A Nonparametric Approach with Marginals for Modeling Consumer Choice: Online Appendix

Yanqiu Ruan[1], Xiaobo Li[2], Karthyek Murthy[1], Karthik Natarajan[1]

[1]Singapore University of Technology and Design

[2]National University of Singapore

E-companion

The E-companion is organized as follows. Sections EC.1 to EC.3 present the proofs of the results, respectively, in Sections 3, 4, and 5. Section EC.4 collects and presents all the illustrative examples mentioned in the paper. Section EC.5 provides an algorithm to evaluate the limit of MDM in (16). Section EC.6 details the implementation of the experiments. Section EC.7 presents the details for the synthetic data generations for experiments in Section 6.2 and Section 6.4. Section EC.8 provides synthetic data experiments that validate the robustness of the numerical findings across diverse assortment structures. Sections EC.9 and EC.10 present additional numerical results with synthetic and real data respectively.

Table of Contents for the E-companion

EC.1. Proofs of the Results in Section 3

EC.1.1. Proof of Theorem 1

Proof. *Necessity of* (5): Suppose $\boldsymbol{p}_{\mathcal{S}}$ is MDM-representable. Then there exist marginal distributions $\{F_i : i \in \mathcal{N}\}$ and deterministic utilities $\{\nu_i : i \in \mathcal{N}\}$ such that for any assortment $S \in \mathcal{S}$, the given choice probability vector $(p_{i,S} : i \in S)$ and the respective Lagrange multipliers $\lambda_S, \{\lambda_{i,S} : i \in S\}$ are obtainable by solving the optimality conditions (4). That is, there exist $\{\lambda_S, \lambda_{i,S} : i \in S, S \in \mathcal{S}\}$ for some fixed choice of $\{F_i : i \in \mathcal{N}\}$ and $\{\nu_i : i \in \mathcal{N}\}$ such that

$$\nu_i + F_i^{-1}(1 - p_{i,S}) - \lambda_S + \lambda_{i,S} \;=\; 0 \quad \forall (i,S) \in \mathcal{I}_{\mathcal{S}}, \tag{EC.1}$$

$$\lambda_{i,S}\, p_{i,S} \;=\; 0 \quad \forall (i,S) \in \mathcal{I}_{\mathcal{S}}. \tag{EC.2}$$

For each product $i \in \mathcal{N}$ and any two assortments $S, T \in \mathcal{S}$ containing i as a common product,

$$\lambda_S - \nu_i = \lambda_{i,S} + F_i^{-1}(1 - p_{i,S}) \quad \text{ and } \quad \lambda_T - \nu_i = \lambda_{i,T} + F_i^{-1}(1 - p_{i,T}).$$

If $p_{i,S} < p_{i,T}$, then $\lambda_{i,S} \geq 0$ and $\lambda_{i,T} = 0$ because of the complementary slackness condition (EC.2). Since $F_i^{-1}(1 - p)$ is a strictly decreasing function over $p \in [0,1]$, by (EC.1), we obtain:

$$\lambda_S - \nu_i \;\geq\; F_i^{-1}(1 - p_{i,S}) \;>\; F_i^{-1}(1 - p_{i,T}) \;=\; \lambda_T - \nu_i.$$

Adding ν_i on both sides, we obtain that the Lagrange multipliers should satisfy $\lambda_S > \lambda_T$. If on the other hand $p_{i,S} = p_{i,T} \neq 0$, we have $\lambda_{i,S} = \lambda_{i,T} = 0$ from the optimality conditions. Then $\lambda_S - \nu_i = F_i^{-1}(1 - p_{i,S}) = F_i^{-1}(1 - p_{i,T}) = \lambda_T - \nu_i$. Again, adding ν_i on both sides, we obtain that the Lagrange multipliers should satisfy $\lambda_S = \lambda_T$. Thus, setting $\lambda(S) = \lambda_S$ for all $S \in \mathcal{S}$, we see that there exists a function $\lambda : \mathcal{S} \to \mathbb{R}$ satisfying (5).

Sufficiency of (5): Given $\boldsymbol{p}_{\mathcal{S}}$ and $\lambda : \mathcal{S} \to \mathbb{R}$ such that (5) holds for all $(i,S), (i,T) \in \mathcal{I}_{\mathcal{S}}$, we next exhibit a construction of marginal distributions $(F_i : i \in \mathcal{N})$ and utilities $(\nu_i : i \in \mathcal{N})$ for MDM. This construction will be such that it yields the given $(p_{i,S} : i \in S)$ as the corresponding choice probabilities from the optimality conditions in (4), for any assortment $S \in \mathcal{S}$.

For any product $i \in \mathcal{N}$, let $\mathcal{S}_i = \{S \in \mathcal{S} : i \in S\}$ denote the subcollection of assortments $S \in \mathcal{S}$ which contain the product i and let $m_i = |\mathcal{S}_i|$. Let l_i denote the number of assortments containing product i for which $p_{i,S} > 0$. Here $l_i = m_i$ when the choice probabilities $\{p_{i,S} : S \in \mathcal{S}_i\}$ are all non-zero. Equipped with this notation, we construct the marginal distribution $F_i(\cdot)$ for any product $i \in \mathcal{N}$ as follows:

(a) Consider any ordering $(S_1, S_2, \ldots, S_{l_i}, S_{l_i+1}, \ldots, S_{m_i})$ over the assortments in $\mathcal{S}_i$ for which $\lambda(S_1) \leq \lambda(S_2) \leq \ldots \leq \lambda(S_{l_i}) < \lambda(S_{l_i+1}) \leq \lambda(S_{l_i+2}) \leq \ldots \leq \lambda(S_{m_i})$. With l_i defined as the number of assortments in $\mathcal{S}_i$ for which $p_{i,S} > 0$, note that it is necessary to have $\lambda(S_{l_i}) < \lambda(S_{l_i+1})$ whenever $l_i < m_i$. This follows from the observations that $\lambda(\cdot)$ satisfies (5) and $p_{i,S_{l_i}} > 0 = p_{i,S_{l_i+1}}$. Further, due to the conditions in (5), the choice probabilities $(p_{i,S} : S \in \mathcal{S}_i)$ must necessarily satisfy the ordering $p_{i,S_1} \geq p_{i,S_2} \geq \ldots \geq p_{i,S_{l_i}} > 0$ and $p_{i,S_{l_i+1}} = p_{i,S_{l_i+2}} = \ldots, p_{i,S_{m_i}} = 0$.

(b) Construct the cumulative distribution function $F_i(\cdot)$ by first setting $F_i(\lambda(S_k)) = 1 - p_{i,S_k}$ for $k = 1, \cdots, l_i$. With this assignment, we complete the construction of the distribution F_i in between these points by connecting them with line segments as follows: For any two consecutive assortments S_k and S_{k+1} in the ordering satisfying $\lambda(S_k) < \lambda(S_{k+1})$, connect the respective points $(\lambda(S_k), 1 - p_{i,S_k})$ and $(\lambda(S_{k+1}), 1 - p_{i,S_{k+1}})$ with a line segment (see Figure EC.1). For $k \le l_i$, note that if the consecutive assortments S_k and S_{k+1} are such that $\lambda(S_k) = \lambda(S_{k+1})$, then the corresponding points $(\lambda(S_k), 1 - p_{i,S_k})$ and $(\lambda(S_{k+1}), 1 - p_{i,S_{k+1}})$ coincide and there is no need to connect them. Further note that $p_{i,S_k} > p_{i,S_{k+1}}$ when $\lambda(S_k) < \lambda(S_{k+1})$, because of (5), and hence the cumulative distribution function F_i is strictly increasing in the interval $[\lambda(S_1), \lambda(S_{l_i})]$.

(c) Lastly we construct the tails of the distribution F_i as follows: For the right tail, connect the points $(\lambda(S_{l_i}), 1 - p_{i,S_{l_i}})$ and $(\lambda(S_{l_i+1}), 1)$ with a line segment if $l_i < m_i$. We then have $F_i(x) = 1$ for every $x \ge \lambda(S_{l_i+1})$ and therefore $F_i^{-1}(1) = \lambda(S_{l_i+1})$. If $l_i = m_i$, connect the points $(\lambda(S_{l_i}), 1 - p_{i,S_{l_i}})$ and $(\lambda(S_{l_i}) + \delta, 1)$ by choosing any arbitrary $\delta > 0$ (see Figure EC.1). In this case, we will have $F_i(x) = 1$ for every $x \ge \lambda(S_{l_i}) + \delta$. For the left tail, if $p_{i,S_1} = 1$, then we have $F_i(x) = 0$ for every $x \le \lambda(S_1)$. Both the cumulative distribution functions drawn in Figure EC.1 illustrate this case. On the other hand, if $p_{i,S_1} < 1$, we use a line segment to connect $(\lambda(S_1), 1 - p_{i,S_1})$ and $(\lambda(S_1) - \delta, 0)$ by choosing an arbitrary $\delta > 0$. In this case, $F_i(x) = 0$ for every $x \le \lambda(S_1) - \delta$.

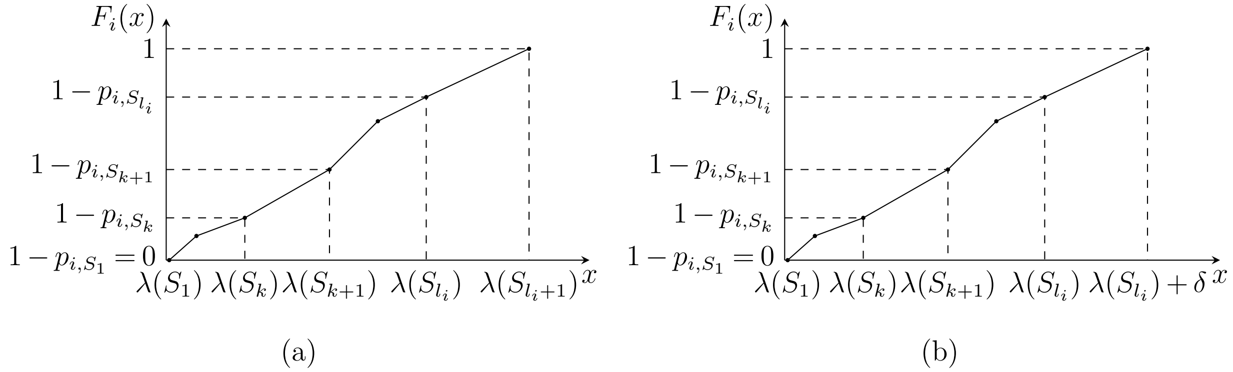


Figure EC.1 An illustration of the construction of the marginal distribution F_i when: (a) there is an assortment S for which $p_{i,S} = 0$ (the case where $l_i < m_i$) and (b) $p_{i,S} > 0$ for all assortments with product i (the case where $l_i = m_i$).

The above construction gives marginal distribution functions $(F_i : i \in \mathcal{N})$ which are absolutely continuous and strictly increasing within its support. We next show that the constructed marginal distributions yield the given choice probabilities $(p_{i,S} : i \in S)$, for any assortment $S \in \mathcal{S}$, when they are used in the optimality conditions (4) together with the assignment $\nu_i = 0$, for $i \in \mathcal{N}$. In other words, given $\boldsymbol{p}_{\mathcal{S}}$, we next verify that

$$F_i^{-1}(1 - p_{i,S}) - \lambda(S) + \lambda_{i,S} = 0, \quad \lambda_{i,S}\ p_{i,S} = 0, \quad \text{and} \quad \lambda_{i,S} \ge 0, \quad \forall (i,S) \in \mathcal{I}_{\mathcal{S}}.$$

For any $(i,S)\in\mathcal{I}_{\mathcal{S}}$ with $p_{i,S}>0$, we have from the construction of F_i that $F_i(\lambda(S))=1-p_{i,S}$. Then for such $p_{i,S}$, we see that the optimality condition $F_i^{-1}(1-p_{i,S})-\lambda(S)+\lambda_{i,S}=0$ readily holds since the optimality conditions also stipulate that $\lambda_{i,S}=0$ when $p_{i,S}>0$.

For any $(i,S)\in\mathcal{I}_{\mathcal{S}}$ such that $p_{i,S}=0$, we have from Steps (a) and (c) of the above construction that $\lambda(S)\geq\lambda(S_{l_i+1})=F_i^{-1}(1)=F_i^{-1}(1-p_{i,S})$. Then if we take $\lambda_{i,S}=\lambda(S)-\lambda(S_{l_i+1})$, we again readily have $F_i^{-1}(1-p_{i,S})-\lambda(S)+\lambda_{i,S}=0$. This completes the verification that for any choice data $\boldsymbol{p}_{\mathcal{S}}$ satisfying (5), there exists marginal distributions $\{F_i:i\in\mathcal{N}\}$ and deterministic utilities $\{\nu_i:i\in\mathcal{N}\}$ which yield $\boldsymbol{p}_{\mathcal{S}}$ as the MDM choice probabilities.

Lastly, checking whether the conditions in (5) are satisfied for given choice data $\boldsymbol{p}_{\mathcal{S}}$ is equivalent to testing if there exists an assignment for variables $(\lambda_S:S\in\mathcal{S})$ and $\epsilon>0$ such that,

$$\lambda_S\geq\lambda_T+\epsilon \quad \text{if} \quad p_{i,S}<p_{i,T},$$

$$\lambda_S=\lambda_T \quad \text{if} \quad p_{i,S}=p_{i,T}\neq 0,$$

for all $(i,S),(i,T)\in\mathcal{I}_{\mathcal{S}}$. This is possible in polynomial time by solving a linear program where the above conditions are formulated as constraints and maximizing ϵ and then checking if the optimal value is strictly positive. This linear program involves $|\mathcal{S}|$ variables for $(\lambda_S:S\in\mathcal{S})$ and one variable for ϵ, and at most $n|\mathcal{S}|$ constraints. ■

EC.1.2. Proof of Theorem 2

To prove MDM possesses positive measure in Theorem 2, we first present Lemma EC.1 below that deals with choice probabilities of the form obtained by the MNL model.

LEMMA EC.1. *For any fixed n, let $S,T\subseteq\mathcal{N}$ with $|S|,|T|\geq 2$. Let $i\in S\cap T$. Then there exists a set of positive integers $x_1,\ldots,x_n$ such that*

$$\frac{x_i}{\sum_{k\in S}x_k}\neq\frac{x_i}{\sum_{k\in T}x_k}, \tag{EC.3}$$

as long as $S\neq T$.

Proof. We prove Lemma EC.1 holds for a set of positive integers such that $x_k=2^k$ when $k\geq 1$. When $S\neq T$, to show $\frac{x_i}{\sum_{k\in S}x_k}\neq\frac{x_i}{\sum_{k\in T}x_k}$, it's equivalent to show $\sum_{k\in S}x_k\neq\sum_{k\in T}x_k$. It's obvious that $\sum_{k\in S}x_k\neq\sum_{k\in T}x_k$ when $S\subset T$ or $T\subset S$ for the selected x_k values. Next, we prove that $\sum_{k\in S}x_k\neq\sum_{k\in T}x_k$ when $S\not\subset T$ or $T\not\subset S$. Let $k_1=\arg\max_{k\in S\setminus T}x_k$ and $k_2=\arg\max_{k\in T\setminus S}x_k$. Without loss of generality, let $k_1>k_2$. Then, we have $k_1-k_2\geq 1$. We have

$$\sum_{k\in S}x_k-\sum_{k\in T}x_k\geq x_{k_1}-\sum_{k\in T\setminus S}x_k\geq x_{k_1}-\sum_{i=1}^{k_2}x_i=2^{k_1}-\frac{2*(1-2^{k_2})}{1-2}=2^{k_1}-2^{k_2+1}+2>0.$$

The first inequality is due to $k_1 \in S \setminus T$, and the second inequality is due to $T \setminus S \subseteq \{1, \cdots k_2\}$. The first equality is due to the formula of the sum of geometric series. This completes the proof. ■

Recall that the probability of choosing product i in assortment S under an MNL model is $p_{i,S} = \frac{e^{\nu_i}}{\sum_{j\in S} e^{\nu_j}}$. Then, there exists $\nu_i = \ln x_i$ for all $i \in \mathcal{N}$, such that the instance in Lemma EC.1 is an MNL instance.

Proof. Equipped with Lemma EC.1, we prove that MDM possesses positive measure as follows: Let $\boldsymbol{p}_{\mathcal{S}}$ be an instance generated as in Lemma EC.1. Then, such $\boldsymbol{p}_{\mathcal{S}}$ is an MNL instance and satisfies $p_{i,S} > 0$, $\forall (i,S) \in \mathcal{I}_{\mathcal{S}}$ and $p_{i,S} \neq p_{i,T}$, $\forall (i,S),(i,T) \in \mathcal{I}_{\mathcal{S}}$. Since MNL is a special case of MDM (Mishra et. al. (2014)), $\boldsymbol{p}_{\mathcal{S}}$ is MDM-representable.

We next show that any instance $\boldsymbol{p}'_{\mathcal{S}}$ that lies in the ball centered at $\boldsymbol{p}_{\mathcal{S}}$ with a specified radius $\epsilon > 0$ is an MDM instance. Let $0 < \epsilon < \min_{(k,S),(k,T)\in\mathcal{I}_{\mathcal{S}}} |p_{k,S} - p_{k,T}|$. We perturb $\boldsymbol{p}_{\mathcal{S}}$ to be $\boldsymbol{p}'_{\mathcal{S}}$ as follows: Arbitrarily choosing $(i,S),(j,S) \in \mathcal{I}_{\mathcal{S}}$, let

$$p'_{i,S} = p_{i,S} + \epsilon, \text{ and } p'_{j,S} = p_{j,S} - \epsilon,$$

and keep other entries of $\boldsymbol{p}'_{\mathcal{S}}$ the same as $\boldsymbol{p}_{\mathcal{S}}$ by setting

$$p'_{k,T} = p_{k,T}, \ \forall (k,T) \in \mathcal{I}_{\mathcal{S}}, \text{ with } (k,T) \neq (i,S) \text{ and } (k,T) \neq (j,S).$$

Then, $\boldsymbol{p}'_{\mathcal{S}}$ lies in the ball centered at $\boldsymbol{p}_{\mathcal{S}}$ with the radius $\epsilon > 0$ and we have $\sum_{i\in S} p'_{i,S} = 1$, for all $S \in \mathcal{S}$. Next, we show $\boldsymbol{p}'_{\mathcal{S}}$ is MDM-representable by showing that $\boldsymbol{p}'_{\mathcal{S}}$ satisfies the MDM-representable conditions (5) in Theorem 1.

Since $\boldsymbol{p}_{\mathcal{S}}$ is MDM-representable, we have

$$\lambda_S > \lambda_T \text{ if } p_{i,S} < p_{i,T} \, \forall (i,S),(i,T) \in \mathcal{I}_{\mathcal{S}}, \text{ and } \lambda_T > \lambda_S \text{ if } p_{j,T} < p_{j,S} \, \forall (j,S),(j,T) \in \mathcal{I}_{\mathcal{S}}.$$

Since $\epsilon < \min_{(k,S),(k,T)\in\mathcal{I}_{\mathcal{S}}} |p_{k,S} - p_{k,T}|$, we have

$$\lambda_S > \lambda_T \text{ if } p_{i,S} + \epsilon < p_{i,T} \, \forall (i,S),(i,T) \in \mathcal{I}_{\mathcal{S}}, \text{ and } \lambda_T > \lambda_S \text{ if } p_{j,T} < p_{j,S} - \epsilon \, \forall (j,S),(j,T) \in \mathcal{I}_{\mathcal{S}}.$$

By the construction of $\boldsymbol{p}'_{\mathcal{S}}$, equivalently, we have

$$\lambda_S > \lambda_T \text{ if } p'_{i,S} < p'_{i,T} \, \forall (i,S),(i,T) \in \mathcal{I}_{\mathcal{S}}, \text{ and } \lambda_T > \lambda_S \text{ if } p'_{j,T} < p'_{j,S} \, \forall (j,S),(j,T) \in \mathcal{I}_{\mathcal{S}}.$$

Thus, $\boldsymbol{p}'_{\mathcal{S}}$ is MDM-representable.

We next prove the choice probabilities represented by the MNL model and nested logit model possess zero Lebesgue measure one by one as follows.

1. *Measure zero of MNL.* To show the choice probabilities represented by MNL possess zero Lebesgue measure, it suffices to show that, for any $n \geq 3$, there exists an assortment collection $\mathcal{S}$ such that $\mu(\mathcal{P}_{\text{MNL}}(\mathcal{S})) = 0$. Consider the nested assortment collection, $\mathcal{S} = \{S_1, S_2, \cdots, S_n\}$ with $S_i = \{1, 2, \cdots, i\}$, $\forall i \leq n$. We have

$$\mathcal{P}_{\text{MNL}}(\mathcal{S}) = \{(x_{i,S} : i \in S, S \in \mathcal{S}) | x_{i,S} \geq 0, \forall i \in S, \forall S \in \mathcal{S}, \sum_{i \in S} x_{i,S} = 1, \forall S \in \mathcal{S},$$
$$\frac{x_{i,S}}{x_{j,S}} = \frac{x_{i,T}}{x_{j,T}}, \forall i, j \in S \cap T, \forall S, T \in \mathcal{S}\}.$$

We define the following set B by reducing the constraints in $\mathcal{P}_{\text{MNL}}(\mathcal{S})$:

$$\mathcal{P}_{\text{MNL}}(\mathcal{S}) \subseteq B := \{(x_{i,S} : i \in S, S \in \mathcal{S}) | x_{i,S} \geq 0, \forall i \in S, \forall S \in \mathcal{S}, \sum_{i \in S} x_{i,S} = 1, \forall S \in \mathcal{S},$$
$$\frac{x_{1,S_{n-1}}}{x_{2,S_{n-1}}} = \frac{x_{1,S_n}}{x_{2,S_n}}\}.$$

Consider $\prod_{S \in \mathcal{S}} \Delta_S$ where $\mathcal{S} = \{S_1, S_2, \cdots, S_n\}$ with $S_i = \{1, 2, \cdots, i\}$, $\forall i \leq n$, as a linear system with $\frac{n(n+1)}{2}$ nonnegative variables and n constraints where the constraints defined by the probability simplexes. We can see that the constraints are linearly independent because of the nested structure of the collection. It follows that $\dim(\prod_{S \in \mathcal{S}} \Delta_S) = \frac{n(n+1)}{2} - n = \frac{n^2 - n}{2}$. $\dim(B) = \frac{n(n+1)}{2} - n = \frac{n^2 - n}{2} - 2$, since $\frac{x_{1,S_{n-1}}}{x_{2,S_{n-1}}} = \frac{x_{1,S_n}}{x_{2,S_n}}$ is a surface in 2-dimensional space. Since $\dim(\mathcal{P}_{\text{MNL}}(\mathcal{S})) \leq \dim(B) < \dim(\prod_{S \in \mathcal{S}} \Delta_S)$, $\mu(\mathcal{P}_{\text{MNL}}(\mathcal{S})) = 0$ on $\prod_{S \in \mathcal{S}} \Delta_S$.

2. *Measure zero of nested logit.* Consider the same nested assortment collection as above with $n \geq 3$. The assumption on the nested logit model is that n alternatives are partitioned into nests. Within a nest, the IIA property holds.
   (a) When the nests are singletons or when all products are in the same nest, the nested logit model reduces to MNL. $\mathcal{P}_{\text{NL}}(\mathcal{S}) = \mathcal{P}_{\text{MNL}}(\mathcal{S})$. Thus, $\mu(\mathcal{P}_{\text{NL}}(\mathcal{S})) = \mu(\mathcal{P}_{\text{MNL}}(\mathcal{S})) = 0$.
   (b) When the number of nests is greater than 2 and less than n, there are at least 2 products in the same nest. This implies that there exists $i, j \in \mathcal{N}$, such that i, j are in the same nest. We have $i, j \in S_i \cap S_j$ because of the nested collection structure. We define the following set C by reducing the constraints in $\mathcal{P}_{\text{NL}}(\mathcal{S})$:

$$\mathcal{P}_{\text{NL}}(\mathcal{S}) \subseteq C := \{(x_{i,S} : i \in S, S \in \mathcal{S}) | x_{i,S} \geq 0, \forall i \in S, \forall S \in \mathcal{S}, \sum_{i \in S} x_{i,S} = 1, \forall S \in \mathcal{S},$$
$$\frac{x_{i,S_i}}{x_{j,S_i}} = \frac{x_{i,S_j}}{x_{j,S_j}}\}.$$

   We see that $\dim(C) = \dim(B)$. In this case, $\mu(\mathcal{P}_{\text{NL}}(\mathcal{S})) = 0$ on $\prod_{S \in \mathcal{S}} \Delta_S$.

To sum up, for all possible nests, $\mu(\mathcal{P}_{\text{NL}}(\mathcal{S})) = 0$. There are at most 2^n nests. The countable union of zero measure sets has measure zero. Thus, $\mu(\mathcal{P}_{\text{NL}}(\mathcal{S})) = 0$ on $\prod_{S \in \mathcal{S}} \Delta_S$.

■

EC.1.3. Proof of Lemma 2

Proof. We first prove a) in Lemma 2 by contradiction. From Theorem 1, if there exists some alternative i such that $p_{i,S} > p_{i,S\cap T}$, then we have $\lambda_S < \lambda_{S\cap T}$, which is equivalent to $\lambda_S < \lambda_{S\cap T}$ for all $j \in S\cap T$. This implies $\lambda_S \leq \lambda_{S\cap T}$ which gives $p_{j,S} \geq p_{j,S\cap T}$ for all $(j,S),(j,S\cap T)\in \mathcal{I}_{\mathcal{S}}$. Since $\sum_{j\in S} p_{j,S} = 1$, we get $\sum_{j\in S\cap T} p_{j,S\cap T} < 1$ contradicting the condition $\sum_{j\in S\cap T} p_{j,S\cap T} = 1$. For b), from Theorem 1, if $i,j \in S\cap T$, we have $p_{j,S} < p_{j,T} \Rightarrow \lambda_S > \lambda_T \Rightarrow \lambda_S \geq \lambda_T \Rightarrow p_{i,S} \leq p_{i,T}$. ■

EC.1.4. Proof of Theorem 3

Proof. *We prove a) of Theorem 3 first.* We use the following notations for the rank list model since any RUM can be described by a rank list model (see, e.g., Block and Marschak 1960). Let Σ_n denote the set of all permutations of n alternatives. Each element $\sigma \in \Sigma_n$ denotes a ranking of n alternatives. For instance, $\sigma = \{1 \succ 2 \succ 3\}$ means alternative 1 is more preferred than alternative 2 which is more preferred than alternative 3. The probability of each ranking is $P(\sigma)$ and $\sum_{\sigma\in\Sigma_n} P(\sigma) = 1$. We prove the result case by case.

1. $n = 2$: Here $\mathcal{P}_{\text{MDM}}(\mathcal{S}) = \mathcal{P}_{\text{RUM}}(\mathcal{S})$. This is straightforward since all probabilities that satisfy $0 \leq p_{1,\{1,2\}} \leq p_{1,\{1\}} = 1$, and $0 \leq p_{2,\{1,2\}} \leq p_{2,\{2\}} = 1$ where $p_{1,\{1,2\}} + p_{2,\{1,2\}} = 1$, are representable by both models.
2. $n = 3$: Lemma 2 implies that $\mathcal{P}_{\text{MDM}}(\mathcal{S}) \subseteq \mathcal{P}_{\text{REG}}(\mathcal{S})$. When $n = 3$, all the possible assortments include $\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}$. The regular model imposes the following regularity constraints:

$$p_{1,\{1,2\}} \geq p_{1,\{1,2,3\}},\ p_{2,\{1,2\}} \geq p_{2,\{1,2,3\}},$$
$$p_{1,\{1,3\}} \geq p_{1,\{1,2,3\}},\ p_{3,\{1,3\}} \geq p_{3,\{1,2,3\}},$$
$$p_{2,\{2,3\}} \geq p_{2,\{1,2,3\}},\ p_{3,\{2,3\}} \geq p_{3,\{1,2,3\}}.$$

$\mathcal{P}_{\text{RUM}}(\mathcal{S}) = \mathcal{P}_{\text{REG}}(\mathcal{S})$ for any given $\mathcal{S}$ since

$$P(\{1\succ 2\succ 3\}) = p_{2,\{2,3\}} - p_{2,\{1,2,3\}} \geq 0 \text{ and } P(\{1\succ 3\succ 2\}) = p_{3,\{2,3\}} - p_{3,\{1,2,3\}} \geq 0,$$
$$P(\{2\succ 1\succ 3\}) = p_{1,\{1,3\}} - p_{1,\{1,2,3\}} \geq 0 \text{ and } P(\{2\succ 3\succ 1\}) = p_{3,\{1,3\}} - p_{3,\{1,2,3\}} \geq 0,$$
$$P(\{3\succ 1\succ 2\}) = p_{1,\{1,2\}} - p_{1,\{1,2,3\}} \geq 0 \text{ and } P(\{3\succ 2\succ 1\}) = p_{2,\{1,2\}} - p_{2,\{1,2,3\}} \geq 0,$$

where $\sum_{\sigma\in\Sigma_n} P(\sigma) = 3 - 2 = 1$.

We next show that $\mathcal{P}_{\text{MDM}}(\mathcal{S}) \subset \mathcal{P}_{\text{RUM}}(\mathcal{S})$ for $n = 3$ by giving an example of choice probabilities with $\mathcal{S} = \{\{1,2,3\},\{1,2\},\{1,3\},\{2,3\}\}$ in Table EC.1 that can be represented by RUM but not by MDM.

Table EC.1 Choice probabilities that cannot be represented by MDM for $n = 3$.

Alternative	A={1,2,3}	B={1,2}	C={1,3}	D={2,3}
1	1/3	5/9	4/9	-
2	1/3	4/9	-	5/9
3	1/3	-	5/9	4/9

This collection of choice probabilities $\boldsymbol{p}_{\mathcal{S}}$ cannot be represented by MDM because $p_{1,B} > p_{1,C}$, $p_{2,D} > p_{2,B}$, $p_{3,C} > p_{3,D}$ implies $\lambda_B < \lambda_C$, $\lambda_D < \lambda_B$ and $\lambda_C < \lambda_D$. This gives $\lambda_D < \lambda_B < \lambda_C < \lambda_D$ which is inconsistent. So, $\boldsymbol{p}_{\mathcal{S}}$ in Table EC.1 cannot be represented by MDM. On the other hand, it is straightforward to check that by setting the ranking probabilities for RUM as follows: $P(\{1 \succ 2 \succ 3\}) = 2/9$, $P(\{1 \succ 3 \succ 2\}) = 1/9$, $P(\{2 \succ 1 \succ 3\}) = 1/9$, $P(\{2 \succ 3 \succ 1\}) = 2/9$, $P(\{3 \succ 1 \succ 2\}) = 2/9$, $P(\{3 \succ 2 \succ 1\}) = 1/9$, we obtain the choice probabilities in Table EC.1. This implies $\boldsymbol{p}_{\mathcal{S}}$ in table EC.1 can be represented by RUM but not MDM.

3. $n \geq 4$: We show $\mathcal{P}_{\text{MDM}}(\mathcal{S}) \not\subset \mathcal{P}_{\text{RUM}}(\mathcal{S})$ and $\mathcal{P}_{\text{RUM}}(\mathcal{S}) \not\subset \mathcal{P}_{\text{MDM}}(\mathcal{S})$ by providing two examples: (1) $\boldsymbol{p}_{\mathcal{S}}$ can be represented by RUM but not MDM and (2) $\boldsymbol{p}_{\mathcal{S}}$ can be represented by MDM but not RUM when $\mathcal{S} = \{\{1,2,3,4\},\{1,2,3\},\{1,2,4\},\{1,2\}\}$. The examples are provided for $n = 4$. For larger n, we can add the alternatives in the assortments and set the choice probabilities for these added alternatives to be zero.

To show $\mathcal{P}_{\mathit{MDM}}(\mathcal{S}) \cap \mathcal{P}_{\mathit{RUM}}(\mathcal{S}) \neq \emptyset$*:* Firstly, we observe that the multinomial logit choice probabilities can be obtained from both RUM and MDM. This follows from using independent and identically distributed Gumbel distributions for the joint distribution of the random parts of utilities for RUM (see, e.g., Ben-Akiva and Lerman 1985) and identical exponential distributions for the marginals of the random parts of utilities for MDM (see, e.g., Mishra et. al. 2014). Hence, the intersection between the two sets is nonempty for any n.

To show $\mathcal{P}_{\mathit{RUM}}(\mathcal{S}) \not\subset \mathcal{P}_{\mathit{MDM}}(\mathcal{S})$*:* Consider the choice probabilities in Table EC.2.

Table EC.2 Choice probabilities can be represented by RUM but not by MDM for $n = 4$.

Alternative	A={1,2,3,4}	B={1,2,3}	C={1,2,4}	D={1,2}
1	3/20	7/20	1/4	1/2
2	3/20	1/4	7/20	1/2
3	7/20	2/5	-	-
4	7/20	-	2/5	-

This can be recreated by RUM using the distribution over the ranking as follows:

$$\begin{array}{lll}
P(\{1 \succ 2 \succ 3 \succ 4\}) = 1/40 & P(\{1 \succ 2 \succ 4 \succ 3\}) = 1/40 & P(\{1 \succ 3 \succ 2 \succ 4\}) = 1/40 \\
P(\{1 \succ 3 \succ 4 \succ 2\}) = 1/40 & P(\{1 \succ 4 \succ 2 \succ 3\}) = 1/40 & P(\{1 \succ 4 \succ 3 \succ 2\}) = 1/40 \\
P(\{2 \succ 1 \succ 3 \succ 4\}) = 1/40 & P(\{2 \succ 1 \succ 4 \succ 3\}) = 1/40 & P(\{2 \succ 3 \succ 1 \succ 4\}) = 1/40 \\
P(\{2 \succ 3 \succ 4 \succ 1\}) = 1/40 & P(\{2 \succ 4 \succ 1 \succ 3\}) = 1/40 & P(\{2 \succ 4 \succ 3 \succ 1\}) = 1/40 \\
P(\{3 \succ 1 \succ 2 \succ 4\}) = 1/20 & P(\{3 \succ 1 \succ 4 \succ 2\}) = 1/20 & P(\{3 \succ 2 \succ 1 \succ 4\}) = 1/10 \\
P(\{3 \succ 2 \succ 4 \succ 1\}) = 1/10 & P(\{3 \succ 4 \succ 1 \succ 2\}) = 1/40 & P(\{3 \succ 4 \succ 2 \succ 1\}) = 1/40 \\
P(\{4 \succ 1 \succ 3 \succ 2\}) = 1/10 & P(\{4 \succ 1 \succ 3 \succ 2\}) = 1/10 & P(\{4 \succ 2 \succ 1 \succ 3\}) = 1/20 \\
P(\{4 \succ 2 \succ 3 \succ 1\}) = 1/20 & P(\{4 \succ 3 \succ 1 \succ 2\}) = 1/40 & P(\{4 \succ 3 \succ 2 \succ 1\}) = 1/40
\end{array}$$

Now $p_{1,B} > p_{1,C}$ implies $\lambda_B < \lambda_C$ and $p_{2,B} > p_{2,C}$ implies $\lambda_B > \lambda_C$. Hence $\boldsymbol{p}_{\mathcal{S}}$ in Table EC.2 is not representable by MDM.

To show $\mathcal{P}_{\text{MDM}}(\mathcal{S}) \not\subset \mathcal{P}_{\text{RUM}}(\mathcal{S})$*:* Next consider the choice probabilities in Table EC.3.

Table EC.3 Choice probabilities can be represented by MDM but not by RUM for $n=4$.

Alternative	A={1,2,3,4}	B={1,2,3}	C={1,2,4}	D={1,2}
1	0.1	0.2	0.2	0.25
2	0.2	0.25	0.25	0.75
3	0.2	0.55	-	-
4	0.5	-	0.55	-

Here $p_{1,A} < p_{1,B} = p_{1,C} < p_{1,D}$ implies $\lambda_A > \lambda_B = \lambda_C > \lambda_D$, and $p_{2,A} < p_{2,B} = p_{2,C} < p_{2,D}$ implies $\lambda_A > \lambda_B = \lambda_C > \lambda_D$, and $p_{3,A} < p_{3,B}$ implies $\lambda_A > \lambda_B$, and $p_{4,A} < p_{4,C}$ implies $\lambda_A > \lambda_C$. So we have $\lambda_A > \lambda_B = \lambda_C > \lambda_D$ which is easy to enforce and so $\boldsymbol{p}_{\mathcal{S}}$ can be represented by MDM. A necessary condition for $\boldsymbol{p}_{\mathcal{S}}$ to be representable by RUM are the Block-Marshak conditions provided in Block and Marschak (1960) (also see Theorem 1 in Fiorini 2004). If the choice probabilities are representable by RUM, one of these conditions is given by $p_{1,A} + p_{1,D} \geq p_{1,B} + p_{1,C}$. Here $p_{1,A} + p_{1,D} = 0.1 + 0.25 = 0.35 < 0.4 = 0.2 + 0.2 = p_{1,B} + p_{1,C}$. So, $\boldsymbol{p}_{\mathcal{S}}$ is not representable by RUM.

We prove b) of Theorem 3 as follows. We know that $\mathcal{P}_{\text{MDM}}(\mathcal{S}) = \mathcal{P}_{\text{RUM}}(\mathcal{S}) = \mathcal{P}_{\text{REG}}(\mathcal{S})$ when $n=2$ and $\mathcal{P}_{\text{RUM}}(\mathcal{S}) \subseteq \mathcal{P}_{\text{REG}}(\mathcal{S})$ and closure$(\mathcal{P}_{\text{MDM}}(\mathcal{S})) \subseteq \mathcal{P}_{\text{REG}}(\mathcal{S})$ for any $\mathcal{S}$. To show b), we just need to show $\mathcal{P}_{\text{REG}}(\mathcal{S}) \subseteq \mathcal{P}_{\text{RUM}}(\mathcal{S})$ and $\mathcal{P}_{\text{REG}}(\mathcal{S}) \subseteq$ closure$(\mathcal{P}_{\text{MDM}}(\mathcal{S}))$ when $\mathcal{S}$ is nested or laminar. It suffices to show $\mathcal{P}_{\text{REG}}(\mathcal{S})' \subseteq \mathcal{P}_{\text{RUM}}(\mathcal{S})$ and $\mathcal{P}_{\text{REG}}(\mathcal{S})' \subseteq$ closure$(\mathcal{P}_{\text{MDM}}(\mathcal{S}))$ when $\mathcal{S}$ is nested or laminar, since $\mathcal{P}_{\text{REG}}(\mathcal{S}) \subseteq \mathcal{P}_{\text{REG}}(\mathcal{S})'$.

$\mathcal{P}_{\text{REG}}(\mathcal{S}) \subseteq$ closure$(\mathcal{P}_{\text{MDM}}(\mathcal{S}))$ with nested or laminar $\mathcal{S}$: Under a nested collection $\mathcal{S} = \{S_1, S_2, \dots, S_m\}$ with $S_1 \subset S_2 \subset \dots \subset S_m$, we have

$$\mathcal{P}_{\text{REG}}(\mathcal{S})' = \Big\{\boldsymbol{x} \in \mathbb{R}^{\mathcal{I}_{\mathcal{S}}} : x_{i,S} \geq 0, \forall (i,S) \in \mathcal{I}_{\mathcal{S}}, \sum_{i \in S} x_{i,S} = 1, \forall S \in \mathcal{S}, \, x_{i,S_k} \leq x_{i,S_j} \; \forall j,k \in [m], j < k, i \in S_i\Big\},$$

where $[m]$ denotes $\{1, 2, \dots, m\}$. Under a laminar collection, we have

$$\mathcal{P}_{\text{REG}}(\mathcal{S})' = \Big\{\boldsymbol{x} \in \mathbb{R}^{\mathcal{I}_{\mathcal{S}}} : x_{i,S} \geq 0, \forall (i,S) \in \mathcal{I}_{\mathcal{S}}, \sum_{i} x_{i,S} = 1, \forall S \in \mathcal{S}, \, x_{i,S} \leq x_{i,T} \; \forall T \subset S, (i,S), (i,T) \in \mathcal{I}_{\mathcal{S}}\Big\}.$$

$\mathcal{P}_{\text{REG}}(\mathcal{S})' \subseteq$ closure$(\mathcal{P}_{\text{MDM}}(\mathcal{S}))$ *with a nested or laminar* $\mathcal{S}$*:* We show that for any $\boldsymbol{p}_{\mathcal{S}} \in \mathcal{P}_{\text{REG}}(\mathcal{S})'$, we can construct $\boldsymbol{\lambda}_{\mathcal{S}}$ such that $(\boldsymbol{p}_{\mathcal{S}}, \boldsymbol{\lambda}_{\mathcal{S}}) \in \Pi'_{\mathcal{S}}$, where

$$\Pi'_{\mathcal{S}} := \Big\{(\boldsymbol{x}, \boldsymbol{\lambda}) \in \mathbb{R}^{\mathcal{I}_{\mathcal{S}}} \times \mathbb{R}^{\mathcal{S}} : x_{i,S} \geq 0, \forall (i,S) \in \mathcal{I}_{\mathcal{S}}, \sum_{i \in S} x_{i,S} = 1, \forall S \in \mathcal{S},$$
$$\lambda_S \geq \lambda_T \text{ if } x_{i,S} \leq x_{i,T}, \forall (i,S), (i,T) \in \mathcal{I}_{\mathcal{S}}\Big\}.$$

Suppose that $\mathcal{S} = \{S_1, S_2, \cdots, S_m\}$ is a nested collection with $S_1 \subset S_2 \subset \ldots \subset S_m$. Then we take any $\boldsymbol{\lambda}_{\mathcal{S}}$ satisfying $\lambda_{S_1} \leq \lambda_{S_2} \leq \cdots \leq \lambda_{S_m}$. Since $p_{i,S_j} \geq p_{i,S_k}$ for any $j \leq k$ (due to $\boldsymbol{p}_{\mathcal{S}} \in \mathcal{P}_{\text{REG}}(\mathcal{S})'$), the resulting $(\boldsymbol{p}_{\mathcal{S}}, \boldsymbol{\lambda}_{\mathcal{S}}) \in \Pi_{\mathcal{S}}^{'}$. As a result, $\boldsymbol{p}_{\mathcal{S}} \in \text{closure}(\mathcal{P}_{\text{MDM}}(\mathcal{S}))$.

If $\mathcal{S}$ is laminar, we take any $\boldsymbol{\lambda}_{\mathcal{S}}$ such that $\lambda_S \leq \lambda_T$ if $S, T \in \mathcal{S}$ with $S \subset T$. Since $p_{i,S} \geq p_{i,T}$ due to the regularity $\boldsymbol{p}_{\mathcal{S}} \in \mathcal{P}_{\text{REG}}(\mathcal{S})'$, the resulting $(\boldsymbol{p}_{\mathcal{S}}, \boldsymbol{\lambda}_{\mathcal{S}}) \in \Pi_{\mathcal{S}}^{'}$. Hence $\boldsymbol{p}_{\mathcal{S}} \in \text{closure}(\mathcal{P}_{\text{MDM}}(\mathcal{S}))$ and $\mathcal{P}_{\text{REG}}(\mathcal{S}) \subseteq \mathcal{P}_{\text{REG}}(\mathcal{S})' \subseteq \text{closure}(\mathcal{P}_{\text{MDM}}(\mathcal{S}))$ with a nested or laminar $\mathcal{S}$.

$\mathcal{P}_{REG}(\mathcal{S})' \subseteq \mathcal{P}_{RUM}(\mathcal{S})$ with a nested or laminar $\mathcal{S}$: We next show that, for any $\boldsymbol{p}_{\mathcal{S}} \in \mathcal{P}_{\text{REG}}(\mathcal{S})'$, there exists a probability distribution $(P(\sigma) : \sigma \in \Sigma_n)$ such that $\boldsymbol{p}_{\mathcal{S}} \in \mathcal{P}_{\text{RUM}}(\mathcal{S})$.

(1) For a nested collection $\mathcal{S}$, we prove $\mathcal{P}_{\text{REG}}(\mathcal{S})' \subseteq \mathcal{P}_{\text{RUM}}(\mathcal{S})$. Without loss of generality, let $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ be $S_k = \{1, \ldots, k\}$ for $k = 1, \cdots, m$. Next, for any $\boldsymbol{p}_{\mathcal{S}} \in \mathcal{P}_{\text{REG}}(\mathcal{S})'$, we prove the existence of a probability distribution $(P(\sigma) : \sigma \in \Sigma_m)$ such that $\boldsymbol{p}_{\mathcal{S}} \in \mathcal{P}_{\text{RUM}}(\mathcal{S})$ from the point of view of polyhedral combinations. To show the existence of a probability distribution $(P(\sigma) : \sigma \in \Sigma_m)$ for $\boldsymbol{p}_{\mathcal{S}} \in \mathcal{P}_{\text{RUM}}(\mathcal{S})$ is equivalent to showing $\boldsymbol{p}_{\mathcal{S}}$ lies in the multiple choice polytope characterized as,

$$\text{convex hull of } \{(\mathbb{I}[\sigma, i, S] : i \in S, S \in \mathcal{S}) \in \{0,1\}^{\sum_{S \in \mathcal{S}} |S|} : \sigma \in \Sigma_m\},$$

where $\mathbb{I}[\sigma, i, S] = 1$ if and only if $i = \arg\min_{j \in S} \sigma(j)$ (see Section 3 in Fiorini 2004 and Lemma 2.5 of Jagabathula and Rusmevichientong 2019).

Now, we show that $\boldsymbol{p}_{\mathcal{S}}$ lies in the multiple choice polytope via a graph representation of the multiple choice polytope following the steps in Section 3 in Fiorini (2004). Let $\boldsymbol{D} = (\mathcal{N}_0, A)$ be a simple, acyclic directed graph, and let $m+1$ be the source node and 0 be the sink node of $\boldsymbol{D}$, where $\mathcal{N}_0 = \{1, \cdots, m\} \cup \{m+1, 0\}$. We encode each $m+1-0$ directed path Π of its arc set A in $\boldsymbol{D}$ by means of the indicator characteristic vector in the set $\{(\mathbb{I}[\sigma, i, S] : i \in S, S \in \mathcal{S}) \in \{0,1\}^{\sum_{S \in \mathcal{S} |S|}} : \sigma \in \Sigma_m\} \in \mathbb{R}^A$, which we denote r^{Π}. The convex hull of the vectors r^{Π}, for a $m+1-0$ directed path Π in $\boldsymbol{D}$, is referred to as the $m+1-0$ directed path polytope of $\boldsymbol{D}$. For a node v of $\boldsymbol{D}$, let $\delta^-(v) = \{(w, v) : w \in \mathcal{N}_0, (w, v) \in A\}$ represent the nodes incoming to node v and $\delta^+(v) = \{(v, w) : w \in \mathcal{N}_0, (v, w) \in A\}$ represent the nodes outgoing from node v. For $B \subseteq A$, let $r(B) = \sum\{r(v, w) : (w, v) \in B\}$. Let M be the matrix whose rows are indexed by modes of $\boldsymbol{D}$ such that the entry corresponding to node v and arc a equals to 1 if a enters v, and -1 if a leaves 0, and 0 else. It's well known that M is totally unimodular (Schrijver 1998). This implies that the polyhedron $\{r \in \mathbb{R}^A : Mr = d, r \geq 0\}$ has all its vertices integer for every integral vector $d \in \mathbb{R}^A$. Assume that $\delta(m+1)^- = \delta(0)^+ = \emptyset$.

Lemma EC.2 **(Theorem 2 in Fiorini 2004)**. *A point $r \in \mathbb{R}^A$ belongs to the $m+1-0$ directed path polytope $\boldsymbol{D}$ if and only if*

$$r(\delta^-(v)) - r(\delta^+(v)) = 0, \quad \forall v \in \mathcal{N}_0 \setminus \{m+1, 0\}, \tag{EC.4}$$

$$r(\delta^-(0)) = 1, \tag{EC.5}$$

$$r(v,w) \geq 0, \quad \forall (w,v) \in A. \tag{EC.6}$$

In network flows, (EC.4)-(EC.6) *define a flow of value 1 in the network* $\boldsymbol{D} = (\mathcal{N}_0, A)$, *with source node* $m+1$ *and sink node* 0.

By Lemma EC.2, to show $\boldsymbol{p}_{\mathcal{S}}$ lies in the multiple choice polytope, we need to show $(r(w,v):(w,v)\in A)$ based on $\boldsymbol{p}_{\mathcal{S}}$ satisfying (EC.4) – (EC.6) in Lemma EC.2. We demonstrate $(r(w,v):(w,v)\in A)$ under $\boldsymbol{p}_{\mathcal{S}}$ as follows:

$$\begin{aligned} r(m+1,j) &= \sum_{\sigma\in\Sigma_m:\arg\min_{i\in S_m}\sigma(i)=j} P(\sigma) = p_{j,S_m}, \quad \forall j = 1,\cdots,m, \\ r(i,j) &= \sum_{\sigma\in\Sigma_m:\arg\min_{k\in S_{i-1}}\sigma(k)=j} P(\sigma) - \sum_{\sigma\in\Sigma_m:\arg\min_{k\in S_i}\sigma(k)=j} P(\sigma) \\ &= p_{j,S_{i-1}} - p_{j,S_i}, \quad \forall i = 2,\cdots,m,\ 1\leq j\leq i-1, \\ r(1,0) &= 1. \end{aligned}$$

We provide Figure EC.2 to illustrate $(r(w,v):(w,v)\in A)$ in the graph $\boldsymbol{D}$.

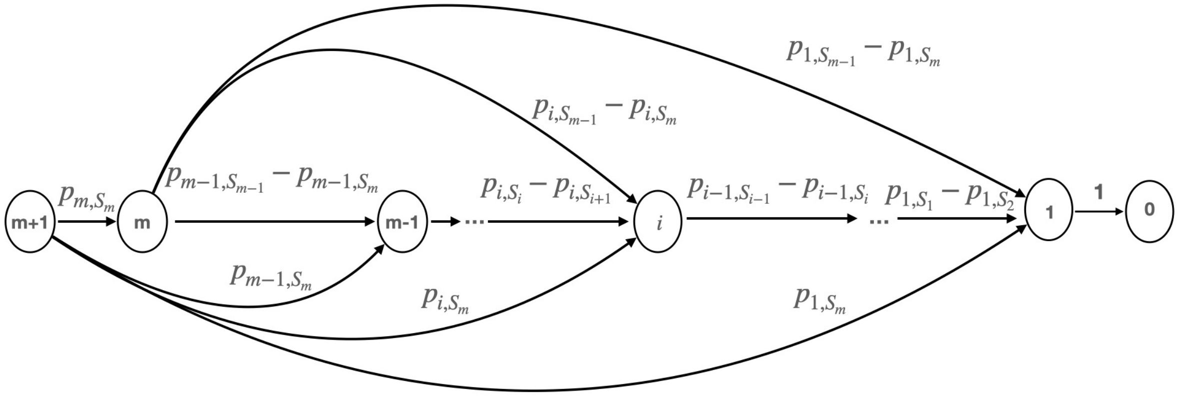


Figure EC.2 An illustration of D given $p_{\mathcal{S}} \in \mathcal{P}_{\mathsf{reg}}(\mathcal{S})'$ with a nested collection $\mathcal{S}$

Next, we verify such $(r(w,v):(w,v)\in A)$ satisfies (EC.4)-(EC.6). For (EC.4), define $p_{j,S_{m+1}} = 0$ for all j. For a node $j\in\{2,\cdots,m\}$,

$$\begin{aligned} r(\delta^-(j)) &= \sum_{k=j+1}^{m+1} r(k,j) \\ &= \sum_{k=j}^{m} p_{j,S_k} - p_{j,S_{k+1}} = p_{j,S_j} - p_{j,S_{j+1}} + p_{j,S_{j+1}} - p_{j,S_{j+2}} + \cdots + p_{j,S_m} - p_{j,S_{m+1}} \\ &= p_{j,S_j} - p_{j,S_{m+1}} = p_{j,S_j}. \\ r(\delta^+(j)) &= \sum_{k=1}^{j-1} p_{k,S_j-1} - p_{k,S_j} \end{aligned}$$

$$
\begin{aligned}
&= p_{1,S_{j-1}} - p_{1,S_j} + p_{2,S_{j-1}} - p_{2,S_j} + \cdots + p_{j-1,S_{j-1}} - p_{j-1,S_j} \\
&= \sum_{k=1}^{j-1} p_{k,S_{j-1}} - \sum_{k=1}^{j-1} p_{k,S_j} = 1 - \sum_{k=1}^{j-1} p_{k,S_j} = p_{j,S_j}.
\end{aligned}
$$

For the node 1,

$$
\begin{aligned}
r(\delta^-(1)) &= \sum_{k=2}^{m+1} r(k,1) \\
&= p_{1,S_1} - p_{1,S_2} + p_{1,S_2} - p_{1,S_3} + \cdots + p_{1,S_m} - p_{1,S_{m+1}} \\
&= p_{1,S_1} - p_{1,S_{m+1}} = 1 - 0 = 1 = r(\delta^+(1)).
\end{aligned}
$$

Therefore, $r(\delta^-(j)) = r(\delta^+(j))$ for $j \in \{1, 2, \cdots, m\}$. (EC.4) is satisfied by $(r(w,v) : (w,v) \in A)$. For (EC.5), $r(\delta^-(0)) = r(1,0) = 1$. For (EC.6), $r(m+1, j) = p_{j,S_m} \geq 0$, $\forall j = 1, \cdots, m$ because of the nonnegativity of choice probabilities. We have $r(i,j) = p_{j,S_{i-1}} - p_{j,S_i} \geq 0$, $\forall i = 2, \cdots, m$, $1 \leq j \leq i-1$ since p_S satisfies regularity,then, $\boldsymbol{p}_{\mathcal{S}} \in \mathcal{P}_{\text{REG}}(\mathcal{S})'$. Further $r(1,0) = 1 > 0$.

The assignment $(r(w,v) : (w,v) \in A)$ satisfies (EC.4)-(EC.6) in Lemma EC.2. Therefore, for any $\boldsymbol{p}_{\mathcal{S}} \in \mathcal{P}_{\text{REG}}(\mathcal{S})'$, there exists a probability distribution $(P(\sigma) : \sigma \in \Sigma_m)$ such that $\boldsymbol{p}_{\mathcal{S}} \in \mathcal{P}_{\text{RUM}}(\mathcal{S})$. This implies $\mathcal{P}_{\text{REG}}(\mathcal{S}) \subseteq \mathcal{P}_{\text{REG}}(\mathcal{S})' \subseteq \mathcal{P}_{\text{RUM}}(\mathcal{S})$ under the nested collection $\mathcal{S}$.

(2) We prove $\mathcal{P}_{\text{REG}}(\mathcal{S})' \subseteq \mathcal{P}_{\text{RUM}}(\mathcal{S})$ under the laminar collection. By the definition of the laminar collection, we know that for $S, T \in \mathcal{S}$, either $S \subset T$, or $T \subset S$, or $S \cap T = \emptyset$. Then, it suffices to construct a distribution $P(\cdot)$ for $\hat{\mathcal{S}} \subset \mathcal{S}$ such that $\hat{\mathcal{S}}$ is a nested collection. Following the proof in (1), we have $\mathcal{P}_{\text{REG}}(\mathcal{S}) \subseteq \mathcal{P}_{\text{REG}}(\mathcal{S})' \subseteq \mathcal{P}_{\text{RUM}}(\mathcal{S})$ under a laminar collection.

■

EC.1.5. Proof of Proposition 1

Proof. *To show the "if" direction:* Since $\boldsymbol{p}_{\mathcal{S}}$ is represented by MDM and $p_{i,S} > 0$ for all $(i,S) \in \mathcal{I}_{\mathcal{S}}$, for each pair of assortment $S, T \in \mathcal{S}$ and alternatives $i \in S \cap T$, there exists marginals F_i, F_j, Lagrange multipliers λ_S and λ_T, $\lambda_{i,S}$ and $\lambda_{i,T}$ and the followings are satisfied:

$$
\begin{aligned}
&\nu_i + F_i^{-1}(1 - p_{i,S}) - \lambda_S + \lambda_{i,S} = 0, \ \lambda_{i,S}\ p_{i,S} = 0, \quad \text{and} \quad \lambda_{i,S} \geq 0, \ \forall i \in S, \quad \text{and} \\
&\nu_i + F_i^{-1}(1 - p_{i,T}) - \lambda_S + \lambda_{i,T} = 0, \ \lambda_{i,T}\ p_{i,T} = 0, \quad \text{and} \quad \lambda_{i,T} \geq 0, \ \forall i \in T.
\end{aligned}
$$

Since $p_{i,S} > 0$ and $p_{i,T} > 0$, we have $\lambda_{i,S} = \lambda_{i,T} = 0$. We then have

$$
\begin{aligned}
&p_{i,S} = 1 - F_i(\lambda_S - \nu_i) \text{ and } p_{j,S} = 1 - F_j(\lambda_S - \nu_j), \quad \text{and,} \\
&p_{i,T} = 1 - F_i(\lambda_T - \nu_i) \text{ and } p_{j,T} = 1 - F_j(\lambda_T - \nu_j).
\end{aligned}
$$

This leads to the following:

$$
\nu_j - \nu_i = F_i^{-1}(1 - p_{i,S}) - F_j^{-1}(1 - p_{j,S}) = F_i^{-1}(1 - p_{i,T}) - F_j^{-1}(1 - p_{j,T}).
$$

Then,

$$\exp(v_j - v_i) = \frac{\exp(F_i^{-1}(1-p_{i,S}))}{\exp(F_j^{-1}(1-p_{j,S}))} = \frac{\exp(F_i^{-1}(1-p_{i,T}))}{\exp(F_j^{-1}(1-p_{j,T}))}.$$

Setting the strictly decreasing functions $f_i : [0,1) \to \mathbb{R}_+$ by $f_i(0) = \infty$ and $f_i(p) = \exp[F_i^{-1}(1-p)]$, for all $p > 0, \forall i \in \mathcal{N}$, we obtain:

$$\exp(v_j - v_i) = \frac{f_i(p_{i,S})}{f_j(p_{j,S})} = \frac{f_i(p_{i,T})}{f_j(p_{j,T})}$$

To show the "only if" direction:. Given $\boldsymbol{p_S}$ and $\{f_i : i \in \mathcal{N}\}$ satisfies Generalized Ordinal IIA, we construct the deterministic utilities $\boldsymbol{\nu} = \{\nu_i : i \in \mathcal{N}\}$ and the marginal distributions $\{F_i : i \in \mathcal{N}\}$ for MDM such that this construction yields the given $(p_{i,S} : i \in S)$ as the corresponding choice probabilities from the KKT optimality conditions in (4), for any assortment $S \in \mathcal{S}$.

Let $F_i(x) = \log^{-1}(f_i(1-x))$, equivalently $F_i^{-1}(1-x) = \log(f_i(x))$. For any fixed product i, set $\nu_i := 0$. For any other products such that $j \neq i$, set $\nu_j := F_i^{-1}(1-p_{i,\{i,j\}}) - F_j^{-1}(1-p_{j,\{i,j\}})$.

Take arbitrary assortment $S \in S$, and pick any two distinct products $j,k \in S$. There are two exclusive cases.

Case 1: $i \in \{j,k\}$. Without loss of generality, set $j = i$. Then,

$$\begin{aligned}\nu_k - \nu_i &= F_i^{-1}(1-p_{i,\{i,k\}}) - F_k^{-1}(1-p_{k,\{i,k\}}) \\ &= \log(\frac{f_i(p_{i,\{i,k\}})}{f_k(p_{k,\{i,k\}})}) \\ &= \log(\frac{f_i(p_{i,S})}{f_k(p_{k,S})}) && \text{(Generalized Ordinal IIA)} \\ &= F_i^{-1}(1-p_{i,S}) - F_k^{-1}(1-p_{k,S}).\end{aligned}$$

We next verify

$$\nu_i + F_i^{-1}(1-p^*_{i,S}) - \lambda_S + \lambda_{i,S} = 0,\ \lambda_{i,S} p_{i,S} = 0, \ \text{ and } \ \lambda_{i,S} \geq 0,\ \forall i \in S.$$

Since $p_{i,S} > 0$ for all $i \in S$, we have $\lambda_{i,S} = 0$ in the KKT conditions. Then, there exists λ_S such that $\nu_k + F_k^{-1}(1-p_{k,S}) - \lambda_S = 0$ and $\nu_i + F_i^{-1}(1-p_{i,S}) - \lambda_S = 0$ since we have proved that $\nu_k + F_k^{-1}(1-p_{k,S}) = \nu_i + F_i^{-1}(1-p_{i,S})$.

Case 2: $i \notin \{j,k\}$. Then,

$$\begin{aligned}\nu_k - \nu_j &= F_i^{-1}(1-p_{i,\{i,k\}}) - F_k^{-1}(1-p_{k,\{i,k\}}) + F_j^{-1}(1-p_{j,\{i,j\}}) - F_i^{-1}(1-p_{i,\{i,j\}}) \\ &= \log(\frac{f_i(p_{i,\{i,k\}})}{f_k(p_{k,\{i,k\}})} \frac{f_j(p_{j,\{i,j\}})}{f_i(p_{i,\{i,j\}})}) \\ &= \log(\frac{f_i(p_{i,\{i,j,k\}})}{f_k(p_{k,\{i,j,k\}})} \frac{f_j(p_{j,\{i,j,k\}})}{f_i(p_{i,\{i,j,k\}})}) && \text{(Generalized Ordinal IIA)}\end{aligned}$$

$$
\begin{aligned}
&= \log(\frac{f_i(p_{i,\{i,j,k\}})}{f_k(p_{k,\{i,j,k\}})}) \\
&= \log(\frac{f_j(p_{j,S})}{f_k(p_{k,S})}) && \text{(Generalized Ordinal IIA)} \\
&= F_j^{-1}(1-p_{j,S}) - F_k^{-1}(1-p_{k,S}).
\end{aligned}
$$

We next verify

$$\nu_i + F_i^{-1}(1-p_{i,S}) - \lambda_S + \lambda_{i,S} = 0,\ \lambda_{i,S}p_{i,S} = 0, \ \text{ and } \ \lambda_{i,S} \geq 0,\ \forall i \in S.$$

Since $p_{i,S} > 0$ for all $i \in S$, we have $\lambda_{i,S} = 0$ in the KKT conditions. Then, there exists λ_S such that $\nu_k + F_k^{-1}(1-p_{k,S}) - \lambda_S = 0$ and $\nu_j + F_j^{-1}(1-p_{j,S}) - \lambda_S = 0$ since we have proved that $\nu_k + F_k^{-1}(1-p_{k,S}) = \nu_j + F_j^{-1}(1-p_{j,S})$.

Therefore, the equalities in the above two cases imply the KKT optimality conditions at S are satisfied. Since this holds for any assortment $S \in \mathcal{S}$, this proves that $\boldsymbol{p}_{\mathcal{S}}$ is represented by MDM.

■

EC.2. Proofs of the Results in Section 4

EC.2.1. Proof of Proposition 2

Proof. Due to Theorem 1, we have $\mathcal{P}_{\text{MDM}}(\mathcal{S}') = \text{Proj}_{\boldsymbol{x}}(\Pi_{\mathcal{S}'})$, where the lifted set $\Pi_{\mathcal{S}'}$ equals

$$\Big\{(\boldsymbol{x},\boldsymbol{\lambda}) \in \mathbb{R}^{\mathcal{I}_{\mathcal{S}'}} \times \mathbb{R}^{\mathcal{S}'} : x_{i,S} \geq 0,\ \sum_i x_{i,S} = 1, \forall S \in \mathcal{S}',$$
$$\lambda_S > \lambda_T \text{ if } x_{i,S} < x_{i,T},\ \lambda_S = \lambda_T \text{ if } x_{i,S} = x_{i,T} \neq 0,\ \forall (i,S),(i,T) \in \mathcal{I}_{\mathcal{S}'} \Big\},$$

following the definition in (7). Since $\mathcal{U}_A := \{\boldsymbol{x}_A : (\boldsymbol{x}_{\mathcal{S}}, \boldsymbol{x}_A, \boldsymbol{\lambda}) \in \Pi_{\mathcal{S}'},\ \boldsymbol{x}_{\mathcal{S}} = \boldsymbol{p}_{\mathcal{S}}\}$, the non-numbered constraints in the formulation in Proposition 2 are obtained by replacing $\boldsymbol{x}_{\mathcal{S}} = \boldsymbol{p}_{\mathcal{S}}$ in the above description of the lifted set $\Pi_{\mathcal{S}'}$. For deducing the remaining constraints (10a) - (10b), we proceed as follows: Consider any $(i,S) \in \mathcal{I}_{\mathcal{S}}$. From the description of $\Pi_{\mathcal{S}'}$, observe that an assignment for $x_{i,A}, \lambda_A, \lambda_S$ in any $(\boldsymbol{x},\boldsymbol{\lambda}) \in \Pi'_{\mathcal{S}}$ satisfying $\boldsymbol{x}_{\mathcal{S}} = \boldsymbol{p}_{\mathcal{S}}$ necessarily satisfies one of the following four cases:

In *Case 1,* we have $\lambda_A > \lambda_S$ and $x_{i,A} < p_{i,S}$: If λ_A, λ_S is such that $\lambda_A > \lambda_S$, this informs the restriction $\{x_{i,A} : x_{i,A} < p_{i,S}\}$ on the values $x_{i,A}$ can take. The closure of this restricted collection $\{x_{i,A} : x_{i,A} < p_{i,S}\}$ equals $\{x_{i,A} : x_{i,A} \leq p_{i,S}\}$.

In *Case 2,* we have $\lambda_A < \lambda_S$ and $x_{i,A} > p_{i,S}$: If λ_A, λ_S is such that $\lambda_A < \lambda_S$, the closure of the corresponding restriction $\{x_{i,A} : x_{i,A} > p_{i,S}\}$ equals $\{x_{i,A} : x_{i,A} \geq p_{i,S}\}$.

In *Case 3,* we have $\lambda_A = \lambda_S$ and $x_{i,A} = p_{i,S} \neq 0$: When λ_A, λ_S is such that $\lambda_A = \lambda_S$ and $p_{i,S} \neq 0$, the corresponding restriction on the values of $x_{i,A}$ is given by the closed set $\{x_{i,A} : x_{i,A} = p_{i,S}\}$.

Finally, in *Case 4,* we have λ_A, λ_S unconstrained and $x_{i,A} = p_{i,S} = 0$: Like in Case 3, the restriction on the values of $x_{i,A}$ corresponding to this case equals $\{x_{i,A} : x_{i,A} = 0\}$. The relationship between

$x_{i,A}, p_{i,S}, \lambda_A, \lambda_S$ in this case is any one of the following sub-cases: Case (4a) $\lambda_A > \lambda_S$ and $0 = x_{i,A} \leq p_{i,S} = 0$, or Case (4b) $\lambda_A < \lambda_S$ and $0 = x_{i,A} \geq p_{i,S} = 0$, or Case (4c) $\lambda_A = \lambda_S$ and $0 = x_{i,A} = p_{i,S} = 0$.

Combining the observations in the cases (1) & (4a), (2) & (4b), and (3) & (4c), we obtain that the closure of $\mathcal{U}_A$ equals the collection of probability vectors $\boldsymbol{x}_A = (x_{i,A} : i \in A)$ for which there exists a function $\lambda : \mathcal{S}' \to \mathbb{R}$ such that

$$\begin{aligned} x_{i,A} &\leq p_{i,S} \quad \text{if } \lambda_A > \lambda_S, \quad \forall i \in A,\ (i,S) \in \mathcal{I}_{\mathcal{S}}, \\ x_{i,A} &\geq p_{i,S} \quad \text{if } \lambda_A < \lambda_S, \quad \forall i \in A,\ (i,S) \in \mathcal{I}_{\mathcal{S}}, \text{ and} \\ x_{i,A} &= p_{i,S} \quad \text{if } \lambda_A = \lambda_S, \quad \forall i \in A,\ (i,S) \in \mathcal{I}_{\mathcal{S}}, \end{aligned}$$

in addition to satisfying $\lambda_S > \lambda_T$ if $p_{i,S} < p_{i,T}$ and $\lambda_S = \lambda_T$ if $p_{i,S} = p_{i,T} \neq 0$, for all $(i,S),(i,T) \in \mathcal{I}_{\mathcal{S}}$. The constraints in the formulation in Proposition 2 exactly specify these conditions describing the closure of $\mathcal{U}_A$. Observe that the objective in $\inf\{\sum_{i\in A} r_i x_{i,A} : \boldsymbol{x}_A \in \mathcal{U}_A\}$ is continuous as a function of $\boldsymbol{x}_A$. Therefore, $\inf\{\sum_{i\in A} r_i x_{i,A} : \boldsymbol{x}_A \in \mathcal{U}_A\} = \min\{\sum_{i\in A} r_i x_{i,A} : \boldsymbol{x}_A \in \text{closure}(\mathcal{U}_A)\}$. ■

EC.2.2. Proof of Proposition 3

Proof. Recall the notation $\mathcal{S}' = \mathcal{S} \cup \{A\}$. Observe that the variables $(\lambda_S : S \in \mathcal{S}')$ influence the value of the formulation in Proposition 2 only via the sign of $\lambda_S - \lambda_T$, for any pair of variables λ_S, λ_T from the collection $(\lambda_S : S \in \mathcal{S}')$. Therefore the optimal value of this optimization formulation is not affected by the presence of the following additional constraints: $0 \leq \lambda_S \leq 1$ for all $S \in \mathcal{S}'$, and

$$\lambda_S - \lambda_T \geq \epsilon \quad \text{if} \quad \lambda_S > \lambda_T, \qquad \forall\, (i,S),(i,T) \in \mathcal{I}_{\mathcal{S}'},$$

for some suitably small value of $\epsilon > 0$. Indeed, this is because the signs of the differences $\{\lambda_S - \lambda_T : S,T \in \mathcal{S}'\}$ are not affected by these additional constraints. Taking ϵ to be smaller than $1/(2|\mathcal{S}|)$, for example, ensures that there is a feasible assignment for $(\lambda_S : S \in \mathcal{S}')$ within the interval $[0,1]$ even if all these variables take distinct values.

Let F denote the feasible values for the variables $(\lambda_S : S \in \mathcal{S}'), (x_{i,A} : i \in A)$ satisfying the constraints introduced in the above paragraph besides those in the formulation in Proposition 2. Equipped with this feasible region F, we have the following deductions from (11a) - (11c) for $(\lambda_S : S \in \mathcal{S}'), (x_{i,A} : i \in A)$ in F: For every $i \in A$ and any $S \in \mathcal{S}$ containing i,

(i) we have $\lambda_A < \lambda_S$ if and only if $\delta_{A,S} = 1$ and $\delta_{S,A} = 0$, due to the constraints (11a) and (11b); in this case, we have from (11c) that $p_{i,S} \leq x_{i,A} \leq 1$;

(ii) likewise, we have $\lambda_A > \lambda_S$ if and only if $\delta_{A,S} = 0$ and $\delta_{S,A} = 1$, due to the constraints (11a) and (11b); in this case, we have from (11c) that $0 \leq x_{i,A} \leq p_{i,S}$.

(iii) finally, $\lambda_A = \lambda_S$ if and only if $\delta_{A,S} = 0$ and $\delta_{S,A} = 0$; here we have from (11d) that $x_{i,A} = p_{i,S}$.

Thus the binary variables $\{\delta_{A,S}, \delta_{S,A} : S \in \mathcal{S}\}$ suitably model the constraints collection (10a) - (10b) and provide an equivalent reformulation in terms of the constraints (11a) - (11d). Therefore the optimal value of the formulations in Propositions 2 and 3 are identical. ■

EC.2.3. Proof of Corollary 1

Proof. When $\mathcal{S}'$ is either nested or laminar, from the proof in Theorem 3, we know that $\mathcal{P}_{\text{MDM}}(\mathcal{S}') = \mathcal{P}_{\text{REG}}(\mathcal{S}')'$. So, we can solve the worst-case expected revenue in (9) with the representable conditions of the regular model which are $x_{i,A} \le p_{i,S}$ if $S \subset A$ and $x_{i,A} \ge p_{i,S}$ if $A \subset S$ for all $i \in A$ and $(i,S) \in \mathcal{I}_{\mathcal{S}}$. This is a linear program with $\mathcal{O}(n)$ continuous variables and $\mathcal{O}(n|\mathcal{S}|)$ constraints. ■

EC.2.4. Proof of Proposition 4

Proof. Suppose that there exists a product i^* being included in every assortment of the collection $\mathcal{S}$ and we have $p_{i^*,S_1} \ge p_{i^*,S_2} \ge \cdots \ge p_{i^*,S_m}$. Recall that $\text{closure}(\mathcal{P}_{\text{MDM}}(\mathcal{S})) = \text{Proj}_{\boldsymbol{x}}(\Pi'_{\mathcal{S}})$ where $\Pi'_{\mathcal{S}}$ is defined as

$$\Pi'_{\mathcal{S}} = \Big\{(\boldsymbol{x},\boldsymbol{\lambda}) \in \mathbb{R}^{\mathcal{I}_{\mathcal{S}}} \times \mathbb{R}^{\mathcal{S}} : x_{i,S} \ge 0, \forall (i,S) \in \mathcal{I}_{\mathcal{S}}, \sum_i x_{i,S} = 1, \forall S \in \mathcal{S}, \\ \lambda_S \ge \lambda_T \text{ if } x_{i,S} \le x_{i,T}, \forall (i,S),(i,T) \in \mathcal{I}_{\mathcal{S}}\Big\}. \tag{EC.7}$$

Thus, we have $\lambda_{S_m} \ge \cdots \lambda_{S_k} \ge \lambda_{S_{k-1}} \ge \cdots \ge \lambda_{S_1}$ for any $\boldsymbol{\lambda}$ such that $(\boldsymbol{x},\boldsymbol{\lambda}) \in \Pi'_{\mathcal{S}}$. For ease of notation, let $\lambda_{S_0} = -\infty$ and $\lambda_{S_{m+1}} = +\infty$. Then for any given A, the corresponding λ_A must satisfy $\lambda_{S_{k+1}} \ge \lambda_A \ge \lambda_{S_k}$ for some $k \in \{0,1,\ldots,m\}$.

From the viewpoint of $\lambda_A \ge \lambda_{S_k} \ge \ldots \ge \lambda_{S_0}$, we deduce the following constraints on $x_{i,A}$: For any $i \in A \cap S_k$, we have the respective MDM feasibility constraints $x_{i,A} \le p_{i,S_k} \le \ldots \le p_{i,S_j}$ for all $j \le k$ such that $i \in S_j$. These constraints can be equivalently summarized by $x_{i,A} \le p_{i,S_k}$, and this comprises the first set of constraints for evaluating $\boldsymbol{R_k}$ in (13).

From the viewpoint of $\lambda_{S_m+1} \ge \cdots \lambda_{S_{k+1}} \ge \lambda_A$, we deduce the following constraints on $x_{i,A}$: For any $i \in A, j \ge k+1, S_j \in \mathcal{S}, i \in S_j \cap A$, we have the respective MDM feasibility constraint $x_{i,A} \ge p_{i,S}$, which comprise the second set of constraints for evaluating $\boldsymbol{R_k}$ in (13). ■

EC.2.5. Proof of Corollary 2

Proof. Suppose $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ is nested as in $S_1 \subset S_2 \subset \cdots \subset S_m$. We have, from the regularity of MDM in Lemma 2, that $p_{i,S_1} \ge p_{i,S_2} \ge \cdots p_{i,S_m}$, for all $i \in S_1$. Recall that $\text{closure}(\mathcal{P}_{\text{MDM}}(\mathcal{S})) = \text{Proj}_{\boldsymbol{x}}(\Pi'_{\mathcal{S}})$ where $\Pi'_{\mathcal{S}}$ is defined as

$$\Pi'_{\mathcal{S}} = \Big\{(\boldsymbol{x},\boldsymbol{\lambda}) \in \mathbb{R}^{\mathcal{I}_{\mathcal{S}}} \times \mathbb{R}^{\mathcal{S}} : x_{i,S} \ge 0, \forall (i,S) \in \mathcal{I}_{\mathcal{S}}, \sum_{i \in S} x_{i,S} = 1, \forall S \in \mathcal{S}, \\ \lambda_S \ge \lambda_T \text{ if } x_{i,S} \le x_{i,T}, \forall (i,S),(i,T) \in \mathcal{I}_{\mathcal{S}}\Big\}. \tag{EC.8}$$

Thus, for the given nested $\mathcal{S}$, we have $\lambda_{S_m} \ge \cdots \lambda_{S_k} \ge \lambda_{S_{k-1}} \ge \cdots \ge \lambda_{S_1}$ for any $\boldsymbol{\lambda}$ such that $(\boldsymbol{x},\boldsymbol{\lambda}) \in \Pi'_{\mathcal{S}}$. For ease of notation, let $\lambda_{S_0} = -\infty$ and $\lambda_{S_{m+1}} = +\infty$. Then for any given A, the corresponding λ_A must satisfy $\lambda_{S_{k+1}} \ge \lambda_A \ge \lambda_{S_k}$ for some $k \in \{0,1,\ldots,m\}$.

From the viewpoint of $\lambda_A \ge \lambda_{S_k} \ge \ldots \ge \lambda_{S_0}$, we deduce the following constraints on $x_{i,A}$: For any $i \in A \cap S_k$, we have the respective MDM feasibility constraints $x_{i,A} \le p_{i,S_k} \le \ldots \le p_{i,S_j}$ for all $j \le k$

such that $i \in S_j$. These constraints can be equivalently summarized by $x_{i,A} \leq p_{i,S_k}$, and this comprises the first set of constraints for evaluating $\boldsymbol{R_k}$ in (14).

From the viewpoint of $\lambda_{S_m+1} \geq \cdots \lambda_{S_{k+1}} \geq \lambda_A$, we deduce the following constraints on $x_{i,A}$: For any $i \in A, (i,S) \in \mathcal{I}_{\mathcal{S}}, S_{k+1} \subseteq S$, we have the respective MDM feasibility constraint $x_{i,A} \geq p_{i,S}$, which comprise the second set of constraints for evaluating $\boldsymbol{R_k}$ in (14).

■

EC.3. Proofs of the Results in Section 5

EC.3.1. Proof of Proposition 5

Proof. Due to Theorem 1, we have $\mathcal{P}_{\text{MDM}}(\mathcal{S}) = \text{Proj}_{\boldsymbol{x}}(\Pi_{\mathcal{S}})$, following the definition in (7). One can argue the closure of $\mathcal{P}_{\text{MDM}}(\mathcal{S})$ by the similar arguments of the proof of Proposition 2. Consider any $(i,S) \in \mathcal{I}_{\mathcal{S}}$. From the description of $\Pi_{\mathcal{S}}$, observe that an assignment for $x_{i,S}, x_{i,T}, \lambda_S, \lambda_T$ in any $(\boldsymbol{x}, \boldsymbol{\lambda}) \in \Pi_{\mathcal{S}}$ necessarily satisfies one of the following four cases:

In *Case 1,* we have $\lambda_S < \lambda_T$ and $x_{i,S} > x_{i,T}$: If λ_S, λ_T is such that $\lambda_S < \lambda_T$, the closure of the corresponding restriction $\{x_{i,S}, x_{i,T} : x_{i,S} > x_{i,T}\}$ equals $\{x_{i,S}, x_{i,T} : x_{i,S} \geq x_{i,T}\}$.

In *Case 2,* we have $\lambda_S > \lambda_T$ and $x_{i,S} < x_{i,T}$: If λ_S, λ_T is such that $\lambda_S > \lambda_T$, the closure of the corresponding restriction $\{x_{i,S}, x_{i,T} : x_{i,S} < x_{i,T}\}$ equals $\{x_{i,S}, x_{i,T} : x_{i,S} \leq x_{i,T}\}$.

In *Case 3,* we have $\lambda_S = \lambda_T$ and $x_{i,S} = x_{i,T} > 0$: When λ_S, λ_T is such that $\lambda_S = \lambda_T$ and $x_{i,S} = x_{i,T} > 0$, the corresponding restriction on the values of $x_{i,S}, x_{i,T}$ is given by the closed set $\{x_{i,S}, x_{i,T} : x_{i,S} = x_{i,T} > 0\}$.

Finally, in *Case 4,* we have λ_S, λ_T unconstrained and $x_{i,S} = x_{i,T} = 0$: Like in Case 3, the restriction on the values of $x_{i,S}, x_{i,T}$ corresponding to this case equals $x_{i,S} = x_{i,T} = 0$. The relationship between $x_{i,S}, x_{i,T}, \lambda_S, \lambda_T$ in this case is any one of the following sub-cases: Case (4a) $\lambda_S > \lambda_T$ and $0 = x_{i,S} \leq x_{i,T} = 0$, or Case (4b) $\lambda_S < \lambda_T$ and $0 = x_{i,S} \geq x_{i,T} = 0$, or Case (4c) $\lambda_S = \lambda_T$ and $0 = x_{i,S} = x_{i,T} = 0$.

Combining the observations in the cases (1) & (4a), (2) & (4b), and (3) & (4c), we obtain that the closure of $\mathcal{P}_{\text{MDM}}(\mathcal{S})$ equals the collection of probability vectors $\boldsymbol{x}$ for which there exists a function $\lambda : \mathcal{S} \to \mathbb{R}$ such that

$$\begin{aligned}
x_{i,S} &\leq x_{i,T} \quad \text{if} \quad \lambda_S > \lambda_T, \quad \forall (i,S),\ (i,T) \in \mathcal{I}_{\mathcal{S}}, \\
x_{i,S} &\geq x_{i,T} \quad \text{if} \quad \lambda_S < \lambda_T, \quad \forall (i,S),\ (i,T) \in \mathcal{I}_{\mathcal{S}}, \\
x_{i,S} &= x_{i,T} \quad \text{if} \quad \lambda_S = \lambda_T, \quad \forall (i,S),\ (i,T) \in \mathcal{I}_{\mathcal{S}}.
\end{aligned}$$

The constraints in the formulation in Proposition 5 exactly specify these conditions describing the closure of $\mathcal{P}_{\text{MDM}}(\mathcal{S})$. Observe that the objective in $\inf\{\text{loss}(\boldsymbol{p}_{\mathcal{S}}, \boldsymbol{x}_{\mathcal{S}}) : \boldsymbol{x}_{\mathcal{S}} \in \mathcal{P}_{\text{MDM}}(\mathcal{S})\}$ is continuous as a function of $\boldsymbol{x}$. Therefore, $\inf\{\text{loss}(\boldsymbol{p}_{\mathcal{S}}, \boldsymbol{x}_{\mathcal{S}}) : \boldsymbol{x}_{\mathcal{S}} \in \mathcal{P}_{\text{MDM}}(\mathcal{S})\} = \min\{\text{loss}(\boldsymbol{p}_{\mathcal{S}}, \boldsymbol{x}_{\mathcal{S}}) : \boldsymbol{x}_{\mathcal{S}} \in \text{closure}(\mathcal{P}_{\text{MDM}}(\mathcal{S}))\}$.

■

EC.3.2. Proof of Theorem 4

Before we formally prove Theorem 4, we first show the following preparatory material. Problem (16) can be solved by the following optimization problem:

$$\inf_{\boldsymbol{\lambda}} \quad \text{loss}(\boldsymbol{x}^*_{\mathcal{S}}(\boldsymbol{\lambda}), \boldsymbol{p}_{\mathcal{S}}) \tag{EC.9}$$

$$\text{s.t.} \quad \boldsymbol{x}^*_{\mathcal{S}}(\boldsymbol{\lambda}) \in \arg\inf_{\boldsymbol{x}_{\mathcal{S}}(\boldsymbol{\lambda}):(\boldsymbol{x}_{\mathcal{S}}(\boldsymbol{\lambda}),\boldsymbol{\lambda})\in\Pi_{\mathcal{S}}} \quad \text{loss}(\boldsymbol{x}_{\mathcal{S}}(\boldsymbol{\lambda}), \boldsymbol{p}_{\mathcal{S}}|\boldsymbol{\lambda}), \tag{EC.10}$$

where $\boldsymbol{x}_{\mathcal{S}}(\boldsymbol{\lambda})$ can be interpreted as a collection of MDM-representable choice probabilities given $\mathcal{S}$ and $\boldsymbol{\lambda}$. Next, we focus on the sub-problem (EC.10). Let $[m] = \{1, 2, \cdots, m\}$ and $[k] = \{1, 2, \cdots, k\}$ for some positive integer m and k.

Assumption EC.1. *Consider $\mathcal{S} = \{S_1, S_2, \cdots, S_m\}$ and $\boldsymbol{p}_{\mathcal{S}}$ as a $n \times m$ matrix with n rows and m columns, satisfying the following properties:*

- *$n = k + m$ with k as a positive integer;*
- *For each $l \in [k]$, there are exactly two elements in row l of $\boldsymbol{p}_{\mathcal{S}}$ and $\kappa = |p_{l,S_i} - p_{l,S_j}| < \frac{1}{2m}$ is a positive constant for $i, j \in [m]$ with $l \in S_i \cap S_j$;*
- *For each $i \in [m]$, there is exactly one element p_{k+i,S_i} in row $k+i$ and $p_{k+i,S_i} = 1 - \sum_{j=1}^{k} p_{j,S_i}$ and $p_{k+i,S_i} > \frac{km\kappa}{2}$.*

Lemma EC.3. *Given $\boldsymbol{p}_{\mathcal{S}}$ that satisfies Assumption EC.1, the sub-problem* (EC.10) *with 1-norm objective function has a closed-form optimal objective value $\sum_{l=1}^{k} \sum\limits_{i,j\in[m],\lambda_{S_i}\le\lambda_{S_j}:l\in S_i\cap S_j} 2|p_{l,S_i} - p_{l,S_j}|\mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \ge 0\}$.*

Intuitively, $\boldsymbol{p}_{\mathcal{S}}$ has exactly one product-assortment pair for product l with $l \in [k]$ and exactly one element for product $k+i$ with $i \in [m]$ which corresponds to the ith assortment in the collection. For product l with $l \in [k]$ the indicator takes the value 1 if and only if the choice probabilities of the product-assortment pair violate the MDM-representable conditions, the minimum loss to make this pair to be MDM-representable under 1-norm loss is $|p_{l,S_i} - p_{l,S_j}|$, which causes the violation of the normalization constraint of the assortments S_i and S_j. To satisfy the normalization conditions of assortment S_i and S_j, the least loss is also $|p_{l,S_i} - p_{l,S_j}|$. Next, we formally prove Lemma EC.3 with three steps: (1) reformulate the sub-problem (EC.10) to a linear program; (2) construct a primal feasible solution $\boldsymbol{x}(\boldsymbol{\lambda})$ such that the desired optimal objective value is achieved, which can be served as an upper bound to (EC.10); (3) derive the dual for the problem and construct a dual feasible solution such that the desired optimal objective value is achieved.

Proof. Step (1): Formulate the sub-problem (EC.10) as a linear program with $\boldsymbol{p}_{\mathcal{S}}$ satisfying Assumption EC.1.

Let f^*_{sub} denote the optimal value of (EC.10). Given $\boldsymbol{p}_{\mathcal{S}}$ that satisfies Assumption EC.1 , we reformulate the sub-problem (EC.10) as the following problem:

$$\begin{aligned} f^*_{\text{sub}}(\boldsymbol{\lambda}) = \min_{\boldsymbol{x}_{\mathcal{S}}} \quad & \sum_{l=1}^{k} \sum_{i,j\in[m],\lambda_{S_i}\le\lambda_{S_j}:l\in S_i\cap S_j} (|x_{l,S_i}-p_{l,S_i}|+|x_{l,S_j}-p_{l,S_j}|)+\sum_{i=1}^{m}|x_{k+i,S_i}-p_{k+i,S_i}| \\ \text{s.t.} \quad & x_{l,S_i}-x_{l,S_j}\ge 0 \text{ if } \lambda_{S_i}\le\lambda_{S_j} \quad \forall l\in[k], i,j,\in[m], l\in S_i\cap S_j, \\ & \sum_{i\in S} x_{i,S}=1, \quad \forall S\in\mathcal{S}, \\ & x_{i,S}\ge 0, \quad \forall (i,S)\in\mathcal{I}_{\mathcal{S}}. \end{aligned} \tag{EC.11}$$

We reformulate Problem (EC.11) as the following linear program (EC.12) by introducing a new variable $\boldsymbol{z}_{\mathcal{S}}$.

$$\begin{aligned} \min_{\boldsymbol{x}_{\mathcal{S}},\boldsymbol{z}_{\mathcal{S}}} \quad & \sum_{l=1}^{k} \sum_{i,j\in[m],\lambda_{S_i}\le\lambda_{S_j}:l\in S_i\cap S_j} (z_{l,S_i}+z_{l,S_j})+\sum_{i=1}^{m} z_{k+i,S_i} \\ \text{s.t.} \quad & z_{l,S_i}-x_{l,S_i}\ge -p_{l,S_i}, \quad z_{l,S_j}-x_{l,S_j}\ge -p_{l,S_j}, \quad \forall l\in[k], i,j\in[m], \lambda_{S_i}\le\lambda_{S_j}: l\in S_i\cap S_j, \\ & z_{l,S_i}+x_{l,S_i}\ge p_{l,S_i}, \quad z_{l,S_j}+x_{l,S_j}\ge p_{l,S_j}, \quad \forall l\in[k], i,j\in[m], \lambda_{S_i}\le\lambda_{S_j}: l\in S_i\cap S_j, \\ & x_{l,S_i}-x_{l,S_j}\ge 0 \text{ if } \lambda_{S_i}\le\lambda_{S_j} \quad \forall l\in[k], i,j\in[m], l\in S_i\cap S_j, \\ & z_{k+i,S_i}-x_{k+i,S_i}\ge -p_{k+i,S_i}, \quad \forall i\in[m], \\ & z_{k+i,S_i}+x_{k+i,S_i}\ge p_{k+i,S_i}, \quad \forall i\in[m], \\ & \sum_{l\in S_i} x_{l,S_i}=1, \quad \forall i\in[m], \\ & x_{i,S},\, z_{i,S}\ge 0, \quad \forall (i,S)\in\mathcal{I}_{\mathcal{S}}. \end{aligned} \tag{EC.12}$$

Step (2): Construct a primal feasible solution to achieve $f^*_{\text{sub}}(\boldsymbol{\lambda})$.

Given any $\boldsymbol{\lambda}$, construct a solution $(\boldsymbol{x}_{\mathcal{S}},\boldsymbol{z}_{\mathcal{S}})$ as follows.

- For $l\in[k]$,$i,j\in[m]$ with $\lambda_{S_i}\le\lambda_{S_j}$ such that $l\in S_i\cap S_j$:
  If $\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\ge 0\}=0$, let

  $$x_{l,S_i}=p_{l,S_i},\ x_{l,S_j}=p_{l,S_j},\ z_{l,S_i}=z_{l,S_j}=0;$$

  If $\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\ge 0\}=1$, let

  $$x_{l,S_i}=x_{l,S_j}=\frac{p_{l,S_i}+p_{l,S_j}}{2},\ z_{l,S_i}=z_{l,S_j}=\frac{|p_{l,S_i}-p_{l,S_j}|}{2}.$$

- For $i \in [m]$, let

$$x_{k+i,S_i} = p_{k+i,S_i} + \sum_{l=1}^{k} \sum_{i,j\in[m]:l\in S_i\cap S_j} \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \geq 0\} \frac{\text{sgn}(p_{l,S_i} - p_{l,S_j})|p_{l,S_i} - p_{l,S_j}|}{2},$$

$$z_{k+i,S_i} = \sum_{l=1,i,j,\in[m],l\in S_i\cap S_j}^{k} \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \geq 0\} \frac{|p_{l,S_i} - p_{l,S_j}|}{2},$$

where $\text{sgn}(x) := \begin{cases} 1 & \text{if} \quad x > 0, \\ -1 & \text{if} \quad x \leq 0. \end{cases}$

We verify the feasibility of the constructed primal solution for (EC.12) as follows: Firstly, we have all the nonnegative constraints for $(x_{i,S},\, z_{i,S} : \forall (i,S) \in \mathcal{I}_{\mathcal{S}})$ are satisfied. For $l \in [k]$,$i, j \in [m]$ with $\lambda_{S_i} \leq \lambda_{S_j}$ such that $l \in S_i \cap S_j$:

If $\mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \geq 0\} = 0$, we have :

$$\begin{aligned}
&z_{l,S_i} - x_{l,S_i} = 0 - p_{l,S_i} = -p_{l,S_i},\ z_{l,S_j} - x_{l,S_j} = 0 - p_{l,S_j} = -p_{l,S_j}, \\
&z_{l,S_i} + x_{l,S_i} = 0 + p_{l,S_i} = p_{l,S_i},\ z_{l,S_j} + x_{l,S_j} = 0 + p_{l,S_j} = p_{l,S_j}, \\
&x_{l,S_i} - x_{l,S_j} = p_{l,S_i} - p_{l,S_j} \geq 0 \text{ if } \lambda_{S_i} \leq \lambda_{S_j}.
\end{aligned}$$

If $\mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \geq 0\} = 1$, we have :

$$\begin{aligned}
z_{l,S_i} - x_{l,S_i} &= \frac{|p_{l,S_i} - p_{l,S_j}|}{2} - \frac{p_{l,S_i} + p_{l,S_j}}{2} \geq \frac{p_{l,S_j} - p_{l,S_i}}{2} - \frac{p_{l,S_i} + p_{l,S_j}}{2} = -p_{l,S_i}, \\
z_{l,S_j} - x_{l,S_j} &= \frac{|p_{l,S_i} - p_{l,S_j}|}{2} - \frac{p_{l,S_i} + p_{l,S_j}}{2} \geq \frac{p_{l,S_i} - p_{l,S_j}}{2} - \frac{p_{l,S_i} + p_{l,S_j}}{2} = -p_{l,S_j}, \\
z_{l,S_i} + x_{l,S_i} &= \frac{|p_{l,S_i} - p_{l,S_j}|}{2} + \frac{p_{l,S_i} + p_{l,S_j}}{2} \geq \frac{p_{l,S_i} - p_{l,S_j}}{2} + \frac{p_{l,S_i} + p_{l,S_j}}{2} = p_{l,S_i}, \\
z_{l,S_j} + x_{l,S_j} &= \frac{|p_{l,S_i} - p_{l,S_j}|}{2} - \frac{p_{l,S_i} + p_{l,S_j}}{2} \geq \frac{p_{l,S_j} - p_{l,S_i}}{2} + \frac{p_{l,S_i} + p_{l,S_j}}{2} = p_{l,S_j}, \\
x_{l,S_i} - x_{l,S_j} &= \frac{p_{l,S_i} + p_{l,S_j}}{2} - \frac{p_{l,S_i} + p_{l,S_j}}{2} = 0 \text{ if } \lambda_{S_i} \leq \lambda_{S_j}.
\end{aligned}$$

For all $i \in [m]$, we also have

$$\begin{aligned}
&z_{k+i,S_i} - x_{k+i,S_i} \\
= &\sum_{l=1,i,j,\in[m],l\in S_i\cap S_j}^{k} \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \geq 0\} \frac{|p_{l,S_i} - p_{l,S_j}|}{2} \\
&- (p_{k+i,S_i} + \sum_{l=1}^{k} \sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j}:l\in S_i\cap S_j} \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \geq 0\} \frac{\text{sgn}(p_{l,S_i} - p_{l,S_j})|p_{l,S_i} - p_{l,S_j}|}{2}) \\
&\geq -p_{k+i,S_i}, \\
&z_{k+i,S_i} + x_{k+i,S_i} \\
= &\sum_{l=1,i,j,\in[m],l\in S_i\cap S_j}^{k} \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \geq 0\} \frac{|p_{l,S_i} - p_{l,S_j}|}{2}
\end{aligned}$$

$$
\begin{aligned}
&+p_{k+i,S_i}+\sum_{l=1}^{k}\sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j}:l\in S_i\cap S_j}\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}\frac{\operatorname{sgn}(p_{l,S_i}-p_{l,S_j})|p_{l,S_i}-p_{l,S_j}|}{2}\\
&\geq p_{k+i,S_i}.
\end{aligned}
$$

Both inequalities hold due to

$$
\frac{|p_{l,S_i}-p_{l,S_j}|}{2}\geq\frac{\operatorname{sgn}(p_{l,S_i}-p_{l,S_j})|p_{l,S_i}-p_{l,S_j}|}{2}.
$$

For the normalization constraints, for all $i\in[m]$, we have:

$$
\begin{aligned}
&\sum_{l\in S_i}x_{l,S_i}=\sum_{l=1}^{k}x_{l,S_i}+x_{k+i,S_i}\\
=&\sum_{l=1}^{k}\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})<0\}p_{l,S_i}+\sum_{l=1}^{k}\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}\frac{p_{l,S_i}+p_{l,S_j}}{2}\\
&+p_{k+i,S_i}+\sum_{l=1}^{k}\sum_{i,j\in[m]:l\in S_i\cap S_j}\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}\frac{\operatorname{sgn}(p_{l,S_i}-p_{l,S_j})|p_{l,S_i}-p_{l,S_j}|}{2}\\
=&\sum_{l\in S_i}p_{l,S_i}=1.
\end{aligned}
$$

Then, the objective value of Problem (EC.12) under the constructed solution $(\boldsymbol{x}_{\mathcal{S}},\boldsymbol{z}_{\mathcal{S}})$ is

$$
\sum_{l=1}^{k}\sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j}:l\in S_i\cap S_j}2|p_{l,S_i}-p_{l,S_j}|\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}.
$$

This implies that

$$
f^*_{\text{sub}}(\boldsymbol{\lambda})\leq\sum_{l=1}^{k}\sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j}:l\in S_i\cap S_j}2|p_{l,S_i}-p_{l,S_j}|\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}.
$$

Step (3): Construct a dual feasible solution to achieve $f^*_{\text{sub}}(\boldsymbol{\lambda})$.

We derive the dual of (EC.12) as follows. For $l\in[k]$,$i,j\in[m]$ with $\lambda_{S_i}\leq\lambda_{S_j}$ such that $l\in S_i\cap S_j$, we introduce the following variables: $\alpha_{l,i},\beta_{l,i},\alpha_{l,j},\beta_{l,j},u_{l,i,j}\geq 0$. For $i\in[m]$, we introduce the following variables: $\alpha_{k+i,S_i},\beta_{k+i,i}\geq 0$, and η_i. The dual problem of (EC.12) is given as:

$$
\begin{aligned}
\max_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{\eta}}\;&\sum_{l=1}^{k}\sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j},l\in S_i\cap S_j}[p_{l,S_i}(\beta_{l,i}-\alpha_{l,i})+p_{l,j}(\beta_{l,j}-\alpha_{l,j})]+\sum_{i=1}^{m}p_{k+i,S_i}(\beta_{k+i,i}-\alpha_{k+i,i})-\sum_{i=1}^{m}\eta_i\\
\text{s.t.}\quad&\alpha_{l,i}+\beta_{l,i}\leq 1,\quad\alpha_{l,j}+\beta_{l,j}\leq 1,\quad\forall l\in[k],i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j},\ l\in S_i\cap S_j,\\
&-\alpha_{l,i}+\beta_{l,i}+u_{l,i,j}+\eta_i\leq 0,\ -\alpha_{l,j}+\beta_{l,j}-u_{l,i,j}+\eta_j\leq 0,\ \forall l\in[k],i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j},l\in S_i\cap S_j,\\
&\alpha_{k+i,i}+\beta_{k+i,i}\leq 1,\quad\forall i\in[m],\\
&-\alpha_{k+i,i}+\beta_{k+i,i}+\eta_i\leq 0,\quad\forall i\in[m],\\
&\alpha_{l,i},\beta_{l,i},\alpha_{l,j},\beta_{l,j},u_{l,i,j}\geq 0,\quad\forall l\in[k],i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j},l\in S_i\cap S_j,\\
&\alpha_{k+i,i},\beta_{k+i,i}\geq 0,\quad\forall i\in[m].
\end{aligned}
\tag{EC.13}
$$

Construct a dual solution for (EC.13) as follows:

For $l \in [k], i, j \in [m]$, with $\lambda_i \le \lambda_j$ and $l \in S_i \cap S_j$,

- If $\mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \ge 0\} = 0$, let $\alpha_{l,i} = \beta_{l,i} = \alpha_{l,j} = \beta_{l,j} = u_{l,i,j} = 0$.
- If $\mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \ge 0\} = 1$, let $\alpha_{l,i} = 1,\ \beta_{l,i} = 0,\ \alpha_{l,j} = 0,\ \beta_{l,j} = 1,\ u_{l,i,j} = 1$.

For $i \in [m]$, let $\alpha_{k+i,i} = \beta_{k+i,i} = \frac{1}{2}$ and $\eta_i = -\sum_{l=1}^k \sum_{j \in [m]: l \in S_i \cap S_j} \frac{|p_{l,S_i} - p_{l,S_j}| \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \ge 0\}}{2}$.

Next, we verify the feasibility of the constructed dual solution for (EC.13). Firstly, we have all the nonnegative constraints for $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{u}$ are satisfied. For $l \in [k]$,$i, j \in [m]$ with $\lambda_{S_i} \le \lambda_{S_j}$ such that $l \in S_i \cap S_j$:

If $\mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \ge 0\} = 0$, we have :

$$\begin{aligned}
&\alpha_{l,i} + \beta_{l,i} = 0 + 0 < 1, \quad \alpha_{l,j} + \beta_{l,j} = 0 + 0 < 1, \\
&-\alpha_{l,i} + \beta_{l,i} + u_{l,i,j} + \eta_i = -0 + 0 + 0 - \sum_{l=1}^k \sum_{j \in [m]: l \in S_i \cap S_j} \frac{|p_{l,S_i} - p_{l,S_j}| \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \ge 0\}}{2} < 0, \\
&-\alpha_{l,j} + \beta_{l,j} - u_{l,i,j} + \eta_j = -0 + 0 - 0 - \sum_{l=1}^k \sum_{j \in [m]: l \in S_i \cap S_j} \frac{|p_{l,S_i} - p_{l,S_j}| \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \ge 0\}}{2} < 0.
\end{aligned}$$

If $\mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \ge 0\} = 1$, we have :

$$\begin{aligned}
&\alpha_{l,i} + \beta_{l,i} = 1 + 0 = 1, \quad \alpha_{l,j} + \beta_{l,j} = 0 + 1 = 1, \\
&-\alpha_{l,i} + \beta_{l,i} + u_{l,i,j} + \eta_i = -1 + 0 + 1 - \sum_{l=1}^k \sum_{j \in [m]: l \in S_i \cap S_j} \frac{|p_{l,S_i} - p_{l,S_j}| \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \ge 0\}}{2} < 0, \\
&-\alpha_{l,j} + \beta_{l,j} - u_{l,i,j} + \eta_j = -0 + 1 - 1 - \sum_{l=1}^k \sum_{j \in [m]: l \in S_i \cap S_j} \frac{|p_{l,S_i} - p_{l,S_j}| \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \ge 0\}}{2} < 0.
\end{aligned}$$

For all $i \in [m]$, we have

$$\begin{aligned}
&\alpha_{k+i,i} + \beta_{k+i,i} = \frac{1}{2} + \frac{1}{2} = 1, \\
&-\alpha_{k+i,i} + \beta_{k+i,i} + \eta_i = -\frac{1}{2} + \frac{1}{2} - \sum_{l=1}^k \sum_{j \in [m]: l \in S_i \cap S_j} \frac{|p_{l,S_i} - p_{l,S_j}| \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \ge 0\}}{2} < 0.
\end{aligned}$$

Then, the objective value of Problem (EC.13) under the constructed solution $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\eta})$ is

$$\begin{aligned}
&\sum_{l=1}^k \sum_{i,j \in [m], \lambda_i \le \lambda_j, l \in S_i \cap S_j} [p_{l,S_i}(\beta_{l,i} - \alpha_{l,i}) + p_{l,j}(\beta_{l,j} - \alpha_{l,j})] + \sum_{i=1}^m p_{k+i,S_i}(\beta_{k+i,i} - \alpha_{k+i,i}) - \sum_{i=1}^m \eta_i \\
=&\sum_{l=1}^k \sum_{i,j \in [m], \lambda_{S_i} \le \lambda_{S_j}, l \in S_i \cap S_j} \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \ge 0\}(p_{l,j} - p_{l,S_i}) - \sum_{i=1}^m \eta_i \qquad \text{(EC.14a)} \\
=&\sum_{l=1}^k \sum_{i,j \in [m], \lambda_{S_i} \le \lambda_{S_j}, l \in S_i \cap S_j} \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \ge 0\}|p_{l,i} - p_{l,S_j}| - \sum_{i=1}^m \eta_i \qquad \text{(EC.14b)}
\end{aligned}$$

$$
\begin{aligned}
=&\sum_{l=1}^{k}\sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j},l\in S_i\cap S_j}\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}|p_{l,i}-p_{l,S_j}| \\
&+\sum_{i=1}^{m}\sum_{l=1}^{k}\sum_{j\in[m]:l\in S_i\cap S_j}\frac{|p_{l,S_i}-p_{l,S_j}|\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}}{2} \qquad \text{(EC.14c)} \\
=&\sum_{l=1}^{k}\sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j},l\in S_i\cap S_j}\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}|p_{l,j}-p_{l,S_i}| \\
&+\sum_{l=1}^{k}\sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j},l\in S_i\cap S_j}\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}|p_{l,i}-p_{l,S_j}| \qquad \text{(EC.14d)} \\
=&\sum_{l=1}^{k}\sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j}:l\in S_i\cap S_j}2|p_{l,S_i}-p_{l,S_j}|\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}.
\end{aligned}
$$

Plugging in the value of $\alpha_{l,i}=\beta_{l,i}=\alpha_{l,j}=\beta_{l,j}$, for $l\in[k]$,$i,j\in[m]$ with $\lambda_{S_i}\leq\lambda_{S_j}$ and $l\in S_i\cap S_j$, we have (EC.14a) . (EC.14b) is equivalent to (EC.14a) because when $\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}=1$ and $\lambda_i\leq\lambda_j$, we have $|p_{l,i}-p_{l,S_j}|=p_{l,j}-p_{l,S_i}>0$. Plugging in the value of η_i for all $i\in[m]$ to (EC.14b), we have (EC.14c). (EC.14d) is equivalent to (EC.14c) because

$$
\begin{aligned}
&\sum_{i=1}^{m}\sum_{l=1}^{k}\sum_{j\in[m]:l\in S_i\cap S_j}\frac{|p_{l,S_i}-p_{l,S_j}|\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}}{2} \\
=&\sum_{l=1}^{k}\sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j},l\in S_i\cap S_j}\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}|p_{l,i}-p_{l,S_j}|.
\end{aligned}
$$

By weak duality, we have $f^*_{\text{sub}}(\boldsymbol{\lambda})\geq\sum_{l=1}^{k}\sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j}:l\in S_i\cap S_j}2|p_{l,S_i}-p_{l,S_j}|\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}$.

Summing up, we have

$$
f^*_{\text{sub}}(\boldsymbol{\lambda})=\sum_{l=1}^{k}\sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j}:l\in S_i\cap S_j}2|p_{l,S_i}-p_{l,S_j}|\mathbb{I}\{(p_{l,S_i}-p_{l,S_j})(\lambda_{S_i}-\lambda_{S_j})\geq 0\}.
$$

■

Before we show the hardness of Problem (16), we provide the following three definitions (see Dwork et. al. (2001)) to describe the Kemeny optimal aggregation problem.

Definition EC.1 (Full lists and partial lists). Let $\mathcal{M}=\{1,\ldots,m\}$ be a finite set of alternatives, called universe. A ranking over $\mathcal{M}$ is an ordered list. If the ranking τ contains all the elements in $\mathcal{M}$, then it is called a full list (ranking). If the ranking τ contains a subset of elements from the universe $\mathcal{M}$, then it is called a partial list (ranking).

Definition EC.2 (Kendall-tau distance (K-distance)). The K-distance, denoted as $K(\sigma,\tau)$, is the number of pairs $i,j\in\mathcal{M}$ such that $\sigma(i)<\sigma(j)$ but $\tau(i)>\tau(j)$ where $\sigma(i)$ stands for the position of i in σ and similar explanations are applied for $\sigma(j), \tau(i)$ and $\tau(j)$. Note that the pair (i,j) has contribution to the K-distance only if both i, j appear in both lists σ, τ.

Definition EC.3 (SK, Kemeny optimal). For a collection of partial lists $\tau_1, \ldots, \tau_k$ and a full list π, we denote

$$SK(\pi, \tau_1, \ldots, \tau_k) = \sum_{i=1}^{k} K(\pi, \tau_i).$$

We say a permutation σ is a Kemeny optimal aggregation of $\tau_1, \ldots, \tau_k$ if it minimizes $SK(\pi, \tau_1, \ldots, \tau_k)$ over all permutations π.

Lemma EC.4 **(see Dwork et. al. (2001))**. *Finding the Kemeny optimal solution for partial lists of length 2 is exactly the same problem as finding a minimum feedback arc set, and hence is NP-hard.*

Now, we are ready to prove the hardness of Problem (16).

Proof. To show Problem (16) is NP-hard, it suffices to show some instances of this problem is NP-hard. We show that Kemeny optimal aggregation of length 2 can be reduced to Problem (16).

The decision version of Kemeny optimal aggregation with a collection of partial lists of all length 2 is stated as follows:

INSTANCE: A finite set $\mathcal{M}$ with $|\mathcal{M}| = m$, a collection of partial lists $\tau_1, \ldots, \tau_k$ with $|\tau_i| = 2$ for $i = 1, \ldots, k$, an upper bound on the loss L.

QUESTION: Is there a full list π, such that $\sum_{i=1}^{k} K(\pi, \tau_i) \leq L$?

The decision version of the limit problem of MDM in Problem (16) is stated as follows:

INSTANCE: A finite set $\mathcal{N}$ with $|\mathcal{N}| = n$, a collection of assortments $\mathcal{S}$ with $|\mathcal{S}| = m$ and $S \subseteq \mathcal{N}$ for all $S \in \mathcal{S}$, the observed choice probabilities $\boldsymbol{p}_{\mathcal{S}} = (p_{i,S} : i \in S, S \in \mathcal{S})$ with $\sum_{i \in S} p_{i,S} = 1$ for all $S \in \mathcal{S}$, an upper bound on the loss $L^{'}$.

QUESTION: Is there a solution $(\boldsymbol{x}_{\mathcal{S}}, \boldsymbol{\lambda})$ to Problem (16) such that loss$(\boldsymbol{x}_{\mathcal{S}}, \boldsymbol{p}_{\mathcal{S}}) \leq L^{'}$?

We then will reduce the Kemeny optimal aggregation problem to Problem (16). Given any instance of Kemeny optimal aggregation problem with partial lists all of length 2, we can construct an instance of Problem (16) as follows.

(a) Let the collection of assortments $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ with $|\mathcal{S}| = m$ and the set of alternatives (products), $\mathcal{N} = \{1, \ldots, k, k+1, \ldots, k+m\}$ with $|\mathcal{N}| = n = k + m$. Given the observed choice data $\boldsymbol{p}_{\mathcal{S}}$, consider $\boldsymbol{p}_{\mathcal{S}}$ as a $n \times m$ matrix with n rows and m columns.

(b) The values of the entries in $\boldsymbol{p}_{\mathcal{S}}$ are set in the following manner. For each $l \in \{1, ..., k\}$, suppose $\tau_l = \{i, j\}$ with $\tau_l(i) < \tau_l(j)$, then we set $p_{l,S_i} = \frac{1}{3 \times k}$ and $p_{l,S_j} = \frac{2}{3 \times k}$. It's easy to see that for each S_i with $1 \leq i \leq m$, $0 < \sum_{j=1}^{k} p_{j,S_i} < 1$.

(c) For each S_i with $1 \leq i \leq m$, let $p_{k+i,S_i} = 1 - \sum_{j=1}^{k} p_{j,S_i}$.

(d) Set other entries of $\boldsymbol{p}_{\mathcal{S}}$ as zero.

(e) Set the loss function in Problem (16) to be 1-norm loss.

We give Example EC.1 and Example EC.2 to illustrate the above instance construction.

EXAMPLE EC.1. Given an instance of Kemeny optimal aggregation with $\mathcal{M} = \{1,2,3\}$ and $\tau_1 = (1 \succ 2), \tau_2 = (1 \succ 3), \tau_3 = (2 \succ 3)$, we construct an instance for Problem (16) with $\boldsymbol{p}_{\mathcal{S}}$ as shown in Table EC.4.

Table EC.4 An example of a representable instance construction

alternative	$S_1 = \{1,2,4\}$	$S_2 = \{1,3,5\}$	$S_3 = \{2,3,6\}$
1	1/9	2/9	-
2	1/9	-	2/9
3	-	1/9	2/9
4	7/9	-	-
5	-	6/9	-
6	-	-	5/9

EXAMPLE EC.2. Given an instance of Kemeny optimal aggregation with $\mathcal{M} = \{1,2,3\}$ and $\tau_1 = (1 \succ 2), \tau_2 = (3 \succ 1), \tau_3 = (2 \succ 3)$, we construct an instance for Problem (16) with $\boldsymbol{p}_{\mathcal{S}}$ as shown in Table EC.5.

Table EC.5 An example of an infeasible instance construction

alternative	$S_1 = \{1,2,4\}$	$S_2 = \{1,3,5\}$	$S_3 = \{2,3,6\}$
1	1/9	2/9	-
2	2/9	-	1/9
3	-	1/9	2/9
4	6/9	-	-
5	-	6/9	-
6	-	-	6/9

In Example EC.1, both the Kemeny optimal aggregation and the limit of MDM instances are feasible to their problems respectively. The optimal solution to the Kemeny optimal aggregation is $\pi = (1 \succ 2 \succ 3)$ and one of the optimal solutions to the limit of MDM is $\boldsymbol{x}_{\mathcal{S}} = \boldsymbol{p}_{\mathcal{S}}$ and $\lambda_{S_1} = 3$, $\lambda_{S_2} = 2$ and $\lambda_{S_3} = 1$. Both instances obtain 0 loss.

In Example EC.2, both the Kemeny optimal aggregation and the limit of MDM instance are infeasible to their problems respectively. It's trivial to see that the optimal solution to the Kemeny optimal aggregation is one of $\{(1 \succ 2 \succ 3), (2 \succ 3 \succ 1), (3 \succ 1 \succ 2)\}$. Each of such solutions obtains $SK = 1$. Given Lemma EC.3, one may make a guess for one of the optimal solutions to the limit of MDM in Example EC.2 to be $x^*_{2,S_1} = \frac{1}{6}$, $x^*_{2,S_3} = \frac{1}{6}$, $x^*_{4,S_1} = \frac{13}{18}$, $x^*_{6,S_3} = \frac{11}{18}$ and the optimal loss to be $\frac{2}{9}$. We will show that this guess is true.

Recall that Problem (16) is equivalent to Problem (EC.9). We next show that for any fixed $\boldsymbol{\lambda}$ in Problem (EC.9), then the sub-problem (EC.10) with optimal $\boldsymbol{x}^*_{\mathcal{S}}(\boldsymbol{\lambda})$ and $f^*_{\text{sub}}(\boldsymbol{\lambda})$ under $\boldsymbol{\lambda}$, there exists π such that $\lambda^*_{S_{\pi^{-1}(1)}} \geq \cdots \geq \lambda^*_{S_{\pi^{-1}(m)}}$ for the Kemeny optimal aggregation and $SK(\pi) = \frac{3\times k}{2} f^*_{\text{sub}}(\boldsymbol{\lambda})$.

By Lemma EC.3, we have

$$\begin{aligned}
&f^*_{\text{sub}}(\boldsymbol{\lambda}) = \text{loss}(\boldsymbol{x}^*_{\mathcal{S}}(\boldsymbol{\lambda}), \boldsymbol{p}_{\mathcal{S}}) \\
=&\sum_{l=1}^{k} \sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j}:l\in S_i\cap S_j} (|x^*_{l,S_i} - p_{l,S_i}| + |x^*_{l,S_j} - p_{l,S_j}|) + \sum_{i=1}^{m} |x^*_{k+i,S_i} - p_{k+i,S_i}| && \text{(EC.15a)} \\
=&\sum_{l=1}^{k} \sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j}:l\in S_i\cap S_j} 2|p_{l,S_i} - p_{l,S_j}| \mathbb{I}\{(p_{l,S_i} - p_{l,S_j})(\lambda_{S_i} - \lambda_{S_j}) \geq 0\} && \text{(EC.15b)} \\
=&\frac{2}{3\times k} \sum_{l=1}^{k} \sum_{i,j\in[m],\lambda_{S_i}\leq\lambda_{S_j}:l\in S_i\cap S_j} \mathbb{I}\{(\pi(i) - \pi(j))(\tau_l(i) - \tau_l(j)) < 0\} && \text{(EC.15c)} \\
=&\frac{2}{3\times k} \sum_{l=1}^{k} K(\pi, \tau_l) \\
=&\frac{2}{3\times k} SK(\pi).
\end{aligned}$$

Equation (EC.15a) is due to the construction of $\boldsymbol{p}_{\mathcal{S}}$. Equation (EC.15b) holds because of $|p_{l,S_i} - p_{l,S_j}| = \frac{1}{3\times k}$ and the closed form objective value in Lemma EC.3. The argument for Equation (EC.15c) is as follows: For $l\in[k]$, $i,j\in[m]: l\in S_i\cap S_j$, by instance construction, we have

$$p_{l,S_i} < p_{l,S_j} \text{ if } \tau_l(i) < \tau_l(j).$$

From the relation between π and $\boldsymbol{\lambda}$, we have $\lambda_{\pi^{-1}(1)} > \ldots > \lambda_{\pi^{-1}(m)}$. Then, we have

$$\pi(i) < \pi(j) \text{ if } \lambda_{S_i} \geq \lambda_{S_j}.$$

The above two inequities imply that

$$\mathbb{I}\{(\lambda_{S_i} \geq \lambda_{S_j})(p_{l,S_i} - p_{l,S_j}) > 0\} = \mathbb{I}\{(\pi(i) - \pi(j))(\tau_l(i) - \tau_l(j)) < 0\}.$$

Setting $L = \frac{3k}{2} L'$. The decision problem of the limit of MDM asks is there $(\boldsymbol{x}_{\mathcal{S}}, \boldsymbol{\lambda})$ such that $f^*_{\text{limit}} \leq L'$ is equivalent to the decision problem of Kemeny optimal aggregation is there a full ranking π such that $f^*_{\text{kemeny}} \leq L$. ■

EC.3.3. Proof of Proposition 6

Proof. Observe that the variables $(\lambda_S : S\in\mathcal{S})$ influence the value of the formulation in Proposition 6 only via the sign of $\lambda_S - \lambda_T$, for any pair of variables λ_S, λ_T from the collection $(\lambda_S : S\in\mathcal{S})$. Therefore the optimal value of this optimization formulation is not affected by the presence of the following additional constraints: $0 \leq \lambda_S \leq 1$ for all $S\in\mathcal{S}$. Indeed, this is because the signs of the differences $\{\lambda_S - \lambda_T : S, T \in \mathcal{S}\}$ are not affected by these additional constraints. Taking ϵ to be smaller than $1/(2|\mathcal{S}|)$, for example, ensures that there is a feasible assignment for $(\lambda_S : S \in \mathcal{S})$ within the interval $[0,1]$ even if all these variables take distinct values.

Let F denote the feasible values for the variables $(\lambda_S : S \in \mathcal{S}), (x_{i,S} : (i,S) \in \mathcal{I}_\mathcal{S})$ satisfying the constraints introduced in the above paragraph besides those in the formulation in Proposition 5. Equipped with this feasible region F, we have the following deductions for $(\lambda_S : S \in \mathcal{S}), (x_{i,S} : (i,S) \in \mathcal{I}_\mathcal{S})$ in F: For any $S, T \in \mathcal{S}$ containing i,

(i) we have $\lambda_S < \lambda_T$ if and only if $\delta_{S,T} = 1$ and $\delta_{T,S} = 0$, due to the first set of constraints of (17); in this case, we have from the second and fourth set of constraints of (17) that $0 \leq x_{i,T} \leq x_{i,S} \leq 1$;

(ii) likewise, we have $\lambda_S > \lambda_T$ if and only if $\delta_{S,T} = 0$ and $\delta_{T,S} = 1$, due to the first set of constraints; in this case, we have from the second and fourth set of constraints of (17) that $0 \leq x_{i,S} \leq x_{i,T} \leq 1$.

(iii) finally, $\lambda_S = \lambda_T$ if and only if $\delta_{S,T} = 0$ and $\delta_{T,S} = 0$; here we have from the third set of constriants of (17) that $x_{i,S} = x_{i,T}$.

Thus the binary variables $\{\delta_{S,T} : S, T \in \mathcal{S}\}$ suitably model the first set of constraints of (16) and provide an equivalent reformulation in terms of the constraints. Therefore the optimal value of the formulations in Propositions 5 and 6 are identical. ■

EC.4. Illustrative Examples in the Paper

EC.4.1. An example to illustrate the representation power of MDM, APU, and MNL

Example EC.3. Note that $\boldsymbol{p}_\mathcal{S}$ can be represented by APU if and only if there exists $\boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{S}|}$ and $\boldsymbol{\nu} \in \mathbb{R}^n$ such that $\lambda_S + \nu_i > \lambda_T + \nu_j$ if and only if $p_{i,S} > p_{i,T}$, for all $(i,S), (j,T) \in \mathcal{I}_\mathcal{S}$ (Fudenberg et. al. 2015). The $\boldsymbol{p}_\mathcal{S}$ in Table EC.6 can be represented by MDM by simply setting $\lambda_A = 10$ and $\lambda_B = 5$. However, it cannot be represented by the APU model because $p_{1,A} < p_{2,A}$ implies $\nu_1 < \nu_2$, while $p_{1,B} > p_{2,B}$ implies $\nu_1 > \nu_2$, which leads to a contradiction. The $\boldsymbol{p}_\mathcal{S}$ in Table EC.7 can be represented by APU, since we can set $\lambda_A = 8$, $\lambda_B = 10$, $\nu_1 = 1$, $\nu_2 = 2$, $\nu_3 = 5$, and $\nu_4 = 0$ such that $p_{3,A} > p_{2,B} > p_{1,B} > p_{2,A} = p_{4,B} > p_{1,A}$ and $\lambda_A + \nu_3 > \lambda_B + \nu_2 > \lambda_B + \nu_1 > \lambda_A + \nu_2 = \lambda_B + \nu_4 > \lambda_A + \nu_1$. Both instances cannot be represented by MNL since IIA property is violated, i.e., $\frac{p_{1,A}}{p_{2,A}} \neq \frac{p_{1,B}}{p_{2,B}}$.

Table EC.6 $\boldsymbol{p}_\mathcal{S}$ **can be represented by MDM but not APU and MNL**

Alternative	A = {1,2,3}	B= {1,2,4}
1	0.20	0.45
2	0.25	0.30
3	0.55	-
4	-	0.25

Table EC.7 $\boldsymbol{p}_\mathcal{S}$ **can be represented by MDM and APU but not MNL**

Alternative	A = {1,2,3}	B= {1,2,4}
1	0.20	0.30
2	0.25	0.45
3	0.55	-
4	-	0.25

EC.4.2. An example to show the non-convexity of MDM feasible region

Example EC.4. $\boldsymbol{x}_{\mathcal{S}}$ is MDM-representable because $x_{1,A} < x_{1,B}$,$x_{2,A} < x_{2,C}$ and $x_{3,B} < x_{3,C}$ implies $\lambda_A > \lambda_B$, $\lambda_A > \lambda_C$ and $\lambda_B > \lambda_C$. The values $\lambda_A = 12, \lambda_B = 10, \lambda_C = 8$ satify this. $\boldsymbol{y}_{\mathcal{S}}$ is MDM-representable because $y_{1,A} > y_{1,B}$, $y_{2,A} > y_{2,C}$ and $y_{3,B} > y_{3,C}$ implies $\lambda_A < \lambda_B$, $\lambda_A < \lambda_C$ and $\lambda_B < \lambda_C$. The values $\lambda_A = 8, \lambda_B = 10, \lambda_C = 12$ satify this. $\boldsymbol{w} = 0.4\boldsymbol{x}_{\mathcal{S}} + 0.6\boldsymbol{y}_{\mathcal{S}}$ is a convex combination of $\boldsymbol{x}_{\mathcal{S}}$ and $\boldsymbol{y}_{\mathcal{S}}$ but it can not be represented by MDM because $w_{1,A} > w_{1,B}$, $w_{1,A} < w_{2,C}$ and $w_{3,B} > w_{3,C}$ which implies $\lambda_A < \lambda_B$, $\lambda_A > \lambda_C$ and $\lambda_B < \lambda_C$, i.e., $\lambda_B < \lambda_C < \lambda_A < \lambda_B$. This means $\boldsymbol{w}$ can not be represented by MDM.

$\boldsymbol{x}_{\mathcal{S}}$

Alternative	A={1,2}	B={1,3}	C={2,3}
1	0.3	0.9	
2	0.7		0.8
3		0.1	0.2

$\boldsymbol{y}_{\mathcal{S}}$

Alternative	A={1,2}	B={1,3}	C={2,3}
1	0.75	0.1	
2	0.25		0.2
3		0.9	0.8

$\boldsymbol{w} = 0.4\boldsymbol{x}_{\mathcal{S}} + 0.6\boldsymbol{y}_{\mathcal{S}}$

Alternative	A={1,2}	B={1,3}	C={2,3}
1	0.57	0.42	
2	0.43		0.44
3		0.58	0.56

EC.4.3. An example to show using (18) to get MDM-representable probabilities

Example EC.5. Consider the observed choice probabilities $\boldsymbol{p}_{\mathcal{S}}$ is given in Table (a) below with $n = 3$ and $\mathcal{S} = \{A = \{1,2\}, B = \{1,3\}, C = \{2,3\}\}$. $\boldsymbol{p}_{\mathcal{S}}$ is not representable by MDM since $p_{1,A} > p_{1,B}$ implies $\lambda_A < \lambda_B$, $p_{2,A} < p_{2,C}$ implies $\lambda_A > \lambda_C$, and $p_{3,B} > p_{3,C}$ implies $\lambda_B < \lambda_C$, that is $\lambda_A < \lambda_B < \lambda_C < \lambda_A$.

Suppose the assortments A, B, C are shown once in the historical data. We choose the limit loss function to be 1-norm $\sum_{(i,S)\in\mathcal{I}_{\mathcal{S}}} |x_{i,S} - p_{i,S}|$, and set $\epsilon = 0.001$ in (17). We then obtain MDM limit probabilities $\boldsymbol{x}^*_{\mathcal{S}}$ in Table (b) below and $\lambda_A = 0$, $\lambda_B = 0.001$, and $\lambda_B = 0.002$, that is $\lambda_A < \lambda_B < \lambda_C$. $\boldsymbol{x}^*_{\mathcal{S}}$ satisfies all the constraints in (16), meaning that $\boldsymbol{x}^*_{\mathcal{S}} \in \text{closure}(\mathcal{P}_{\text{MDM}}(\mathcal{S}))$. However, $\boldsymbol{x}^*_{\mathcal{S}} \notin \mathcal{P}_{\text{MDM}}(\mathcal{S})$ since $x^*_{2,A} = x^*_{2,C}$ when $\lambda_A < \lambda_C$.

To obtain feasible MDM probabilities, we apply (18) by setting $\delta = 10^{-5}$. We then obtain MDM representable choice probabilities in Table (c) below.

(a) $\boldsymbol{p}_{\mathcal{S}}$ that is not representable by MDM

Alternative	A={1,2}	B={1,3}	C={2,3}
1	0.57	0.42	-
2	0.43	-	0.44
3	-	0.58	0.56

(b) $\boldsymbol{x}^*_{\mathcal{S}}$ obtained by (17)

Alternative	A={1,2}	B={1,3}	C={2,3}
1	0.56	0.42	-
2	0.44	-	0.44
3	-	0.58	0.56

(c) feasible MDM probabilities obtained by (18)

Alternative	A={1,2}	B={1,3}	C={2,3}
1	0.564823	0.42	-
2	0.435177	-	0.435176
3	-	0.58	0.564824

EC.5. An Algorithm for Evaluating the Limit of MDM when $|\mathcal{S}|$ is Small

In Algorithm 1, for a fixed $\boldsymbol{\lambda}$, we just need to solve a convex optimization problem with $\mathcal{O}(n|\mathcal{S}|)$ continuous variables and $\mathcal{O}(n|\mathcal{S}|^2)$ linear constraints to compute the limit loss. There are $m!$ possible $\boldsymbol{\lambda}$. Thus, Algorithm 1 is polynomial in the alternative size n.

```
Algorithm 1: An algorithm solves the limit of MDM polynomial in $n$
Input: Observed choice probabilities $\boldsymbol{p}_{\mathcal{S}}$, collection $\mathcal{S}$, product universe $\mathcal{N}$.
Output: MDM choice probabilities $\boldsymbol{x}^*_{\mathcal{S}}$, optimal loss $f^*$, optimal ranking of assortments $\tau^*$.
$\boldsymbol{T} \leftarrow \{$all rankings of $(S : S \in \mathcal{S})\}$;
$f^* \leftarrow +\infty$ keeps tracking of the optimal value of Problem (16);
$\boldsymbol{x}^*_{\mathcal{S}} \leftarrow \boldsymbol{0}$ keeps tracking of the optimal solution;
for $\tau \in \boldsymbol{T}$ do
    Solve   $\min_{\boldsymbol{x}_{\mathcal{S}}}$ loss$(\boldsymbol{x}_{\mathcal{S}}, \boldsymbol{p}_{\mathcal{S}})$
            s.t. $x_{i,S} - x_{i,T} \geq 0$,   if $\tau(T) < \tau(S)$   $\forall (i,S),(i,T) \in \mathcal{I}_{\mathcal{S}}$,
                 $\sum_{i \in S} x_{i,S} = 1$,   $\forall S \in \mathcal{S}$,
                 $x_{i,S} \geq 0$,   $\forall (i,S) \in \mathcal{I}_{\mathcal{S}}$.        (Limit-LP)
    $f \leftarrow$ the output optimal objective value of (Limit-LP);
    $\boldsymbol{x}_{\mathcal{S}} \leftarrow$ the output optimal solution of (Limit-LP);
    if $f < f^*$ then
        $\boldsymbol{x}^*_{\mathcal{S}} \leftarrow \boldsymbol{x}_{\mathcal{S}}$;
        $f^* \leftarrow f$ ;
        $\tau^* \leftarrow \tau$;
    end
end
```

EC.6. Additional Useful Details on the Experiments in the Paper

In this section, we give details on the implementation of the experiments. We used a MacBook Pro Laptop with a 2 GHz 4 core Intel Core i5 processor for all experiments.

EC.6.1. Representation test experiment implementation details

EC.6.1.1. Checking the representability of MDM For a collection of observed choice probabilities $\boldsymbol{p}_{\mathcal{S}}$, by Theorem 1, we check the representability of MDM with the following linear program:

$$\begin{aligned} \max_{\epsilon} \quad & \epsilon \\ \text{s.t.} \quad & \lambda_S - \lambda_T - \epsilon \geq 0, \text{ if } p_{i,S} < p_{i,T} \quad \forall (i,S),(i,T) \in \mathcal{I}_{\mathcal{S}}, \\ & \lambda_S - \lambda_T = 0, \text{ if } p_{i,S} = p_{i,T} > 0 \quad \forall (i,S),(i,T) \in \mathcal{I}_{\mathcal{S}}. \end{aligned} \tag{EC.16}$$

If the optimal value of (EC.16) is strictly positive, then $\boldsymbol{p}_{\mathcal{S}}$ can be represented by MDM. Otherwise, $\boldsymbol{p}_{\mathcal{S}}$ cannot be represented by MDM.

EC.6.1.2. Checking the representability of RUM For a collection of observed choice probabilities $\boldsymbol{p}_{\mathcal{S}}$, we check the representability of RUM with the following linear program:

$$\begin{aligned}
\max_{\boldsymbol{\lambda}} \quad & 0 \\
\text{s.t.} \quad & \sum_{\sigma\in\Sigma_n} \lambda(\sigma)\mathbb{I}[\sigma,i,S] - p_{i,S} = 0, \quad \forall (i,S)\in\mathcal{I}_{\mathcal{S}}, \\
& \sum_{\sigma\in\Sigma_n} \lambda(\sigma) = 1, \quad \lambda(\sigma)\geq 0, \quad \forall\sigma\in\Sigma_n.
\end{aligned} \tag{EC.17}$$

If (EC.17) is feasible, $\boldsymbol{p}_{\mathcal{S}}$ can be represented by RUM. Otherwise, $\boldsymbol{p}_{\mathcal{S}}$ cannot be represented by RUM.

EC.6.1.3. Checking the representability of MNL For a collection of observed choice probabilities $\boldsymbol{p}_{\mathcal{S}}$, we check the representability of MNL with the following linear program:

$$\begin{aligned}
\max_{\boldsymbol{\nu}} \quad & \sum_{i\in\mathcal{N}} \nu_i \\
\text{s.t.} \quad & p_{i,S}\sum_{j\in\nu_j} \nu_j - \nu_i = 0, \quad \forall (i,S)\in\mathcal{I}_{\mathcal{S}}, \\
& \nu_i \geq 0, \quad \forall i\in\mathcal{N}.
\end{aligned} \tag{EC.18}$$

If the optimal value of (EC.18) is strictly positive, $\boldsymbol{p}_{\mathcal{S}}$ can be represented by MNL. Otherwise, $\boldsymbol{p}_{\mathcal{S}}$ cannot be represented by MNL.

EC.6.2. Limit experiment implementation details

EC.6.2.1. Limit of MDM Given an instance $\boldsymbol{p}_{\mathcal{S}}$ with collection $\mathcal{S}$, we solve the limit of MDM with (17). When the loss function is chosen to be the average absolute deviation loss, we set $\sum_{S\in\mathcal{S}} \text{loss}(\boldsymbol{p}_S, \boldsymbol{x}_S) = \sum_{S\in\mathcal{S}} n_S \sum_{i\in S} p_{i,S}|x_{i,S}-p_{i,S}|/\sum_{S\in\mathcal{S}} n_S$ in (17).

When the loss function is chosen to be the average KL loss function, the limit of MDM can not be easily solved by existing solvers including Gurobi, Mosek, and CVXPY. We take the following 2-stage approach to solve the limit of MDM with the average KL loss function: (1) We first solve (17) with the average absolute deviation loss and get $\boldsymbol{\lambda}^*$. (2) We then solve (17) with the average KL loss function and $\boldsymbol{\lambda}^*$ as follows:

$$\begin{aligned}
\min_{\boldsymbol{x}} \; & -\sum_{S\in\mathcal{S}} n_S \sum_{i\in S} p_{i,S}\log(x_{i,S}/p_{i,S})/\sum_{S\in\mathcal{S}} n_S \\
\text{s.t.} \quad & x_{i,S}\geq x_{i,T} \text{ if } \lambda_S^*\leq\lambda_T^*, \qquad \forall (i,S),(i,T)\in\mathcal{I}_{\mathcal{S}}, \\
& \sum_{i\in S} x_{i,S} = 1, \ \forall S\in\mathcal{S}, \quad x_{i,S}\geq 0, \ \forall (i,S)\in\mathcal{I}_{\mathcal{S}}.
\end{aligned}$$

EC.6.2.2. Limit of RUM Given an instance $\boldsymbol{p}_{\mathcal{S}}$ with collection $\mathcal{S}$, we solve the limit of RUM with the following convex program:

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{\lambda}} \quad & \sum_{S\in\mathcal{S}} \text{loss}(\boldsymbol{p}_S, \boldsymbol{x}_S) \\
\text{s.t.} \quad & x_{i,S} - \sum_{\sigma\in\Sigma_n} \lambda(\sigma)\mathbb{I}[\sigma,i,S] = 0, \quad \forall (i,S)\in\mathcal{I}_{\mathcal{S}}, \\
& \sum_{\sigma\in\Sigma_n} \lambda(\sigma) = 1, \quad \lambda(\sigma)\geq 0, \quad \forall\sigma\in\Sigma_n.
\end{aligned} \tag{EC.19}$$

We set the loss function of (EC.19) based on the chosen loss function. In our experiments, the limit of RUM is computed when the loss function is chosen as the average absolute deviation loss $\sum_{S\in\mathcal{S}} \text{loss}(\boldsymbol{p}_S, \boldsymbol{x}_S) = \sum_{S\in\mathcal{S}} n_S \sum_{i\in S} p_{i,S}|x_{i,S} - p_{i,S}|/\sum_{S\in\mathcal{S}} n_S$, which can be reformulated as a linear program, and we can solve it in the Gurobi solver. The limit losses of RUM of all instances reported in the paper are solved to optimality.

EC.6.2.3. Limit of MNL For the parametric model MNL, given an instance $\boldsymbol{p}_{\mathcal{S}}$ with collection $\mathcal{S}$, we first estimate the parameters of MNL by solving the maximum loglikelihood with the given data and then compute the best-fitting choice probabilities of MNL to $\boldsymbol{p}_{\mathcal{S}}$ with the estimated parameters. The parameters of MNL are estimated as follows:

$$\begin{aligned}\boldsymbol{\nu}^* = \text{argmax}_{\boldsymbol{\nu}}\, ll(\boldsymbol{\nu}|\boldsymbol{p}_{\mathcal{S}}) &= \sum_{S\in\mathcal{S}} n_S \sum_{i\in S} p_{i,S} \log(\frac{\exp v_i}{\sum_{j\in S}\exp v_j}) \\ &= \sum_{S\in\mathcal{S}} n_S (\sum_{i\in S} p_{i,S}\, v_i - \log \sum_{j\in S} \exp v_j).\end{aligned} \tag{EC.20}$$

When the limit loss function is chosen to be the average absolute deviation loss, we compute the loss between the choice probability collection and the probability collection with the estimated MLE of MNL as $\sum_{S\in\mathcal{S}} n_S \sum_{i\in S} p_{i,S}|p_{i,S} - \frac{\exp v_i^*}{\sum_{j\in S}\exp v_j^*}|/\sum_{S\in\mathcal{S}} n_S$. When the limit loss function is chosen to be the average KL loss, we first solve (EC.20) and record the optimal values as ll^*, and then compute the average KL loss of MNL as $(-ll^* + \sum_{S\in\mathcal{S}} n_S \sum_{i\in S} p_{i,S} \log p_{i,S})/\sum_{S\in\mathcal{S}} n_S$.

EC.6.2.4. Limit of MCCM By the definition of MCCM, the choice probabilities satisfy the constraints in (EC.21) below, so the limit of MCCM can be solved by the following continuous optimization problem:

$$\begin{aligned}\min_{\boldsymbol{\lambda},\boldsymbol{\rho},\boldsymbol{x}_{\mathcal{S}},\boldsymbol{y}_{\mathcal{S}}} \quad & \text{loss}(\boldsymbol{p}_S, \boldsymbol{x}_S) \\ \text{s.t.} \quad & x_{i,S} = \lambda_i + \sum_{j\in\mathcal{N}_0\setminus S} y_{j,S}\rho_{j,i}, \quad \forall i \in S\cup\{0\}, S\in\mathcal{S}, \\ & y_{i,S} = \lambda_i + \sum_{j\in\mathcal{N}_0\setminus S} y_{j,S}\rho_{j,i}, \quad \forall i \in \mathcal{N}\setminus S, S\in\mathcal{S}, \\ & x_{i,S} = 0, \quad \forall i\in\mathcal{N}\setminus S, S\in\mathcal{S}, \\ & y_{i,S} = 0 \quad \forall i\in S\cup\{0\}, S\in\mathcal{S}, \\ & \sum_{i=0}^{n} \lambda_i = 1, \quad \lambda_i \geq 0,\ \forall i\in\mathcal{N}_0, \\ & \sum_{j=0}^{n} \rho_{ij} = 1,\ \forall i\in\mathcal{N}_0, \quad \rho_{0i} = 0, \rho_{ii} = 0,\ \forall i\in\mathcal{N}, \quad \rho_{ij}\geq 0,\ \forall i,j\in\mathcal{N}_0, \\ & \sum_{i\in S\cup\{0\}} x_{i,S} = 1,\ S\in\mathcal{S}, \quad x_{i,S}\geq 0,\ \forall i\in S\cup\{0\}, S\in\mathcal{S}, \\ & y_{i,S}\geq 0, \quad \forall i\in\mathcal{N}\setminus S, S\in\mathcal{S},\end{aligned} \tag{EC.21}$$

where 0 denotes for the outside option and $\mathcal{N}_0 = \mathcal{N} \cup \{0\}$. We solve (EC.21) with 'ipopt' in the continuous optimization solver pyomo. We set the loss function of (EC.21) based on the chosen loss function.

When the outside option is not included in the assortments, the limit of MCCM is computed as follows:

$$\begin{aligned}
\min_{\boldsymbol{\lambda},\boldsymbol{\rho},\boldsymbol{x}_{\mathcal{S}},\boldsymbol{y}_{\mathcal{S}}} \quad & \text{loss}(\boldsymbol{p}_S, \boldsymbol{x}_S) \\
\text{s.t.} \quad & x_{i,S} = \lambda_i + \sum_{j\in\mathcal{N}\setminus S} y_{j,S}\rho_{j,i}, \quad \forall i \in S, S \in \mathcal{S}, \\
& y_{i,S} = \lambda_i + \sum_{j\in\mathcal{N}\setminus S} y_{j,S}\rho_{j,i}, \quad \forall i \in \mathcal{N}\setminus S, S \in \mathcal{S}, \\
& x_{i,S} = 0, \quad \forall i \in \mathcal{N}\setminus S, S \in \mathcal{S}, \\
& y_{i,S} = 0 \quad \forall i \in S, S \in \mathcal{S}, \\
& \sum_{i\in\mathcal{N}} \lambda_i = 1, \quad \lambda_i \geq 0, \ \forall i \in \mathcal{N}, \\
& \sum_{j\in\mathcal{N}} \rho_{ij} = 1, \ \rho_{ii} = 0, \ \forall i \in \mathcal{N}, \quad \rho_{ij} \geq 0, \ \forall i,j \in \mathcal{N}, \\
& \sum_{i\in S} x_{i,S} = 1, \ S \in \mathcal{S}, \quad x_{i,S} \geq 0, \ \forall i \in S, S \in \mathcal{S}, \\
& y_{i,S} \geq 0, \quad \forall i \in \mathcal{N}\setminus S, S \in \mathcal{S}.
\end{aligned} \tag{EC.22}$$

EC.6.2.5. Limit of LC-MNL The limit computation of the limit of LC-MNL is similar to the limit of MNL. Given an instance $\boldsymbol{p}_{\mathcal{S}}$ with collection $\mathcal{S}$, we first estimate the parameters of LC-MNL and then compute the best-fitting choice probabilities of LC-MNL to $\boldsymbol{p}_{\mathcal{S}}$ with the estimated parameters. Lastly, we compute the limit of LC-MNL with the best-fitting choice probabilities and the chosen loss function. The parameters are estimated by the following loglikelihood of LC-MNL:

$$\begin{aligned}
\boldsymbol{w}^*, \boldsymbol{v}^* = \arg\max_{\boldsymbol{w},\boldsymbol{v}} & LL(\boldsymbol{w},\boldsymbol{v}|\boldsymbol{p}_{\mathcal{S}}) = \sum_{S\in\mathcal{S}}\sum_{i\in S} n_S p_{i,S} \log \sum_{l=1}^{L} w_l \frac{\exp v_{li}}{\sum_{j\in S}\exp v_{lj}} \\
\text{s.t.} \quad & \sum_{l=1}^{L} w_l = 1, \\
& w_l \geq 0, \quad \forall l \in [L],
\end{aligned} \tag{EC.23}$$

where we assume that the population is described by a mixture of MNL models consisting of L classes, with w_l denoting the fraction of customers in class l, and $\boldsymbol{v}_l = (v_{l0}, v_{l1}, \cdots, v_{ln})$ denoting the parameters of the corresponding MNL model of customers in class l. (EC.23) is non-convex in the model parameters $\boldsymbol{w}$ and $\boldsymbol{v}$. We obtain an approximate solution by using the expectation-maximization (EM) algorithm, see Train (2008) for details. The number of iterations has been set to 50 for the estimation problem of LC-MNL with 10 classes and 40 times for the estimation problem of LC-MNL with 2 classes. For some rare randomly generated instances, we allow the iteration up to 100 to ensure convergence.

EC.6.3. Prediction experiment implementation details

EC.6.3.1. Revenue prediction with MDM using estimate-then-predict For an unseen assortment A, we follow the revenue prediction procedure in Figure 1 to obtain the revenue estimate of the assortment A.

EC.6.3.2. Revenue prediction with MDM under data uncertainty Considering the data uncertainty, for any given $\boldsymbol{p}_{\mathcal{S}}$, the worst case revenue prediction for assortment A is computed as follows:

$$\begin{array}{lll}
\min\limits_{\boldsymbol{x}_A,\boldsymbol{\lambda},\boldsymbol{\delta}} & \sum\limits_{i\in A} r_i x_{i,A} & \\
\text{s.t.} & -\delta_{A,S} \le \lambda_A - \lambda_S \le 1-(1+\epsilon)\delta_{A,S}, & \forall i\in A,\ (i,S)\in\mathcal{I}_{\mathcal{S}}, \\
 & -\delta_{S,A} \le \lambda_S - \lambda_A \le 1-(1+\epsilon)\delta_{S,A}, & \forall i\in A,\ (i,S)\in\mathcal{I}_{\mathcal{S}}, \\
 & \delta_{A,S}-1 \le x_{i,A}-x_{i,S} \le 1-\delta_{S,A}, & \forall i\in A,\ (i,S)\in\mathcal{I}_{\mathcal{S}}, \\
 & -(\delta_{A,S}+\delta_{S,A}) \le x_{i,A}-x_{i,S} \le \delta_{A,S}+\delta_{S,A}, & \forall i\in A,\ (i,S)\in\mathcal{I}_{\mathcal{S}}, \\
 & \lambda_S-\lambda_T \ge \epsilon, \quad \forall (i,S),(i,T)\in\mathcal{I}_{\mathcal{S}} \ \text{s.t.}\ p_{i,S}<p_{i,T}, & \\
 & \lambda_S-\lambda_T = 0, \quad \forall (i,S),(i,T)\in\mathcal{I}_{\mathcal{S}} \ \text{s.t.}\ p_{i,S}=p_{i,T}\neq 0, & \\
 & \sum\limits_{i\in A} x_{i,A} = 1, & \\
 & 0\le\lambda_A\le 1, \quad x_{i,A}\ge 0,\ \forall i\in A, \quad 0\le\lambda_S\le 1, \quad \delta_{A,S},\delta_{S,A}\in\{0,1\},\ \forall S\in\mathcal{S}, & \\
 & p_{i,S}(1-z\alpha_{i,S}) \le x_{i,S} \le p_{i,S}(1+z\alpha_{i,S}), & \forall (i,S)\in\mathcal{I}_{\mathcal{S}},
\end{array} \tag{EC.24}$$

where $\alpha_{i,S}\sqrt{\frac{1-p_{i,S}}{n_{i,S}}}$ with $n_{i,S}=n_S * p_{i,S}$ representing the number of customers chose product i when assortment S is offered. Here, $p_{i,S}\alpha_{i,S}$ is the standard error, and z is a constant multiplier that determines the width of the confidence interval. By replacing the "min" operator with the "max" operator, we obtain the best-case revenue estimate for assortment A.

EC.6.3.3. Revenue prediction with RUM Given the training data $\boldsymbol{p}_{\mathcal{S}}$, for an unseen assortment A, we estimate the revenue of A under the RUM by following the robust prediction method in (Farias et. al. 2013). To obtain the revenue estimate of A by solving the following linear optimization problem:

$$\begin{array}{ll}
\min\limits_{\boldsymbol{x}_A,\boldsymbol{x}_{\mathcal{S}},\boldsymbol{\lambda}} & \boldsymbol{r}^T\boldsymbol{x}_A \\
\text{s.t.} & \sum\limits_{\sigma\in\Sigma_n}\lambda(\sigma)\mathbb{I}[\sigma,i,S]-x_{i,S}=0, \quad \forall (i,S)\in\mathcal{I}_{\mathcal{S}}, \\
 & p_{i,S}(1-z\epsilon_{i,S})\le x_{i,S}\le p_{i,S}(1+z\epsilon_{i,S}), \quad \forall (i,S)\in\mathcal{I}_{\mathcal{S}}, \\
 & \sum\limits_{\sigma\in\Sigma_n}\lambda(\sigma)\mathbb{I}[\sigma,i,S]-x_{i,A}=0, \quad \forall i\in A, \\
 & \sum\limits_{\sigma\in\Sigma_n}\lambda(\sigma)=1, \quad \lambda(\sigma)\ge 0, \quad \forall\sigma\in\Sigma_n,
\end{array} \tag{EC.25}$$

where $\epsilon_{i,S} = \sqrt{\frac{1-p_{i,S}}{n_{i,S}}}$ with $n_{i,S} = n_S * p_{i,S}$ representing the number of customers chose product i when assortment S is offered and z is a constant multiplier that determines the width of the confidence interval. For our real data experiments, we have set z to be 2.5758, which corresponded to the smallest value of z of which (EC.25) was feasible over all tested instances; incidentally, this value of z corresponds to approximately 99% confidence interval for $x_{i,S}$. For the synthetic data experiment, we have set z to be 2.807, which corresponds to approximately 99.5% confidence interval for $x_{i,S}$, which corresponds z to be 2.807. We report the infeasible proportion of (EC.25) in the tested instances.

EC.6.3.4. Revenue prediction with MNL, LC-MNL For parametric models MNL and LC-MNL, we first estimate the parameters of the models with maximum likelihood estimation. With the estimators of the model parameters, we then compute the choice probabilities of the products in the given unseen assortment A and evaluate the revenue of assortment A.

EC.6.3.5. Revenue prediction with MCCM For MCCM, we first estimate the parameters of MCCM, the arrival rates $\boldsymbol{\lambda}$ and transition matrix $\boldsymbol{\rho}$ with maximum likelihood estimation. We then obtain the estimated revenue for an assortment A by solving the following linear optimization problem (Feldman and Topaloglu 2017):

$$\begin{aligned}
\max_{\boldsymbol{x},\boldsymbol{y}} \quad & \sum_{i\in\mathcal{N}_0} r_i x_i \\
\text{s.t.} \quad & x_i = \lambda_i + \sum_{j\in\mathcal{N}_0\setminus A} y_j \rho_{j,i}, \quad \forall i \in A, \\
& y_i = \lambda_i + \sum_{j\in\mathcal{N}_0\setminus A} y_j \rho_{j,i}, \quad \forall i \in \mathcal{N}_0 \setminus A, \\
& x_i = 0, \quad \forall i \in \mathcal{N}_0 \setminus A, \\
& y_i = 0, \quad \forall i \in A, \\
& \sum_{i\in A} x_i = 1, \\
& x_i, y_i \geq 0, \ \forall i \in \mathcal{N}_0.
\end{aligned} \tag{EC.26}$$

EC.6.3.6. Average Ranking of Models Based on Prediction Performance For the prediction experiments, we evaluate and rank five different models, including MDM, RUM, MNL, MCCM, and LC-MNL, based on their prediction performance across 50 testing instances. Specifically, we assess each model using a designated performance metric (Kendall Tau distance or relative revenue regret of the top-ranked assortment) and rank them from 1 to 5 for each testing instance. Rank 1 indicates the best performance, while rank 5 indicates the worst. For each of the 50 testing instances, the models are ranked individually. In cases where two or more models achieve identical performance scores within a testing instance, they are considered tied. Tied models are assigned an average rank that reflects their shared position. For example, if two models are tied for the second-best performance, they both receive a rank of 2.5 (the average of ranks 2 and 3). The remaining models are ranked accordingly, with subsequent positions adjusted to account for any ties. After ranking the models across all 50 testing instances, we calculate the average ranking for each model. This average rank provides an overall performance summary, indicating which models consistently perform better or worse across the multiple testing instances.

EC.7. Details on Synthetic Data Generation

EC.7.1. A general procedure of generating instances for synthetic data experiments in Section 6.2

Generate 50 instances with underlying models, either HEV or Probit with negative correlations in utilities. For each instance $\boldsymbol{p}_{\mathcal{S}}$, we generate with the following Step 1 - 3.

Step 1: Generate true underlying choice probabilities of the chosen model (details to be provided in the next two subsections).

Step 2: Generate multinomial samples for purchased records based on step 1.

Input: The collection of choice probabilities from Step 1, the historical assortments, and their offer times.

Procedure: For each instance: generate multinomial samples based on the choice probabilities, the assortments, and the offer times.

Output: A collection of multinomial samples representing product choices within each assortment.

Step 3: Obtain choice probabilities $\boldsymbol{p}_{\mathcal{S}}$ by computing average purchase frequencies.

Input: The multinomial samples from Step 2, the collection of historical assortments.

Procedure: For each instance, compute the average purchase frequency of each product in each assortment based on the samples generated in Step 2. These frequencies form the observed choice data $\boldsymbol{p}_{\mathcal{S}}$ for the experiments.

Output: A collection of average purchase frequencies for each product across the assortments.

EC.7.1.1. HEV instance in Section 6.2 Recall that the HEV model is a parametric subclass of the random utility model and assumes independent extreme value error distributions with nonidentical scales. There is no closed-form expression for computing choice probabilities. For each underlying HEV instance, we generate with following procedure:

Step 1: Generate deterministic utilities $\{\nu_i : i \in \mathcal{N} \cup \{0\}\}$ and scales parameters $\{\beta_i : i \in \mathcal{N} \cup \{0\}\}$

Input: A collection of products and the corresponding price of each product $\{r_i : i \in \mathcal{N}\}$.

Procedure for generating deterministic utilities $\{\nu_i : i \in \mathcal{N} \cup \{0\}\}$

- Compute the mean of prices $\bar{r} = \sum_{i \in \mathcal{N}} r_i$.

- Compute the standard deviation of $\{r_i : i \in \mathcal{N}\}$ and denote it as δ.
- Let $z_i = (r_i - \bar{r})/\delta$, for all $i \in \mathcal{N}$.
- Generate $\nu_i = \rho\, z_i + \sqrt{1-\rho^2}\, w_i$, for all $i \in \mathcal{N}$, where $\rho = -0.5$ and w_i is a random variable that follows the standard normal distribution, that is $w_i \sim N(0,1)$.
- Let the deterministic utility of the outside option be greater than the deterministic utility of each product by setting $\nu_0 = \max_{i\in\mathcal{N}} \nu_i + 1$.

Procedure for generating scales parameters $\{\beta_i : i \in \mathcal{N} \cup \{0\}\}$**:** Let $\beta_0 = 10$, and β_i be generated from a uniform distribution with range $[0.04, 1]$, for all $i \in \mathcal{N}$. This indicates that the utility of the outside option has the largest variance.

Output: Deterministic utilities $(\nu_i : i \in \mathcal{N} \cup \{0\})$ and scales parameters $(\beta_i : i \in \mathcal{N} \cup \{0\})$.

Step 2: Generate 10000 utility samples.

Input: 10000 repetitions of Step 1 to generate deterministic utilities $(\nu_i^k : i \in \mathcal{N} \cup \{0\})$ and scales parameters $(\beta_i^k : i \in \mathcal{N} \cup \{0\})$, for $k = 1, \ldots, 10000$.

Procedure: Generate 10000 utility samples. For each utility sample $\{\tilde{U}_i^k : i \in \mathcal{N} \cup \{0\}\}$, it is draw from a Gumbel distribution with mean $\{\nu_i^k : i \in \mathcal{N} \cup \{0\}\}$ and scales $\{\beta_i^k : i \in \mathcal{N} \cup \{0\}\}$.

Output: Utility samples $\{\tilde{U}_i^k : i \in \mathcal{N} \cup \{0\}\}$ for $k = 1, \ldots, 10000$.

Step 3: Compute the number of times each alternative in each assortment is purchased.

Input: Utility samples $\{\tilde{U}_i^k : i \in \mathcal{N} \cup \{0\}\}$ for $k = 1, \ldots, 10000$, and the collection of historical assortments.

Procedure:

- For each utility sample $\{\tilde{U}_i^k : i \in \mathcal{N} \cup \{0\}\}$, determine the product purchased by the customer based on the utility maximization principle. For every assortment S in the collection, the product $i \in S$ is purchased if $i = \arg\max_{j\in S\cup\{0\}} \tilde{U}_j^k$.
- Obtain the number of purchasing times of each product in each assortment based on the 10000 utility samples.

Output: The number of purchasing times of each product in each assortment given 10000 purchasing records.

Step 3: Obtain choice probabilities $\boldsymbol{p}_{\mathcal{S}}$ by dividing the number of purchasing times of each product in each assortment by 10000.

EC.7.1.2. Probit instance with negative correlations in utilities in Section 6.2 Recall that the Probit model is a parametric subclass of the random utility model and assumes that the stochastic components of utilities follow a Gaussian distribution. There is no closed-form expression for computing choice probabilities. For each underlying Probit instance with negative correlations in utilities, the instance generation procedure is similar to the above for the HEV instances. The key difference lies in Step 1 to generate model-dependent parameters and a slight difference in Step 2 to generate utility samples while Steps 3 and 4 are the same. Hence, we only give details for Steps 1 and 2 in the following.

Step 1: Generate deterministic utilities $\{\nu_i : i \in \mathcal{N} \cup \{0\}\}$ and covariance matrix $\{\Sigma_{ij} : i, j \in \mathcal{N} \cup \{0\}\}$.

Input: A collection of products.

Procedure for generating deterministic utilities $\{\nu_i : i \in \mathcal{N} \cup \{0\}\}$

- Generate ν_i for all $i \in \mathcal{N}$, such that ν_i is a random draw from the uniform distribution with range $[-2, 2]$.
- Let the deterministic utility of the outside option be greater than the deterministic utility of each product, where ν_0 is a random draw from the uniform distribution with range $[6, 10]$.

Procedure for generating the covariance matrix $\{\Sigma_{ij} : i, j \in \mathcal{N} \cup \{0\}\}$

- Generate $\{\Sigma_{ij} : i, j \in \mathcal{N} \cup \{0\}\}$ as a positive semidefinite matrix with negative correlation.
- For the diagonal entries which represent variances of the utilities of products, let Σ_{00} be a random draw from the uniform distribution with range $[30, 40]$, and for each $i \in \mathcal{N}$, let Σ_{ii} be a random draw from the uniform distribution with range $[0.5, 2]$.

Output: The deterministic utilities $\{\nu_i : i \in \mathcal{N} \cup \{0\}\}$ and the covariance matrix $\{\Sigma_{ij} : i, j \in \mathcal{N} \cup \{0\}\}$.

Step 2: Generate 10000 utility samples.

Input: 10000 repetitions of Step 1 to generate deterministic utilities $(\nu_i^k : i \in \mathcal{N} \cup \{0\})$ and the covariance matrix $\{\Sigma_{ij}^k : i, j \in \mathcal{N} \cup \{0\}\}$, for $k = 1, \ldots, 10000$.

Procedure: Generate 10000 utility samples. For each utility sample $\{\tilde{U}_i^k : i \in \mathcal{N} \cup \{0\}\}$, it is draw from a multivariate Gaussian distribution with mean $\{\nu_i^k : i \in \mathcal{N} \cup \{0\}\}$ and covariance matrix $\{\Sigma_{ij}^k : i, j \in \mathcal{N} \cup \{0\}\}$.

Output: Utility samples $\{\tilde{U}_i^k : i \in \mathcal{N} \cup \{0\}\}$ for $k = 1, \ldots, 10000$.

EC.7.2. Data Generation for experiments in Section 6.3

HEV instances We generate HEV instances where the deterministic utilities of products are randomly drawn from a uniform distribution on the interval $[-2, 2]$. The scale parameters of the Gumbel distributions are randomly drawn from a uniform distribution on the interval $[0.04, 2]$.

Probit instances We generate Probit instances where the deterministic utilities of products are randomly drawn from a uniform distribution on the interval $[-2, 2]$. The off diagonals of the covariance matrix, which capture the correlation of utilities, are randomly drawn from a uniform distribution on the interval $[-0.9, -0.1]$. The diagonals of the covariance matrix, which capture the variances of utilities, are randomly drawn from a uniform distribution on the interval $[0.5, 2]$.

EC.7.3. Data generation for experiments in Section 6.4

We generate parametric RUM instances where the deterministic utilities of products are randomly drawn from a uniform distribution on the interval $[-2, 2]$ and the scale parameters of the Gumbel distributions are randomly drawn from a uniform distribution on the interval [0.04,2].

Instances with Independent Copula: 50 instances are generated using independent copulas, following the procedure outlined in Section EC.7.1.1. The deterministic utilities and scale parameters are set according to the aforementioned uniform distributions.

Instances with Comonotonic Copula: 50 instances are generated with Comonotonic copulas, following a modified version of the procedure in SectionEC.7.1.1. The key difference lies in Step 2: instead of generating utility samples independently, a single uniform random variable $U \in (0, 1)$ is drawn for each utility sample. The utilities for all products are then generated by applying the inverse transform of the marginals for each product.

EC.8. Robustness of numerical results in Section 6.2

In Section 6.2, we use the product-assortment structure from the JD.com dataset to generate random instances for our experiments. A natural question is whether the performance of the compared models remains robust under different assortment structures. To investigate this, we generate random HEV-based instances with varied assortment structures and evaluate the models' estimation and prediction performance. The results reported under this randomly generated assortment setting are consistent with the findings in Section 6.2.

Data generation. Similar to the JD.com dataset, we consider $n = 8$ products, with the outside option always included in each assortment. The price of each product is set to the same as in the JD.com dataset. We construct eight distinct test scenarios that primarily differ in the assortment structures of the collection $\mathcal{S}$, specifically in how likely assortments of different sizes are to be included in $\mathcal{S}$. We consider two generation strategies for the assortment structures of the collections: one that favors the inclusion of smaller sized assortments and another that favors larger sized assortments. Let $k \in \{1, 2, 3, 4\}$ denote the number of products in an assortment. We define a probability distribution mass function $p(k)$ parameterized by $\alpha \in \{0.6, 0.7, 0.8, 0.9\}$. Under the strategy favoring the inclusion of assortments with fewer products in $\mathcal{S}$, we set $p(k) = \alpha^k / \sum_{j=1}^4 \alpha^j$. Under the strategy favoring the inclusion of assortments with more products in $\mathcal{S}$, we set $p(k) = \alpha^{5-k} / \sum_{j=1}^4 \alpha^j$. Specifically, given $p(k)$ for $k \in \{1, 2, 3, 4\}$, for each $\alpha \in \{0.6, 0.7, 0.8, 0.9\}$ and each fixed collection size $|\mathcal{S}| \in \{30, 25, 20, 15, 10\}$, we generate a random assortment collection $\mathcal{S}$ as follows. For $i = 1, \ldots, |\mathcal{S}|$,

- generate an independent sample X_i with probability mass function $P(X_i = k) = p(k)$, for $k = 1, 2, 3, 4$.

- given X_i generated as above, take assortment S_i to be the subset selected uniformly at random from all subsets of $\{1,2,\ldots,8\}$ whose size is X_i.
- add the outside option to S_i and include the resulting S_i in the collection $\mathcal{S}$.

For every collection $\mathcal{S}$ generated above, we generate 20 random HEV-based instances under $\mathcal{S}$. Every such instance $\boldsymbol{p}_{\mathcal{S}}$ is generated by following the procedure described in Section EC.7.1 with $\mu_0 = \max_{i\in\mathcal{N}} \mu_i + 10$ to specify the deterministic part of the utility of the outside option. The number of offer times n_S of each assortment $S \in \mathcal{S}$ is sampled based on its size $|S|$ as follows: if $|S| = 2$, then $n_S \sim \text{Uniform}(20, 2000)$; if $|S| = 3$, then $n_S \sim \text{Uniform}(20, 250)$; if $|S| = 4$, then $n_S \sim \text{Uniform}(20, 60)$; if $|S| = 5$, then $n_S \sim \text{Uniform}(20, 30)$. Overall, we evaluate the performance of models with a total of 800 HEV-based instances, consisting of 20 random instances for each combination of collection size in $\{30, 25, 20, 15, 10\}$ and each of the 8 test scenarios with different assortment structures of collections.

Comparison of explanatory, predictive, and prescriptive abilities. The performance of the compared models in estimation and prediction is almost consistent across different test scenarios, so we present results from a representative scenario. Specifically, we select the case with $\alpha = 0.8$ and $p(k) = \alpha^k / \sum_{j=1}^{4} \alpha^j$, for $k \in \{1,2,3,4\}$, where assortments with smaller sizes are more likely to be included in the assortment collection.

We report the average absolute deviation loss and standard error over the 20 randomly generated HEV-based instances under each considered collection size for the MDM, RUM, MNL, MCCM, and LC-MNL in Table EC.8. The results show that the models perform similarly to their performance on the JD.com dataset. Notably, the nonparametric MDM and RUM models achieve significantly lower average absolute deviation loss when approximating HEV-based instances, indicating superior explanatory power compared to the parametric models. Specifically, MDM exhibits a loss reduction of approximately 85% compared to the best-fitting MNL, 47% compared to MCCM, and 37% compared to LC-MNL for the largest collection size considered. Additionally, similar to experiments in Section 6.3, due to the nonconvex feasible region of the limit of MCCM, some rare randomly generated instances encounter convergence issues when solved using a continuous solver (IPOPT). The absolute deviation loss that we report for MCCM in Table EC.8 is only for the instances that successfully converged to an acceptable level. To further validate MDM's estimation performance, we also compare the average KL loss across MDM, MNL, and LC-MNL. Although we report sub-optimal KL losses for MDM due to computational challenges, the results in Table EC.9 demonstrate that both MDM and LC-MNL significantly outperform MNL across all tested scenarios, further highlighting MDM's robustness in accurately estimating choice probabilities. We do not report the computational time of the models, as the estimations were performed on different machines, making direct comparisons unreliable. However, we note that solving the limit of MDM can be computationally intensive, particularly when $|\mathcal{S}| = 30$

or when collections are with larger size. We leave it to future research to explore more efficient formulations and algorithms that can improve the scalability of solving the limit of MDM.

Next, we report the quality of predicted rankings of the tested collection of assortments captured by Kendall Tau distances in Table EC.10 and the average relative error in the revenue of the top-ranked assortment in Table EC.11. For executing the robust approach to RUM developed in Farias et. al. (2013) in these tables, we have taken the instances to be lying in the RUM representable region with 99.99% confidence. Robust revenue prediction using RUM has not always been feasible with this choice despite 99.99% being a very high confidence level to use: we find 20% of instances in the dataset obtained by restricting to $|\mathcal{S}| = 30$ turn out to be infeasible under RUM; 10% of instances are not feasible when $|\mathcal{S}| = 20$; and 5% instances are infeasible for $|\mathcal{S}| = 15$ and $|\mathcal{S}| = 10$. Additionally, similar to experiments in Section 6.3, due to the nonconvex feasible region of the MLE formulation of MCCM, some rare randomly generated instances encounter convergence issues when solved using a continuous solver (IPOPT). Therefore, in Tables EC.10 - EC.11, the numbers reported for RUM are obtained by restricting only to feasible instances, and the numbers reported for MCCM are obtained by restricting only to instances that successfully converged to an acceptable level. Overall, from Tables EC.10 and EC.11, we infer that MCCM and LC-MNL perform comparably well, followed by MDM, then MNL, with RUM ranking near the last among the models considered.

Table EC.8 Average absolute deviation loss (10^{-3}) comparison with HEV-based instances when $\alpha = 0.8$

$\lvert\mathcal{S}\rvert$	MDM	RUM	MNL	MCCM	LC-MNL
30	4.90 (0.44)	4.76 (0.39)	33.61 (1.52)	9.20 (3.52)	7.77 (0.71)
25	4.97 (0.45)	4.59 (0.43)	32.61 (1.76)	10.59 (3.74)	7.39 (0.58)
20	4.74 (0.73)	4.19 (0.71)	28.33 (1.39)	5.73 (1.05)	7.12 (0.85)
15	5.09 (0.77)	4.17 (0.64)	26.31 (2.34)	4.78 (0.78)	7.81 (1.04)
10	3.42 (0.55)	2.60 (0.48)	18.43 (1.80)	3.45 (0.68)	5.93 (1.05)

Notes. Standard errors are reported in parentheses.

Table EC.9 Comparison on average KL loss (10^{-3}) among MDM, MNL, and LC-MNL with HEV-based instances when $\alpha = 0.8$

$\lvert\mathcal{S}\rvert$	MDM	MNL	LC-MNL
30	4.39 (0.43)	513.30 (4.12)	2.54 (0.33)
25	7.52 (1.19)	514.76 (5.08)	2.59 (0.35)
20	9.59 (4.82)	515.60 (6.30)	2.22 (0.31)
15	6.44 (1.07)	511.15 (8.17)	2.44 (0.48)
10	4.48 (1.19)	520.00 (8.53)	1.34 (0.25)

Notes. Standard errors are reported in parentheses.

Table EC.10 Kendall Tau distance comparison with HEV-based instances when $\alpha = 0.8$

$\|\mathcal{S}\|$	$\|\mathcal{A}\|$	MDM	RUM	MNL	MCCM	LC-MNL
24	6	4.85 (0.49)	6.06 (0.68)	5.70 (0.49)	4.06 (0.41)	**4.05** (0.41)
20	5	4.05 (0.46)	5.30 (0.49)	4.60 (0.45)	**3.21** (0.44)	3.40 (0.29)
16	4	1.65 (0.30)	2.33 (0.36)	1.65 (0.27)	1.41 (0.35)	**1.05** (0.22)
12	3	1.50 (0.18)	1.63 (0.20)	1.50 (0.21)	**1.05** (0.17)	**1.05** (0.15)
8	2	**0.15** (0.08)	0.42 (0.11)	0.30 (0.10)	0.29 (0.11)	0.30 (0.10)

Notes. Standard errors are reported in parentheses. Bold values indicate the best-performing model in each row.

Table EC.11 Comparison of relative error (%) in revenue of the top-ranked assortment with HEV-based instances when $\alpha = 0.8$

$\|\mathcal{S}\|$	$\|\mathcal{A}\|$	MDM	RUM	MNL	MCCM	LC-MNL
24	6	25.59 (6.40)	30.64 (5.84)	28.47 (4.99)	12.89 (3.86)	**10.17** (3.41)
20	5	23.79 (5.04)	30.90 (6.32)	**22.30** (5.11)	16.76 (4.93)	23.53 (5.17)
16	4	14.56 (4.84)	27.89 (6.46)	21.04 (5.21)	14.48 (5.40)	**6.06** (2.81)
12	3	25.99 (5.61)	26.43 (4.97)	26.91 (5.88)	**13.37** (4.83)	17.01 (5.67)
8	2	**2.52** (1.61)	14.78 (5.11)	6.57 (2.92)	8.84 (4.15)	5.18 (2.67)

Notes. Standard errors are reported in parentheses. Bold values indicate the best-performing model in each row.

Comparison of overall predictive and prescriptive abilities. To assess the overall predictive and prescriptive performance of the models across all instances of all testing scenarios, we summarize the distributions of their rankings based on the average Kendall Tau distance of testing assortments and the average relative error in revenue of the top-ranked assortment in Tables EC.12 and EC.13, respectively. In total, 800 instances were tested. After excluding instances where nonparametric RUM prediction was not feasible or where the MLE estimation of MCCM failed to converge, 634 instances remained for comparison. When ties occur in model rankings, we assign the same rank to the tied models. Based on the results, all models demonstrate a reasonable ability to predict both the Kendall Tau distances of testing assortments and the best assortments in the test sets, meaning that the models are often tied at the Rank 1 position. However, their effectiveness varies in terms of how frequently they achieve top rankings, with some models consistently outperforming others in securing the first rank. Overall, both tables indicate that MCCM and LC-MNL perform comparably well, followed by MDM, while RUM and MNL exhibit the weakest performance among the models considered.

EC.9. Additional numerical results with synthetic data

In Experiments 1 - 2 below, we compare the representational ability of MDM with RUM and MNL model. Experiment 3 compares the prediction performance offered by the nonparametric approach proposed in this paper with that offered by models involving parametric assumptions. Experiment 4 compares the limit of approximating choice probabilities with MDM, RUM, and MNL models. Additional useful details on the precise setup of all the experiments are furnished in EC.6.

Table EC.12 Distributions (%) of Models' Ranking based on Kendall Tau Distance

Rank	1	2	3	4	5
MDM	45.74	18.45	17.67	11.99	6.15
RUM	38.33	10.57	12.30	14.35	24.45
MNL	39.27	13.72	15.14	20.98	10.88
MCCM	58.04	18.77	10.88	9.15	3.15
LC-MNL	57.89	22.87	10.41	6.62	2.21

Notes. Each row represents the distribution of rankings of each model. The sum of values along each column does not equal 100% due to ties between models. These ties occur most frequently at the top rank.

Table EC.13 Distributions (%) of Models' Ranking based on Relative Error in Revenue of the Top-ranked Assortment

Rank	1	2	3	4	5
MDM	52.84	8.68	9.15	19.24	10.09
RUM	46.37	5.36	8.83	18.30	21.14
MNL	50.79	10.88	12.46	11.67	14.20
MCCM	65.30	11.99	11.04	7.57	4.10
LC-MNL	66.72	12.62	11.36	7.10	2.21

Notes. Each row represents the distribution of rankings of each model. The sum of values along each column does not equal 100% due to ties between models. These ties occur most frequently at the top rank.

EC.9.1. The representation power and tractability of MDM compared to RUM and MNL.

In Experiment 1, we investigate the representational power of MDM for a large number of alternatives ($n = 1000$) by randomly perturbing choice probabilities obtained from an underlying MNL model. We test for the fraction of instances that can be represented by MDM where the parameter α controls the fraction of choice probabilities that are perturbed from the MNL model (a larger value indicates more entries are modified from the underlying MNL model). While checking the representability of these models can be done by solving linear programs, RUM quickly becomes intractable as n increases. In Figure EC.3, we see that even with small perturbations to the choice probabilities of the underlying MNL model, none of the MNL models can represent the perturbed choice data. However, MDM which subsumes the MNL model can capture many of these instances. This shows that MDM is a much more robust model than MNL model. The runtimes for these large instances were less than 1 second. The computational requirements for RUM make it impossible to run at this scale.

In Experiment 2, we compare the representational power and computational time for MDM and RUM for a small number of alternatives. We find that both MDM and RUM show good representational power: In particular, with the collection size $|\mathcal{S}| = 20$, round 80 percent of the instances can still be represented by MDM when 25 percent of the choice probability entries are perturbed; this drops to 60 percent when 100 percent of entries are perturbed, RUM has better representation power in these examples (see Figure EC.4). However, this comes at a significant run time cost even at this scale as seen in Figure EC.4 as compared to MDM.

EC.9.2. Revenue and choice probability prediction with nonparametric MDM.

In Experiment 3, we generate 20 random instances with a product size of 7 and a collection size $|\mathcal{S}|$ ranging among {20, 40, 80}, using nonidentical exponential distributions for the marginal distributions

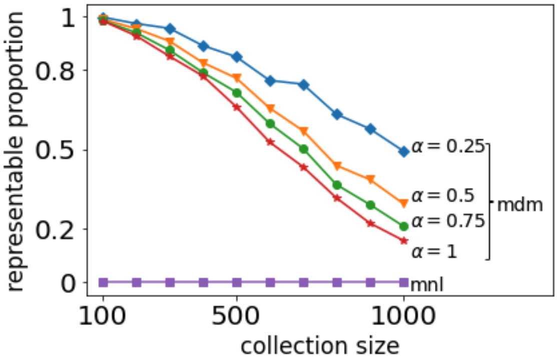


Figure EC.3 **The representational power of MDM**

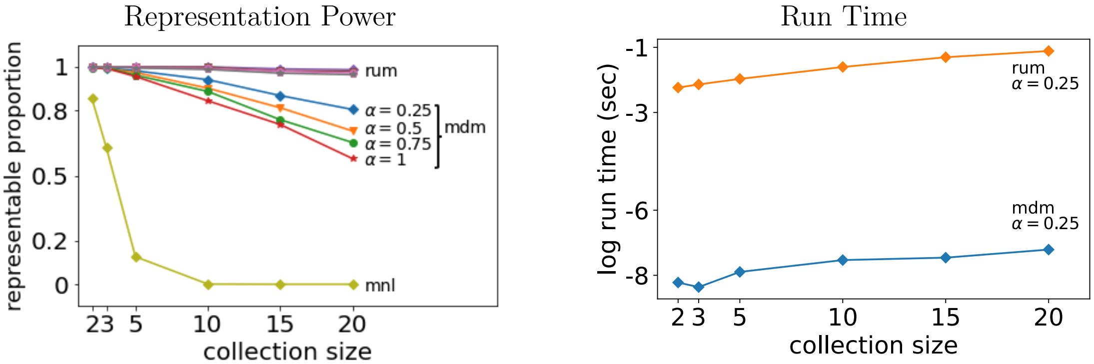
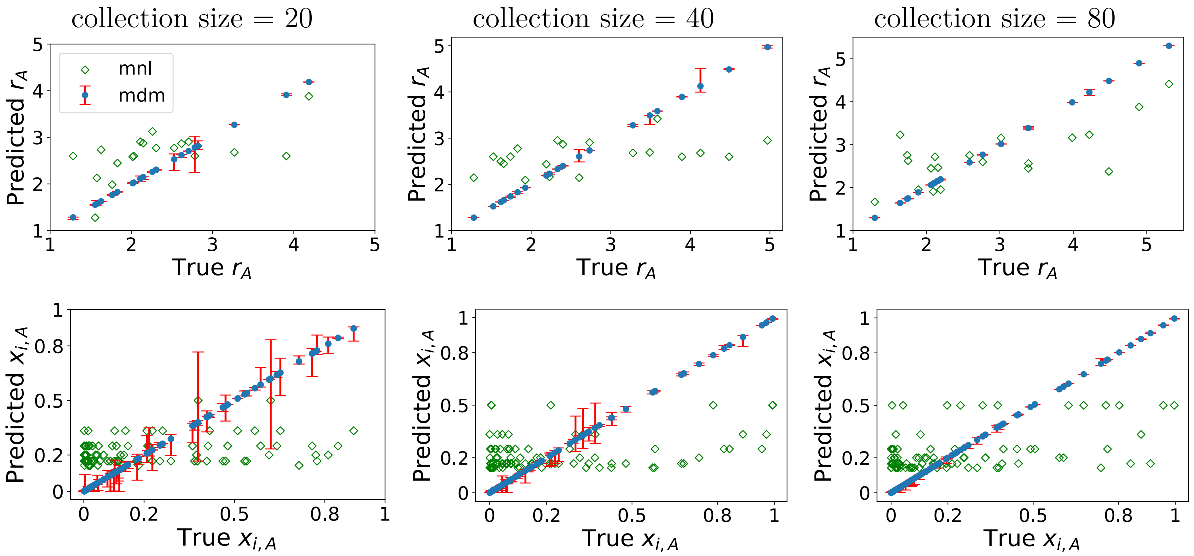


Figure EC.4 **Comparison of the performance of MDM and RUM**

to generate the underlying choice probabilities. In Figure EC.5, we compare the predictions offered by the following two methods: (1) computing the nonparametric MDM lower and upper bounds of revenue and choice probabilities for each instance by solving $\underline{r}(A)$ and $\bar{r}(A)$; and (2) restricting the marginal distributions for MDM to be identical exponential distributions (which leads to the underlying choice model being MNL), and estimating the preference parameters using maximum likelihood estimation (MLE); we then using the estimated MNL model to predict revenue for unseen assortments in each instance. While Figure EC.5 reveals the proposed nonparametric approach to be correctly predicting the true revenue or choice probabilities, the MLE of the parametric approach with mis-specified marginals is often found to lead to inaccurate predictions which are far from the truth and also far out of the nonparametric MDM prediction intervals. When more assortments are offered, the prediction under nonparametric MDM becomes more accurate while the prediction results made under the incorrect parametric model become worse. Thus, besides revealing the benefits of the proposed nonparametric data-driven approach for prediction based on MDM, Experiment 3 brings out the risks in stipulating apriori distributional assumptions on the model.



Notes. In each figure, the blue dots represent the true revenues or choice probabilities, while the red ranges represent the predicted revenue intervals or choice probability intervals with the nonparametric MDM, and the green squares represent the predicted revenues or choice probabilities using the MLE of the MNL model.

Figure EC.5 Comparison of prediction accuracy between MDM and MNL with randomly generated instances

EC.9.3. Estimation performance of MDM compared to RUM and MNL.

In Experiment 4, we compare the explanatory ability of MDM, RUM and MNL models by examining the cumulative absolute deviation loss suffered in fitting them to uniformly generated choice data instances. Figure EC.6 reveals that nonparametric MDM and RUM models are competitive and have much higher explanatory ability than MNL with increasing collection sizes. In particular, MDM incurs about 44% lesser loss, on average, than the best-fitting MNL model for the largest collection size considered.

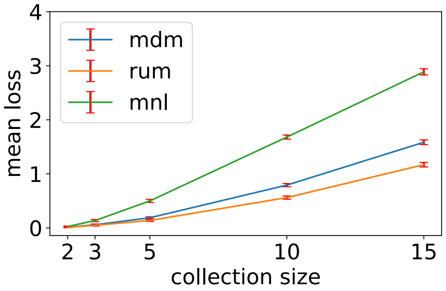


Figure EC.6 The limit loss comparison among MDM, RUM, and MNL

EC.9.4. Impact on explanatory abilities of models with common alternatives in the assortments.

We examine the explanatory power of the compared models by excluding the outside option in the assortments. We utilize the assortment data from the JD.com dataset, excluding both the outside option and any assortments containing only one product. The performance of models is evaluated by focusing on assortments offered at least $n_{\mathcal{S}}$ times, with $n_{\mathcal{S}} = 15$ resulting in $|\mathcal{S}| = 30$ and $n_{\mathcal{S}} = 20$ resulting in $|\mathcal{S}| = 21$. For each scenario, we randomly generate 50 underlying HEV and Probit instances. Since the exact characterization of MCCM and LC-MNL is not fully explored in the literature, we compare the representational power of each model by reporting the fraction of instances with an average KL loss below 10^{-6}. Since the MLE formulation of MCCM has a nonconvex feasible region, there are convergence issues for some rare randomly generated instances when the formulation is directly solved with a continuous solver. Thus, we report the average KL loss for all models only for those instances that converged to an acceptable level.

Table EC.14 and Table EC.15 report the fraction of instances with average KL loss being less than 10^{-6} for each model, approximating the representable proportions of the testing instances for each model, and the average KL losses. The results show that MDM maintains good representability and results in a low average KL loss across both HEV and Probit, while LC-MNL with 10 classes achieves slightly better representability and a lower average KL loss in the HEV setting and MCCM performs better in the Probit setting. In contrast, MNL and LC-MNL with 2 classes exhibit much higher average KL losses than the other models.

Table EC.14 Representation and average KL loss comparison with $|\mathcal{S}| = 30$

Model	HEV		Probit	
	fraction of instances with average KL loss $< 10^{-6}$	average KL loss	fraction of instances with average KL loss $< 10^{-6}$	average KL loss
MDM	0.66	1.05×10^{-5} (3.3×10^{-6})	0.71	1.47×10^{-6} (3.6×10^{-7})
MNL	0	1.07×10^{-2} (1.0×10^{-3})	0	6.84×10^{-4} (5.8×10^{-5})
MCCM	0.14	3.60×10^{-4} (1.62×10^{-4})	0.90	3.65×10^{-7} (2.36×10^{-7})
LC-MNL (2 classes)	0	2.16×10^{-3} (2.6×10^{-4})	0	2.82×10^{-4} (2.0×10^{-5})
LC-MNL (10 classes)	0.86	2.64×10^{-6} (1.92×10^{-6})	0.73	1.05×10^{-4} (2.9×10^{-5})

Notes. Standard errors are reported in parentheses.

EC.10. Additional Experiment Results with Real-World Data

In Experiments 5-7 below, we provide additional experiment results by using the dataset from JD.com (introduced in Shen et. al. 2020) to evaluate the feasibility of representing it with an MDM, the efficacy of predictions obtained by the proposed nonparametric approach, and the explanatory ability captured by the limit formulations. In Experiments 5-7, we ignore the constraints of the models on the outside option.

Table EC.15 Representation and average KL loss comparison with $|\mathcal{S}| = 21$

Model	HEV		Probit	
	fraction of instances with average KL loss $< 10^{-6}$	average KL loss	fraction of instances with average KL loss $< 10^{-6}$	average KL loss
MDM	0.98	1.62×10^{-7} (1.61×10^{-7})	0.98	2.88×10^{-7} (2.82×10^{-7})
MNL	0	7.18×10^{-3} (8.3×10^{-4})	0	5.03×10^{-5} (2.24×10^{-5})
MCCM	0.56	1.05×10^{-4} (3.6×10^{-5})	0.88	3.24×10^{-7} (3.20×10^{-7})
LC-MNL (2 classes)	0	1.22×10^{-3} (1.9×10^{-4})	0	1.55×10^{-4} (1.41×10^{-4})
LC-MNL (10 classes)	0.98	5.01×10^{-6} (4.95×10^{-6})	0.42	1.08×10^{-4} (1.79×10^{-4})

Notes. Standard errors are reported in parentheses.

Representational power comparison among several models. Experiment 5 compares the representation power of MDM with the MNL model and the class of regular choice models. The tested instances feature assortments which are offered at least n_S times, with n_S values ranging from 60 to 100. If we include data on the outside option, none of the models considered are found to exactly represent the data even when $n_S = 100$. By focusing on the sales data of the products offered by the firm, Table EC.16 shows that the nonparametric MDM and regular choice models are able to represent the choice data obtained from $n_S = 75$ and 100, whereas MNL models fail to represent any of the instances. We could not report the results for RUM here because of its intractability.

Table EC.16 The representability of MNL, MDM, and the class of regular choice models

$n_S \geq$	$\lvert\mathcal{S}\rvert$	MNL	MDM	Regular Model
60	13	0	0	0
75	12	0	1	1
100	11	0	1	1

Notes. 1 denotes an instance that can be represented by the tested model, while 0 denotes the opposite.

Estimation performance comparison between MDM and MNL. In Experiment 6, we compare the explanatory ability of the MDM and MNL model by computing the limit loss over choice data obtained by considering assortments that are offered at least n_S times, where n_S is set to vary from 1 to 100. Using 1-norm as the loss function, the results in Table EC.17 show that nonparametric MDM suffers much lesser cost in approximating the choice data, and hence greater explanatory ability. We also observe that the run time of solving the limit of MDM grows when the size of the assortment collection becomes larger. We further assess the accuracy of the nonparametric MDM and MNL models by comparing the observed (true) and estimated choice probabilities via scatter

Table EC.17 **Comparison of the estimation performance of MDM and MNL model**

n_S	$\lvert\mathcal{S}\rvert$	MDM		MNL	
$\geq$		loss	time (sec)	loss	time (sec)
1	134	0.223	3600	0.19	0.839
10	42	0.027	3600	0.16	0.253
20	29	0.019	13.654	0.15	0.193
30	24	0.017	6.298	0.15	0.318
40	19	0.016	2.379	0.14	0.149
50	15	0.012	0.237	0.14	0.146
60	13	0.011	0.094	0.13	0.114
75	12	0.0097	0.060	0.13	0.120
100	11	0.0098	0.046	0.13	0.112

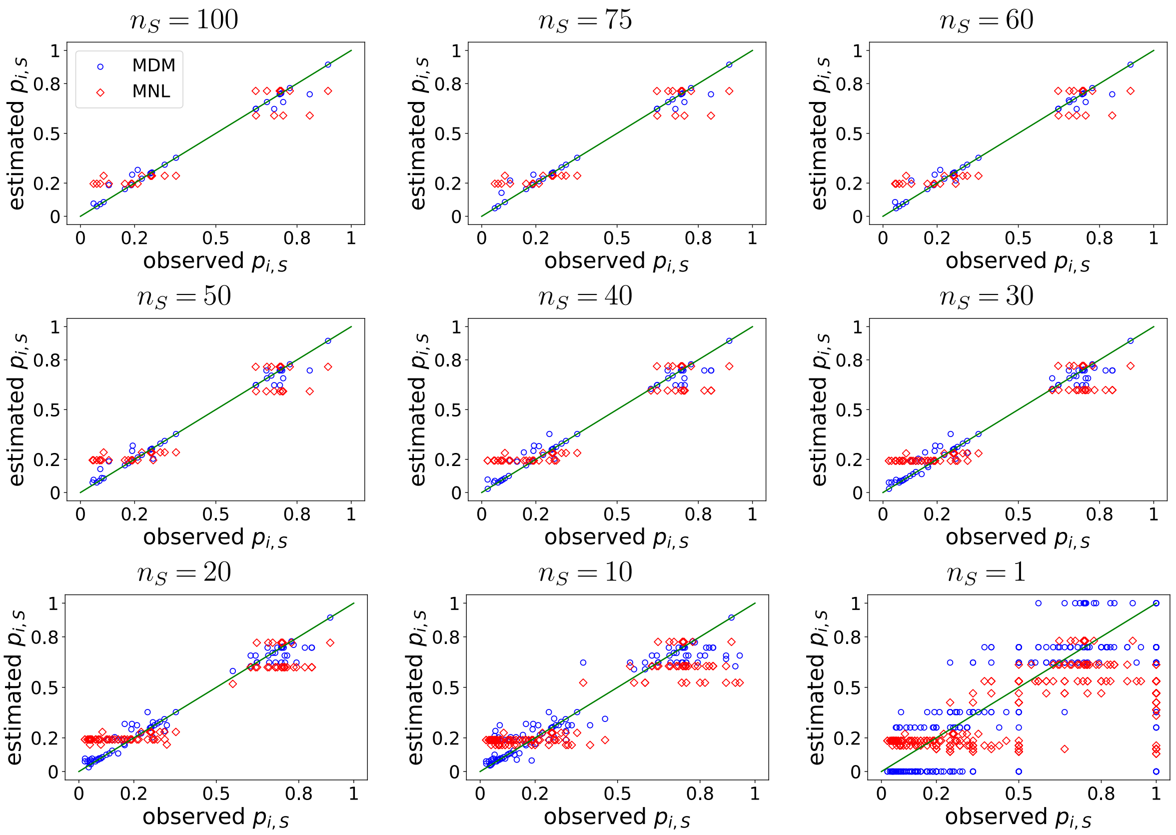
plots. Figure EC.7 shows these scatter plots, where each point represents an observed and estimated choice probability pair. The horizontal axis shows the observed probability and the vertical axis shows the estimated probability. The closer the points are to the 45-degree line segment (green segment in Figure EC.7), the better the estimation accuracy. The scatter plots reveal the following findings:

(i) When $n_S = \{50, 60, 75, 100\}$, MDM is seen to correctly estimate most data points due to its proximity to the 45^o line while most points from MNL estimation are still away from the 45^o line.

(ii) When $n_S = \{10, 20, 30, 40\}$, although both MDM and MNL models are limited in their abilities to exactly represent the choice data, MDM shows much better estimation accuracy than MNL model with most points by MDM being much closer to the 45^o line than the MNL model.

(iii) In the noisy environment where many assortments are just shown once (corresponding to $n_S = 1$), both MDM and MNL fail understandably with most points falling away from the 45^o line.

Prediction performance comparison between MDM and MNL. Experiment 7 evaluates the predictive-cum-prescriptive abilities of the nonparametric MDM and the MNL model by comparing their accuracies in identifying (i) a ranking over unseen test assortments based on their expected revenues, and (ii) the average revenue of the assortment identified to offer the largest revenue in the test set. Considering assortments that are shown at least n_S times (with n_S taken to vary from 20 to 50), we report the average out-of-sample performance over instances generated by randomly picking $\max\{2, \lfloor 0.2|\mathcal{S}|\rfloor\}$ fraction of the assortments to be the test set and the remaining to be the training set. For MNL, we use the Maximum Likelihood Estimator (MLE) obtained from training data to estimate choice probabilities for the test assortments and use them subsequently to rank the test assortments in a decreasing order of expected revenues. For MDM, we compute the robust revenue $\underline{r}(A)$ and the optimistic revenue $\bar{r}(A)$ and record the corresponding choice probabilities for each tested assortment A in the test set if the training data can be represented by MDM. If we find



Notes. In each plot, each point corresponds to the coordinate (true choice probability, estimated choice probability) of each observation of the processed data with n_S. The green line is the 45^o line. The blue dots represent the estimation results with the nonparametric MDM and the red squares represent the estimation results with the MNL model.

Figure EC.7 Scatter plots to compare the estimation accuracy of MDM and MNL

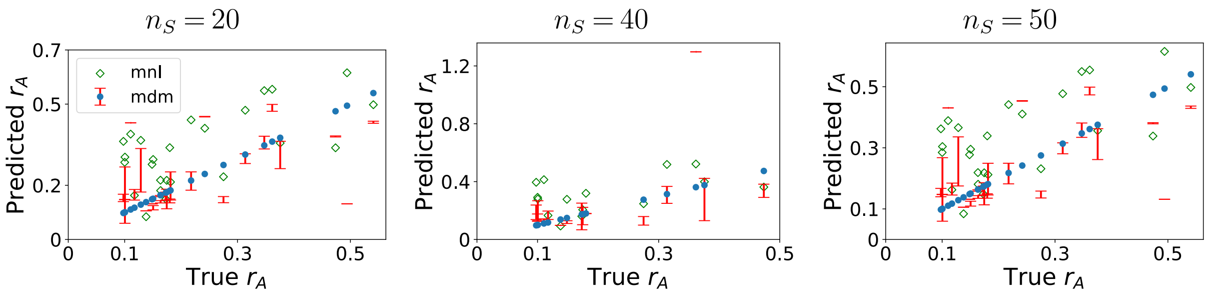
the training data to be not exactly representable by MDM, we solve the limit of MDM (Problem (17)) with the training data and use the choice probabilities yielded by solving (18) to proceed as before with ranking the assortments in a decreasing order of expected revenues.

For comparing the quality of rankings offered by the MDM and the MNL model, we take the well-known Kendall Tau distance (see Definition EC.2) as a natural metric for evaluating the closeness of the predicted ranking with the ground truth hidden from training. For both models, we also compare the true revenues of the assortments which are predicted to rank at the top. The average of these out-of-sample performance metrics across randomly generated train-test splits are reported in Table EC.18. The results in Table EC.18 show that the optimistic prediction results of nonparametric MDM outperform the MNL model, yielding uniformly lower average Kendall tau distances and higher average revenues for the predicted best assortments across all scenarios. Similarly, the robust prediction results of nonparametric MDM outperform the MNL models in most scenarios, except for instances with $n_S = 30$ in terms of average Kendall tau distances and instances with $n_S = 50$ in terms of average revenue predictions. Figure EC.8 illustrates that the nonparametric MDM approach predicts the true revenue more accurately than MNL, as the predicted intervals include the true revenue or are closer to it.

Table EC.18 The prediction performance of MDM and MNL

$n_S \geq$	$\lvert\mathcal{S}\rvert$	#test assortments	Average Kendall Tau Distance			Average Revenue of the Predicted Best Assortments		
			MNL	MDM_LB	MDM_UB	MNL	MDM_LB	MDM_UB
20	23	6	4.9	3.1	3.9	0.385	0.416	0.422
30	20	4	3.0	3.1	2.5	0.389	0.420	0.403
40	15	4	1.8	1.2	1.1	0.225	0.271	0.275
50	12	3	1.2	0.9	0.6	0.257	0.249	0.271

Notes. MDM_LB represents the results by solving $\underline{r}(A)$ while MDM_UB represents the results by solving $\bar{r}(A)$.



Notes. In each figure, the blue dots represent the true revenues, while the red ranges represent the predicted revenue intervals with the nonparametric MDM, & the green squares represent the predicted revenues using the MNL model.

Figure EC.8 Revenue predictions vs. true revenue and for the nonparametric MDM and the MNL model

References

Ben-Akiva, M., S. R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand.* The MIT Press, Cambridge, MA.

H. D. Block and J. Marschak. Random orderings and stochastic theories of response. In: *Economic Information, Decision, and Prediction. Theory and Decision Library*, Vol. 7-1. Springer, Dordrecht, 1960.

C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In: *Proceedings of the 10th International Conference on World Wide Web*, pp. 613–622, 2001.

V. F. Farias, S. Jagabathula, and D. Shah. A nonparametric approach to modeling choice with limited data. *Management Science*, 59(2):305–322, 2013.

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

S. Fiorini. A short proof of a theorem of Falmagne. *Journal of Mathematical Psychology*, 48(1):80–82, 2004.

D. Fudenberg, R. Iijima, and T. Strzalecki. Stochastic choice and revealed perturbed utility. *Econometrica*, 83(6):2371–2409, 2015.

S. Jagabathula and P. Rusmevichientong. The limit of rationality in choice modeling: Formulation, computation, and implications. *Management Science*, 65(5):2196–2215, 2019.

V. K. Mishra, K. Natarajan, D. Padmanabhan, C.-P. Teo, and X. Li. On theoretical and empirical aspects of marginal distribution choice models. *Management Science*, 60(6):1511–1531, 2014.

A. Schrijver. *Theory of Linear and Integer Programming.* John Wiley & Sons, Chichester, UK, 1998.

M. Shen, C. S. Tang, D. Wu, R. Yuan, and W. Zhou. JD.com: Transaction-level data for the 2020 MSOM data driven research challenge. *Manufacturing & Service Operations Management*, 2020.

K. E. Train. EM algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling*, 1(1):40–69, 2008.