

# Electronic Companion to “A Feature-based Consideration Set Choice Model for Online Retailing”

Mohammad Amin Farzaneh and Ashwin Venkataraman

Naveen Jindal School of Management, The University of Texas at Dallas, Richardson, TX, 75080, {amin, ashwin.venkataraman}@utdallas.edu

Sajad Modaresi

Kenan-Flagler Business School, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, sajad\_modaresi@kenan-flagler.unc.edu

## Appendix A: EM Algorithm for FCM Estimation

Algorithm 1 presents the Expectation-Maximization (EM) algorithm for estimating the parameters of the FCM from data.

---

### Algorithm 1 EM Algorithm for FCM Estimation with $K$ Customer Types

---

1: **Input:** Data  $\{(j_t, \mathcal{C}_t, S_t, \mathbf{Z}_t)\}_{t=1}^T$ , number of customer types  $K$

2: **Initialize:**  $\alpha_0^{(k)}, \theta_0^{(k)}, \omega_0^{(k)}$  for all  $k \in [K]$ ; set  $i = 0$

3: **while** stopping condition is not met **do**

4:   **E-Step:** For each  $t \in [T]$  and  $k \in [K]$ , compute

$$\gamma_{tk}^{(i)} = \frac{\alpha_i^{(k)} \cdot \mathbb{P}(\mathcal{C}_t | S_t; \theta_i^{(k)}) \cdot \mathbb{P}_{\text{MNL}}(j_t | \mathcal{C}_t; \omega_i^{(k)})}{\sum_{\ell=1}^K \alpha_i^{(\ell)} \cdot \mathbb{P}(\mathcal{C}_t | S_t; \theta_i^{(\ell)}) \cdot \mathbb{P}_{\text{MNL}}(j_t | \mathcal{C}_t; \omega_i^{(\ell)})}$$

5:   **M-Step:** For each  $k \in [K]$ ,

6:     Update proportions:

$$\alpha_{i+1}^{(k)} = \frac{1}{T} \sum_{t=1}^T \gamma_{tk}^{(i)}$$

7:     Update consideration parameters:

$$\theta_{i+1}^{(k)} = \arg \max_{\theta} \sum_{t=1}^T \gamma_{tk}^{(i)} \cdot \log \mathbb{P}(\mathcal{C}_t | S_t; \theta)$$

8:     Update purchase parameters:

$$\omega_{i+1}^{(k)} = \arg \max_{\omega} \sum_{t=1}^T \gamma_{tk}^{(i)} \cdot \log \mathbb{P}_{\text{MNL}}(j_t | \mathcal{C}_t; \omega)$$

9:     Increment  $i \leftarrow i + 1$

10: **end while**

11: **Output:** Estimated parameters  $\{\alpha_i^{(k)}, \theta_i^{(k)}, \omega_i^{(k)}\}_{k=1}^K$

---

The updates in Steps 7 and 8 correspond to the weighted maximum log-likelihood estimation of two MNL models for the consideration and purchase stages; see (5) for the details of the reduction to MNL estimation in Step 7. These optimizations can be carried out efficiently considering the concavity of the log-likelihood function. The stopping condition is met when the relative change in the log-likelihood falls below a threshold for a predefined number of consecutive iterations.

## Appendix B: Characterizing Feasible Consideration Sets and Consideration Utilities

In this section, we first present an example illustrating the infeasibility of certain consideration sets. In Section B.2, we develop a set of structural results that characterize the feasible consideration set structures. Section B.3 formalizes the space of feasible consideration utilities via a set of linear constraints. Finally, in Section B.4, we establish a condition under which the objective function of the joint problem (18) exhibits strict concavity, enabling the use of efficient solution methods.

### B.1. Example: Infeasibility of Arbitrary Consideration Set Structures

Following the definition of consideration sets under D-FCM in (8), not all consideration set structures across customer types are feasible. The following example illustrates such an infeasibility:

**EXAMPLE EC.1 (INFEASIBLE CONSIDERATION SETS).** Suppose  $N = K = 2$ ,  $\beta_1 = \beta_1^{(1)} = \beta_1^{(2)}$  and  $\beta_2 = \beta_2^{(1)} = \beta_2^{(2)}$ , such that  $\beta_1 \neq \beta_2$ . Additionally,  $\theta^{(1)} \neq \theta^{(2)}$ . Then,  $\mathcal{C}^{(1)} = \mathcal{C}^{(2)} = \{1, 2\}$  is infeasible as follows:

$$\begin{aligned} \mathcal{C}^{(1)} = \{1, 2\} &\implies \bar{u}_1^{(1)} = \bar{u}_2^{(1)} \implies \beta_1 - \theta^{(1)}p_1 = \beta_2 - \theta^{(1)}p_2 \stackrel{(a)}{\implies} \theta^{(1)} = \frac{\beta_1 - \beta_2}{p_1 - p_2} \\ &\implies \theta^{(2)} \neq \frac{\beta_1 - \beta_2}{p_1 - p_2} \implies \beta_1 - \theta^{(2)}p_1 \neq \beta_2 - \theta^{(2)}p_2 \implies \bar{u}_1^{(2)} \neq \bar{u}_2^{(2)} \implies \mathcal{C}^{(2)} \neq \{1, 2\}, \end{aligned}$$

where in (a),  $p_1 - p_2 \neq 0$  because  $\beta_1 \neq \beta_2$ .

### B.2. Space of Feasible Consideration Set Structures

In this section, we leverage the structure induced by Assumption 1 to characterize the set of feasible consideration set structures across the customer types. First, w.l.o.g., we index products and customer types in decreasing order of consideration utilities and increasing order of price sensitivities, respectively. That is, we have that  $\beta_1 > \beta_2 > \dots > \beta_N$ , and  $\theta^{(1)} < \theta^{(2)} < \dots < \theta^{(K)}$ . Additionally, for each  $j \in [N]$ , we define  $S_j = \{j + 1, \dots, N\}$ , and  $S'_j = [j - 1] = \{1, \dots, j - 1\}$ . Note that by definition,  $S_N = S'_1 = \emptyset$ . The following results establish the necessary conditions for the structure of feasible consideration sets.

**Structural Result 1 (Feasibility).** If  $j \in \mathcal{C}^{(l)}$  for any  $l \in [K]$ , then  $S_j \cap \mathcal{C}^{(h)} = \emptyset$ , for any  $h < l$ , and  $S'_j \cap \mathcal{C}^{(k)} = \emptyset$ , for any  $k > l$ .

An immediate corollary of Structural Result 1 is that  $|\mathcal{C}^{(k)} \cap \mathcal{C}^{(k')}| \leq 1$  for any two types  $k \neq k'$ ; that is, two customer types can share at most one product in their consideration sets.

**Structural Result 2 (Feasibility).** If  $\mathcal{C}^{(l)} \neq \emptyset$  for any  $l \in [K]$ , then  $\mathcal{C}^{(k)} \neq \emptyset$  for any  $k < l$ . Conversely, if  $\mathcal{C}^{(l)} = \emptyset$  for any  $l \in [K]$ , then  $\mathcal{C}^{(k)} = \emptyset$  for any  $k > l$ .

To better understand the implications of the above results, we illustrate their application using a *consideration matrix*, where the inclusion or exclusion of products in the consideration sets of customer types are indicated with 1 and 0, respectively. Figure EC.1 shows two examples with  $N = 6$  products and  $K = 3$  types. The consideration matrix

in Figure EC.1 (left) represents the consideration sets  $\mathcal{C}^{(1)} = \{1, 3\}$ ,  $\mathcal{C}^{(2)} = \emptyset$ , and  $\mathcal{C}^{(3)} = \{2, 6\}$ . Because  $3 \in \mathcal{C}^{(1)}$ , Structural Result 1 states that it is infeasible to include product 2 in  $\mathcal{C}^{(3)}$ . Furthermore, since  $\mathcal{C}^{(2)} = \emptyset$ , Structural Result 2 indicates that it is infeasible to include any product in  $\mathcal{C}^{(3)}$ . Consequently, this consideration matrix cannot be feasible. In contrast, the consideration matrix in Figure EC.1 (right) does not violate any of the constraints in Structural Results 1 and 2, making it (potentially) feasible—it might still be infeasible because Structural Results 1 and 2 provide only necessary conditions for feasibility.

**Figure EC.1 Examples of Two Candidate Consideration Set Structures**

		Products					
		1	2	3	4	5	6
Customer types	$k = 1$	1	0	1	0	0	0
	$k = 2$	0	0	0	0	0	0
	$k = 3$	0	1	0	0	0	1

		Products					
		1	2	3	4	5	6
Customer types	$k = 1$	1	0	1	0	0	0
	$k = 2$	0	0	1	1	0	0
	$k = 3$	0	0	0	0	0	1

The next three results provide insights on the structure of the optimal consideration sets. Let  $\mathcal{C}_*^{(l)}$  denote the optimal consideration set of customer type  $l \in [K]$ .

**Structural Result 3 (Optimality).** Suppose  $\mathcal{C}_*^{(l)} \neq \emptyset$  and let  $a^{(l)} = \arg \min_{j \in \mathcal{C}_*^{(l)}} j$ , and  $b^{(l)} = \arg \max_{j \in \mathcal{C}_*^{(l)}} j$ , denote the lowest and highest index product in the optimal consideration set of customer type  $l \in [K]$ , respectively. If  $b^{(l)} - a^{(l)} > 1$ , then it must be the case that  $\{a^{(l)} + 1, \dots, b^{(l)} - 1\} \subset \mathcal{C}_*^{(l)}$ .

**Structural Result 4 (Optimality).** For any  $k \in [K - 1]$ , there are no two consecutive customer types  $k$  and  $k + 1$  such that  $a^{(k+1)} - b^{(k)} > 1$ , where  $a^{(l)}, b^{(l)}$  are as defined in Structural Result 3.

**Structural Result 5 (Optimality).** Product 1 is included in the consideration set of type 1, i.e.,  $1 \in \mathcal{C}_*^{(1)}$ .

Structural Result 3 states that, in any optimal solution to (18), the indices of products included in the consideration set of any customer type must be contiguous. This implies that the consideration matrix in Figure EC.1 (right) cannot be optimal, as product 2 is not included in  $\mathcal{C}^{(1)}$ , but products 1 and 3 are. Note that this result identifies only an optimality criterion; non-contiguous consideration sets might still be feasible. Structural Result 4 implies that in any optimal consideration matrix, there should not be any “gap” (sequence of zeros) between the largest index product in a row and the smallest index product in the next row. In the example in Figure EC.1 (right), the consideration matrix again cannot be optimal because the condition is violated between the second and the third row, corresponding to type 2 and 3 customers, respectively.

The proofs of these structural results are established as part of the proof of Theorem 4 in Appendix C.6.

### B.3. Feasible Region of Consideration Utilities

We term any consideration matrix that satisfies the Structural Results 1–5 in B.2, a “valid” consideration matrix. Given any valid consideration matrix, the consideration utilities  $\bar{u}$  of the customer types should be set in such a way that guarantees the inclusion and exclusion of the products in the respective consideration sets. We introduce four sets of constraints that ensure the proper assignment of the consideration utilities. The first set of constraints is analogous to constraint (16b), which ensure the non-negativity of the consideration utilities as well as product prices:

$$0 \leq \bar{u}^{(k)} \leq \beta_j \quad \forall j \in \mathcal{C}^{(k)}, \forall k \in [K] \text{ s.t. } \mathcal{C}^{(k)} \neq \emptyset. \quad (\text{B-1})$$

The second set of constraints is associated with products that are not included in the consideration set of a particular customer type. To ensure that product  $j$  is not included in the consideration set of type  $k$ , product  $j$ 's consideration utility should be smaller than  $\bar{u}^{(k)}$ ; that is,  $\bar{u}_j^{(k)} < \bar{u}^{(k)}$ . Let  $l_j \neq k$  be any customer type that includes product  $j$  in their consideration set, which implies that the price  $p_j$  of product  $j$  can be computed as  $(\beta_j - \bar{u}^{(l_j)})/\theta^{(l_j)}$ . Moreover, if  $j \notin \mathcal{C}^{(k')}$  for all  $k' \in [K]$ , then we can simply charge a very high price for product  $j$  that results in a negative consideration utility for all customer types, and solve the problem for the remaining products. With this, the constraints can be expressed as:

$$\beta_j - \theta^{(k)} \left( \frac{\beta_j - \bar{u}^{(l_j)}}{\theta^{(l_j)}} \right) < \bar{u}^{(k)} \implies \bar{u}^{(k)} > \left( \frac{\theta^{(k)}}{\theta^{(l_j)}} \right) \bar{u}^{(l_j)} + \left( \frac{\theta^{(l_j)} - \theta^{(k)}}{\theta^{(l_j)}} \right) \beta_j \quad \forall j \notin \mathcal{C}^{(k)} \quad \forall k \in [K] \text{ s.t. } \mathcal{C}^{(k)} \neq \emptyset. \quad (\text{B-2})$$

The third set of constraints focuses on products that belong to consideration sets of more than one customer type. In this case, we need to ensure that the resulting price of the shared product remains the same across all those customer types. Specifically, if customer types  $k$  and  $k'$  both include product  $j$  in their consideration sets, we must have that:

$$\begin{aligned} \frac{\beta_j - \bar{u}^{(k)}}{\theta^{(k)}} &= \frac{\beta_j - \bar{u}^{(k')}}{\theta^{(k')}} \implies \\ \bar{u}^{(k)} &= \left( \frac{\theta^{(k)}}{\theta^{(k')}} \right) \bar{u}^{(k')} + \left( \frac{\theta^{(k')} - \theta^{(k)}}{\theta^{(k')}} \right) \beta_j \quad \forall j \in \mathcal{C}^{(k)} \cap \mathcal{C}^{(k')}, \quad \forall k \neq k' \in [K] \text{ s.t. } \mathcal{C}^{(k)}, \mathcal{C}^{(k')} \neq \emptyset. \end{aligned} \quad (\text{B-3})$$

The last set of constraints corresponds to customer types with empty consideration sets. Although we can ignore the consideration utility  $\bar{u}^{(k)}$  for any customer type  $k \in [K]$  such that  $\mathcal{C}^{(k)} = \emptyset$  when solving the problem, we still need to ensure that the consideration utility of all products is negative for such customers. This leads to the following constraints:

$$\beta_j - \theta^{(k)} \left( \frac{\beta_j - \bar{u}^{(l_j)}}{\theta^{(l_j)}} \right) < 0 \implies \bar{u}^{(l_j)} < \left( \frac{\theta^{(k)} - \theta^{(l_j)}}{\theta^{(k)}} \right) \beta_j \quad \forall j \in [N], \forall k \in [K] \text{ s.t. } \mathcal{C}^{(k)} = \emptyset, \quad (\text{B-4})$$

where again  $l_j \neq k$  is any customer type that includes product  $j$  in their consideration set.

We make a few remarks regarding the above constraints. First, many of the constraints in (B-2) are redundant, and it suffices to enforce the constraints for only those combinations that are not deemed infeasible by Structural Result 1. This is because the infeasibility of a particular cell in the consideration matrix already implies that there is *no* consideration utility vector that can guarantee the inclusion of that product (column) in the consideration set of the corresponding customer type (row). Second, following Structural Result 2, constraint (B-4) is only required for the lowest-index customer type with an empty consideration set.

#### B.4. A Condition for Strict Concavity of the Objective Function in (18)

Let  $\mathcal{U}^{\mathcal{C}} \subset \mathcal{U}$  denote the feasible region for the consideration utilities associated with some (feasible) consideration set structure  $\mathcal{C}$ . The following result provides a sufficient condition for strict concavity of the objective function in (18) with respect to the consideration utilities:

**PROPOSITION EC.1.** *Given any feasible consideration set structure  $\mathcal{C}$ , suppose that for each  $k \in [K]$  with  $\mathcal{C}^{(k)} \neq \emptyset$ , there exists some product  $j \in \mathcal{C}^{(k)}$  such that  $\lambda_j^{(k)} - \beta_j \geq 0$ . Then, the objective function in (18) is strictly concave w.r.t.  $\bar{\mathbf{u}}$  over  $\mathcal{U}^{\mathcal{C}}$ .*

Proof. Let  $\gamma_j^{(k)} = \lambda_j^{(k)} - \beta_j$ , for all  $j \in [N]$  and all  $k \in [K]$ . The second derivative of the revenue function with respect to the consideration utility of customer type  $l \in [K]$  such that  $\mathcal{C}^{(l)} \neq \emptyset$  is as follows:

$$\begin{aligned}
& \frac{\partial^2}{\partial (\bar{u}^{(l)})^2} \left( \sum_{k \in [K]} \alpha^{(k)} \sum_{j \in \mathcal{C}^{(k)}} \left( \frac{\beta_j - \bar{u}^{(k)}}{\theta^{(k)}} \right) \frac{\exp(\gamma_j^{(k)} + \bar{u}^{(k)})}{1 + \sum_{i \in \mathcal{C}^{(k)}} \exp(\gamma_i^{(k)} + \bar{u}^{(k)})} \right) \\
&= \frac{\partial^2}{\partial (\bar{u}^{(l)})^2} \left( \alpha^{(l)} \sum_{j \in \mathcal{C}^{(l)}} \left( \frac{\beta_j - \bar{u}^{(l)}}{\theta^{(l)}} \right) \frac{\exp(\gamma_j^{(l)} + \bar{u}^{(l)})}{1 + \sum_{i \in \mathcal{C}^{(l)}} \exp(\gamma_i^{(l)} + \bar{u}^{(l)})} \right) \\
&= \alpha^{(l)} \sum_{j \in \mathcal{C}^{(l)}} \frac{-2}{\theta^{(l)}} \left( \frac{\exp(\gamma_j^{(l)} + \bar{u}^{(l)})}{\left(1 + \sum_{i \in \mathcal{C}^{(l)}} \exp(\gamma_i^{(l)} + \bar{u}^{(l)})\right)^2} \right) \\
&+ \alpha^{(l)} \sum_{j \in \mathcal{C}^{(l)}} \frac{\beta_j - \bar{u}^{(l)}}{\theta^{(l)}} \left( \frac{\exp(\gamma_j^{(l)} + \bar{u}^{(l)}) \cdot \left(1 - \left(\sum_{i \in \mathcal{C}^{(l)}} \exp(\gamma_i^{(l)} + \bar{u}^{(l)})\right)^2\right)}{\left(1 + \sum_{i \in \mathcal{C}^{(l)}} \exp(\gamma_i^{(l)} + \bar{u}^{(l)})\right)^4} \right) \\
&< 0,
\end{aligned}$$

where the last inequality follows because (i)  $\sum_{i \in \mathcal{C}^{(l)}} \exp(\gamma_i^{(l)} + \bar{u}^{(l)}) > 1$  as there exists some  $t \in \mathcal{C}^{(l)}$  such that  $\gamma_t^{(l)} \geq 0$  by hypothesis, and (ii)  $(\beta_j - \bar{u}^{(l)})/\theta^{(l)} \geq 0$ , for all  $j \in \mathcal{C}^{(l)}$  from constraint (B-1).

Because the objective function in (18) is (additively) separable in the consideration utilities  $(\bar{u}^{(k)}: k \in [K])$ , the result follows.  $\square$

Proposition EC.1 shows that as long as, for each customer type, there is at least one product in the consideration set for which the non-price purchase utility does not decrease relative to the non-price consideration utility, the revenue function in (18) exhibits strict concavity. Returning to the example in Section 1, this condition will be satisfied as long as customers' perceived utility from the additional information visible on the product page does not decrease, for some product in the consideration set—given that there is much more information available on the product pages (i.e., at the purchasing stage) relative to the search page (i.e., the consideration stage), it is reasonable to imagine that this assumption holds for at least one product for each customer type. Considering that the revenue function is differentiable, Proposition EC.1 implies that the optimal consideration utilities can be obtained by solving a concave program with linear constraints. The Frank-Wolfe (or conditional gradient) algorithm (Jaggi 2013) is ideally suited for solving such optimization problems.

## Appendix C: Proof of Technical Results

### C.1. Proof of Theorem 1

It is well known that the assortment optimization problem under the LC-MNL model with  $K = 2$  customer types is NP-complete (Rusmevichientong et al. 2014). Given the structural similarity between our assortment problem in (7) and the one under LC-MNL, the NP-completeness of our problem can be established via a reduction from the LC-MNL assortment problem.

Under the LC-MNL model, the utility of each product  $j \in [N]$  for a customer of type  $k \in \{1, 2\}$  is given by:

$$U_j^{(k)} = v_j^{(k)} + \xi_j^{(k)},$$

where  $v_j^{(k)}$  is the deterministic component of the utility, and  $\xi_j^{(k)}$  is a Gumbel-distributed random variable. The deterministic component of the no-purchase option is set to 0 for all  $k \in \{1, 2\}$ , so its utility is given by  $U_0^{(k)} = \xi_0^{(k)}$ , where  $\xi_0^{(k)}$  is also a Gumbel-distributed random variable. Each product  $j \in [N]$  is associated with a price  $p_j$ , and each customer type  $k \in \{1, 2\}$  has a mixing proportion  $\lambda^{(k)}$ , with  $\lambda^{(1)} + \lambda^{(2)} = 1$ . The assortment optimization problem under the LC-MNL model is then formulated as:

$$\arg \max_{S \subseteq [N]} \sum_{k \in \{1, 2\}} \lambda^{(k)} \sum_{j \in S} p_j \frac{\exp(v_j^{(k)})}{1 + \sum_{i \in S} \exp(v_i^{(k)})}. \quad (\text{C-5})$$

We construct an instance of our problem with  $N$  products and  $K = 2$  customer types, where each type  $k \in \{1, 2\}$  has proportion  $\alpha^{(k)} = \lambda^{(k)}$ . We consider 5 features, with only the first three displayed in the consideration stage. The first feature corresponds to price. We define the feature vector of each product  $j \in [N]$  as  $\mathbf{z}_j = (2p_j, 2p_j + v_j^{(1)}, 2p_j + v_j^{(2)}, -v_j^{(1)}, -v_j^{(2)})$ , which implies that the price of each product is  $2p_j$ . For the feature preference parameters, we set

$$\boldsymbol{\theta}^{(1)} = \boldsymbol{\omega}^{(1)} = (-1, 1, 0, 1, 0), \quad \text{and} \quad \boldsymbol{\theta}^{(2)} = \boldsymbol{\omega}^{(2)} = (-1, 0, 1, 0, 1).$$

This construction ensures that  $\bar{u}_j^{(k)} = v_j^{(k)}$ ,  $u_j^{(k)} = 0$ , and  $b_j^{(k)} = p_j$  for all  $j \in [N]$  and  $k \in \{1, 2\}$ . Hence, in this instance, the assortment optimization problem under our model is equivalent to the assortment optimization problem under the LC-MNL model, as formulated in (C-5).

Note that the same proof extends to show the equivalence between the assortment problem under FCM and LC-MNL when the customer population consists of  $K$  types. For this case, we construct an instance with  $N$  products,  $K$  customer types, and  $2K + 1$  features, where only the first  $K + 1$  features are displayed in the consideration stage. The feature vector for each product  $j \in [N]$  is set as:

$$\mathbf{z}_j = (2p_j, 2p_j + v_j^{(1)}, \dots, 2p_j + v_j^{(K)}, -v_j^{(1)}, \dots, -v_j^{(K)}).$$

For each customer type  $k \in [K]$ , we set  $\boldsymbol{\theta}^{(k)} = \boldsymbol{\omega}^{(k)}$  as a vector where the first element equals -1, the  $(k + 1)$ -th element equals 1, and the  $(K + k + 1)$ -th element also equals 1 (all other entries are zero). This construction ensures that  $\bar{u}_j^{(k)} = v_j^{(k)}$ ,  $u_j^{(k)} = 0$ , and  $b_j^{(k)} = p_j$ , for all  $j \in [N]$  and  $k \in [K]$ . Hence, the assortment optimization problem under our model is equivalent to the LC-MNL formulation with  $K$  customer types. □

## C.2. Proof of Proposition 1

The assortment optimization problem under the D-FCM with  $K = 1$  can be written as:

$$\begin{aligned} \arg \max_{S \subseteq [N]} \quad & \sum_{j \in \mathcal{C}} p_j \cdot \frac{\exp(u_j)}{1 + \sum_{i \in \mathcal{C}} \exp(u_i)} \\ \text{s.t. } \mathcal{C} = \quad & \{j \in S \mid \bar{u}_j \geq \max(\bar{u}_{(L,S)}, 0)\}. \end{aligned} \quad (\text{C-6})$$

If all products have distinct consideration utilities, then for any assortment  $S$ , the consideration set  $\mathcal{C}$  comprises the (up to)  $L$  products in  $S$  with the highest consideration utilities, while the remaining products do not contribute to the revenue—here, w.l.o.g., we assume that  $\bar{u}_j \geq 0$  for all  $j \in [N]$  because any products with negative consideration utilities can simply be dropped from the analysis. Therefore, it suffices to restrict attention to subsets  $S$  of size at most  $L$ . In that case, all offered products are included in the consideration set, and problem (C-6) simplifies to:

$$\arg \max_{S \subseteq [N], |S| \leq L} \sum_{j \in S} p_j \cdot \frac{\exp(u_j)}{1 + \sum_{i \in S} \exp(u_i)}, \quad (\text{C-7})$$

which is exactly the assortment optimization problem under the MNL model with a cardinality constraint of  $L$ .

### C.3. Proof of Theorem 2

Motivated by the similarity between the assortment problem under the FCM and that under the LC-MNL model with consideration sets, we adapt the proof technique of Bront et al. (2009) to our setting. Theorem 1 in Bront et al. (2009) establishes the NP-hardness of the following special case of 0-1 fractional programming problem:

$$\max_{\mathbf{y} \in \{0,1\}^N} \sum_{j=1}^N w_j y_j \left( \sum_{m=1}^M \frac{\gamma_m v_{mj}}{\sum_{i \in C_m} v_{mi} y_i + v_{m0}} \right), \quad (\text{C-8})$$

where  $\mathbf{y} = (y_j : j \in [N])$  is a binary vector,  $w_j > 0$  for all  $j \in [N]$ ,  $v_{mj} > 0$  and  $v_{m0} > 0$  for all  $j \in [N]$  and all  $m \in [M]$ , and  $C_m \subset [N]$  for all  $m \in [M]$ . The above problem is shown to be NP-hard via a polynomial-time reduction from the *minimum vertex cover* problem, as discussed next.

Consider an instance of the minimum vertex cover problem with graph  $G = (V, E)$ , where  $V = \{1, \dots, v\}$  represents the nodes, and  $E \subset \{(i, j) \in V \times V, i < j\}$  represents the edges. Bront et al. (2009) reduce the given instance of the minimum vertex cover problem to the following instance of problem (C-8). The binary decision variables are  $y_1, \dots, y_v, y_{v+1}$ , and the objective function consists of  $M = |E| + |V|$  summands. For each edge  $(i, j) \in E$ , the set  $C_a = \{i, j\}$  is defined for  $a = 1, \dots, |E|$ , while for each node  $i \in V$ , the set  $C_{|E|+i} = \{v+1, i\}$ . Next,  $w_j$  is set to 1 for  $1 \leq j \leq v$ , and  $w_{v+1} = 2$ . Furthermore,  $\gamma_m = 1$ ,  $v_{m0} = 1/v^2$ , and finally  $v_{mj} = \mathbb{1}\{j \in V \cap C_m\}$  for all  $m \in [M]$ . The authors show that a solution to the 0-1 fractional programming problem for the constructed instance can be converted to a solution for the minimum vertex cover problem, thereby establishing the desired reduction and proving NP-hardness.

Below, we outline how a similar reduction can be adopted in our context. Let the binary variable  $y_j$  represent the presence or absence of product  $j$  in the offered assortment, indicated by 1 and 0, respectively. The assortment optimization problem under FCM with  $M$  customer types and  $L = 2$  can then be expressed as:

$$\begin{aligned} \max_{\mathbf{y} \in \{0,1\}^N} \quad & \sum_{j=1}^N p_j \left( \sum_{m=1}^M \frac{\alpha^{(m)} \cdot \mathbb{1}[j \in \mathcal{C}^{(m)}] \cdot \exp(u_j^{(m)})}{1 + \sum_{i \in \mathcal{C}^{(m)}} \exp(u_i^{(m)})} \right), \\ \text{s.t. } \mathcal{C}^{(m)} = \quad & \left\{ j \in [N] : y_j = 1, \bar{u}_j^{(m)} \geq \max(\bar{u}_{(2;S(\mathbf{y}))}^{(m)}, 0) \right\}. \end{aligned} \quad (\text{C-9})$$

where  $S(\mathbf{y})$  denotes the assortment corresponding to vector  $\mathbf{y}$ .

Given the instance  $(V, E)$  of the minimum vertex cover problem described above, we create an instance of the D-FCM as follows. We have  $N = v + 1$  products,  $M = |E| + v$  customer types. We set the type proportions as  $\alpha^{(m)} = 1/M$  for each customer type  $m \in [M]$ . We set the prices as follows:  $p_j = 1$  for all  $j = 1, \dots, v$  and  $p_{v+1} = 2$ . Next, for each edge  $(i, j) \in E$  and corresponding type  $k \in \{1, \dots, |E|\}$ , we set  $\bar{u}_i^{(k)} \geq 0$ ,  $\bar{u}_j^{(k)} \geq 0$  and  $\bar{u}_r^{(k)} < 0$  for any  $r \in [N] \setminus \{i, j\}$ . This setup ensures that only products  $\{i, j\}$  can be included in the consideration set of type  $k$  customers. Further, for each  $t \in [N]$ , we set  $\exp(u_t^{(k)}) = v^2 \cdot \mathbb{1}[t \in \{i, j\}]$ . In a similar vein, for each customer type  $k = |E| + i$ , where  $i \in \{1, \dots, v\}$ , we set  $\bar{u}_{v+1}^{(k)} \geq 0$ ,  $\bar{u}_i^{(k)} \geq 0$  and  $\bar{u}_r^{(k)} < 0$  for any  $r \in [N] \setminus \{v+1, i\}$ , which again guarantees that only products  $\{v+1, i\}$  can be included in the consideration set of type  $|E| + i$  customers. Finally, for each  $t \in [N]$ , we set  $\exp(u_t^{(k)}) = v^2 \cdot \mathbb{1}[t \in \{v+1, i\}]$ .

It is not too hard to verify that for the above setting of the problem parameters, problem (C-9) is identical to the particular instance of problem (C-8) constructed by Bront et al. (2009). The claim then follows.

Finally, we note that in the above proof, we specified only the values of the consideration utility  $\bar{u}_j^{(m)}$  and purchase utility  $u_j^{(m)}$  for each product  $j \in [N]$  and customer type  $m \in [M]$ , without explicitly defining the underlying product

feature vectors  $z_j$  or the corresponding feature preference parameters  $\theta^{(m)}, \omega^{(m)}$  for the consideration and purchase stages. This was done to keep the proof succinct and avoid excessive notation. It is a straightforward but tedious exercise to construct the underlying product feature vectors and feature preference parameters of customer types that result in the aforementioned utility values. See the proof of Theorem 1 in Appendix C.1 for an illustration.  $\square$

#### C.4. Proof of Lemma 1

Below, we demonstrate the log-concavity of the revenue function in (16a), for any fixed consideration set. Consider the second derivative of the natural logarithm of the revenue function:

$$\begin{aligned} & \frac{d^2}{d\bar{u}^2} \log \left( \sum_{j \in \mathcal{C}} \left( \frac{\beta_j - \bar{u}}{\theta} \right) \frac{\exp(\lambda_j - \beta_j + \bar{u})}{1 + \sum_{i \in \mathcal{C}} \exp(\lambda_i - \beta_i + \bar{u})} \right) \\ &= \frac{-(\sum_{i \in \mathcal{C}} \exp(\lambda_i - \beta_i + \bar{u}))^2}{\sum_{i \in \mathcal{C}} (\exp(\lambda_i - \beta_i + \bar{u}))^2} - \frac{\sum_{i \in \mathcal{C}} \exp(\lambda_i - \beta_i + \bar{u})}{(1 + \sum_{i \in \mathcal{C}} \exp(\lambda_i - \beta_i + \bar{u}))^2} < 0 \quad \forall \bar{u} \geq 0, \end{aligned}$$

which implies the log-concavity of the revenue function. The claim then follows because log-concavity implies quasi-concavity, and in  $\mathbb{R}$ , quasi-concavity implies unimodality (Uhrin 1984).  $\square$

#### C.5. Proof of Lemma 2

First, for any given consideration utility  $\bar{u} \geq 0$ , we identify the collection of products that satisfy constraint (16b), and therefore, can be included in the consideration set. Denote this set as  $S(\bar{u}) := \{j \in [N] \mid \beta_j \geq \bar{u}\}$ , which implies that  $\mathcal{C} \subseteq S(\bar{u})$ . The reformulated joint optimization problem (16a) then reduces to:

$$\arg \max_{\mathcal{C} \subseteq S(\bar{u})} \sum_{j \in \mathcal{C}} \left( \frac{\beta_j - \bar{u}}{\theta} \right) \frac{\exp(\lambda_j - \beta_j + \bar{u})}{1 + \sum_{i \in \mathcal{C}} \exp(\lambda_i - \beta_i + \bar{u})}.$$

The problem above is equivalent to the assortment optimization problem under an MNL model, where the price and mean utility associated with each product  $j \in S(\bar{u})$  are given by  $(\beta_j - \bar{u})/\theta$  and  $\lambda_j - \beta_j + \bar{u}$ , respectively.  $\square$

#### C.6. Proof of Theorem 4

Below, we first prove the structural results presented in Section B.2, and then show how they imply that the search space of consideration sets reduces to  $\mathcal{O}(N^K)$ .

**C.6.1. Proof of Structural Result 1. Part 1.** Note that the result is trivially true if  $j = N$ . So, suppose  $j < N$ . Since  $j \in \mathcal{C}^{(l)}$  and following the definition of consideration set in (8), it follows that the consideration utility of product  $j$  for customer type  $l$  is at least that of product  $t$  for any  $t \in S_j$ :

$$\begin{aligned} \bar{u}_j^{(l)} \geq \bar{u}_t^{(l)} &\implies \beta_j - \theta^{(l)} p_j \geq \beta_t - \theta^{(l)} p_t \\ &\implies \beta_j - \beta_t \geq \theta^{(l)} (p_j - p_t). \end{aligned}$$

Note that by definition  $\beta_j > \beta_t$ . Now, for any  $h < l$ , it follows that:

Case 1: If  $p_j - p_t > 0$ ,

$$\begin{aligned} \frac{\beta_j - \beta_t}{p_j - p_t} &\geq \theta^{(l)} > \theta^{(h)} \\ \implies \beta_j - \beta_t &> \theta^{(h)} p_j - \theta^{(h)} p_t \\ \implies \bar{u}_j^{(h)} &> \bar{u}_t^{(h)}. \end{aligned}$$

Case 2: If  $p_j - p_t \leq 0$ ,

$$\begin{aligned}\bar{u}_j^{(h)} - \bar{u}_t^{(h)} &= \beta_j - \beta_t + \theta^{(h)}(p_t - p_j) > 0 \\ \implies \bar{u}_j^{(h)} &> \bar{u}_t^{(h)}.\end{aligned}$$

From cases 1 and 2, it can be concluded that for any  $h < l$  and any  $t \in S_j$ , the consideration utility of product  $t$  is strictly lower than that of product  $j$ ; hence, not included in the consideration set, i. e.,  $t \notin \mathcal{C}^{(h)}$ .

**Part 2.** Again, the result is trivially true for  $j = 1$ . So, we suppose  $j > 1$ . Note that based on the definition,  $\beta_j < \beta_q$ , for any  $q \in S'_j$ . Then, it must be the case that  $p_j - p_q < 0$ . Otherwise,

$$p_j \geq p_q \implies \beta_j - \theta^{(l)}p_j \leq \beta_j - \theta^{(l)}p_q < \beta_q - \theta^{(l)}p_q \implies \bar{u}_j^{(l)} < \bar{u}_q^{(l)},$$

which contradicts the fact that product  $j \in \mathcal{C}^{(l)}$ .

Now, because  $j \in \mathcal{C}^{(l)}$ , it must follow that:

$$\begin{aligned}\bar{u}_j^{(l)} \geq \bar{u}_q^{(l)} &\implies \beta_j - \theta^{(l)}p_j \geq \beta_q - \theta^{(l)}p_q \\ &\implies \beta_j - \beta_q \geq \theta^{(l)}(p_j - p_q) \\ &\stackrel{(a)}{\implies} \frac{\beta_j - \beta_q}{p_j - p_q} \leq \theta^{(l)} < \theta^{(k)} \\ &\implies \beta_j - \beta_q > \theta^{(k)}(p_j - p_q) \\ &\implies \bar{u}_j^{(k)} > \bar{u}_q^{(k)},\end{aligned}$$

where (a) follows as  $p_j < p_q$ . This implies that for any  $k > l$  and any  $q \in S'_j$ , the consideration utility of product  $q$  is strictly lower than that of product  $j$ ; hence, not included in the consideration set, i.e.,  $q \notin \mathcal{C}^{(k)}$ .  $\square$

**C.6.2. Proof of Structural Result 2.** For any  $l \in [K]$ , suppose  $j \in \mathcal{C}^{(l)}$  for some  $j \in [N]$ . From the definition of the consideration set (8), it follows that  $\bar{u}_j^{(l)} \geq 0$ . Then, it follows that  $\bar{u}_j^{(k)} = \beta_j - \theta^{(k)}p_j \geq \beta_j - \theta^{(l)}p_j = \bar{u}_j^{(l)} \geq 0$ , for any  $k < l$ , because  $\theta^{(k)} < \theta^{(l)}$ . This implies that at least the consideration utility of product  $j$  is non-negative for any customer type  $k < l$ . Consequently, the consideration set of type  $k < l$  is non-empty.

In contrast, if  $\mathcal{C}^{(l)} = \emptyset$ , it implies that  $\bar{u}_j^{(l)} < 0$ , for all  $j \in [N]$ . Consequently, for any  $k > l$ , it follows that  $\bar{u}_j^{(k)} = \beta_j - \theta^{(k)}p_j \leq \beta_j - \theta^{(l)}p_j < 0$ , for any  $j \in [N]$ , because now  $\theta^{(k)} > \theta^{(l)}$ . Therefore, it can be concluded that  $\mathcal{C}^{(k)} = \emptyset$ , for any  $k > l$ .  $\square$

**C.6.3. Proof of Structural Result 3.** First, we introduce the required notation. Let  $\bar{u}_*^{(k)}$  and  $\bar{u}_{j,*}^{(k)}$  represent the consideration utility of the consideration set and the consideration utility of product  $j \in [N]$ , associated with customer type  $k \in [K]$ , in the optimal solution. Note that for any  $k \in [K]$ ,  $\bar{u}_{j,*}^{(k)} = \bar{u}_*^{(k)}$ , for all  $j \in \mathcal{C}_*^{(k)}$ . Below, we prove the result by contradiction.

Suppose that in the optimal consideration set of some customer type  $l \in [K]$ , there exists some product  $m \in \{a^{(l)} + 1, a^{(l)} + 2, \dots, b^{(l)} - 1\}$  such that  $m \notin \mathcal{C}_*^{(l)}$ . For the universe of products being  $\{a^{(l)}, \dots, b^{(l)}\}$ , it directly follows from Lemma (2) that while maintaining the consideration utility of  $\bar{u}_*^{(l)}$ , one of the  $\beta$ -ordered consideration sets among  $\{a^{(l)}\}$ ,  $\{a^{(l)}, a^{(l)} + 1\}, \dots$ , or  $\{a^{(l)}, a^{(l)} + 1, \dots, b^{(l)}\}$  yields a strictly higher revenue for customer type  $l$ , compared to  $\mathcal{C}_*^{(l)}$ —the revenue is strictly higher based on the analysis in Talluri and Van Ryzin (2004). Let this consideration set be denoted with  $\mathcal{C}_{**}^{(l)}$ , and let  $\Delta^{(l)} = R^{(l)}(\mathcal{C}_{**}^{(l)}, \bar{u}_*^{(l)}) - R^{(l)}(\mathcal{C}_*^{(l)}, \bar{u}_*^{(l)}) > 0$ , represent the incremental revenue obtained

from customer type  $l$ , due to the transition from  $\mathcal{C}_*^{(l)}$  to  $\mathcal{C}_{**}^{(l)}$ . Note that this itself does not immediately lead to a contradiction as by making this change, the revenue from other types might also be impacted. However, we will show that while increasing the obtained revenue from customer type  $l$  by modifying its consideration set to  $\mathcal{C}_{**}^{(l)}$ , the total revenue obtained from other types either does not change or decreases by an amount strictly smaller than  $\Delta^{(l)}$ . In either scenario, this contradicts the optimality of the original consideration set  $\mathcal{C}_*^{(l)}$ .

We split the rest of the proof into two cases depending on the structure of the new consideration set  $\mathcal{C}_{**}^{(l)}$ . In both cases, to facilitate the comparison of the revenues obtained from customer types other than  $l$ , we will adjust product prices to keep their consideration sets the same as in the original optimal solution. In particular, these price adjustments will only be limited to the products in the set  $\{a^{(l)}, \dots, b^{(l)} - 1\}$ .

**Case 1:**  $\mathcal{C}_{**}^{(l)} = \{a^{(l)}, \dots, b^{(l)}\}$ .

To construct  $\mathcal{C}_{**}^{(l)}$ , it suffices to set the price of any product  $j \in \{a^{(l)} + 1, \dots, b^{(l)} - 1\} \setminus \mathcal{C}_*^{(l)}$  equal to  $(\beta_j - \bar{u}_*^{(l)})/\theta^{(l)}$ , so that it obtains a consideration utility of  $\bar{u}_*^{(l)}$ , which is the same as the consideration utility of the products of  $\mathcal{C}_*^{(l)}$ . Note that the proposed prices are non-negative:

$$\frac{\beta_j - \bar{u}_*^{(l)}}{\theta^{(l)}} > \frac{\beta_{b^{(l)}} - \bar{u}_*^{(l)}}{\theta^{(l)}} \geq 0, \quad \forall j \in \{a^{(l)} + 1, \dots, b^{(l)} - 1\} \setminus \mathcal{C}_*^{(l)},$$

as  $\beta_{a^{(l)}} > \beta_{a^{(l)}+1} > \dots > \beta_{b^{(l)}}$ . Since  $a^{(l)}, b^{(l)} \in \mathcal{C}_*^{(l)} \cap \mathcal{C}_{**}^{(l)}$ , it follows from Structural Result 1 that it is infeasible for any product in set  $\{a^{(l)} + 1, \dots, b^{(l)} - 1\}$  to be in the consideration set of any customer type other than  $l$ , both in the original optimal solution and after the price adjustments to construct  $\mathcal{C}_{**}^{(l)}$ . This implies that the inclusion of any product  $j \in \{a^{(l)} + 1, \dots, b^{(l)} - 1\} \setminus \mathcal{C}_*^{(l)}$  into the consideration set of customer type  $l$  does not change the consideration sets of other customer types. Therefore, the revenue obtained from customer type  $l$  increases by  $\Delta^{(l)}$ , while the revenue obtained from other customer types does not change. This contradicts the optimality of  $\mathcal{C}_*^{(l)}$ .

**Case 2:**  $\mathcal{C}_{**}^{(l)} = \{a^{(l)}, \dots, t\}$ , where  $t < b^{(l)}$ .

Below, we first discuss how to construct  $\mathcal{C}_{**}^{(l)}$  and how its construction impacts the consideration sets of other customer types. Next, we analyze the differences in the total revenue obtained in the original optimal solution and the modified solution. We introduce some additional notations. Let  $\hat{p}_j = (\beta_j - \bar{u}_*^{(l)})/\theta^{(l)}$  denote the transition price of product  $j \in \{a^{(l)}, \dots, t\}$ , which is the price that maintains the consideration utility of  $\bar{u}_*^{(l)}$ . Note that for products in the set  $\{a^{(l)}, \dots, t\} \cap \mathcal{C}_*^{(l)}$ , the transition price matches the product's price in the original optimal solution. Additionally, let  $\hat{u}_j^{(k)}$ , denote the transition utility of product  $j \in \{a^{(l)}, \dots, t\}$  for customer type  $k \in [K]$ , defined as:

$$\hat{u}_j^{(k)} = \beta_j - \theta^{(k)} \hat{p}_j = \beta_j - \theta^{(k)} \left( \frac{\beta_j - \bar{u}_*^{(l)}}{\theta^{(l)}} \right).$$

**1) Construction of  $\mathcal{C}_{**}^{(l)}$ .** To construct  $\mathcal{C}_{**}^{(l)}$ , we first set the price of each product  $j \in \{a^{(l)}, \dots, t\} \setminus \mathcal{C}_*^{(l)}$  to its transition price, ensuring that it is included in the consideration set of customer type  $l$ . However,  $\mathcal{C}_{**}^{(l)}$  is not fully formed yet, as at least product  $b^{(l)}$  remains in the consideration set of customer type  $l$ . To address this, we slightly reduce the prices of products in the set  $\{a^{(l)}, \dots, t\}$  by  $\delta/\theta^{(l)}$ . Indeed, the new price of product  $j \in \{a^{(l)}, \dots, t\}$  is equal to  $(\beta_j - (\bar{u}_*^{(l)} + \delta))/\theta^{(l)}$ . This adjustment ensures that the products in the set  $\{a^{(l)}, \dots, t\}$  achieve the consideration utility of  $\bar{u}_*^{(l)} + \delta$ , which exceeds the consideration utility of products in  $\{t + 1, \dots, b^{(l)}\}$ . Below, we show that for a properly chosen value of  $\delta > 0$ , the new prices of products in  $\mathcal{C}_{**}^{(l)}$  are non-negative.

*Non-negativity of new prices:* Let

$$P := \min_{j \in \{a^{(l)}, \dots, t\}} \frac{\beta_j - \bar{u}_*^{(l)}}{\theta^{(l)}} = \frac{\beta_t - \bar{u}_*^{(l)}}{\theta^{(l)}}$$

be the minimum transition price across the products in the set  $\{a^{(l)}, \dots, t\}$ , which corresponds to product  $t$  as  $\beta_{a^{(l)}} > \beta_{a^{(l)+1}} > \dots > \beta_t$ . Note that since  $b^{(l)} \in \mathcal{C}_*^{(l)}$ , it follows that:

$$P = \frac{\beta_t - \bar{u}_*^{(l)}}{\theta^{(l)}} > \frac{\beta_{b^{(l)}} - \bar{u}_*^{(l)}}{\theta^{(l)}} \geq 0.$$

To ensure the non-negativity of the new prices,  $\delta$  should be chosen such that:

$$\min_{j \in \{a^{(l)}, \dots, t\}} \frac{\beta_j - (\bar{u}_*^{(l)} + \delta)}{\theta^{(l)}} = \frac{\beta_t - (\bar{u}_*^{(l)} + \delta)}{\theta^{(l)}} = P - \frac{\delta}{\theta^{(l)}} > 0,$$

which holds true for any  $\delta < P\theta^{(l)}$ .

Next, let  $l' = \min \{k \in [K] : a^{(l)} \in \mathcal{C}_*^{(k)}\}$  denote the lowest index customer type that includes product  $a^{(l)}$  in its consideration set in the original optimal solution. Additionally, for any product  $j \in [N]$  and customer type  $k \in [K]$ , let  $\bar{u}_{j,**}^{(k)}$  denote the consideration utility of product  $j$  for customer type  $k$  after all the price adjustments mentioned above. Similarly, let  $\bar{u}_{**}^{(k)}$  represent the consideration utility of customer type  $k \in [K]$  following the price adjustments. Below, we provide two upper bounds for the value of  $\delta$  that need to be satisfied in order to keep the consideration sets of the customer types in  $[K] \setminus \{l', \dots, l-1\}$  the same as in the original optimal solution. The discussion on the consideration sets of customer types belonging to the set  $\{l', \dots, l-1\}$  is forwarded to the next part of the proof where we analyze the differences in the total revenue.

*Consistent consideration set of customer type  $m$ , where  $m > l$ :* Note if  $l = K$ , we can skip the discussion below. Because  $t \in \mathcal{C}_{**}^{(l)}$ , none of the products of the set  $\{a^{(l)}, \dots, t-1\}$  can be part of the consideration set of any customer type  $m > l$ . However, we need to ensure that price adjustments do not result in the inclusion of product  $t$  in the consideration set of customer type  $m$ . Since  $b^{(l)} \in \mathcal{C}_*^{(l)}$  in the original optimal solution, it follows from Part 2 of the proof of Structural Result 1 that under transition prices:

$$\bar{u}_{b^{(l)},*}^{(m')} - \hat{u}_t^{(m')} > 0, \quad \forall m' \in \{l+1, \dots, K\}.$$

Let

$$U := \min_{m' \in \{l+1, \dots, K\}} \bar{u}_{b^{(l)},*}^{(m')} - \hat{u}_t^{(m')} > 0,$$

represents the minimum difference between the consideration utility of product  $b^{(l)}$  in the original optimal solution, and the consideration utility of product  $t$  under transition prices, across all customer types  $m' > l$ . To avoid the inclusion of product  $t$  into the consideration set of customer type  $m$  following the price adjustments, it suffices to choose  $\delta$  such that:

$$\begin{aligned} \bar{u}_{b^{(l)},**}^{(m)} - \bar{u}_{t,**}^{(m)} > 0 &\iff \bar{u}_{b^{(l)},*}^{(m)} - \bar{u}_{t,**}^{(m)} > 0 \\ &\iff \bar{u}_{b^{(l)},*}^{(m)} - \beta_t + \theta^{(m)} \left( \frac{\beta_t - (\bar{u}_*^{(l)} + \delta)}{\theta^{(l)}} \right) > 0 \\ &\iff \bar{u}_{b^{(l)},*}^{(m)} - \beta_t + \theta^{(m)} \left( \frac{\beta_t - \bar{u}_*^{(l)}}{\theta^{(l)}} \right) - \frac{\delta\theta^{(m)}}{\theta^{(l)}} > 0 \\ &\iff \bar{u}_{b^{(l)},*}^{(m)} - \hat{u}_t^{(m)} - \frac{\delta\theta^{(m)}}{\theta^{(l)}} > 0 \\ &\iff U - \frac{\delta\theta^{(m)}}{\theta^{(l)}} > 0 \\ &\iff \delta < U\theta^{(l)}/\theta^{(m)}, \end{aligned}$$

where the first implication follows because the price of product  $b^{(l)}$  did not change when going from  $\mathcal{C}_*^{(l)}$  to  $\mathcal{C}_{**}^{(l)}$ . Note that since  $\theta^{(1)} < \theta^{(2)} < \dots < \theta^{(K)}$ , it is sufficient to have  $\delta < (U\theta^{(l)}/\theta^{(K)})$ .

The above sequence of arguments implies that for a small enough value of  $\delta$ , the consideration utility of product  $b^{(l)}$  remains higher than that of product  $t$ , and therefore, this product cannot be included in the consideration set of customer types with indices greater than  $l$ . Therefore, we can conclude that the price adjustments do not change the consideration sets of the customer types with indices larger than  $l$ .

*Consistent consideration set of customer type  $q$ , where  $q < l'$ :* Note if  $l' = 1$ , we can skip the discussion below. Since  $a^{(l)} \in \mathcal{C}_{**}^{(l)}$ , none of the products of the set  $\{a^{(l)} + 1, \dots, t\}$  can be part of the consideration set of any customer type  $q < l$ . However, we still need to ensure that decreasing the price of product  $a^{(l)}$  does not result in its inclusion into the consideration set of customer type  $q$ . Let

$$S := \min_{q' \in [l' - 1]} \bar{u}_*^{(q')} - \bar{u}_{a^{(l)},*}^{(q')} > 0,$$

represents the minimum difference between the consideration utility of product  $a^{(l)}$  and the consideration utility values of customer type  $q' \in [l' - 1]$ , in the original optimal solution. Note that the inequality follows as  $a^{(l)} \notin \mathcal{C}_*^{(q)}$  for any  $q < l'$ , by the definition of  $l'$ . To avoid the inclusion of product  $a^{(l)}$  into the consideration set of customer type  $q$  following the price adjustments, it suffices to choose  $\delta$  such that:

$$\begin{aligned} \bar{u}_{**}^{(q)} - \bar{u}_{a^{(l)},**}^{(q)} > 0 &\iff \bar{u}_*^{(q)} - \bar{u}_{a^{(l)},**}^{(q)} > 0 \\ &\iff \bar{u}_*^{(q)} - \beta_{a^{(l)}} + \theta^{(q)} \left( \frac{\beta_{a^{(l)}} - (\bar{u}_*^{(l)} + \delta)}{\theta^{(l)}} \right) > 0 \\ &\iff \bar{u}_*^{(q)} - \beta_{a^{(l)}} + \theta^{(q)} \left( \frac{\beta_{a^{(l)}} - \bar{u}_*^{(l)}}{\theta^{(l)}} \right) - \frac{\theta^{(q)}\delta}{\theta^{(l)}} > 0 \\ &\iff \bar{u}_*^{(q)} - \bar{u}_{a^{(l)},*}^{(q)} - \frac{\theta^{(q)}\delta}{\theta^{(l)}} > 0 \\ &\iff S - \frac{\theta^{(q)}\delta}{\theta^{(l)}} > 0 \\ &\iff \delta < S\theta^{(l)}/\theta^{(q)}, \end{aligned}$$

where the first implication follows because the price adjustments do not impact the consideration utility of customer type  $q < l'$ . Note that since  $\theta^{(1)} < \theta^{(2)} < \dots < \theta^{(K)}$ , it is sufficient to have  $\delta < (S\theta^{(l)}/\theta^{(l'-1)})$ .

Note that this bound on the value of the  $\delta$  serves only as a necessary condition. Indeed, depending on the composition of  $\mathcal{C}_*^{(l')}$ , the consideration set of customer types with indices lower than  $l'$  might still change, even if  $\delta$  is chosen within the specified bound. This scenario might occur when  $|\mathcal{C}_*^{(l')}| > 1$ , where price adjustments may potentially result in changes to the consideration sets of multiple customer types. We will discuss how to prevent such scenarios in further detail in Subcase 3 later in the proof.

**2) Differences in the total revenue:** We now demonstrate that the overall revenue does not decrease after constructing  $\mathcal{C}_{**}^{(l)}$  through a reduction in the prices of products in the set  $\{a^{(l)}, \dots, t\}$ . Depending on the values of  $l'$  and  $l$ , and the composition of the consideration set of customer type  $l'$  in the original optimal solution, we partition the analysis into three subcases. It can be verified that any consideration set structure other than these three subcases violates Structural Result 1.

**Subcase 1:**  $l' = l$ .

In this case, customer type  $l$  does not share an identical product in its consideration set with any of the customer types with indices lower than  $l$ . Therefore, the price adjustments of the products of the set  $\{a^{(l)}, \dots, t\}$  do not change the revenue obtained from any customer type  $k \in [K] \setminus \{l\}$ . Therefore, we only need to guarantee that after the price adjustments, the revenue obtained from customer type  $l$  increases. Note that this is not immediate because we decreased the prices of products in the set  $\{a^{(l)}, \dots, t\}$ . Consider the change in the revenue obtained from customer type  $l$  after price adjustments:

$$\begin{aligned}
& R^{(l)}(\mathcal{C}_{**}^{(l)}, \bar{u}_{**}^{(l)}) - R^{(l)}(\mathcal{C}_*^{(l)}, \bar{u}_*^{(l)}) \\
&= R^{(l)}(\mathcal{C}_{**}^{(l)}, \bar{u}_{**}^{(l)}) - R^{(l)}(\mathcal{C}_{**}^{(l)}, \bar{u}_*^{(l)}) + R^{(l)}(\mathcal{C}_{**}^{(l)}, \bar{u}_*^{(l)}) - R^{(l)}(\mathcal{C}_*^{(l)}, \bar{u}_*^{(l)}) \\
&= \sum_{j \in \{a^{(l)}, \dots, t\}} \frac{\beta_j - (\bar{u}_*^{(l)} + \delta)}{\theta^{(l)}} \left( \frac{\exp(\lambda_j^{(l)} - \beta_j + \bar{u}_*^{(l)} + \delta)}{1 + \sum_{i \in \{a^{(l)}, \dots, t\}} \exp(\lambda_i^{(l)} - \beta_i + \bar{u}_*^{(l)} + \delta)} \right) \\
&\quad - \sum_{j \in \{a^{(l)}, \dots, t\}} \frac{\beta_j - \bar{u}_*^{(l)}}{\theta^{(l)}} \left( \frac{\exp(\lambda_j^{(l)} - \beta_j + \bar{u}_*^{(l)})}{1 + \sum_{i \in \{a^{(l)}, \dots, t\}} \exp(\lambda_i^{(l)} - \beta_i + \bar{u}_*^{(l)})} \right) + \Delta^{(l)} \\
&> \Delta^{(l)} - \sum_{j \in \{a^{(l)}, \dots, t\}} \frac{\delta}{\theta^{(l)}} \left( \frac{\exp(\lambda_j^{(l)} - \beta_j + \bar{u}_*^{(l)} + \delta)}{1 + \sum_{i \in \{a^{(l)}, \dots, t\}} \exp(\lambda_i^{(l)} - \beta_i + \bar{u}_*^{(l)} + \delta)} \right),
\end{aligned}$$

where the inequality follows as for any  $a_i \in \mathbb{R}$  for  $i \in [n]$ , and any  $x > 0$ , it is easy to verify that:

$$\frac{\exp(a_j)}{1 + \sum_{i \in [n]} \exp(a_i)} < \frac{\exp(a_j + x)}{1 + \sum_{i \in [n]} \exp(a_i + x)}, \quad \forall j \in [n]. \quad (\text{C-10})$$

Now as long as the value of  $\delta$  satisfies the following condition:

$$\sum_{j \in \{a^{(l)}, \dots, t\}} \frac{\delta}{\theta^{(l)}} \left( \frac{\exp(\lambda_j^{(l)} - \beta_j + \bar{u}_*^{(l)} + \delta)}{1 + \sum_{i \in \{a^{(l)}, \dots, t\}} \exp(\lambda_i^{(l)} - \beta_i + \bar{u}_*^{(l)} + \delta)} \right) < \Delta^{(l)},$$

we are guaranteed that the total revenue increases. Note that the function on the left-hand side of the above inequality is continuous and takes the value 0 at  $\delta = 0$ . Therefore, there must exist a  $\delta > 0$  for which the inequality holds, which contradicts the optimality of the original optimal solution.

In the following two subcases,  $l' < l$ . Here, since  $a^{(l)} \in \mathcal{C}_*^{(l)} \cap \mathcal{C}_*^{(l')}$ , it can be concluded from Structural Result 1 that  $\mathcal{C}_*^{(z)} = \{a^{(l)}\}$ , for any  $z \in \{l' + 1, \dots, l - 1\}$ . We distinguish between two cases based on the composition of  $\mathcal{C}_*^{(l')}$ .

**Subcase 2:**  $l' < l$  and  $\mathcal{C}_*^{(z)} = \{a^{(l)}\}$ , for all  $z \in \{l', \dots, l - 1\}$ .

Decreasing the price of products in the set  $\{a^{(l)}, \dots, t\}$  to construct  $\mathcal{C}_{**}^{(l)}$  increases their consideration utility. Consequently, product  $a^{(l)}$  continues to be the only product in the consideration set of customer types  $\{l', \dots, l - 1\}$ , i.e.,  $\mathcal{C}_{**}^{(z)} = \mathcal{C}_*^{(z)}$  for all  $z \in \{l', \dots, l - 1\}$ . Moreover, we already established that the consideration set of any customer type  $q \notin \{l', \dots, l\}$  is not impacted as long as  $\delta$  is chosen suitably. Therefore, we only need to guarantee that after the price adjustments, the total revenue obtained from customer types in  $\{l', \dots, l\}$  increases. Following the price adjustments, the new consideration utility of customer type  $z \in \{l', \dots, l\}$  is as follows:

$$\bar{u}_{**}^{(z)} = \beta_{a^{(l)}} - \theta^{(z)} \left( \frac{\beta_{a^{(l)}} - \bar{u}_{**}^{(l)}}{\theta^{(l)}} \right)$$

$$\begin{aligned}
&= \beta_{a^{(l)}} - \theta^{(z)} \left( \frac{\beta_{a^{(l)}} - (\bar{u}_*^{(l)} + \delta)}{\theta^{(l)}} \right) \\
&= \beta_{a^{(l)}} - \theta^{(z)} \left( \frac{\beta_{a^{(l)}} - \bar{u}_*^{(l)}}{\theta^{(l)}} \right) + \frac{\theta^{(z)} \delta}{\theta^{(l)}} \\
&= \bar{u}_*^{(z)} + \frac{\theta^{(z)} \delta}{\theta^{(l)}}. \tag{C-11}
\end{aligned}$$

Now, the difference in the total revenue obtained from customer types in  $\{l', \dots, l\}$  in the original optimal solution and after the price adjustments can be written as follows:

$$\begin{aligned}
&\sum_{z=l'}^l \alpha^{(z)} (R^{(z)}(\mathcal{C}_{**}^{(z)}, \bar{u}_{**}^{(z)}) - R^{(z)}(\mathcal{C}_*^{(z)}, \bar{u}_*^{(z)})) \\
&= \sum_{z=l'}^{l-1} \alpha^{(z)} (R^{(z)}(\mathcal{C}_{**}^{(z)}, \bar{u}_{**}^{(z)}) - R^{(z)}(\mathcal{C}_*^{(z)}, \bar{u}_*^{(z)})) + \alpha^{(l)} (R^{(l)}(\mathcal{C}_{**}^{(l)}, \bar{u}_{**}^{(l)}) - R^{(l)}(\mathcal{C}_*^{(l)}, \bar{u}_*^{(l)})) \\
&= \sum_{z=l'}^{l-1} \alpha^{(z)} (R^{(z)}(\mathcal{C}_{**}^{(z)}, \bar{u}_{**}^{(z)}) - R^{(z)}(\mathcal{C}_*^{(z)}, \bar{u}_*^{(z)})) + \alpha^{(l)} (R^{(l)}(\mathcal{C}_{**}^{(l)}, \bar{u}_{**}^{(l)}) - R^{(l)}(\mathcal{C}_*^{(l)}, \bar{u}_*^{(l)}) + R^{(l)}(\mathcal{C}_{**}^{(l)}, \bar{u}_*^{(l)}) - R^{(l)}(\mathcal{C}_*^{(l)}, \bar{u}_*^{(l)})) \\
&\stackrel{(a)}{=} \sum_{z=l'}^{l-1} \alpha^{(z)} \frac{\beta_{a^{(l)}} - (\bar{u}_*^{(l)} + \delta)}{\theta^{(l)}} \left( \frac{\exp(\lambda_{a^{(l)}}^{(z)} - \beta_{a^{(l)}} + \bar{u}_*^{(z)} + \frac{\theta^{(z)} \delta}{\theta^{(l)}})}{1 + \exp(\lambda_{a^{(l)}}^{(z)} - \beta_{a^{(l)}} + \bar{u}_*^{(z)} + \frac{\theta^{(z)} \delta}{\theta^{(l)}})} \right) \\
&\quad - \sum_{z=l'}^{l-1} \alpha^{(z)} \frac{\beta_{a^{(l)}} - \bar{u}_*^{(l)}}{\theta^{(l)}} \left( \frac{\exp(\lambda_{a^{(l)}}^{(z)} - \beta_{a^{(l)}} + \bar{u}_*^{(z)})}{1 + \exp(\lambda_{a^{(l)}}^{(z)} - \beta_{a^{(l)}} + \bar{u}_*^{(z)})} \right) \\
&\quad + \alpha^{(l)} \sum_{j \in \{a^{(l)}, \dots, t\}} \frac{\beta_j - (\bar{u}_*^{(l)} + \delta)}{\theta^{(l)}} \left( \frac{\exp(\lambda_j^{(l)} - \beta_j + \bar{u}_*^{(l)} + \delta)}{1 + \sum_{i \in \{a^{(l)}, \dots, t\}} \exp(\lambda_i^{(l)} - \beta_i + \bar{u}_*^{(l)} + \delta)} \right) \\
&\quad - \alpha^{(l)} \sum_{j \in \{a^{(l)}, \dots, t\}} \frac{\beta_j - \bar{u}_*^{(l)}}{\theta^{(l)}} \left( \frac{\exp(\lambda_j^{(l)} - \beta_j + \bar{u}_*^{(l)})}{1 + \sum_{i \in \{a^{(l)}, \dots, t\}} \exp(\lambda_i^{(l)} - \beta_i + \bar{u}_*^{(l)})} \right) + \alpha^{(l)} \Delta^{(l)} \\
&\stackrel{(b)}{>} \alpha^{(l)} \Delta^{(l)} - \sum_{z=l'}^{l-1} \alpha^{(z)} \frac{\delta}{\theta^{(l)}} \left( \frac{\exp(\lambda_{a^{(l)}}^{(z)} - \beta_{a^{(l)}} + \bar{u}_*^{(z)} + \frac{\theta^{(z)} \delta}{\theta^{(l)}})}{1 + \exp(\lambda_{a^{(l)}}^{(z)} - \beta_{a^{(l)}} + \bar{u}_*^{(z)} + \frac{\theta^{(z)} \delta}{\theta^{(l)}})} \right) \\
&\quad - \alpha^{(l)} \sum_{j \in \{a^{(l)}, \dots, t\}} \frac{\delta}{\theta^{(l)}} \left( \frac{\exp(\lambda_j^{(l)} - \beta_j + \bar{u}_*^{(l)} + \delta)}{1 + \sum_{i \in \{a^{(l)}, \dots, t\}} \exp(\lambda_i^{(l)} - \beta_i + \bar{u}_*^{(l)} + \delta)} \right),
\end{aligned}$$

where (a) follows from C-11 and the fact that  $\mathcal{C}_*^{(z)} = \mathcal{C}_{**}^z = \{a^l\}$  for any  $z \in \{l', \dots, l-1\}$ . In addition, (b) follows from C-10. Again, as long as  $\delta$  is chosen to satisfy the following condition:

$$\begin{aligned}
&\sum_{z=l'}^{l-1} \frac{\alpha^{(z)} \delta}{\alpha^{(l)} \theta^{(l)}} \left( \frac{\exp(\lambda_{a^{(l)}}^{(z)} - \beta_{a^{(l)}} + \bar{u}_*^{(z)} + \frac{\theta^{(z)} \delta}{\theta^{(l)}})}{1 + \exp(\lambda_{a^{(l)}}^{(z)} - \beta_{a^{(l)}} + \bar{u}_*^{(z)} + \frac{\theta^{(z)} \delta}{\theta^{(l)}})} \right) \\
&\quad + \sum_{j \in \{a^{(l)}, \dots, t\}} \frac{\delta}{\theta^{(l)}} \left( \frac{\exp(\lambda_j^{(l)} - \beta_j + \bar{u}_*^{(l)} + \delta)}{1 + \sum_{i \in \{a^{(l)}, \dots, t\}} \exp(\lambda_i^{(l)} - \beta_i + \bar{u}_*^{(l)} + \delta)} \right) < \Delta^{(l)},
\end{aligned}$$

we are guaranteed that the total revenue increases. Similar to subcase 1, the left-hand side of the above inequality is continuous in  $\delta$  and takes on a value of 0 when  $\delta = 0$ . Thus, the existence of a  $\delta > 0$  satisfying the above inequality is guaranteed, which again contradicts the optimality of the original solution.

**Subcase 3:**  $l' < l$  and  $\mathcal{C}_*^{(z)} = \{a^{(l)}\}$ , for all  $z \in \{l' + 1, \dots, l - 1\}$ , with  $|\mathcal{C}_*^{(l')}| > 1$ .

In this case, apart from product  $a^{(l)}$ , there are some other products in the consideration set  $\mathcal{C}_*^{(l')}$  of customer type  $l'$  in the original optimal solution. Therefore, the price adjustments result in product  $a^{(l)}$  attaining a higher consideration utility and becoming the only product in the new consideration set  $\mathcal{C}_{**}^{(l')}$  of customer type  $l'$ . To keep the consideration set of customer type  $l'$  the same as in the original optimal solution, we need to also adjust the prices of any product  $j \in \mathcal{C}_*^{(l')} \setminus \{a^{(l)}\}$  to reintroduce them into the consideration set of customer type  $l'$ . We achieve this by simply reducing the price of each product  $j \in \mathcal{C}_*^{(l')} \setminus \{a^{(l)}\}$  by the amount  $\delta/\theta^{(l)}$ . It can be verified that this decrease in the price ensures that all the consideration set of type  $l'$  is unchanged from before, that is,  $\mathcal{C}_{**}^{(l')} = \mathcal{C}_*^{(l')}$ . Moreover, the modified prices are non-negative because the price of any product  $j \in \mathcal{C}_*^{(l')} \setminus \{a^{(l)}\}$  in the original optimal solution must be strictly greater than the price of product  $a^{(l)}$ , as was established in Part 2 of the proof of Structural Result 1. With the above price changes in place, we consider two scenarios:

*Subcase 3.1*  $l' = 1$ , or  $l' > 1$  but customer type  $l'$  does not share any product in the consideration sets of customer types with indices lower than  $l'$ .

In this case, an argument analogous to that in Subcase 2 earlier can be applied to show that the total revenue increases for an appropriate value of  $\delta$ .

*Subcase 3.2:*  $l' > 1$  and customer type  $l'$  shares product  $a^{(l')}$  in its consideration set with some customer types having indices lower than  $l'$ , where recall that  $a^{(l')} = \arg \min_{j \in \mathcal{C}_*^{(l')}} j$ .

Note that only product  $a^{(l')}$  can be shared as a consequence of Structural Result 1. Let  $l'' = \min \{k \in [K] : a^{(l')} \in \mathcal{C}_*^{(k)}\}$ . Depending on the value of  $l''$  and the composition of  $\mathcal{C}_*^{(l'')}$ , the arguments outlined above in Subcases 2 and 3.1 can be applied recursively to guarantee that the total revenue after all the price adjustments increases, for a small enough value of  $\delta$ .

In summary, with proper selection of  $\delta > 0$ , the total revenue increases in either case, again contradicting the optimality of the original solution.  $\square$

**C.6.4. Proof of Structural Result 4.** Suppose that in the optimal solution, there exist two consecutive customer types  $k$  and  $k+1$ , where  $k \in [K-1]$ , such that  $a^{(k+1)} - b^{(k)} > 1$ . Consequently,  $b^{(k)} + 1 \notin \mathcal{C}_*^{(k+1)}$ . Similar to the argument at the beginning of the proof of Structural Result 3, one can conclude that while keeping the consideration utility of customer type  $k+1$  the same as in the original optimal solution, one of the consideration sets among  $\{b^{(k)} + 1\}$ ,  $\{b^{(k)} + 1, b^{(k)} + 2\}$ ,  $\dots$ ,  $\{b^{(k)} + 1, \dots, b^{(k+1)}\}$  provides a strictly higher revenue compared to  $\mathcal{C}_*^{(k+1)}$ . Following the sequence of steps outlined in the proof of Structural Result 3, we can then adjust product prices to exhibit an alternate solution that provides strictly higher revenue compared to the original optimal solution, resulting in a contradiction.

$\square$

**C.6.5. Proof of Structural Result 5.** Suppose there exists an optimal solution such that product  $1 \notin \mathcal{C}_*^{(1)}$ . Note that  $\mathcal{C}_*^{(1)} \neq \emptyset$  because otherwise, following Structural Result 2, all the consideration sets will be empty. Recall that  $b^{(1)} = \arg \max_{j \in \mathcal{C}_*^{(1)}} j$ . Similar to the argument at the beginning of the proof of Structural Result 3, one can conclude that while keeping the consideration utility of customer type 1 the same as in the original optimal solution, one of the consideration sets among  $\{1\}$ ,  $\{1, 2\}$ ,  $\dots$ ,  $\{1, 2, \dots, b^{(1)}\}$  provides a strictly higher revenue compared to  $\mathcal{C}_*^{(1)}$ . We can then follow the sequence of steps outlined in the proof of Structural Result 3 to adjust product prices and exhibit an alternate solution that provides strictly higher revenue compared to the original optimal solution, which leads to a contradiction.  $\square$

We now turn to quantifying the number of valid consideration sets implied by the structural results. Let  $T(N, K)$  denote the total number of valid consideration sets with  $N$  products and  $K$  customer types. We bound  $T(N, K)$  via a recursive relation. First, note that  $T(N, 1) = N$ , as directly implied by Structural Results 3 and 5.

Next, suppose product 1 has the largest index in the consideration set of customer type  $K - 1$ . Following Structural Result 4, if the consideration set of customer type  $K$  is non-empty, then the product with the lowest index in this set must be either 1 or 2. Following Structural Result 3, there can be at most  $N$  or  $N - 1$  consideration sets for customer type  $K$ , respectively. Including the possibility of an empty consideration set, there are  $2N$  possible consideration sets for customer type  $K$ . Further, if any product other than 1 has the largest index in the consideration set of customer type  $K - 1$ , the total number of possible consideration sets for customer type  $K$  will be less than  $2N$ . Therefore, it follows that:

$$\begin{aligned} T(N, K) &\leq T(N, K - 1) \cdot (2N) \\ &\leq (T(N, K - 2) \cdot (2N)) \cdot 2N \\ &\leq T(N, 1) \cdot (2N)^{K-1} \\ &\leq N \cdot (2N)^{K-1}. \end{aligned}$$

This implies that the total number of valid consideration sets is  $\mathcal{O}(N^K)$ . □

### C.7. Proof of Proposition 2

It is evident that the number of constraints in (B-1), (B-2), and (B-4) each have a complexity of  $\mathcal{O}(NK)$ . For constraint (B-3), note that first, by the corollary stated after Structural Result 1, any two customer types can have at most one product in common within their consideration sets. Moreover, this constraint needs to be written only for types  $k \in [K - 1]$  and  $k + 1$  when they share a product in their consideration sets. This is because if customer type  $k$  shares a product with any customer type  $k' > k + 1$ , Structural Result 1 implies that the shared product must also belong to the consideration sets of all customer types in  $\{k + 1, \dots, k' - 1\}$ . Consequently, this constraint appears at most  $K - 1$  times, resulting in a complexity of  $\mathcal{O}(K)$ . Therefore, it follows that the total number of constraints is:

$$\mathcal{O}(NK) + \mathcal{O}(NK) + \mathcal{O}(NK) + \mathcal{O}(K) = \mathcal{O}(NK).$$

□

### C.8. Proof of Theorem 5

We prove this result by establishing a reduction from the single-product design problem studied in Akchen and Mišić (2025), where the goal is to design a product by selecting a subset of features to maximize the purchase probability under a binary logit model. In their setting, a product is described by a binary feature vector  $\mathbf{a} = (a_1, \dots, a_n) \in \{0, 1\}^n$ , indicating the presence of each of  $n$  features. The customer population consists of  $K$  types, and the utility of customer type  $k \in [K]$ , with a mixing proportion of  $\lambda_k \geq 0$ , for a product with feature vector  $\mathbf{a}$  is of the form:

$$u_k(\mathbf{a}) = \beta_{k,0} + \sum_{i=1}^n \beta_{k,i} a_i. \tag{C-12}$$

Akchen and Mišić (2025, Theorem 1) show that the following product design problem, which aims to maximize the aggregate purchase probability across customer types, is NP-hard:

$$\max_{\mathbf{a} \in \{0,1\}^n} \sum_{k=1}^K \lambda_k \cdot \frac{\exp(u_k(\mathbf{a}))}{1 + \exp(u_k(\mathbf{a}))}. \quad (\text{C-13})$$

Moreover, in Theorem 2 of their paper, they prove that this problem is NP-hard to approximate within a factor of  $\mathcal{O}(1/n^{1-\epsilon})$  for any  $\epsilon > 0$ .

We construct an instance of our feature selection problem (22) such that solving it is equivalent to solving the product design problem (C-13). We consider a single product, that is  $N = 1$ , and therefore omit product-specific subscripts. The customer population consists of  $K$  customer types, with mixing proportions given by  $\alpha^{(k)} = \lambda_k$  for all  $k \in [K]$ . We consider a total of  $n + 1$  features, where the last feature is always displayed in the first stage. The firm decides which of the remaining features to display. Therefore, the constraint set  $\mathcal{G} = \{\mathbf{g} \in \{0,1\}^{n+1} : g_{n+1} = 1\}$ .

We set all the feature values equal to 1, i.e.,  $z_d = 1$  for all  $d \in [n + 1]$ . The feature preference parameters in the consideration stage are set to match the utility coefficients in (C-12): for the last feature,  $\theta_{n+1}^{(k)} = \beta_{k,0}$ ; for all other features  $d \in [n]$ , we set  $\theta_d^{(k)} = \beta_{k,d}$ . For all customer types  $k \in [K]$ , this yields  $c_d^{(k)} = \beta_{k,d}$  for all features  $d \in [n]$ , and  $c_{n+1}^{(k)} = \beta_{k,0}$ . In the purchase stage, we assume that all customer types share the same feature preference parameters, i.e.,  $\omega_d^{(k)} = \omega_d$  for all  $k \in [K]$  and  $d \in [n + 1]$ , where the values of  $\omega_d$  are chosen arbitrarily. As a result, it can be verified that  $b^{(k)} = b$  for all  $k \in [K]$  for some  $b \geq 0$ .

Under this construction, the feature selection problem (22) reduces to:

$$\arg \max_{\mathbf{g} \in \mathcal{G}} \sum_{k \in [K]} \lambda_k \cdot b \cdot \frac{\exp(\sum_{d \in [n+1]} c_d^{(k)} \cdot g_d)}{1 + \exp(\sum_{d \in [n+1]} c_d^{(k)} \cdot g_d)} \equiv \arg \max_{\mathbf{g}' \in \{0,1\}^n} \sum_{k \in [K]} \lambda_k \cdot b \cdot \frac{\exp(\beta_{k,0} + \sum_{d \in [n]} \beta_{k,d} \cdot g'_d)}{1 + \exp(\beta_{k,0} + \sum_{d \in [n]} \beta_{k,d} \cdot g'_d)},$$

□

### C.9. Proof of Theorem 6

We prove the NP-completeness of the feature selection problem using a reduction from the well-known NP-complete problem, Subset Sum. In the Subset Sum problem, we are given a set of  $B$  positive integers  $\{a_1, \dots, a_B\}$  and a target value  $T$ . The goal is to determine whether there exists a subset  $S \subseteq [B]$  such that

$$\sum_{i \in S} a_i = T.$$

We reduce this problem to the decision version of our feature selection problem: given a set of features and a target value  $R^*$ , is there a subset of features such that the expected revenue is at least  $R^*$ ? We construct an instance of the feature selection problem so that a solution exists if and only if the corresponding Subset Sum instance has a solution.

We consider  $N = 2$  products,  $B + 2$  features, and  $K = 1$  customer type. Therefore, we drop type-specific superscripts. Features  $B + 1$  and  $B + 2$  are always displayed in the first stage, and the platform selects which of the remaining features to display. Feature  $B + 2$  corresponds to price. We set  $\theta_{B+2} = -1$  and  $\theta_d = 1$ , for any feature  $d \in [B + 1]$ . For product 1, we set  $z_{1,B+1} = T + \epsilon$ ,  $z_{1,B+2} = T$ , and  $z_{1,d} = 0$  for any  $d \in [B]$ , for some  $\epsilon > 0$ . This yields:  $c_{1,B+1} = T + \epsilon$  and  $c_{1,B+2} = -T$  and  $c_{1,d} = 0$  for all  $d \in [B]$ . Therefore, for any binary selection vector over the remaining features, denoted with  $\mathbf{g}' \in \{0,1\}^B$ , the consideration utility of product 1 is as follows:

$$\bar{u}_1 = T + \epsilon - T + \sum_{d \in [B]} c_{1,d} \cdot g'_d = \epsilon + \sum_{d \in [B]} 0 \cdot g'_d = \epsilon.$$

For product 2, we set  $z_{2,B+1} = \epsilon$ ,  $z_{2,B+2} = T$ , and  $z_{2,d} = a_d$  for any  $d \in [B]$ . Thus  $c_{2,B+1} = \epsilon$  and  $c_{2,B+2} = -T$  and  $c_{2,d} = a_d$  for all  $d \in [B]$ . The consideration utility of product 2 is then computed as follows:

$$\bar{u}_2 = \epsilon - T + \sum_{d \in [B]} c_{2,d} \cdot g'_d = \epsilon - T + \sum_{d \in [B]} a_d \cdot g'_d.$$

Let  $\omega_d = 0$  for all  $d \in [B + 1]$  and  $\omega_{B+2} = -1$ . This ensures that  $u_1 = u_2 = -T$ . Now consider the target expected revenue:

$$R^* = T \left( \frac{2 \exp(-T)}{1 + 2 \exp(-T)} \right).$$

Note that the total number of features is  $B + 2$ , and by construction, the last two features are always displayed in the first stage. Therefore, any feasible selection vector must satisfy  $g_{B+1} = g_{B+2} = 1$ , and the problem reduces to selecting a binary vector  $\mathbf{g}' \in \{0, 1\}^B$  corresponding to the first  $B$  features. We now argue that such a vector  $\mathbf{g}'$  yields an expected revenue at least  $R^*$  if and only if the Subset Sum problem admits a solution, that is, there exists a subset  $S \subseteq [B]$  with  $\sum_{i \in S} a_i = T$ .

In the case where both products are included in the consideration set, it can be verified that the expected revenue in (23) reduces to

$$T \left( \frac{2 \exp(-T)}{1 + 2 \exp(-T)} \right).$$

In contrast, if only one product  $j \in \{1, 2\}$  is included in the consideration set, the expected revenue is

$$T \left( \frac{\exp(-T)}{1 + \exp(-T)} \right).$$

Note that the expected revenue when both products are included in the consideration set is exactly equal to the target  $R^*$ , and is strictly greater than the revenue from including each one of the products individually. Therefore, to achieve revenue at least  $R^*$ , it is necessary to ensure that both products are included in the consideration set. This occurs only when their consideration utilities are equal, i.e.,  $\bar{u}_1 = \bar{u}_2$ . Since  $\bar{u}_1 = \epsilon$  is fixed by construction, we must have:

$$\bar{u}_2 = \epsilon \Leftrightarrow \sum_{d \in [B]} a_d \cdot g'_d = T.$$

That is, the sum of the weights  $a_d$ , corresponding to the selected features, must equal  $T$ , which is exactly the Subset Sum condition. □

## Appendix D: MILP Formulations

In Section D.1, we present the formulation for the assortment optimization problem under the D-FCM. In Section D.2, we extend the MILP formulation (24) for the feature selection problem to also incorporate assortment decisions under the D-FCM with consideration limit  $L = 1$ .

### D.1. MILP for Assortment Optimization under D-FCM

Let  $a_j^{(k)} = \exp(u_j^{(k)})$  denote the exponentiated mean purchase utility, commonly referred to as the ‘‘attractiveness’’, of product  $j \in [N]$  for customer type  $k \in [K]$ . Let  $\mathbf{x} = (x_1, \dots, x_N)$  be a binary vector indicating which products are offered, with  $x_j = 1$  if product  $j \in [N]$  is included in the assortment. To capture which products enter each type’s consideration set, we introduce a binary matrix  $\mathbf{H} = (h_j^{(k)})$  where  $h_j^{(k)} = 1$  if product  $j \in [N]$  is included in the

consideration set of customer type  $k \in [K]$ , and  $h_j^{(k)} = 0$  otherwise. For simplicity, we start by assuming that all consideration utilities are distinct within each type; we extend it to accommodate ties in the consideration utilities later. Let  $\mathcal{A}^{(k)} := \{j \in [N] \mid \bar{u}_j^{(k)} \geq 0\}$  denote the set of products with non-negative consideration utilities for type  $k \in [K]$ , and let  $\pi^{(k)} = \langle \pi_1^{(k)}, \dots, \pi_{|\mathcal{A}^{(k)}|}^{(k)} \rangle$  denote the ordering of the products in  $\mathcal{A}^{(k)}$  sorted in decreasing order of the consideration utilities. Lastly, let  $\mathbf{F} = (f_j^{(k)})$  denote the matrix of purchase probabilities for all  $j \in [N] \cup \{0\}$  and  $k \in [K]$ , where  $f_j^{(k)}$  denotes the purchase probability of product  $j$  for customer type  $k$ .

With the above notation in place, problem (9) can be reformulated as follows:

$$\arg \max_{\mathbf{x}, \mathbf{H}, \mathbf{F}} \sum_{k=1}^K \alpha^{(k)} \sum_{j=1}^N p_j \cdot f_j^{(k)} \quad (\text{D-14a})$$

$$\text{s.t. } x_j \geq h_j^{(k)}, \quad \forall j \in [N], \forall k \in [K] \quad (\text{D-14b})$$

$$h_{\pi_1^{(k)}}^{(k)} = x_{\pi_1^{(k)}}, \quad \forall k \in [K], \quad (\text{D-14c})$$

$$h_{\pi_r^{(k)}}^{(k)} \geq x_{\pi_r^{(k)}} - \left( \sum_{t=1}^{r-1} h_{\pi_t^{(k)}}^{(k)} \right) / L, \quad \forall r \in [|\mathcal{A}^{(k)}|] \setminus \{1\}, \forall k \in [K], \quad (\text{D-14d})$$

$$h_j^{(k)} = 0, \quad \forall j \notin \mathcal{A}^{(k)}, \forall k \in [K], \quad (\text{D-14e})$$

$$\sum_{j \in \mathcal{A}^{(k)}} h_j^{(k)} \leq L, \quad \forall k \in [K], \quad (\text{D-14f})$$

$$f_0^{(k)} + \sum_{j \in [N]} f_j^{(k)} = 1, \quad \forall k \in [K], \quad (\text{D-14g})$$

$$f_j^{(k)} \leq a_j^{(k)} \cdot h_j^{(k)}, f_j^{(k)} \leq a_j^{(k)} \cdot f_0^{(k)}, f_j^{(k)} \geq a_j^{(k)} \cdot (f_0^{(k)} - 1 + h_j^{(k)}), \quad \forall j \in [N], \forall k \in [K] \quad (\text{D-14h})$$

$$x_j, h_j^{(k)} \in \{0, 1\}, \quad \forall j \in [N], \forall k \in [K], \quad (\text{D-14i})$$

$$0 \leq f_j^{(k)} \leq 1, \quad \forall j \in [N] \cup \{0\}, \forall k \in [K]. \quad (\text{D-14j})$$

In the above, constraint (D-14b) ensures that every product included in a consideration set must also be part of the offered assortment. Constraints (D-14c)-(D-14f) ensure that the consideration sets in (8) are satisfied—we note that it was not trivial to come up with such a compact formulation. Constraints (D-14g) and (D-14h) ensure the correct assignment of purchase probabilities to the offered products following existing formulations, see, e.g., Bernstein et al. (2024). It can be verified that the above formulation involves  $\mathcal{O}(NK)$  variables and constraints.

To formulate the assortment optimization problem under the D-FCM without assuming distinct consideration utility values, we introduce additional notation. For each customer type  $k \in [K]$ , let  $E^{(k)} \leq N$  denote the number of distinct positive consideration utility values across all products. For each  $e \in [E^{(k)}]$  and  $k \in [K]$ , let  $S_e^{(k)}$  be the group of products that share the  $e$ -th highest positive consideration utility value. For each  $e \in [E^{(k)}]$  and  $k \in [K]$ , we define the binary matrix  $\mathbf{Q} = (q_e^{(k)})$ , where  $q_e^{(k)} = 1$  if any product in  $S_e^{(k)}$  is offered. Similarly, we define the binary matrix  $\mathbf{Y} = (y_e^{(k)})$ , where  $y_e^{(k)} = 1$  if any product in  $S_e^{(k)}$  is included in the consideration set of customer type  $k$ . The assortment optimization problem is then formulated as follows:

$$\arg \max_{\mathbf{x}, \mathbf{H}, \mathbf{F}, \mathbf{Q}, \mathbf{Y}} \sum_{k=1}^K \alpha^{(k)} \sum_{j=1}^N p_j \cdot f_j^{(k)} \quad (\text{D-15a})$$

$$q_e^{(k)} \leq x_j \quad \forall j \in S_e^{(k)}, \forall e \in [E^{(k)}], \forall k \in [K] \quad (\text{D-15b})$$

$$q_e^{(k)} \geq \left( \sum_{j \in S_e^{(k)}} x_j \right) / |S_e^{(k)}| \quad \forall e \in [E^{(k)}], \forall k \in [K] \quad (\text{D-15c})$$

$$y_1^{(k)} = q_1^{(k)}, \quad \forall k \in [K] \quad (\text{D-15d})$$

$$y_{e+1}^{(k)} \geq q_{e+1}^{(k)} - \left( \sum_{t=1}^e q_t^{(k)} \right) / L \quad \forall e \in [E^{(k)} - 1], \forall k \in [K] \quad (\text{D-15e})$$

$$\sum_{e \in [E^{(k)}]} y_e^{(k)} \leq L, \quad \forall k \in [K] \quad (\text{D-15f})$$

$$x_j \geq h_j^{(k)}, y_e^{(k)} \geq h_j^{(k)}, h_j^{(k)} \geq x_j + y_e^{(k)} - 1, \quad \forall j \in S_e^{(k)}, \forall e \in [E^{(k)}], \forall k \in [K] \quad (\text{D-15g})$$

$$f_0^{(k)} + \sum_{j \in [N]} f_j^{(k)} = 1, \quad \forall k \in [K], \quad (\text{D-15h})$$

$$f_j^{(k)} \leq a_j^{(k)} \cdot h_j^{(k)}, f_j^{(k)} \leq a_j^{(k)} \cdot f_0^{(k)}, f_j^{(k)} \geq a_j^{(k)} \cdot (f_0^{(k)} - 1 + h_j^{(k)}), \quad \forall j \in [N], \forall k \in [K] \quad (\text{D-15i})$$

$$x_j, h_j^{(k)} \in \{0, 1\} \quad \forall j \in [N], \forall k \in [K] \quad (\text{D-15j})$$

$$y_e^{(k)}, q_e^{(k)} \in \{0, 1\} \quad \forall e \in [E^{(k)}], \forall k \in [K] \quad (\text{D-15k})$$

To determine the consideration set for each customer type in the formulation above, we follow a three-step procedure. First, for each product group  $S_e^{(k)}$ , we determine whether it contains at least one offered product. Second, among the groups with (some) offered products, we identify those corresponding to the  $L$  highest utility values for each customer type. Finally, we ensure that all offered products within those groups with top- $L$  consideration utilities are included in the consideration set.

Constraints (D-15b)–(D-15c) implement the first step by identifying, for each  $e \in [E^{(k)}]$  and  $k \in [K]$ , whether the product group  $S_e^{(k)}$  contains at least one offered product. Constraints (D-15d)–(D-15f) implement the second step by determining which of these groups—with at least one offered product—correspond to the  $L$  highest consideration utility values and should therefore be included in the consideration sets. Finally, Constraint (D-15g) enforces the third step by ensuring that all offered products belonging to the top- $L$  utility groups are included in the consideration set. The remaining constraints are the same as in formulation (D-14). It can be verified that the above formulation involves  $\mathcal{O}(NK)$  variables and constraints.

## D.2. MILP for Joint Assortment and Feature Selection under the D-FCM with Consideration Limit $L = 1$

We extend the MILP formulation for the feature selection problem from Section 7 to jointly incorporate assortment decisions. To this end, we introduce a binary variable  $x_j$ , which equals 1 if product  $j \in [N]$  is offered and 0 otherwise. Below, we present the joint assortment and feature selection formulation under the D-FCM with consideration limit  $L = 1$ :

$$\arg \max_{\mathbf{x}, \mathbf{g}, \mathbf{V}, \bar{\mathbf{u}}, \mathbf{H}, \mathbf{F}} \sum_{k \in [K]} \alpha^{(k)} \sum_{j \in [N]} p_j \cdot f_j^{(k)} \quad (\text{D-16a})$$

$$\text{s.t.} \quad \sum_{d \in D} c_{j,d}^{(k)} \cdot v_{j,d} \leq \bar{u}^{(k)} - \epsilon + M \cdot (1 - x_j + h_j^{(k)}), \quad \forall j \in [N], \forall k \in [K], \quad (\text{D-16b})$$

$$\sum_{d \in D} c_{j,d}^{(k)} \cdot v_{j,d} \geq \bar{u}^{(k)} - M \cdot (1 - h_j^{(k)}), \quad \forall j \in [N], \forall k \in [K], \quad (\text{D-16c})$$

$$\sum_{d \in D} c_{j,d}^{(k)} \cdot v_{j,d} \leq \bar{u}^{(k)} + M \cdot (1 - h_j^{(k)}), \quad \forall j \in [N], \forall k \in [K], \quad (\text{D-16d})$$

$$x_j \geq h_j^{(k)} \quad \forall j \in [N], \forall k \in [K], \quad (\text{D-16e})$$

$$\sum_{d \in D} g_d \leq D_{\text{cap}}, \quad (\text{D-16f})$$

$$f_0^{(k)} + \sum_{j \in [N]} f_j^{(k)} = 1, \quad \forall k \in [K], \quad (\text{D-16g})$$

$$f_j^{(k)} \leq a_j^{(k)} \cdot h_j^{(k)}, \quad f_j^{(k)} \leq a_j^{(k)} \cdot f_0^{(k)}, \quad f_j^{(k)} \geq a_j^{(k)} \cdot (f_0^{(k)} - 1 + h_j^{(k)}), \quad \forall j \in [N], \forall k \in [K], \quad (\text{D-16h})$$

$$g_d = 1 \quad \forall d \in D', \quad g_d \in \{0, 1\} \quad \forall d \in D \setminus D', \quad (\text{D-16i})$$

$$v_{j,d} \leq g_d, \quad v_{j,d} \leq x_j, \quad v_{j,d} \geq g_d + x_j - 1, \quad \forall j \in [N], \forall d \in D, \quad (\text{D-16j})$$

$$x_j \in \{0, 1\}, \quad h_j^{(k)} \in \{0, 1\}, \quad v_{j,d} \in \{0, 1\}, \quad \forall j \in [N], \forall k \in [K], \forall d \in D, \quad (\text{D-16k})$$

$$0 \leq \bar{u}^{(k)} \leq B^{(k)}, \quad 0 \leq f_j^{(k)} \leq 1, \quad \forall j \in [N] \cup 0, \forall k \in [K], \quad (\text{D-16l})$$

In the formulation above, we define the binary matrix  $V = (v_{j,d})$ , where  $v_{j,d} := x_j \cdot g_d$  for all  $j \in [N]$  and  $d \in D$ . Based on this definition, the left-hand side of Constraints (D-16b)–(D-16d) computes the consideration utility for each *offered* product. These constraints ensure that, for each customer type  $k \in [K]$ , only those offered products with the highest consideration utility values are included in the consideration set. The product of binary variables in this definition is linearized through Constraints (D-16j). Constraint (D-16e) guarantees that every product included in a consideration set must be offered. The remaining constraints are the same as in formulation (24). Note that the formulation above involves  $\mathcal{O}(|D| + NK)$  binary variables,  $\mathcal{O}(NK)$  continuous variables and  $\mathcal{O}(N(K + |D|))$  constraints.

We employ the Big- $M$  approach to enforce equal (maximum) utilities among products included in the consideration sets in Constraints (D-16b)–(D-16d) (similar constraints also appear as (24b)–(24d) in the MILP formulation (24)). A valid choice for this constant  $M$  is an upper bound on the consideration utility values. Specifically, we choose  $M \geq \max_{k \in [K]} B^{(k)}$ , where  $B^{(k)}$  is defined as follows:

$$B^{(k)} := \max_{j \in [N]} \sum_{d \in D} c_{j,d}^{(k)} \cdot \mathbb{1}[c_{j,d}^{(k)} > 0],$$

where  $\mathbb{1}[\cdot]$  denotes the indicator function. Following the definition of the consideration set in (13),  $B^{(k)}$  also bounds the consideration utility values from above for each customer type  $k \in [K]$ .

## Appendix E: Assortment Optimization Decision on Synthetic Data

In this section, we discuss the results of experiments where the ground-truth is the rank-based model (Farias et al. 2013). The model assumes that the customer population comprises  $K$  distinct types, with each type associated with a specific ranking of the products. For simplicity, we suppose that the proportion of each type is the same, that is, the proportion of type  $k \in [K]$  is  $1/K$ . When presented with an assortment of products, we suppose that customers sequentially add products to their consideration set based on their preference order, including only products that are offered and preferred over the no-purchase option. Existing literature has found that customers typically evaluate only a small number of products before making a purchase decision (Hauser and Wernerfelt 1990, Feldman et al. 2019). To capture this behavior, we adjust the position of the no-purchase option in the ground-truth rankings to enforce a consideration limit of 2 or 3 in different settings. After forming the consideration set, the customer purchases the highest-ranked product within the consideration set, or leaves without a purchase if the consideration set is empty.

**Setup and performance metric.** Under the aforementioned ground-truth, we generate different problem instances where each problem instance corresponds to a distinct underlying rank-based model. For each problem instance, we first sample  $K$  different rankings over a universe of  $N = 10$  products. Then, we simulate  $T$  transactions under

the sampled ground-truth model as follows. For each transaction  $t \in [T]$ , the assortment  $S_t$  is randomly generated by including each product in the assortment with probability 0.7. We record the consideration sequence  $\mathcal{C}_t$  and the purchased product  $j_t$  resulting from a randomly selected customer type  $k_t \in [K]$ . We then fit the FCM with  $K = 5$  customer types, as well as the 5-class LC-MNL model to the transaction data, where the latter does not use the consideration set information. For brevity, we denote the FCM with  $k$  customer types as  $\text{FCM}^{(k)}$ . When fitting the  $\text{FCM}^{(5)}$ , we set  $L$  equal to the largest consideration set size in the observed transactions. Because the rank-based choice model does not capture variations in product features, we estimate only the mean consideration and purchase utilities under the FCM, and the mean purchase utilities under the LC-MNL model, similar to the setup in Section 6.1. Moreover, because the  $\text{FCM}^{(5)}$  estimates twice as many parameters as the 5-class LC-MNL model, we also compare against a 10-class LC-MNL model fit to the transaction data.

After fitting the different choice models to the transaction data, we evaluate the ground-truth revenues achieved using the optimal decisions derived from each model. In the rank-based model, each product is viewed as a fixed bundle of features and consequently, the model cannot account for changes in product prices. Therefore, we focus on solving the (pure) assortment optimization problem under the estimated FCM and LC-MNL models. Moreover, to isolate the performance improvement due to modeling consideration set behavior, independent of the computational difficulty of solving the decision problem, we use brute-force search to determine the optimal assortment for both models.

Specifically, for each problem instance  $b$ , we randomly sample the price for each product uniformly at random from the interval  $[20, 50]$ . We compare the ground-truth revenues obtained under the optimal assortments recommended by both models. Let  $S_{\mathcal{M}}^{*,b}$  denote the optimal assortment under the estimated model  $\mathcal{M} \in \{\text{FCM}^{(5)}, \text{LC-MNL}^{(5)}, \text{LC-MNL}^{(10)}\}$  for problem instance  $b$ , where  $\text{LC-MNL}^{(5)}$  and  $\text{LC-MNL}^{(10)}$  refer to the 5-class and 10-class LC-MNL models, respectively. Additionally, denoting  $\mathbf{p}_b$  as the vector of product prices for problem instance  $b$ , let  $R^{\text{true},b}(S, \mathbf{p}_b)$  denote the expected revenue for any assortment  $S$  and price vector  $\mathbf{p}_b$  under the ground-truth rank-based model for problem instance  $b$ . We measure the average percentage improvement in the revenues obtained under the optimal assortment suggested by the FCM compared to the benchmark LC-MNL models  $\mathcal{M} \in \{\text{LC-MNL}^{(5)}, \text{LC-MNL}^{(10)}\}$  across  $B$  different problem instances, as follows:

$$100 \times \left( \frac{1}{B} \sum_{b=1}^B \frac{R^{\text{true},b}(S_{\text{FCM}^{(5)}}^{*,b}, \mathbf{p}_b) - R^{\text{true},b}(S_{\mathcal{M}}^{*,b}, \mathbf{p}_b)}{R^{\text{true},b}(S_{\mathcal{M}}^{*,b}, \mathbf{p}_b)} \right). \quad (\text{E-17})$$

**Results.** Table EC.1 reports the average performance improvement across  $B = 100$  instances, as a function of the consideration limit ( $L$ ), the number of rankings in the ground-truth model ( $K$ ), and the number of transactions ( $T$ ) made available for fitting the models. The table clearly demonstrates that the  $\text{FCM}^{(5)}$  consistently outperforms both LC-MNL models across all scenarios. Moreover, the percentage improvement is greater when the number of ground-truth rankings  $K$  is smaller. This observation aligns with prior findings in the literature, which show that the LC-MNL model better approximates the underlying choice behavior as the number of rankings in the ground-truth increases, see e.g., Berbeglia et al. (2022), Gallego and Li (2024). Consequently, the value of the FCM becomes more evident when the ground truth includes a smaller number of rankings. Similarly, the performance improvements are larger when the consideration limit is smaller. This is again expected because models that do not explicitly account for consideration set formation are less affected as customers consider a larger number of products.

**Table EC.1 Revenue comparison of the FCM compared to LC-MNL benchmark**

Consideration limit ( $L$ )	Number of rankings ( $K$ )	Number of transactions ( $T$ )	% Improvement over LC-MNL <sup>(5)</sup>	% Improvement over LC-MNL <sup>(10)</sup>
2	10	1,000	12.5	12.2
		2,000	15.4	15.0
		5,000	16.8	16.6
	20	1,000	6.2	6.1
		2,000	7.9	7.1
		5,000	7.3	6.6
3	10	1,000	11.0	11.2
		2,000	8.1	8.1
		3,000*	11.2	9.4
	20	1,000	5.0	5.1
		2,000	4.5	3.7
		3,000*	5.0	4.5

\*: Numbers reported for 3,000 transactions instead of 5,000 due to computational constraints during estimation. For  $Q \in \{5, 10\}$ , LC-MNL<sup>(Q)</sup> refers to the Q-class LC-MNL model.

## References

- Akchen I, Mišić VV (2025) Exact logit-based product design URL <https://arxiv.org/abs/2106.15084>, Working paper, UCLA Anderson School of Management.
- Berbeglia G, Garassino A, Vulcano G (2022) A comparative empirical study of discrete choice models in retail operations. *Management Science* 68(6):4005–4023.
- Bernstein F, Modaresi S, Saure D (2024) Exploration optimization for dynamic assortment personalization under linear preferences. *Available at SSRN 4115721*.
- Bront JJM, Méndez-Díaz I, Vulcano G (2009) A column generation algorithm for choice-based network revenue management. *Operations research* 57(3):769–784.
- Farias VF, Jagabathula S, Shah D (2013) A nonparametric approach to modeling choice with limited data. *Management Science* 59(2):305–322.
- Feldman J, Paul A, Topaloglu H (2019) Assortment optimization with small consideration sets. *Operations Research* 67(5):1283–1299.
- Gallego G, Li A (2024) A random consideration set model for demand estimation, assortment optimization, and pricing. *Operations Research*.
- Hausser JR, Wernerfelt B (1990) An evaluation cost model of consideration sets. *Journal of consumer research* 16(4):393–408.
- Jaggi M (2013) Revisiting frank-wolfe: Projection-free sparse convex optimization. *International conference on machine learning*, 427–435 (PMLR).
- Rusmevichientong P, Shmoys D, Tong C, Topaloglu H (2014) Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management* 23(11):2023–2039.
- Talluri K, Van Ryzin G (2004) Revenue management under a general discrete choice model of consumer behavior. *Management Science* 50(1):15–33.
- Uhrin B (1984) Some remarks about the convolution of unimodal functions. *The Annals of Probability* 640–645.