\{ \sum_{q\in M} \Bigg(\mu_q(z)\, y_q + \sum_{i\in B_q} \theta_{iq}(z)\, x_i \Bigg) \;:\; (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{P}(H^k) \Bigg\}. \tag{39}$$

At the first iteration of the iterative rounding algorithm, we have $H^1 = \varnothing$. Also, as discussed earlier, there exists an optimal solution $(\boldsymbol{x}^*,\boldsymbol{y}^*)$ to problem (35), where we have $x_i^* = 0$ for all $i \notin \hat{N}$ and $y_q^* = 0$ for all $q \notin \hat{M}$. Therefore, we have $f^1 = f^R$.

As the iterations of the iterative rounding algorithm progress, we fix additional variables at the value $\frac{1}{2}$. Therefore, the optimal objective value of problem (39) degrades from iteration k to $k+1$. We can use the same approach in the proof of Lemma 5.3 to upper bound the degradation in the optimal objective value. In particular, we can show that $f^k - f^{k+1} \leq \frac{1}{2}\sum_{q\in M} \mathbf{1}(i_k \in B_q)\,\theta_{i_k,q}$, where i_k is the product that we choose in Step 3 of the iterative rounding algorithm at iteration k. To see this result, we define the solution $(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}})$ to problem (39) at iteration $k+1$ as follows. Letting $(\boldsymbol{x}^k,\boldsymbol{y}^k)$ be an optimal solution to problem (39) at iteration k and i_k be the product that we choose in Step 3 of the iterative rounding algorithm at iteration k, we set $\tilde{x}_i = x_i^k$ for all $i \in \hat{N} \setminus \{i_k\}$, $\tilde{x}_{i_k} = \frac{1}{2}$ and $\tilde{x}_i = 0$ for all $i \notin \hat{N}$. Also, we set $\tilde{y}_q = \left[\sum_{i \in B_q} \tilde{x}_i - |B_q| + 1\right]^+$ for all $q \in \hat{M}$ and $\tilde{y}_q = 0$ for all $q \notin \hat{M}$. Using the same approach in the proof of Lemma 5.3, we can show that the solution $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ is feasible to problem (39) at iteration $k+1$. Furthermore, we can show that $y_q^k \geq \tilde{y}_q$ for all $q \in M$. In this case, since $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ is a feasible but not necessarily an optimal solution problem (39) at iteration $k+1$, we obtain

$$
\begin{aligned}
f^k - f^{k+1} \;&\leq\; f_k - \sum_{q \in M} \left(\mu_q \, \tilde{y}_q + \sum_{i \in B_q} \theta_{iq} \, \tilde{x}_i \right) \;=\; f_k - \sum_{q \in M} \mu_q \, \tilde{y}_q - \sum_{i \in N} \sum_{q \in M} \mathbf{1}(i \in B_q) \, \theta_{iq} \, \tilde{x}_i \\
&=\; \sum_{q \in M} \mu_q \, (y_q^k - \tilde{y}_q) \;+\; \sum_{q \in M} \mathbf{1}(i_k \in B_q) \, \theta_{i_k,q} \left(x_{i_k}^k - \frac{1}{2} \right) \;+ \sum_{i \in N \setminus \{i_k\}} \sum_{q \in M} \mathbf{1}(i \in B_q) \, \theta_{iq} \, (x_i^k - \tilde{x}_i) \\
&\leq\; \frac{1}{2} \sum_{q \in M} \mathbf{1}(i_k \in B_q) \, \theta_{i_k,q}.
\end{aligned}
$$

In the chain of inequalities above, the first equality holds since $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ is an optimal solution to problem (39) at iteration k. To see the last inequality, note that $\mu_q \leq 0$ and $y_q^k - \tilde{y}_q \geq 0$ for all $\in \hat{M}$, whereas we have $y_q^k = 0$ for all $q \notin \hat{M}$ by the definition of $\mathcal{P}(H)$ and $\tilde{y}_q = 0$ for all $q \notin \hat{M}$. Similarly, we have $\tilde{x}_i = x_i^k$ for all $i \in \hat{N} \setminus \{i_k\}$ and $x_i^k = 0 = \tilde{x}_i$ for all $i \notin \hat{N}$. Lastly, we have $x_{i_k}^k \leq 1$ and $i_k \in \hat{N}$ in the iterative rounding algorithm, so $\theta_{i_k,q} \geq 0$ for all $q \in M$ such that $i_k \in B_q$.

Building on the fact that $f^k - f^{k+1} \leq \frac{1}{2} \sum_{q \in M} \mathbf{1}(i_k \in B_q) \, \theta_{i,k_q}$, we can use the same approach in the proof of Lemma 5.4 to show that if $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to problem (39) at the last iteration of the iterative rounding algorithm, then $\sum_{q \in M} (\mu_q \, y_q^* + \sum_{i \in B_q} \theta_{iq} \, x_i^*) \geq \frac{1}{2} f^R$.

M.4.2. Feasible Subsets Through Coupled Randomized Rounding We let $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ be an optimal solution to problem (39) at the last iteration of the iterative rounding algorithm. At the last iteration, there is no product $i \in \hat{N}$ such that $\frac{1}{2} < x_i^* < 1$, in which case, by the analogue of Lemma 5.1 for the polyhedron $\mathcal{P}(H)$ in (38), we have $x_i^* \in \{0, \frac{1}{2}, 1\}$. We apply the coupled randomized rounding approach in Section 5.2 without any modifications to obtain a random subset of products $\hat{\boldsymbol{X}} = \{\hat{X}_i : i \in \hat{N}\}$. In this case, through minor modifications in the proof of Theorem 5.2, we can show that $\sum_{q \in M} \mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q} \, (R_q(\hat{\boldsymbol{X}}) - \hat{z})\} \geq \frac{1}{2} \sum_{q \in M} (\mu_q \, y_q^* + \sum_{i \in B_q} \theta_{iq} \, x_i^*)$. Therefore, noting the discussion at the end of the previous section, we get $\sum_{q \in M} \mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q} \, (R_q(\hat{\boldsymbol{X}}) - \hat{z})\} \geq \frac{1}{2} \sum_{q \in M} (\mu_q \, y_q^* + \sum_{i \in B_q} \theta_{iq} \, x_i^*) \geq \frac{1}{4} f^R$. We describe the modifications in the proof of Theorem 5.2. For each nest q, we will show that $\mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q} \, (R_q(\hat{\boldsymbol{X}}) - \hat{z})\} \geq \frac{1}{2} (\mu_q \, y_q^* + \sum_{i \in B_q} \theta_{iq} \, x_i^*)$.

By the construction of the coupled randomized rounding approach, $\mathbb{E}\{\hat{X}_i\} = x_i^*$. Noting (33), we have $V_q(\hat{\boldsymbol{X}})^{\gamma_q} \, (R_q(\hat{\boldsymbol{X}}) - \hat{z}) = \mu_q \prod_{i \in B_q} \hat{X}_i + \sum_{i \in B_q} \theta_{iq} \, \hat{X}_i$. If $|B_q| = 1$, then the definition of μ_q implies that $\mu_q = 0$, in which case, we obtain $\mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q} \, (R_q(\hat{\boldsymbol{X}}) - \hat{z})\} = \sum_{i \in B_q} \theta_{iq} \, x_i^* \geq \frac{1}{2} \sum_{i \in B_q} \theta_{iq} \, x_i^* =$

$\frac{1}{2}\,(\mu_q\, y_q^* + \sum_{i\in B_q} \theta_{iq}\, x_i^*)$. To see the inequality, note that if $i \in \hat{N}$, then $\theta_{iq} \geq 0$, but if $i \notin \hat{N}$, then $x_i^* = 0$. Thus, we have $\mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q}\,(R_q(\hat{\boldsymbol{X}}) - \hat{z})\} \geq \frac{1}{2}\,(\mu_q\, y_q^* + \sum_{i\in B_q} \theta_{iq}\, x_i^*)$ whenever $|B_q| = 1$. In the rest of the discussion, we assume that $|B_q| = 2$ and consider three cases.

Case 1: Suppose that the two products in B_q are paired in the coupled randomized rounding approach. Therefore, we have $\prod_{i\in B_q} \hat{X}_i = 0$. Since the products in B_q are paired, we have $x_i^* = \frac{1}{2}$ for all $i \in B_q$, in which case, noting problem (39), we must have $i \in \hat{N}$ for all $i \in B_q$. Thus, we obtain $q \in \hat{M}$. In this case, we have $\mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q}\,(R_q(\hat{\boldsymbol{X}}) - \hat{z})\} = \sum_{i\in B_q} \theta_{iq}\, x_i^* \geq \frac{1}{2}\,(\mu_q\, y_q^* + \sum_{i\in B_q} \theta_{iq}\, x_i^*)$, where the inequality holds because $\mu_q \leq 0$ and $\theta_{iq} \geq 0$ for all $i \in B_q$ for any nest $q \in \hat{M}$.

Case 2: Suppose that the two products in B_q are not paired in the coupled randomized rounding approach and $q \in \hat{M}$. In this case, $\{\hat{X}_i : i \in B_q\}$ are independent so that $\mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q}\,(R_q(\hat{\boldsymbol{X}}) - \hat{z})\} = \mu_q \prod_{i\in B_q} x_i^* + \sum_{i\in B_q} \theta_{iq}\, x_i^*$. Furthermore, since $q \in \hat{M}$, we have $\mu_q \leq 0$, in which case, $y_q^* = \big[\sum_{i\in B_q} x_i^* - |B_q| + 1\big]^+ = \big[\sum_{i\in B_q} x_i^* - 1\big]^+$. Noting that $\sum_{i\in B_q} x_i^* \leq 2$, if $x_j^* = 0$ for some $j \in B_q$, then we have $y_q^* = \big[\sum_{i\in B_q} x_i^* - 1\big]^+ = 0 = \prod_{i\in B_q} x_i^*$. Thus, it follows that

$$\begin{aligned}\mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q}\,(R_q(\hat{\boldsymbol{X}}) - \hat{z})\} \;&=\; \mu_q \prod_{i\in B_q} x_i^* + \sum_{i\in B_q} \theta_{iq}\, x_i^* \\ &\geq\; \frac{1}{2}(\mu_q \prod_{i\in B_q} x_i^* + \sum_{i\in B_q} \theta_{iq}\, x_i^*) \;=\; \frac{1}{2}\,(\mu_q\, y_q^* + \sum_{i\in B_q} \theta_{iq}\, x_i^*), \quad (40)\end{aligned}$$

where the inequality holds because if $q \in \hat{M}$, then we have $\rho_q \geq 0$ and $\theta_{iq} \geq 0$ for all $i \in B_q$, so get $\mu_q \prod_{i\in B_q} x_i^* + \sum_{i\in B_q} \theta_{iq}\, x_i^* = \rho_q \prod_{i\in B_q} x_i^* + \sum_{i\in B_q} \theta_{iq}\,\big(1 - \prod_{j\in B_q\setminus\{i\}} x_j^*\big)\, x_i^* \geq 0$. Similarly, if $x_j^* = 1$ for some $j \in B_q$, then we can show that $y_q^* = \big[\sum_{i\in B_q} x_i^* - 1\big]^+ = \prod_{i\in B_q} x_i^*$ and we can follow the same line of reasoning in (40) to get $\mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q}\,(R_q(\hat{\boldsymbol{X}}) - \hat{z})\} \geq \frac{1}{2}\,(\mu_q\, y_q^* + \sum_{i\in B_q} \theta_{iq}\, x_i^*)$. Lastly, if $x_i^* = \frac{1}{2}$ for all $i \in B_q$, then we have $y_q^* = \big[\sum_{i\in B_q} x_i^* - |B_q| + 1\big]^+ = \big[\sum_{i\in B_q} x_i^* - 1\big]^+ = 0$, so

$$\begin{aligned}\mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q}\,(R_q(\hat{\boldsymbol{X}}) - \hat{z})\} \;&=\; \mu_q \prod_{i\in B_q} x_i^* + \sum_{i\in B_q} \theta_{iq}\, x_i^* \;=\; \rho_q \prod_{i\in B_q} x_i^* + \sum_{i\in B_q} \theta_{iq} \left(1 - \prod_{j\in B_q\setminus\{i\}} x_j^*\right) x_i^* \\ &\geq\; \sum_{i\in B_q} \theta_{iq} \left(1 - \prod_{j\in B_q\setminus\{i\}} x_j^*\right) x_i^* \;=\; \frac{1}{2}\,(\mu_q\, y_q^* + \sum_{i\in B_q} \theta_{iq}\, x_i^*),\end{aligned}$$

where the last equality in the chain of inequalities above holds because we have $y_q^* = 0$ and $x_j^* = \frac{1}{2}$ for all $j \in B_q$, along with the fact that $|B_q| = 2$.

Case 3: Suppose that the two products in B_q are not paired in the coupled randomized rounding approach and $q \notin \hat{M}$. Since $q \notin \hat{M}$, noting problem (39), we get $y_q^* = 0$. Furthermore, since $q \notin \hat{M}$, we have $p_j < \hat{z}$ for some $j \in B_q$, in which case, $j \notin \hat{N}$. Therefore, we have $x_j^* = 0$, which implies that $\prod_{i\in B_q} x_i^* = 0$. In this case, we get $\mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q}\,(R_q(\hat{\boldsymbol{X}}) - \hat{z})\} = \mu_q \prod_{i\in B_q} x_i^* + \sum_{i\in B_q} \theta_{iq}\, x_i^* =$

$\sum_{i\in B_q}\theta_{iq}\,x_i^* \geq \frac{1}{2}\sum_{i\in B_q}\theta_{iq}\,x_i^* = \frac{1}{2}\,(\mu_q\,y_q^* + \sum_{i\in B_q}\theta_{iq}\,x_i^*)$, where the inequality holds because if $i\in\hat{N}$, then $\theta_{iq}\geq 0$, but if $i\notin\hat{N}$, then $x_i^*=0$. In all of the three cases considered above, we have $\mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q}\,(R_q(\hat{\boldsymbol{X}})-\hat{z})\}\geq\frac{1}{2}\,(\mu_q\,y_q^*+\sum_{i\in B_q}\theta_{iq}\,x_i^*)$, as desired. Therefore, the random subset of products $\hat{\boldsymbol{X}}$ satisfies $\sum_{q\in M}\mathbb{E}\{V_q(\hat{\boldsymbol{X}})^{\gamma_q}\,(R_q(\hat{\boldsymbol{X}})-\hat{z})\}\geq 0.25\,f^R$. Also, we have $\sum_{i\in N}\hat{X}_i\leq c$ by the construction of the coupled randomized rounding approach. As in Section 5.2, we can de-randomize the subset of products $\hat{\boldsymbol{X}}$ by using the method of conditional expectations to obtain a deterministic subset of products $\hat{\boldsymbol{x}}$ that satisfies $\sum_{q\in M}V_q(\hat{\boldsymbol{x}})^{\gamma_q}(R_q(\hat{\boldsymbol{x}})-\hat{z})\geq 0.25\,f^R$ and $\sum_{i\in N}\hat{x}_i\leq c$.

References

Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*. Springer, New York, NY.

Bodea, T., M. Ferguson, L. Garrow. 2009. Choice-based revenue management: Data from a major hotel chain. *Manufacturing & Service Operations Management* **11**(2) 356–361.

Cardell, N. S. 1997. Variance components structure for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory* **13**(2) 185–213.

de la Vega, W. Fernandez, M. Karpinski. 2006. Approximation complexity of nondense instances of max-cut. *Electronic Colloquium on Computational Complexity* Report No. 101.

Feldman, J. B., H. Topaloglu. 2017. Revenue management under the Markov chain choice model. *Operations Research* **65**(5) 1322–1342.

Garey, M. R., D. S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, New York, NY.

Goemans, M. X., D. P. Willamson. 1995. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM* **42**(6) 1115–1145.

Gunn, H., J. Bates. 1982. Statistical aspects of travel demand modeling. *Transportation Research Part A* **16A**(5-6) 371–382.

Hastie, T., R. Tibshirani, J. Friedman. 2017. *The Elements of Statistical Learning*. Springer, New York, NY.

Hausman, J., D. McFadden. 1984. Specification tests for the multinomial logit model. *Econometrica* **52**(5) 1219–1240.

Hochbaum, D. S. 1998. Instant recognition of half-integrality and 2-approximations. K. Jansen, J. Rolim, eds., *Approximation Algorithms for Combinatorial Optimization*. International Workshop APPROX'98, Aalborg, Denmark, 99–110.

Horowitz, J. 1982. Specification tests for probabilistic choice models. *Transportation Research Part A* **16A**(5-6) 383–394.

Kleinberg, J., E. Tardos. 2005. *Algorithm Design*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA.

Mahajan, S., H. Ramesh. 1999. Derandomizing approximation algorithms based on semidefinite programming. *SIAM Journal on Computing* **28**(5) 1641–1663.

McFadden, D. 1978. Modeling the choice of residential location. A. Karlqvist, L. Lundqvist, F. Snickars, J. Weibull, eds., Spatial Interaction Theory and Planning Models. North-Holland, Amsterdam, 75–96.

Meeker, W. Q., G. J. Hahn, L. A. Escobar. 2017. *Statistical Intervals*. Wiley Series in Probability and Statistics, Wiley, Hoboken, NJ.

Padberg, M. 1989. The boolean quadric polytope: Some characteristics, facets and relatives. *Mathematical Programming* **45**(1) 139–172.

van Ryzin, G., Gustavo Vulcano. 2015. A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Science* **61**(2) 281–300.

Vulcano, G., G. J. van Ryzin, R. Ratliff. 2012. Estimating primary demand for substitutable products from sales transaction data. *Operations Research* **60**(2) 313–334.