

---

# The Limit of the Marginal Distribution Model in Consumer Choice

---

**Yanqiu Ruan**  
SUTD, Singapore  
yanqiu\_ruan@mymail.sutd.edu.sg

**Xiaobo Li**  
NUS, Singapore  
iselix@nus.edu.sg

**Karthyek Murthy**  
SUTD, Singapore  
karthyek\_murthy@sutd.edu.sg

**Karthik Natarajan**  
SUTD, Singapore  
karthik\_natarajan@sutd.edu.sg

## Abstract

Given data on choices made by consumers for different assortments, a key challenge is to develop parsimonious models that describe and predict consumer choice behavior. One such choice model is the marginal distribution model which requires only the specification of the marginal distributions of the random utilities of the alternatives to explain choice data. In this paper, we develop an exact characterization of the set of choice probabilities which are representable by the marginal distribution model consistently across any collection of assortments. Allowing for the possibility of alternatives to be grouped based on the marginal distribution of their utilities, we show (a) verifying consistency of choice probability data with this model is possible in polynomial time and (b) finding the closest fit reduces to solving a mixed integer convex program. Our results show that the marginal distribution model provides much better representational power as compared to multinomial logit and much better computational performance as compared to the random utility model.

## 1 Introduction

Discrete choice models have been used extensively in economics (see Allenby and Ginter (1995)), marketing (see McFadden (1986)), healthcare (see de Bekker-Grob et al. (2018)), transportation (see Ben-Akiva and Lerman (1985)) and operations management (see Talluri and Van Ryzin (2004)). Such models describe the behavior of one or more consumers who choose their most preferred alternative from a finite and discrete set of alternatives. The most popular model of discrete choice behavior is the random utility model (RUM) which postulates that the utilities of the alternatives are random variables and the consumers are utility maximizers. Modeling the joint probability distribution of the random utilities gives rise to the computation of the expected consumer utility (welfare) and the choice probabilities. Given the choice probabilities for a set of assortments, verifying if it is consistent with rational behaviour stipulated by RUM is however known to be NP-hard (see Jagabathula and Rusmevichientong (2019)). Jagabathula and Rusmevichientong (2019) showed that going beyond the RUM family is often required to obtain good prediction accuracy of choice behavior. There has been recent interest in developing choice models using machine learning techniques; examples include the decision forest choice model (Chen and Mišić, 2022) and neural network based choice models (Wang et al., 2020). The decision forest choice model, for example, has been shown to capture any choice probability data for a set of assortments; however this expressive power comes at the cost of requiring lots of data to describe the model. So a natural question is to study the representational power of different choice models (possibly not contained in the RUM family) and corresponding computational techniques.

In this work, we consider a class of choice models called marginal distribution model (MDM) proposed in Natarajan et al. (2009), which subsumes the additive perturbed utility (APU) model proposed in Fudenberg et al. (2015) as a special case. Unlike specifying a single joint distribution for the random utility as in RUM, a set of distributions with given marginal distributions are considered in MDM. The choice probabilities are computed for the extremal distribution in this set which maximizes expected consumer utility. We provide details in Section 2. A key advantage of this model is that choice probabilities are efficiently computable using convex optimization and the model is not contained within the RUM class.

**Key Contributions:**

- (1) We propose a group marginal distribution model (G-MDM) which allows alternatives to be grouped suitably based on the marginal distributions of their utilities. Based on this grouping, G-MDM spans the spectrum of models interpolating between MDM and the APU model. Analogous to the nested logit model (see, e.g., McFadden, 1980), the additional flexibility provided by G-MDM compared with MDM and APU allows one to incorporate domain knowledge on the similarities of different alternatives. Given choice data over a collection of assortments, we develop necessary and sufficient conditions to verify if a G-MDM can represent the observed choice probabilities (see Section 3). Unlike RUM, checking these conditions is possible in polynomial time.
- (2) In Section 4, we define the limit of the model as the smallest loss that can be obtained by fitting G-MDM to given choice data. By computing the limit, one can obtain an MDM which offers the best fit to given choice data. This estimation approach is novel, contrasting with existing approaches which need to make specific parametric assumptions on the marginal distributions to proceed with estimation (e.g. Natarajan et al. (2009); Mishra et al. (2014) and Yan et al. (2022)). Our formulation provides the first procedure to obtain an MDM with best fit to choice data nonparametrically, while utilizing grouping information available (if any).
- (3) To solve this limit problem, we develop a mixed integer convex program which is applicable generally. We also propose an algorithm which is polynomial in the number of the alternatives and can be of use when the assortment collection is not too large.
- (4) We utilize the feasibility conditions developed in Section 3 to develop novel prediction intervals for choice probabilities for assortments unseen in past data. (see Proposition 2 and Corollary 3).
- (5) In Section 5, we provide numerical results to show that G-MDM is computationally tractable in comparison to RUM and provides good representation power when the assumptions underlying parametric models such as the multinomial logit model are violated. Besides demonstrating the significant improvement in explanatory power offered by MDM when compared to best fitting MNL models, the experiments showcase how models utilizing grouping information can lead to much narrower prediction intervals than models which ignore group structure while training.

**2 Related Work and Preliminaries**

Several parametric choice models have been developed over the past few decades including the multinomial logit model (see, e.g., Luce, 1959), the nested logit model (see, e.g., McFadden, 1980), the generalized extreme value model (see, e.g., McFadden, 1978), the multinomial probit model (see, e.g., Daganzo, 1979) and the mixed logit model (see, e.g., McFadden and Train, 2000). Non-parametric models such as the rank list model (see, e.g., Block and Marschak, 1960) and the Markov chain choice model (Blanchet et al., 2016) have also been proposed. The parametric choice models in RUM vary in the specification of the joint distributions of the random utilities. McFadden and Train (2000) showed that any RUM choice model can be approximated closely by a mixed logit model. The class of RUM and the class of rank list models have been shown to be equivalent (Block and Marschak, 1960).

**2.1 Preliminaries**

Let  $\mathcal{N} = \{1, \dots, n\}$  be the set of products. Let  $S \subset \mathcal{N}$  be an assortment, i.e., a collection of products. Here we do not explicitly model the outside option, instead treat the outside option as one of the products in  $\mathcal{N}$ . Choice models have been developed to describe the choice probabilities for consumers facing any assortment  $S$ . The random utility model (RUM) assumes that the utility of each

alternative  $i \in \mathcal{N}$  takes the form  $\tilde{u}_i = \nu_i + \tilde{\epsilon}_i$ , where  $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)$  and  $\tilde{\boldsymbol{\epsilon}} = (\tilde{\epsilon}_1, \dots, \tilde{\epsilon}_n)$  denote the deterministic and stochastic parts of the utilities. Assuming a joint distribution  $\theta$  on the random part  $\tilde{\boldsymbol{\epsilon}}$ , the probability of choosing product  $i$  in assortment  $S$  is  $p_{i,S} = \mathbb{P}_{\tilde{\boldsymbol{\epsilon}} \sim \theta}(i = \arg \max_{j \in S} \nu_j + \tilde{\epsilon}_j)$ . In particular, if  $\tilde{\boldsymbol{\epsilon}}$  consist of independent and identically distributed Gumbel random variables, we obtain the well-known multinomial logit (MNL) model with the choice probability formula  $p_{i,S} = \frac{e^{\nu_i}}{\sum_{j \in S} e^{\nu_j}}$ .

**Representative agent model** In the representative agent model, a single agent makes a choice on behalf of the entire population. To make her choice, the agent takes into account the expected utility while preferring some degree of diversification. More precisely, given an assortment  $S$ , the representative agent solves an optimization problem of the form:

$$\max \left\{ \boldsymbol{\nu}^T \mathbf{x} - C(\mathbf{x}) \mid \mathbf{x} \in \Delta_{n-1}, x_i = 0 \forall i \notin S \right\}, \quad (1)$$

where  $\Delta_{n-1} = \{\mathbf{x} \in \mathbb{R}_+^n \mid \sum_{i \in \mathcal{N}} x_i = 1\}$  is the unit simplex. Here  $C(\mathbf{x}) : \Delta_{n-1} \mapsto \mathbb{R}$  is a convex perturbation function that rewards diversification in (1). The optimal  $x_i$  value provides the fraction of the population that chooses alternative  $i$  in assortment  $S$ . Hofbauer and Sandholm (2002) show that all RUM can be expressed using a representative agent model where the function  $C(\mathbf{x})$  satisfies additional requirements (see also Feng et al. (2017)).

**Additive perturbed utility model** The additive perturbed utility model was proposed by Fudenberg et al. (2015) and is a special case of the representative agent model. In this model, the perturbation is additive and separable with  $C(\mathbf{x}) = \sum_i c(x_i)$  where  $c(x) : [0, 1] \mapsto \mathbb{R}$  is a strictly convex function:

$$\max \left\{ \boldsymbol{\nu}^T \mathbf{x} - \sum_{i \in \mathcal{N}} c(x_i) \mid \mathbf{x} \in \Delta_{n-1}, x_i = 0 \forall i \notin S \right\}. \quad (2)$$

**Marginal distribution model** The marginal distribution model (MDM) is a semiparametric choice model which uses only limited information on the joint distribution of the random utilities. In the MDM, the joint distribution of the random vector  $\tilde{\boldsymbol{\epsilon}}$  is not specified, rather only the marginal distributions are specified. Let  $\Theta$  denote this set of joint distributions  $\theta$  for  $\tilde{\boldsymbol{\epsilon}}$ . Given an assortment  $S$ , the MDM computes the maximum expected consumer utility over all distributions in the set:

$$\sup_{\theta \in \Theta} \mathbb{E}_{\tilde{\boldsymbol{\epsilon}} \sim \theta} \left[ \max_{i \in S} \nu_i + \tilde{\epsilon}_i \right]. \quad (3)$$

The choice probability  $x_i^* = \mathbb{P}_{\tilde{\boldsymbol{\epsilon}} \sim \theta^*}(i = \arg \max_{j \in S} \nu_j + \tilde{\epsilon}_j)$  is evaluated for the distribution  $\theta^*$  which attains the maximum in (3).

**Assumption 1.** Each random term  $\tilde{\epsilon}_i, i \in \mathcal{N}$  is an absolutely continuous random variable with a strictly increasing marginal distribution  $F_i(\cdot)$  on its support and  $\mathbb{E}|\tilde{\epsilon}_i| < \infty$ .

**Lemma 1.** (Natarajan et al., 2009; Mishra et al., 2014; Chen et al., 2022) Under Assumption 1, the choice probabilities for a distribution which attains the maximum in (3) is unique and given by the optimal solution  $\mathbf{x}^*$  of a strictly concave maximization problem over the simplex:

$$\max \left\{ \sum_{i \in S} \nu_i x_i + \sum_{i \in S} \int_{1-x_i}^1 F_i^{-1}(t) dt \mid \sum_{i \in S} x_i = 1, x_i \geq 0 \forall i \in S \right\}, \quad (4)$$

with the convention that  $F_i^{-1}(0) = \lim_{t \downarrow 0} F_i^{-1}(t)$  and  $F_i^{-1}(1) = \lim_{t \uparrow 1} F_i^{-1}(t)$ .

The formulation of MDM in (4) shows that it is a special case of (1) where the perturbation function is strictly convex and separable of the form  $C(\mathbf{x}) = -\sum_i \int_{1-x_i}^1 F_i^{-1}(t) dt$  and when the marginal distributions are the identical, it reduces to (2). MDM thus provides a probabilistic utility interpretation of the additive perturbed utility model. The necessary and sufficient optimality conditions of MDM are given by:

$$\begin{aligned} \nu_i + F_i^{-1}(1 - x_i^*) - \lambda + \lambda_i &= 0, \quad \forall i \in S, \\ \lambda_i x_i^* &= 0, \quad \forall i \in S, \\ \sum_{i \in S} x_i &= 1, \\ x_i \geq 0, \lambda_i &\geq 0, \quad \forall i \in S, \end{aligned} \quad (5)$$

where  $\lambda$  and  $\lambda_i$  are the Lagrange multipliers associated with the constraints defining the simplex. An additional assumption that  $F_i^{-1}(1) = \infty$  (the random terms are unbounded to the right; for example the normal or the exponential distribution) is often made in prior work (see e.g., Mishra et al. (2014)) to guarantee strictly positive choice probabilities. However, in real datasets sometimes alternatives offered in an assortment might never be chosen by consumers. In this paper, we allow for this possibility. In the literature, MDM has been used to estimate parameters from individual choice data (see Natarajan et al. (2009); Mishra et al. (2014)) and solve product pricing (see Yan et al. (2022)), assortment optimization (see Ahipasaoglu et al. (2020)) and traffic equilibrium problems (see Ahipasaoglu et al. (2016, 2019)), often assuming specific marginal distributions. MDM is also known to recreate certain choice probabilities such as the multinomial logit choice probability formula under appropriately selected marginal distributions (see Mishra et al. (2014)).

### 3 Group Marginal Distribution Model

In this section, we introduce the group marginal distribution Model (G-MDM). The model specifies groups of alternatives in which the stochastic part of the utilities of alternatives within a group have the same marginal distribution.

**Assumption 2.** There exists a partition  $\mathcal{G} = \{G_1, G_2, \dots, G_K\}$  of the set of alternatives  $\mathcal{N}$  such that the marginal distribution of  $\tilde{\epsilon}_i$ , for any  $i \in G_l$  is given by  $F_l(\cdot)$ .

**A motivating example.** Consider a population of consumers choosing among computers. The available options are given by {Desktop, Laptop, Tablet}. It is natural to postulate that the distribution of the random utilities of the last two alternatives are more similar since they provide more flexibility in terms of movability and should be grouped together. Another example is travel mode choice for travellers among four options {AIR, TRAIN, BUS, CAR}. Here the last three alternatives can be potentially put in one group based on being ground travel modes.

Let  $g : \mathcal{N} \mapsto \mathcal{G}$  be a function that maps an alternative in  $\mathcal{N}$  to a group in  $\mathcal{G}$ . For example,  $g(i) = l$  means  $i \in G_l$ . Then one can write the objective function in the convex formulation for G-MDM as:

$$\sum_{i \in S} \nu_i x_i + \sum_{i \in S} \int_{1-x_i}^1 F_{g(i)}^{-1}(t) dt = \sum_{i \in S} \nu_i x_i + \sum_{l=1}^K \sum_{i \in S: g(i)=l} \int_{1-x_i}^1 F_l^{-1}(t) dt. \quad (6)$$

#### 3.1 A characterization of the choice probabilities described by G-MDM

Let  $\mathcal{S}$  denote a collection of subsets of  $\mathcal{N}$  with  $|\mathcal{S}| = m$ . Each  $S \in \mathcal{S}$  is an assortment of the alternatives presented to the consumers. For each assortment  $S \in \mathcal{S}$ , let  $p_{i,S}$  be the fraction of population who choose alternative  $i \in S$  where  $(p_{i,S} : i \in S) \in \Delta_{|S|-1}$ . Let  $\mathcal{I}$  denote the set of all pairs  $(i, S)$  with  $i \in S$  and  $S \in \mathcal{S}$ . Then the observed choice probability collection  $\mathbf{p}_{\mathcal{S}} = (p_{i,S} : i \in S, S \in \mathcal{S})$  is a vector in  $\mathbb{R}^{|\mathcal{I}|}$ . The key question we ask is: Given a choice probability collection  $\mathbf{p}_{\mathcal{S}}$ , does it belong to the set of choice probabilities that can be obtained from G-MDM? Is this verifiable in polynomial time? The following theorem provides an affirmative answer to this question.

**Theorem 1. (Feasibility conditions for G-MDM).** A choice probability collection  $\mathbf{p}_{\mathcal{S}}$  is feasible for G-MDM satisfying Assumptions 1 - 2 if and only if there exists  $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^n$  such that for all  $(i, S), (j, T) \in \mathcal{I}$  with  $g(i) = g(j)$ :

$$\begin{aligned} \lambda_S - \nu_i &> \lambda_T - \nu_j & \text{if } p_{i,S} < p_{j,T}, \\ \lambda_S - \nu_i &= \lambda_T - \nu_j & \text{if } 0 < p_{i,S} = p_{j,T}. \end{aligned} \quad (7)$$

Checking these conditions is possible in polynomial time.

*Proof.* Suppose  $\mathbf{p}_{\mathcal{S}}$  is feasible for G-MDM. From the optimality conditions (5), for any group  $l$ , alternatives  $i, j \in G_l$ , assortments  $S, T$  with  $i, j \in S \cap T$ , there exists a set of  $\lambda$  such that:

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

If  $p_{i,S} < p_{j,T}$ , we obtain  $\lambda_{i,S} \geq 0$  and  $\lambda_{j,T} = 0$  from the optimality conditions. Since  $F_l^{-1}(1-p)$  is a strictly decreasing function of  $p \in [0, 1]$ , we obtain:

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

If  $0 < p_{i,S} = p_{j,T}$ , we have  $\lambda_{i,S} = \lambda_{j,T} = 0$  from the optimality conditions. Then:

$$\lambda_S - \nu_i = F_l^{-1}(1 - p_{i,S}) = F_l^{-1}(1 - p_{j,T}) = \lambda_T - \nu_j.$$

Conversely, given  $\mathbf{p}_S$  and  $\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\nu} \in \mathbb{R}^n$  such that for all  $(i, S), (j, T) \in \mathcal{I}$  with  $g(i) = g(j)$ , (7) holds, one can construct marginal distributions for G-MDM that yields the choice probabilities. Details are provided in the appendix. □

In Theorem 1 the feasibility condition of G-MDM can be specified with  $\mathcal{O}(n+m)$  variables and  $\mathcal{O}(nm)$  constraints. This is possible in polynomial time using a linear program. By setting  $K = 1$  and  $K = n$ , we obtain Corollary 1 and Corollary 2 with identical marginals and nonidentical marginals respectively.

**Corollary 1. (Feasibility condition under identical marginals).** When  $K = 1$ ,  $\mathbf{p}_S$  is feasible for G-MDM if and only if there exists  $\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\nu} \in \mathbb{R}^n$  such that for all  $(i, S), (j, T) \in \mathcal{I}$ :

$$\begin{aligned} \lambda_S - \nu_i &> \lambda_T - \nu_j & \text{if } p_{i,S} < p_{j,T}, \\ \lambda_S - \nu_i &= \lambda_T - \nu_j & \text{if } 0 < p_{i,S} = p_{j,T}. \end{aligned} \quad (8)$$

The condition in Corollary 1 is equivalent to those derived in Fudenberg et al. (2015) and describe the probabilities that can be obtained with the additive perturbed utility model.

**Corollary 2. (Feasibility condition under nonidentical marginals).** When  $K = n$ ,  $\mathbf{p}_S$  is feasible for G-MDM if and only if there exists  $\boldsymbol{\lambda} \in \mathbb{R}^m$  such that for all  $(i, S), (i, T) \in \mathcal{I}$ :

$$\begin{aligned} \lambda_S &> \lambda_T & \text{if } p_{i,S} < p_{i,T}, \\ \lambda_S &= \lambda_T & \text{if } 0 < p_{i,S} = p_{i,T}. \end{aligned} \quad (9)$$

In particular,  $p_{i,S} \leq p_{i,T}$  if  $p_{j,S} < p_{j,T}$ , for all  $S, T \in \mathcal{S}, i, j \in S \cap T$ .

The conditions in Corollary 2 describe the probabilities that obtained with the most general version of MDM. In this case, the deterministic utilities  $\boldsymbol{\nu}$  do not provide additional modeling power. Corollary 2 illustrates that the feasibility conditions for MDM corresponds to finding an ordinal ranking of assortments common across products. Jagabathula and Rusmevichientong (2019) show that the feasibility conditions for RUM, on the other hand, correspond to keeping a consistent preference list (an ordinal ranking of products) across assortments. The key contribution of Theorem 1 is the polynomial time verifiability of choice probabilities with MDM while for RUM this is known to be NP-hard. A useful regularity property of G-MDM is provided next (see the appendix for the proof).

**Lemma 2.** If a collection of choice probabilities  $\mathbf{p}_S$  is feasible for G-MDM, then  $p_{i,S} \leq p_{i,S \cap T}$  for all  $(i, S), (i, S \cap T) \in \mathcal{I}$ .

## 3.2 Representational power with G-MDM

### 3.2.1 Effect of grouping of the marginals

The results from the previous section imply that if we have two partitions of the alternatives denoted by  $\mathcal{G}_1$  and  $\mathcal{G}_2$  where each group in  $\mathcal{G}_1$  is a subset of a group in  $\mathcal{G}_2$ , then the set of choice probabilities that G-MDM captures with  $\mathcal{G}_2$  is a subset of that captured with  $\mathcal{G}_1$ . This provides a natural nesting structure as groups merge. In the Appendix, we provide an example to show that having a larger number of groups can provide strictly more representational power.

### 3.2.2 Comparison of representation between MDM and RUM

We compare the representational power of RUM with G-MDM when the number of groups  $K = n$ . The theorem is stated next (see the figure below) and the proof is provided in the appendix.

**Theorem 2. (Representation power of MDM and RUM).** When  $n = 2$ , RUM and MDM are equivalent; when  $n = 3$ , RUM subsumes MDM; when  $n \geq 4$ , there exist choice probabilities that can be represented by both RUM and MDM and neither of the models subsumes the other.

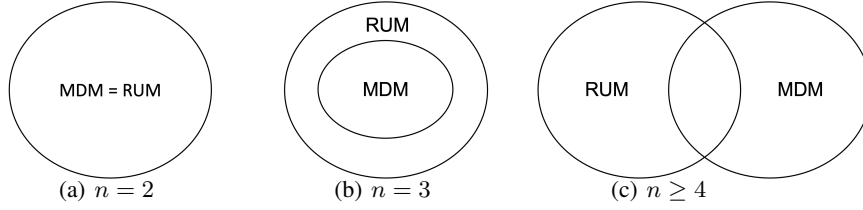


Figure 1: Comparison of the representation between MDM and RUM for different values of  $n$

#### 4 Limit of G-MDM: Fitting the best model to choice data

In this section, we investigate the problem of finding the G-MDM that fits given choice probability data as closely as possible. We measure the fit between the observed choice probabilities  $\mathbf{p}_S$  and the choice probabilities  $\mathbf{x}_S$  obtained from G-MDM by means of a nonnegative, strictly convex loss function. The loss function is assumed to satisfy the property that  $\text{loss}(\mathbf{p}_S, \mathbf{x}_S) = 0$  if and only if  $\mathbf{p}_S = \mathbf{x}_S$ ; examples included log loss and norm-based loss functions. We define the limit of the G-MDM as the smallest value of  $\text{loss}(\mathbf{p}_S, \mathbf{x}_S)$  attainable by fitting observed data with G-MDM. The formulation of computing the limit of the G-MDM is given by:

$$\begin{aligned}
& \min_{\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}} \sum_{S \in \mathcal{S}} \text{loss}(\mathbf{p}_S, \mathbf{x}_S) \\
& \text{s.t. } \lambda_S - \nu_i > \lambda_T - \nu_j \text{ if } x_{i,S} < x_{j,T}, \quad \forall (i, S), (j, T) \in \mathcal{I} : g(i) = g(j), \\
& \lambda_S - \nu_i = \lambda_T - \nu_j \text{ if } 0 < x_{i,S} = x_{j,T}, \quad \forall (i, S), (j, T) \in \mathcal{I} : g(i) = g(j), \quad (10) \\
& \sum_{i \in S} x_{i,S} = 1, \quad \forall S \in \mathcal{S}, \\
& x_{i,S} \geq 0, \quad \forall (i, S) \in \mathcal{I},
\end{aligned}$$

where the decision variables are the choice probabilities  $x_{i,S}$  associated with alternative-assortment pairs, the Lagrange multipliers  $\lambda_S$  associated with assortments, and the deterministic utilities  $\nu_i$  associated with the alternatives. An example is provided in the appendix to show that the set of choice probabilities that are representable by G-MDM is a nonconvex set in general.

##### 4.1 A Mixed integer convex program for the limit of G-MDM

We formulate (10) as a mixed integer convex program (MICP) next.

**Proposition 1.** Problem (10) is equivalent to the MICP:

$$\begin{aligned}
& \min_{\mathbf{x}_S, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\delta}, \mathbf{y}} \sum_{S \in \mathcal{S}} \text{loss}(\mathbf{p}_S, \mathbf{x}_S) \\
& \text{s.t. } -\delta_{i,j,S,T} \leq x_{i,S} - x_{j,T} \leq 1 - (1 + \epsilon)\delta_{i,j,S,T}, \quad \forall (i, S), (j, T) \in \mathcal{I} : g(i) = g(j), \\
& \lambda_S - \nu_i - \lambda_T + \nu_j \geq -1 + (1 + \epsilon)\delta_{i,j,S,T}, \quad \forall (i, S), (j, T) \in \mathcal{I} : g(i) = g(j), \\
& x_{i,S} \leq 1 - y_{i,S}, \quad \forall (i, S) \in \mathcal{I}, \\
& -y_{i,S} - y_{j,T} - \delta_{i,j,S,T} - \delta_{j,i,T,S} \leq \lambda_S - \nu_i - \lambda_T + \nu_j, \quad \forall (i, S), (j, T) \in \mathcal{I} : g(i) = g(j), \\
& \lambda_S - \nu_i - \lambda_T + \nu_j \leq y_{i,S} + y_{j,T} + \delta_{i,j,S,T} + \delta_{j,i,T,S}, \quad \forall (i, S), (j, T) \in \mathcal{I} : g(i) = g(j), \\
& \sum_{i \in S} x_{i,S} = 1, \quad \forall S \in \mathcal{S}, \\
& x_{i,S} \geq 0, \quad \forall (i, S) \in \mathcal{I}, \\
& 0 \leq \lambda_S - \nu_i \leq 1, \quad \forall (i, S) \in \mathcal{I}, \\
& \delta_{i,j,S,T} \in \{0, 1\}, \quad \forall (i, S), (j, T) \in \mathcal{I} : g(i) = g(j), \\
& y_{i,S} \in \{0, 1\}, \quad \forall (i, S) \in \mathcal{I},
\end{aligned} \tag{11}$$

for some small positive number  $\epsilon > 0$ .

Solving Problem (10) can NP-hard in general when considering dependence on both the number of alternatives and the size of collection of assortments and if the assortments are totally unstructured. However, solving the limit of G-MDM with nonidentical marginals may be efficient if the given size of the collection of assortments is small.

**Theorem 3.** There is an algorithm to solve Problem (11) with nonidentical marginals that is polynomial in the product size  $n$ .

## 4.2 Prediction on new unseen assortments

Besides yielding a fit to choice data, an optimal solution  $(\mathbf{x}_S, \boldsymbol{\lambda}, \boldsymbol{\nu})$  of the formulation can also be used to make predictions on choice probabilities for assortments unseen in the entire dataset. To accomplish this, define the following upper and lower envelopes:

$$\begin{aligned}\bar{F}_l(z) &= \min \{1 - x_{i,S} \mid (i, S) \in \mathcal{I}, g(i) = l, \lambda_S - \nu_i \geq z\} \text{ and} \\ \underline{F}_l(z) &= \max \{1 - x_{i,S} \mid (i, S) \in \mathcal{I}, g(i) = l, \lambda_S - \nu_i \leq z\},\end{aligned}$$

for any marginal distribution collection  $(F_l : l = 1, \dots, K)$  capable of yielding  $(\mathbf{x}_S, \boldsymbol{\lambda}, \boldsymbol{\nu})$  as a G-MDM. In defining  $\bar{F}_l$  and  $\underline{F}_l$  above, we follow the usual convention that  $\min \emptyset = +\infty$  and  $\max \emptyset = -\infty$ . For any marginal distribution  $F_l$  consistent with G-MDM  $(\mathbf{x}_S, \boldsymbol{\lambda}, \boldsymbol{\nu})$ , we have the bounds  $\underline{F}_l \leq F_l \leq \bar{F}_l$  as a consequence of the monotonicity of  $F_l, \underline{F}_l$  and  $\bar{F}_l$ : To see this, observe that a marginal distribution  $F_l$  violating  $F_l(z) \geq \underline{F}_l(z)$  cannot satisfy the optimality condition  $x_{i,S} = 1 - F_l(\lambda_S - \nu_i)$  in (5) for  $(i, S) \in \mathcal{I}$  such that  $\lambda_S - \nu_i = \underline{F}_l^{-1}(\underline{F}_l(z))$ . An analogous argument can be made for the upper bound  $F_l(\cdot) \leq \bar{F}_l(\cdot)$ .

Given an assortment  $A \subseteq \mathcal{N}$  unseen in the choice data, an explicit characterization of the set of choice probabilities  $(x_{k,A} : k \in A)$  which are consistent with the G-MDM fit  $(\mathbf{x}_S, \boldsymbol{\lambda}, \boldsymbol{\nu})$  is given by Proposition 2 below. To state Proposition 2, let

$$\begin{aligned}e_l &= \inf\{z : \underline{F}_l(z) > 0\} \quad \text{and} \quad r_l = \sup\{z : \bar{F}_l(z) < 1\}, \\ \bar{e}_l &= \inf\{z : \bar{F}_l(z) > 0\} \quad \text{and} \quad \bar{r}_l = \sup\{z : \underline{F}_l(z) < 1\},\end{aligned}$$

denote the left and right end-points of the envelope  $\underline{F}_l, \bar{F}_l$ , respectively, for  $l \leq K$ . Observe from the definitions of  $\underline{F}_l, \bar{F}_l$  that these endpoints satisfy  $\bar{e}_l < e_l \leq \bar{r}_l < r_l$ .

**Proposition 2.** Together with the G-MDM instance  $(\mathbf{x}_S, \boldsymbol{\lambda}, \boldsymbol{\nu})$ , a choice probability vector  $(x_{i,A} : i \in A)$  over a new assortment  $A \notin \mathcal{S}$  is feasible for G-MDM satisfying Assumptions 1 - 2 if and only if it is of the form,

$$x_{i,A} = 1 - [\alpha_i \underline{F}_{g(i)}(\lambda - \nu_i) + (1 - \alpha_i) \bar{F}_{g(i)}(\lambda - \nu_i)], \quad i \in A, \quad (12)$$

for some  $\lambda \in \mathbb{R}$  such that,

$$\sum_{i \in A} [1 - \bar{F}_{g(i)}(\lambda - \nu_i)] \leq 1 \leq \sum_{i \in A} [1 - \underline{F}_{g(i)}(\lambda - \nu_i)], \quad (13)$$

and  $(\alpha_i : i \in A) \in [0, 1]^{|A|}$  satisfying  $\sum_{i \in A} x_{i,A} = 1$  along with the following conditions:

- a)  $\alpha_i \in (0, 1)$  if  $i \in A$  is such that  $e_{g(i)} \leq \lambda - \nu_i \leq \bar{r}_{g(i)}$ ; and
- b) for any  $i, j \in G_l \cap A$  such that  $\underline{F}_l(\lambda - \nu_i) = \underline{F}_l(\lambda - \nu_j)$  and  $\bar{F}_l(\lambda - \nu_i) = \bar{F}_l(\lambda - \nu_j)$ ,
  - $\alpha_i > \alpha_j$  if  $\nu_i > \nu_j$  and  $\lambda - \nu_i$  or  $\lambda - \nu_j \in [e_l, \bar{r}_l]$ ,
  - $\alpha_i = \alpha_j$  if  $\nu_i = \nu_j$  and  $\lambda - \nu_i = \lambda - \nu_j \in [e_l, \bar{r}_l]$ ,
  - $\alpha_i > \alpha_j$  or  $\alpha_i = \alpha_j = 1$  if  $\nu_i > \nu_j$  and  $\lambda - \nu_j < e_l$ ,
  - $\alpha_i > \alpha_j$  or  $\alpha_i = \alpha_j = 0$  if  $\nu_i > \nu_j$  and  $\lambda - \nu_i > \bar{r}_l$ .

A consequence of Proposition 2 is the following simpler characterization of prediction intervals for unseen choice probabilities consistent with given G-MDM.

**Corollary 3.** Given a G-MDM instance  $(\mathbf{x}_S, \boldsymbol{\lambda}, \boldsymbol{\nu})$  and a product  $k$  from a new assortment  $A \notin \mathcal{S}$ , a prediction interval for the choice probability  $x_{k,A}$  which includes all G-MDM consistent with the instance  $(\mathbf{x}_S, \boldsymbol{\lambda}, \boldsymbol{\nu})$  is given by,

$$[1 - \bar{F}_{g(k)}(\bar{\lambda} - \nu_k), 1 - \underline{F}_{g(k)}(\lambda - \nu_k)] \quad (15)$$

where  $\lambda$  and  $\bar{\lambda}$ , respectively, are the supremum and infimum values of the collection of  $\lambda \in \mathbb{R}$  satisfying (13).

Observe that the width of the interval in (15) decreases with increased grouping of products. Indeed this follows from noting that the set  $z_S^l = \{(i, S) \in \mathcal{I} : g(i) = l\}$  defining the marginal distribution  $F_l$  includes more product-assortment pairs for a larger group, thereby leading to a more narrow gap  $\bar{F}_l - \underline{F}_l$  between the upper and lower envelopes. Recall that MDM with non-identical marginals has a greater modeling power as it yields a larger feasible region in (10). The benefits of carefully grouping the products to have common marginals, on the other hand, comes from a narrower range in (15) for predicting choice probabilities for assortments unseen in the dataset.

## 5 Experiments

In Experiments 1 - 3 below, we compare the representational and explanatory abilities of G-MDM with RUM and MNL model. Experiment 4 is devoted to exploring how imparting domain knowledge into the model via grouping reduces the length of prediction intervals (thereby pointing to lesser data requirements).

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 G-MDM where the parameter  $\alpha$  controls the fraction of choice probabilities that are perturbed from MNL (a larger value indicate more entries are modified from the underlying MNL model). While the feasibility check of these models can be done by solving linear programs, RUM quickly becomes intractable as  $n$  increases. The details of the data generation are provided in the appendix. In Figure 2, we see that even with small perturbations to the choice probabilities of the underlying MNL model, the class of all MNL models is almost always infeasible. On the other hand, MDM which subsumes MNL can capture many of these instances. This shows that MDM is a much more robust model than MNL. 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. MDM shows good representational power. For example with  $m = 20$ , where 25 percent of the choice probability entries are perturbed, around 80 percent of the instances can still be represented by G-MDM and when 100 percent of entries are perturbed, this drops to 60 percent. RUM has better representation power in these examples (see Figure 3(a)). However this comes at significant run time cost even at this scale as seen in Figure 3(b) as compared to MDM.

In Experiment 3, we compare the explanatory ability of MDM, APU, RUM and MNL by examining the cumulative absolute deviation loss suffered by each of them in fitting uniformly generated choice data instances. Figure 4 reveals that nonparametric MDM and RUM models 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 ( $m = 15$ ) considered.

The results of Experiment 4, reported in Figure 5, show how grouping alternatives leads to significantly narrower prediction intervals for choice probabilities in unseen assortments. Varying  $m =$  the number of assortments for which choice data are available, we utilize Corollary 3 to construct and report prediction interval lengths averaged over unseen choice probabilities. As expected, both models lead to tighter intervals when more choice data is made available for training. However, the predicted intervals of the model using grouping information are 70% - 88% narrower than the model assuming no groups. Notably, the quality of prediction interval obtained by considering grouping information with choice data over just  $m = 2$  assortments is unmatched by that obtained with  $m = 15$  seen assortments when grouping information is ignored while training.



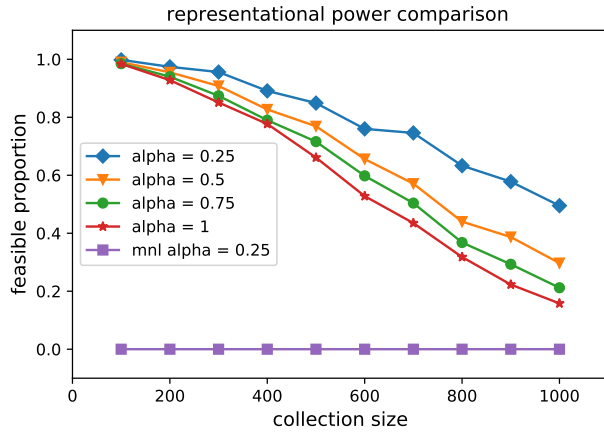


Figure 2: The representational power of MDM in Experiment 1

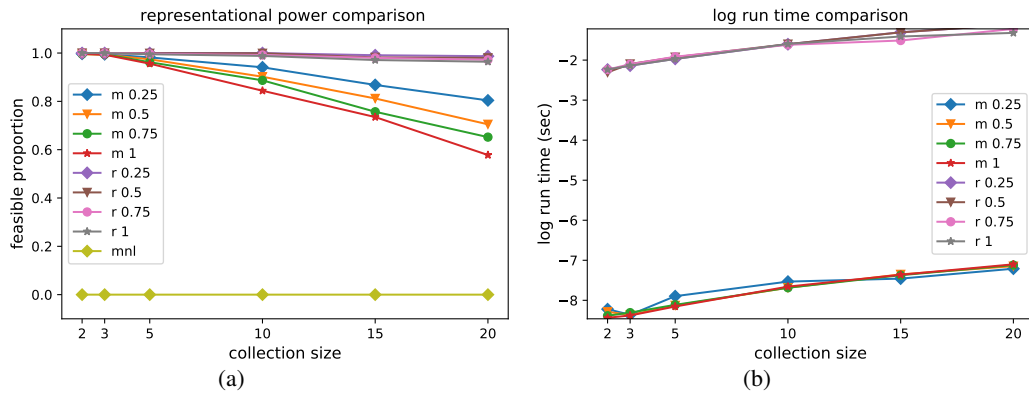


Figure 3: Comparison of the performance of MDM and RUM in Experiment 2, m stands for MDM, r stands for RUM, numbers stand for the perturbation parameters

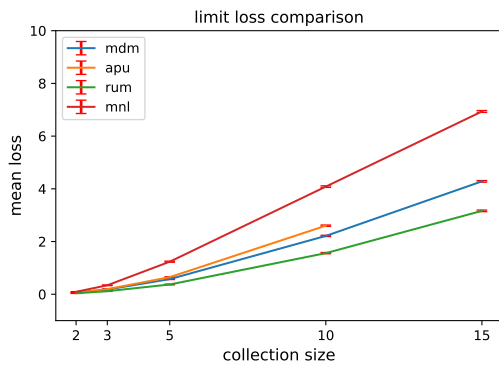


Figure 4: The limit loss comparison

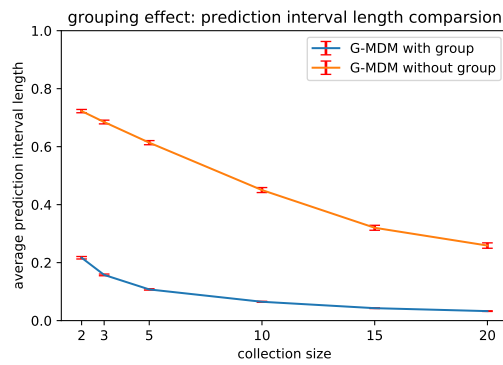


Figure 5: The grouping effect in G-MDM

## References

S. Ahipasaoglu, X. Li, Z. Sun, and Y. Yuan. Assortment optimization under heteroscedastic data. Available at SSRN 3638783, 2020.

- S. D. Ahipasaoglu, U. Arikan, and K. Natarajan. On the flexibility of using marginal distribution choice models in traffic equilibrium. *Transportation Research Part B: Methodological*, 91:130–158, 2016.
- S. D. Ahipasaoglu, U. Arikan, and K. Natarajan. Distributionally robust markovian traffic equilibrium. *Transportation Science*, 53(6):1546–1562, 2019.
- G. M. Allenby and J. L. Ginter. Using extremes to design products and segment markets. *Journal of Marketing Research*, 32(4):392–403, 1995.
- M. Ben-Akiva and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, 1985.
- J. Blanchet, G. Gallego, and V. Goyal. A markov chain approximation to choice modeling. *Operations Research*, 64(4):886–905, 2016.
- H. D. Block and J. Marschak. Random orderings and stochastic theories of response. In: *Economic Information, Decision, and Prediction. Theory and Decision Library*, 7-1, 1960.
- L. Chen, W. Ma, K. Natarajan, D. Simchi-Levi, and Z. Yan. Distributionally robust linear and discrete optimization with marginals. *To appear in Operations Research*, 2022.
- Y.-C. Chen and V. V. Mišić. Decision forest: A nonparametric approach to modeling irrational choice. *To appear in Management Science*, 2022.
- C. Daganzo. *Multinomial probit: the theory and its application to demand forecasting*. Academic Press, New York, 1979.
- E. W. de Bekker-Grob, J. Veldwijk, M. Jonker, B. Donkers, J. Huisman, S. Buis, J. Swait, E. Lancsar, C. L. Witteman, G. Bonsel, et al. The impact of vaccination and patient characteristics on influenza vaccination uptake of elderly people: a discrete choice experiment. *Vaccine*, 36(11):1467–1476, 2018.
- G. Feng, X. Li, and Z. Wang. Technical note—on the relation between several discrete choice models. *Operations Research*, 65(6):1429–1731, 2017.
- D. Fudenberg, R. Iijima, and T. Strzalecki. Stochastic choice and revealed perturbed utility. *Econometrica*, 83(6):2371–2409, 2015.
- J. Hofbauer and W. H. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, 2002.
- S. Jagabathula and P. Rusmevichientong. The limit of rationality in choice modeling: Formulation, computation, and implications. *Management Science*, 65(5):2196–2215, 2019.
- R. D. Luce. *Individual choice behavior*. John Wiley, 1959.
- D. McFadden. Modelling the choice of residential location. *Spatial Interaction Theory and Residential Location*, pages 75–96, 1978.
- D. McFadden. Econometric models for probabilistic choice among products. *Journal of Business*, pages S13–S29, 1980.
- D. McFadden. The choice theory approach to market research. *Marketing science*, 5(4):275–297, 1986.
- D. McFadden and K. Train. Mixed mnl models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470, 2000.
- 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.
- K. Natarajan, M. Song, and C.-P. Teo. Persistency model and its applications in choice modeling. *Management Science*, 55(3):453–469, 2009.

- K. Talluri and G. Van Ryzin. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1):15–33, 2004.
- S. Wang, Q. Wang, and J. Zhao. Deep neural networks for choice analysis: Extracting complete economic information for interpretation. *Transportation Research Part C*, 118, 2020.
- Z. Yan, K. Natarajan, C.-P. Teo, and C. Cheng. A representative consumer model in data-driven multi-product pricing optimization. *To appear in Management Science*, 2022.