Robust Maximum Capture Facility Location under Random Utility Maximization Models

Tien Thanh Dam[1], Thuy Anh Ta[2], and Tien Mai[3]

[1]*ORLab, Faculty of Computer Science, Phenikaa University, Yen Nghia, Ha Dong, Hanoi, VietNam, anh.tathuy@phenikaa-uni.edu.vn*

[2]*ORLab, Faculty of Computer Science, Phenikaa University, Yen Nghia, Ha Dong, Hanoi, VietNam, thanh.damtien@phenikaa-uni.edu.vn*

[3]*School of Computing and Information Systems, Singapore Management University, 80 Stamford Rd, Singapore 178902, atmai@smu.edu.sg*

February 14, 2023

Robust Maximum Capture Facility Location under Random Utility Maximization Models

Abstract

We study a robust version of the maximum capture facility location problem in a competitive market, assuming that each customer chooses among all available facilities according to a random utility maximization (RUM) model. We employ the generalized extreme value (GEV) family of models and assume that the parameters of the RUM model are not given exactly but lie in convex uncertainty sets. The problem is to locate new facilities to maximize the worst-case captured user demand. We show that, interestingly, our robust model preserves the monotonicity and submodularity from its deterministic counterpart, implying that a simple greedy heuristic can guarantee a $(1-1/e)$ approximation solution. We further show the concavity of the objective function under the classical multinomial logit (MNL) model, suggesting that an outer-approximation algorithm can be used to solve the robust model under MNL to optimality. We conduct experiments comparing our robust method to other deterministic and sampling approaches, using instances from different discrete choice models. Our results clearly demonstrate the advantages of our robust model in protecting the decision-maker from worst-case scenarios.

Keywords: Facilities planning and design, maximum capture, random utility maximization, robust optimization, local search, outer-approximation

1 Introduction

Facility location is an active research area in operations research and has been attracting researchers for decades. Facility location problems play important roles in many decision-making tasks such as installation of new retail or service facilities in a market, launching new products to the market, or developing optimal customer segmentation policies. In facility location, a firm aims at selecting a set of locations to locate their facilities to maximize a profit or minimize a cost function. In this context, to make good decisions, one may need to build a good model to predict customers' behavior with respect to each possible facility location decision. The random utility maximization (RUM) discrete choice framework (McFadden, 1978) has become useful in the context due to its flexibility in capturing human behavior. To the best of our knowledge, existing works on facility location under RUM all assume that the parameters of the RUM model are known with certainty and ignore any uncertainty associated with the estimates, with a tacit understanding that the parameters have to be estimated in practice. Such an estimation can cause errors and the decision-maker needs to cope with the fact that the estimates of the choice parameters can significantly deviate from their true values. Ignoring such estimation errors would lead to bad decisions, as shown in several studies in the robust optimization literature (see Bertsimas et al., 2011, for instance). In this paper, we address this uncertainty issue by studying a robust version of the facility location problem under RUM.

We consider the problem of how to locate new facilities in a competitive market such that the captured demand of users is maximized, assuming that each individual chooses among all available facilities according to a RUM model. This problem is called as the maximum capture problem (MCP) (Benati and Hansen, 2002a). We formulate and solve the MCP under uncertainty in a robust manner. That is, we assume that customers' behavior is driven by the well-known generalized extreme value (GEV) family of models, but the parameters of the RUM model cannot be determined with certainty and belong to some uncertainty sets. These uncertainty sets can represent a partial knowledge of the decision-makers about the RUM model and can be inferred from data. The goal here is to maximize the worst-case expected captured customer demand when the RUM parameters vary in the uncertainty sets. We will study theoretical properties and develop algorithms for the robust MCP under any GEV model and, in particular, the robust MCP under the popular multinomial logit (MNL) (Train, 2003).

Before presenting our contributions in detail, we note that, when mentioning the "GEV" model, we refer to any RUM (or discrete choice) model in the GEV family. This family covers most of the discrete choice (or RUM) models in the literature (Train, 2003).

Our contributions: We study a robust version of the MCP under a GEV model, assuming that the parameters of the GEV model are not known with certainty but can take any values in some uncertainty sets, the uncertainty sets are customer-wise independent, and the objective is to maximize the worst-case expected captured customer demand. We will show that, under our uncertainty settings, the inner minimization problem can be solved by convex optimization. We then leverage the properties of the GEV family to show that the worst-case objective function is monotonic and submodular for any GEV model, noting that the monotonicity and submodularity have been shown for the deterministic MCP (Dam et al., 2021) and in this work, we show that the robust model preserves both properties. Here, it is important to note that a robust submodular maximization problem is generally inapproximable, i.e. there is no polynomial-time algorithm that can guarantee a positive fraction of the optimal value, unless $\mathsf{P} = \mathsf{NP}$ (Krause et al., 2008). Our results, however, show that, in the context of the MCP, a simple polynomial-time greedy algorithm can achieve $(1 - 1/e)$ approximation solutions.

The monotonicity and submodularity of the robust problem imply that the robust MCP, under a cardinality constraint, always admits a $(1 - 1/e)$ approximation algorithm (Nemhauser et al., 1978). That is, we can simply start from an empty set and iteratively select locations, one at a time, taking at each step the location that increases the worst-case objective function the most, until the maximum capacity is reached. A solution from this simple procedure will yield an objective value being at least $(1 - 1/e)$ times the optimal value of the robust problem. We then further adapt the local search procedure proposed by Dam et al. (2021) to efficiently solve the robust MCP under GEV models. Our results generally hold for any RUM model in the GEV family, and under any convex uncertainty sets. We further consider the robust MCP under the MNL model and show that, under the assumption that the uncertainty sets are independent over customer zones, the robustness preserves the concavity of the relaxation of the worst-case objective function, implying that an outer-approximation algorithm can be used to exactly solve the MCP under MNL. A multicut outer-approximation algorithm is then presented for this purpose. We finally conduct experiments based on the MNL and nested logit (two most popular RUM models) to demonstrate the advantages of our robust model in protecting decision-makers from worst-case scenarios, as compared to other deterministic and sampling-based baselines.

Literature review: The GEV family (McFadden, 1978, Train, 2003) covers many popular RUM models in the demand modeling and operations research literature. In this family, the simplest and most popular member is the MNL (McFadden, 1978) and it is well-known that the MNL model retains the independence from irrelevant alternatives (IIA) property, which is often regarded as a limitation of the MNL model. The literature has seen several other GEV models that relax this property and provide flexibility in modeling the correlation between choice alternatives. Some examples are the nested logit (Ben-Akiva and Lerman, 1985, Ben-Akiva, 1973), the cross-nested logit (Vovsha and Bekhor, 1998), and network GEV (Daly and Bierlaire, 2006, Mai et al., 2017a) models. The cross-nested and network GEV models are considered as being fully flexible as they can approximate any RUM model (Fosgerau et al., 2013). It is worth noting that in the context of descriptive representation (i.e., modeling human behavior) the MNL and nested logit models are found the most empirically applicable among existing RUM models, but in prescriptive optimization (i.e., decision-making), apart from the MNL, the use of other GEV models is limited due to their complicated structures. Besides the GEV family, we note that the mixed logit model (MMNL) (McFadden and Train, 2000) is popular due to its flexibility in capturing utility correlations. In the context of the MCP, the use of the MMNL however yields the same problem structure as the one from the MNL model (Dam et al., 2021, Mai and Lodi, 2020).

Besides the GEV family and MMNL model, the literature has seen other discrete choice models that would be useful for people demand modeling. For example, Blanchet et al. (2016) propose a *Markov-chain choice model* under which the substitution from one product to another is modeled as a state transition in a Markov chain. Gallego and Wang (2019) propose a *threshold utility model* where consumers buy any product whose net utility exceeds a non-negative, product-specific threshold. Farias et al. (2013) propose a *non-parametric choice model* where the choice probabilities are modelled based on a distribution over all permutations of the choice alternative preferences. (Mishra et al., 2014) develop the *marginal distribution choice* (MDC) models based on the assumption that the distribution of the random utilities are not given exactly but belong to an ambiguity set with marginal information. With techniques from distributionally robust optimization, the estimation of MDC models in decision-making can be handled by convex optimization (Yan et al., 2022). The difference between the MDC and our robust models lies in the sources of uncertainty each model captures. More precisely, under the RUM principle, the utility of an alternative j is modelled as $u_j = v_j + \xi_j$, where v_j is deterministic and ξ_j is assumed to follow a given distribution. While our model captures uncertainties associated with the deterministic term v_j, the MDC model assumes that the distribution of ξ_j is ambiguous. In fact, our robust model is based on the well-studied GEV family while the MDC is not, and our robust model is capable of naturally capturing uncertainties caused by, for instance, estimation errors, limited data, or lack of information about facilities and customers when specifying the utility function. The MDC model, to the best of our knowledge, is limited in this aspect. Moreover, while the application of GEV models in choice-based optimization has had a lot of success, the application of the MDC model has just received attention recently and is still limited.

In the context of facility location under RUM, there are a number of works making use of the MNL model to capture customers' demand. For example, Benati and Hansen (2002a) formulate the first MCP under MNL and propose methods based on mixed-integer linear programming (MILP) and variable neighborhood search (VNS). Afterward, some alternative MILP models have been proposed by Zhang et al. (2012) and Haase (2009). Haase and Müller (2013) then provide a comparison of existing MILP models and conclude that the MILP from Haase (2009) gives the best performance. Freire et al. (2015) strengthen the MILP reformulation of (Haase, 2009) using a branch-and-bound algorithm with some tight inequalities. Ljubić and Moreno (2018) propose a branch-and-cut method combining two types of cutting planes, namely, outer-approximation and submodular cuts, and Mai and Lodi (2020) propose a multicut outer-approximation algorithm to efficiently solve large instances. All the above papers employ the MNL or MMNL models, leveraging the linear fractional structures of the objective functions to develop solution algorithms. Recently, Dam et al. (2021) make the first effort to bring general GEV models into the MCP. In this work, we show that the objective function of the MCP under GEV is monotonic submodular, leading to the development of an efficient local search algorithm with a performance guarantee.

Our work belongs to the general literature of robust facility location where the problem is to open new (or reopen available) facilities under uncertainty. For example, Averbakh and Berman (1997) considers a minimax-regret formulation of the weighted p-center problem and shows that the problem can be solved by solving a sequence of deterministic p-center problems. Subsequently, Averbakh and Berman (2000a) study a minimax-regret 1-center problem with uncertain node weights and edge lengths and Averbakh and Berman (2000b) consider a minimax-regret 1-median problem with interval uncertainty of the nodal demands. Some distributionally robust facility location models have been studied, for instance, Lu et al. (2015) study a distributionally robust reliable facility location problem by optimizing over worst-case distributions based on a given distribution of random facility disruptions, Liu et al. (2019) study a distributionally robust model for optimally locating emergency medical service stations under demand uncertainty, and Hugosson and Algers (2012) study a facility location problem where the distribution of customer demand is dependent on location decisions. We refer the reader to Snyder (2006) for a review. Our work differs from the above papers as we employ RUM models, thus customers' behavior is captured by a probabilistic discrete choice model.

Our work relates to the rich literature of robust optimization (e.g. Ben-Tal and Nemirovski, 1998, 1999, Bertsimas et al., 2004, EI-Ghaoui and Lebret, 1997) and distributionally robust optimization (Rahimian and Mehrotra, 2019, Wiesemann et al., 2014). Most of the works in the robust/distributionally robust optimization literature are focused on linear or general convex programs, thus the existing methods do not apply to our robust MCP. It is worth mentioning some robust models in the assortment and/or pricing optimization literature, where RUM models are made use to capture customers' behavior. Some examples are Rusmevichientong and Topaloglu (2012), Chen et al. (2019), and Mai and Jaillet (2019), noting that, in the context of assortment and/or pricing optimization, the objective function is often based on one fraction, instead of a sum of fractions as in the context of the MCP. Moreover, the objective function of an assortment problem is typically not submodular, even when the choice model is MNL. Our robust MCP model under MNL closely relates to the robust fractional 0-1 program studied in Mehmanchi et al. (2020) but differs by the fact that our approach works with any convex uncertainty sets while the approaches of Mehmanchi et al. (2020) rely on a particular uncertainty structure introduced by Bertsimas et al. (2004).

Paper outline: We organize the paper as follows. Section 2 provides a background of the deterministic MCP under RUM models. Section 3 presents our main results for the robust MCP. Section 4 describes our algorithms used to solve the robust problems. Section 5 provides some numerical experiments, and Section 6 concludes. Missing proofs and additional experiments are provided in the appendix.

Notation: Boldface characters represent matrices (or vectors), and a_i denotes the i-th element of vector $\mathbf{a}$. We use $[m]$, for any $m \in \mathbb{N}$, to denote the set $\{1, \ldots, m\}$.

2 Background: Deterministic MCP under RUM

In this section, we first revisit the RUM framework and then describe the deterministic MCP under RUM models.

2.1 The RUM Framework and the GEV family

The RUM framework (McFadden, 1978) consists of prominent discrete choice models for modeling human behavior when faced with a set of discrete choice alternatives. Under the RUM principle, the customers are assumed to associate a random utility u_j with each choice alternative j in a given choice set of available alternatives S. The additive RUM (Fosgerau et al., 2013, McFadden, 1978) assumes that each random utility is a sum of two parts $u_j = v_j + \xi_j$, where the term v_j is deterministic and can include values representing the characteristics of the choice alternative j and/or the customers, and ξ_j is random and unknown to the analyst. Assumptions then can be made for the random terms ξ_j, leading to different discrete choice models, i.e., the MNL model relies on the assumption that ξ_j are i.i.d *Extreme Value type I*. The RUM principle then assumes that a choice is made by maximizing the random utilities, thus the probability that an alternative j is selected can be computed as $P(u_j \geq u_k, \forall k \in S)$.

Among RUM models, the MNL model is the simplest one and it is well known that this model fails to capture the correlation between choice alternatives, due to the i.i.d. assumption imposed on the random terms. This drawback is called as the IIA property (Train, 2003). Efforts have been made to relax this property, leading to more advanced choice models with flexible correlation structures such as the nested logit or cross-nested logit models. Among them, the GEV family (Daly and Bierlaire, 2006, McFadden, 1981) is regarded as one of the most general families of discrete choice models in the econometrics and operations research literature. We describe this family of models in the following.

Assume that the choice set contains m alternatives indexed as $\{1, \ldots, m\}$ and let $\{v_1, \ldots, v_m\}$ be the vector of deterministic utilities of the m alternatives. A GEV model can be represented by a choice probability generating function (CPGF) $G(\mathbf{Y})$ (Fosgerau et al., 2013, McFadden, 1981), where $\mathbf{Y}$ is a vector of size m with entries $Y_j = e^{v_j}$, $j \in [m]$. Given $j_1, \ldots, j_k \in [m]$, let $\partial G_{j_1 \ldots j_k}$, be the mixed partial derivatives of G with respect to $Y_{j_1}, \ldots, Y_{j_k}$. The following basic properties hold for any CPGF $G(\cdot)$ in the GEV family (McFadden, 1978).

Remark 1 (Basic properties of GEV's CPGF) *The following properties hold for any GEV probability generating function.*

(i) $G(\boldsymbol{Y}) \geq 0, \ \forall \boldsymbol{Y} \in \mathbb{R}^m_+$,

(ii) $G(\boldsymbol{Y})$ *is homogeneous of degree one, i.e.,* $G(\lambda \boldsymbol{Y}) = \lambda G(\boldsymbol{Y})$, *for any scalar* $\lambda > 0$

(iii) $G(\boldsymbol{Y}) \to \infty$ *if* $Y_j \to \infty$, *for any* $j \in [m]$

(iv) Given $j_1, \ldots, j_k \in [m]$ *distinct from each other,* $\partial G_{j_1,\ldots,j_k}(\boldsymbol{Y}) > 0$ *if* k *is odd, and* ≤ 0 *if* k *is even.*

The above properties are standard for the GEV family. The economic intuition behind these properties is however limited (Section 4.6 in Train, 2003), especially for Property (iv). In fact, these properties (or conditions) are to ensure that the corresponding choice model is consistent with the RUM principle and would be useful to design a new RUM model. To support our later exposition, we present some additional properties of the GEV family in Proposition 1 below. These new properties can be verified easily using the basic properties introduced above.

Proposition 1 (Some additional properties of GEV's CPGF) *The following properties hold for any CPGF under the GEV family:*

(i) $G(\boldsymbol{Y}) = \sum_{j \in [m]} Y_j \partial G_j(\boldsymbol{Y})$

(ii) $\partial G_j(\lambda \boldsymbol{Y}) = \partial G_j(\boldsymbol{Y})$ *for any scalar* $\lambda > 0$

(iii) $\sum_{k \in [m]} Y_k \partial G_{jk}(\boldsymbol{Y}) = 0, \ \forall j \in [m]$.

Proof. Property (i) can be obtained by taking the derivatives with respect to λ on both sides of $G(\lambda\mathbf{Y}) = \lambda G(\mathbf{Y})$ (Property (ii) of Remark 1) to have $G(\mathbf{Y}) = \sum_{j\in[m]} Y_j \partial G_j(\lambda\mathbf{Y})$. We then let $\lambda = 1$ to obtain the desired equality.

For Property (ii), we take derivatives on both sides of the equality $G(\lambda\mathbf{Y}) = \lambda G(\mathbf{Y})$ with respect to Y_j to have

$$\lambda\partial G_j(\lambda\mathbf{Y}) = \lambda\partial G_j(\mathbf{Y}).$$

By removing λ from both sides of the equality, we can obtain the desired equality $\partial G_j(\lambda\mathbf{Y}) = \partial G_j(\mathbf{Y})$.

For Property (iii), we further take the derivatives with respect to λ on both sides of (ii), we get $\sum_{k\in[m]} Y_k \partial G_{jk}(\lambda\mathbf{Y}) = 0$. By letting $\lambda = 1$, we obtain (iii). ■

Under a GEV model specified by a CPGF $G(\mathbf{Y})$, the choice probabilities are given as

$$P(j|\mathbf{Y}, G) = \frac{Y_i \partial G_i(\mathbf{Y})}{G(\mathbf{Y})}.$$

If the choice model is MNL, the CPGF has a linear form $G(\mathbf{Y}) = \sum_{j\in[m]} Y_j$ and the corresponding choice probabilities are given by a linear fraction $P(j|\mathbf{Y}, G) = Y_j/(\sum_{j\in[m]} Y_j)$. On the other hand, if the choice model is a nested logit model, the choice set can be partitioned into L nests, which are disjoint subsets of the alternatives. Let denote by $n_1, \ldots, n_L$ the L nests. The corresponding CPGF has a nonlinear form as $G(\mathbf{Y}) = \sum_{l\in L} \left(\sum_{j\in n_l} Y_j^{\mu_l}\right)^{1/\mu_l}$, where $\mu_l \geq 1$, $l \in [L]$, are the parameters of the nested logit model. The choice probabilities have the more complicated form

$$P(j|\mathbf{Y}, G) = \frac{\left(\sum_{j'\in n_l} Y_{j'}^{\mu_l}\right)^{1/\mu_l}}{\sum_{l\in[L]} \left(\sum_{j'\in n_l} Y_{j'}^{\mu_l}\right)^{1/\mu_l}} \frac{Y_j^{\mu_l}}{\sum_{j'\in n_l} Y_{j'}^{\mu_l}}, \ \forall l \in [L], j \in n_l.$$

In a more general setting, a CPGF can be represented by a rooted network (Daly and Bierlaire, 2006), for which the choice probabilities may have no closed-form and need to be computed by recursion or dynamic programming (Mai et al., 2017a).

2.2 The Deterministic MCP

In the context of the MCP under a RUM model, a firm would like to select some locations from a set of available locations to set up new facilities, assuming that there exist facilities from the competitor in the market. The firm then aims to maximize an expected market share achieved by attracting customers to the new facilities. We suppose that there are m available locations and we let $[m] = \{1, 2, \ldots, m\}$ be the set of locations. Let I be the set of geographical zones where customers are located and q_i be the number of customers in zone $i \in I$. For each customer zone i, let v_{ij} be the corresponding deterministic utility of location $j \in [m]$. These utility values can be inferred by estimating the discrete choice model using observational data of how customers made decisions. For each customer zone $i \in I$, the corresponding discrete choice model can be represented by a CPGF $G^i(\mathbf{Y}^i)$, where $\mathbf{Y}^i$ is a vector of size m with entries $Y^i_j = e^{v_{ij}}$. The choice probability of a location $j \in [m]$ can then be computed as

$$P(j|\mathbf{Y}^i) = \frac{Y_j \partial G^i_j(\mathbf{Y}^i)}{U^i + G^i(\mathbf{Y}^i)},$$

where U^i is the total utility of the competitor for zone $i \in I$. Such an utility is analogous to a non-purchase utility used in the context of assortment optimization (Talluri and Van Ryzin, 2004). Note that we can write

$$P(j|\mathbf{Y}^i) = \frac{\frac{Y_j}{U^i}\partial G^i_j(\mathbf{Y}^i)}{1 + \frac{1}{U^i} G^i(\mathbf{Y}^i)} \overset{(a)}{=} \frac{\frac{Y_j}{U^i}\partial G^i_j\left(\frac{\mathbf{Y}^i}{U^i}\right)}{1 + G^i\left(\frac{\mathbf{Y}^i}{U^i}\right)},$$

where *(a)* is due to Property (ii) of Proposition 1. Thus, $P(j|\mathbf{Y}^i)$ are the choice probabilities given by the set of utilities $v'_{ij} = v_{ij} - \ln U^i$. In other words, one can subtract the utilities v_{ij} by $\ln U^i$ to force the competitor's utilities to be 1, without any loss of generality. Therefore, from now on, for the sake of simplicity, we assume $U^i = 1$, for all $i \in I$.

In the context of the MCP, we are interested in choosing a subset of locations $S \subset [m]$ to locate new facilities. Hence, the conditional choice probabilities of choosing a location $j \in S$ can be written as

$$P(j|\mathbf{Y}^i, S) = \frac{Y^i_j \partial G^i_j(\mathbf{Y}^i|S)}{1 + G^i(\mathbf{Y}^i|S)}, \ \forall j \in S,$$

where the $G^i(\mathbf{Y}^i|S)$ is defined as $G^i(\mathbf{Y}^i|S) = G^i(\overline{\mathbf{Y}}^i)$, where $\overline{\mathbf{Y}}^i$ is a vector of size m with entries $\overline{Y}^i_j = Y^i_j$ if $j \in S$ and $\overline{Y}^i_j = 0$ otherwise. This can be interpreted as if a location j is not selected, then its utility should be very low, i.e., $v_{ij} = -\infty$, then $Y^i_j = e^{v_{ij}} = 0$. The deterministic MCP under GEV models specified by CPGFs $G^i(\mathbf{Y}^i)$, $i \in I$, can be defined as

$$\max_{S \in \mathcal{K}} \left\{ f(S) = \sum_{i \in I} q_i \sum_{j \in S} P(j|\mathbf{Y}^i, S) \right\}, \tag{MCP}$$

where $\mathcal{K}$ is the set of feasible solutions. Under a conventional cardinality constraint $|S| \le C$, $\mathcal{K}$ can be defined as $\mathcal{K} = \{S \subset [m] | \, |S| \le C\}$, for a given constant C such that $1 \le C \le m$. (MCP) is generally NP-hard, even under the MNL choice model. However, it is possible to obtain $(1 - 1/e)$ approximation solutions using a greedy local search algorithm (Dam et al., 2021).

Under the MNL model, the MCP can be formulated as a linear fractional program as

$$\max_{S \in \mathcal{K}} \left\{ f(S) = \sum_{i \in I} q_i \sum_{i \in I} \frac{\sum_{j \in S} Y^i_j}{1 + \sum_{j \in S} Y^i_j} \right\}.$$

The fractional structure allows to formulate the MCP under MNL as a MILP, thus a MILP solver, e.g., CPLEX or GUROBI, can be used (Freire et al., 2015, Haase, 2009, Haase and Müller, 2013). It is well-known that the objective function $f(S)$ is submodular, leading to some approaches based on submodular cuts or local search heuristic (Dam et al., 2021, Ljubić and Moreno, 2018). Under more general GEV models, e.g., the nested logit or cross-nested logit models (Train, 2003), the objective function becomes much more complicated. In fact, it would be possible to reformulate the MCP under a GEV model as a MILP with submodular cuts (Benati and Hansen, 2002b, Ljubić and Moreno, 2018, Nemhauser et al., 1978). However, such a MILP approach involves an exponential number of constraints, thus would be not tractable to solve. In addition, under a general GEV model, the choice probabilities would not be expressed in a closed form (Mai et al., 2017b), making the computation of $f(S)$ intractable and the corresponding MILP intractable as well.

It is relevant to connect the MCP formulation to the context of assortment optimization under discrete choice models. In fact, (MCP) under MNL shares a close structure with assortment optimization problems under the mixed logit model (Rusmevichientong et al., 2014). However, (MCP) is more tractable, in the sense it can be written as a binary program with a convex objective function, allowing for some convex optimization techniques (e.g. the outer-approximation algorithm) to be applied , while it is not the case in assortment optimization. Moreover, under the GEV family, the assortment problem becomes even more challenging to solve and, to the best of our knowledge, there is no algorithm with performance guarantees for such a problem. On the other hand, one can achieve $(1-1/e)$ approximation solutions to the MCP by just using a simple greedy heuristic (Dam et al., 2021).

There would be relevant situations where the algorithms developed in this paper can apply to assortment optimization problems. For example, one can think of a situation where a seller needs to select a set of products to sell together with a competitor, and the objective is to maximize the expected number of customers that come to purchase their products, instead of maximizing an expected revenue in a conventional assortment optimization problem. In this context, the assortment optimization model shares the same structure with (MCP) and the methods developed in this paper can apply.

3 Robust MCP

We first consider the robust MCP under any GEV model and show that the robust model preserves the monotonicity and submodularity from the deterministic one, which will ensure that a simple greedy heuristic will always return a $(1-1/e)$ approximation solution. Moreover, we show that under the MNL model, the robust model preserves the concavity, implying that an outer-approximation algorithm could be used to efficiently solve the robust MCP to optimality.

3.1 Robust MCP under GEV models

In our uncertainty setting, it is assumed that the vectors of utilities are not known with certainty but belong to some uncertainty sets. That is, we assume that for each customer zone $i \in I$, the corresponding deterministic utilities $\mathbf{v}^i = \{v_{ij} |\ j \in [m]\}$ can vary in an uncertainty set $\mathcal{V}_i$ and these uncertainty sets are independent across $i \in I$. This setting is natural in the context, in the sense that there is one discrete choice model with choice utilities $\mathbf{v}^i$ for customers in each zone i, and these utilities are typically inferred from observations of how people in that zone made choices. Uncertainty sets could be constructed from this, leading to independent uncertainty sets over customer zones. Moreover, if we relax this assumption, i.e., $\mathcal{V}_i$ are no longer independent over $i \in I$, the adversary's minimization problem [1] will become much more difficult to solve, even under the classical MNL model. In contrast, under the assumption that the uncertainty sets are separable by zones, we will show later that the adversary's problem can be efficiently handled by convex optimization.

We further assume that $\mathcal{V}_i$ are convex and bounded for all $i \in I$. Convexity and boundedness are typical assumptions in the robust optimization literature (Bertsimas et al., 2011). For later investigations, we assume that the uncertainty set $\mathcal{V}_i$ can be defined by a set of constraints $\{g^i_t(\mathbf{v}^i) \leq 0;\ t = 1, \ldots, T\}$ where $g^i_t(\mathbf{v}^i)$ are convex functions in $\mathbf{v}^i$. When the choice parameters $\mathbf{v}^i$, $i \in I$, cannot be identified exactly, we are interested in the worst-case scenario. The robust version of the classical MCP then can be formulated as

$$\max_{S \in \mathcal{K}} \min_{\mathbf{v}^i \in \mathcal{V}_i} \left\{ f(S, \mathbf{V}) = \sum_{i \in I} q_i \sum_{j \in S} P(j|\mathbf{v}^i, S) \right\}, \tag{1}$$

where $P(j|\mathbf{v}^i, S)$ is the choice probability of location $j \in [m]$ given utilities $\mathbf{v}^i$ and set of locations $S \in \mathcal{K}$. Under a GEV choice model with CPGFs G^i, $i \in I$, we write the choice probabilities as

$$P(j|\mathbf{v}^i, S) = \frac{Y^i_j \partial G^i_j(\mathbf{Y}(\mathbf{v}^i)|S)}{1 + G^i(\mathbf{Y}(\mathbf{v}^i)|S)},$$

where $\mathbf{Y}(\mathbf{v}^i)$ is a vector of size m with entries $Y(\mathbf{v}^i)_j = e^{v_{ij}}$. The objective function can be further simplified as

$$\begin{aligned} f(S, \mathbf{V}) &= \sum_{i \in I} q_i \sum_{j \in S} P(j|\mathbf{v}^i, S) \\ &= \sum_{i \in I} q_i \frac{\sum_{j \in S} Y^i_j \partial G^i_j(\mathbf{Y}(\mathbf{v}^i)|S)}{1 + G^i(\mathbf{Y}(\mathbf{v}^i)|S)} \\ &\overset{(a)}{=} \sum_{i \in I} q_i - \sum_{i \in I} \frac{q_i}{1 + G^i(\mathbf{Y}(\mathbf{v}^i)|S)}, \end{aligned} \tag{2}$$

where (a) is due to Property (i) of Proposition 1, i.e., $\sum_{j \in S} Y^i_j \partial G^i_j(\mathbf{Y}(\mathbf{v}^i)|S) = G^i(\mathbf{Y}(\mathbf{v}^i)|S)$.

[1] When we say "adversary", we refer to the worst-case minimization problem, not the competitor in the market.

Thus, the robust problem can be reformulated as

$$\max_{S\in\mathcal{K}} \left\{ f^{\mathrm{WC}}(S) = \sum_{i\in I} q_i - \sum_{i\in I} \frac{q_i}{1+\min_{\mathbf{v}^i\in\mathcal{V}_i}\{G^i(\mathbf{Y}(\mathbf{v}^i)|S)\}} \right\}, \tag{RMCP}$$

where $f^{\mathrm{WC}}(S)$ is referred to as the worst-case objective function when the choice parameters of the GEV model vary in the uncertainty sets.

In (RMCP) we assume that there is no uncertainty associated with the specification of the competitor's utilities. In a more general setting, it is possible that the competitor's utilities are not known with certainty and would need to be taken into consideration in the robust model. Nevertheless, we will show, in the following, that such a general uncertainty structure can be converted into the same uncertainty structure in (RMCP) with shifted (convex) uncertainty sets. To facilitate our exposition, let us assume that the utility of the competitor is v_0^i for customer zone $i \in I$, and both $v_0^i, \mathbf{v}^i$ can vary in an uncertainty set $\mathcal{V}_i$. The robust problem then becomes:

$$\max_{S\in\mathcal{K}} \min_{\substack{(v_0^i,\mathbf{v}^i)\in\mathcal{V}_i\\ \forall i\in I}} \left\{ \sum_{i\in I} q_i \frac{\sum_{j\in[m]} Y_j^i(v_j^i)\partial G_j^i(\mathbf{Y}(\mathbf{v}^i)|S)}{e^{v_0^i} + G^i(\mathbf{Y}(\mathbf{v}^i)|S)} \right\}. \tag{3}$$

Proposition 2 shows that (3) can be converted equivalently into a robust problem of the same structure as (RMCP). The proof can be found in Appendix A.

Proposition 2 (RMCP) *is equivalent to*

$$\max_{S\in\mathcal{K}} \min_{\substack{\widetilde{\boldsymbol{v}}^i\in\widetilde{\mathcal{V}}_i\\ \forall i\in I}} \left\{ \sum_{i\in I} q_i \frac{\sum_{j\in[m]} Y_j^i(\widetilde{v}_j^i)\partial G_j^i(\mathbf{Y}(\widetilde{\boldsymbol{v}}^i)|S)}{1 + G^i(\mathbf{Y}(\widetilde{\boldsymbol{v}}^i)|S)} \right\}, \tag{4}$$

where $\widetilde{\mathcal{V}}_i = \left\{ \widetilde{\boldsymbol{v}}^i \in \mathbb{R}^m \middle| \ \exists \boldsymbol{v}^i \in \mathcal{V}_i \text{ s.t. } \widetilde{v}_j^i = v_j^i - v_0^i, \ \forall j \in [m] \right\}$.

It can be seen that $\widetilde{\mathcal{V}}_i$ is convex if $\mathcal{V}_i$ is convex. The inner minimization problem of (3) differs from the inner minimization of (RMCP) just by some additional linear constraints, i.e.,

$$\begin{aligned} \min \quad & \sum_{i\in I} q_i \frac{\sum_{j\in[m]} Y_j^i(\widetilde{v}_j^i)\partial G_j^i(\mathbf{Y}(\widetilde{\mathbf{v}}^i)|S)}{1 + G^i(\mathbf{Y}(\widetilde{\mathbf{v}}^i)|S)} & \\ \text{subject to} \quad & \widetilde{v}_j^i = v_j^i - v_0^i & \forall i \in I, j \in [m] \\ & (v_j^0, \mathbf{v}_j^i) \in \mathcal{V}_i & \forall i \in I. \end{aligned}$$

Thanks to Proposition 2, for the sake of simplicity, we will keep the assumption that the utilities of the competitor are deterministically equal to 1 throughout the rest of the paper.

We now explore some properties of the robust program in the following. Proposition 3 below first shows that the inner minimization problem (adversary's problem) can be solved efficiently via convex optimization. Before discussing the result, let us define a binary representation of

$f^{\text{WC}}(S)$ as follows. For any set $S \in \mathcal{K}$, let $\mathbf{x}^S \in \{0,1\}^m$ such that $x_j^S = 1$ if $j \in S$ and $x_j^S = 0$ otherwise. Then, we can write $G^i(\mathbf{Y}(\mathbf{v}^i)|S) = G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x}^S)$, where $\circ$ is the element-by-element operator.

Proposition 3 (Convexity of the adversary's minimization problems) *Given any $i \in I$ and $S \subset [m]$, $G^i(\boldsymbol{Y}(\boldsymbol{v}^i)|S)$ is strictly convex in $\boldsymbol{v}^i$.*

As shown in Dam et al. (2021), the objective function of the deterministic MCP is monotonic increasing, i.e., adding more facilities always yields better objective values. The following proposition shows that the robust model preserves the monotonicity.

Theorem 1 (Robustness Preserves the Monotonicity) *Given $S \subset [m]$, for any $j \in [m]\backslash S$, we have*

$$f^{\text{WC}}(S \cup \{j\}) > f^{\text{WC}}(S).$$

Proof. Let $\mathbf{x} \in \{0,1\}^m$ be the binary presentation of set $S \subset [m]$. We write the objective function as

$$f^{\text{WC}}(\mathbf{x}) = \sum_{i \in I} q_i - \sum_{i \in I} \frac{q_i}{1 + \min_{\mathbf{v}^i \in \mathcal{V}_i}\{G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})\}}$$

We need to prove that, for any $\mathbf{x} \in X$ and $j \in [m]$ such that $x_j = 0$, we have $f^{\text{WC}}(\mathbf{x} + \mathbf{e}^j) > f^{\text{WC}}(\mathbf{x})$, where e^j is a vector of size m with all zero elements except the j-th one which is equal to 1. To prove this, let us consider $G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})$. Taking the derivative of $G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})$ w.r.t. x_j we have

$$\frac{\partial G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})}{\partial x_j} = Y_j^i(\mathbf{v}^i) \partial G_j^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x}) \overset{(b)}{>} 0,$$

where the (strict) inequality (b) is because $Y_j^i(\mathbf{v}^i) = e^{v_{ij}} > 0$ and $\partial G_j^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x}) > 0$ (Property (iv) of Remark 1). Thus,

$$G^i(\mathbf{Y}(\mathbf{v}^i) \circ (\mathbf{x} + \mathbf{e}^j)) > G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x}),$$

and consequently,

$$\min_{\mathbf{v}^i \in \mathcal{V}_i}\{G^i(\mathbf{Y}(\mathbf{v}^i) \circ (\mathbf{x} + \mathbf{e}^j))\} > \min_{\mathbf{v}^i \in \mathcal{V}_i}\{G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})\}$$

which directly leads to the desired inequality $f^{\text{WC}}(\mathbf{x} + \mathbf{e}^j) > f^{\text{WC}}(\mathbf{x})$. ■

Similar to the deterministic case, the monotonicity implies that an optimal solution S^* to (RMCP) always reaches its maximum capacity, i.e., $|S^*| = C$.

It is known that the objective function of the deterministic MCP is submodular (Dam et al., 2021). Typically, a robust version of a monotonic submodular function is not submodular (Orlin et al., 2018). However, in the theorem below, we show that the inclusion of the adversary in our robust MCP preserves the submodularity.

Theorem 2 (Robustness Preserves the Submodularity) *$f^{\text{WC}}(S)$ is submodular, i.e., for two sets $A \subset B \subset [m]$ and for any $j \in [m]\backslash B$ we have*

$$f^{\text{WC}}(A \cup \{j\}) - f^{\text{WC}}(A) \geq f^{\text{WC}}(B \cup \{j\}) - f^{\text{WC}}(B).$$

In the following we provide essential steps to prove the submodularity result. We will employ the binary representation of $f^{\text{WC}}(S)$ to prove the claim. Let

$$\phi^i(\mathbf{x}) = \min_{\mathbf{v}^i \in \mathcal{V}_i} G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x}). \tag{5}$$

Since $G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})$ is strictly convex and differentiable, $\phi^i(\mathbf{x})$ is continuous and differentiable (Corollary 8.2 in Hogan, 1973). We first introduce the following lemma.

Lemma 1 (First and second order derivatives of $\phi^i(\mathbf{x})$) *Given $\boldsymbol{x} \in [0,1]^m$, let $\boldsymbol{v}^{i*}$ be an optimal solution to the problem $\min_{\boldsymbol{v}^i \in \mathcal{V}_i} G^i(\boldsymbol{Y}(\boldsymbol{v}^i) \circ \boldsymbol{x})$, then for any $j, k \in [m]$,*

$$\frac{\partial \phi^i(\boldsymbol{x})}{\partial x_j} = \frac{\partial G^i(\boldsymbol{Y}(\boldsymbol{v}^{i*}) \circ \boldsymbol{x})}{\partial x_j}$$
$$\frac{\partial^2 \phi^i(\boldsymbol{x})}{\partial x_j \partial x_k} = \frac{\partial^2 G^i(\boldsymbol{Y}(\boldsymbol{v}^{i*}) \circ \boldsymbol{x})}{\partial x_j \partial x_k}.$$

Note that $\min_{\mathbf{v}^i \in \mathcal{V}_i} G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})$ always yields a unique optimal solution due to the strict concavity of $G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})$ (Proposition 3). Thus, $\mathbf{v}^{i*}$ mentioned in Lemma 1 is always unique.

We need one more lemma to complete the proof of the submodularity result. The following lemma shows a monotonicity behavior of $\partial \phi^i(\mathbf{x})/\partial x_k$ as a function of $\mathbf{x}$.

Lemma 2 *Given $\boldsymbol{x} \in \{0,1\}^m$ and any $j, k \in [m]$ such that $x_j = x_k = 0$ and $j \neq k$, we have*

$$\frac{\partial \phi^i(\boldsymbol{x} + \boldsymbol{e}^j)}{\partial x_k} \leq \frac{\partial \phi^i(\boldsymbol{x})}{\partial x_k}.$$

Proof. We define $\psi(\mathbf{x}) = \partial \phi^i(\mathbf{x})/\partial x_k$. Taking the first-order derivative of $\psi(\mathbf{x})$ w.r.t. x_j we get

$$\begin{aligned}\frac{\partial \psi(\mathbf{x})}{\partial x_j} &= \frac{\partial^2 \phi^i(\mathbf{x})}{\partial x_k \partial x_j} \\ &\overset{(a)}{=} \frac{\partial^2 G^i(\mathbf{Y}(\mathbf{v}^{i*}) \circ \mathbf{x})}{\partial x_j \partial x_k} \\ &= Y_j(\mathbf{v}^{i*}) Y_k(\mathbf{v}^{i*}) \partial^2 G^i_{jk}(\mathbf{Y}(\mathbf{v}^{i*}) \circ \mathbf{x}) \overset{(b)}{\leq} 0\end{aligned} \tag{6}$$

where (a) is from Lemma 1 and (b) is due to Property (iv) of Remark 1. So, we have $\partial \psi(\mathbf{x})/\partial x_j \leq 0$, implying that $\psi(\mathbf{x})$ is monotonically decreasing in x_j. Thus, $\partial \phi^i(\mathbf{x} + \mathbf{e}^j)/\partial x_k \leq \partial \phi^i(\mathbf{x})/\partial x_k$, as desired. ■

We are now ready for the main proof of Theorem 2.

Proof of Theorem 2. From Theorem 1, we have

$$\phi^i(\mathbf{x}+\mathbf{e}^j) \geq \phi^i(\mathbf{x}) \tag{7}$$

$$\phi^i(\mathbf{x}+\mathbf{e}^j+\mathbf{e}^k) \geq \phi^i(\mathbf{x}+\mathbf{e}^k). \tag{8}$$

Moreover, from Lemma 2 we see that, for any $\mathbf{x} \in \{0,1\}^m$ and any $j,k \in [m]$ such that $x_j = x_k = 0$, function $\psi(\mathbf{x}) = \phi^i(\mathbf{x}+\mathbf{e}^j) - \phi^i(\mathbf{x})$ is monotonically decreasing in x_k. Thus,

$$\phi^i(\mathbf{x}+\mathbf{e}^j) - \phi^i(\mathbf{x}) \geq \phi^i(\mathbf{x}+\mathbf{e}^j+\mathbf{e}^k) - \phi^i(\mathbf{x}+\mathbf{e}^k) \geq 0. \tag{9}$$

Moreover, from (7) and (8), we have

$$(1+\phi^i(\mathbf{x}+\mathbf{e}^j))(1+\phi^i(\mathbf{x})) \leq (1+\phi^i(\mathbf{x}+\mathbf{e}^j+\mathbf{e}^k))(1+\phi^i(\mathbf{x}+\mathbf{e}^k)). \tag{10}$$

Combine (9) and (10) we get

$$\begin{aligned}
&\frac{\phi^i(\mathbf{x}+\mathbf{e}^j) - \phi^i(\mathbf{x})}{(1+\phi^i(\mathbf{x}+\mathbf{e}^j))(1+\phi^i(\mathbf{x}))} \geq \frac{\phi^i(\mathbf{x}+\mathbf{e}^j+\mathbf{e}^k) - \phi^i(\mathbf{x}+\mathbf{e}^k)}{(1+\phi^i(\mathbf{x}+\mathbf{e}^j+\mathbf{e}^k))(1+\phi^i(\mathbf{x}+\mathbf{e}^k))} \\
\Leftrightarrow\ &\frac{1}{1+\phi^i(\mathbf{x})} - \frac{1}{1+\phi^i(\mathbf{x}+\mathbf{e}^j)} \geq \frac{1}{1+\phi^i(\mathbf{x}+\mathbf{e}^k)} - \frac{1}{1+\phi^i(\mathbf{x}+\mathbf{e}^j+\mathbf{e}^k)}.
\end{aligned} \tag{11}$$

Now we note that

$$f^{\text{WC}}(\mathbf{x}) = \sum_{i\in I} q_i - \sum_{i\in I} \frac{q_i}{1+\phi^i(\mathbf{x})}.$$

Thus, from (11) we have

$$f^{\text{WC}}(\mathbf{x}+\mathbf{e}^j) - f^{\text{WC}}(\mathbf{x}) \geq f^{\text{WC}}(\mathbf{x}+\mathbf{e}^j+\mathbf{e}^k) - f^{\text{WC}}(\mathbf{x}+\mathbf{e}^j). \tag{12}$$

Now, given $A \subset B \subset [m]$, let $\mathbf{x}^A$ and $\mathbf{x}^B$ be the binary representations of A, B, respectively. From (12) we have, for any $j \in [m]\backslash B$

$$\begin{aligned}
f^{\text{WC}}(\mathbf{x}^A+\mathbf{e}^j) - f^{\text{WC}}(\mathbf{x}^A) &\geq f^{\text{WC}}\left(\mathbf{x}^A+\mathbf{e}^j+\sum_{k\in B\backslash A}\mathbf{e}^k\right) - f^{\text{WC}}\left(\mathbf{x}^A+\sum_{k\in B\backslash A}\mathbf{e}^k\right) \\
&= f^{\text{WC}}(\mathbf{x}^B+\mathbf{e}^j) - f^{\text{WC}}(\mathbf{x}^B),
\end{aligned}$$

which is equivalent to $f^{\text{WC}}(A\cup\{j\}) - f^{\text{WC}}(A) \geq f^{\text{WC}}(B\cup\{j\}) - f^{\text{WC}}(B)$, implying the submodularity as desired. ■

The submodularity and monotonicity imply that, under a cardinality constraint $|S| \leq C$, a greedy heuristic can guarantee a $(1-1/e)$ approximation solution to the robust problem (Corollary 1 below) (Nemhauser et al., 1978). Such a greedy heuristic can start from an empty set and keep adding locations, one at a time, taking at each step a location that increases the worst-case objective function $f^{\text{WC}}(S)$ the most. The algorithm stops when the maximum capacity is reached, i.e., $|S| = C$. This greedy algorithm runs in $\mathcal{O}(mC)$.

Corollary 1 (Performance guarantee for a greedy heuristic) *For the robust MCP under a cardinality constraint $|S| \leq C$, there exists a greedy heuristic that always returns a solution S^* such that*

$$f^{\text{WC}}(S^*) \geq (1 - 1/e) \max_{|S| \leq C} \{f^{\text{WC}}(S)\}.$$

If the uncertainty set is rectangular, i.e., each v^i_j can freely deviate within an interval, the following proposition shows that the robust MCP can be further converted into a deterministic MCP.

Proposition 4 (Rectangular uncertainty sets) *If the uncertainty sets are rectangular, i.e., $\mathcal{V}_i = \{\boldsymbol{v}^i |\ \underline{\boldsymbol{v}}^i \leq \boldsymbol{v}^i \leq \overline{\boldsymbol{v}}^i\}$, for all $i \in I$, then the robust problem* (RMCP) *is equivalent to*

$$\max_{S \in \mathcal{K}} \left\{ \sum_{i \in I} q_i - \sum_{i \in I} \frac{q_i}{1 + G^i(\mathbf{Y}(\underline{\boldsymbol{v}}^i)|S)} \right\},$$

which is a deterministic MCP with parameters $\boldsymbol{v}^i = \underline{\boldsymbol{v}}^i$, $\forall i \in I$.

Proof. We will simply show that $G^i(\mathbf{Y}(\mathbf{v}^i)|S)$ is minimized over $\mathbf{v}^i \in \mathcal{V}_i$ at $\mathbf{v}^i = \underline{\mathbf{v}}^i$. To prove this, we consider the equivalent binary representation $G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x}^S)$ and take its derivatives w.r.t. v^i_j, for any $j \in [m]$, to have

$$\frac{G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x}^S)}{\partial v^i_j} = Y_j(\mathbf{v}^i) x_j \partial G^i_j(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x}^S),$$

and see that the right hand side is non-negative, because $Y_j(\mathbf{v}^i) \geq 0$, $x_j \geq 0$ and $\partial G^i_j(\mathbf{Y}(\mathbf{v}^i)) \geq 0$, where the later is due to Property (iv) of Remark 1. So, $G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x}^S)$ is monotonically increasing in $\mathbf{v}^i$. Thus,

$$G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x}^S) \geq G^i(\mathbf{Y}(\underline{\mathbf{v}}^i) \circ \mathbf{x}^S),\ \forall \mathbf{v}^i \in \mathcal{V}_i,$$

implying that $G^i(\mathbf{Y}(\mathbf{v}^i)|S)$ is minimized at $\mathbf{v}^i = \underline{\mathbf{v}}^i$ as desired. ■

3.2 Robust MCP under MNL

Under the MNL choice model, the deterministic MCP has a concave objective function (Benati and Hansen, 2002a), making it solvable by an exact method such as the outer-approximation algorithms (Ljubić and Moreno, 2018, Mai and Lodi, 2020). In the following, we show that it is the case for our robust model. First, recall that under the MNL model, the CPGF becomes

$$G^i(Y(\mathbf{v}^i)|S) = \sum_{j \in S} e^{v_{ij}}, \text{ and } G^i(Y(\mathbf{v}^i) \circ \mathbf{x}) = \sum_{j \in [m]} e^{v_{ij}} x_j.$$

Proposition 5 (Concavity under the MNL model) *$f^{\text{WC}}(\boldsymbol{x})$ is concave in $\boldsymbol{x}$.*

The claim can be obviously verified based on the fact that minimization will preserve the concavity. That is, if we define (for ease of notation)

$$\gamma(\mathbf{x}, \mathbf{V}) = \sum_{i \in I} q_i - \sum_{i \in I} \frac{q_i}{1 + G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})},$$

then we know that $\gamma(\mathbf{x}, \mathbf{V})$ is concave in $\mathbf{x}$ (Benati and Hansen, 2002a). Note that $f^{\text{WC}}(\mathbf{x}) = \min_{\mathbf{v}^i \in \mathbf{V}, \forall i} \gamma(\mathbf{x}, \mathbf{V})$. To prove the concavity of $f^{\text{WC}}(\mathbf{x})$ we will show that for any $\mathbf{x}^1, \mathbf{x}^2 \in [0,1]^m$ and $\alpha \in [0,1]$, we have $\alpha f^{\text{WC}}(\mathbf{x}^1) + (1-\alpha) f^{\text{WC}}(\mathbf{x}^2) \leq f^{\text{WC}}(\alpha \mathbf{x}^1 + (1-\alpha)\mathbf{x}^2)$. This is verified by the following chain of inequalities

$$\begin{aligned}
\alpha f^{\text{WC}}(\mathbf{x}^1) + (1-\alpha) f^{\text{WC}}(\mathbf{x}^2) &= \min_{\mathbf{v}^i \in \mathcal{V}_i, \forall i} \left\{ \alpha \gamma(\mathbf{x}^1, \mathbf{V}) \right\} + \min_{\mathbf{v}^i \in \mathcal{V}_i, \forall i} \left\{ (1-\alpha) \gamma(\mathbf{x}^2, \mathbf{V}) \right\} \\
&\leq \min_{\mathbf{v}^i \in \mathcal{V}_i, \forall i} \left\{ \alpha \gamma(\mathbf{x}^1, \mathbf{V}) + (1-\alpha) \gamma(\mathbf{x}^2, \mathbf{V}) \right\} \\
&\overset{(a)}{\leq} \min_{\mathbf{v}^i \in \mathcal{V}_i, \forall i} \left\{ \gamma(\alpha \mathbf{x}^1 + (1-\alpha)\mathbf{x}^2, \mathbf{V}) \right\} \\
&= f^{\text{WC}}(\alpha \mathbf{x}^1 + (1-\alpha)\mathbf{x}^2)
\end{aligned}$$

where (a) is due to the concavity of $\gamma(\mathbf{x}, \mathbf{V})$.

Note that the concavity is preserved with any uncertainty set, not necessarily with convex and customer-wise decomposable sets. That is, under any nonempty uncertainty set $\mathcal{V}$ such that the worst-case objective function

$$f^{\text{WC}}(\mathbf{x}) = \min_{\{\mathbf{v}^1, \ldots, \mathbf{v}^{|I|}\} \in \mathcal{V}} \left\{ \sum_{i \in I} q_i - \sum_{i \in I} \frac{q_i}{1 + G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})} \right\}$$

is finite, then $f^{\text{WC}}(\mathbf{x})$ is concave in $\mathbf{x}$. However, under this general setting, the above minimization problem is difficult to solve, as its objective function is highly non-convex in $\mathbf{V} = (\mathbf{v}^i,\ i \in I)$.

4 Robust Algorithms

For the MCP under GEV models, Dam et al. (2021) proposes a local search procedure, named as GGX, which does not only provide a performance-guaranteed solution but performs well in practice. On the other hand, if the choice model is MNL, it is possible to efficiently solve the deterministic MCP using a multi-cut outer-approximation algorithm (Mai and Lodi, 2020). As shown above, our robust models preserve some main properties of the deterministic versions, i.e., the submodularity and monotonicity for the MCP under GEV models, and concavity for the MCP under MNL, making the local search and outer-approximation approaches still useful. In the following, we discuss how such approaches can be adapted to handle the robust MCP.

4.1 Local Search for GEV-based Robust MCP

The submodularity and monotonicity of the objective function of (RMCP) shown above are sufficient to guarantee that a simple greedy heuristic can always yield a $(1-1/e)$ approximation. Such a greedy can simply start with a null set and iteratively add locations, one at a time, until the capacity $|S| = C$ is reached. The second and last phases of the GGX algorithm apply due to that the fact that the objective function $f^{\rm WC}(\mathbf{x})$ is differentiable in $\mathbf{x}$. Such derivatives can be computed as, for any $j \in [m]$,

$$\begin{aligned}\frac{\partial f^{\rm WC}(\mathbf{x})}{\partial x_j} &= \sum_{i\in I} q_i \frac{\partial \phi^i(\mathbf{x})/\partial x_j}{(1+\phi^i(\mathbf{x}))^2} \\ &\overset{(a)}{=} \sum_{i\in I} q_i \frac{Y_j(\mathbf{v}^{i*})\partial G^i_j(\mathbf{Y}(\mathbf{v}^{i*})\circ \mathbf{x})}{(1+G^i(\mathbf{Y}(\mathbf{v}^{i*})\circ\mathbf{x}))^2},\end{aligned} \tag{13}$$

where $\phi^i(\mathbf{x})$ is defined in (5) and *(a)* is due to Lemma 1. We can apply the second phase of GGX (i.e., gradient-based local search) to further improve the solution candidate from the greedy heuristic. At each iteration of this phase, we need to solve the following subproblem

$$\begin{aligned} \max_{\mathbf{x}\in\{0,1\}^m} \quad & \nabla f^{\rm WC}(\overline{\mathbf{x}})^{\rm T}\mathbf{x} && \text{(Sub-Prob)} \\ \text{subject to} \quad & \sum_j x_j = C \\ & \sum_{j\in[m],\overline{x}_j=1} (1-x_j) + \sum_{j\in[m],\overline{x}_j=0} x_j \le \Delta, && (14)\end{aligned}$$

where $\overline{\mathbf{x}}$ is the current solution candidate, $\Delta > 0$ is a positive integer scalar used to define a local area around $\overline{\mathbf{x}}$ in which we want to find the next solution candidate, and (14) is referred to as a local branching constraint typically used to exploit the neighborhood of a given binary solution (Fischetti and Lodi, 2003). The use of Δ in the subproblem is motivated by the trust-region method in the continuous optimization literature (Conn et al., 2000). (Dam et al., 2021) show that their subproblem can be solved efficiently in $\mathcal{O}(m\Delta)$ but their algorithm requires that all the coefficients of the objective function of the subproblem are non-negative. The following proposition tells us that it is the case under our robust model, making it possible to apply Algorithm 1 in Dam et al. (2021) to solve (Sub-Prob).

Proposition 6 *Given any $\overline{\mathbf{x}} \in \{0,1\}^m$, all the coefficients of the objective function of* (Sub-Prob) *are non-negative. As a result,* (Sub-Prob) *can be solved in $\mathcal{O}(m\Delta)$.*

Proof. We simply use Lemma 1 and Property *(iv)* of Remark 1 to see that

$$\frac{\partial f^{\rm WC}(\overline{\mathbf{x}})}{\partial x_j} = \sum_{i\in I} q_i \frac{Y_j(\mathbf{v}^{i*}(\overline{\mathbf{x}}))\partial G^i_j(\mathbf{Y}(\mathbf{v}^{i*}(\overline{\mathbf{x}}))\circ\overline{\mathbf{x}})}{(1+G^i(\mathbf{Y}(\mathbf{v}^{i*}(\overline{\mathbf{x}}))\circ\overline{\mathbf{x}}))^2} \ge 0,$$

where $\mathbf{v}^{i*}(\overline{\mathbf{x}}) = \text{argmax}_{\mathbf{v}^i} G^i(\mathbf{Y}(\mathbf{v}^i)\circ\overline{\mathbf{x}})$. With the positiveness of the coefficients, Dam et al.

(2021) show that (Sub-Prob) can be effciently solved in $\mathcal{O}(m\Delta)$. ■

The third phase of GGX is based on steps of adding/removing locations, thus it can be applied straightforwardly to the robust problem. We briefly describe our adapted GGX in Algorithm 1 below and we refer reader to Dam et al. (2021) for more details. Here we note that, to compute $f^{\text{WC}}(\mathbf{x})$, we need to solve the inner adversary problems $\min_{\mathbf{v}^i} G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})$, which are convex optimization ones (Proposition 3) and can be solved efficiently using a convex optimization solver. Moreover, let $\mathbf{v}^{i*} = \text{argmax}_{\mathbf{v}^i \in \mathcal{V}_i} G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})$, then using Lemma 1 we can compute the first derivatives of $f^{\text{WC}}(\mathbf{x})$ used in (Sub-Prob) as in (14).

Algorithm 1: Local Search

*# 1: **Greedy heuristics***
- *Start from $S = \emptyset$*
- *Iteratively add locations to S, one at a time, taking locations that increase $f^{\text{WC}}(S)$ the most*
- *Stop when $|S| = C$*

*# 2: **Gradient-based local search***
- *Iteratively solve* (Sub-Prob) *to find new solution candidates*
- *Move to a new candidate solution if it yields a better objective value $f^{\text{WC}}(S)$*
- *If* (Sub-Prob) *yields a worse solution, then reduce the size of the searching area Δ*

*# 3: **Exchanging phase***
- *Swap one (or two) locations in S with one (or two) locations in $[m] \backslash S$, taking locations that increase the objective function the most*
- *Stop when no further improvements can be made.*

So far we assume that there is only a cardinality constraint $|S| \leq C$. In a real-life application, one may require some side constraints on $\mathbf{x}$ to capture, for instance, traveling costs between facilities. In this context, for Algorithm 1 to work, one can modify Steps #1 and #2 to account for the additional constraints. For Step #3, we only need to add the side constraints to the subproblem (Sub-Prob). Such additional constraints, however, would break the performance guarantee secured by the greedy heuristic.

4.2 Outer-Approximation for the Robust MCP under MNL

Under the MNL model, the concavity shown in Proposition 5 suggests that one can use a multicut outer-approximation algorithm to exactly solve the MCP. The idea is to divide the set of zones I into $\mathcal{L}$ disjoint groups $\mathcal{D}_1, \ldots, \mathcal{D}_{\mathcal{L}}$ such that $\bigcup_{l \in [\mathcal{L}]} \mathcal{D}_l = I$. We then write the objective function of (RMCP) as

$$f^{\text{WC}}(\mathbf{x}) = \sum_{l \in [\mathcal{L}]} \delta_l(\mathbf{x}),$$

where

$$\delta_l(\mathbf{x}) = \sum_{i \in \mathcal{D}_l} q_i - \sum_{i \in \mathcal{D}_l} \frac{q_i}{1 + \min_{\mathbf{v}^i \in \mathcal{V}_i} \left\{ \sum_{j \in [m]} e^{v_{ij}} x_j \right\}}, \ \forall l \in [\mathcal{L}].$$

A multicut outer-approximation algorithm executes by iteratively adding sub-gradient cuts to a master problem and solve it until reaching an optimal solution. Such a master problem can be defined as

$$\max_{\mathbf{x}\in\{0,1\}^m, \boldsymbol{\theta}\in\mathbb{R}^L} \quad \sum_{l\in[L]} \theta_l \tag{sub-OA}$$

$$\text{subject to} \quad \sum_{j\in[m]} x_j = C$$

$$\mathbf{A}\mathbf{x} - \mathbf{B}\boldsymbol{\theta} \geq \mathbf{c} \tag{15}$$

$$\boldsymbol{\theta} \geq 0,$$

where the linear constraints (15) are linear cuts of the form $\theta_l \leq \nabla\delta_l(\mathbf{x})(\mathbf{x} - \overline{\mathbf{x}}) + \delta_l(\overline{\mathbf{x}})$ added to the master problem (sub-OA) at each iteration of the outer-approximation algorithm with a solution candidate $\overline{\mathbf{x}}$. The algorithm stops when we find a solution $(\mathbf{x}^*, \boldsymbol{\theta}^*)$ such that $\sum_{l\in[L]} \theta_l^* \leq \sum_{l\in[L]} \delta_l(\mathbf{x}^*)$. Here we note that the outer-approximation algorithm works with any side linear constraints on $\mathbf{x}$ and always returns an optimal solution after a finite number of iterations due to the concavity of $\delta_l(\mathbf{x})$ and the fact that the feasible set of $\mathbf{x}$ is finite. Similar to (14), the gradients of $\delta_l(\overline{\mathbf{x}})$ can be computed as

$$\frac{\delta_l(\overline{\mathbf{x}})}{\partial x_j} = \sum_{i\in\mathcal{D}_l} \frac{q_i \partial G_j^i(\mathbf{Y}(\mathbf{v}^{i*}) \circ \overline{\mathbf{x}})}{\Big(1 + G_j^i(\mathbf{Y}(\mathbf{v}^{i*}) \circ \overline{\mathbf{x}})\Big)^2}, \; \forall j \in [m].$$

We briefly describe our multicut outer-approximation approach in Algorithm 2 below.

Algorithm 2: Multicut Outer-Approximation

Initialize the master problem (sub-OA).
Choose a small threshold $\tau > 0$ *as a stopping criteria.*
do
  Solve (sub-OA) *to get a solution* $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{\theta}})$
  if $\sum_{l\in[L]} \overline{\theta}_l > \sum_{l\in[L]} \delta_l(\overline{\boldsymbol{x}}) + \tau$ **then**
    Add constraints $\overline{\theta}_l \leq \nabla\overline{\delta}_l(\overline{\boldsymbol{x}})(\boldsymbol{x} - \overline{\boldsymbol{x}}) + \delta_l(\overline{\boldsymbol{x}})$ *to* (sub-OA).
until $\sum_{l\in[L]} \overline{\theta}_l \leq \sum_{l\in[L]} \delta_l(\overline{\boldsymbol{x}}) + \tau$;
Return $\overline{\boldsymbol{x}}$.

It is possible to use the Branch-and-Cut algorithm proposed in Ljubić and Moreno (2018) to solve the robust MCP under MNL. Our multicut outer-approximation described here is similar to the one used in Mai and Lodi (2020) and differs from the Branch-and-Cut method of Ljubić and Moreno (2018) by the fact that it generates one cut per a group of multiple demand points instead of one cut per every single demand point, and it is a Cutting Plane approach instead of a Branch-and-Cut. Mai and Lodi (2020) show that the Cutting Plane approach is more efficient than the Branch-and-Cut in handling large-scale instances.

5 Numerical Experiments

We provide numerical experiments showing the performance of our robust approach in protecting us from worst-case scenarios when the choice parameters are not known with certainty. We first discuss our approach to construct uncertainty sets to capture the issue of choice parameter uncertainties. We then provide experiments with MCP instances under two popular discrete choice models in the GEV family, namely, the MNL and nested logit models.

5.1 Constructing Uncertainty Sets

We first discuss our approach to build uncertainty sets to capture uncertainty issues when identifying choice parameters in the MCP. In the deterministic setting, it is assumed that there is only one vector of customer choice utilities $\mathbf{v}^i$ for each zone $i \in I$. However, the real market typically has many different types of customers characterized by, for instance, age or income. The choice parameters may significantly vary across different customer types. Let us assume that, for each zone $i \in I$, there are N types of customers with N utility vectors $\widetilde{\mathbf{v}}^{i1}, ..., \widetilde{\mathbf{v}}^{iN}$, respectively. We let $\tau_{i1}, ..., \tau_{iN} \in [0,1]$ be the actual proportions of the customer types with $\sum_{n\in[N]} \tau_{in} = 1$. If these proportions are known with certainty, then one can define a mean value vector $\widetilde{\mathbf{v}}^i = \sum_{n\in[N]} \tau^{in}\widetilde{\mathbf{v}}^{in}$ and solve the corresponding deterministic MCP problem. Our assumption here is that these proportions cannot be identified exactly, but we know that the actual proportions are not too far from some estimated proportions. In this context, uncertainty sets can be defined as

$$\mathcal{V}_i = \left\{ \mathbf{v}^i = \sum_{n\in[N]} \eta_n \widetilde{\mathbf{v}}^{in} \middle| \eta_n \geq 0, \forall n \in [M]; \sum_{n\in[N]} \eta_n = 1; \quad \text{and} \quad ||\boldsymbol{\eta} - \widetilde{\boldsymbol{\tau}}^i|| \leq \epsilon \right\}, \tag{16}$$

where $\widetilde{\boldsymbol{\tau}}^i = \{\widetilde{\tau}^{in},\ n \in [N]\}$ are some estimates of the actual proportions $\boldsymbol{\tau}^i$ and $\epsilon \in [0,1]$ represents the "uncertainty level" of the uncertainty set. Larger ϵ values lead to larger uncertainty sets, thus resulting in more conservative models that may help provide better protection against worst-case scenarios, but may give a low average performance. In contrast, smaller ϵ values provide smaller uncertainty sets, which would lead to better average performance but may give a bad performance in protecting against worst-case scenarios. The firm could adjust ϵ to balance the worst-case protection and average performance. Clearly, if $\epsilon = 0$, the uncertainty sets become singleton and the robust MCP becomes a deterministic MCP with mean-value choice utilities, i.e., the actual proportions are known perfectly, or the firm just treats their estimated proportions as the actual ones. On the other hand, if we select $\epsilon > N$, then the uncertainty sets will cover all possible affine combinations of $\{\widetilde{\mathbf{v}}^i,\ i \in I\}$. This reflects the situation that the firm knows nothing about the actual proportions and has to ignore the pre-computed values $\widetilde{\boldsymbol{\tau}}^i$. This way of constructing uncertainty sets is similar to the approach employed in Rusmevichientong and Topaloglu (2012) in the context of assortment optimization under uncertainty.

5.2 Baseline Approaches and Other Settings

We will compare our robust approach (denoted as RO) against other baseline approaches under the MNL and nested logit models. The first baseline approach, which is a deterministic one (denoted as DET1), relies on mean-value choice parameters, i.e., we solve the MCP with average values of the choice utilities, i.e. $\widetilde{\mathbf{v}}^i = \sum_{n\in[N]} \tau_{in}\widetilde{\mathbf{v}}_{in}$. This is the case of $\epsilon = 0$, i.e., no uncertainty in the robust model. For the second baseline approach (denoted as DET2), we utilize all the vectors of choice parameters $\widetilde{\mathbf{v}}_{in}$, for all $n \in [N]$ and solve the following mixed version of the MCP

$$\max_{S\in\mathcal{K}} \left\{ \sum_{i\in I} q_i \sum_{n\in[N]} \tau^{in} \left(\sum_{j\in S} P(j|\widetilde{\mathbf{v}}_{in}, S) \right) \right\}. \tag{17}$$

So, for the DET1 and DET2 approaches, we acknowledge that there are N customer types in the market but ignore the uncertainty issue. Note that the DET2 approach can be viewed as an MCP under a mixed logit model. We consider another baseline that accounts for the issue that the proportion of the customer types maybe not be determined with certainty. We perform this approach by sampling over the uncertainty sets $\mathcal{V}_i$. Then, for each sample of the utilities, we solve the corresponding MCP to get a solution. Since we aim at covering the worst-case scenarios, for each solution, we sample again from the uncertainty sets to get an approximation of the worst expected captured demand value. We then pick a solution that gives the best worst-case objective value over samples. This way allows us to find solutions that are capable of guaranteeing some protection against worst-case scenarios. This approach can be viewed as a sampling-based method to solve the robust problem. We denote it as SA.

For all the MNL and nested logit instances, we employ the local search (Algorithm 1) to solve the robust/deterministic MCP, for all the RO, DET1, DET2, and SA approaches, noting that for MNL instances, the outer-approximation (Algorithm 2) can be used as an exact method, but it is generally outperformed by the local search in terms of both solution quality and running time. For nested logit instances, the local search is also more convenient to use, as (i) the multicut outer-approximation algorithm becomes heuristic and does not offer us any guarantees, and (ii) Dam et al. (2021) already show that the local search outperforms the multicut outer-approximation algorithm for nested instances. Thanks to the submodularity, the local search always gives us at least $(1 - 1/e)$ approximation solutions (Corollary 1).

We use instances from the three datasets HM14, ORlib, and NYC to generate instances for our robust problems and we refer the reader to Freire et al. (2016) for a detailed description. These datasets have been used in previous MCP studies (Dam et al., 2021, Ljubić and Moreno, 2018, Mai and Lodi, 2020). We select $N = 5$ (i.e., there are 5 customer types in each zone). We then randomly choose the pre-determined proportions $\{\tau_1, ..., \tau_5\}$ such that $\sum_{n\in[N]} \tau_n = 1$. For each deterministic instance from the datasets (HM14, ORlib, or NYC) to generate the underlying utility vectors $\{\widetilde{\mathbf{v}}^{1i}, ..., \widetilde{\mathbf{v}}^{5i}\}$ to construct uncertainty sets, we take the set of utility values $\{v^i_j,\ i \in I, j \in [m]\}$ from the data and sample (randomly and uniformly) each element $\widetilde{v}^{ki}_j$ in range $[0.7 \times v^i_j; 1.3 \times v^i_j]$, for all $k = 1, \ldots, 5$. To compare the performances of the RO, DET, and SA approach in protecting us from worst-case scenarios, we vary the uncertainty level ϵ and compare the robust solutions with those from the DET and SA approaches. More precisely, we will perform the following steps for each $\epsilon > 0$.

(i) Define an uncertainty set as in (16).

(ii) For RO, we solve the robust problem (RMCP) and obtain a robust solution $\mathbf{x}^{\text{RO}}$.

(iii) For DET1, we solve the deterministic MCP (MCP) with the weighted average utilities $\widetilde{\mathbf{v}} = \sum_{n\in[n]} \tau_n \widetilde{v}_n$ and obtain a solution $\mathbf{x}^{\text{DET1}}$.

(iv) For DET2, we solve (17) to get a solution $\mathbf{x}^{\text{DET2}}$.

(v) For SA, we take 10 samples from the uncertainty sets to obtain 10 candidate solutions.[2] Then, for each solution obtained, we use 1000 samples of the utilities, taken from the uncertainty sets, to get an approximate worst-case objective value. We then pick the solution $\mathbf{x}^{\text{SA}}$ that gives the best worst-case objective. All the sampling is done randomly and uniformly.

(vi) We assess the value and the price of the robust approach by checking the performances of different solutions obtained by the robust, sampling-based, and deterministic approaches in the uncertain environment, and in the nominal (or deterministic) setting. That is, we plug $\mathbf{x}^{\text{DET1}}$, $\mathbf{x}^{\text{DET2}}$, and $\mathbf{x}^{\text{SA}}$ into the robust problem to compute relative decreases in terms of worst-case objective, and plug $\mathbf{x}^{\text{RO}}$ and $\mathbf{x}^{\text{SA}}$ into the deterministic MCP to evaluate relative decreases in terms of average objective. These two measurements are often referred to as the value and the price of robustness and have been popularly used to assess robust optimization methods (Bertsimas and Sim, 2004, Mehmanchi et al., 2020)

(vii) We additionally evaluate the performances of $\mathbf{x}^{\text{RO}}$, $\mathbf{x}^{\text{DET1}}$, $\mathbf{x}^{\text{DET2}}$, and $\mathbf{x}^{\text{SA}}$ by looking at the empirical distributions of the objective values yielded by different vectors of utilities sampled from the uncertainty sets. To this end, we randomly and uniformly sample 2000 utility vectors $\{\mathbf{v}^i,\ i \in I\}$ from the uncertainty sets and compute, for each utility sample, the corresponding objective values given by $\mathbf{x}^{\text{RO}}$, $\mathbf{x}^{\text{DET1}}$, $\mathbf{x}^{\text{DET2}}$, and $\mathbf{x}^{\text{SA}}$. This allows us to plot and analyze empirical distributions of the objective values given by different solutions when the uncertain utilities $\mathbf{v}^i$ vary randomly within the uncertainty sets.

We use MATLAB 2020 to implement and run the algorithms. We use the maximum norm to define the uncertainty sets in (16). Under this setting, the adversary's optimization problems can be formulated as convex optimization ones with (strictly) convex objective functions and linear constraints. We use the nonlinear optimization solver *fmincon* from MATLAB, under default settings, to solve the adversary's convex optimization problems (i.e., $\min_{\mathbf{v}^i\in} G^i(\mathbf{Y}(\mathbf{v}^i)\circ\mathbf{x})$). The experiments were done on a PC with processor AMD-Ryzen 7-3700X CPU-3.80 GHz and with 16 gigabytes of RAM.

[2] More samples can be taken, but we restrict ourselves to 10 samples, as the SA approach is expensive to run as we will show later.

5.3 Comparison Results

In this section, we present comparison results for MNL and nested logit instances to assess the value of the robust approach. Computing time comparisons will also be presented.

5.3.1 MNL Instances

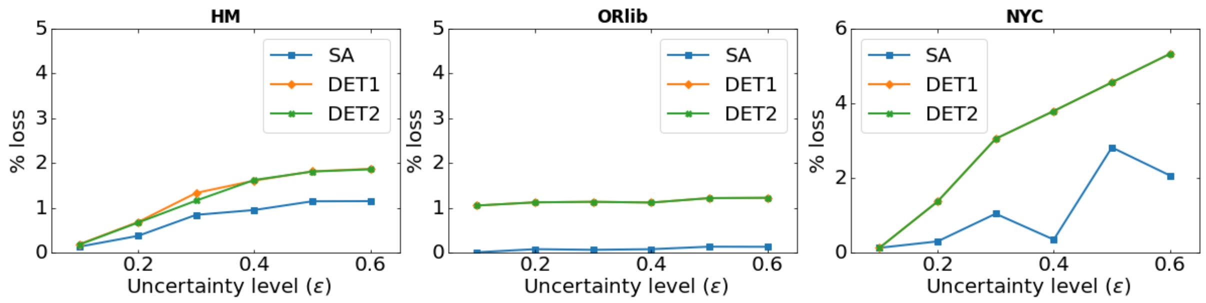


Figure 1: Value of robustness for MNL instances; the performances of DET1 and DET2 are almost identical.

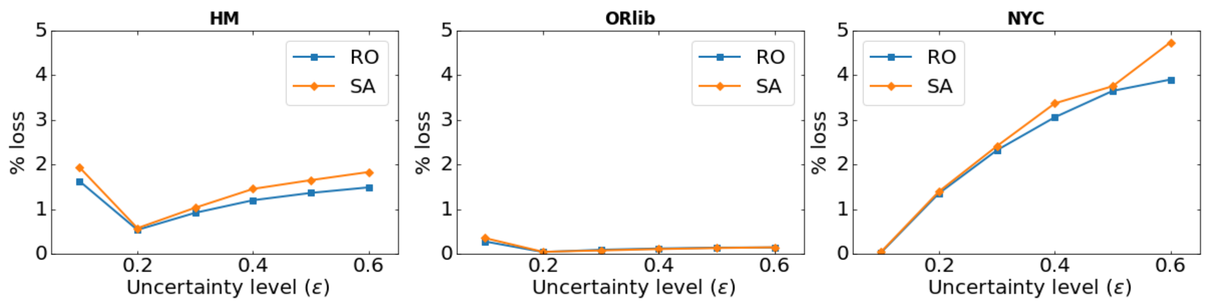


Figure 2: Price of robustness for MNL instances.

We first solve MNL instances using Algorithm 1 and plot, in Figure 1 and Figure 2, the average relative decreases ("% loss") w.r.t. the robust and deterministic MCP for the three datasets (i.e. HM, ORlib, and NYC). In both figures, higher "% loss" implies worse results. The results generally show that the percentage loss, as expected, seems to increase as the uncertainty level ϵ increases. The percentage losses for ORlib are remarkably smaller than the losses for the other datasets, and the losses for the largest dataset NYC are, as expected, significantly higher than those from ORlib and HM. Moreover, Figure 1 shows that SA performs better than DET1 and DET2 in terms of the value of robustness. Looking at both figures, it can be observed that the deterministic solution performs worse in the robust setting than the robust and SA solutions do in the deterministic setting for the ORlib dataset, and performs comparably for the other datasets.

To further assess the performance of the robust approach, we pick instances of size $(|I|, m) = (100, 50)$ and plot, in Figure 3, empirical distributions of the objective values given by the four approaches and utilities sampled from the uncertainty sets. More plots can be found in Appendix

C. The first row of the figure shows the histograms for small ϵ ($\epsilon \leq 0.08$). Under these small *uncertainty levels*, the distributions given by the four approaches are quite similar and there is no clear advantage of the robust approach. This is an expected observation, as when ϵ is small, the corresponding uncertainty sets become small and $\mathbf{x}^{\text{RO}}$ should be close to $\mathbf{x}^{\text{DET1}}$, $\mathbf{x}^{\text{DET2}}$, and $\mathbf{x}^{\text{SA}}$. When we increase ϵ, the second row of Figure 3 shows clear differences between the four approaches. That is, the distributions given by $\mathbf{x}^{\text{RO}}$ always have lower variances, and shorter tails, as compared to the those from the other approaches. In terms of worst-case protection, the RO performs the best, followed by the SA and then the two deterministic approaches. This clearly demonstrates the capability of the RO approach in giving "*not-too-low*" expected captured demands. Moreover, the protection against "low" objective values seems higher as ϵ increases. This is also an expected observation, as the RO and SA approaches are essentially designed for this purpose. We also observe that the two DET approaches perform better in terms of best-case scenarios, indicating a trade-off of being conservative when making robust decisions under uncertainty.

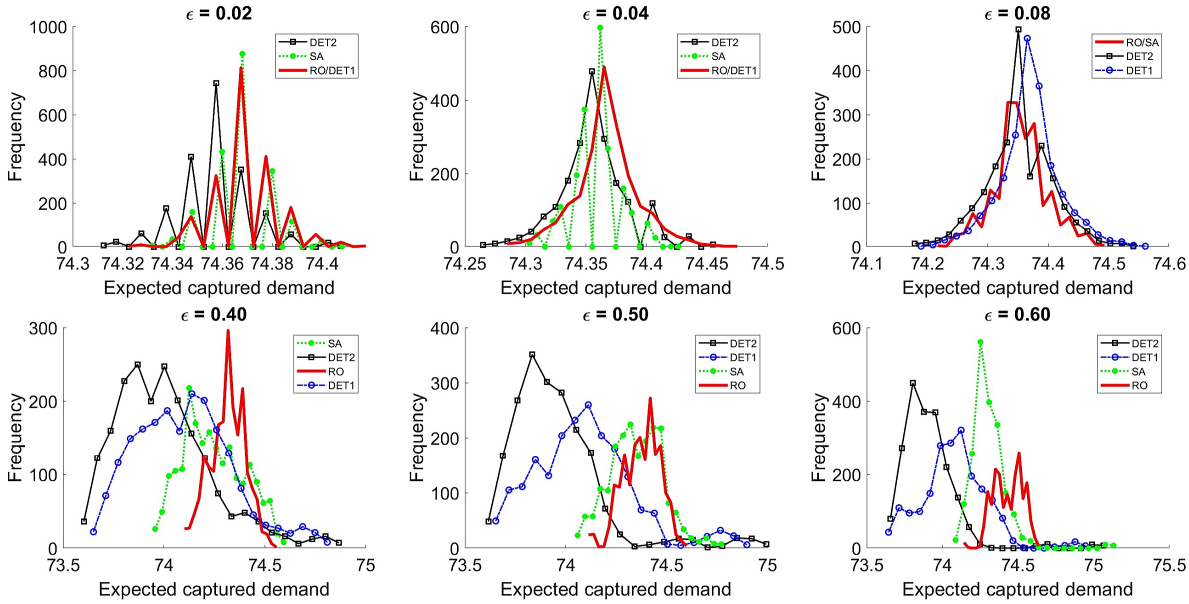


Figure 3: Comparison of the distributions of the objective values given by solutions from RO, SA, DET1 and DET2 approaches, under the MNL choice model and instances of size $|I| = 100$ and $m = 50$.

Before moving to nested logit instances, we note that the robust MCP under MNL is relevant to the robust fractional 0-1 program studied in Mehmanchi et al. (2020). We provide a comparison with this work in Appendix B.

5.3.2 Nested Logit Instances

We now provide comparison results for nested logit instances. As mentioned, to create the instances, we divide the set of locations into 5 groups of the same size (i.e, $L = 5$) and choose the nested logit parameters as $\mu = (1.1, 1.2, 1.3, 1.4, 1.5)$, noting that these parameters are just selected at random for a testing purpose and other values can be chosen. Similar to the MNL instances, we first report the value and price of robustness in Figures 4 and 5 below. We first notice that the percentage losses reported in Figure 4 are remarkably smaller than those reported for the MNL instances. This would be because we solve the nested logit instances by the local search algorithm, thus the solutions $\mathbf{x}^{\text{RO}}$ obtained may not be optimal for the robust problem, affecting the value of robustness. Moreover, in analogy to MNL instances, SA performs slightly better than DET1 and DET2, in terms of the value of robustness. We also see that the percentage losses for ORlib are smaller, compared to losses for the other datasets. Especially, the percentage losses of the RO and SA solutions in the deterministic environment almost vanish. The losses for NYC instances, similarly to the MNL case, are also significantly larger than the losses for ORlib and HM. Moreover, while the robust solutions perform better in the deterministic problem than the DET solutions do in the robust setting for the ORlib dataset, they perform slightly worse for the other datasets.

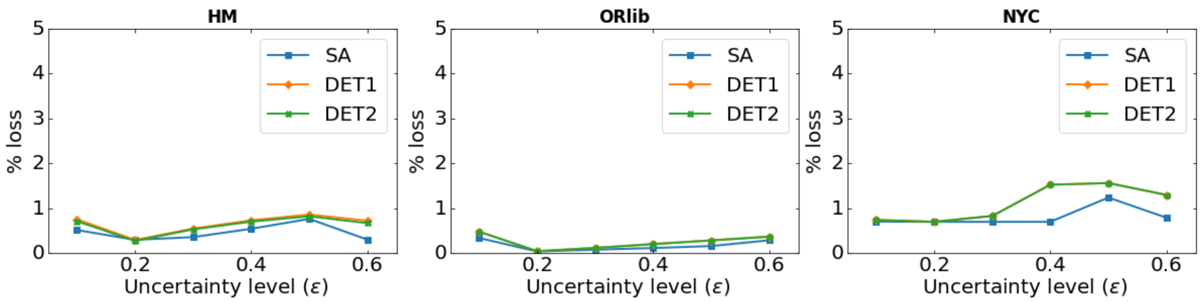


Figure 4: Value of robustness for nested logit instances; the performances of DET1 and DET2 are almost identical.

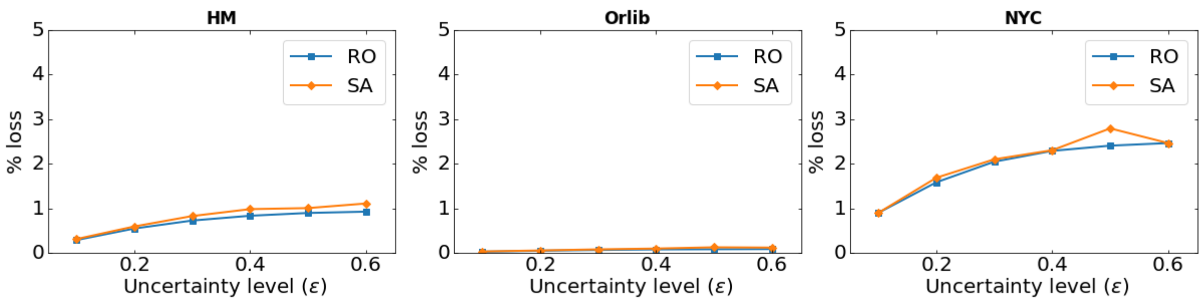


Figure 5: Price of robustness for nested logit instances.

We also pick instances of size $(|I|, m) = (100, 50)$ to plot empirical distributions of the objective values given by the four approaches and utilities sampled from the uncertainty sets. In the first row of Figure 6, we plot the histograms for $\epsilon \in \{0.02, 0.04, 0.08\}$. As we can see, histograms given by RO, DET1, and SA are identical. For $\epsilon = 0.02$ or $\epsilon = 0.04$, the histograms look very similar, but for $\epsilon = 0.08$ we start seeing that the histogram given by RO has smaller variance and shorter tail. Moreover, even though the differences are not clear with these small uncertainty set levels, some protection from the RO approach against bad scenarios can still be observed.

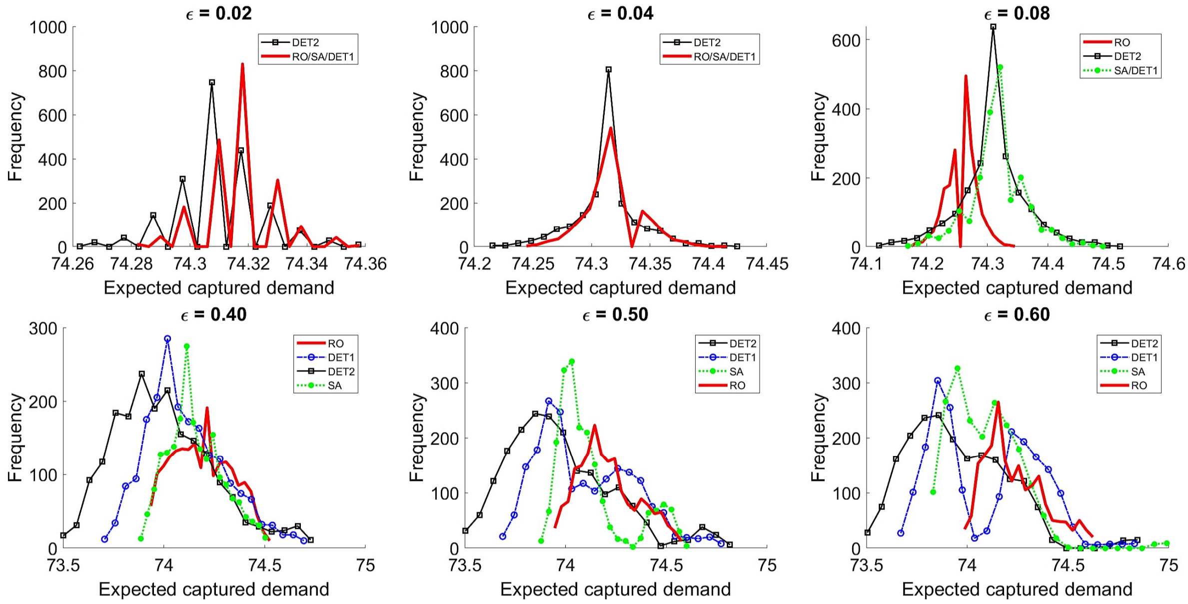


Figure 6: Comparison of the distributions of the objective values given by solutions from RO, SA, DET1 and DET2 approaches, under the nested logit choice model and instances of size $|I| = 100$ and $m = 50$.

Histograms with larger ϵ ($\epsilon \in \{0.4, 0.5, 0.6\}$) are plotted in the second row of Figure 6. There is no surprise, as similar to the MNL instances, the distributions given by the RO approach have small variances, shorter tails, and larger worst-case objective values, as compared to those given by the other approaches. In particular, we can see that the two deterministic approaches can give many low objective values. The SA approach seems to do better in protecting the objective value from being too low, and RO performs the best in pushing its worst-case scenarios to higher objective values. It is worth noting that the DET1 approach (when average values of the choice utilities are made use of) performs better than the mixed deterministic approach DET2 in terms of worst-case protection. A trade-off between having high worst-case objective values and having low best-case objective values can also be observed. This is also consistent with remarks from prior studies in the context of assortment and pricing optimization where discrete choice models are also employed (Li and Ke, 2019, Mai and Jaillet, 2019, Rusmevichientong and Topaloglu, 2012).

5.3.3 Percentile Ranks of the RO's Worst-case Objective Values

We look closely into the distributions of the objective values given by the four approaches to see where the RO's worst-case values are located in other distributions. This would help evaluate how much RO can provide protection when ϵ increases. To this end, we compute the percentile ranks of the RO's worst-case objective values in the distributions given by the DET and SA solutions. Such a percentile rank can tell us the percentage of the values in the DET1, DET2, or SA distributions that are lower than or equal to the worst objective values given by the RO approach. We plot the percentile ranks of the RO's worst objective values in Figure 7 for both the MNL and nested logit models. For the MNL instances, when $\epsilon > 0.1$, the percentile ranks are significant and increasing from about 0% to almost 100% for DET2, from 0% to about 60% for DET1, and are quite small for SA (less than 20%). For the nested logit instances, the percentile ranks for DET1 and DET2 can go up to more than 40% and can go up to 30% for the SA approach. This clearly indicates how much RO can provide protection, as compared to the other approaches, for instance, for the nested model with $\epsilon = 0.6$, the RO's worst objective values are larger than more than 30% of the SA's sample objective values, and more than 40% of the DET1's and DET2's sample objective values. It can be seen that, among the DET1, DET2 and SA approaches, SA performs better than DET1 and DET2, and DET1 is better than DET2, in terms of worst-case protection.

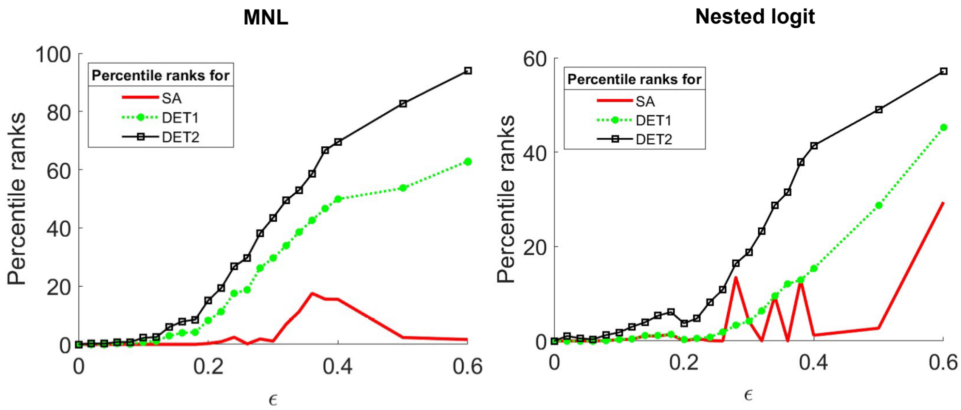


Figure 7: The percentile ranks of the RO's worst-case values in the distributions given by the SA, DET1 and DET2 solutions under the MNL and nested choice models for instance of $|I| = 100$ and $m = 50$.

5.4 Computing Time Comparison

We look at the computing times required by the different approaches to terminate. In Table 1, we report the average computing times for the four approaches, under the two discrete choice models (the MNL and nested logit) and with two instance sizes. It clearly shows that DET1 require the least amount of running time. It is interesting to see that RO requires less computing times than DET2, except for the instances of size $|I| = 82341$. SA is much more time-consuming, as compared to the other three approaches since it requires sampling and solving several deterministic problems. On the other hand, for the largest problem, RO requires the largest computing times under the MNL model due to the fact that it needs to solve a large number of inner minimization problems (i.e., solving $\min_{\mathbf{v}^i \in \mathcal{V}_i} G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})$ for $i = 1, 2, ..., 82341$).

Choice model	$\|I\|$	m	RO	SA	DET1	DET2
MNL	100	50	14.0	18.8	0.3	3.6
	100	100	12.4	21.6	0.3	1.5
	200	100	26.7	25.09	0.3	13.7
	1000	100	60.0	479.72	45.3	680.4
	82341	59	2861.1	885.1	0.39	6.5
Nested logit	100	50	5.6	17.0	1.4	3.3
	100	100	11.3	88.8	8.2	6.6
	200	100	16.40	116.4	11.0	23.3
	1000	100	49.0	311.6	23.2	155.7
	82341	59	2262.4	4386.9	311.5	2586.7

Table 1: Average computing times (in seconds).

As shown in Dam et al. (2021), such a local search procedure as in Algorithm 1 achieves the best performance in terms of objective value and computing time, as compared to outer-approximation algorithms or MILP reformulations. Thus, we provide here a comparison of Algorithm 1 and Algorithm 2. Note that, for the MCP under the nested logit model, the objective function is highly non-concave and the outer-approximation method often performs much worse than the local search heuristic (Dam et al., 2021). Thus, we only provide a comparison of the two algorithms with MNL instances. We use Algorithm 1 and Algorithm 2 to solve the robust MCP with MNL instances of different sizes with $\epsilon \in \{0.02, 0.04, 0.08, 0.4, 0.5, 0.6\}$ and report the objective values and the computing times in Table 2. For ease of exposition, we denote Algorithm 1 as LS and Algorithm 2 as MOA. We observe that the LS and MOA always return the same objective values for all the instances, implying that LS always give us optimal solutions. In addition, the computing times required by MOA and LS are similar for the instances of sizes $(|I|, m) = (100, 50)$, $(|I|, m) = (100, 100)$ and $(|I|, m) = (200, 100)$. However, for instances of size $(|I|, m) = (1000, 100)$, LS requires shorter computing times than MOA. In particular, the computing times of LS are always below 40 seconds when MOA always needs more than 60 seconds to finish. For the largest instances $(|I|, m) = (82341, 59)$, LS requires shorter computing times for small ϵ value (i.e, $\epsilon \leq 0.08$). In contrast, MOA requires shorter computing times for the largest instances with large ϵ values (i.e., $\epsilon \geq 0.4$).

$\|I\|$	m	ϵ	MOA		LS	
			Objective values	Time	Objective values	Time
100	50	0.02	74.33	**2.24**	74.33	2.79
100	50	0.04	74.29	**2.38**	74.29	2.71
100	50	0.08	74.22	**2.68**	74.22	2.72
100	50	0.4	74.01	**2.21**	74.01	2.27
100	50	0.5	74.01	**1.83**	74.01	1.93
100	50	0.6	74.02	**1.76**	74.02	1.82
100	100	0.02	74.63	**3.08**	74.63	3.22
100	100	0.04	74.60	3.46	74.60	**3.44**
100	100	0.08	74.55	4.04	74.55	**3.99**
100	100	0.4	74.40	**2.40**	74.40	2.47
100	100	0.5	74.40	**2.07**	74.40	2.24
100	100	0.6	74.39	**2.03**	74.39	2.08
200	100	0.02	148.71	**5.92**	148.71	6.29
200	100	0.04	148.64	**6.53**	148.64	6.89
200	100	0.08	148.51	**7.28**	148.51	7.68
200	100	0.4	148.15	**4.62**	148.15	4.83
200	100	0.5	148.14	**4.13**	148.14	4.23
200	100	0.6	148.14	**3.84**	148.14	4.00
1000	100	0.02	39236.30	60.10	39236.30	**29.83**
1000	100	0.04	39228.52	78.44	39228.52	**31.78**
1000	100	0.08	39213.34	82.75	39213.34	**36.19**
1000	100	0.4	39153.40	75.73	39153.40	**29.17**
1000	100	0.5	39150.54	70.80	39150.54	**24.28**
1000	100	0.6	39149.27	69.19	39149.27	**22.81**
82341	59	0.02	71682.01	2915.80	71682.01	**2833.67**
82341	59	0.04	71603.14	3034.34	71603.14	**2962.36**
82341	59	0.08	71458.12	3181.14	71458.12	**3109.69**
82341	59	0.4	70996.35	**2384.12**	70996.35	2465.88
82341	59	0.5	70985.87	**2303.91**	70985.87	2395.21
82341	59	0.6	70982.53	**2237.14**	70982.53	2326.52

Table 2: Comparison of the MOA and LS algorithms.

6 Conclusion

We have formulated and studied a robust version of the MCP under GEV models. We have shown that the adversary's minimization problem can be solved by convex optimization, and the robust model preserves the monotonicity and submodularity from its deterministic counterpart, leading to the fact that a simple greedy heuristic can guarantee $(1-1/e)$ approximation solutions. We have then introduced the GGX algorithm that works with the robust MCP under GEV, and a multicut outer-approximation algorithm that can exactly solve the robust MCP under MNL. Our numerical experiments based on the MNL and nested logit models have shown the advantages of our model and algorithms in providing protection against worst-case scenarios, as compared to other deterministic and sampling-based baseline approaches.

Our robust model assumes that the choice parameters can vary uniformly in the uncertainty sets, which might be conservative in some contexts where the distribution of these parameters may be partially known. Therefore, it would be interesting to consider a distributionally robust version of the MCP. It is interesting to look at a model where the firm and the competitor make decisions in a Stackelberg game setting. Moreover, in this paper we assume that the (expected) number of customer in each zone is fixed and ignore any uncertainty associated with the fact that customers may travel between different zones. Accounting for this would require a new model to predict how customers would move between zones any maybe a new stochastic or robust optimization model for the MCP. This would be an interesting direction for future work.

As mentioned earlier, our model only targets uncertainties associated with the deterministic parts of the utilities. There would be other sources of uncertainty that would come from, for instance, the structure of the GEV model or the distribution of the random parts of the utilities. Models and algorithms addressing such uncertainties would be challenging, but worth being investigated. Moreover, since the objective functions of the MCP and robust MCP under GEV models are all submodular, a MILP approach based on submodular cuts would be interesting for future explorations.

Acknowledgments

We thank the Editor and the three referees for their detailed and thoughtful comments and suggestions, which substantially improved the paper. Dr. Tien Mai is supported by the National Research Foundation Singapore and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2- RP-2020-017)

References

Averbakh, I. and Berman, O. Minimax regret p-center location on a network with demand uncertainty. *Location Science*, 5(4):247–254, 1997.

Averbakh, I. and Berman, O. Algorithms for the robust 1-center problem on a tree. *European Journal of Operational Research*, 123(2):292–302, 2000a.

Averbakh, I. and Berman, O. Minmax regret median location on a network under uncertainty. *INFORMS Journal on Computing*, 12(2):104–110, 2000b.

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

Ben-Akiva, M. *The structure of travel demand models.* PhD thesis, MIT, 1973.

Ben-Tal, A. and Nemirovski, A. Robust convex optimization. *Mathematics of operations research*, 23(4):769–805, 1998.

Ben-Tal, A. and Nemirovski, A. Robust solutions of uncertain linear programs. *Operations research letters*, 25(1):1–13, 1999.

Benati, S. and Hansen, P. The maximum capture problem with random utilities: Problem formulation and algorithms. *European Journal of Operational Research*, 143(3):518–530, 2002a.

Benati, S. and Hansen, P. The maximum capture problem with random utilities: Problem formulation and algorithms. *European Journal of Operational Research*, 143:518–530, 12 2002b. doi: 10.1016/S0377-2217(01)00340-X.

Bertsimas, D. and Sim, M. The price of robustness. *Operations research*, 52(1):35–53, 2004.

Bertsimas, D., Pachamanova, D., and Sim, M. Robust linear optimization under general norms. *Operations Research Letters*, 32(6):510–516, 2004.

Bertsimas, D., Brown, D. B., and Caramanis, C. Theory and applications of robust optimization. *SIAM review*, 53(3):464–501, 2011.

Blanchet, J., Gallego, G., and Goyal, V. A markov chain approximation to choice modeling. *Operations Research*, 64(4):886–905, 2016.

Chen, X., Krishnamurthy, A., and Wang, Y. Robust dynamic assortment optimization in the presence of outlier customers. *arXiv preprint arXiv:1910.04183*, 2019.

Conn, A. R., Gould, N. I., and Toint, P. L. *Trust region methods*. SIAM, 2000.

Daly, A. and Bierlaire, M. A general and operational representation of generalised extreme value models. *Transportation Research Part B: Methodological*, 40:285–305, 05 2006. doi: 10.1016/j.trb.2005.03.003.

Dam, T. T., Ta, T. A., and Mai, T. Submodularity and local search approaches for maximum capture problems under generalized extreme value models. *European Journal of Operational Research*, 2021.

De Klerk, E. *Aspects of semidefinite programming: interior point algorithms and selected applications*, volume 65. Springer Science & Business Media, 2006.

EI-Ghaoui, L. and Lebret, H. Robust solutions to least-square problems to uncertain data matrices. *Sima Journal on Matrix Analysis and Applications*, 18:1035–1064, 1997.

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

Fischetti, M. and Lodi, A. Local branching. *Mathematical programming*, 98(1):23–47, 2003.

Fosgerau, M., McFadden, D., and Bierlaire, M. Choice probability generating functions. *Journal of Choice Modelling*, 8:1–18, 2013. ISSN 1755-5345. doi: 10.1016/j.jocm.2013.05.002.

Freire, A., Moreno, E., and Yushimito, W. A branch-and-bound algorithm for the maximum capture problem with random utilities. *European Journal of Operational Research*, 252, 12 2015. doi: 10.1016/j.ejor.2015.12.026.

Freire, A. S., Moreno, E., and Yushimito, W. F. A branch-and-bound algorithm for the maximum capture problem with random utilities. *European Journal of Operational Research*, 252 (1):204–212, 2016.

Gallego, G. and Wang, R. Threshold utility model with applications to retailing and discrete choice models. *Available at SSRN 3420155*, 2019.

Haase, K. Discrete location planning. 11 2009.

Haase, K. and Müller, S. A comparison of linear reformulations for multinomial logit choice probabilities in facility location models. *European Journal of Operational Research*, 232, 08 2013. doi: 10.1016/j.ejor.2013.08.009.

Hogan, W. W. Point-to-set maps in mathematical programming. *SIAM review*, 15(3):591–603, 1973.

Hugosson, M. B. and Algers, S. Accelerated introduction of "clean" cars in sweden. In Zachariadis, T. I., editor, *Cars and Carbon*, pages 247–268. Springer Netherlands, 2012.

Krause, A., McMahan, H. B., Guestrin, C., and Gupta, A. Robust submodular observation selection. *Journal of Machine Learning Research*, 9(12), 2008.

Li, X. and Ke, J. Robust assortment optimization using worst-case cvar under the multinomial logit model. *Operations Research Letters*, 47(5):452–457, 2019. ISSN 0167-6377.

Liu, K., Li, Q., and Zhang, Z.-H. Distributionally robust optimization of an emergency medical service station location and sizing problem with joint chance constraints. *Transportation research part B: methodological*, 119:79–101, 2019.

Ljubić, I. and Moreno, E. Outer approximation and submodular cuts for maximum capture facility location problems with random utilities. *European Journal of Operational Research*, 266(1):46–56, 2018.

Lu, M., Ran, L., and Shen, Z.-J. M. Reliable facility location design under uncertain correlated disruptions. *Manufacturing & Service Operations Management*, 17(4):445–455, 2015.

Mai, T. and Jaillet, P. Robust multi-product pricing under general extreme value models. *arXiv preprint arXiv:1912.09552*, 2019.

Mai, T. and Lodi, A. A multicut outer-approximation approach for competitive facility location under random utilities. *European Journal of Operational Research*, 284, 01 2020.

Mai, T., Frejinger, E., Fosgerau, M., and Bastin, F. A dynamic programming approach for quickly estimating large network-based mev models. *Transportation Research Part B*, 98: 179–197, 2017a.

Mai, T., Frejinger, E., Fosgerau, M., and Bastin, F. A dynamic programming approach for quickly estimating large network-based mev models. *Transportation Research Part B: Methodological*, 98:179–197, 2017b.

McCormick, G. P. Computability of global solutions to factorable nonconvex programs: Part i—convex underestimating problems. *Mathematical programming*, 10(1):147–175, 1976.

McFadden, D. Modelling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F., and Weibull, J., editors, *Spatial Interaction Theory and Residential Location*, pages 75–96. North-Holland, Amsterdam, 1978.

McFadden, D. Econometric models of probabilistic choice. In Manski, C. and McFadden, D., editors, *Structural Analysis of Discrete Data with Econometric Applications*, chapter 5, pages 198–272. MIT Press, 1981.

McFadden, D. and Train, K. Mixed MNL models for discrete response. *Journal of applied Econometrics*, pages 447–470, 2000.

Mehmanchi, E., Gillen, C. P., Gómez, A., and Prokopyev, O. A. On robust fractional 0-1 programming. *INFORMS Journal on Optimization*, 2(2):96–133, 2020.

Mirrlees, J. A. An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208, 1971.

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

Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. An analysis of approximations for maximizing submodular set functions—i. *Mathematical programming*, 14(1):265–294, 1978.

Orlin, J. B., Schulz, A. S., and Udwani, R. Robust monotone submodular function maximization. *Mathematical Programming*, 172(1):505–537, 2018.

Rahimian, H. and Mehrotra, S. Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*, 2019.

Rusmevichientong, P. and Topaloglu, H. Robust assortment optimization in revenue management under the multinomial logit choice model. *Operations research*, 60(4):865–882, 2012.

Rusmevichientong, P., Shmoys, D., Tong, C., and Topaloglu, H. Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management*, 23(11):2023–2039, 2014.

Snyder, L. V. Facility location under uncertainty: a review. *IIE transactions*, 38(7):547–564, 2006.

Talluri, K. and Van Ryzin, G. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1):15–33, 2004.

Train, K. *Discrete Choice Methods with Simulation.* Cambridge University Press, 2003.

Vovsha, P. and Bekhor, S. Link-nested logit model of route choice Overcoming route overlapping problem. *Transportation Research Record*, 1645:133–142, 1998.

Wiesemann, W., Kuhn, D., and Sim, M. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

Yan, Z., Natarajan, K., Teo, C. P., and Cheng, C. A representative consumer model in data-driven multiproduct pricing optimization. *Management Science*, 2022.

Zhang, Y., Berman, O., and Verter, V. The impact of client choice on preventive healthcare facility network design. *OR Spectrum*, 34, 04 2012. doi: 10.1007/s00291-011-0280-1.

Appendix

A Proofs

A.1 Proof of Lemma 1

Proof. We will make use of the assumption that $\mathcal{V}_i$ can be defined by a set of constraints $\{g^i_t(\mathbf{v}^i) \le 0;\ t = 1, \ldots, T\}$. The Lagrangian expression of the minimization problem is

$$\mathcal{L}(\mathbf{x}, \mathbf{v}^i, \boldsymbol{\lambda}) = G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x}) + \sum_{t \in [T]} \lambda_t g^i_t(\mathbf{v}^i)$$

Let $\boldsymbol{\lambda}^*$ be the saddle point of Lagrangian (solution that minimize $G^i(\mathbf{Y}(\mathbf{v}^i) \circ \mathbf{x})$ subjects to the constraints), the Envelop theorem (Mirrlees, 1971) implies that

$$\begin{aligned} \frac{\partial \phi^i(\mathbf{x})}{\partial x_j} &= \frac{\partial \mathcal{L}(\mathbf{x}, \mathbf{v}^i, \boldsymbol{\lambda})}{\partial x_j}\bigg|_{(\mathbf{v}^i, \boldsymbol{\lambda}) = (\mathbf{v}^{i*}, \boldsymbol{\lambda}^*)} \\ &= \frac{\partial G^i(\mathbf{Y}(\mathbf{v}^{i*}) \circ \mathbf{x})}{\partial x_j} + \frac{\partial \left(\sum_{t=1}^T \lambda_t g^t(\mathbf{v}^i) \right)}{\partial x_j}\Bigg|_{(\mathbf{v}^i, \boldsymbol{\lambda}) = (\mathbf{v}^{i*}, \boldsymbol{\lambda}^*)} \\ &\overset{(a)}{=} \frac{\partial G^i(\mathbf{Y}(\mathbf{v}^{i*}) \circ \mathbf{x})}{\partial x_j}, \end{aligned} \tag{18}$$

where (a) is due to the fact that $\sum_{t=1}^T \lambda_t g^t(\mathbf{v}^i) = 0$ (complementary slackness from the KKT conditions) does not involve $\mathbf{x}$. We take the second derivative of $\phi^i(\mathbf{x})$ w.r.t. x_k we have

$$\begin{aligned} \frac{\partial^2 \phi^i(\mathbf{x})}{\partial x_j \partial x_k} &= \frac{\partial^2 \mathcal{L}(\mathbf{x}, \mathbf{v}^{i*}, \boldsymbol{\lambda}^*)}{\partial x_j \partial x_k} + \sum_{j \in [m]} \frac{\partial^2 \mathcal{L}(\mathbf{x}, \mathbf{v}^i(\mathbf{x}), \boldsymbol{\lambda}^*)}{\partial x_j \partial \mathbf{v}^i_j(\mathbf{x})} \frac{\partial v^i_j(\mathbf{x})}{\partial x_k} \\ &\quad + \sum_{t \in [T]} \frac{\partial^2 \mathcal{L}(\mathbf{x}, \mathbf{v}^{i*}, \boldsymbol{\lambda}(\mathbf{x}))}{\partial x_j \partial \lambda_t(\mathbf{x})} \frac{\partial \lambda_t(\mathbf{x})}{\partial x_k}, \end{aligned} \tag{19}$$

where $(\mathbf{v}^i(\mathbf{x}), \boldsymbol{\lambda}(\mathbf{x}))$ is the saddle point of the Lagrangian as functions of $\mathbf{x} \in [0,1]^m$. We know that, for any $\mathbf{x} \in [0,1]^m$ and for any $j \in [m], t \in [T]$, the KKT conditions imply that

$$\frac{\partial \mathcal{L}(\mathbf{x}, \mathbf{v}^i(\mathbf{x}), \boldsymbol{\lambda}(\mathbf{x}))}{\partial v^i_j(\mathbf{x})} = 0; \tag{20}$$

and $\lambda_t(\mathbf{x}) g^i_t(\mathbf{v}^i(\mathbf{x})) = 0$. There are two cases to consider here. If $\lambda_t(\mathbf{x}) = 0$ then $\partial \lambda_t(\mathbf{x}) / \partial x_k = 0$ and if $g^i_t(\mathbf{v}^i(\mathbf{x})) = 0$ then

$$\frac{\partial \mathcal{L}(\mathbf{x}, \mathbf{v}^{i*}, \boldsymbol{\lambda})}{\partial \lambda_t}\bigg|_{\lambda = \lambda(\mathbf{x})} = \left(\frac{\partial G^i(\mathbf{Y}(\mathbf{v}^{i*}) \circ \mathbf{x})}{\partial \lambda_t} + g^i_t(\mathbf{v}^{i*}) \right)\bigg|_{\lambda = \lambda(\mathbf{x})} = 0.$$

Thus

$$\left.\frac{\partial^2 \mathcal{L}(\mathbf{x}, \mathbf{v}^i, \boldsymbol{\lambda})}{\partial \lambda_t \partial x_j}\right|_{\lambda=\lambda(\mathbf{x})} = 0.$$

Combine the two cases, we have

$$\sum_{t\in[T]} \frac{\partial^2 \mathcal{L}(\mathbf{x}, \mathbf{v}^{i*}, \boldsymbol{\lambda}(\mathbf{x}))}{\partial x_j \partial \lambda_t(\mathbf{x})} \frac{\partial \lambda_t(\mathbf{x})}{\partial x_k} = 0. \tag{21}$$

Combine (19), (20) and (21) we have

$$\begin{aligned}
\frac{\partial^2 \phi^i(\mathbf{x})}{\partial x_j \partial x_k} &= \frac{\partial^2 \mathcal{L}(\mathbf{x}, \mathbf{v}^{i*}, \boldsymbol{\lambda}^*)}{\partial x_j \partial x_k} \\
&= \frac{\partial^2 G^i(\mathbf{Y}(\mathbf{v}^{i*}) \circ \mathbf{x})}{\partial x_j \partial x_k} + \left.\frac{\partial^2 \left(\sum_{t=1}^T \lambda_t g^t(\mathbf{v}^i)\right)}{\partial x_j \partial x_k}\right|_{(\mathbf{v}^i, \boldsymbol{\lambda})=(\mathbf{v}^{i*}, \boldsymbol{\lambda}^*)} \\
&= \frac{\partial^2 G^i(\mathbf{Y}(\mathbf{v}^{i*}) \circ \mathbf{x})}{\partial x_j \partial x_k}.
\end{aligned} \tag{22}$$

We complete the proof. ■

A.2 Proof of Proposition 2

Proof. Recall that $Y_j^i(v_j^i) = e^{v_j^i}$, then by dividing the numerator and denominator of each fraction by $e^{v_0^i}$, we write the objective function of (3) as

$$\sum_{i\in I} q_i \frac{\sum_{j\in[m]} Y_j^i(v_j^i - v_0^i)\partial G_j^i(\mathbf{Y}(\mathbf{v}^i)|S)}{1 + \frac{1}{e^{v_0^i}} G^i(\mathbf{Y}(\mathbf{v}^i)|S)}. \tag{23}$$

Moreover, from Property (ii) of Remark 1 and (ii) of Proposition 1 we have

$$\begin{aligned}
\partial G_j^i(\mathbf{Y}(\mathbf{v}^i)|S) &= \partial G_j^i(\mathbf{Y}(\mathbf{v}^i)/e^{v_0^i}|S) = \partial G_j^i(\mathbf{Y}(\mathbf{v}^i - v_0^i\mathbf{e})|S) \\
\frac{1}{e^{v_0^i}} G^i(\mathbf{Y}(\mathbf{v}^i)|S) &= G^i(\mathbf{Y}(\mathbf{v}^i)/e^{v_0^i}|S) = G^i(\mathbf{Y}(\mathbf{v}^i - v_0^i\mathbf{e})|S),
\end{aligned}$$

where $\mathbf{e}$ is an all-ones vector of size m. Thus, the inner minimization problem of (3) is equivalent to

$$\min_{\substack{(v_0^i, \mathbf{v}^i)\in\mathcal{V}_i \\ \forall i\in I}} \left\{ \sum_{i\in I} q_i \frac{\sum_{j\in[m]} Y_j^i(\widetilde{v}_j^i)\partial G_j^i(\mathbf{Y}(\widetilde{\mathbf{v}}^i)|S)}{1 + G^i(\mathbf{Y}(\widetilde{\mathbf{v}}^i)|S)} \right\}. \tag{24}$$

where $\widetilde{\mathbf{v}}^i$ is a vector of size m with entries $\widetilde{v}_j^i = v_j^i - v_0^i$, for all $i \in I$. Now, by defining new uncertainty sets $\widetilde{\mathcal{V}}_i$, $\forall i \in I$, we can write (24) equivalently as

$$\min_{\substack{\widetilde{\mathbf{v}}^i\in\widetilde{\mathcal{V}}_i \\ \forall i\in I}} \left\{ \sum_{i\in I} q_i \frac{\sum_{j\in[m]} Y_j^i(\widetilde{v}_j^i)\partial G_j^i(\mathbf{Y}(\widetilde{\mathbf{v}}^i)|S)}{1 + G^i(\mathbf{Y}(\widetilde{\mathbf{v}}^i)|S)} \right\},$$

as desired. ∎

A.3 Proof of Proposition 3

Proof. It is more convenient to use the binary representation $G^i(\mathbf{Y}(\mathbf{v}^i)\circ\mathbf{x})$ to prove the claim. To prove the convexity, we take the second derivatives of $G^i(\mathbf{Y}(\mathbf{v}^i)\circ\mathbf{x})$ w.r.t. $\mathbf{v}^i$ and prove that the Hessian matrix is positive definite. Let $\rho(\mathbf{v}^i) = G^i(\mathbf{Y}(\mathbf{v}^i)\circ\mathbf{x})$. We take the first and second order derivatives of $\rho(\mathbf{v}^i)$ with respect to $\mathbf{v}^i$ and have

$$\frac{\partial\rho(\mathbf{v}^i)}{\partial v^i_j} = \frac{\partial G^i(\mathbf{Y}(\mathbf{v}^i)\circ\mathbf{x})}{\partial v^i_j} = (x_jY_j)\partial G^i_j(\mathbf{Y}(\mathbf{v}^i)\circ\mathbf{x}),\ \forall j\in[m]$$
$$\frac{\partial^2\rho(\mathbf{v}^i)}{\partial v^i_j\partial v^i_k} = (x_jY_j)(x_kY_k)\partial G^i_{jk}(\mathbf{Y}(\mathbf{v}^i)\circ\mathbf{x}),\ \forall j,k\in[m], j\neq k$$
$$\frac{\partial^2\rho(\mathbf{v}^i)}{\partial v^i_j\partial v^i_j} = (x_jY_j)\partial G^i_j(\mathbf{Y}(\mathbf{v}^i)\circ\mathbf{x}) + (x_jY_j)^2\partial G^i_{jj}(\mathbf{Y}(\mathbf{v}^i)\circ\mathbf{x}),\ \forall j\in[m].$$

So, we define $\overline{\mathbf{Y}}\in\mathbb{R}^m$ such that $\overline{Y}_j = e^{v^i_j}x_j$ and write

$$\nabla^2\rho(\mathbf{v}^i) = \text{diag}(\overline{\mathbf{Y}})\nabla^2 G^i(\overline{\mathbf{Y}})\text{diag}(\overline{\mathbf{Y}}) + \text{diag}(\nabla G^i(\overline{\mathbf{Y}})\circ\overline{\mathbf{Y}})$$

where $\nabla^2 G^i(\overline{\mathbf{Y}})$ is a matrix of size $(m\times m)$ with entries $\partial G^i_{jk}(\overline{\mathbf{Y}})$, for all $j,k\in[m]$, and $\text{diag}(\overline{\mathbf{Y}})$ is the square diagonal matrix with the elements of vector $\overline{\mathbf{Y}}$ on the main diagonal. We see that $\text{diag}(\nabla G^i(\overline{\mathbf{Y}})\circ\overline{\mathbf{Y}})$ is positive definite. Moreover, $\text{diag}(\overline{\mathbf{Y}})\nabla^2 G^i(\overline{\mathbf{Y}})\text{diag}(\overline{\mathbf{Y}})$ is symmetric and its (j,k)-th element is $\overline{Y}_j\overline{Y}_k\partial G^i_{jk}(\overline{\mathbf{Y}})$. For $j\neq k$ we have $\partial G^i_{jk}(\overline{\mathbf{Y}})\leq 0$ (Property *(iv)* of Remark 1), thus all the off-diagonal entries of the matrix are non-positive. Moreover, from Property *(iii)* of Proposition 1, we see that $\sum_{k\in[m]}\overline{Y}_j\overline{Y}_k\partial G^i_{jk}(\overline{\mathbf{Y}}) = 0$ for any $j\in[m]$, thus each row of the matrix sums up to zero. Using Theorem A.6 of De Klerk (2006), we see that $\text{diag}(\overline{\mathbf{Y}})\nabla^2 G^i(\overline{\mathbf{Y}})\text{diag}(\overline{\mathbf{Y}})$ is positive semi-definite. Since $\text{diag}(\nabla G^i(\overline{\mathbf{Y}})\circ\overline{\mathbf{Y}})$ is positive definite, $\nabla^2\rho(\mathbf{v}^i)$ is positive definite, implying that $\rho(\mathbf{v}^i)$ is strictly convex in $\mathbf{v}^i$, as desired. ∎

B Comparison with Mehmanchi et al. (2020)

Mehmanchi et al. (2020) study a robust fractional 0-1 program based on the uncertainty structure introduced by (Bertsimas and Sim, 2004, Bertsimas et al., 2004), which can be well applied to the MCP under MNL. The work of Mehmanchi et al. (2020) differs from our robust MCP under MNL by the fact that our methods work with any convex uncertainty sets while Mehmanchi et al. (2020) employ rectangular uncertainty sets where each unknown coefficient (i.e., utility in the MCP context) lies in a symmetric interval centered on a nominal value. Thus, our algorithms (i.e., both the local search and outer-approximation) can work well with their uncertainty setting. In this section, we consider a robust MCP under the Mehmanchi et al. (2020)'s uncertainty setting and make a numerical comparison between our algorithms and their solution approach which is based on a MILP reformulation.

Under Mehmanchi et al. (2020)'s uncertainty setting, each uncertainty set $\mathcal{V}_i$ is defined as $\mathcal{V}_i = \{\mathbf{V}^i | V_{ij} \in [\overline{V}_{ij} - d_{ij}; \overline{V}_{ij} + d_{ij}],\ |S_i(\mathbf{V})| \leq \Gamma_i\}$ where $\overline{V}_{ij}$ are nominal utilities, coefficients $d_{ij} \geq 0$ denote potential deviation from the nominal utilities, $S_i(\mathbf{V})$ is the set of indices of the uncertain parameters $\mathbf{V}$ whose values are different from the nominal values $\overline{\mathbf{V}}$. The constraints $S_i(\mathbf{V}) \leq \Gamma_i$ imply that there are at most Γ_i unknown coefficients taking values different from their nominal values. The robust MCP under MNL can be formulated as:

$$\max_{\mathbf{x}\in\{0,1\}^m} \left\{ f^{\mathrm{WC}}(\mathbf{x}) = \sum_{i\in I} \min_{\mathbf{V}^i\in\mathcal{V}_i} \left\{ q_i - \sum_{i\in I} \frac{q_i}{1+\sum_j V_{ij}x_j} \right\} \right\} \tag{RO*}$$
$$\text{subject to} \sum_{j\in[m]} x_j = C.$$

Since the inner problem is a minimization problem, we want to maximize the number of times that V_{ij} is equal to its lower bound. Thus, we can use additional binary variables $u_{ij} \in \{0,1\},\ \forall i \in [I], j \in [m]$ to write the MCP problem as:

$$\max_{\mathbf{x}\{0,1\}^m} \left\{ f^{\mathrm{WC}}(\mathbf{x}) = \sum_{i\in I} q_i - \sum_{i\in I} \frac{q_i}{1+\sum_j \overline{V}_{ij}x_j - \max_{\substack{u_{ij},\ \forall j\in[m] \\ \sum_j u_{ij}\leq\Gamma_i}} \left\{ \sum_j d_{ij}x_j u_{ij} \right\}} \right\}$$
$$\text{subject to} \sum_{j\in[m]} x_j = C.$$

To reformulate the above fractional program as a MILP, we let

$$y_i = 1/\left(1+\sum_j \overline{V}_{ij}x_j - \max_{\substack{u_{ij}\in\{0,1\} \\ \sum_j u_{ij}\leq\Gamma_i}} \left\{ \sum_j d_{ij}x_j u_{ij} \right\}\right)$$

and denote $z_{ij} = y_i x_j$. We further linearize these bi-linear terms using McCormick inequalities

(McCormick, 1976) to obtain:

$$
\begin{aligned}
\max_{\mathbf{x}\in\{0,1\}^m} \quad & \left\{ f^{\mathrm{WC}}(\mathbf{x}) = \sum_{i\in I} q_i - \sum_{i\in I} q_i y_i \right\} \\
\text{subject to} \quad & \sum_{j\in[m]} x_j = C \\
& y_j + \sum_j \overline{V}_{ij} z_{ij} - \max_{\substack{u_{ij}\in\{0,1\} \\ \sum_j u_{ij}\le\Gamma_i}} \left\{ \sum_j d_{ij} z_{ij} u_{ij} \right\} \ge 1 \\
& z_{ij} \ge y_i - \overline{y}_i(1-x_j) && \forall i\in I, \forall j\in[m] \\
& z_{ij} \le y_i + \underline{y}_i(x_j-1) && \forall i\in I, \forall j\in[m] \\
& z_{ij} \le \overline{y}_i x_j && \forall i\in I, \forall j\in[m] \\
& z_{ij} \ge \underline{y}_i x_j && \forall i\in I, \forall j\in[m] \\
& y_j \ge 0, z_{ij} \ge 0,\ \forall i\in I, \forall j\in[m],
\end{aligned}
$$

where $\overline{y}_i$, $\underline{y}_i$, $\forall i\in I$, are some upper and lower bounds of y_i, respectively. Using Equation (9) of Mehmanchi et al. (2020), we take the dual of the inner maximization problem and formulate the robust MCP as the following MILP:

$$
\begin{aligned}
\max_{\mathbf{x}} \quad & \left\{ f^{\mathrm{WC}}(\mathbf{x}) = \sum_{i\in I} q_i - \sum_{i\in I} q_i y_i \right\} && \text{(MILP*)} \\
\text{subject to} \quad & \sum_{j\in[m]} x_j = C \\
& y_j + \sum_j \overline{V}_{ij} z_{ij} - 1 \ge \Gamma_i \alpha + \sum_j p_j \\
& p_j + \alpha \ge d_{ij} z_{ij} \\
& z_{ij} \le y_i + \underline{y}_i(x_j-1) && \forall i\in I, \forall j\in[m] \\
& z_{ij} \le \overline{y}_i x_j && \forall i\in I, \forall j\in[m] \\
& z_{ij} \ge \underline{y}_i x_j && \forall i\in I, \forall j\in[m] \\
& \alpha, p_j, y_j, z_{ij} \ge 0, && \forall i\in I, \forall j\in[m],
\end{aligned}
$$

where α and p_j, $j\in[m]$ are dual variables of the inner maximization problem.

To conduct the experiment, we take instances from the three datasets HM, ORlib, and NYC. We choose the deviation coefficients as $d_{ij} = 0.5\overline{V}_{ij}$, $\forall i\in I, j\in[m]$ and the level of uncertainty $\Gamma_i \in [1,2,3,4,5]$ and $C=5$. We set a time limit of 600 seconds for all the methods and select lower and uppers bounds for y_i as $\overline{y}_i = 1$ and $\underline{y}_i = 0$, for all $i\in I$, similarly to the setup in Mehmanchi et al. (2020). We will compare (MILP*) against our local search algorithm (i.e. Algorithm 1), noting that the MOA can be used as well but it is generally less efficient than the local search algorithm, in terms of both computing time and solution quality. For each step of LS, we compute the worst-case objective value by solving the inner maximization

$\max_{\substack{u_{ij}\in\{0,1\} \\ \sum_j u_{ij}\leq\Gamma_i}} \left\{\sum_j d_{ij}x_j u_{ij}\right\}$ ($\mathbf{x}$ is fixed). This can be done efficiently by relaxing the binary variables u_{ij} and using CPLEX to solve the resulting linear program. Here, we only focus on comparison in terms of solving the robust problem, as the value and price of robustness under this uncertainty setting were intensively assessed in previous work (Bertsimas and Sim, 2004, Mehmanchi et al., 2020).

Table 3 below reports the performance of (MILP*) solved by CPLEX and our LS algorithm, in terms of running time and the number of times each method returns better objective values within the time budget. The results generally show that LS outperforms (MILP*) in terms of solution quality and is also faster, especially for large-sized instances.

Dataset	$\lvert I\rvert$	m	# instances with better objectives		Average running time (seconds)	
			MILP*	LS (ours)	MILP*	LS (ours)
HM14	50	25	5	5	0.4	12.1
HM14	50	50	5	5	1.1	9.6
HM14	50	100	5	5	4.3	12.7
HM14	100	25	5	5	1.0	11.7
HM14	100	50	5	5	3.8	11.7
HM14	100	100	5	5	13.5	13.6
HM14	200	25	5	5	4.8	14.9
HM14	200	50	5	4	22.4	16.3
HM14	200	100	5	3	152.4	21.7
HM14	400	25	5	5	14.9	15.5
HM14	400	50	5	5	78.4	18.9
HM14	400	100	3	4	600	24.0
HM14	800	25	5	5	136.2	17.9
HM14	800	50	2	4	600	15.1
HM14	800	100	2	3	600	22.7
ORlib	50	25	5	4	27.1	11.8
ORlib	50	25	5	4	27.1	10.2
ORlib	50	25	5	5	25.4	11.5
ORlib	50	25	5	5	25.6	10.2
ORlib	50	50	5	4	600	12.4
ORlib	50	50	5	4	600	13.7
ORlib	50	50	5	4	600	12.5
ORlib	50	50	5	5	600	13.8
ORlib	1000	100	0	5	600	16.3
ORlib	1000	100	0	5	600	14.3
ORlib	1000	100	0	5	600	15.0
NYC	82341	59	0	5		
Average			4.0	4.6		

Table 3: Comparison between the MILP approach proposed in Mehmanchi et al. (2020) and Algorithm 1 for the MCP under MNL.

C Additional Experiments

C.1 Comparing across the MNL and Nested Logit Models

We will assess the performance of the RO approach across the MNL and nested logit instances. The aim is to see how the RO performs, as compared to the other approaches, when the correlation structure of the choice model changes. To this end, we will provide results under the following two settings. First, we solve the robust MCP problem under MNL and nested logit by Algorithm 2 and 1 and compare the obtained solutions by injecting them into the corresponding nested logit instances and comparing the corresponding distributions of the objective values, in a similar way as in the above sections. The aim is to evaluate gains that we can get if the choice model is well-specified, under our robust settings. Second, we solve MNL instances by the RO, DET1, DET2 and SA approaches. We then take the obtained solutions to test on the corresponding nested logit instances in the same manner as in the previous sections. In other words, we solve the robust MCP problem under MNL but test the solutions obtained on nested logit instances. By doing this, we aim to explore how different approaches protect us from worst-case scenarios when the choice model is misspecified.

For the first setting, we plot in Figure 8 the distributions of the 2000 samples of the objective values given by MNL and nested-logit solutions with $\epsilon \in \{0.02, 0.04, 0.08, 0.4, 0.5, 0.6\}$, for instances of size $|I| = 100$ and $m = 50$. The figure clearly shows that the nested-logit solutions always return better objective values than those given by the MNL solutions. Moreover, the difference seems smaller as ϵ increases.

For the second setting, in Table 4, we report the averages of 2% of the worst-case nested logit objective values for each ϵ value. It is quite clear that RO is not able to retain the same advantages as in the cases considered above, especially for some large-scale instances where the nested correlation structures become more complex. The SA and DET2 approaches seem to provide better protection in this case, suggesting that more investigations would be needed if the correlation structure of the choice model is not well specified.

In summary, if the choice model is well-specified in terms of correlation structure, our results show gains from our robust model in protecting decision-makers from expected user demand that would be too low. The histograms given by the robust approach have high peaks, small variances, and high worst-case values, as compared to their deterministic and sampling-based counterparts. The sampling-based SA approach can provide some protection as well, but it is much more expensive than the RO. If the correlation of the choice model is not well specified, we observe that the RO is limited in maintaining the same advantage, suggesting that more investigations, or possibly, new robust models would be needed to address such a misspecification issue. This is out-of-scope of the paper and we keep this for future work.

$\lvert I \rvert$	m	ϵ	RO	DET1	DET2	SA
100	50	0.02	73.48	73.48	73.85	**73.87**
100	50	0.04	73.45	73.45	**73.82**	72.95
100	50	0.08	71.62	73.38	**73.73**	72.91
100	50	0.40	73.27	72.80	73.13	**73.53**
100	50	0.50	**73.29**	72.78	73.13	71.91
100	50	0.60	**73.37**	72.79	73.16	72.55
100	100	0.02	69.51	69.51	**73.12**	**73.12**
100	100	0.04	69.50	69.50	**73.09**	**73.09**
100	100	0.08	69.47	69.47	**73.05**	**73.05**
100	100	0.40	70.48	69.34	72.78	**72.94**
100	100	0.50	70.51	69.36	**72.81**	71.01
100	100	0.60	70.55	69.40	72.85	**72.93**
200	100	0.02	140.63	140.63	140.63	140.63
200	100	0.04	140.56	140.56	140.56	140.56
200	100	0.08	**140.41**	140.38	140.38	134.13
200	100	0.40	**145.09**	139.10	139.10	133.83
200	100	0.50	145.15	138.89	138.89	140.17
200	100	0.60	**140.23**	138.71	138.71	134.11
1000	100	0.02	38104.29	38266.32	38104.29	38104.29
1000	100	0.04	38099.74	38260.89	38099.74	**38831.59**
1000	100	0.08	38087.11	38244.76	38087.11	**38823.69**
1000	100	0.40	38281.47	38135.82	38008.28	**38963.94**
1000	100	0.50	38291.25	38124.55	38003.93	**38362.69**
1000	100	0.60	38300.33	38114.96	38002.34	**39164.15**
82341	59	0.02	68497.10	**70338.96**	**70338.96**	68497.10
82341	59	0.04	68442.18	**70246.72**	**70246.72**	68442.18
82341	59	0.08	68345.19	**70076.93**	**70076.93**	68345.19
82341	59	0.40	67974.98	68772.24	68772.24	**69141.38**
82341	59	0.50	67999.14	**68579.18**	**68579.18**	67999.14
82341	59	0.60	68069.08	**68442.86**	**68442.86**	68069.08

Table 4: Comparison of nested logit objective values given by solutions obtained by solving MNL instances.

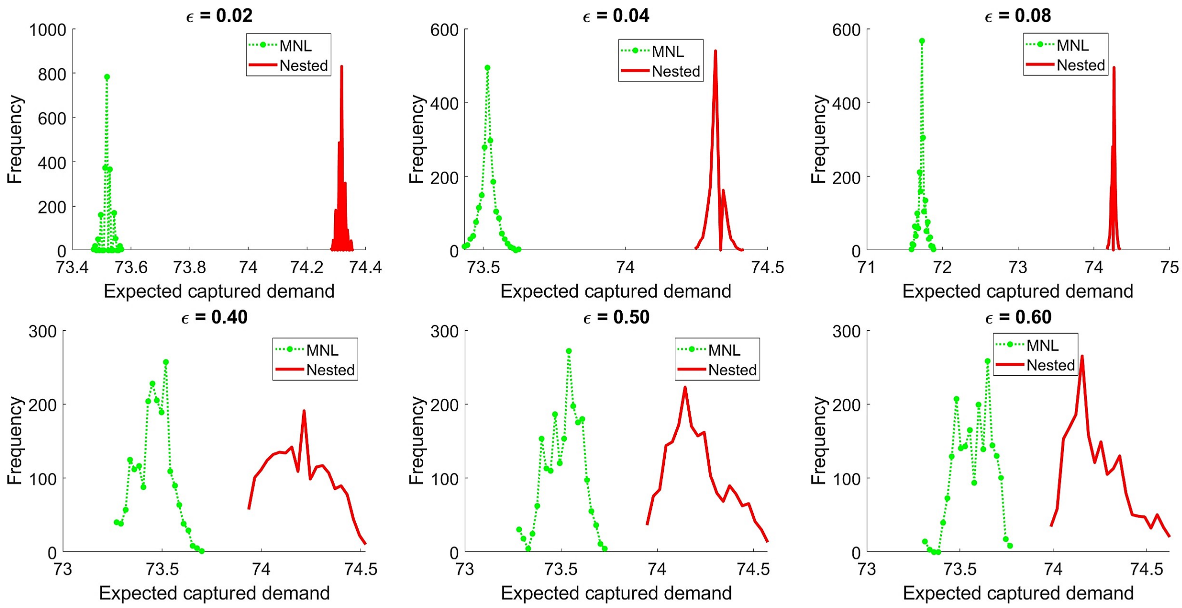


Figure 8: Comparison of the distributions of the *nested-logit* objective values given by solutions from the MNL and the nested logit model for instance of $|I| = 100$ and $m = 50$.

C.2 Objective Value Distributions

Instances of $|I| = 100$ and $m = 100$. Figure 9 shows the histograms given by the four approaches with MNL instances. When ϵ is small, the histograms are very similar across the four approaches. When $\epsilon \geq 0.4$, in analogy to the experiments shown in the main part of the paper, the histograms given by the RO solutions have small variance and shorted left tails, indicating the capability of RO to cover too-low expected demand values.

Figure 10 shows comparison results for nested logit instances, where the histograms from the four approaches are identical for $\epsilon \leq 0.08$. When $\epsilon \geq 0.4$, the superiority of RO in terms of worst-case protection becomes much clearer. We note that RO seems to give better protection with nested logit instances, as compared to the MNL instances shown in Figure 9.

In Figure 11 we plot the percentile ranks of the RO's worst-case objective values in the distributions given by the other approaches. We see that the ranks only become significant with $\epsilon > 0.4$ for MNL instances and with $\epsilon > 0.2$ for nested logit instances. The solutions given by DET1 and DET2 are identical for these testing instances. We see that the percentile ranks are higher for nested logit instances, indicating that the RO would give better protection under the nested logit.

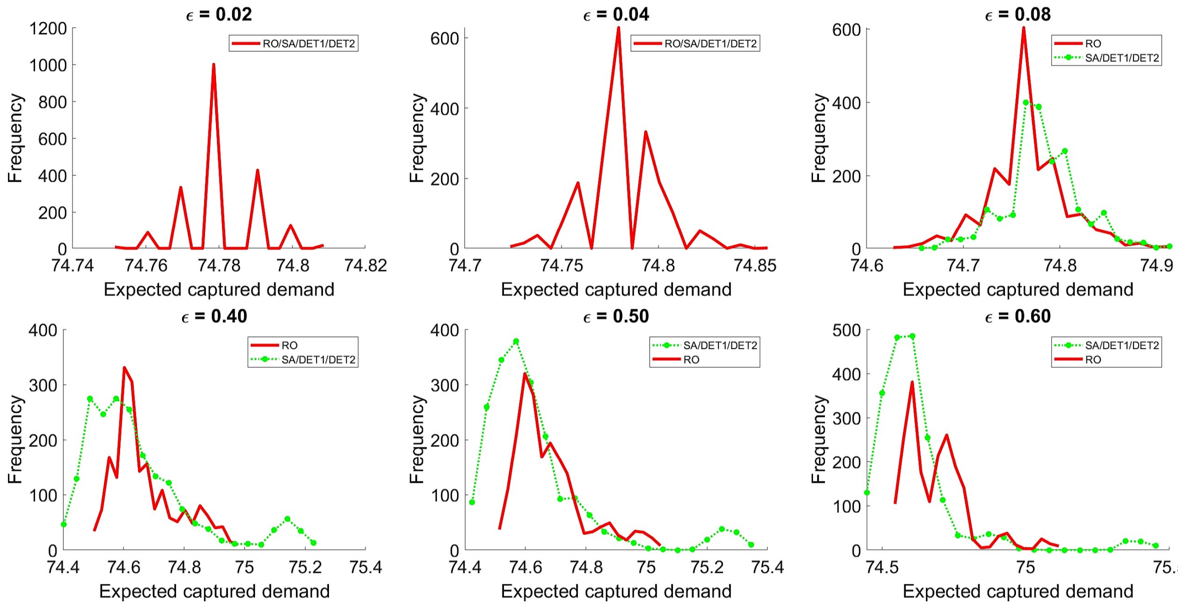


Figure 9: Comparison results between the distributions of the objective values given by solutions from the RO, SA, DET1, and DET2 approaches, under the MNL choice model and with instances of size $|I| = 100$ and $m = 100$.

Instances of $|I| = 200$ and $m = 100$. Figure 12 shows the histograms given by the four approaches for MNL instances, where the ability of RO to provide protection is clearly shown. When $\epsilon \geq 0.4$, the histograms given by RO always have higher peaks, shorter tails, and higher worst-case objective values. The trade-off of being robust can also be seen. Figure 13 shows the histograms for nested logit instances. Figure 14 shows the percentile ranks of the RO's worst-case objective values. In analogous to the previous experiments, the ability of RO in protecting the decision-maker from low objective values is clearly demonstrated.

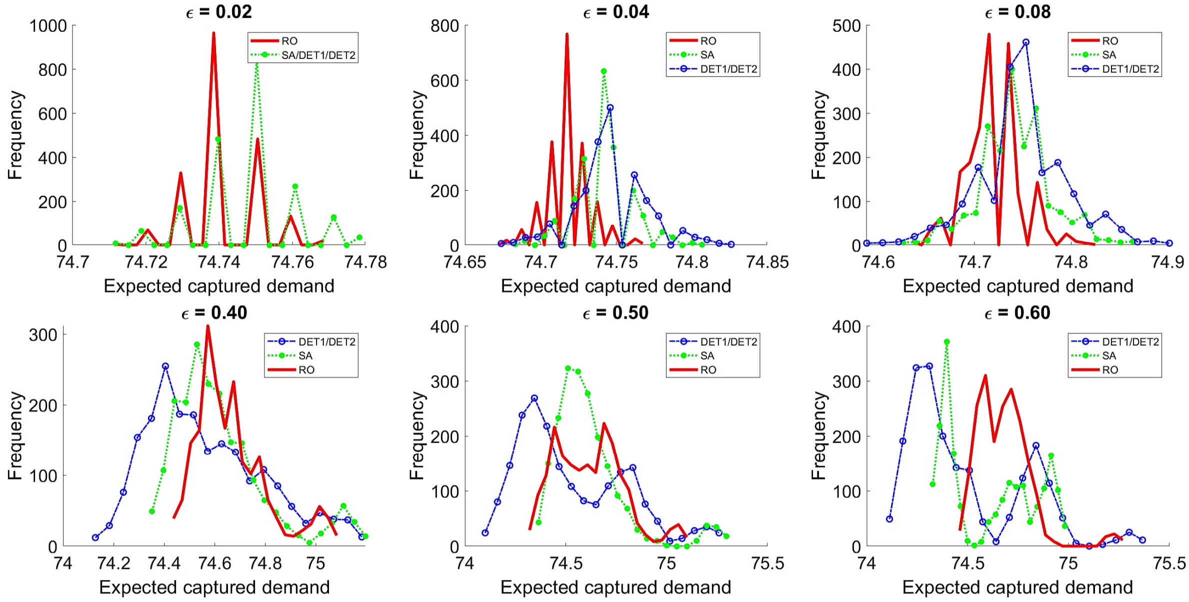


Figure 10: Comparison of the distributions of the objective values given by solutions from RO, SA, DET1, and DET2 approaches, under the nested choice model and with instances of size $|I| = 100$ and $m = 100$.

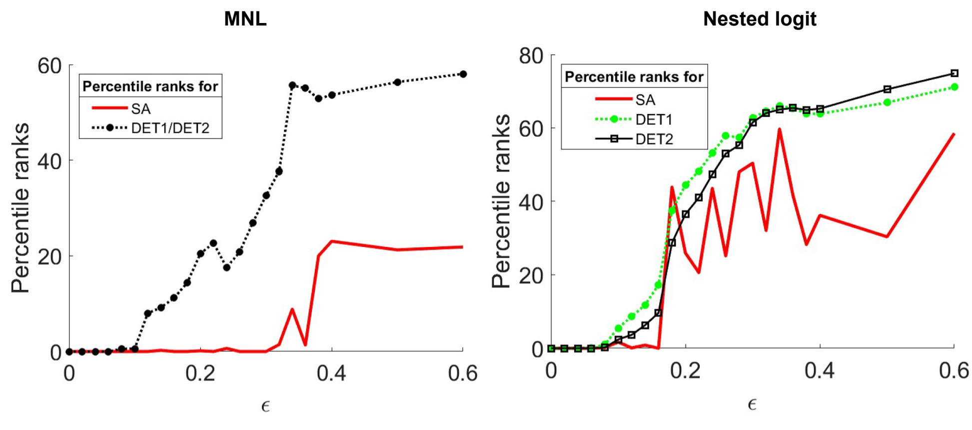


Figure 14: The percentile ranks of RO worst value in the distributions given by the SA, DET1, and DET2 solutions under the nested logit choice model for instance with $|I| = 200$ and $m = 100$.

Instances of $|I| = 1000$ and $m = 100$. Figure 15 shows the histograms given by the four approaches under the MNL choice model. The difference between the histograms of the four

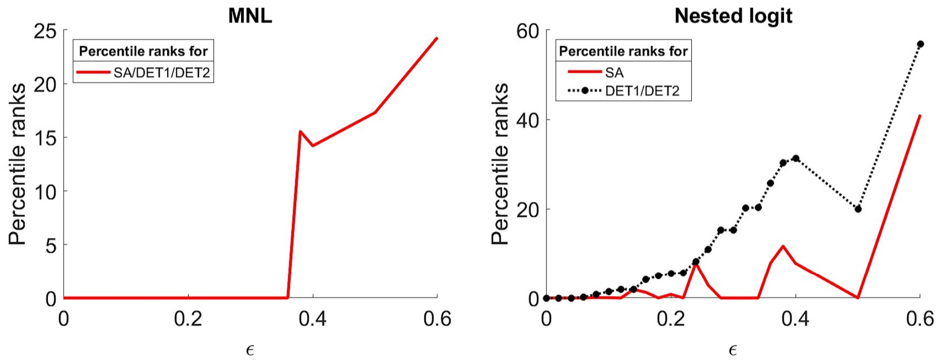


Figure 11: The percentile ranks of RO worst value in the distributions given by the SA, DET1, and DET2 solutions under the MNL choice model for instance with $|I| = 100$ and $m = 100$.

approaches is not clear when ϵ is small. With $\epsilon \geq 0.4$, the histograms given by the RO solutions have very small variance, shorted left tails, and particularly high peaks, as compared to the other approaches. Figure 16 below shows the histograms given by the four approaches under the nested logit choice model. The histograms of RO, SA, and DET1 are similar at $\epsilon \in \{0.02, 0.04\}$. When ϵ becomes larger, RO gives histograms of higher peaks, smaller variances, and shorter tails, as compared to other approaches. It clearly shows the ability of RO in protecting the decision-makers against worse-case situations. In Figure 17, we plot the percentile ranks of the RO's worst objective value. We see that the percentile ranks only become significant when $\epsilon > 0.26$ for the MNL and $\epsilon > 0.2$ for the nested logit instances.

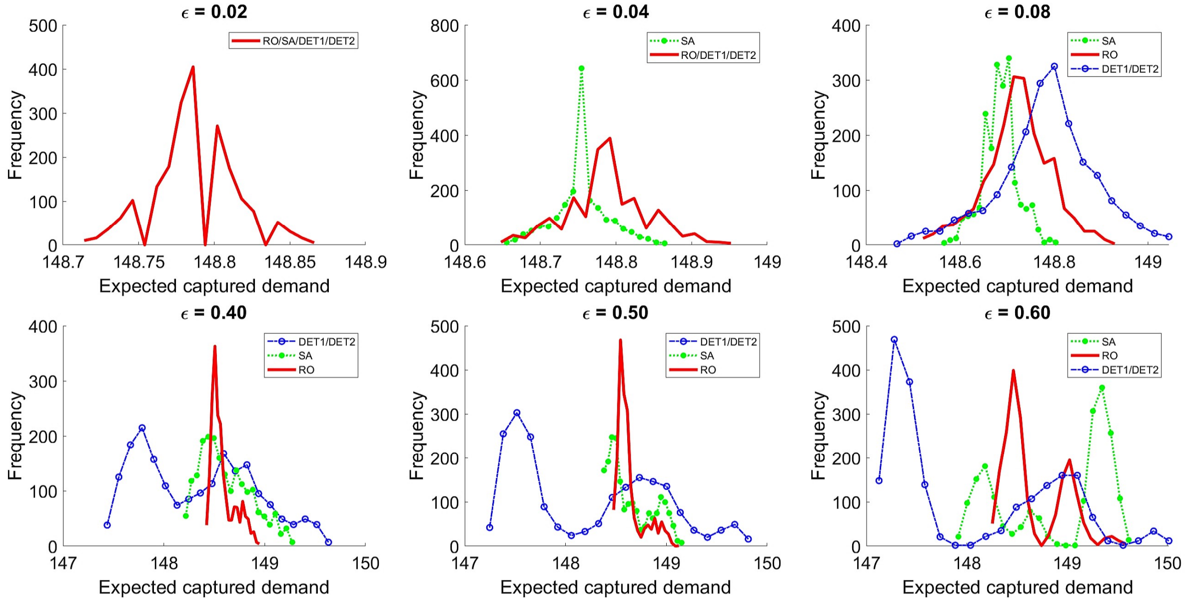
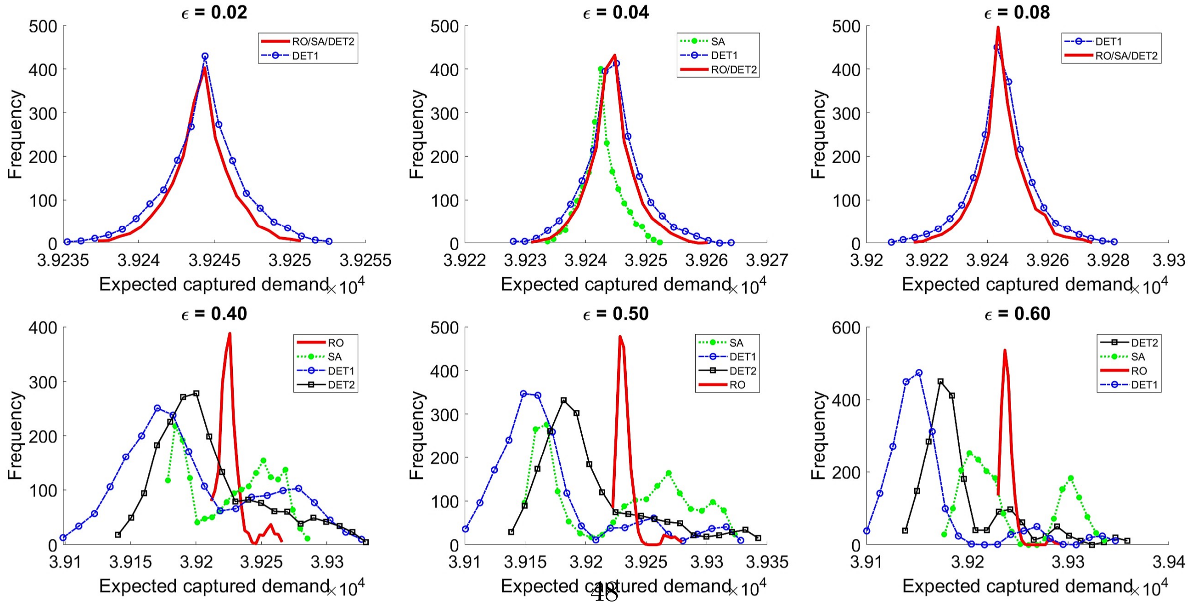


Figure 12: Comparison between the distributions of the objective values given by solutions from RO, SA, DET1, and DET2 approaches, under the MNL choice model and with instances of size $|I| = 200$ and $m = 100$.



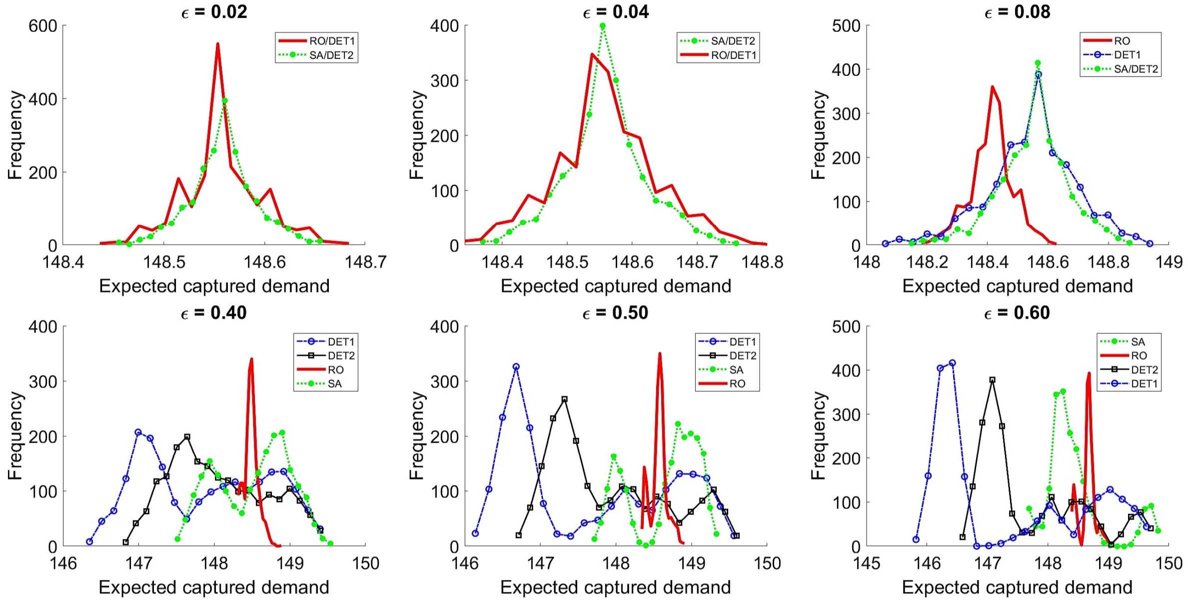
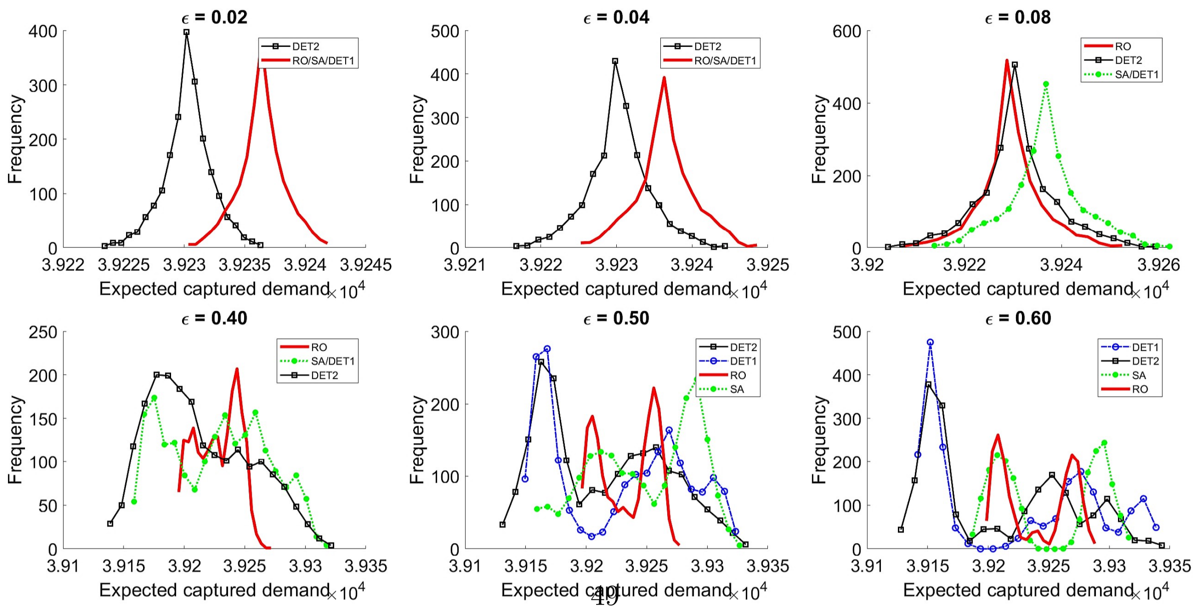


Figure 13: Comparison between the distributions of the objective values given by solutions from RO, SA, DET1, and DET2 approaches, under the nested logit choice model and with instances of size $|I| = 200$ and $m = 100$.



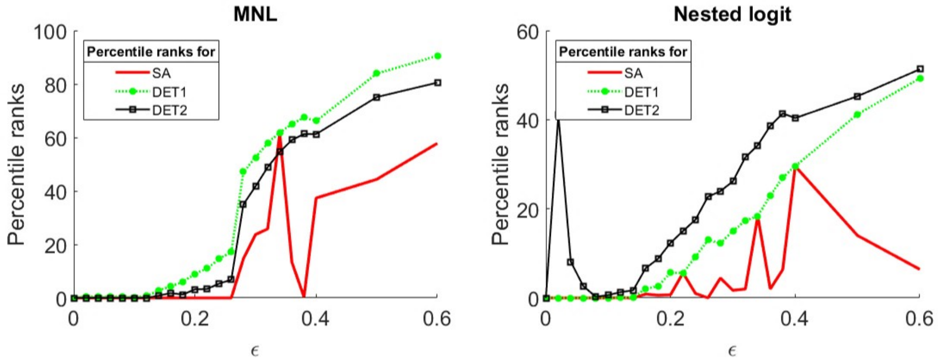


Figure 17: The percentile ranks of RO worst value in the distributions given by the SA, DET1, and DET2 solutions under the nested logit choice model for instance with $|I| = 1000$ and $m = 100$.

Instances of $|I| = 82341$ **and** $m = 59$ In Figures 18, 19, and 20, we plot similar figures as in the previous sections for large-scale instances of size $|I| = 82341$ and $m = 59$, which give analogous observations, i.e., the histograms of RO always have lower variances and shorter tails and the percentiles ranks of the RO's worst-case objective values are always significant, especially when ϵ becomes large.

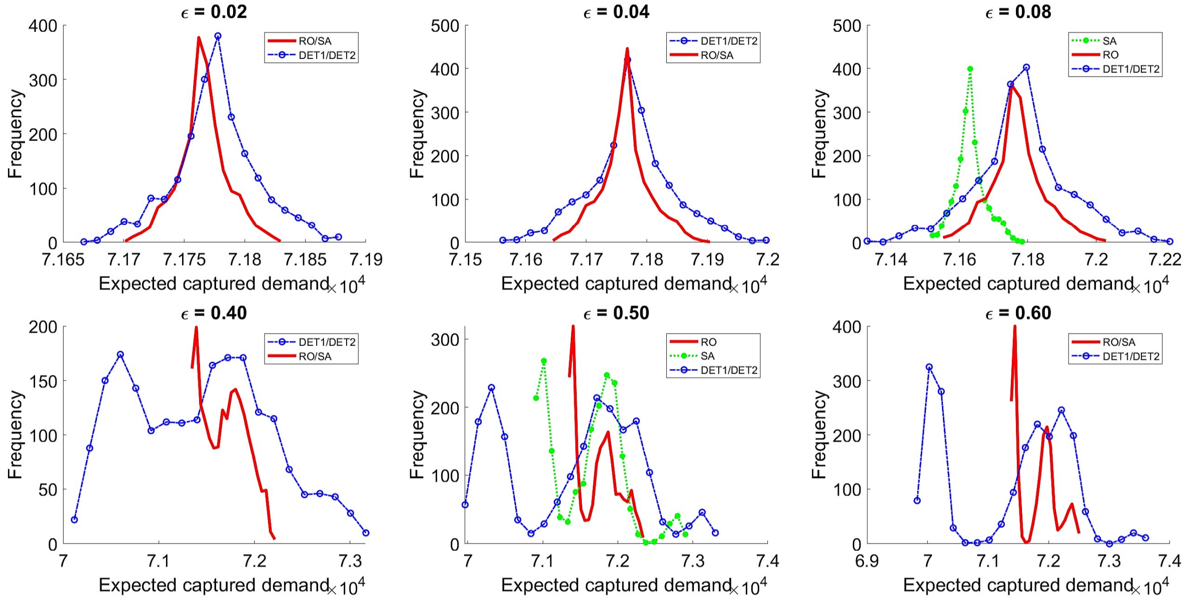


Figure 18: Comparison between the distributions of the objective values given by solutions from RO, SA, DET1, and DET2 approaches, under the MNL choice model and with instances of size $|I| = 82341$ and $m = 59$.

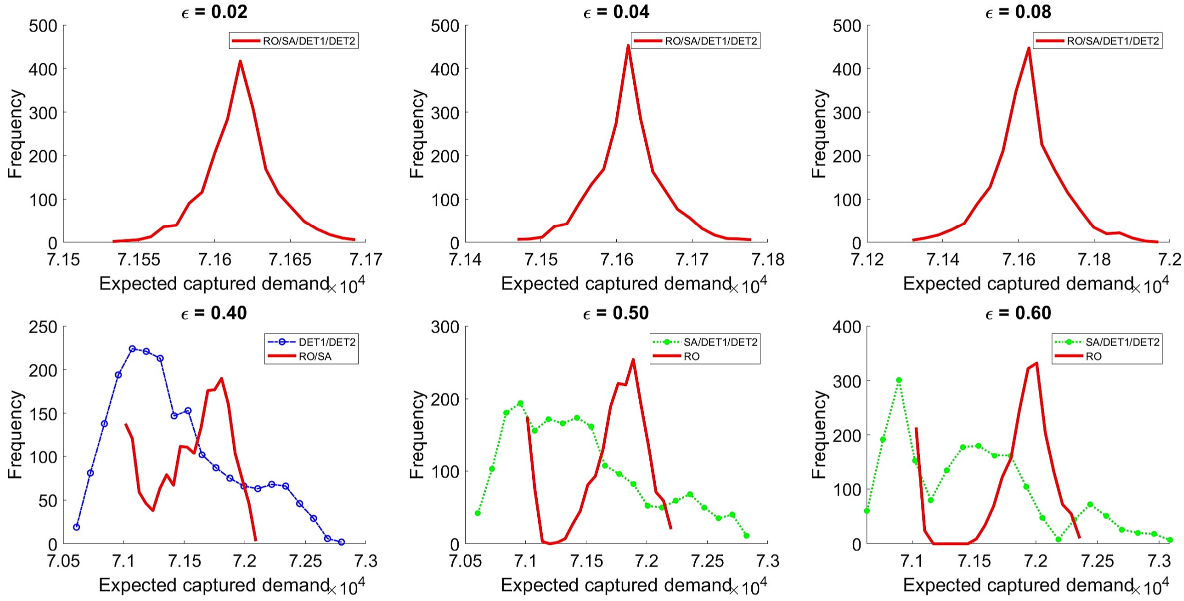


Figure 19: Comparison between the distributions of the objective values given by solutions from RO, SA, DET1, and DET2 approaches, under the nested logit choice model and with instances of size $|I| = 82341$ and $m = 59$.

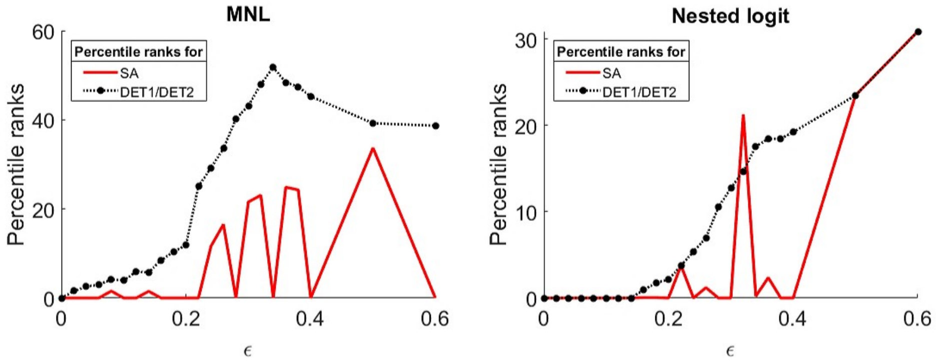


Figure 20: The percentile ranks of RO worst value in the distributions given by the SA, DET1, and DET2 solutions under the nested logit choice model for instance with $|I| = 82341$ and $m = 59$.