Article submitted

Robust Multi-product Pricing under General Extreme Value Models

Tien Mai
Singapre-MIT Alliance for Research and Technologies (SMART), mai.tien@smart.mit.edu

Patrick Jaillet
EECS, Massachusetts Institute of Technologies, jaillet@mit.edu

We study robust versions of pricing problems where customers choose products according to a general extreme value (GEV) choice model, and the choice parameters are not given exactly but lie in an uncertainty set. We show that, when the robust problem is unconstrained and the price sensitivity parameters are homogeneous, the robust optimal prices have a constant markup over products and we provide formulas that allow to compute this constant markup by binary search. We also show that, in the case that the price sensitivity parameters are only homogeneous in each subset of the products and the uncertainty set is rectangular, the robust problem can be converted into a deterministic pricing problem and the robust optimal prices have a constant markup in each subset, and we also provide explicit formulas to compute them.

For constrained pricing problems, we argue that the formulation where the aim is to find purchase probabilities that maximize the expected revenue while satisfying some expected sale constraints, even-though convenient to use when the choice parameters are exactly known, is not appropriate in our uncertainty setting, as there may be no fixed prices under which the resulting purchase probabilities always satisfy the expected sale constraints when the choice parameters vary in an uncertainty set. Thus, we propose an alternative formulation where, instead of requiring that the expected sale constraints be satisfied, we add a penalty cost to the objective function for violated constraints. We then show that the robust pricing problem with over-expected-sale penalties can be reformulated as a convex optimization program where the purchase probabilities are the decision variables. We provide numerical results for the logit and nested logit model to illustrate the advantages of our approach. Our results generally hold for any arbitrary GEV model, including the multinomial logit, nested or cross-nested logit.

Key words: Robust optimization, multi-product pricing, general extreme value model

1. Introduction

In revenue management, pricing is an important problem that refers to the selection of prices for a set of products in order to maximize an expected revenue. This is motivated by the fact that prices are key features that may significantly affect demand for products. The literature of multi-product pricing has seen a large amount of studies focusing on how to set prices when customers purchase products according to a discrete choice model (e.g. Talluri and Van Ryzin 2004, Gallego and Wang 2014, Zhang et al. 2018). In general, finding prices is challenging as the discrete choice model becomes more complicated. This is a well-known trade-off between choice model complexity and operational tractability in revenue management. In a recent work, Zhang et al. (2018) study the pricing problem under the general extreme value (GEV) family of choice models (McFadden 1980). Their results are general, as they apply to any choice model in the GEV family, e.g., the multinomial logit (MNL) and nested logit models, to name a few. To the best of our knowledge, existing studies all assume that the parameters of the choice models are known in advance or can be estimated exactly from data. Thus, the corresponding pricing optimization models are built based on the pre-determined parameters and ignore any uncertainty associated with the estimates. Nevertheless, in practice, the parameter estimates may vary significantly for different customer types or in different purchasing periods of the year. Thus, ignoring such uncertainties may lead to bad pricing decisions.

In this paper, we formulate pricing optimization models explicitly taking into consideration uncertainties occurring in the determination of choice parameters when customers make purchases according to choice models in the GEV family. That is, we assume customers' behavior is driven by any choice model in the GEV family such as the MNL or nested logit model, and the parameters of the choice model are not given exactly but belong to an uncertainty set. In other words, we consider robust versions of the pricing problem under GEV models considered in Zhang et al. (2018). The goal here is to maximize the worst-case expected revenue when the estimates vary in their support set. We consider both unconstrained and constrained problems where the price sensitivity parameters (PSP) are homogeneous or partition-wise homogeneous, i.e., the set of products can be separated into disjoint subsets and the PSP are the same in each subset but can be different over subsets. Our results for unconstrained problems hold for any choice model in the GEV family. For the constrained problem, we argue that the constrained model relying on purchase probabilities as decision variables is not appropriate for a robust version. Therefore, we propose an alternative formulation by adding a penalty term for violated constraints to the objective function. Our results in this context holds for the MNL model and for any choice model in the GEV family where the choice probability generating function (Fosgerau et al. 2013) has a separable structure.

Form now on, when saying "a GEV model", we refer to any choice model in the GEV family. Each GEV model can be represented by a choice probability generating function (CPGF) $G(\cdot)$ (see our detailed definition in the next section). To relax the homogeneity of the PSP, we need to assume that the CPGF has a separable structure, which means that $G(\cdot)$ can be written as a sum of sub-CPGFs, each corresponding to a subset of products.

Our contributions: We consider robust versions of the standard pricing optimization problem under GEV models. The setting here is to assume that the parameters of the choice model is not known with certainty and the aim is to find optimal prices associated with products, which maximize the worst-case expected revenue when the choice parameters vary in an uncertainty set. For the unconstrained problem with homogeneous PSP, we show that if the uncertainty set is convex and compact, the robust optimal prices have a constant markup with respect to the products costs, i.e., the robust optimal price of a product is equal to its unit cost plus a constant that is the same over all products. We also provide formulas that allow to efficiently compute that constant markup by binary search. This finding generalizes the results for the deterministic unconstrained problem with homogeneous PSP considered in Zhang et al. (2018). We also provide comparative insights showing how the robust optimal revenue and the robust optimal constant markups change as functions of the uncertainty level (i.e., the size of the uncertainty set). Our results in this case holds for any GEV model and with any convex, compact and bounded uncertainty set.

For the unconstrained problem with *non-homogeneous* PSP, similarly to previous studies (Zhang et al. 2018), we need to assume that CPGF is partition-wise separable and in each partition, the PSP are homogeneous. We show that if the uncertainty set is rectangular, then the robust problem can be converted equivalently into a deterministic pricing problem with partition-wise homogeneous PSP. As a result, the robust optimal prices have partition-wise constant markups, i.e., in each partition, the robust optimal prices have a constant markup with respect to their costs, and these constant markups can be computed by explicit formulas. We further show that, with a general uncertainty set, a partition-wise constant markup solution might be optimal to the robust problem. However, such solutions might not exist or are not easy to compute in a tractable way. We also provide comparative insights for the robust optimal prices and solutions when the size of the uncertainty set varies.

For the constrained pricing problem, as motivated by the applications with inventory considerations (Gallego and Van Ryzin 1997), previous studies (Zhang et al. 2018, Song and Xue 2007, Zhang and Lu 2013) also look at constraints on the expected sales. In this context, the aim is to select prices that maximize the expected revenue while requiring that the expected sales of products lie in a convex set. The advantage of such constraints is that the pricing problem can be reformulated equivalently as a convex program where the decision variables are the purchase probabilities. However, the final decision is a vector of prices and there may be no fixed prices under which the resulting purchase probabilities always satisfy the expected sale constraints when the choice parameters vary. For this reason, we consider that the use of the constrained formulation is not appropriate in our setting. Thus, we propose an alternative formulation in which, instead of requiring that the expected sale constraints are satisfied, we add a penalty cost to the objective function for violated constraints. Our formulation, called pricing with over-expected-sale penalties, is more general than the constrained formulation, in the sense that we show that if the penalty parameters increase to infinity, then the corresponding optimal solutions will converge to those from the constrained problem, and with *zero* penalty parameters, the pricing problem becomes the unconstrained one. Since the robust version under over-expected-sale penalties does not seek a price solution that satisfies the expected sale constraints under choice parameter uncertainty, it is more appropriate to use than the constrained version. We show that if the choice model is MNL and the uncertainty set is rectangular, then the robust problem can be converted into an equivalent deterministic pricing problem with over-expected-sale penalties, and this deterministic version can be solved by convex optimization. Our results also hold for any GEV model with a separable structure and under some restrictions on the parameters of the expected sale penalties.

In summary, we show that robust versions of the unconstrained problem are tractable, in the sense that the robust optimal solutions are shown to have constant markups with respect to the product costs, and we provide formulas to compute these constant markups efficiently. For the constrained version, we propose the formulation with over-expected-sale penalties, which is rational to used in our uncertainty setting, and show that the corresponding robust problem can be reformulated as a convex optimization problem, which is indeed tractable. Note that one may consider a stochastic approach to deal with the uncertainty issue, i.e., a model aiming at maximizing an average expected revenue over a finite number of scenarios of the choice parameters. However, such an objective function is difficult to handle, as one can show that there may be no constant-markup-style solutions that are optimal, and the problem optimization is not convex under formulations where the decision variables are the purchase probabilities.

Literature review: The GEV family includes most of the parametric discrete choice models available in the demand modeling and operations research literatures. The simplest and most popular member is the MNL (McFadden 1978, 1980) and it is well-known that the MNL model retains the independence from irrelevant alternatives (IIA) property, which does not hold in many contexts. There are a number of GEV models that allows to relax this property and provide flexibility in modeling the correlation between alternatives, for example, the nested logit model (Ben-Akiva et al. 1985, Ben-Akiva 1973), the cross-nested logit (Vovsha and Bekhor 1998), the generalized nested logit (Wen and Koppelman 2001), the paired combinatorial logit (Koppelman and Wen 2000), the ordered generalized extreme value (Small 1987), the specialized compound generalized extreme value models (Bhat 1998, Whelan et al. 2002) and network-based GEV (Daly and Bierlaire 2006, Mai et al. 2017) models. Fosgerau et al. (2013) show that the cross-nested logit model and its generalized version (i.e. network-based GEV) are fully flexible in the sense that they can approximate arbitrarily close any random utility maximization model. Beside the GEV family, it is worth noting that the mixed logit model (McFadden and Train 2000) is also popular due to its flexibility in capturing utility correlation. There is a fundamental trade-off between the flexibility and the generality of the choice models and the complexity of their estimation andapplication in operational problems. For the case of GEV models, even being flexible in modeling choice behavior, the resulting operational problems (e.g., product assortment or pricing) are often nonlinear and non-convex, leading to difficulties solving them in practice.

There is a large amount of research on unconstrained pricing under different discrete choice models. For example, Hopp and Xu (2005) and Dong et al. (2009) consider the pricing problem under the MNL model, Li and Huh (2011) consider the the nested logit model, Li et al. (2015) consider the pricing problem under the paired combinatorial logit model, and Zhang et al. (2018) consider the pricing problem under any choice model in the GEV family. Under the assumption that the PSP are the same over product, these authors show that the prices have a constant markup with respect to the product costs and provide formulas to explicitly computed this constant markup.

There are some studies trying to get over the assumption that the PSP are homogeneous over products. Li and Huh (2011) study the pricing problem under the nested logit model and assume that the PSP are homogeneous only in each nest and can be different over nests. They then show that the PSP in each nest have a constant markup. Zhang et al. (2018) generalize these results by considering the pricing problem under GEV models, in which the CPGF is partition-wise separable and the PSP are assumed to be homogeneous in each partition. The authors also show that, in this case, the optimal prices have a constant markup in each partition.

The literature has also seen studies considering the pricing problem with arbitrary PSP. Gallego and Hu (2014) show that the the pricing optimization problem unde the nested logit model can have multiple local optimal solutions if the PSP are arbitrarily heterogeneous and provide sufficient conditions to ensureunimodality of the expected revenue function. Li et al. (2015), Huh and Li (2015) consider the pricing problem under the d-nested and paired combinatorial logit models and also provide sufficient conditions on the PSP to ensure unimodality of the expected revenue function.

The constrained pricing problem where the prices are required to lie in a feasible set is difficult to solve as the expected revenue function is nonlinear and non-concave in the prices. Motivated by the applications with inventory considerations (Gallego and Van Ryzin 1997) and the observation that the expected revenue function is concave in the purchase probabilities, researchers have consider the pricing problem with constraints on the expected sales. For example, Song and Xue (2007), Zhang and Lu (2013) consider the pricing problem under the MNL model and show that the expected revenue is concave in the purchase probabilities if the PSP are homogeneous. Keller (2013) consider the pricing problem under the MNL and nested logit models and show that the expected revenue function is concave in the purchase probabilities under the MNL and arbitrary PSP, and establish sufficient conditions on the PSP to ensure that the expected revenue under the nested logit model is concave. Zhang et al. (2018) also generalizes all these results by showing that, under any GEV model, if the PSP are homogeneous or partition-wise homogeneous, then the expected revenue is concave in purchasing probabilities, making the pricing problem with expected sale constraints tractable.

All above studies assume that the parameters of the choice model is given in advance and ignore any uncertainty associated with such parameters in the pricing problem. However, the choice parameters typically need to be inferred from data and uncertainties may occur. In this work, we explicitly take into consider this issue by considering robust versions of the unconstrained and constrained pricing problems, with homogeneous and partition-wise homogeneous PSP. Our results directly generalize the results for deterministic pricing from Zhang et al. (2018), which already covers most of the the pricing optimization studies in the literature.

Our work is concerned with robust solutions for the pricing problem under uncertainty, so it is directly related to the concept of robust optimization, an important research area in operations research which has received a growing attention over the past two decades. Robust optimization is motivated by the fact that many real-world decision problems arising in engineering and management science have uncertain parameters due to limited data or noisy measurements. The literature on robust optimization includes a larger number of excellent studies (see Ben-Tal and Nemirovski 1998, 2000, Ben-Tal et al. 2006, for instance). Most of the studies in the literature of robust optimization focus on linear, piece-wise linear or convex objective functions. In our context, the expected revenue is nonlinear and non-convex/non-concave in the prices, implying that existing robust optimization results do not apply (except the part where we consider the constrained pricing problem under uncertain expected-sale constraints in Section A), and making our robust problem challenging to solve in a tractable way. It is worth noting that our work is relevant to Rusmevichientong and Topaloglu (2012) where the authors consider robust versions of the assortment planing problem. The decision variables in their work are discrete (i.e., a set of assortment) and the authors were able to obtain some nice results, e.g., they show that if the problem is unconstrained, then a revenue-ordered assortment is optimal to their robust problem.

Paper outline: We organize the paper as follows. In Section 2, we present the deterministic pricing problem under GEV models and recall some results from Zhang et al. (2018). In Section 3, we present our results for the robust unconstrained pricing problem under homogeneous and partition-wise homogeneous PSP. In Sections 4, we study the pricing optimization formulation with over-expected-sale penalties and its robust version. In Section 5 we provide some experimental results and in Section 6 we conclude. In the Appendix, we provide discussions onrobust constrained problems, some detailed proofs and numerical results.

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 $\{1,\ldots,m\}$. For any vector $\mathbf{b}$ with all equal elements, we use $\langle \mathbf{b} \rangle$ to denote the value of one element of the vector. Given two vectors of the same size $\mathbf{a},\mathbf{b} \in \mathbb{R}^m$, $\mathbf{a} \succeq \mathbf{b}$ is equivalent to $\mathbf{a}-\mathbf{b} \in \mathbb{R}^m_+$, and $\mathbf{a} \preceq \mathbf{b}$ is equivalent to $\mathbf{b} \succeq \mathbf{a}$.

2. Deterministic Pricing under Generalized Extreme Value Models

We denote by $\mathcal{V} = \{1,\ldots,m\}$ the set of m available products. There is a *non-purchase item* indexed by 0, so the set of all possible *products* is $\mathcal{V} \cup \{0\}$. We also denote by x_i and c_i the price and the cost of product i, respectively. The random utility maximization (RUM) framework (McFadden 1978) is the most popular approach to model discrete choice behavior. Under this framework, each product $i \in \mathcal{V}$ is assigned with a random utility U_i and the additive RUM framework (Fosgerau et al. 2013, McFadden 1978) assumes that each random utility can be expressed as a sum of two part $U_i = u_i + \varepsilon_i$, where the term u_i is deterministic and can include values representing characteristics of the product, and the term ε_i is unknown to the analyst. The RUM principle then assume that the selections are made by maximizing these utilities and the probability that an product i (including the non-purchase item) is selected can be computed as $P(U_i \geq U_j,\ \forall j \in \mathcal{V} \cup \{0\})$.

In our context, we are interested in the effect of the prices on the expected revenue. So we assume that the deterministic terms u_i, $\forall i \in \mathcal{V}$, can be expressed as $u_i = a_i - b_i x_i$, where b_i is the PSP associated with product i and a_i can include other information that may affect customer's demand such as the brand, size or color of the items. These values can be obtained by fitting the choice model with observation data. As mentioned above, the estimation process may cause uncertainties associated with such estimates. Note that here we assume that the utilities u_i depend linearly on the prices, which is a popular assumption in most of the existing pricing studies. Nonlinear formulation would make the pricing problem much more difficult to deal with, but would be interesting to look at in future research.

A GEV model can be represented by a CPGF $G(\mathbf{Y})$, where $\mathbf{Y}$ is a vector of size m with entries $Y_i = e^{u_i}$, for all $i \in \mathcal{V}$. To be consistent with the RUM principle, $G(\cdot)$ needs to satisfy the following properties (McFadden 1978, Ben-Akiva et al. 1985).

Remark 1 (**Properties of GEV-CPGF**). *A GEV-CPGF $G(\boldsymbol{Y})$ has the following properties.*

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

(ii) G is homogeneous of degree one, i.e., $G(\lambda \boldsymbol{Y}) = \lambda G(\boldsymbol{Y})$

(iii) $G(\boldsymbol{Y}) \to \infty$ *if* $Y_i \to \infty$

(iv) Given $i_1,\ldots,i_k$ *distinct from each other,* $\partial G_{i_1,\ldots,i_k}(\boldsymbol{Y}) \geq 0$ *if* k *is odd, and* $\leq$ *if* k *is even*

(v) $G(\boldsymbol{Y}) = \sum_{i \in \mathcal{V}} Y_i \partial G_i(\boldsymbol{Y})$

(vi) $\sum_{j \in \mathcal{V}} Y_j \partial G_{ij}(\boldsymbol{Y}) = 0$, $\forall i \in \mathcal{V}$.

where $\partial G_i(\boldsymbol{Y}) = \partial G(\boldsymbol{Y})/\partial Y_i$

Under a GEV model specified by CPGF G, given any vector $\mathbf{Y} \in \mathbb{R}^m$, the choice probability of product $i \in \mathcal{V}$ is given by

$$P_i(\mathbf{Y}|G) = \frac{Y_i \partial G_i(\mathbf{Y})}{1 + G(\mathbf{Y})}.$$

Note that the above formulation also implies that the choice probability of the *non-purchase* item is $P_0(\mathbf{Y}|G) = 1/(1+G(\mathbf{Y}))$. The GEV becomes the MNL model if $G(\mathbf{Y}) = \sum_{i=1}^m Y_i$, and it becomes the nested logit model if $G(\mathbf{Y}) = \sum_{n\in\mathcal{N}} \left(\sum_{i\in C_n} (\sigma_{in} Y_i)^{\mu_n}\right)^{\mu/\mu_n}$, where $\mathcal{N}$ is the set of nests, C_n is the set of items in nest n and $\sigma_{in}, \mu > 0, \mu_n > 0$ are the parameters of the nested logit model. In the generalized version of the nested logit model proposed by Daly and Bierlaire (2006), called the network GEV, the corresponding CPGF can be computed recursively based on a rooted and cycle-free graph representing the correlation structure of the items.

Under a GEV model specified by a CPGF $G(\cdot)$, the deterministic version of the pricing problem is stated as

$$\max_{\mathbf{x}\in\mathbb{R}^m} R(\mathbf{x}) = \sum_{i=1}^m (x_i - c_i) P_i(\mathbf{Y}(\mathbf{x},\mathbf{a},\mathbf{b})|G), \tag{P1}$$

where $\mathbf{Y}(\mathbf{x},\mathbf{a},\mathbf{b}) \in \mathbb{R}^m$ with entries $Y_i(\mathbf{x},\mathbf{a},\mathbf{b}) = \exp(a_i - b_i x_i)$. The expected revenue $R(\mathbf{x})$ becomes more difficult to handle as the GEV model becomes more complicated. By leveraging the properties of GEV models stated in Remark 1, Zhang et al. (2018) manage to show that if the PSP are homogeneous, i.e., $b_i = b_j$ for all $i,j \in \mathcal{V}$ and if $\mathbf{x}^*$ is an optimal solution to (P1), then

$$x_i^* - c_i = \frac{1}{\langle \mathbf{b} \rangle} + R(\mathbf{x}^*), \forall i \in \mathcal{V} \text{ and } R(\mathbf{x}^*) = \frac{W(\gamma e^{-1})}{\langle \mathbf{b} \rangle} \tag{1}$$

where $\gamma = G(Y_1(c_1), \ldots, Y_m(c_m))$ and $W(\cdot)$ is the Lambert-W function. The results in (1) indeed imply that a constant markup solution is optimal to (P1) and this constant markup can be computed explicitly. Moreover, if the PSP are partition-wise homogeneous and G is separable, then Zhang et al. (2018) show that the optimal prices have a constant markup in each partition. These results also provide an explicit way to compute optimal prices for the pricing problem under the MNL with arbitrary PSP. Zhang et al. (2018) also show that the expected revenue function is concave in the purchasing probabilities under any GEV model, making the pricing problem with expected sale constraints tractable.

3. Robust Unconstrained Pricing

In this section, we study a robust version of the unconstrained pricing problem, under the setting that the choice parameters $(\mathbf{a},\mathbf{b})$ are not given exactly but belong to an uncertainty set. We first present our results for the case of homogeneous PSP. We then switch to the case of partition-wise PSP later on. In our robust model, we aim at maximizing the worst-case expected revenue over all parameters in the uncertainty set. The robust unconstrained pricing problem (P1) can be formulated as

$$\max_{\mathbf{x}\in\mathbb{R}^m}\left\{g(x)=\min_{(\mathbf{a},\mathbf{b})\in\mathcal{A}}\sum_{i=1}^{m}(x_i-c_i)P_i(\mathbf{Y}(\mathbf{x},\mathbf{a},\mathbf{b})|G),\right\}, \tag{RO}$$

where $\mathcal{A}$ is the uncertainty set of the parameters $(\mathbf{a},\mathbf{b})$. We also assume that $\mathcal{A}$ is compact, convex and bounded (Assumption 1). The convexity and compactness assumptions are useful later in the section, as we need to show that, under a constant-markup style vector of prices, the objective function of the adversary's problem is convex on $\mathcal{A}$, which in turn helps to identify a saddle point of the robust problem. The boundedness assumption is rational in the context, as the choice parameters are often inferred from data and it is expected that they are finite. We also assume that the PSP are positive, i.e., $\underline{b}>0$, which is rational from a behavior point of view.

Assumption 1. *$\mathcal{A}$ is convex, compact and bounded, i.e., there exists $(\underline{\boldsymbol{a}},\underline{\boldsymbol{b}})$ and $(\overline{\boldsymbol{a}},\overline{\boldsymbol{b}})$ in $\mathbb{R}^{2m}$ such that $(\underline{\boldsymbol{a}},\underline{\boldsymbol{b}})\preceq(\boldsymbol{a},\boldsymbol{b})\preceq(\overline{\boldsymbol{a}},\overline{\boldsymbol{b}})$, $\forall(\boldsymbol{a},\boldsymbol{b})\in\mathcal{A}$. Moreover, assume that $\underline{\boldsymbol{b}}\succeq 0$ and $\overline{\boldsymbol{b}}\neq 0$.*

3.1. Homogeneous Price Sensitivity Parameters

When the PSP are the same over all the products, we will show that the robust optimal prices have a constant markup and this constant markup can be computed efficiently by binary search. To prove the results, we will consider the robust unconstrained pricing problem with constant-markup prices, i.e., we only look at prices $\mathbf{x}$ such that $x_i-c_i=x_j-c_j$ for all $i,j\in\mathcal{A}$. Then we show that there exist constant-markup prices $\mathbf{x}^*$ such that if $(\mathbf{a}^*,\mathbf{b}^*)$ is an optimal solution to the adversary's problem under prices $\mathbf{x}^*$, then $\mathbf{x}^*$ is also optimal to the deterministic unconstrained problem with choice parameters $(\mathbf{a}^*,\mathbf{b}^*)$. In this context, $(\mathbf{x}^*,\mathbf{a}^*,\mathbf{b}^*)$ is a saddle point of the robust problem and we show that $\mathbf{x}^*$ is also an optimal solution to the robust problem.

Given constant-markup prices $\mathbf{x}\in\mathbb{R}^m$ and choice parameters $(\mathbf{a},\mathbf{b})\in\mathcal{A}$, the expected revenue becomes

$$\begin{aligned}\sum_{i=1}^{m}(x_i-c_i)P_i(\mathbf{Y}(\mathbf{x},\mathbf{a},\mathbf{b})|G)&=\frac{z\sum_{i\in\mathcal{V}}Y_i\partial G_i(\mathbf{Y})}{1+G(\mathbf{Y})}\\&=z\left(1-\frac{1}{1+G(\mathbf{Y})}\right),\end{aligned}$$

where $z=x_i-c_i$, $\forall i\in\mathcal{V}$ and $\mathbf{Y}$ is a vector with entries $Y_i=\exp(a_i-b_i(z+c_i))$ for all $i\in\mathcal{V}$. As a result, the expected revenue is a function of z and $(\mathbf{a},\mathbf{b})$, and if $(\mathbf{a}^*(z),\mathbf{a}^*(z))$ is an optimal solution to the adversary's problem, then we also have

$$(\mathbf{a}^*(z),\mathbf{b}^*(z))=\operatorname*{argmin}_{\mathbf{a},\mathbf{b}\in\mathcal{A}}\quad G(\mathbf{Y}|z,\mathbf{a},\mathbf{b}), \tag{2}$$

where $G(\mathbf{Y}|z,\mathbf{a},\mathbf{b})=G(Y_1,\ldots,Y_m)$ with $Y_i=e^{a_i-b_i(z+c_i)}$. First, we show that, given $z\geq 0$, $G(\mathbf{Y}|z,\mathbf{a},\mathbf{b})$ is strictly convex in $(\mathbf{a},\mathbf{b})$ and $(\mathbf{a}^*(z),\mathbf{b}^*(z))$ is always uniquely determined (Proposition 1).

Proposition 1. *Under Assumption 1, given any $z \in \mathbb{R}_+$, $G(\boldsymbol{Y}|z,\boldsymbol{a},\boldsymbol{b})$ is strictly convex on $\mathcal{A}$ and Problem 2 always has a unique solution.*

Proof: First, we consider function $f^G(\mathbf{s}) : \mathbb{R}^m \rightarrow \mathbb{R}_+$

$$f^G(\mathbf{s}) = G(Y_1,\ldots,Y_m), \text{ where } Y_i = e^{s_i}, \ \forall i \in \mathcal{V}$$

We will prove that $f^G(\mathbf{s})$ is convex. Taking the first and second derivatives of $f^G(\mathbf{s})$ we obtain

$$\frac{\partial f^G(\mathbf{s})}{\partial s_i} = \partial G_i(\mathbf{Y})Y_i,$$

and

$$\frac{\partial^2 f^G(\mathbf{s})}{\partial s_i \partial s_i} = \partial G_{ii}(\mathbf{Y})Y_i^2 + \partial G_i(\mathbf{Y})Y_i,$$
$$\frac{\partial^2 f^G(\mathbf{s})}{\partial s_i \partial s_j} = \partial G_{ij}(\mathbf{Y})Y_iY_j.$$

So we have

$$\nabla^2 f^G(\mathbf{s}) = \text{diag}(\mathbf{Y})\nabla^2 G(\mathbf{Y})\text{diag}(\mathbf{Y}) + \text{diag}(\nabla G(\mathbf{Y}) \circ \mathbf{Y}),$$

where $\text{diag}(\mathbf{Y})$ is the square diagonal matrix with the elements of vector $\mathbf{Y}$ on the main diagonal. The second term $\text{diag}(\nabla G(\mathbf{Y}) \circ \mathbf{Y})$ is always positive definite. Moreover, $\text{diag}(\mathbf{Y})\nabla^2 G(\mathbf{Y})\text{diag}(\mathbf{Y})$ is symmetric and its (i,j)-th component is given by $Y_i \partial G_{ij}(\mathbf{Y})Y_j$. For $i \neq j$, we have $\partial G_{ij}(\mathbf{Y}) \leq 0$ by the property of the GEV-CPGF G, so all off-diagonal entries of the matrix are non-positive. In addition, $\sum_{j\in\mathcal{V}} Y_j \partial G_{ij}(\mathbf{Y}) = 0$, so that each row of the matrix sums to zero. Thus, $\text{diag}(\mathbf{Y})\nabla^2 G(\mathbf{Y})\text{diag}(\mathbf{Y})$ is positive semi-definite (see Theorem A.6 in De Klerk 2006). So, $\nabla^2 f^G(\mathbf{s})$ is positive definite, or equivalently, $f^G(\mathbf{s})$ is strictly convex in $\mathbf{s}$. This lead to the following inequality, for all $\mathbf{s}^1, \mathbf{s}^2 \in \mathbb{R}^m$ and $\lambda \in (0,1)$

$$\lambda f^G(\mathbf{s}^1) + \lambda f^G(\mathbf{s}^2) > f^G(\lambda \mathbf{s}^1 + (1-\lambda)\mathbf{s}^2).$$

For all $(\mathbf{a}^1,\mathbf{b}^1),(\mathbf{a}^2,\mathbf{b}^2) \in \mathcal{A}$, replace s_i^1 by $a_i^1 - b_i^1(z+c_i)$ and s_i^2 by $a_i^2 - b_i^2(z+c_i)$ we have

$$\lambda G(\mathbf{Y}|z,\mathbf{a}^1,\mathbf{b}^1) + \lambda G(\mathbf{Y}|z,\mathbf{a}^2,\mathbf{b}^2) > G(\mathbf{Y}|z,\lambda\mathbf{a}^1 + (1-\lambda)\mathbf{a}^2, \lambda\mathbf{b}^1 + (1-\lambda)\mathbf{b}^2), \ \forall \lambda \in (0,1)$$

which means that $G(\mathbf{Y}|z,\mathbf{a},\mathbf{b})$ is strictly convex in $\mathbf{a},\mathbf{b}$. This completes the proof. □

Next, we further show that $(\mathbf{a}^*(z),\mathbf{b}^*(z))$ determined in (2) are not only unique given any $z \geq 0$, but also are continuous in z (Lemma 1).

Lemma 1. $(\boldsymbol{a}^*(z),\boldsymbol{b}^*(z))$ *determined in (2) is continuous in* $z \in \mathbb{R}_+$.

Proof: This is a direct result from Proposition 1, i.e., $(\mathbf{a}^*(z),\mathbf{b}^*(z)$ are uniquely determined, and the Corollary 8.2 of Hogan (1973). □

In the next lemma, we show that, given any $z \in \mathbb{R}_+$, if the uncertainty set $\mathcal{A}$ is bounded, the the function $G(\mathbf{Y}|z, \underline{\mathbf{a}}, \overline{\mathbf{b}})$ is also bounded for all parameters $(\mathbf{a}, \mathbf{b}) \in \mathcal{A}$. The lemma allows us to determine an finite interval where we can search the robust optimal constant markup.

Lemma 2. *Under Assumption 1 we have*

$$G(\boldsymbol{Y}|z, \underline{\boldsymbol{a}}, \overline{\boldsymbol{b}}) \le G(\boldsymbol{Y}|z, \boldsymbol{a}, \boldsymbol{b}) \le G(\boldsymbol{Y}|z, \overline{\boldsymbol{a}}, \underline{\boldsymbol{b}}), \ \forall z \in \mathbb{R}_+, \ (\boldsymbol{a}, \boldsymbol{b}) \in \mathcal{A}.$$

Proof: Again, consider $f^G(\mathbf{s}) = G(Y_1, \ldots, Y_m)$, where $Y_i = e^{s_i}$, $\forall i = 1, \ldots, m$. Taking the derivative of $f^G(\mathbf{s})$ w.r.t. s_i we have

$$\frac{\partial f^G(\mathbf{s})}{\partial s_i} = \partial G_i(\mathbf{Y}) Y_i \ge 0$$

So, $f^G(\mathbf{s})$ is monotonic in every coordinate, meaning that given any $\mathbf{s}, \mathbf{s}_0 \in \mathbb{R}^m$, $\mathbf{s} \succeq s_0$, we have $f^G(\mathbf{s}) \ge f^G(\mathbf{s}_0)$. Moreover, it is clear from Assumption 1 that

$$\underline{\mathbf{a}} - \overline{\mathbf{b}} \circ (\mathbf{c} + z\mathbf{e}) \preceq \mathbf{a} - \mathbf{b} \circ (\mathbf{c} + z\mathbf{e}) \preceq \overline{\mathbf{a}} - \underline{\mathbf{b}} \circ (\mathbf{c} + z\mathbf{e}), \ \forall z \in \mathbb{R}_+, (\mathbf{a}, \mathbf{b}) \in \mathcal{A}.$$

So, we obtain the following inequality

$$G(\mathbf{Y}|z, \underline{\mathbf{a}}, \overline{\mathbf{b}}) \le G(\mathbf{Y}|z, \mathbf{a}, \mathbf{b}) \le G(\mathbf{Y}|z, \overline{\mathbf{a}}, \underline{\mathbf{b}}), \ \forall (\mathbf{a}, \mathbf{b}) \in \mathcal{A},$$

which completes the proof. □

Proposition 2 below is a key result of this section where we show that there is a constant markup z^* such that if we solve the corresponding adversary's problem under z^* and obtain a solution $(\mathbf{a}^*(z), \mathbf{b}^*(z))$, then z^* is also the optimal constant markup for the deterministic pricing problem under parameters $(\mathbf{a}^*(z), \mathbf{b}^*(z))$. In other words, we show that there exist $z^* \in \mathbb{R}_+$ such that $(z^* + \mathbf{c}, \mathbf{a}^*(z), \mathbf{b}^*(z))$ is a saddle point of the robust problem.

Proposition 2. *For any $i \in \mathcal{V}$, there exists $z^* \in \mathbb{R}_+$ such that*

$$z^* = \frac{1 + W(\tau(z^*))}{\langle \boldsymbol{b}^*(z^*) \rangle} \in \left[\underline{Z}^0, \overline{Z}^0\right]$$

where

$$\underline{Z}^0 = \frac{1 + W(G(\boldsymbol{Y}|0, \underline{\boldsymbol{a}}, \overline{\boldsymbol{b}})e^{-1})}{\langle \overline{\boldsymbol{b}} \rangle}$$
$$\overline{Z}^0 = \frac{1 + W(G(\boldsymbol{Y}|0, \overline{\boldsymbol{a}}, \underline{\boldsymbol{b}})e^{-1})}{\langle \underline{\boldsymbol{b}} \rangle}$$
$$\tau(z^*) = G(\boldsymbol{Y}|0, \boldsymbol{a}^*(z^*), \boldsymbol{b}^*(z^*))e^{-1}$$

and $W(\cdot)$ is the is the Lambert-W function.

Proof: Let

$$f(z) = z - \frac{1 + W(\tau(z))}{\langle \mathbf{b}^*(z) \rangle}.$$

From Lemma 2, we have the following chain of inequalities

$$\underline{Z}^0 \leq \frac{1 + W(\tau(z))}{\langle \mathbf{b}^*(z) \rangle} \leq \overline{Z}^0, \ \forall z \in \mathbb{R}_+.$$

Which means

$$f(\underline{Z}^0) \leq 0; \ f(\overline{Z}^0) \geq 0$$

Since $f(z)$ is continuous in z (Lemma 1), equation $f(z) = 0$ always has a solution in the interval $\left[\underline{Z}^0, \overline{Z}^0\right]$. □

In Proposition 2, we make use of the boundedness assumption on $\mathcal{A}$ to identify an interval where we can find z^*. Without this assumption, one can simply choose 0 as a lower bound, as $f(0)$ is always less than 0. However, to identify an upper bound, one needs some limits from the uncertainty set. This is because even in the deterministic case, if the choice parameters $\mathbf{b}$ approach zero, or $\mathbf{a}$ increase to infinity, then the optimal constant markup will go to infinity (see Equation 1 for details).

We are now ready to establish the main result of the section. Theorem 1 below shows that the robust optimal prices have a constant markup and it provides a formula to compute this constant markup by binary search. The proof of the theorem is quite obvious, as we already show that there isa saddle point of the robust problem that has the constant-markup style (Lemma 2).

Theorem 1. ***(Constant markup is optimal to the robust problem when the PSP are homogeneous).*** *Assume that Assumption 1 holds and the PSP are homogeneous, i.e., for any* $(\boldsymbol{a}, \boldsymbol{b}) \in \mathcal{A}$, $b_i = b_j$, $\forall i, j = 1, \ldots, m$, *then* $\boldsymbol{x}^{\text{RO}*}$ *defined below is the unique optimal solution of the robust problem* (RO)

$$x_i^{\text{RO}*} = z^{\text{RO}*} + c_i, \ \forall i = 1, \ldots, m \tag{3}$$

where $z^{\text{RO}*} \in \mathbb{R}_+$ *is the* ***unique*** *value of* z *satisfying*

$$z = \frac{1 + W(G(\boldsymbol{Y}|0, \boldsymbol{a}^*(z), \boldsymbol{b}^*(z))e^{-1})}{\langle \boldsymbol{b}^*(z) \rangle}, \tag{4}$$

where $W(\cdot)$ *is the Lambert-W function.*

Proof: To prove that $\mathbf{x}^{\text{RO}*}$ is optimal to the robust problem, we will show that $g(\mathbf{x}^{\text{RO}*}) \geq g(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^m$. Let us first define

$$\Phi(\mathbf{x}, \mathbf{a}, \mathbf{b}) = \sum_{i=1}^{m} (x_i - c_i) P_i(\mathbf{Y}(\mathbf{x}, \mathbf{a}, \mathbf{b})|G) \tag{5}$$

Given $\mathbf{x}^{\text{RO}*}$ and $z^{\text{RO}*}$ defined in (3) and (4), we first remark that $(\mathbf{a}^*(z^{\text{RO}*}), \mathbf{b}^*(z^{\text{RO}*}))$ is also the unique solution of the adversary's problem

$$\underset{(\mathbf{a},\mathbf{b})\in\mathcal{A}}{\text{argmin}} \qquad \Phi(\mathbf{x}^{\text{RO}*}, \mathbf{a}, \mathbf{b})$$

So, $g(\mathbf{x}^{\text{RO}*}) = \Phi(\mathbf{x}^{\text{RO}*}, \mathbf{a}^*(z^{\text{RO}*}), \mathbf{b}^*(z^{\text{RO}*}))$. Using the results from Zhang et al. (2018) and according to the way $\mathbf{x}^{\text{RO}*}$ is computed, $\mathbf{x}^{\text{RO}*}$ is optimal to the following problem

$$\max_{\mathbf{x}\in\mathbb{R}^m} \Phi(\mathbf{x}, \mathbf{a}^*(z^{\text{RO}*}), \mathbf{b}^*(z^{\text{RO}*})).$$

This leads to the following inequalities, for any $\mathbf{x}\in\mathbb{R}^m$,

$$\begin{aligned} g(\mathbf{x}^{\text{RO}*}) &= \Phi(\mathbf{x}^{\text{RO}*}, \mathbf{a}^*(z^{\text{RO}*}), \mathbf{b}^*(z^{\text{RO}*})) \\ &\geq \Phi(\mathbf{x}, \mathbf{a}^*(z^{\text{RO}*}), \mathbf{b}^*(z^{\text{RO}*})) \\ &\geq \min_{(\mathbf{a},\mathbf{b})\in\mathcal{A}} \qquad \Phi(\mathbf{x}, \mathbf{a}, \mathbf{b}) \\ &= g(x). \end{aligned} \tag{6}$$

So, $g(\mathbf{x}^{\text{RO}*}) \geq g(\mathbf{x})$ for all $\mathbf{x}\in\mathbb{R}^m$, i.e., $\mathbf{x}^{\text{RO}*}$ is an optimal solution to the robust problem.

Note that the deterministic version of the unconstrained pricing problem always has a unique solution, which is the constant markup one (Zhang et al. 2018). So, the inequality in (6) is strict if $\mathbf{x} \neq \mathbf{x}^{\text{RO}*}$. In other words, $g(\mathbf{x}^{\text{RO}*}) > g(\mathbf{x})$ if $\mathbf{x} \neq \mathbf{x}^{\text{RO}*}$, meaning that there is only one solution to the robust pricing problem (RO) and there is only solution to the equation (4), as required. □

Theorem 1 implies that a solution to the robust problem can be found by solving the equation

$$\frac{1 + W(G(\mathbf{Y}|0, \mathbf{a}^*(z), \mathbf{b}^*(z))e^{-1})}{\langle \mathbf{b}^*(z)\rangle} - z = 0, \tag{7}$$

in the interval $\left[\underline{Z}^0, \overline{Z}^0\right]$, in which $\underline{Z}^0, \overline{Z}^0$ are defined in Lemma 2. This could be done efficiently via binary search. In comparison with its deterministic counter part, the robust problem would requires about $\log_2\left((\overline{Z}^0 - \underline{Z}^0)\right)/\epsilon)$ steps to obtain a constant markup that is in the ϵ-neighbourhood of the optimal constant markup, while the deterministic problem requires one step to get its optimal solution.

Theorem 1 also allows us to give comparative statistics that describe how the robust optimal value and the optimal prices change as a function of the size of the uncertainty set. To facilitate our exposition, let us denote by $\Gamma(\mathcal{A})$ the optimal value of the robust problem (RO) andby $\mathbf{x}^*(\mathcal{A})$ the robust constant markup given by Theorem 1. We also need the following definition and lemma to establish the comparative results.

DEFINITION 1. An uncertainty set $\mathcal{A}$ is called "SEPARATED-BY-PRICE-SENSITIVITY-PARAMETERS" if $\mathcal{A}$ is a Cartesian product of two uncertainty sets of $\mathbf{a}$ and $\mathbf{b}$, i.e., $\mathcal{A} = \{(\mathbf{a},\mathbf{b})|\ \mathbf{a}\in\mathcal{A}^{\mathbf{a}}\subset\mathbb{R}^m, \mathbf{b}\in\mathcal{A}^{\mathbf{b}}\subset\mathbb{R}^m_+\}$.

Lemma 3. *Assume that the PSP are homogeneous, given an uncertainty set $\mathcal{A}$, satisfying Assumption 1 and the "SEPARATED-BY-PRICE-SENSITIVITY-PARAMETERS"*

PROPERTY 1. , then the set $\mathcal{A}$ can be represented as $\mathcal{A} = \mathcal{A}^{\mathbf{a}}\times\mathcal{A}^{\mathbf{b}}$ where $\mathcal{A}^{\mathbf{b}} = \{\mathbf{b}|b_i = b_j \in [\underline{b},\overline{b}], \forall i,j\in[m]\}$, where $\underline{b},\overline{b}\in\mathbb{R}_+$, and $\mathbf{x}^*(\mathcal{A})_i = c_i + 1/\overline{b} + \Gamma(\mathcal{A})$.

Proof: Since the PSP are homogeneous and the uncertainty sets are compact, there always exists $\underline{b},\overline{b}\in\mathbb{R}_+$, $\underline{b}\le\overline{b}$ such that $\mathcal{A}^{\mathbf{b}} = \{\mathbf{b}|b_i = b_j\in[\underline{b},\overline{b}]\}$. Moreover, from Theorem 1, the robust optimal prices $\mathbf{x}^*(\mathcal{A})$ is constant markup, i.e., $\exists z^*\in\mathbb{R}_+$ such that $\mathbf{x}^*(\mathcal{A})_i - c_i = z^*$. Let $(\mathbf{a}^*,\mathbf{b}^*) = \text{argmin}_{(\mathbf{a},\mathbf{b})\in\mathcal{A}} G(\mathbf{Y}|z^*,\mathbf{a},\mathbf{b})$. Using the result that $f^G(\mathbf{s}) = G(e^{e_1},\ldots,e^{e_m})$ is monotonic in every coordinate of $\mathbf{s}\in\mathbb{R}^m$ (Lemma 2) and $(\mathbf{a}^*,\mathbf{b}^*)$ is always uniquely determined (Proposition 1), we can show that $b^*_i = \overline{b}$, for all $i\in[m]$. So, from Theorem 1 and Proposition 3.2 of Zhang et al. (2018) we obtain

$$\begin{aligned}\mathbf{x}^*(\mathcal{A})_i - c_i = z^* &= \frac{1}{\overline{b}} + \frac{W(G(\mathbf{Y}|0,\mathbf{a}^*,\mathbf{b}^*)e^{-1})}{\overline{b}}\\ &= \frac{1}{\overline{b}} + \Gamma(\mathcal{A}),\quad \forall i\in[m],\end{aligned}$$

which is the desired result. □

In the next proposition, we show that if the uncertainty set becomes larger, then we obtain smaller robust expected value and the robust prices decreases as well.

Proposition 3. (Larger uncertainty set leads to smaller robust expected revenue and smaller robust optimal prices). *If the PSP are homogeneous, then given two convex uncertainty sets $\mathcal{A}_1,\mathcal{A}_2$ satisfying Assumption 1 such that $\mathcal{A}_1\subset\mathcal{A}_2$, we have $\Gamma(\mathcal{A}_1)\ge\Gamma(\mathcal{A}_2)$. Moreover, if $\mathcal{A}_1$ and $\mathcal{A}_2$ are "SEPARATED-BY-PRICE-SENSITIVITY-PARAMETERS", then $\boldsymbol{x}^*(\mathcal{A}_1)\succeq\boldsymbol{x}^*(\mathcal{A}_2)$.*

Proof: For notational convenience, we denote the objective value of the adversary as $g(\mathbf{x}|\mathcal{A}) = \min_{(\mathbf{a},\mathbf{b})\in\mathcal{A}}\sum_i(x_i - c_i)P_i(\mathbf{Y}(\mathbf{x},\mathbf{a},\mathbf{b})|G)$. It is obvious that $g(\mathbf{x}|\mathcal{A}_1)\ge g(\mathbf{x}|\mathcal{A}_2)$ for any $\mathbf{x}\in\mathbb{R}^m_+$, so we have

$$g(\mathbf{x}^*(\mathcal{A}_2)|\mathcal{A}_2)\le g(\mathbf{x}^*(\mathcal{A}_2)|\mathcal{A}_1)\le g(\mathbf{x}^*(\mathcal{A}_1)|\mathcal{A}_1),$$

which means $\Gamma(\mathcal{A}_1)\ge\Gamma(\mathcal{A}_2)$, as required.

Now, assume that the uncertainty sets $\mathcal{A}_1$ and $\mathcal{A}_2$ are "SEPARATED-BY-PRICE-SENSITIVITY-PARAMETERS". Using Lemma 3, we can write $\mathcal{A}_1 = \mathcal{A}^{\mathbf{a}}_1\times\mathcal{A}^{\mathbf{b}}_1$ where $\mathcal{A}^{\mathbf{b}}_1 = \{\mathbf{b}|b_i = b_j\in[\underline{b}_1,\overline{b}_1],\ \forall i,j\}$, and $\mathcal{A}_2 = \mathcal{A}_2^{\mathbf{a}} \times \mathcal{A}_2^{\mathbf{b}}$ where $\mathcal{A}_2^{\mathbf{b}} = \{\mathbf{b}|b_i = b_j \in [\underline{b}_2, \overline{b}_2],\ \forall i, j\}$, where $\underline{b}_1, \overline{b}_1, \underline{b}_2, \overline{b}_2$ are non-negative constants such that $\underline{b}_1 \leq \overline{b}_1$ and $\underline{b}_2 \leq \overline{b}_2$. From Theorem 1 we can write

$$\mathbf{x}^*(\mathcal{A}_1)_i = c_i + \frac{1}{\overline{b}_1} + \Gamma(\mathcal{A}_1), \quad \forall i \in [m]$$
$$\mathbf{x}^*(\mathcal{A}_2)_i = c_i + \frac{1}{\overline{b}_2} + \Gamma(\mathcal{A}_2), \quad \forall i \in [m].$$

Moreover, since $\mathcal{A}_1 \subset \mathcal{A}_2$, we have $\overline{b}_1 \leq \overline{b}_2$. Using the result $\Gamma(\mathcal{A}_1) \geq \Gamma(\mathcal{A}_2)$ we have $\mathbf{x}^*(\mathcal{A}_1) \succeq \mathbf{x}^*(\mathcal{A}_2)$, which completes the proof. □

The proof of Proposition 3 contains more information than the proposition itself, as we we can write $\mathbf{x}^*(\mathcal{A}_1)_i - \mathbf{x}^*(\mathcal{A}_2)_i = \frac{1}{\overline{b}_1} - \frac{1}{\overline{b}_2} + \Gamma(\mathcal{A}_1) - \Gamma(\mathcal{A}_2) \geq \frac{1}{\overline{b}_1} - \frac{1}{\overline{b}_2}$, which indicates that if we increase the upper bound of the PSP in the uncertainty set, the robust optimal prices will strictly decrease.

3.2. Partially Heterogeneous Price Sensitivity Parameters

In this section, we try to relax the assumption that the PSP are homogeneous. Similar to the deterministic version considered in Zhang et al. (2018), we need additional assumptions to derive solutions to the robust problem. More precisely, we require that the products can be partitioned into disjoint subsets, the generating function is separable by the partitions, and the products in each partition share the same price sensitivity parameter. We partition the set of all products $\mathcal{V}$ into N non-empty subsets $\mathcal{V}_1, \ldots, \mathcal{V}_N$ such that $\mathcal{V} = \bigcup_{n=1}^N \mathcal{V}_n$ and $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset,\ \forall i \neq j, i, j \in [N]$. Moreover, we separate the vector $\mathbf{Y}$ into sub-vectors $\mathbf{Y}^1, \ldots, \mathbf{Y}^N$ such that $\mathbf{Y}^n = \{Y_i |\ i \in \mathcal{V}_n\}$ for all $n \in [N]$. We also separate vector $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}_+^{2m}$ into sub-vectors $(\mathbf{a}^1, \mathbf{b}^1), \ldots, (\mathbf{a}^N, \mathbf{b}^N)$ such that $(\mathbf{a}^n, \mathbf{b}^n) = \{(a_i, b_i) |\ i \in \mathcal{V}_n\}$, $\forall n \in [N]$. Let denote by $\mathcal{A}^n \subset \mathcal{A}$ the uncertainty set for the sub-vector $(\mathbf{a}^n, \mathbf{b}^n)$, for any $n \in [N]$. We further assume that the GEV-CPGF $G(\mathbf{Y})$ can be separated into N GEV-CPGFs as

$$G(\mathbf{Y}) = \sum_{n=1}^N G^n(\mathbf{Y}^n).$$

In this context, we need to further assume that the uncertainty set is rectangular, because of the following reason. When the PSP is not homogeneous, the optimal prices to the deterministic pricing problem do not have a single constant markup over all products (Zhang et al. 2018). As a consequence, the robust optimal prices to (RO) would generally not have a single constant markup over all the products. Moreover, if the prices do not have constant-markup style as in the previous section, the corresponding adversary's problem would be non-concave in $(\mathbf{a}, \mathbf{b})$ and solutions to the adversary's problem may be not unique. For this reason, we can not apply the techniques used in the previous version to identify a saddle point of the the robust problem. Furthermore, in Proposition 5 below, we show that under a general uncertainty set, a partition-wise constant-markup solution would be optimal to the robust problem, but such a solution may not exist or may not be tractable to obtain.

Assumption 2. *The PSP are homogeneous in each subset* $\mathcal{V}_n$*,* $n\in[N]$*, and the uncertainty set* $\mathcal{A}$ *is rectangular, i.e.,* $\mathcal{A}=\left\{(\boldsymbol{a},\boldsymbol{b})|\ \boldsymbol{a}\in[\underline{\boldsymbol{a}},\overline{\boldsymbol{a}}],\ \boldsymbol{b}\in[\underline{\boldsymbol{b}},\overline{\boldsymbol{b}}],\ b_i=b_j,\ \forall i,j\in\mathcal{V}_n,\ \ \forall n\right\}$.

To deal with the robust problem in this context, we first consider the problem where we only seek prices that have a constant markup in each partition, i.e., $\mathbf{x}\in\mathbb{R}^m$ such that $x_i-c_i=x_j-c_j$ for all $i,j\in\mathcal{V}_n$, $n\in[N]$. Similar to the previous section, we will show that there is a solution $\mathbf{x}^*$ of this style such that if $(\mathbf{a}^*,\mathbf{b}^*)$ is an optimal solution to the adversary's problem under prices $\mathbf{x}^*$, then $(\mathbf{x}^*,\underline{\mathbf{a}},\overline{\mathbf{b}})$ is a saddle point of the robust problem and $\mathbf{x}^*$ is a robust optimal solution. In this context, Assumption 2 allows us to characterize solutions of the adversary's problem under the robust optimal prices. In brief, we show in the following that if the uncertainty set is rectangular, then under any prices that have a constant markup in each partition, there is an adversary solution whose each coordinate is equal to either its upper bound or its lower bound. Then, we show that if the vector of prices is optimal to the robust problem, then there is an solution to the corresponding adversary's problem $(\mathbf{a}^*,\mathbf{b}^*)$ such that $\mathbf{a}^*$ is equal to its lower bound and $\mathbf{b}^*$ is equal to its upper bound, which is nice because it allows us to convert the robust problem into an equivalent deterministic one.

Consider a robust pricing problem, in which we require the prices $\mathbf{x}$ to have a constant markup in each partition, i.e., $x_i-c_i=x_j-c_j$ for all $i,j\in\mathcal{V}_n$, $n\in[N]$. Let $z_n=x_i-c_i$ for all $i\in\mathcal{V}_n$ and $n\in[N]$. The robust problem becomes

$$\max_{\mathbf{z}\in\mathbb{R}^N}\left\{\min_{(\mathbf{a},\mathbf{b})\in\mathcal{A}}\sum_{n\in[N]}\sum_{i\in\mathcal{V}_n}z_nP_i(\mathbf{Y}^n(z_n+\mathbf{c},\mathbf{a},\mathbf{b})|G^n)\right\},$$

or equivalently

$$\max_{\mathbf{z}\in\mathbb{R}^N}\left\{\min_{(\mathbf{a},\mathbf{b})\in\mathcal{A}}\frac{\sum_{n\in[N]}z_nG^n(\mathbf{Y}^n|z_n,\mathbf{a},\mathbf{b})}{1+\sum_{n\in[N]}G^n(\mathbf{Y}^n|z_n,\mathbf{a},\mathbf{b})}\right\},\tag{8}$$

where $G^n(\mathbf{Y}^n|z,\mathbf{a},\mathbf{b})=G^n(Y_i,\ i\in\mathcal{V}_n)$ with $Y_i=e^{a_i-b_i(z_n+c_i)}$, for all $i\in\mathcal{V}_n$. For notational brevity, let

$$\rho(\mathbf{z},\mathbf{a},\mathbf{b})=\frac{\sum_{n\in[N]}z_nG^n(\mathbf{Y}^n|z_n,\mathbf{a},\mathbf{b})}{1+\sum_{n\in[N]}G^n(\mathbf{Y}^n|z_n,\mathbf{a},\mathbf{b})}.$$

In Lemma 4 below, we show that if we aim at maximizing function $G^n(Y^n|z_n,\mathbf{a}^n,\mathbf{b}^n)$, for any $n\in[N]$, over the uncertainty set $\mathcal{A}$, then we can pick the upper bound vector of $\mathbf{a}^n$ and lower bound vector of $\mathbf{b}^n$. On the other hand, if the goal is to minimize $G^n(Y^n|z_n,\mathbf{a}^n,\mathbf{b}^n)$, we can just take the lower bound vector of $\mathbf{a}^n$ and upper bound vector of $\mathbf{b}^n$. The role of this lemma is to support the following lemma where we show that if $(\mathbf{a}^*,\mathbf{b}^*)$ is optimal to the adversary problem, then each coordinate of $(\mathbf{a}^*,\mathbf{b}^*)$ is equal to either its upper bound or its lower bound.

Lemma 4. *Assume that Assumption 2 holds, given any $n \in [N]$, if $(\boldsymbol{a}^{n*}, \boldsymbol{b}^{n*})$ is a solution to the maximization problem $\max_{(\boldsymbol{a}^n, \boldsymbol{b}^n)\in\mathcal{A}^n} G^n(\boldsymbol{Y}^n|z, \boldsymbol{a}^n, \boldsymbol{b}^n)$, then $\boldsymbol{a}^{n*} = \overline{\boldsymbol{a}}^n$ and $\boldsymbol{b}^{n*} = \underline{\boldsymbol{b}}^n$. On the other hand, if $(\boldsymbol{a}^{n*}, \boldsymbol{b}^{n*})$ is a solution to minimization problem $\min_{(\boldsymbol{a}^n, \boldsymbol{b}^n)\in\mathcal{A}^n} G^n(\boldsymbol{Y}^n|z, \boldsymbol{a}^n, \boldsymbol{b}^n)$, then $\boldsymbol{a}^{n*} = \underline{\boldsymbol{a}}^n$ and $\boldsymbol{b}^{n*} = \overline{\boldsymbol{b}}^n$, where $\overline{\boldsymbol{a}}^n, \underline{\boldsymbol{a}}^n, \overline{\boldsymbol{b}}^n, \underline{\boldsymbol{a}}^n$ are sub-vectors of $\overline{\boldsymbol{a}}, \underline{\boldsymbol{a}}, \overline{\boldsymbol{b}}, \underline{\boldsymbol{b}}$, respectively.*

Proof: Using the same technique as in the proof of Lemma 2, we can show that function $f^{Gn}(\mathbf{s}) = G^n(\mathbf{Y}^n(\mathbf{s}))$, where $Y_i(\mathbf{s}) = e^{s_i}$, $\forall i \in \mathcal{V}_n$, is a monotonically increasing function in every coordinate. Since $\mathcal{A}^n$ is rectangular and $G^n(\mathbf{Y}^n|z, \mathbf{a}^n, \mathbf{b}^n) = G^n(Y_i = e^{a_i - b_i(z+c_i)}|\ i \in \mathcal{V}_n)$, we easily have the following result

$$G^n(\mathbf{Y}^n|z, \overline{\mathbf{a}}^n, \underline{\mathbf{b}}^n) \geq G^n(\mathbf{Y}^n|z, \mathbf{a}^n, \mathbf{b}^n),\ \forall (\mathbf{a}^n, \mathbf{b}^n) \in \mathcal{A}^n,$$

Or, equivalently,

$$(\overline{\mathbf{a}}^n, \underline{\mathbf{b}}^n) = \text{argmax}_{(\mathbf{a}^n, \mathbf{b}^n)\in\mathcal{A}^n} G^n(\mathbf{Y}^n|z, \mathbf{a}^n, \mathbf{b}^n) \tag{9}$$

Moreover, from Proposition 1, we know that $G^n(\mathbf{Y}^n|z, \mathbf{a}^n, \mathbf{b}^n)$ is strictly convex in $\mathcal{A}^n$, meaning that $(\overline{\mathbf{a}}^n, \underline{\mathbf{b}}^n)$ is the unique solution of the maximization problem in (9). By a similar way, we can also show that $(\underline{\mathbf{a}}^n, \overline{\mathbf{b}}^n)$ is the unique solution to the minimization problem $\min_{(\mathbf{a}^n, \mathbf{b}^n)\in\mathcal{A}^n} G^n(\mathbf{Y}^n|z, \mathbf{a}^n, \mathbf{b}^n)$. This completes the proof. □

Lemma 5 below shows that under any prices, one of the vector in $\{(\mathbf{a}, \mathbf{b}) \in \mathcal{A}|\ a_i \in \{\underline{a}_i, \overline{a}_i\},\ b_i \in \{\underline{b}_i, \overline{b}_i\}\}$ is optimal to the adversary's problem, which also means that we can solve the adversary problem by searching over 2^{m+n} possible solutions. This lemma is an important step towards the result showing that under the robust optimal prices, vector $(\underline{\mathbf{a}}, \overline{\mathbf{b}})$ is optimal to the adversary's problem.

Lemma 5. *Assume that Assumption 2 holds and let $(\boldsymbol{a}^*, \boldsymbol{b}^*)$ be a solution to the corresponding adversary's problem of (8), then for any $n \in [N]$, we have*

$$(a_i^*, b_i^*) = \begin{cases} (\underline{a}_i, \overline{b}_i) & \text{if } \rho(\boldsymbol{z}, \boldsymbol{a}^*, \boldsymbol{b}^*) < z_n \\ (\overline{a}_i, \underline{b}_i) & \text{if } \rho(\boldsymbol{z}, \boldsymbol{a}^*, \boldsymbol{b}^*) > z_n \end{cases} \qquad \forall i \in \mathcal{V}_n.$$

Moreover, if $\rho(\boldsymbol{z}, \boldsymbol{a}^*, \boldsymbol{b}^*) = z_n$ then all the solutions in the following set are optimal to the adversary's problem

$$S^* = \{(\boldsymbol{a}, \boldsymbol{b}) \in \mathcal{A}|\ (a_j, b_j) = (a_j^*, b_j^*), \forall j \notin \mathcal{V}_n\}$$

Proof: We denote $(\mathbf{a}^{n*}, \mathbf{b}^{n*})$ as the sub-vectors of $(\mathbf{a}^*, \mathbf{b}^*)$ associated with partition n-th, i.e., $(\mathbf{a}^{n*}, \mathbf{b}^{n*}) = \{(a_i^*, b_i^*),\ i \in \mathcal{V}_n\}$. Now, given any $n \in [N]$, let

$$A = \sum_{n' \in [N], n' \neq n} z_n G^n(\mathbf{Y}^n|z_n, \mathbf{a}^*, \mathbf{b}^*)$$
$$B = 1 + \sum_{n' \in [N], n' \neq n} G^n(\mathbf{Y}^n|z_n, \mathbf{a}^*, \mathbf{b}^*).$$

We can write

$$\rho(\mathbf{z},\mathbf{a}^*,\mathbf{b}^*) = \frac{A + z_n G^n(\mathbf{Y}^n|z_n,\mathbf{a}^*,\mathbf{b}^*)}{B + G^n(\mathbf{Y}^n|z_n,\mathbf{a}^*,\mathbf{b}^*)}$$

We now prove the lemma by considering the following three cases:

(i) If $\rho(\mathbf{z},\mathbf{a}^*,\mathbf{b}^*) < z_n$, then for any $\gamma < G^n(\mathbf{Y}^n|z_n,\mathbf{a}^*,\mathbf{b}^*)$ one can easily show the following inequality

$$\rho(\mathbf{z},\mathbf{a}^*,\mathbf{b}^*) = \frac{A + z_n G^n(\mathbf{Y}^n|z_n,\mathbf{a}^*,\mathbf{b}^*)}{B + G^n(\mathbf{Y}^n|z_n,\mathbf{a}^*,\mathbf{b}^*)} > \frac{A + z_n\gamma}{B+\gamma}$$

Since $(\mathbf{a}^*,\mathbf{b}^*)$ is a solution to the adversary problem (8),$(\mathbf{a}^{n*},\mathbf{b}^{n*})$ must be a solution to the minimization problem $\min_{(\mathbf{a}^n,\mathbf{b}^n)\in\mathcal{A}^n} G^n(\mathbf{Y}^n|z_n,\mathbf{a}^n,\mathbf{b}^n)$. According to Lemma 4, we have $(a_i^*,b_i^*) = (\underline{a}_i,\overline{b}_i)$ for all $i\in\mathcal{V}_n$.

(ii) If $\rho(\mathbf{z},\mathbf{a}^*,\mathbf{b}^*) > z_n$, then similarly to the previous case, we can show that, for any $\gamma > G^n(\mathbf{Y}^n|z_n,\mathbf{a}^*,\mathbf{b}^*)$ one can easily show the following inequality

$$\rho(\mathbf{z},\mathbf{a}^*,\mathbf{b}^*) > \frac{A + z_n\gamma}{B+\gamma}.$$

Thus, $(\mathbf{a}^{n*},\mathbf{b}^{n*})$ must be a solution to the maximization problem $\max_{(\mathbf{a}^n,\mathbf{b}^n)\in\mathcal{A}^n} G^n(\mathbf{Y}^n|z_n,\mathbf{a}^n,\mathbf{b}^n)$. Again, using Lemma 4, we have $(a_i^*,b_i^*) = (\overline{a}_i,\underline{b}_i)$ for all $i\in\mathcal{V}_n$.

(iii) If $\rho(\mathbf{z},\mathbf{a}^*,\mathbf{b}^*) = z_n$, then for any $\gamma\in\mathbb{R}$ we have

$$\rho(\mathbf{z},\mathbf{a}^*,\mathbf{b}^*) = \frac{A + z_n\gamma}{B+\gamma},$$

meaning that any solution $(\mathbf{a},\mathbf{b})\in\mathcal{A}$ such that $(a_i,b_i) = (a_i^*,b_i^*)$, for all $i\neq\mathcal{V}_n$, is optimal to the adversary problem (8).

Combining the above three cases, we obtain the desired result. □

Since we want to prove that under robust optimal prices $\mathbf{z}^*$, the solution $(\underline{\mathbf{a}},\overline{\mathbf{b}})$ should be optimal to the adversary's problem, Lemma 5 tells us that we need to show $\rho(\mathbf{z}^*,\underline{\mathbf{a}},\overline{\mathbf{b}}) \le z_n^*$ for all $n\in[N]$. Before presenting this result, we need the following lemma.

Lemma 6. *Given $\overline{\boldsymbol{z}}\in\mathbb{R}^N$, $(\boldsymbol{a},\boldsymbol{b})\in\mathcal{A}$ and $k\in[N]$, if $\rho(\overline{\boldsymbol{z}},\boldsymbol{a},\boldsymbol{b})\ge\overline{z}_k$, then*

$$\left.\frac{\partial\rho(\boldsymbol{z},\boldsymbol{a},\boldsymbol{b})}{\partial z_k}\right|_{\boldsymbol{z}=\overline{\boldsymbol{z}}} > 0.$$

Proof: For notational brevity, let $A = \sum_{n\in[N]}\overline{z}_n G^n(\mathbf{Y}^n|\overline{z}_n,\mathbf{a},\mathbf{b})$ and $B = 1 + \sum_{n\in[N]} G^n(\mathbf{Y}^n|\overline{z}_n,\mathbf{a},\mathbf{b})$. Taking the first derivative of $\rho(\mathbf{z},\mathbf{a},\mathbf{b})$ w.r.t. z_k we obtain

$$\begin{aligned}\left.\frac{\partial\rho(\mathbf{z},\mathbf{a},\mathbf{b})}{\partial z_k}\right|_{\mathbf{z}=\overline{\mathbf{z}}} &= \frac{\left(G^k(\overline{z}_k,\mathbf{a},\mathbf{b}) - \overline{z}_k\sum_{i\in\mathcal{V}_k} b_iY_i\partial G_i^k(\mathbf{Y}^k)\right)B + A\left(\sum_{i\in\mathcal{V}_k} b_iY_i\partial G_i^k(\mathbf{Y}^k)\right)}{B^2}\\ &= \frac{1}{B^2}\left(G^k(\overline{z}_k,\mathbf{a},\mathbf{b}) + \left(\sum_{i\in\mathcal{V}_k} b_iY_i\partial G_i^k(\mathbf{Y}^k)\right)(A-\overline{z}_kB)\right) > 0.\end{aligned}$$

The last inequality is due to $\rho(\overline{\mathbf{z}},\mathbf{a},\mathbf{b}) = A/B \ge \overline{z}_k$. □

We are now ready to show that $(\underline{\mathbf{a}}, \overline{\mathbf{b}})$ is an optimal solution to the adversary's problem under the robust optimal prices (Lemma 7).

Lemma 7. *Under Assumption 2, $(\underline{\mathbf{a}}, \overline{\mathbf{b}})$ is the* ***unique*** *optimal solution to the adversary's problem of (8) under robust optimal prices.*

Proof: Let $\mathbf{z}^*$ be an optimal solution to (8), we first prove that $(\underline{\mathbf{a}}, \overline{\mathbf{b}})$ is optimal to the adversary's problem under $\mathbf{z}^*$. Let $\mathcal{A}^*$ be the set of optimal solutions to the adversary's problem of (8) under $\mathbf{z}^*$, i.e.,

$$\mathcal{A}^* = \underset{(\mathbf{a},\mathbf{b})\in\mathcal{A}}{\text{argmin}} \quad \rho(\mathbf{z}^*, \mathbf{a}, \mathbf{b}).$$

Let also $f(\mathbf{z}) = \text{argmin}_{(\mathbf{a},\mathbf{b})\in\mathcal{A}}\ \rho(\mathbf{z}, \mathbf{a}, \mathbf{b})$. We will prove that $f(\mathbf{z}^*) < z^*_n$ for all $n \in [N]$. By contradiction, assume that there exists a set $n \in [N]$ such that $\rho(\mathbf{x}^*, \mathbf{a}^*, \mathbf{b}^*) \geq z^*_n$. Let

$$k = \text{argmax}\{z^*_n|\ n \in [N], f(\mathbf{z}^*) > z^*_n\}$$
$$h = \text{argmin}\{z^*_n|\ n \in [N], f(\mathbf{z}^*) < z^*_n\}.$$

Indeed, h always exists because $\rho(\mathbf{z}^*, \mathbf{a}^*, \mathbf{b}^*) < \max_{n\in[N]} z^*_n$. We consider two following cases

(i) If such k exists. We have $z^*_h > \rho(\mathbf{z}^*, \mathbf{a}^*, \mathbf{b}^*) > z^*_k$ and for any $n \neq h$ and $n \neq k$ we have either $\rho(\mathbf{z}^*, \mathbf{a}^*, \mathbf{b}^*) = z^*_n$ or $z^*_k \geq z^*_n$ or $z^*_h \leq z^*_n$. Moreover, the function $f(\mathbf{z})$ is continuous in $\mathbf{x}$ (Theorem 7, Hogan 1973). So, there is $\delta > 0$ such that

$$z^*_h > f(\mathbf{z}^* + t\mathbf{e}^k) > z^*_k + t,\ \forall t \in [0, \delta],$$

where $\mathbf{e}^k$ is a vector of size N with zero entries except the k-th element is equal to 1. As a result, for any $n \in [N]$ such that $z^*_n \geq z^*_h$ we have $z^*_n > f(\mathbf{z}^* + t\mathbf{e}^k)$ and if $z^*_n \leq z^*_k$ we have $z^*_n < f(\mathbf{z}^* + t\mathbf{e}^k)$. From Lemma (5), this means that there is a solution $(\mathbf{a}^*, \mathbf{b}^*) \in \mathcal{A}^*$ that is optimal to the adversary's problem $\{\text{argmin}_{(\mathbf{a},\mathbf{b})\in\mathcal{A}} \rho(\mathbf{z}^* + t\mathbf{e}^k, \mathbf{a}, \mathbf{b})\}$, for all $t \in [0, \delta]$. This leads to

$$f(\mathbf{z}^* + t\mathbf{e}^k) = \rho(\mathbf{z}^* + t\mathbf{e}^k, \mathbf{a}^*, \mathbf{b}^*) \leq \rho(\mathbf{z}^*, \mathbf{a}^*, \mathbf{b}^*),\ \forall t \in [0, \delta],$$

which is contradictory to the result of Lemma 6, which says that the partial derivative of $\rho(\mathbf{z}^*, \mathbf{a}^*, \mathbf{b}^*)$ is positive.

(ii) If such k does not exist, we also have that for any $n \in [N]$, either $f(\mathbf{z}^*) < z^*_h \leq z^*_n$ or $f(z^*) = z^*_n$. Similar to the previous case, we also have the result that there exists $\delta > 0$ such that $z^*_h > f(\mathbf{a}^* + t\mathbf{e}^k)$ for all $t \in (0, \delta)$, which also leads to the result that there is $(\mathbf{a}^*, \mathbf{b}^*) \in \mathcal{A}^*$ being optimal to the adversary's problem under $\mathbf{z}^* + t\mathbf{e}^k$. Using Lemma 6, there is $t \in (0, \delta]$ such that $\rho(\mathbf{z}^*, \mathbf{a}^*, \mathbf{b}^*) < \rho(\mathbf{z}^* + t\mathbf{e}^k, \mathbf{a}^*, \mathbf{b}^*) = f(\mathbf{z}^* + t\mathbf{e}^k)$, which is contradictory to our initial assumption that $\mathbf{a}^*$ is optimal to the robust problem.

So in summary, we can claim that $f(\mathbf{z}^*) < z_n^*$, for all $n \in [N]$ and Lemma 5 tells us that $(\underline{\mathbf{a}}, \overline{\mathbf{b}})$ is the unique optimal solution to the adversary's problem of (8), which is the desired result. □

Lemma 8. *The robust problem* (8) *is equivalent to* $\max_{\boldsymbol{z}} \rho(\boldsymbol{z}, \underline{\boldsymbol{a}}, \overline{\boldsymbol{b}})$.

Proof: We use the same notations as in the proof of Lemma 7. We need to prove that if $\mathbf{z}^*$ is an robust optimal solution to (8), then it is also optimal to $\max_{\mathbf{z}} \rho(\mathbf{z}, \underline{\mathbf{a}}, \overline{\mathbf{b}})$. To facilitate our exposition, let $\overline{\mathcal{A}} = \{(\mathbf{a}, \mathbf{b}) |\ a_i \in \{\underline{a_i}, \overline{a_i}\}$ and $b_i \in \{\underline{b_i}, \overline{b_i}\}\}$ (i.e. $\overline{\mathcal{A}}$ contains points whose each coordinate is either equal to its lower bound or upper bound). We also define

$$\tau = \min_{(\mathbf{a},\mathbf{b}) \in \overline{\mathcal{A}}} \left\{ \rho(\mathbf{z}^*, \mathbf{a}, \mathbf{b}) - \rho(\mathbf{z}^*, \underline{\mathbf{a}}, \overline{\mathbf{b}}) \middle|\ \rho(\mathbf{z}^*, \mathbf{a}, \mathbf{b}) > \rho(\mathbf{z}^*, \underline{\mathbf{a}}, \overline{\mathbf{b}}) \right\} \tag{10}$$

Now, by contradiction, assume that $\mathbf{z}^* \notin \max_{\mathbf{z}} \rho(\mathbf{z}, \underline{\mathbf{a}}, \overline{\mathbf{b}})$. since the problem $\max_{\mathbf{z}} \rho(\mathbf{z}, \underline{\mathbf{a}}, \overline{\mathbf{b}})$ has a unique local optimum (Theorem C.1 in Zhang et al. (2018)), $\nabla_{\mathbf{z}} \rho(\mathbf{z}^*, \underline{\mathbf{a}}, \overline{\mathbf{b}}) \neq 0$. Thus, there always exits a vector $\boldsymbol{\epsilon} \in \mathbb{R}^n \neq 0$ and a constant $\delta > 0$ such that

$$\rho(\mathbf{z}^*, \underline{\mathbf{a}}, \overline{\mathbf{b}}) < \rho(\mathbf{z}^* + t\boldsymbol{\epsilon}, \underline{\mathbf{a}}, \overline{\mathbf{b}}), \forall t \in (0, \delta). \tag{11}$$

Moreover, since $f(\mathbf{z})$ and $\rho(\mathbf{z}, \mathbf{a}, \mathbf{b})$ are continuous in $\mathbf{z}$ (Theorem 7, Hogan 1973), we can choose $t \in (0, \delta)$ such that

$$\begin{cases} |f(\mathbf{z}^*) - f(\mathbf{z}^* + t\boldsymbol{\epsilon})| < \tau/2 \\ |\rho(\mathbf{z}^*, \mathbf{a}^t, \mathbf{b}^t) - \rho(\mathbf{z}^* + t\boldsymbol{\epsilon}, \mathbf{a}^t, \mathbf{b}^t)| < \tau/2 \end{cases} \tag{12}$$

where $(\mathbf{a}^t, \mathbf{b}^t)$ is an optimal solution to the adversary's problem under prices $\mathbf{z}^* + t\boldsymbol{\epsilon}$. From Lemma 5 we can choose $(\mathbf{a}^t, \mathbf{b}^t) \in \overline{\mathcal{A}}$. By the selection of t in (12) we have

$$|\rho(\mathbf{z}^*, \underline{\mathbf{a}}, \overline{\mathbf{b}}) - \rho(\mathbf{z}^*, \mathbf{a}^t, \mathbf{b}^t)| \leq |\rho(\mathbf{z}^*, \underline{\mathbf{a}} - \rho(\mathbf{z}^* + t\boldsymbol{\epsilon}, \mathbf{a}^t, \mathbf{b}^t)| + |\rho(\mathbf{z}^* + t\boldsymbol{\epsilon}, \mathbf{a}^t, \mathbf{b}^t) - \rho(\mathbf{z}^*, \mathbf{a}^t, \mathbf{b}^t)| < \tau.$$

Combine with the definition of τ in (10), we have that $(\mathbf{a}^t, \mathbf{b}^t)$ is also an optimal solution to the adversary's problem under $\mathbf{z}^*$. Lemma 5 tells us that $(\underline{\mathbf{a}}, \overline{\mathbf{b}})$ is the unique optimal solution to the adversary's problem under $\mathbf{z}^*$, so we have $(\mathbf{a}^t, \mathbf{b}^t) = (\underline{\mathbf{a}}, \overline{\mathbf{b}})$. Combine with (11) we have $f(\mathbf{z}^*) < f(\mathbf{z}^* + t\boldsymbol{\epsilon})$, which is contradictory to the assumption that $\mathbf{z}^*$ is optimal to the robust problem (8). This completes the proof. □

Now we are ready for the main result. The following theorem shows that there is a partition-wise constant markup solution that is optimal to the robust problem (RO). This is a direct outcome from Lemmas 7 and (8).

Theorem 2. (If the PSP are partition-wise homogeneous, a partition-wise constant markup solution is optimal). *Assume that Assumption 2 holds and let* R^* *be the unique solution of the equation*

$$R = \sum_{n \in [N]} \frac{1}{\langle \overline{\boldsymbol{b}}^n \rangle} e^{-\langle \overline{b}^n \rangle R - 1} G^n(\boldsymbol{Y}^n | 0, \underline{\boldsymbol{a}}, \overline{\boldsymbol{b}})$$

then $\boldsymbol{x}^* \in \mathbb{R}^m$ *such that* $x_i^* = c_i + 1/\langle \overline{\boldsymbol{b}}^n \rangle + R^*$, $\forall i \in \mathcal{V}_n$, $\forall n \in [N]$, *is the* ***unique*** *optimal solution to the robust problem* (RO).

Proof: According to Theorem C.1 of Zhang et al. (2018) we see that $\mathbf{x}^*$ is also an optimal solution the problem $\max_{\mathbf{x}} \Phi(\mathbf{x}, \underline{\mathbf{a}}, \overline{\mathbf{b}})$. Moreover, if we define $\mathbf{z}^* \in \mathbb{R}^N$ such that $z_n^* = x_i^* - c_i$ for all $n \in [N], i \in \mathcal{V}_n$, then $\mathbf{z}^*$ is also optimal to the problem $\max_{\mathbf{z}} \rho(\mathbf{z}, \underline{\mathbf{a}}, \overline{\mathbf{b}})$. Hence, $\mathbf{z}^*$ is also a robust optimal solution to (8) (Lemma 8). As a result, from Lemma 7 we see that $(\underline{\mathbf{a}}, \overline{\mathbf{b}})$ is an optimal solution to the adversary's problem of (RO) under prices $\mathbf{x}^*$. So, we have $g(\mathbf{x}^*) = \Phi(\mathbf{x}^*, \underline{\mathbf{a}}, \overline{\mathbf{b}})$ (recall that $g(\mathbf{x})$ is the adversary's optimal objective value under prices $\mathbf{x} \in \mathbb{R}^m$). Under the assumption that the PSP are homogeneous in each partition $\mathcal{V}_n$, Zhang et al. (2018) (Theorem C.1) show that $\mathbf{x}^*$ is also optimal to the deterministic problem

$$\max_{\mathbf{x} \in \mathbb{R}^m} \quad \Phi(\mathbf{x}, \underline{\mathbf{a}}, \overline{\mathbf{b}}).$$

So, for any $\mathbf{x} \in \mathbb{R}^m$, we have

$$\begin{aligned} g(\mathbf{x}^*) = \Phi(\mathbf{x}^*, \underline{\mathbf{a}}, \overline{\mathbf{b}}) &\geq \Phi(\mathbf{x}, \underline{\mathbf{a}}, \overline{\mathbf{b}}) \\ &\geq \min_{(\mathbf{a},\mathbf{b}) \in \mathcal{A}} \Phi(\mathbf{x}, \mathbf{a}, \mathbf{b}) = g(\mathbf{x}). \end{aligned}$$

So, $g(\mathbf{x}^*) \geq g(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^m$. Moreover, if $\mathbf{x} \neq \mathbf{x}^*$, then from Zhang et al. (2018) we have $\Phi(\mathbf{x}^*, \underline{\mathbf{a}}, \overline{\mathbf{b}}) > \Phi(\mathbf{x}, \underline{\mathbf{a}}, \overline{\mathbf{b}})$, which means $g(\mathbf{x}^*) > g(\mathbf{x})$. So, $\mathbf{x}^*$ is the unique optimal solution to the robust problem, as required. □

Theorem 2 indicates that to solve the robust problem under a rectangular uncertainty set and partition-wise PSP, we just need to pick $(\underline{\mathbf{a}}, \overline{\mathbf{b}})$ for the choice parameters and solve the corresponding deterministic pricing problem. Thus, the robust problem in this context is not difficult to solve as compared to the deterministic counterpart.

In the next two corollaries, we provide formulas to compute the robust optimal prices for the case of the MNL and nested logit models. These formulas are a direct result from Theorem 2 and generalize previous studies on the pricing problem under the MNL and nested logit models (Keller 2013, Li and Huh 2011).

Corollary 1. (Robust solutions for the MNL-based robust pricing problem with heterogeneous price sensitivity parameters). *If the choice model is MNL, the PSP are heterogeneous over products and the uncertainty set is rectangular, i.e.,* $\mathcal{A} = \{(\boldsymbol{a}, \boldsymbol{b}) |\ \boldsymbol{a} \in [\underline{\boldsymbol{a}}, \overline{\boldsymbol{a}}],\ \boldsymbol{b} \in [\underline{\boldsymbol{b}}, \overline{\boldsymbol{b}}]\}$, *then* $\boldsymbol{x}^*$ *defined below is an optimal solution to the robust problem* (RO): $x_i^* = c_i + \frac{1}{/}\overline{b}_i + R^*$, *where* R^* *is a unique value of* R *satisfying* $R = \sum_{n \in [N]} \frac{1}{\overline{b}_i} \exp(\underline{a}_i - \overline{b}_i(c_i + R) - 1)$.

If the choice model is a nested-logit model of N nests with GEV-CPGF function

$$G(\mathbf{Y}) = \sum_{n=1}^{N} \left(\sum_{i\in\mathcal{V}_n} \sigma_{in} Y_i^{\mu_n} \right)^{\mu/\mu_n}, \tag{13}$$

where $\mu, \mu_n > 0, \sigma_{in}, \forall n \in [N], i \in \mathcal{V}_n$, are some model parameters.

Corollary 2. (Robust solutions for the robust pricing problem under nested-logit when the PSP are partially heterogeneous). *If the choice model is a nested-logit with GEV-CPGF (13), the PSP are homogeneous in each nest and the uncertainty set is rectangular, i.e.,* $\mathcal{A} = \{(\boldsymbol{a},\boldsymbol{b})|\ \boldsymbol{a}\in[\underline{\boldsymbol{a}},\overline{\boldsymbol{a}}],\ \boldsymbol{b}\in[\underline{\boldsymbol{b}},\overline{\boldsymbol{b}}]\}$*, then* $\boldsymbol{x}^*$ *defined below is an optimal solution to the robust problem* (RO)*:* $x_i^* = c_i + 1/\langle \overline{\boldsymbol{b}}^n\rangle + R^*, \forall i\in\mathcal{V}_n, n = 1,\ldots,N$*, where* R^* *is a unique value of* R *satisfying*

$$R = \sum_{n\in[N]} \frac{1}{\langle \overline{\boldsymbol{b}}^n\rangle} \left(\sum_{i\in\mathcal{V}_n} \sigma_{in} e^{(\underline{a}_i - \langle \overline{\boldsymbol{b}}^n\rangle c_i)\mu_n} \right)^{\mu/\mu_n} e^{-\langle \overline{\boldsymbol{b}}^n\rangle R - 1}.$$

We also give some comparative insights that describe how the robust optimal value and solution change as a function of the uncertainty set. To facilitate the comparison, let $\widetilde{\mathbf{x}}^*(\mathcal{A})$ denote the robust solution given in Theorem 2, the following theorem shows that, in case that the PSP are partition-wise homogeneous, a larger uncertainty set leads to smaller robust prices.

Proposition 4. (When the PSP are partition-wise homogeneous, larger uncertainty set leads to smaller robust optimal prices). *Given two convex uncertainty set* $\mathcal{A}_1, \mathcal{A}_2$ *satisfying Assumption 2, if* $\mathcal{A}_1 \subset \mathcal{A}_2$*, then* $\Gamma(\mathcal{A}_1) \ge \Gamma(\mathcal{A}_2)$ *and* $\widetilde{\boldsymbol{x}}^*(\mathcal{A}_1) \succeq \widetilde{\boldsymbol{x}}^*(\mathcal{A}_2)$.

Proof: Similarly to the proof of Theorem 3, the inequality $\Gamma(\mathcal{A}_1) \ge \Gamma(\mathcal{A}_2)$ is easy to verify. To prove the inequality associated with the optimal prices, we write $\mathcal{A}_1 = \{(\mathbf{a},\mathbf{b})|\ \mathbf{a}\in[\underline{\mathbf{a}}^1,\overline{\mathbf{a}}^1],\ \mathbf{b}\in[\underline{\mathbf{b}}^1,\overline{\mathbf{b}}^1]\}$ and $\mathcal{A}_2 = \{(\mathbf{a},\mathbf{b})|\ \mathbf{a}\in[\underline{\mathbf{a}}^2,\overline{\mathbf{a}}^2],\ \mathbf{b}\in[\underline{\mathbf{b}}^2,\overline{\mathbf{b}}^2]\}$. For each partition $n\in[N]$, let $\overline{\mathbf{b}}^{1,n}$ and $\overline{\mathbf{b}}^{2,n}$ be the corresponding sub-vectors of $\overline{\mathbf{b}}^1$ and $\overline{\mathbf{b}}^2$, respectively. Using the results from Theorem 2 and Theorem C.1 of Zhang et al. (2018), we can show that, for any partition $n\in[N]$,

$$\widetilde{\mathbf{x}}^*(\mathcal{A}_1)_i = c_i + \frac{1}{\langle \overline{\mathbf{b}}^{1,n}\rangle} + \Gamma(\mathcal{A}_1),\ \forall i\in\mathcal{V}_n$$
$$\widetilde{\mathbf{x}}^*(\mathcal{A}_2)_i = c_i + \frac{1}{\langle \overline{\mathbf{b}}^{2,n}\rangle} + \Gamma(\mathcal{A}_2),\ \forall i\in\mathcal{V}_n.$$

Moreover, since $\mathcal{A}_1 \subset \mathcal{A}_2$, we have $\langle \overline{\mathbf{b}}^{1,n}\rangle \le \langle \overline{\mathbf{b}}^{2,n}\rangle$. Using the inequality $\Gamma(\mathcal{A}_1)\ge\Gamma(\mathcal{A}_2)$, we obtain the desired result $\widetilde{\mathbf{x}}^*(\mathcal{A}_1) \succeq \widetilde{\mathbf{x}}^*(\mathcal{A}_2)$ □

The proof of Proposition 4 also tells us that $\widetilde{\mathbf{x}}^*(\mathcal{A}_1)_i - \widetilde{\mathbf{x}}^*(\mathcal{A}_2)_i \ge 1/\langle \overline{\mathbf{b}}^{1,n}\rangle - 1/\langle \overline{\mathbf{b}}^{2,n}\rangle$, meaning that the robust price of each item will strictly decrease if we increase the upper bound of its PSP in the uncertainty set.

In Proposition 5 below, we try to relax Assumption 2, i.e., the rectangularity of the uncertainty set. In general, if the uncertainty set is not rectangular, we are able to show that there might be a vector of partition-wise constant markup prices being optimal to the robust problem. However, such a vector might not exist due to the fact that the adversary's problem is no-longer convex.

Proposition 5. *If there exist* $\boldsymbol{z}^* \in \mathbb{R}^N$, $\boldsymbol{a}^*, \boldsymbol{b}^* \in \mathbb{R}^m$, *and* $R^* \in \mathbb{R}_+$ *such that*

$$(\boldsymbol{a}^*, \boldsymbol{b}^*) = argmax_{(\boldsymbol{a},\boldsymbol{b})\in\mathcal{A}} \rho(\boldsymbol{z}^*, \boldsymbol{a}, \boldsymbol{b}) \tag{14}$$

$$R^* = \sum_{n\in[N]} \frac{1}{\langle \boldsymbol{b}^{n*}\rangle} e^{-\langle b^{n*}\rangle R^* - 1} G^n(\boldsymbol{Y}^n|0, \boldsymbol{a}^*, \boldsymbol{b}^*) \tag{15}$$

$$z_i^* = \frac{1}{\langle \boldsymbol{b}^{n*}\rangle} + R^*, \;\; \forall i \in \mathcal{V}_n, n\in[N], \tag{16}$$

then vector $\boldsymbol{x}^* = \boldsymbol{c} + \boldsymbol{z}^*$ *is an optimal solution to the robust problem* (8).

Proof: Using Theorem C.1 in Zhang et al. (2018), we see that $\mathbf{x}^*$ is an optimal solution to the deterministic problem $\max_{\mathbf{x}\in\mathbb{R}^m} \rho(\mathbf{x}, \mathbf{a}^*, \mathbf{b}^*)$. Thus, we have the following chain of inequalities for any price vector $\mathbf{x} \in \mathbb{R}^m$

$$\begin{aligned} g(\mathbf{x}^*) &= \rho(\mathbf{z}^*, \mathbf{a}^*, \mathbf{b}^*) \\ &\geq \rho(\mathbf{x} - \mathbf{c}, \mathbf{a}^*, \mathbf{b}^*) \\ &\geq \min_{(\mathbf{a},\mathbf{b})\in\mathcal{A}} \rho(\mathbf{x} - \mathbf{c}, \mathbf{a}, \mathbf{b}) \\ &= g(\mathbf{x}), \end{aligned}$$

which implies that $\mathbf{x}^*$ is optimal to the robust problem (8), as desired. □

Proposition 5 suggests that we might solve the robust problem in (8) under a general uncertainty set by solving the system of equations (14)-(16). However, this is not computational tractable, as the function $\rho(\mathbf{z}^*, \mathbf{a}, \mathbf{b})$ in (14) is not convex in $(\mathbf{a}, \mathbf{b})$ under the heterogeneity setting, even when the choice model is MNL. Note that when there is only one partition (the PSP are homogeneous), Proposition 5 becomes Theorem 1. In this context, the function $\rho(\mathbf{z}^*, \mathbf{a}, \mathbf{b})$ is strictly convex in $(\mathbf{a}, \mathbf{b})$ and there always exist $\mathbf{z}^*, \mathbf{a}^*, \mathbf{b}^*$ satisfying (14)-(16). Furthermore, if the uncertainty set $\mathcal{A}$ is rectangular (Assumption 2), then we have shown that $(\mathbf{a}^*, \mathbf{b}^*)$ can be determined under the optimal constant markups $\mathbf{z}^*$, and Proposition 5 becomes Theorem 2. In the case that the uncertainty set is singleton, the proposition becomes Theorem C.1 in Zhang et al. (2018).

4. Robust Pricing with Over-expected-sale Penalties

Motivated by applications in inventory considerations (Gallego and Hu 2014), we are interested in a robust model for the pricing problem with expected sale requirements under uncertain choice parameters $(\mathbf{a}, \mathbf{b})$. Given the fact that there may be no fixed prices such that the corresponding expected sale constraints are always satisfied when the choice parameters vary in the uncertainty set, (see Appendix A), we propose the version with over-expected-sale penalties, which allows us to handle both the expected sale requirements and the uncertainty issue. Our idea is to put the expected sale constraints to the objective function, i.e., we do not force the purchase probabilities to be in a feasible set, but instead we add penalties for purchase probabilities violating the constraints. More precisely, we consider the objective function $\Phi(\mathbf{x},\mathbf{a},\mathbf{b}) - \sum_{t=1}^{T}\lambda_t \max\{0,(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p} - r_t\}$, where $\lambda_t \geq 0$, $t = 1,\ldots,T$, are penalty parameters. In this objective function, if a constraint is violated, i.e., $(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p} > r_t$, then a cost $-\lambda_t \max\{0,(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p} - r_t\}$ is added to the expected revenue. In general, if we choose λ_t large enough, we will need a vector of purchase probabilities satisfying all the expected sale constraints to obtain high objective values. The deterministic pricing problem under the above objective function is

$$\max_{\mathbf{x}\in\mathbb{R}^m} \left\{ \Phi(\mathbf{x},\mathbf{a},\mathbf{b}) - \sum_{t=1}^{T} \lambda_t \max\{0,(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p} - r_t\} \right\}, \tag{17}$$

which can be formulated as the convex optimization problem

$$\begin{aligned} \max_{\mathbf{p},\mathbf{y}} \quad & \sum_{i\in\mathcal{V}} \left(\mathbf{x}(\mathbf{p}|\mathbf{a},\mathbf{b},G)_i - c_i\right) p_i - \sum_{t=1}^{T}\lambda_t y_t \\ \text{subject to} \quad & (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p} - y_t \leq r_t \\ & \sum_{i\in\mathcal{V}} p_i \leq 1 \\ & \mathbf{p},\mathbf{y} \geq 0. \end{aligned} \tag{18}$$

Before moving to a robust version, we investigate some characteristics of the deterministic pricing problem with penalties (18). First, let us denote by v^* and $\mathbf{p}^*$ the optimal value and optimal solution of the standard pricing problem under expected sale constraints (28) and $v^{\boldsymbol{\lambda}}$ and $\mathbf{p}^{\boldsymbol{\lambda}}$ the optimal value and optimal solution to the pricing problem with over-expected-sale penalties (18). Theorem 3 below shows that the expected value given by (18) will converges to the optimal expected revenue given by the constrained pricing problem (28) when λ_t, $\forall t \in [T]$, increase to infinity.

Theorem 3. *For any $\epsilon > 0$, we have*

(i) For any $\boldsymbol{\lambda}^1, \boldsymbol{\lambda}^2 \in \mathbb{R}^T_+$ such that $\boldsymbol{\lambda}^1 - \boldsymbol{\lambda}^2 = \epsilon \mathbf{1}$,

$$\sum_t \max\{0,(\boldsymbol{\alpha}^t)^T\boldsymbol{p}^{\boldsymbol{\lambda}^1} - r_t\} \leq \sum_t \max\{0,(\boldsymbol{\alpha}^t)^T\boldsymbol{p}^{\boldsymbol{\lambda}^2} - r_t\},$$

where $\mathbf{1}$ is a unit vector of appropriate size.

(ii) $v^{\boldsymbol{\lambda}} \geq v^*$ for all $\boldsymbol{\lambda} \in \mathbb{R}^T_+$ and if $\lambda_0 = \min_{t\in[T]} \lambda_t \geq (\Delta^* - v^*)/\epsilon$ then $\sum_t \max\{0,(\boldsymbol{\alpha}^t)^T\boldsymbol{p}^{\boldsymbol{\lambda}} - r_t\} \leq \epsilon$, where $\Delta^* = \max_{\boldsymbol{x}} \Phi(\boldsymbol{x},\boldsymbol{a},\boldsymbol{b})$.

(iii) Assume that there are positive constant L_i, l_i, $i \in \mathcal{V}$ such that $Y_i \partial G_i(\boldsymbol{Y})$ is bounded from above by $L_i Y_i^{li}$ for all prices $\boldsymbol{x} \geq 0$, then for any ϵ such that

$$\epsilon \leq \min_{t,i}\{\alpha_i^t |\ \alpha_i^t > 0\} \min_t \left\{ \frac{r_t}{(\boldsymbol{\alpha}^t)^T \boldsymbol{1}} \right\},$$

then if we choose $\lambda_0 \geq (\Delta^* - v^*)/\epsilon$, we can upper-bound $|v^{\boldsymbol{\lambda}} - v^*|$ as

$$|v^{\boldsymbol{\lambda}} - v^*| \leq \max\left\{ \max_i \left\{ \frac{a_i}{b_i} - \frac{1}{b_i l_i} \log \frac{\delta(\epsilon)}{L_i} \right\}, 0 \right\} \frac{m\epsilon}{\min_{t,i}\{\alpha_i^t |\ \alpha_i^t > 0\}},$$

where $\delta(\epsilon) = \min_t \left\{ \frac{r_t}{(\boldsymbol{\alpha}^t)^T \boldsymbol{1}} \right\} - \frac{\epsilon}{\min_{t,i}\{\alpha_i^t |\ \alpha_i^t > 0\}}$, and this upper bound converges to zero linearly when ϵ tends to zero.

The proof of Theorem 3 can be found in Appendix C. It is not difficult to validate that the assumption in Theorem 3–(iii) holds for all the well-known GEV models in the literatures. For examples, for the MNL, $Y_i \partial G_i(\mathbf{Y}) = Y_i$. For a nested logit mode specified by $G(\mathbf{Y}) = \sum_{n \in \mathcal{N}} \left(\sum_{i \in C_n} \sigma_{in} Y_i^{\mu_n} \right)^{1/\mu_n}$, where $\mathcal{N}$ is the set of nests, C_n is the corresponding nest and μ, μ_n are some parameters, we have $Y_i \partial G_i(\mathbf{Y}) = Y_i^{\mu_n} \left(\sum_{j \in C_n} \sigma_{jn} Y_j^{\mu_n} \right)^{1/\mu_n - 1}$. If $\mu_n > 1$ then $Y_i \partial G_i(\mathbf{Y}) \leq \sigma_{in}^{1/\mu_n - 1} Y_i$ and if $\mu_n < 1$ then $Y_i \partial G_i(Y) \leq L_n Y_i^{\mu_n}$, where L_n is an upper bound of $\left(\sum_{j \in C_n} \sigma_{jn} Y_j^{\mu_n} \right)^{1/\mu_n - 1}$ for all $\mathbf{x} \in \mathbb{R}_i^m$, which always exists. For a more general GEV model, we note that $\partial G_{ij}(\mathbf{Y}) \leq 0$ and $Y_j \geq 0$ for all $i, j \in \mathcal{V}, i \neq j$. As a result, we have $\partial G_i(\mathbf{Y}) \leq \partial G_i(\widetilde{\mathbf{Y}}^i)$, where $\widetilde{\mathbf{Y}}^i$ is a vector of size m with entries $\widetilde{Y}_i^i = Y_i$ and $\widetilde{Y}_j^i = 0$ for all $j \neq i$. Thus,$\partial G_i(\widetilde{\mathbf{Y}}^i)$ is a function of only Y_i. For a more complicated GEV model such as the network GEV model (Daly and Bierlaire 2006), we can easily upper-bound $Y_i \partial G_i(\widetilde{\mathbf{Y}}^i)$ by a function of form $L_i Y_i^{l_i}$, where $L_i, l_i > 0$.

In Theorem 3, the first result (i) indicates that the penalty term $\sum_{t=1}^T \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}} \mathbf{p} - r_t\}$ is monotonically decreasing as a function of $\boldsymbol{\lambda}$. As a result, this term will converges to a non-negative constant when $\boldsymbol{\lambda}$ increases. The second statement (ii) tells us explicitly that the penalty term will converge to zero when the parameters $\boldsymbol{\lambda}$ are large enough. It also provides an estimate for $\min_t\{\lambda_t\}$ to get arbitrarily small penalty costs. The third result (iii) provides an upper-bound for the gap between the optimal expected revenues given by the constrained pricing problem and the pricing problem with over-expected-sale penalties, and this upper bound converges to zero linearly when ϵ goes to zero. So in general, Problem 17 can be viewed as a generalized version of the constrained pricing problem in (28), in the sense that if we select the penalty parameters $\boldsymbol{\lambda}$ large enough, then we will get a solution that is similar to the one from the constrained problem, and if we set $\boldsymbol{\lambda} = 0$ then we come back to the unconstrained problem. Thus, the formulation in (18) provides a more flexible way to handle expected sale requirements.

Theorem 3 also allows us to answer the question whether an optimal solution to (17) has the constant-markup style. In general, we can show that a solution to (17) does not have a constant markup over products if the PSP are all homogeneous, or has a constant markup in each partition if the PSP are partition-wise homogeneous and the CPGF is partition-wise separable (Corollary 3, the proof is given in Appendix B). For this reason, the results presented in this section are not a generalized version of those shown in Sections 3 and 3.2 when the penalty parameters $\boldsymbol{\lambda}$ equals zero.

Corollary 3. *A solution to the unconstrained pricing problem with penalties (17) does not have a constant-markup style.*

Since we consider the pricing problem with over-expected-sale penalties, we do not face the issue of violating the expected sale constraints when the choice parameters vary in the uncertainty set.We consider the robust version of (17) under $(\mathbf{a},\mathbf{b})$ uncertainty

$$\max_{\mathbf{x}\in\mathbb{R}^m}\min_{(\mathbf{a},\mathbf{b})\in\mathcal{A}}\left\{\Phi(\mathbf{x},\mathbf{a},\mathbf{b})-\sum_{t=1}^{T}\lambda_t\max\{0,(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}-r_t\}\right\} \tag{19}$$

The adversarial problem of (19) is still difficult to solve as the objective function is not convex on $\mathcal{A}$. Surprisingly, in Theorem 4 below, we show that if under the robust optimal prices, an optimal solution to the adversary problem can be identified, then it allows us to convert the robust problem into a deterministic pricing problem with expected-sale-penalties. To obtain this result, we need to assume that the uncertainty set is rectangular and the reason is similar to the case of not-homogeneous PSP in Section 3.2. To avoid overly complicated proofs, we will first provide results for the MNL model, then we will show results for a general GEV model with additional assumptions in the next proposition. Theorem 4 shows results for the pricing problem with over-expected-penalties under the MNL model.

Theorem 4. (A tractable solution for MNL-based robust pricing with over-expected-sale penalties and rectangular uncertainty sets). *If the choice model is MNL and the uncertainty set $\mathcal{A}$ is rectangular, i.e., $\mathcal{A}=\{(\boldsymbol{a},\boldsymbol{b})|\ (\underline{\boldsymbol{a}},\underline{\boldsymbol{b}})\preceq(\boldsymbol{a},\boldsymbol{b})\preceq(\overline{\boldsymbol{a}},\overline{b})\}$, then the robust problem (19) is equivalent to*

$$v^*=\max_{\boldsymbol{x}}\left\{\Phi(\boldsymbol{x},\underline{\boldsymbol{a}},\overline{\boldsymbol{b}})-\sum_{t=1}^{T}\lambda_t\max\{0,(\boldsymbol{\alpha}^t)^T\boldsymbol{p}(\boldsymbol{x},\underline{\boldsymbol{a}},\overline{\boldsymbol{b}})-r_t\}\right\},$$

and the maximization problem can be formulated as a convex optimization one.

For notational convenience, let

$$\mathcal{L}(\mathbf{x},\mathbf{a},\mathbf{b})=\Phi(\mathbf{x},\mathbf{a},\mathbf{b})-\sum_{t=1}^{T}\lambda_t\max\{0,(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}-r_t\},\text{ and } f(\mathbf{x})=\max_{\mathbf{x}\in\mathbb{R}^m}\mathcal{L}(\mathbf{x},\mathbf{a},\mathbf{b}).$$

We will also use the set $\overline{\mathcal{A}}\{(\mathbf{a},\mathbf{b})|\ a_i\in\{\underline{a}_i,\overline{a_i}\},\ b_i\in\{\underline{b}_i,\overline{b_i}\}\}$, which is already defined in the proof of Lemma 8. In general, the robust problem in (19) is challenging to handle because of the term

$\sum_{t=1}^{T} \lambda_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p} - r_t\}$, which makes the objective function no-longer differentiable in $\mathbf{x}$. However, the good thing here is that if we consider the subset $\mathcal{T} \in [T]$ such that the constraints are violated only in $\mathcal{T}$, we can write the objective function as

$$\begin{aligned}\mathcal{L}(\mathbf{x},\mathbf{a},\mathbf{b}) &= \Phi(\mathbf{x},\mathbf{a},\mathbf{b}) - \sum_{t\in\mathcal{T}} \lambda_t (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p} + \sum_{t\in\mathcal{T}} \lambda_t r_t \\ &= \sum_{i\in\mathcal{V}} (x_i - \sum_{t\in\mathcal{T}} \lambda_t \boldsymbol{\alpha}_i^t - c_i) p_i(\mathbf{x},\mathbf{a},\mathbf{b}) + \sum_{t\in\mathcal{T}} \lambda_t r_t \end{aligned} \tag{20}$$

and note that $\sum_{i\in\mathcal{V}} (x_i - \sum_{t\in\mathcal{T}} \lambda_t \boldsymbol{\alpha}_i^t - c_i) p_i(\mathbf{x},\mathbf{a},\mathbf{b})$ is also the expected revenue with shifted item costs $\sum_{t\in\mathcal{T}} \lambda_t \boldsymbol{\alpha}_i^t + c_i$, $\forall i \in \mathcal{V}$. We will leverage this observation to prove the theorem.

In Lemma 9 below, we show that given any prices $\mathbf{x} \in \mathbb{R}^m$, one of the solutions in the set $\overline{\mathcal{A}}$ is optimal to the adversary's problem. This result is similar to the unconstrained case with partition-wise PSP considered in Section 3.2.

Lemma 9. *Given any $\boldsymbol{x} \in \mathbb{R}^m$, there is at least one solution $(\boldsymbol{a},\boldsymbol{b}) \in \overline{\mathcal{A}}$ that is optimal to the corresponding adversary's problem of (19).*

Proof: Given $\mathbf{x} \in \mathbb{R}^m$, let $(\mathbf{a}^{\mathbf{x}}, \mathbf{b}^{\mathbf{x}})$ be optimal to the adversary problem. We also denote by $\mathcal{T}$ a subset of $[T]$ such that $(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}(\mathbf{x},\mathbf{a}^{\mathbf{x}},\mathbf{b}^{\mathbf{x}}) \geq r_t$ for all $t \in \mathcal{T}$ and $(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}(\mathbf{x},\mathbf{a}^{\mathbf{x}},\mathbf{b}^{\mathbf{x}}) < r_t$ if $t \notin \mathcal{T}$. The robust optimal value at $\mathbf{x}$ becomes

$$\begin{aligned} f(\mathbf{x}) &= \sum_i (x_i - c_i)\, p_i(\mathbf{x},\mathbf{a}^{\mathbf{x}},\mathbf{b}^{\mathbf{x}}) - \sum_{t\in\mathcal{T}} \lambda_t (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}(\mathbf{x},\mathbf{a}^{\mathbf{x}},\mathbf{b}^{\mathbf{x}}) + \sum_{t\in\mathcal{T}} \lambda_t r_t \\ &= \frac{\sum_i \left(x_i - c_i - \sum_{t\in\mathcal{T}} \lambda_t \alpha_i^t\right) Y_i^{\mathbf{x}}}{1 + \sum_i Y_i^{\mathbf{x}}} + \sum_{t\in\mathcal{T}} \lambda_t r_t, \end{aligned}$$

where $Y_i^{\mathbf{x}} = \exp(a_i^{\mathbf{x}} - b_i^{\mathbf{x}} x_i)$. For notational brevity, let

$$\begin{aligned} &\rho^{\mathbf{x}} = \frac{\sum_i \left(x_i - c_i - \sum_{t\in\mathcal{T}} \lambda_t \alpha_i^t\right) Y_i^{\mathbf{x}}}{1 + \sum_i Y_i^{\mathbf{x}}} \\ &\mathcal{I}_1 = \{i \in \mathcal{V} |\ \rho^{\mathbf{x}} < x_i - c_i - \sum_{t\in\mathcal{T}} \lambda_t \alpha_i^t\} \\ &\mathcal{I}_2 = \{i \in \mathcal{V} |\ \rho^{\mathbf{x}} > x_i - c_i - \sum_{t\in\mathcal{T}} \lambda_t \alpha_i^t\} \\ &\mathcal{A}^{\mathbf{x}} = \{(\mathbf{a},\mathbf{b}) \in \mathcal{A} |\ (a_i, b_i) = (\underline{a}_i, \overline{b}_i) \text{ if } i \in \mathcal{I}_1,\ (a_i, b_i) = (\overline{a}_i, \underline{b}_i) \text{ if } i \in \mathcal{I}_2\} \end{aligned}$$

From Lemma 5, if $(\mathbf{a}^{\mathbf{x}}, \mathbf{b}^{\mathbf{x}}) \notin \mathcal{A}^{\mathbf{x}}$ then for any $(\mathbf{a},\mathbf{b}) \in \mathcal{A}^{\mathbf{x}}$ we have

$$\begin{aligned} f(\mathbf{x}) &> \sum_i (x_i - c_i)\, p_i(\mathbf{x},\mathbf{a},\mathbf{b}) - \sum_{t\in\mathcal{T}} \lambda_t (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}(\mathbf{x},\mathbf{a},\mathbf{b}) + \sum_{t\in\mathcal{T}} \lambda_t r_t \\ &\geq \Phi(\mathbf{x},\mathbf{a},\mathbf{b}) - \sum_{t=1}^{T} \lambda_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}(\mathbf{x},\mathbf{a},\mathbf{b}) - r_t\}, \end{aligned}$$

which is contradictory to the assumption that $(\mathbf{a}^{\mathbf{x}}, \mathbf{b}^{\mathbf{x}})$ is optimal to the adversary's problem. So we have $(\mathbf{a}^{\mathbf{x}}, \mathbf{b}^{\mathbf{x}}) \in \mathcal{A}^{\mathbf{x}}$. On the other hand, Lemma 5 tells us that if we take any point $(\mathbf{a}, \mathbf{b}) \in \mathcal{A}^{\mathbf{x}}$ such that $(\mathbf{a}, \mathbf{b}) \in \overline{\mathcal{A}}$ we also have

$$\begin{aligned} f(\mathbf{x}) &= \sum_i \left(x_i - c_i - \sum_{t\in\mathcal{T}} \lambda_t \alpha_i^t \right) p_i(\mathbf{x}, \mathbf{a}, \mathbf{b}) + \sum_{t\in\mathcal{T}} \lambda_t r_t \\ &\geq \sum_i (x_i - c_i)\, p_i(\mathbf{x}, \mathbf{a}, \mathbf{b}) - \sum_{t=1}^{T} \lambda_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}} \mathbf{p}(\mathbf{x}, \mathbf{a}, \mathbf{b}) - r_t\} = \mathcal{L}(\mathbf{x}, \mathbf{a}, \mathbf{b}). \end{aligned}$$

Since $f(\mathbf{x}) = \min_{(\mathbf{a},\mathbf{b})\in\mathcal{A}} \mathcal{L}(\mathbf{x}, \mathbf{a}, \mathbf{b})$, we have $f(\mathbf{x}) = \mathcal{L}(\mathbf{x}, \mathbf{a}, \mathbf{b})$, meaning that $(\mathbf{a}, \mathbf{b})$ is also optimal to the adversary's problem under prices $\mathbf{x}$. This completes the proof. □

The lemma above tells us that a solution in which each coordinate is at a bound of the uncertainty set is optimal to the adversary's problem, given any prices. This allows us to reduce the searching of the space for the adversary's problem from an infinite to a finite set, which is helpful. The number of elements in $\overline{\mathcal{A}}$ is still huge (2^{2m}). The next lemma characterizes an important property of the robust optimal prices, which allows us to further reduce the adversary's searching space into a singleton set, which greatly simplifies the robust problem. Let $\mathbf{x}^*$ be a robust optimal solution to the robust problem and $(\mathbf{a}^*, \mathbf{b}^*)$ be an optimal solution to the adversary problem such that $(\mathbf{a}^*, \mathbf{b}^*) \in \overline{\mathcal{A}}$, i.e., $(\mathbf{a}^*, \mathbf{b}^*)$ are at the bounds of the the uncertainty set. We also denote by $\mathcal{T}^*$ a subset of $[T]$ such that $(\boldsymbol{\alpha}^t)^{\mathrm{T}} \mathbf{p}(\mathbf{x}^*, \mathbf{a}^*, \mathbf{b}^*) \geq r_t$ for all $t \in \mathcal{T}^*$ and $(\boldsymbol{\alpha}^t)^{\mathrm{T}} \mathbf{p}(\mathbf{x}^*, \mathbf{a}^*, \mathbf{b}^*) < r_t$ if $t \notin \mathcal{T}^*$, and let

$$\rho^* = \frac{\sum_i \left(x_i^* - c_i - \sum_{t\in\mathcal{T}^*} \lambda_t \alpha_i^t \right) Y_i^*}{1 + \sum_i Y_i^*}, \text{ where } Y_i^* = \exp(a_i^* - b_i^* x_i^*).$$

Lemma 10. *$\rho^* < x_i^* - c_i - \sum_{t\in\mathcal{T}^*} \lambda_t \alpha_i^t$, for all $i \in \mathcal{V}$.*

Proof: By contradiction, assume that there exits $k \in \mathcal{V}$ such that $\rho^* \geq x_k^* - c_k - \sum_{t\in\mathcal{T}^*} \lambda_t \alpha_k^t$. Let

$$\delta = \min_{(\mathbf{a},\mathbf{b})\in\overline{\mathcal{A}}} \left\{ \mathcal{L}(\mathbf{x}^*, \mathbf{a}, \mathbf{b}) - \mathcal{L}(\mathbf{x}^*, \mathbf{a}^*, \mathbf{b}^*) \middle|\ \mathcal{L}(\mathbf{x}^*, \mathbf{a}, \mathbf{b}) > \mathcal{L}(\mathbf{x}^*, \mathbf{a}^*, \mathbf{b}^*) \right\}, \tag{21}$$

with a note that we set $\delta = +\infty$ if the corresponding searching set is empty. Using Lemma 9 we can choose $(\mathbf{a}^*, \mathbf{b}^*)$ such that $(a_k^*, b_k^*) = (\overline{a}_k, \underline{b}_k)$. So, if we select $\epsilon > 0$ such that $\epsilon < (\overline{a}_k - \underline{a}_k)/\underline{b}_k$ then $a_k^* - \epsilon \underline{b}_k \geq \underline{a}_k$, meaning that $(\mathbf{a}^* - \epsilon b_k^* \mathbf{e}^k, \mathbf{b}^*) \in \mathcal{A}$ and we have the following claim

$$\begin{aligned} \mathcal{L}(\mathbf{x}^*, \mathbf{a}^*, \mathbf{b}^*) &\leq \mathcal{L}(\mathbf{x}^*, \mathbf{a}^* - \epsilon b_k^* \mathbf{e}^k, \mathbf{b}^*) \\ &= (x_i^* - c_i)\, p_i(\mathbf{x}^*, \mathbf{a}^* - \epsilon b_k^* \mathbf{e}^k, \mathbf{b}^*) - \sum_{t=1}^{T} \lambda_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}} \mathbf{p}(\mathbf{x}^*, \mathbf{a}^* - \epsilon b_k^* \mathbf{e}^k, \mathbf{b}^*) - r_t\} \\ &= (x_i^* - c_i)\, p_i(\mathbf{x}^* + \epsilon \mathbf{e}^k, \mathbf{a}^*, \mathbf{b}^*) - \sum_{t=1}^{T} \lambda_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}} \mathbf{p}(\mathbf{x}^* + \epsilon \mathbf{e}^k, \mathbf{a}^*, \mathbf{b}^*) - r_t\} \\ &< \mathcal{L}(\mathbf{x}^* + \epsilon \mathbf{e}^k, \mathbf{a}^*, \mathbf{b}^*). \end{aligned}$$

Moreover, combining with the fact that $\mathcal{L}(\mathbf{x},\mathbf{a}^*,\mathbf{b}^*)$ and $f(\mathbf{x})$ are continuous in $\mathbf{x}$, we always can select $\epsilon > 0$ small enough such that

$$\mathcal{L}(\mathbf{x},\mathbf{a}^*,\mathbf{b}^*) < \mathcal{L}(\mathbf{x}+\epsilon\mathbf{e}^k,\mathbf{a}^*,\mathbf{b}^*) \tag{22}$$

$$|f(\mathbf{x}^*) - f(\mathbf{x}^*+\epsilon\mathbf{e}^k)| < \delta/2 \tag{23}$$

$$\left|\mathcal{L}(\mathbf{x}^*+\epsilon\mathbf{e}^k,\mathbf{a}^\epsilon,\mathbf{b}^\epsilon) - \mathcal{L}(\mathbf{x}^*,\mathbf{a}^\epsilon,\mathbf{b}^\epsilon)\right| < \delta/2, \tag{24}$$

where $(\mathbf{a}^\epsilon,\mathbf{b}^\epsilon)$ is an optimal solution to the adversary's problem under prices $\mathbf{x}+\epsilon\mathbf{e}^k$. Using (23) and (24) we have

$$|f(\mathbf{x}^*) - \mathcal{L}(\mathbf{x}^*,\mathbf{a}^\epsilon,\mathbf{b}^\epsilon)| \le |f(\mathbf{x}^*) - f(\mathbf{x}^*+\epsilon\mathbf{e}^k)| + |f(\mathbf{x}^*+\epsilon\mathbf{e}^k) - \mathcal{L}(\mathbf{x}^*,\mathbf{a}^\epsilon,\mathbf{b}^\epsilon)| < \delta.$$

So, according to the definition of δ in (21), we have $f(\mathbf{x}^*) = \mathcal{L}(\mathbf{x}^*,\mathbf{a}^\epsilon,\mathbf{b}^\epsilon)$, which means that $(\mathbf{a}^\epsilon,\mathbf{b}^\epsilon)$ is also an optimal solution to the adversary's problem under prices $\mathbf{x}^*$. So, we always can choose an optimal solution $(\mathbf{a}^*,\mathbf{b}^*)$ in the set of optimal solutions of the adversary's problem under $\mathbf{x}^*$ such that $(\mathbf{a}^*,\mathbf{b}^*) = (\mathbf{a}^\epsilon,\mathbf{b}^\epsilon)$. Together with (22), we have

$$f(\mathbf{x}^*) = \mathcal{L}(\mathbf{x}^*,\mathbf{a}^*,\mathbf{b}^*) < \mathcal{L}(\mathbf{x}^*+\epsilon\mathbf{e}^k,\mathbf{a}^*,\mathbf{b}^*) = f(\mathbf{x}^*+\epsilon\mathbf{e}^k),$$

which is contradictory to our initial assumption that $\mathbf{x}^*$ is a robust optimal solution. So our contradiction hypothesis is untrue and this completes the proof. □

We are now ready for the proof of Theorem 4.

Proof of Theorem 4. From Lemma 9 and 10, we have that if $\mathbf{x}^*$ is a robust optimal solution, then $(\underline{\mathbf{a}},\overline{\mathbf{b}})$ is the unique optimal solution to the adversary's problem under robust optimal prices $\mathbf{x}^*$. We need to prove that $\mathbf{x}^*$ is also optimal to the maximization problem $\max_{\mathbf{x}} \mathcal{L}(\mathbf{x},\underline{\mathbf{a}},\overline{\mathbf{b}})$. In this case, the function $\mathcal{L}(\mathbf{x},\underline{\mathbf{a}},\overline{\mathbf{b}})$ is not differentiable in $\mathbf{x}$, so we cannot use the techniques in the proof of Lemma 8. Fortunately, if we consider the objective function $\mathcal{L}(\mathbf{x},\underline{\mathbf{a}},\overline{\mathbf{b}})$ as a function of the purchase probabilities $\mathbf{p}$, then it is strictly concave in $\mathbf{p}$. To facilitate this point, lets us define

$$\mathcal{F}(\mathbf{p},\mathbf{a},\mathbf{b}) = \mathcal{L}(\mathbf{x},\mathbf{a},\mathbf{b}) = (\mathbf{x}(\mathbf{p}|\mathbf{a},\mathbf{b}) - \mathbf{c})^{\mathrm{T}}\,\mathbf{p} - \sum_{t=1}^{T} \lambda_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p} - r_t\}$$

We know that the first term $(\mathbf{x}(\mathbf{p}|\mathbf{a},\mathbf{b}) - \mathbf{c})^{\mathrm{T}}\,\mathbf{p}$ is strictly concave in $\mathbf{p}$ (Zhang et al. 2018) and it is not difficult to show that $-\sum_{t=1}^{T} \lambda_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p} - r_t\}$ is concave in $\mathbf{p}$. As a result, $\mathcal{F}(\mathbf{p},\mathbf{a},\mathbf{b})$ is strictly concave in $\mathbf{p}$. Now, let $\mathbf{p}^*$ be the purchase probabilities given by prices $\mathbf{x}^*$ and choice parameters $(\underline{\mathbf{a}},\overline{\mathbf{b}})$. We will prove that $\mathbf{p}^* = \text{argmax}_{\mathbf{p}} \mathcal{K}(\mathbf{p},\underline{\mathbf{a}},\overline{\mathbf{b}})$. By contradiction, assume that $\widetilde{\mathbf{p}} = \text{argmax}_{\mathbf{p}} \mathcal{K}(\mathbf{p},\underline{\mathbf{a}},\overline{\mathbf{b}})$ and $\mathcal{K}(\widetilde{\mathbf{p}},\underline{\mathbf{a}},\overline{\mathbf{b}}) > \mathcal{K}(\mathbf{p}^*,\underline{\mathbf{a}},\overline{\mathbf{b}})$. Since $\mathcal{K}(\mathbf{p},\underline{\mathbf{a}},\overline{\mathbf{b}})$ is strictly concave in $\mathbf{p}$, we have, for any $t \in (0,1)$,

$$t\mathcal{K}(\widetilde{\mathbf{p}},\underline{\mathbf{a}},\overline{\mathbf{b}}) + (1-t)\mathcal{K}(\mathbf{p}^*,\underline{\mathbf{a}},\overline{\mathbf{b}}) < \mathcal{K}(t\widetilde{\mathbf{p}} + (1-t)\mathbf{p}^*,\underline{\mathbf{a}},\overline{\mathbf{b}})$$

Since $\mathcal{K}(\widetilde{\mathbf{p}},\underline{\mathbf{a}},\overline{\mathbf{b}}) \geq \mathcal{K}(t\widetilde{\mathbf{p}}+(1-t)\mathbf{p}^*,\underline{\mathbf{a}},\overline{\mathbf{b}})$, we have $\mathcal{K}(\mathbf{p}^*,\underline{\mathbf{a}},\overline{\mathbf{b}}) < \mathcal{K}(t\widetilde{\mathbf{p}}+(1-t)\mathbf{p}^*,\underline{\mathbf{a}},\overline{\mathbf{b}})$ for all $t \in (0,1)$. This also mean that for any $\epsilon > 0$, we always can find a point $\mathbf{p} > 0$, $\sum_i p_i < 1$, such that $||\mathbf{p}^* - \mathbf{p}|| \leq \epsilon$ and $\mathcal{K}(\mathbf{p}^*,\underline{\mathbf{a}},\overline{\mathbf{b}}) < \mathcal{K}(\mathbf{p},\underline{\mathbf{a}},\overline{\mathbf{b}})$. Since $\mathbf{p}(\mathbf{x},\underline{\mathbf{a}},\overline{\mathbf{b}})$ is continuous in $\mathbf{x}$, this also means that given any $\epsilon > 0$, there always exists $\mathbf{x} \in \mathbb{R}^m$ such that $||\mathbf{x}-\mathbf{x}^*|| \leq \epsilon$ and $\mathcal{L}(\mathbf{x}^*,\underline{\mathbf{a}},\overline{\mathbf{b}}) < \mathcal{L}(\mathbf{x},\underline{\mathbf{a}},\overline{\mathbf{b}})$.

Now, similarly to the proof of Lemma 8, let

$$\tau = \min_{(\mathbf{a},\mathbf{b})\in\overline{\mathcal{A}}} \left\{ \mathcal{L}(\mathbf{x}^*,\mathbf{a},\mathbf{b}) - \mathcal{L}(\mathbf{x}^*,\underline{\mathbf{a}},\overline{\mathbf{b}}) \middle|\ \mathcal{L}(\mathbf{x}^*,\mathbf{a},\mathbf{b}) > \mathcal{L}(\mathbf{x}^*,\underline{\mathbf{a}},\overline{\mathbf{b}}) \right\} \tag{25}$$

Since $f(\mathbf{x})$ and $\mathcal{L}(\mathbf{x},\mathbf{a},\mathbf{b})$ are continuous in $\mathbf{z}$, there is an $\epsilon > 0$ such that, for all $\mathbf{x}\in\mathbb{R}^m$, $||\mathbf{x}^*-\mathbf{x}|| \leq \epsilon$

$$\begin{cases} |f(\mathbf{x}^*) - f(\mathbf{x})| < \tau/2 \\ |\mathcal{L}(\mathbf{x}^*,\mathbf{a}^{\mathbf{x}},\mathbf{b}^{\mathbf{x}}) - \mathcal{L}(\mathbf{x},\mathbf{a}^{\mathbf{x}},\mathbf{b}^{\mathbf{x}})| < \tau/2, \end{cases} \tag{26}$$

where $(\mathbf{a}^{\mathbf{x}},\mathbf{b}^{\mathbf{x}})$ is an optimal solution to the adversary's problem (19) under prices $\mathbf{x}$. As a result, for all $\mathbf{x}$ such that $||\mathbf{x}^*-\mathbf{x}|| \leq \epsilon$

$$|\mathcal{L}(\mathbf{x}^*,\underline{\mathbf{a}},\overline{\mathbf{b}}) - \mathcal{L}(\mathbf{x}^*,\mathbf{a}^{\mathbf{x}},\mathbf{b}^{\mathbf{x}})| < \tau.$$

Combine this with (25) we have, for all $\mathbf{x}$ such that $||\mathbf{x}^*-\mathbf{x}|| \leq \epsilon$, $(\mathbf{a}^{\mathbf{x}},\mathbf{b}^{\mathbf{x}})$ is optimal to the adversary's problem under prices $\mathbf{x}^*$. Since $(\underline{\mathbf{a}},\overline{\mathbf{b}})$ is the unique solution to the adversary's problem under prices $\mathbf{x}^*$, we have $(\mathbf{a}^{\mathbf{x}},\mathbf{b}^{\mathbf{x}}) = (\underline{\mathbf{a}},\overline{\mathbf{b}})$. Moreover, we have shown that given any $\epsilon > 0$, there exists $\overline{\mathbf{x}}$ such that $||\overline{\mathbf{x}}-\mathbf{x}^*|| \leq \epsilon$ and $\mathcal{L}(\overline{\mathbf{x}},\underline{\mathbf{a}},\overline{\mathbf{b}}) > \mathcal{L}(\mathbf{x}^*,\underline{\mathbf{a}},\overline{\mathbf{b}})$. If we choose ϵ small enough, $(\underline{\mathbf{a}},\overline{\mathbf{b}})$ is also optimal to the adversary's problem under $\overline{\mathbf{x}}$, which leads to $f(\overline{\mathbf{x}}) > f(\mathbf{x}^*)$. This is contradictory to the fact that $\mathbf{x}^*$ is a robust optimal solution. So, our contradiction hypothesis that $\mathbf{p}^*$ is not optimal to $\max_{\mathbf{p}} \mathcal{K}(\mathbf{p},\underline{\mathbf{a}},\overline{\mathbf{b}})$ is untrue, meaning that $\mathbf{p}^*$ is optimal to $\max_{\mathbf{p}} \mathcal{K}(\mathbf{p},\underline{\mathbf{a}},\overline{\mathbf{b}})$, or equivalently, $\mathbf{x}^*$ is optimal to $\max_{\mathbf{x}} \mathcal{L}(\mathbf{x},\underline{\mathbf{a}},\overline{\mathbf{b}})$. We obtain the desired results. □

It is important to note that $(\underline{\mathbf{a}},\overline{\mathbf{b}})$ is not necessary an adversary's optimal solution under any price vector $\mathbf{x}$. To illustrate this observation, we simply select $\boldsymbol{\lambda} = 0$ and a price vector $\mathbf{x}$ in such a way that there is $i \in \mathcal{V}$ such that $x_i = c_i$. So, we have $f(x) > x_i - c_i$. Lemma 5 tells us that if $(\mathbf{a}^{\mathbf{x}},\mathbf{b}^{\mathbf{x}})$ is optimal to the adversary's problem under prices $\mathbf{x}$, then $(a_i,b_i) = (\overline{a}_i,\underline{b_i})$, meaning that $(\underline{\mathbf{a}},\overline{\mathbf{b}})$ is not optimal to the adversary's problem. So, basically, to solve the adversary's problem under any price solution, one might need to go through all possible solutions in $\overline{\mathcal{A}}$ or use a nonlinear optimization solver, which is not computationally tractable.

An interesting question here is how the robust optimal value and optimal solution change when the penalty parameters $\boldsymbol{\lambda}$ increase. To answer this, let $\varphi^{\boldsymbol{\lambda}}$ be the optimal value of the robust problem in (19) under penalty parameters λ, $\overline{\varphi}$ be the optimal value of the constrained problem with parameters $(\underline{\mathbf{a}},\overline{\mathbf{b}})$, $\mathbf{x}^{\boldsymbol{\lambda}}$ is an robust solution to (19) and $\mathbf{p}^{\boldsymbol{\lambda}}$ is the purchase probabilities given by the robust solution $\mathbf{x}^{\boldsymbol{\lambda}}$ in the worst-case. The following corollary is a direct result from Theorem 3 and 4, noting that in the case of the MNL model, L_i, l_i (defined in Theorem 3) are equal to 1 for all $i \in \mathcal{V}$.

Corollary 4. (Convergence of the robust optimal values when the penalty parameters $\boldsymbol{\lambda}$ increase). *Assume that choice model is MNL and the uncertainty set is rectangular ($\mathcal{A} = \{(\boldsymbol{a}, \boldsymbol{b})|\ (\underline{\boldsymbol{a}}, \underline{\boldsymbol{b}}) \preceq (\boldsymbol{a}, \boldsymbol{b}) \preceq (\overline{\boldsymbol{a}}, \overline{b})\}$). Given any $\epsilon > 0$, if we select $\boldsymbol{\lambda}$ such that $\min_t \lambda_t \geq (\Delta^* - \overline{\varphi})/\epsilon$ then $\sum_t \max\{0, (\boldsymbol{\alpha}^t)^T \boldsymbol{p}^{\boldsymbol{\lambda}} - r_t\} \leq \epsilon$, where $\Delta^* = \max_{\boldsymbol{x}} \Phi(\boldsymbol{x}, \underline{\boldsymbol{a}}, \overline{\boldsymbol{b}})$. Moreover, if we select ϵ such that*

$$\epsilon \leq \min_{t,i}\{\alpha_i^t|\ \alpha_i^t > 0\} \min_t \left\{ \frac{r_t}{(\boldsymbol{\alpha}^t)^T \boldsymbol{1}} \right\},$$

then $\varphi^{\boldsymbol{\lambda}} - \overline{\varphi}$ can be bounded as

$$0 \leq \varphi^{\boldsymbol{\lambda}} - \overline{\varphi} \leq \max\left\{ \max_i \left\{ \frac{\underline{a}_i}{\overline{b}_i} - \frac{1}{\overline{b}_i} \log \delta(\epsilon) \right\}, 0 \right\} \frac{m\epsilon}{\min_{t,i}\{\alpha_i^t|\ \alpha_i^t > 0\}},$$

where $\delta(\epsilon)$ is defined as in Theorem 3.

We can also extend the results of Theorem 4 for robust pricing problems under GEV models, under the assumption that the CPGF is partition-wise separable and we require that the expected sale parameters are the same in each partition. Since the robust prices in this context would not have the constant-markup style (Corollary 3) , i.e., robust optimal prices would not have a constant markup in each partition. The techniques used in Section 3.2 cannot be used and we are unable to identify robust solutions in this context. However, it we only seek constant-markup solutions, i.e., $\mathbf{x} \in X := \{\mathbf{x} \in \mathbf{R}_+^m|\ x_i - c_i = x_j - c_j,\ \forall i, j \in \mathcal{V}_n, n \in [N]\}$, then the robust problem can be converted into a convex optimization problem (Proposition 6). We refer the reader to Section 3.2 for the definitions of $G^n(\cdot)$, $\mathbf{Y}^n$, $\forall n \in [N]$.

Proposition 6. (Robust solutions for the robust pricing problem under GEV models with over-expected-sale penalties). *Assume that the uncertainty set is rectangular ($\mathcal{A} = \{(\boldsymbol{a}, \boldsymbol{b})|\ (\underline{\boldsymbol{a}}, \underline{\boldsymbol{b}}) \preceq (\boldsymbol{a}, \boldsymbol{b}) \preceq (\overline{\boldsymbol{a}}, \overline{b})\}$) and the CPGF $G(\boldsymbol{Y})$ is partition-wise separable as defined in Section 3.2 and $\alpha_i^t = \alpha_j^t$ for all $i, j \in \mathcal{V}_n$, $n \in [N]$. If $\boldsymbol{z}^*$ is optimal to the problem*

$$\max_{\boldsymbol{z} \in \mathbb{R}^N} \left\{ \mathcal{H}(\boldsymbol{z}, \underline{\boldsymbol{a}}, \overline{\boldsymbol{b}})) = \frac{\sum_{n\in[N]} z_n G^n(\boldsymbol{Y}^n | z_n, \underline{\boldsymbol{a}}, \overline{\boldsymbol{b}})}{1 + \sum_{n\in[N]} G^n(\boldsymbol{Y}^n | z_n, \underline{\boldsymbol{a}}, \overline{\boldsymbol{b}})} - \sum_{t=1}^T \lambda_t \max\{0, (\boldsymbol{d}^t)^T \boldsymbol{p}^G - r_t\} \right\}, \tag{27}$$

where $\boldsymbol{d}^t$ is a vector of size N with entries $d_n^t = \alpha_i^t$ for all $i \in \mathcal{V}_n$, $n \in [N]$, $t \in [T]$ and $\boldsymbol{p}^G$ is of size N with entries $p_n^G = G^n(\boldsymbol{Y}^n | z_n, \underline{\boldsymbol{a}}, \overline{\boldsymbol{b}}) \Big/ \left(1 + \sum_{j\in[N]} G^j(\boldsymbol{Y}^j | z_j, \underline{\boldsymbol{a}}, \overline{\boldsymbol{b}})\right)$, then $\boldsymbol{x}^* \in \mathbb{R}^m$ such that $x_i^* = c_i + z_n^*$, $\forall n \in [n], i \in \mathcal{V}_n$ is optimal to the robust problem $\max_{\boldsymbol{x}\in X} \min_{(\boldsymbol{a},\boldsymbol{b})\in\mathcal{A}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{b})$. Moreover, (27) can be reformulated as a convex optimization problem.

We provide the proof of Proposition 6 in Appendix D. This proof is done in a similar way as in the proof of Theorem 4, where we try to show that under the robust optimal prices, $(\underline{\mathbf{a}}, \overline{\mathbf{b}})$ is a unique solution to the adversary's problem. To show that the equivalent optimization problem (40) is convex, we convert the problem into an MNL-based problem with heterogeneous PSP, and results from previous studies (Zhang et al. 2018) can be used. The limitation of the Proposition 6 is that it only returns best solutions among those that have a constant markup in each partition, and all the expected sale penalties in each partition need to be the same. Relaxing these assumption would make the robust problem challenging to handle (see the discussion after Corollary 3). Moreover, we believe that Proposition 6 is still useful in some contexts, where the firm only wants to make pricing decisions for each group of products and only impose expected sale requirements for the whole groups instead of each single product in the groups.

5. Numerical experiments

In this section we provide experimental results to show how the robust models considered in Sections 3 (i.e, robust unconstrained pricing with homogeneous or partition-wise homogeneous PSP) and 4 (i.e., robust pricing with over-expected-sale penalties) protect us from choice parameter uncertainties.

When the choice parameters are not given exactly, one can consider a deterministic approach where the mean values of the choice parameters are employed. In this context, we know that the pricing problem is computationally tractable (Zhang et al. 2018). Alternatively, one may look at different possibilities of the choice parameters and define a mixed version where the market is divided into a finite number of market segments and each segment is governed by a scenario of the choice parameters. However, one can show that the expected revenue in this context is no longer unimodal and the *constant-markup* property identified for the GEV pricing problem no-longer holds, even if there are only two market segments (Li et al. 2018). As a result, this mixed version is not computationally tractable.

Another baseline approach that can be used to account for parameters uncertainty is to sample some choice parameters from the uncertainty set and use simulation to select a solution that provides best protection from worst-case scenarios. More precisely, let assume that the firm needs to make a pricing decision while being aware that the choice parameters may vary in an uncertainty set. In this context, the firm can sample some points from the uncertainty set and compute the corresponding optimal prices for each selection, using the deterministic approach from Zhang et al. (2018). Then, for each price vector, the firm can sample a *sufficiently large* number of vector of choice parameters from the uncertainty set, in order to evaluate how each price vector obtained performs when the choice parameters vary in the uncertainty set. This can be done by simply selecting the solution that gives the best worst-case profit among the samples. This approach may be computationally tractable with a reasonable number of samples, but would be much more computationally expensive than the robust and deterministic approaches. We refer to this as the sampling-based pricing approach. One can show that solutions given by the sampling-based approach will converge to those from the robust counterpart when the sample sizes grow to infinity.

In these experiments, we will compare our robust models, which are computationally tractable, against the sampling-based approach and the deterministic counterparts with mean-value choice parameters. In the sampling-based approach, we sample points uniformly from the uncertainty set since we do not make any assumption about the distribution of the choice parameters. One can argue that the uniform distribution may not be the best choice in the case that the firm believes that it has some ideas (perhaps via estimation) about the distribution of the choice parameters. Nevertheless, estimating such a distribution is not easy in practice. A common approach in choice modeling is to assume that the parameters follow some distributions (e.g. normal distribution) with unknown coefficients and try to estimate these coefficients by maximum likelihood estimation (McFadden and Train 2000). This approach, even though popular, does not guarantee that the distribution obtained is the *true* distribution of the choice parameters, assuming that there exists a true distribution. As such, the distribution of the choice parameters is typically only known ambiguously. Distributionally robust optimization is a *robust* approach that is explicitly designed to handle this ambiguity (Shapiro 2018), which we keep for future research.

A crucial decision for our robust models and the sampling-based approach is to define the uncertainty set. A common approach in the robust optimization literature is to describe an uncertainty set as an ellipsoidal set

$$\mathcal{A} = \left\{ (\mathbf{a}, \mathbf{b}) \middle| \; ||(\mathbf{a}, \mathbf{b}) - (\mathbf{a}_0, \mathbf{b}_0)|| \leq \epsilon \right\},$$

where $||\cdot||$ is an appropriate norm, $(\mathbf{a}_0, \mathbf{b}_0)$ is the center of the ellipsoid and can be interpreted as the most likely parameter vector associated with the underlying deterministic counterparts, and $\epsilon \in \mathbb{R}_+$ reflects an "uncertainty level" of the uncertainty set. Larger ϵ values provide larger uncertainty sets, corresponding to more conservative models that may help protect well against worst-case scenarios, but may lead to low average performance. On the other hand, smaller ϵ values provide smaller uncertainty sets and would lead to less conservative robust solutions, which may perform well in terms of average performance but would be less beneficial in protecting bad scenarios of the choice parameters. We will show these in detail in the following sections.

An uncertainty set can be constructed by considering different possibilities of the choice parameter estimates. For example, there may be a setting where the firm operates in a market with Q heterogeneous customer types, each corresponding to a vector of choice parameter $(\mathbf{a}_q, \mathbf{b}_q)$, $q \in Q$, and the firm is uncertain about the proportion of each customer type. This suggests an idea to construct an ellipsoidal uncertainty set as $\mathcal{A}^\epsilon = \left\{ (\mathbf{a}, \mathbf{b}) \middle| \; ||(\mathbf{a}, \mathbf{b}) - (\mathbf{a}_0, \mathbf{b}_0)|| \leq \epsilon \right\}$ where $(\mathbf{a}_0, \mathbf{b}_0) = 1/Q \sum_{q \in [Q]} (\mathbf{a}_q, \mathbf{b}_q)$ and ϵ is chosen as the minimum possible value such that $(\mathbf{a}_q, \mathbf{b}_q) \in \mathcal{A}^\epsilon$ for all

$q \in [Q]$. In our context, we choose the L-1 norm for the case of the unconstrained robust problem with homogeneous PSP, as it allows to linearize the constraints of the adversary's problem. In the other cases (unconstrained pricing with partition-wise homogeneous PSP or pricing with over-expected-sale penalties), since our results require rectangular uncertainty sets, we choose the L-infinity norm.

5.1. Unconstrained Pricing with Homogeneous PSP

In this section we work with the robust unconstrained pricing problems with homogeneous PSP considered in Section 3.1. The goal here is to compare the robust (RO) approach against the standard deterministic (DET) and the sampling-based (SA) approaches under two popular GEV models in the literature, i.e., the MNL and nested logit models, when the PSP are the same over all the products. In this context, we know that the optimal prices for all the cases (RO, DET, and SA) have a constant markup with respect to the product costs, and this constant markup can be computed by a closed-form formula for the DET and SA approaches and by binary search for the RO approach.

We choose $m = 50$ and manually choose a vector of item costs $\mathbf{c}$ where $c_i \in [10, 30]$, $\forall i \in \mathcal{V}$, and the mean choice parameters $(\mathbf{a}_0, b_0)$ where $a_{0,i} \in [7, 23]$, $\forall i \in \mathcal{V}$, and $b_0 = 0.535$, noting that in this experiment, there is only one price sensitive parameter for all the products. Given an uncertainty level $\epsilon > 0$, we define the polyhedron uncertainty $\mathcal{A} = \{(\mathbf{a}, b)|\ ||\mathbf{a} - \mathbf{a}_0||_1 + \beta||b - b_0||_1 \leq \epsilon\}$, where $||.||_1$ stands for the L1 norm and β is the scale of the PSP b with respect to $\mathbf{a}$. The comparison is done as follows. For each ϵ, we solve the corresponding robust problem and obtain a robust solution $\mathbf{x}^{\text{RO}}$. For the DET, we solve the deterministic model with the mean-value parameters $(\mathbf{a}_0, \mathbf{b}_0)$ and obtain an optimal solution $\mathbf{x}^{\text{DET}}$. For the SA approach, we sample randomly and uniformly n_1 points from the uncertainty set, and for each point compute the corresponding optimal prices, which have a constant markup over products. For each pricing solution, we again sample randomly and uniformly $n_2 = 1000$ choice parameters from $\mathcal{A}$, and compute and pick a pricing solution with the largest worst-case expected revenue among the 1000 samples. We test this approach with $n_1 = 10$ and $n_1 = 50$ and denote the corresponding solutions as $\mathbf{x}^{\text{SA10}}$, $\mathbf{x}^{\text{SA50}}$, respectively. Larger n_1 can be chosen, but it would mean that the SA becomes way more expensive as compared to the RO and DET approaches. For example, if we choose $n_1 = 100$, the SA requires to solve 100 deterministic problems and compute 10^5 expected revenues to obtain a pricing solution.

To evaluate the performance of the three approaches when the choice parameters vary, given the uncertainty set defined above, we randomly and uniformly sample 1000 parameters $(\mathbf{a}, \mathbf{b})$ from the set $\mathcal{A}$, and compute the expected revenues given by $\mathbf{x}^{\text{RO}}$, $\mathbf{x}^{\text{SA10}}$, $\mathbf{x}^{\text{SA50}}$, and $\mathbf{x}^{\text{DET}}$. So, for each

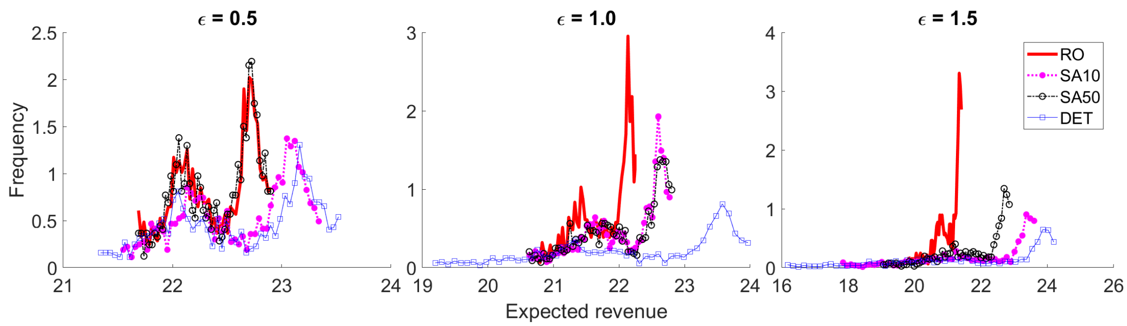

Figure 1 Comparison between revenue distributions given by two optimal price vectors given by the robust (RO) and deterministic (DET) approaches, under the MNL choice model and different uncertainty level ϵ.

solution, we get a distribution of expected revenues over 1000 samples, and we draw the histograms of the distributions obtained in Figure 1 for $\epsilon \in \{0.5, 1.0, 1.5\}$.

When $\epsilon = 0.5$, the distributions given by the RO and SA50 are similar, but with larger ϵ we see that the distributions given by the RO approach have higher peaks, lower variances and shorter tails, as compared to the other approaches. In addition, the sampling-based approach (SA10 and SA50) perform better then the DET in terms of protecting us against too low revenues. In this aspect, the the SA50 also performs better than the SA10, especially when ϵ is large.

In Table 1 in Appendix E, we provide more details about the maximum, average and worst-case values of the distributions given by the three approach. In particular, we compute the "*percentile ranks*" of the RO worst-case revenues, which indicates the percentages that the expected revenues given by the baseline approaches (DET, SA10 and SA50) are lower than the corresponding worst-case expected revenues given by the RO. For example, for $\epsilon = 1.5$, there are 19% of the revenues given by the DET (over 1000 sampled revenues) are less than the corresponding RO worst-case revenue (i.e. 19.8). Over $\epsilon \in \{0.5, \ldots, 10\}$, the average percentile ranks of the RO worst-case revenues are 34%, 27.3% and 26.1% for the DET, SA10 and SA50 approaches, respectively, which clearly indicates gains from the use of the RO approach, especially when the uncertainty is high.

We also provide experimental results for the robust model under the nested logit model, which was done under the same settings as for the MNL case, i.e., there are 50 items with the same mean-value choice parameters $(\mathbf{a}_0, \mathbf{b}_0)$ and the same item costs $\mathbf{c}$. The CPGF of the nested logit model is given as $G(\mathbf{Y}) = \sum_{n\in[N]} \left(\sum_{i\in C_n} Y_i^{\mu_n}\right)^{\mu/\mu_n}$, where $[N]$ is the set of nests and for each $n \in [N]$, C_n is the corresponding subset of the items, μ and μ_n, $n \in \mathcal{N}$ are the positive parameters of the nested logit model. In this experiment, we separate the whole item set into 5 nests of the same size (10 items per each nest), i.e. $N = 5$ and $|C_n| = 10$ for all $n \in [N]$. Moreover, we select $\mu = 1$ and μ_n, $n \in \mathcal{N}$ are 0.88, 0.61, 0.92, 0.95 and 1.0, respectively. The corresponding distributions are plotted in Figure 2, which are quite similar to the case of the MNL model. We further refer the reader to Table 2 in Appendix E for more details about the average, maximum, worst-case revenues and percentile ranks when ϵ increases, for which a similar observation as in the case of the MNL model applies. It is interesting to note that the average percentile ranks in this case are remarkably higher than those from the MNL case, meaning that the RO approach does a better job when the choice model is the nested logit. In general, we also see the advantages of the RO approach in protecting us from "bad" scenarios of the parameters $(\mathbf{a}, \mathbf{b})$.

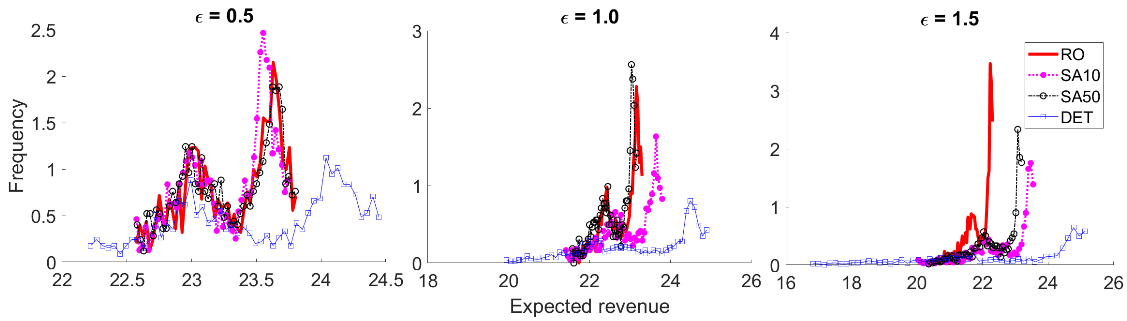

Figure 2 Comparison between revenue distributions given by two optimal price vectors given by the robust (RO) and deterministic (DET) approaches, under the nested logit model and different uncertainty level ϵ.

To further illustrate the protection of the RO approach in the worst-case scenarios, in Figure 3 we plot the worst-case revenues given by $\mathbf{x}^{\text{RO}}$, $\mathbf{x}^{\text{SA10}}$, $\mathbf{x}^{\text{SA50}}$ and $\mathbf{x}^{\text{DET}}$ when ϵ varies from 0 to 10. The worst-case revenues given by the DET and SA solutions decrease fast when ϵ increases and become close to zero when $\epsilon \geq 5$. On the other hand, the worst-case revenues given by the RO approach are still significant even with highest values of ϵ considered ($\epsilon = 10$). It is also interesting to see the the curves given by the MNL is more smooth than those from the nested logit model.

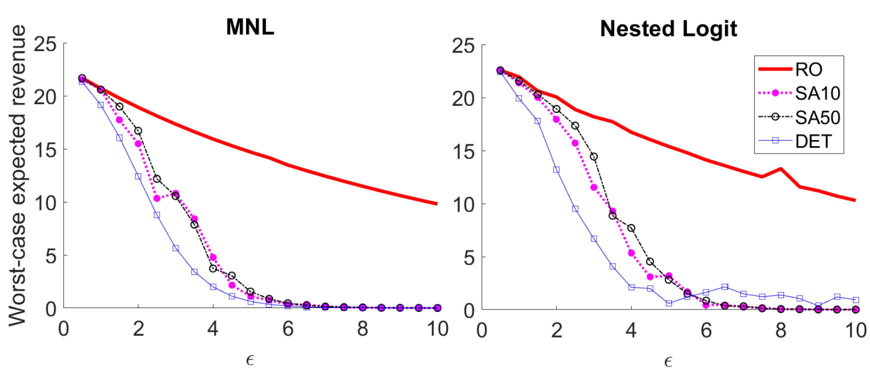

Figure 3 Worst-case expected revenues as functions of ϵ

In summary, the baseline approaches (DET, SA10, SA50) always give higher average and maximum revenues, but lower worst-case revenues, which clearly indicates that the RO approach does a better job in protecting us from worst-case situations, but also show the trade-off of being robust. Moreover, the results in Tables 1 and 2 also tell us that if the firm cares more about the worst cases, a large ϵ can be chosen to have better protection against too low expected revenues. On the other hand, if average performance is of concern, then by choosing a small ϵ, one can still get a protection from the robust solutions, but also get an average performance that is comparable to that of the solutions by the deterministic approach. This observation is also consistent with other robust approaches in the revenue management literature (Li and Ke 2019, Rusmevichientong and Topaloglu 2012).

5.2. Unconstrained Pricing with Partition-wise Homogeneous PSP

In this section we provide comparison results for the case of partition-wise homogeneous PSP considered in Section 3.2. We use the nested logit model specified above but the PSP are the same in each nest but different across nests. Note that in this case, we only provide results for the robust model with rectangular uncertainty sets. In this context, we know that the robust problem can be converted equivalently into the deterministic problem with parameters $(\underline{\mathbf{a}}, \overline{\mathbf{b}})$. So basically, the RO and DET approaches can be done by solving two deterministic problems, one with the choice parameters $(\underline{\mathbf{a}}, \overline{\mathbf{b}})$ and one with the mean-value parameters $(\mathbf{a}_0, \mathbf{b}_0)$. On the other hand, for the SA approach, if we select n_1 vectors of choice parameters from the uncertainty set, we need to solve n_1 deterministic problems.

For each uncertainty level $\epsilon > 0$, we define an rectangular uncertainty set $\mathcal{A}^\epsilon = \{(\mathbf{a}, \mathbf{b})|\ \mathbf{a} \in [\mathbf{a}_0 - \epsilon\mathbf{1}, \mathbf{a}_0 + \epsilon\mathbf{1}];\ \mathbf{b} \in [\mathbf{b}_0 - \epsilon\mathbf{1}/\beta, \mathbf{b}_0 + \epsilon\mathbf{1}/\beta]\}$, where $\mathbf{1}$ is a unit vector of appropriate size. Vector $\mathbf{a}_0$ is chosen similarly as in the previous section and $\mathbf{b}_0$ is chosen as $(0.535, 0.635, 0.335, 0.735, 0.454)$. The latter selection is due to that fact that the nested logit model has 5 nests and the PSP are the same in each nest. Here, to simplify the experiments, we only consider uncertainty sets where the differences between the bounds and the mean values are the same over all the coordinates. Similarly to the previous section, we first solve the deterministic problem with the mean-value parameters $(\mathbf{a}_0, \mathbf{b}_0)$ to obtain a solution $\mathbf{x}^{\text{DET}}$. Then, for each set $\mathcal{A}^\epsilon$ we solve the RO problem to obtain a robust solution $\mathbf{x}^{\text{RO}}$ (i.e., solve the deterministic pricing problem with choice parameters $(\mathbf{a}_0 - \epsilon\mathbf{1}, \mathbf{b}_0 + \epsilon\mathbf{1}/\beta)$). We also sample $n_1 = 10$ and $n_1 = 50$ points from $\mathcal{A}^\epsilon$ for the the SA approach.

To evaluate the performance of the solutions obtained, we also sample 1000 points randomly and uniformly from $\mathcal{A}^\epsilon$ and compute the expected revenues given by $\mathbf{x}^{\text{RO}}$, $\mathbf{x}^{\text{SA10}}$, $\mathbf{x}^{\text{SA50}}$, and $\mathbf{x}^{\text{DET}}$. The distributions of the expected revenue over 1000 samples with $\epsilon \in \{0.5, 1.0, 1.5\}$ are plotted in Figure 4. There is nothing surprising, as similarly to the previous experiments, distributions given by $\mathbf{x}^{\text{RO}}$ have small variances, higher peaks, shorter tails and higher worst-case revenues, as compared to those from $\mathbf{x}^{\text{SA10}}$, $\mathbf{x}^{\text{SA50}}$ and $\mathbf{x}^{\text{DET}}$. We also refer the reader to Table 3 in Appendix E for more details about the average, maximum and worst-case revenues when ϵ increases from 0.2 to 4.0. We also see that the RO approach always gives higher worst-case revenues but lower average revenues, and the SA approaches also provide some protections against low revenues. However, in this case, even-though the percentile ranks for the DET approach are still high (30.6 on average), those from the SA50 are significantly lower (5.6 on average). In particular, we see that there are some instances where the percentile ranks are only 3-*th*, which means that only 3% of the revenues are lower than the corresponding RO worst-case revenues. Nevertheless , the average revenues given by the SA50 are remarkably higher than those from the RO, especially when ϵ is large. As such, the RO seems too conservative in this case (partially homogeneous PSP with rectangular uncertainty sets), and large ϵ should not be chosen.

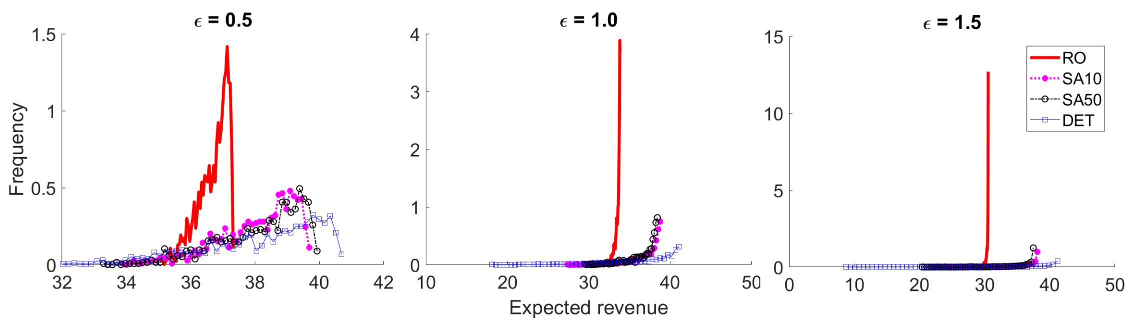

Figure 4 Distributions of the expected revenues under over-expected-sale penalties given by $\mathbf{x}^{\text{RO}}$ and $\mathbf{x}^{\text{DET}}$ under a nested logit model with partition-wise homogeneous PSP.

5.3. Pricing with Over-expected-sale Penalties

We first provide experiments to show how the constrained pricing model performs when the choice parameters are uncertain, and how the deterministic pricing model with over-expected-sale penalties works, as compared to the constrained counterpart. We take the nested logit model with homogeneous PSP considered above. We create one expected sale constraint (i.e., $T = 1$) in such a way that the optimal prices from the unconstrained problem do not satisfy the expected sale constraint.

In the first experiment, our goal is to show how the constrained pricing model (28) performs when the choice parameters are uncertain. To this end, we solve the deterministic problem with the mean-value parameters $(\mathbf{a}_0, \mathbf{b}_0)$ to obtain a solution $\mathbf{x}^{\text{DET}}$. Then, we select an uncertainty level $\epsilon > 0$ and assume that the choice parameters vary in the uncertainty set $\mathcal{A}^\epsilon = \{(\mathbf{a}, \mathbf{b})|\ \mathbf{a} \in [\mathbf{a}_0 - \epsilon\mathbf{1}, \mathbf{a}_0 + \epsilon\mathbf{1}];\ \mathbf{b} \in [\mathbf{b}_0 - \epsilon\mathbf{1}/\beta, \mathbf{b}_0 + \epsilon\mathbf{1}/\beta]\}$. We then randomly (and uniformly) sample $\mathbf{a} \in [\mathbf{a}_0 - \epsilon\mathbf{1}, \mathbf{a}_0 + \epsilon\mathbf{1}]$ and $\mathbf{b} \in [\mathbf{b}_0 - \epsilon\mathbf{1}/\beta, \mathbf{b}_0 + \epsilon\mathbf{1}/\beta]$. For each sample, we compute the corresponding vector of purchase probabilities and check whether that vector satisfies the expected sale constraints. We also compute the penalty costs as percents of the thresholds r_t, which can explain how the new purchase probabilities violate the constraints. These percentages based on 1000 samples are plotted in the left sub-figure of Figure 5. We see that there are about 50% of the samples for which the purchase probabilities given by $\mathbf{x}^{\text{DET}}$ violate the expected sale constraints. On the other hand, the penalty costs increase significantly when ϵ increases.

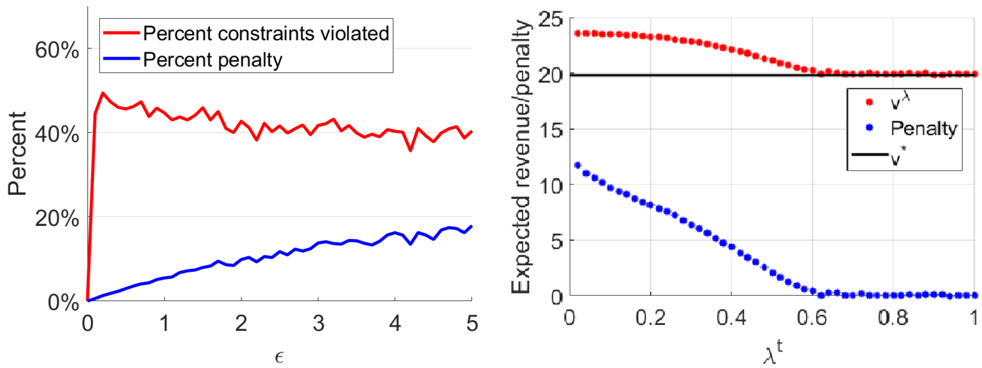

Figure 5 **Left figure:** Percentages of times that the purchase probabilities given by $\mathbf{x}^{\text{DET}}$ violate the expected sale constraints, and penalty costs as percents of r_t when ϵ varies. **Right figure:** The expected revenue values and penalty costs given by (18) when λ_t vary.

To illustrate the convergence of the optimal values of the pricing problem with over-expected-sale penalties when the penalty parameters $\boldsymbol{\lambda}$ increase, in Figure 5 we plot $v^{\boldsymbol{\lambda}}$, the optimal value of the constrained pricing problem v^*, and the penalty costs $\max\{0, (\boldsymbol{\alpha}^1)^{\text{T}}\mathbf{p} - r_1\}$ as functions of λ_1, noting that in this experiment $T = 1$. We see that when $\lambda_1 \geq 0.6$, the objective values $v^{\boldsymbol{\lambda}}$ become very close to the optimal value of the constrained pricing problem, and the penalty cost is also close to zero. This observation is indeed in line with the claims established in Theorem 3.

Now, we move to the robust version of the pricing problem under choice parameter uncertainty. We select $m = 50$ and consider a MNL-based pricing problem with one expected sale constraint. The expected sale constraint and the mean-value parameters $(\mathbf{a}_0)$ are the same as in the previous experiments and the parameters $\mathbf{b}_0$ are chosen similarly as in the case of pricing under the nested logit model with partition-wise homogeneous PSP. For each uncertainty level $\epsilon > 0$, we define the rectangular uncertainty set $\mathcal{A}^{\epsilon} = \{(\mathbf{a}, \mathbf{b}) |\ \mathbf{a} \in [\mathbf{a}_0 - \epsilon\mathbf{1}, \mathbf{a}_0 + \epsilon\mathbf{1}];\ \mathbf{b} \in [\mathbf{b}_0 - \epsilon\mathbf{1}/\beta, \mathbf{b}_0 + \epsilon\mathbf{1}/\beta]\}$.

Our goal here is to illustrate how the robust model with over-expected-sale penalties performs, as compared to other baseline approaches, i.e., deterministic and sampling-based counterparts and the deterministic constrained pricing problem. To this end, we solve the deterministic constrained pricing problem with the mean-value parameters $(\mathbf{a}_0, \mathbf{b}_0)$ to obtain a solution $\mathbf{x}^{\text{DET-CON}}$, solve the deterministic pricing problem under over-expected-sale penalties with the mean-value parameters $(\mathbf{a}_0, \mathbf{b}_0)$ to obtain a solution $\mathbf{x}^{\text{DET-PEN}}$. For the SA approach, we only sample $n_1 = 10$ points from the uncertainty set, due to the fact that the number of points n_1 is also the number of convex optimization problems to be solved, and these optimization problems, even-though computationally tractable, are much more expensive to solve, as compared to closed-form solutions for the case of unconstrained pricing considered in Section 3.1. As such, for each ϵ, we sample 10 points from $\mathcal{A}^\epsilon$ and use simulation of 1000 samples to select a solution $\mathbf{x}^{\text{SA10}}$. For the RO approach, we solve the robust problem (19) with the uncertainty set $\mathcal{A}^\epsilon$ to obtain a robust solution $\mathbf{x}^{\text{RO}}$. Theorem 4 tells use that for the DET-PEN, RO and DET-CON approaches, we only need to solve one convex optimization problem, while the SA10 requires to solve 10 convex nonlinear optimization problems.

To evaluate the performance of the solutions obtained, similarly to the previous sections, we sample randomly and uniformly 1000 points from $\mathcal{A}$ and compute the corresponding expected revenues given by the four solutions $\mathbf{x}^{\text{DET-CON}}$, $\mathbf{x}^{\text{DET-PEN}}$, $\mathbf{x}^{\text{SA10}}$ and $\mathbf{x}^{\text{RO}}$. The distributions of the profit values (the expected revenue minus the penalty cost) over 1000 choice parameter samples for different λ and ϵ are plotted in Figure 6, where similar observations apply. The histograms given by $\mathbf{x}^{\text{RO}}$ have higher peaks, smaller variances, shorter tails and get tighter as ϵ increases. When $\lambda = 0.2$, the histograms given by $\mathbf{x}^{\text{SA10}}$ have higher peaks, smaller variances and better worst-case values than those from $\mathbf{x}^{\text{DET-CON}}$, $\mathbf{x}^{\text{DET-PEN}}$. This seems however not the case for $\lambda = 0.6$, especially when $\lambda = 0.6$ and $\epsilon = 1.5$, the SA10 gives remarkably low profit values as compared to the other approach. More detailed results are provided in Table 4 in Appendix E. In general, we also see that the RO approach always gives higher worst-case but lower average profits. The SA10 also provides some protections against worst-case scenarios, The DET-PEN and SA10 seems to have similar performance, which may be due to the fact that the number of samples points n_1 is small. The percentile ranks of the RO worst-case profits are small for the DET-PEN and SA10 approaches, as compared to those reported in previous experiments, which indicates the conservativeness of the RO approach, especially when ϵ is large.

We also observe that $\mathbf{x}^{\text{DET-CON}}$ gives lower-value histogram when $\lambda = 0.2$ and similar histogram when $\lambda = 0.6$, as compared to $\mathbf{x}^{\text{DET-PEN}}$, which is in line with the claims of Theorem 3 stating that when λ increases, solution given by the pricing problem with over-expected-sale penalties converge to solutions given by the constrained pricing problem. To further illustrate this, we fix $\epsilon = 0.5$ and increase λ from 0.1 to 0.4 and plot the corresponding histograms in Figure 7, which clearly shows that $\mathbf{x}^{\text{DET-CON}}$ always gives lower-value histograms, and these histograms become similar to those given by $\mathbf{x}^{\text{DET-PEN}}$ when λ increase.

In summary, our experiments for the three cases (robust unconstrained pricing with homogeneous and partition-wise homogeneous PSP, and robust pricing with over-expected-sale penalties) show

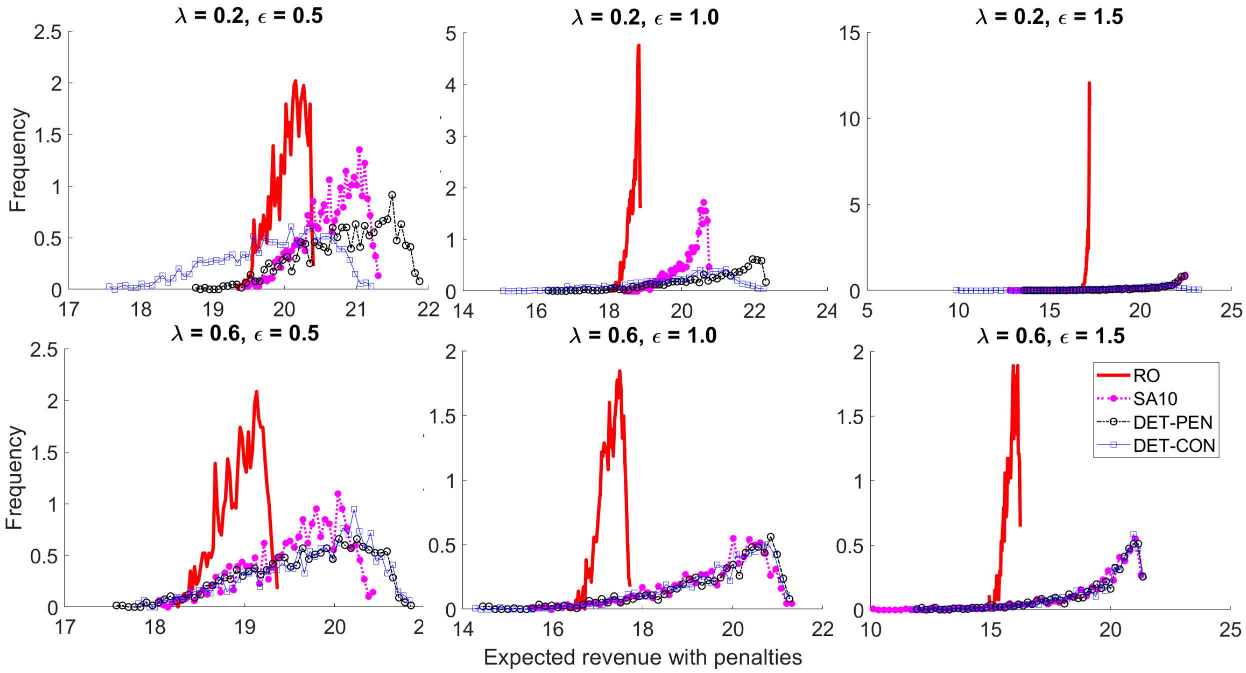

Figure 6 Distributions of the profit values under over-expected-sale penalties given by $\mathbf{x}^{\text{RO}}$, $\mathbf{x}^{\text{SA10}}$, $\mathbf{x}^{\text{DET-PEN}}$ and $\mathbf{x}^{\text{DET-CON}}$

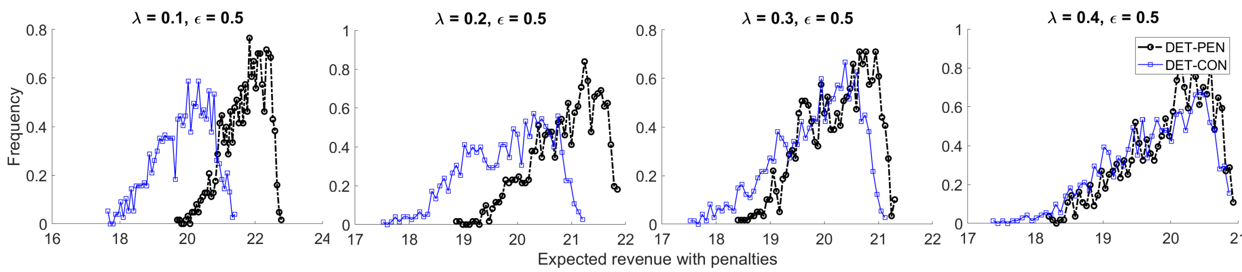

Figure 7 Distributions of the profits given by $\mathbf{x}^{\text{DET-PEN}}$ and $\mathbf{x}^{\text{DET-CON}}$ when $\lambda \in \{0.1, 0.2, 0.3, 0.4\}$.

gains from our robust models in protecting us from revenues that would be too low. The histograms given by the robust models have higher peaks, smaller variances, higher worst-case revenues, but lower averages, as compared to their deterministic counterparts. This observation also shows the trade-off in being robust in making pricing decisions when the choice parameters are uncertain, and also consistent with observations from other robust studies in the revenue management literature (Rusmevichientong and Topaloglu 2012, Li and Ke 2019).

6. Conclusion

In this paper, we have considered robust versions of the pricing problem under GEV choice models, in which the choice parameters are not given in advance but lie in an uncertainty set. These robust models are motivated by the fact that uncertainties may occur in the estimation procedure of the choice parameters. We have shown that when the problem is unconstrained and the PSP are the same over all the products, the robust optimal prices have a constant markup with respect to the product costs and we have shown how to efficiently compute this constant markup by binary search. When the PSP are partition-wise homogeneous and the CPGF are also partition-wise separable, we have shown that if the uncertainty set is rectangular, we can convert the robust problem into an equivalent deterministic pricing problem, in which the optimal prices can be computed by explicit formulas.

We have also considered the pricing problem with over-expected-revenue-penalties as an alternative to the constrained pricing problem. We have shown that, when the penalty parameters goes to infinity, the penalty term converges to zero and the optimal value converges to the expected revenue given by the constrained pricing problem. Since there may be no fixed prices under which the purchase probabilities always satisfy the expected sale constraints when the choice parameters vary, the robust version of the pricing problem under over-expected-sale penalties is more appropriate to use in this context, as opposed to the robust constrained version. We have shown that if the uncertainty set is rectangular and the choice model is MNL, the robust problem can be converted equivalently into a deterministic one that can be solve efficiently by convex optimization. This results can be extended to robust pricing problems under GEV models under some additional assumptions on the expected sale parameters and the feasible set of the prices. Experimental results based on the MNL and nested logit models have shown the advantages of our robust model in providing protection against bad-case revenues. In future research, it would be interesting to look at distributionally robust versions of the pricing problem, which may help provide less conservative robust solutions as compared to the standard robust optimization approaches. We are also interested in robust approaches for the joint assortment and pricing problem under GEV choice models.

Acknowledgments

This research is supported by the National Research Foundation, Prime Ministers Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) program, Singapore-MIT Alliance for Research and Technology (SMART) Future Urban Mobility (FM) IRG.

References

Moshe E Ben-Akiva. *Structure of passenger travel demand models.* PhD thesis, Massachusetts Institute of Technology, 1973.

Moshe E Ben-Akiva, Steven R Lerman, and Steven R Lerman. *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press, 1985.

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

Aharon Ben-Tal and Arkadi Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical programming*, 88(3):411–424, 2000.

Aharon Ben-Tal, Stephen Boyd, and Arkadi Nemirovski. Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Mathematical Programming*, 107(1-2):63–89, 2006.

Chandra R Bhat. Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transportation Research Part A: Policy and Practice*, 32(7):495–507, 1998.

Andrew Daly and Michel Bierlaire. A general and operational representation of generalised extreme value models. *Transportation Research Part B: Methodological*, 40(4):285–305, 2006.

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

Lingxiu Dong, Panos Kouvelis, and Zhongjun Tian. Dynamic pricing and inventory control of substitute products. *Manufacturing & Service Operations Management*, 11(2):317–339, 2009.

Mogens Fosgerau, Daniel McFadden, and Michel Bierlaire. Choice probability generating functions. *Journal of Choice Modelling*, 8:1–18, 2013.

Guillermo Gallego and Ming Hu. Dynamic pricing of perishable assets under competition. *Management Science*, 60(5):1241–1259, 2014.

Guillermo Gallego and Garrett Van Ryzin. A multiproduct dynamic pricing problem and its applications to network yield management. *Operations research*, 45(1):24–41, 1997.

Guillermo Gallego and Ruxian Wang. Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. *Operations Research*, 62(2):450–461, 2014.

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

Wallace J Hopp and Xiaowei Xu. Product line selection and pricing with modularity in design. *Manufacturing & Service Operations Management*, 7(3):172–187, 2005.

Woonghee Tim Huh and Hongmin Li. Pricing under the nested attraction model with a multistage choice structure. *Operations Research*, 63(4):840–850, 2015.

Philipp Wilhelm Keller. *Tractable multi-product pricing under discrete choice models*. PhD thesis, Massachusetts Institute of Technology, 2013.

Frank S Koppelman and Chieh-Hua Wen. The paired combinatorial logit model: properties, estimation and application. *Transportation Research Part B: Methodological*, 34(2):75–89, 2000.

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

Hongmin Li and Woonghee Tim Huh. Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications. *Manufacturing & Service Operations Management*, 13(4):549–563, 2011.

Hongmin Li, Scott Webster, Nicholas Mason, and Karl Kempf. Product-line pricing under discrete mixed multinomial logit demand. *Manufacturing & Service Operations Management*, 21(1):14–28, 2018.

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

Tien Mai, Emma Frejinger, Mogens Fosgerau, and Fabian Bastin. A dynamic programming approach for quickly estimating large network-based mev models. *Transportation Research Part B: Methodological*, 98:179–197, 2017.

Daniel McFadden. Modeling the choice of residential location. *Transportation Research Record*, (673), 1978.

Daniel McFadden. Econometric models for probabilistic choice among products. *Journal of Business*, pages S13–S29, 1980.

Daniel McFadden and Kenneth Train. Mixed mnl models for discrete response. *Journal of applied Econometrics*, 15(5):447–470, 2000.

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

Alexander Shapiro. Tutorial on risk neutral, distributionally robust and risk averse multistage stochastic programming. *Optimization Online http://www. optimization-online. org/DB_HTML/2018/02/6455. html*, 2018.

Kenneth A Small. A discrete choice model for ordered alternatives. *Econometrica: Journal of the Econometric Society*, pages 409–424, 1987.

Jing-Sheng Song and Zhengliang Xue. Demand management and inventory control for substitutable products. *Working paper*, 2007.

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

Peter Vovsha and Shlomo Bekhor. Link-nested logit model of route choice: overcoming route overlapping problem. *Transportation research record*, 1645(1):133–142, 1998.

Chieh-Hua Wen and Frank S Koppelman. The generalized nested logit model. *Transportation Research Part B: Methodological*, 35(7):627–641, 2001.

GRTA Whelan, R Batley, T Fowkes, and A Daly. Flexible models for analyzing route and departure time choice. *Publication of: Association for European Transport*, 2002.

Dan Zhang and Zhaosong Lu. Assessing the value of dynamic pricing in network revenue management. *INFORMS Journal on Computing*, 25(1):102–115, 2013.

Heng Zhang, Paat Rusmevichientong, and Huseyin Topaloglu. Multiproduct pricing under the generalized extreme value models with homogeneous price sensitivity parameters. *Operations Research*, 66(6): 1559–1570, 2018.

Appendix A: Robust Constrained Pricing

Motivated by the fact that the expected profit is concave in the purchase probabilities, previous studies (Zhang et al. 2018, Keller 2013) show that it is convenient to consider the pricing problem with expected sale constraints. Technically speaking, given a GEV-CPGF $G(\mathbf{Y})$, price vector $\mathbf{x} \in \mathbb{R}^m$ andparameters $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{2m}$, let us define the vector of purchase probabilities of products $\mathbf{p}$ with entries $p_i = P_i(\mathbf{x}, \mathbf{a}, \mathbf{b}|G)$. We also let $\mathbf{x}(\mathbf{p}|\mathbf{a}, \mathbf{b}, G)$ be the denote the prices that achieve the purchase probabilities $\mathbf{p}$. The deterministic version of the constrained pricing problem can be formulated as

$$\max_{\mathbf{p}\in\mathcal{P}} \quad \left(\sum_{i\in\mathcal{V}} \mathbf{x}(\mathbf{p}|\mathbf{a}, \mathbf{b})_i - c_i\right) p_i. \tag{28}$$

where $\mathcal{P} \in \mathbb{R}^m$ is a convex set such that for all $\mathbf{p} \in \mathcal{P}$, $\sum_{i\in\mathcal{V}} p_i \leq 1$. One the optimal purchase probabilities $\mathbf{p}$ is specified, we can obtain the *optimal prices* $\mathbf{x}(\mathbf{p}|\mathbf{a}, \mathbf{b}, G)$ by solving a convex optimization problem. A natural robust version of the constrained pricing problem can be formulated as

$$\max_{\mathbf{p}\in\mathcal{P}} \left\{ \phi(\mathbf{p}) = \min_{(\mathbf{a},\mathbf{b})\in\mathcal{A}} \sum_{i\in\mathcal{V}} \left(\mathbf{x}(\mathbf{p}|\mathbf{a}, \mathbf{b}, G)_i - c_i\right) p_i \right\}. \tag{29}$$

Even though it is not difficult to show (29) is computationally tractable under rectangular or some polyhedrons uncertainty sets, the issue here is that the final decision is a price vector, not purchase probabilities. So even if we get an optimal purchase probabilities $\mathbf{p}$ from the robust model, it is not clear how to compute the corresponding optimal prices under $(\mathbf{a}, \mathbf{b})$ uncertainty. On the other hand, one can show that given any prices $\mathbf{x}$, there may be $(\mathbf{a}, \mathbf{b}) \in \mathcal{A}$ such the resulting purchase probability vector $\mathbf{p} = P(\mathbf{x}, \mathbf{a}, \mathbf{b}|G)$ that does not belong to the feasible set (i.e. the expected sale constraints are not satisfied). All these make the robust version in (29) inappropriate to use. This is the reasonwe propose an alternative robust model in Section 4, in which instead of requiring the purchasing probabilities to satisfy some constraints, we add a penalty cost to the objective function.

Alternatively, in some situations the firm may face uncertainties occurring in the inventory, leading to uncertain expected sale constraints. A robust model may require that the expected sales constraints are satisfied for all the scenarios that may occur, i.e., $\mathbf{p} \in \mathcal{P}(\xi)$, for all $\xi \in \Xi$. Such a robust model can be formulated as

$$\begin{aligned} \max_{\mathbf{p}} \quad & \left(\sum_{i\in\mathcal{V}} \mathbf{x}(\mathbf{p}|\mathbf{a}, \mathbf{b})_i - c_i\right) p_i \\ \text{subject to} \quad & (\boldsymbol{\alpha}^t(\xi))^{\mathrm{T}}\mathbf{p} \leq r_t(\xi) \qquad \forall \xi \in \Xi \\ & \sum_{i\in\mathcal{V}} p_i \leq 1, \;\; \mathbf{p} \geq 0 \end{aligned} \tag{30}$$

where $(\boldsymbol{\alpha}^t(\xi), r_t(\xi))$, $\forall t$, are the parameters of the expected sale constraints, which are not certain in the context and depend on a random vector $\xi \in \Xi$. Since the objective function is concave and all the constraints are linear in $\mathbf{p}$, the above problem is generally tractable (Ben-Tal and Nemirovski 1998). A simple but useful setting is that the parameter of the expected sale constraints vary in a rectangular uncertainty set, i.e., $\underline{\boldsymbol{\alpha}}^t \preceq \boldsymbol{\alpha}^t \preceq \underline{\boldsymbol{\alpha}}^t$ and $\underline{r}^t \leq r^t \leq \overline{r}^t$ for all $t \in [T]$. In this context, one can show that (30) is equivalent to the following convex optimization problem

$$\begin{aligned} \max_{\mathbf{p}} \quad & \left(\sum_{i \in \mathcal{V}} \mathbf{x}(\mathbf{p}|\mathbf{a}, \mathbf{b})_i - c_i \right) p_i \\ \text{subject to} \quad & (\overline{\boldsymbol{\alpha}}^t)^{\mathrm{T}} \mathbf{p} \leq \underline{r}_t \\ & \sum_{i \in \mathcal{V}} p_i \leq 1, \;\; \mathbf{p} \geq 0 \end{aligned} \tag{31}$$

Other uncertainty sets may be considered, i.e., polyhedron or ellipsoidal ones, and we refer the reader to Ben-Tal and Nemirovski (1998) for details.

Appendix B: Proof of Corollary 3

We will give a counter example to illustrate the claim. For the sake of illustration, we only consider a pricing problem under the MNL model with 2 products and the PSP are homogeneous. We also consider only one expected sale constraint as $\alpha_1 p_1 \leq r_t$, where r_t/α_1 is very small. Let $\mathbf{p}^\lambda = (p_1^\lambda, p_2^\lambda)$ be a solution to (28) under penalty parameter λ. When λ goes to infinity, Theorem 3 tells us that a solution to the pricing problem with penalties converge to a solution to the constrained pricing problem. Thus, for any $\epsilon > 0$ arbitrarily small, we can chose λ large enough such that $\alpha_1 p_1^\lambda \leq r_t + \epsilon$. So, if we choose ϵ and r_t/α_1 to be very small, then p_1^λ would be very close to zero. Since $p_1^\lambda = \exp(a_1 - bx_1^\lambda) \Big/ (1 + \exp(a_1 - bx_1^\lambda) + \exp(a_2 - bx_2^\lambda))$ (where b is the PSP of the two products, $\mathbf{x}^\lambda$ is an optimal price solution to the pricing problem under penalty parameter λ), we have $\lim_{p_1^\lambda \to 0} x_1^\lambda(\mathbf{p}) = +\infty$, meaning that to have an arbitrarily small probability p_1^λ, we need to increase the price of Product 1 to infinity. On the other hand, p_2^λ does not affect the penalty term and we have $\lim_{x_2 \to +\infty} (x_2 - c_2) \exp(a_2 - bx_2) \Big/ (1 + \exp(a_1 - bx_1^\lambda) + \exp(a_2 - bx_2)) = 0$. Thus, to maximize the objective function, the solution x_2^λ needs to be finite. So, in summary, we can create an example yielding a solution $(x_1^\lambda, x_2^\lambda)$ such that x_1^λ can be arbitrarily large and x_2^λ is bounded from above. Thus, $(x_1^\lambda, x_2^\lambda)$ would not have the constant-markup style. If the PSP not not homogeneous, but partition-wise homogeneous, we just consider a model that has only one partition, and the above example can be reused.

Appendix C: Proof of Theorem 3

First, for notational simplicity we denote $R(\mathbf{p}) = \sum_{i\in\mathcal{V}} (\mathbf{x}(\mathbf{p}|\mathbf{a},\mathbf{b},G)_i - c_i)\, p_i$. For (i), we have the following inequalities

$$\begin{aligned}
R(\mathbf{p}^{\boldsymbol{\lambda}^1}) - \sum_t \lambda_t^1 \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}^1} - r_t\} &\geq R(\mathbf{p}^{\boldsymbol{\lambda}^2}) - \sum_t \lambda_t^1 \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}^2} - r_t\}\\
&= R(\mathbf{p}^{\boldsymbol{\lambda}^2}) - \sum_t \lambda_t^2 \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}^2} - r_t\} - \epsilon\sum_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}^2} - r_t\}\\
&\geq R(\mathbf{p}^{\boldsymbol{\lambda}^1}) - \sum_t \lambda_t^2 \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}^1} - r_t\} - \epsilon\sum_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}^2} - r_t\}
\end{aligned}$$

So, we have

$$\epsilon\sum_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}^2} - r_t\} \geq \sum_t (\lambda_t^1 - \lambda_t^2) \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}^1} - r_t\},$$

which leads to the desired inequality.

For (ii), since $(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^* \leq r_t$ for all t, given $\boldsymbol{\lambda} \in \mathbb{R}_+^T$, we have

$$v^{\boldsymbol{\lambda}} \geq R(\mathbf{p}^*) - \sum_t \lambda_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^* - r_t\} = R(\mathbf{p}^*) = v^*.$$

Moreover, since $v^{\boldsymbol{\lambda}} - v^* = R(\mathbf{p}^{\boldsymbol{\lambda}}) - v^* - \sum_t \lambda_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}} - r_t\}$, we have

$$R(\mathbf{p}^{\boldsymbol{\lambda}}) - v^* \geq \sum_t \lambda_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}} - r_t\} \geq \lambda_0 \sum_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}} - r_t\} \tag{32}$$

The left hand side of (32) is less than $\Delta^* - v^*$, so if we choose $\lambda_0 \geq (\Delta^* - v^*)/\epsilon$ then $\sum_t \max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}} - r_t\} \leq \epsilon$ as desired.

We move to (iii). As shown previously, we can choose λ_0 such that $\max\{0, (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}} - r_t\} \leq \epsilon$ or $(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}} \leq r_t + \epsilon$ for all $t \in [T]$. We now consider the following problem

$$\max_{\substack{\mathbf{p}\geq 0\\ \sum_i p_i \leq 1}} \left\{ R(\mathbf{p}) \middle|\ (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p} \leq r_t + \epsilon,\ \forall t \right\} \tag{33}$$

and denote by $\mathbf{p}^\epsilon$ as an optimal solution to (33). Since $\mathbf{p}^{\boldsymbol{\lambda}}$ is feasible to (33) we have $R(\mathbf{p}^\epsilon) \geq R(\mathbf{p}^{\boldsymbol{\lambda}})$. Moreover, if we define $\mathcal{P} := \{\mathbf{p} \in \mathbb{R}^m |\ p_i \geq 0, \sum_i p_i \leq 1,\ (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^{\boldsymbol{\lambda}} \leq r_t, \forall t \in [T]\}$, then $v^* \geq R(\mathbf{p})$ for all $\mathbf{p} \in \mathcal{P}$. Therefore, we have

$$|v^{\boldsymbol{\lambda}} - v^*| \leq R(\mathbf{p}^\epsilon) - R(\mathbf{p}),\ \forall \mathbf{p} \in \mathcal{P}. \tag{34}$$

We will show that there is $\mathbf{p} \in \mathcal{P}$ such that $||\mathbf{p}^\epsilon - \mathbf{p}||$ can be arbitrarily small when ϵ decreases, which allows us to use the *Mean Value Theorem* to bound $|R(\widetilde{\mathbf{p}}^\epsilon) - R(\mathbf{p})|$. If $\mathbf{p}^\epsilon \in \mathcal{P}$, then the result is obvious and we have $|v^{\boldsymbol{\lambda}} - v^*|$. Now assume that $\mathbf{p}^\epsilon \notin \mathcal{P}$, let $\mathcal{T} := \{t \in [T]|\ (\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^\epsilon > r_t\}$ and for any $t\in\mathcal{T}$ we select $i_t=\text{argmax}_{i\in\mathcal{V}}\{p^\epsilon_i|\ \alpha^t_i>0\}$. Then we denote $\mathcal{I}=\{i_t|\ t\in\mathcal{T}\}$. We pick a $\widetilde{\mathbf{p}}$ such that

$$\begin{cases}\widetilde{p}_i=p^\epsilon_i-\epsilon/(\min_{t,j}\{\alpha^t_j|\ \alpha^t_j>0\}),\ \forall i\in\mathcal{I}\\ \widetilde{p}_j=p^\epsilon_j,\ \forall j\notin\mathcal{I}.\end{cases} \tag{35}$$

With this selection, we see that, for any $t\in\mathcal{T}$

$$\begin{aligned}(\boldsymbol{\alpha}^t)^{\mathrm{T}}\widetilde{\mathbf{p}}&\le(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^\epsilon-\alpha^t_{i_t}\epsilon/(\min_{t,i}\{\alpha^t_i|\ \alpha^t_i>0\})\\ &\le(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^\epsilon-\epsilon\le r_t.\end{aligned} \tag{36}$$

And indeed for any $t\notin\mathcal{T}$ we have $(\boldsymbol{\alpha}^t)^{\mathrm{T}}\widetilde{\mathbf{p}}\le(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^\epsilon\le r_t$. Furthermore, for any $t\in\mathcal{T}$, we have $(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^\epsilon>r_t$. Combine this with the fact that $i_t=\text{argmax}_{i\in\mathcal{V}}\{p^\epsilon_i|\ \alpha^t_i>0\}$ we have

$$\left(\sum_i\alpha^t_i\right)p^\epsilon_{t_t}\ge(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{p}^\epsilon>r_t.$$

So, under the assumption on the selection of ϵ, we have the chain of inequalities

$$p^\epsilon_{i_t}>\frac{r_t}{(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{1}}\ge\min_t\left\{\frac{r_t}{(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{1}}\right\}\ge\frac{\epsilon}{\min_{t,i}\{\alpha^t_i|\ \alpha^t_i>0\}},$$

meaning that $\widetilde{\mathbf{p}}>0$. So, combine with (36) we have $\widetilde{\mathbf{p}}\in\mathcal{P}$.

Moreover, for any point $\mathbf{p}'\in[\widetilde{\mathbf{p}},\mathbf{p}^\epsilon]$ and any $i\in\mathcal{I}$, we have

$$\begin{aligned}p'_i\ge\widetilde{p}_i=p^\epsilon_i&-\frac{\epsilon}{\min_{t,i}\{\alpha^t_i|\ \alpha^t_i>0\}}\\ &>\min_t\left\{\frac{r_t}{(\boldsymbol{\alpha}^t)^{\mathrm{T}}\mathbf{1}}\right\}-\frac{\epsilon}{\min_{t,i}\{\alpha^t_i|\ \alpha^t_i>0\}}:=\delta(\epsilon).\end{aligned} \tag{37}$$

So, if we denote $\mathbf{x}'=\mathbf{x}(\mathbf{p}',G)$ (i.e., the prices that result in purchase probabilities $\mathbf{p}'$). For any $i\in\mathcal{I}$, under the assumption that $Y_i\partial G_i(\mathbf{Y})\le L_iY_i^{l_i}$ we have

$$\begin{aligned}L_iY_i(\mathbf{x}')^{l_i}&\ge p'_i(1+G(\mathbf{Y}(\mathbf{x}')))\delta(\epsilon)\\ &\ge p'_i\ge\delta(\epsilon),\end{aligned}$$

where $\mathbf{Y}(\mathbf{x}')$ is a vector of size m with entries $Y_j(\mathbf{x}')=\exp(a_j-b_jx'_j)$, $\forall j\in\mathcal{V}$. So we have

$$x'_i\le\frac{a_i}{b_i}-\frac{1}{b_il_i}\log\frac{\delta(\epsilon)}{L_i}.$$

Moreover, if we look at the gradient of $R(\mathbf{p})$ at p'_i. According to Theorem 4.3 in Zhang et al. (2018) we have

$$\nabla_{\mathbf{p}}R(\mathbf{p}')_i\le x'_i\le\max_i\left\{\frac{a_i}{b_i}-\frac{1}{b_il_i}\log\frac{\delta(\epsilon)}{L_i}\right\}. \tag{38}$$

Now, we look at $|R(\mathbf{p}^\epsilon)-R(\widetilde{\mathbf{p}})|$ and by combining (35), (38), the *Mean Value Theorem* tells us that there is $\mathbf{p}'\in[\widetilde{\mathbf{p}},\mathbf{p}^\epsilon]$

$$\begin{aligned}|R(\mathbf{p}^\epsilon)-R(\widetilde{\mathbf{p}})|&=\sum_{i\in\mathcal{I}}\nabla_{\mathbf{p}}R(\mathbf{p}')_i|\widetilde{p}_i-p^\epsilon_i|\\ &\le\max_i\left\{\frac{a_i}{b_i}-\frac{1}{b_il_i}\log\frac{\delta(\epsilon)}{L_i}\right\}\frac{m\epsilon}{\min_{t,i}\{\alpha^t_i|\ \alpha^t_i>0\}}\end{aligned} \tag{39}$$

Combine (39) with (34) and recall that $\widetilde{\mathbf{p}} \in \mathcal{P}$, we have

$$|v^{\boldsymbol{\lambda}} - v^*| \leq \max_i \left\{ \frac{a_i}{b_i} - \frac{1}{b_i l_i} \log \frac{\delta(\epsilon)}{L_i} \right\} \frac{m\epsilon}{\min_{t,i}\{\alpha_i^t | \ \alpha_i^t > 0\}}.$$

Combine with the case $\mathbf{p}^\epsilon \in \mathcal{P}$, we obtain the desired bound, which definitely converge to zero when ϵ tends to zero, as desired.

Appendix D: Proof of Proposition 6

For any price solution $\mathbf{x} \in X$, let $\mathbf{z} \in \mathbb{R}^n$ with entries $z_n = x_i - c_i$ for all $i \in \mathcal{V}_n$, $n \in [N]$. We can write expected revenue as

$$\Phi(\mathbf{x}, \mathbf{a}, \mathbf{b}) = \frac{\sum_{n\in[N]} z_n G^n(\mathbf{Y}^n | z_n, \mathbf{a}, \mathbf{b})}{1 + \sum_{n\in[N]} G^n(\mathbf{Y}^n | z_n, \mathbf{a}, \mathbf{b})},$$

Moreover, the assumption requiring that $\alpha_i^t = \alpha_j^t$ for all $i, j \in \mathcal{V}_n$, $n \in [N]$ help write the term $(\boldsymbol{\alpha}^t)^{\mathrm{T}} \mathbf{p}(\mathbf{x}, \mathbf{a}, \mathbf{b})$ as

$$(\boldsymbol{\alpha}^t)^{\mathrm{T}} \mathbf{p}(\mathbf{x}, \mathbf{a}, \mathbf{b}) = \frac{\sum_{n\in N} d_n^t G^n(\mathbf{Y}^n | z_n, \mathbf{a}, \mathbf{b})}{1 + \sum_{n\in[N]} G^n(\mathbf{Y}^n | z_n, \mathbf{a}, \mathbf{b})}$$

Thus, we can write the objective function of the robust problem as

$$\mathcal{L}(\mathbf{x}, \mathbf{a}, \mathbf{b}) = \mathcal{H}(\mathbf{z}, \mathbf{a}, \mathbf{b}) = \frac{\sum_{n\in[N]} z_n G^n(\mathbf{Y}^n | z_n, \mathbf{a}, \mathbf{b})}{1 + \sum_{n\in[N]} G^n(\mathbf{Y}^n | z_n, \mathbf{a}, \mathbf{b})} - \sum_{t=1}^{T} \lambda_t \max\{0, (\mathbf{d}^t)^{\mathrm{T}} \mathbf{p}^G - r_t\},$$

which is also the objective function given by a MNL model with N items. Lemma 4 tells us that $(\underline{\mathbf{a}}^n, \overline{\mathbf{b}}^n)$ is the unique solution to the minimization problem $\min_{(\mathbf{a}^n, \mathbf{b}^n)\in\mathcal{A}^n} G^n(\mathbf{Y}^n | z_n, \mathbf{a}, \mathbf{b})$ and $(\overline{\mathbf{a}}^n, \underline{\mathbf{b}}^n)$ is the unique solution to the maximization problem $\max_{(\mathbf{a}^n, \mathbf{b}^n)\in\mathcal{A}^n} G^n(\mathbf{Y}^n | z_n, \mathbf{a}, \mathbf{b})$ (we refer the reader to Section 3.2 for the definitions of $\mathcal{A}^n$, $\mathbf{a}^n$, $\mathbf{b}^n$). Then, we can follow the same way as in the proof of Lemmas 9, 10 and Theorem 4 to show that the robust problem $\max_{\mathbf{z}\in\mathbb{R}^N} \min_{(\mathbf{a},\mathbf{b})\in\mathcal{A}} \mathcal{H}(\mathbf{z}, \mathbf{a}, \mathbf{b})$ is equivalent to $\max_{\mathbf{z}\in\mathbb{R}^N} \mathcal{H}(\mathbf{z}, \underline{\mathbf{a}}, \overline{\mathbf{b}})$, which also means that the problem $\max_{\mathbf{x}\in\mathbb{R}^m} \min_{(\mathbf{a},\mathbf{b})\in\mathcal{A}} \mathcal{H}(\mathbf{z}, \mathbf{a}, \mathbf{b})$ is equivalent to $\max_{\mathbf{z}\in\mathbb{R}^N} \mathcal{H}(\mathbf{z}, \underline{\mathbf{a}}, \overline{\mathbf{b}})$.

Now we show that (27) can be converted equivalently into a convex optimization problem. For any vector $\mathbf{p}^G \in \Delta^N := \{\mathbf{p} \in \mathbb{R}_+^N | \ \sum_n p_n < 1\}$ such that $\sum_i p_i^G < 1$, we denote by $\mathbf{z}(\mathbf{p}^G)$ a vector of prices that results in probabilities $\mathbf{p}^G$ and let $\Theta(\mathbf{z}) : \mathbb{R}^N \to \mathbb{R}^N$ such that $\Theta(\mathbf{z})_n = G^n(\mathbf{Y}^n | z_n) \Big/ \left(1 + \sum_{j\in[N]} G^j(\mathbf{Y}^j | z_j)\right)$, where $G^n(\mathbf{Y}^n | z_n) = G^n(\mathbf{Y}^n | z_n, \underline{\mathbf{a}}, \overline{\mathbf{b}})$ but we omit the choice parameters $(\underline{\mathbf{a}}, \overline{\mathbf{b}})$ for notational simplicity. We also denote by $\widetilde{\mathbf{b}}$ a vector of size N with entries $\widetilde{b}_n = \langle \overline{\mathbf{b}}^n \rangle$. The lemma below show that given any $\mathbf{p}^G$, $\mathbf{z}(\mathbf{p}^G)$ can be uniquely determined by solving a strictly convex optimization problem. The structure of the problem presented in this lemma is slightly different with those considered in Theorem 4.1 in Zhang et al. (2018), so even though the proof of the lemma is quite similar, we provide its own proof for the sake of self-contained.

Lemma 11. *Given $\boldsymbol{p}^G \in \Delta^N$, there is a unique vector $\boldsymbol{z}(\boldsymbol{p}^G) \in \mathbb{R}^N$ such that $\Theta(\boldsymbol{z}(\boldsymbol{p}^G)) = \boldsymbol{p}^G$, and this $\boldsymbol{z}(\boldsymbol{p}^G)$ is the unique solution to the convex optimization problem.*

$$\min_{\boldsymbol{z}\in\mathbb{R}^N} \left\{ \ln\left(1 + \sum_{n\in[N]} G^n(\boldsymbol{Y}^n|z_n)\right) + \sum_{n\in[N]} p_n^G \widetilde{b}_n z_n \right\} \tag{40}$$

Proof: We first prove that (40) is a strictly convex optimization problem. To this end, we will show that $\nabla^2 \mathcal{Q}(\mathbf{z})$ is a positive definite matrix, where $\mathcal{Q}(\mathbf{z}) = \ln\Big(1 + \sum_{n\in[N]} G^n(\mathbf{Y}^n|z_n)\Big)$. To simplify the proof and make use of the results in Zhang et al. (2018), let us denote $\mathbf{u}(\mathbf{z}) : \mathbb{R}^N \to \mathbb{R}^N$ such that $u(\mathbf{z})_n = -\ln G^n(\mathbf{Y}^n|z_n)/\widetilde{b}_n$. with this definition we have

$$\frac{\partial u(\mathbf{z})_n}{\partial z_n} = \frac{\sum_{i\in\mathcal{V}_n} \partial G_i^n(\mathbf{Y}^n|z_n) Y_i \widetilde{b}_n}{G^n(\mathbf{Y}^n|z_n)\widetilde{b}_n} = 1.$$

The objective function now can be written as

$$\mathcal{Q}(\mathbf{u}(\mathbf{z})) = \ln\left(1 + \sum_{n\in[N]} \exp(-u(\mathbf{z})_n).\right)$$

Taking the derivative of $\mathcal{Q}$ with respect to z_n we obtain

$$\frac{\partial \mathcal{Q}(\mathbf{u}(\mathbf{z}))}{\partial z_n} = \frac{\partial \mathcal{Q}(\mathbf{u})}{\partial u_n}\bigg|_{\mathbf{u}=\mathbf{u}(\mathbf{z})} \frac{\partial u(\mathbf{z})_n}{\partial z_n} = \frac{\partial \mathcal{Q}(\mathbf{u})}{\partial u_n}\bigg|_{\mathbf{u}=\mathbf{u}(\mathbf{z})}.$$

And if we take the second derivative with respective to z_n, z_k, $n,k \in [N]$ we get

$$\frac{\partial^2 \mathcal{Q}(\mathbf{u}(\mathbf{z}))}{\partial z_n \partial z_k} = \frac{\partial^2 \mathcal{Q}(\mathbf{u})}{\partial u_n \partial u_k}\bigg|_{\mathbf{u}=\mathbf{u}(\mathbf{z})},$$

or equivalently $\nabla^2 \mathcal{Q}(\mathbf{z}) = \nabla^2_{\mathbf{u}} \mathcal{Q}(\mathbf{u})$, where $\mathbf{u} = \mathbf{u}(\mathbf{z})$. Moreover, $\mathcal{Q}(\mathbf{u})$ is just a special objective function under the MNL model with N products and all the PSP are equal to 1. As a result, $\nabla^2_{\mathbf{u}} \mathcal{Q}(\mathbf{u})$ is positive definite (see Theorem 4.1 Zhang et al. 2018), so $\nabla^2 \mathcal{Q}(\mathbf{z})$ is also positive definite, as desired.

Now we know that (40) is strictly convex, so it yields a unique solution. Moreover, one can show that (40) have finite optimal solutions. For any $n \in [N]$, taking the derivative of $\mathcal{Q}(\mathbf{z})$ with respect to z_n and set it to zero we obtain

$$\frac{\sum_{n\in[N]} \sum_{i\in\mathcal{V}_n} -\partial G_i^n(\mathbf{Y}^n|z_n) Y_i \widetilde{b}_n}{1 + G(\mathbf{Y})} = p_n^G \widetilde{b}_n,$$

or equivalently, $p_n^G = \Theta(\mathbf{z})$. So,if $\mathbf{z}(\mathbf{p}^G)$ is the unique solution to (40), we always have $\mathbf{p}^G = \Theta(\mathbf{z}^P)$ as desired. □

Next, we will show that the first part of $\mathcal{H}(\mathbf{z}, \underline{\mathbf{a}}, \overline{\mathbf{b}})$ is a concave function of $\mathbf{p}^G$. We also omit the choice parameters for notational convenience and denote $\mathcal{W}(\mathbf{z}) = \sum_{n\in[N]} z_n p_n^G$. We have the following lemma.

Lemma 12. *Function $\mathcal{W}(\boldsymbol{z}(\boldsymbol{p}^G))$ is concave in $\boldsymbol{p}^G$, for all $\boldsymbol{p}^G \in \Delta^N$.*

Proof: Similar to the previous proof, we also denote $u(\mathbf{z})_n = -\ln G^n(\mathbf{Y}^n|z_n)$. We also see that $\mathbf{p}^G$ is also a choice probability vector given by a MNL model with N products with the utility vector $-\mathbf{u}(\mathbf{z}(\mathbf{p}^G)) \circ \widetilde{\mathbf{b}}$. So, if we denote $\mathbf{u}'(\mathbf{p}^G)$ be a mapping from $\mathbb{R}^N$ to $\mathbb{R}^N$ such that $p_n^G = \exp(-\widetilde{b}_n u'(\mathbf{p}^G)_n) \Big/ \left(\sum_{n\in[N]} \exp(-\widetilde{b}_n u'(\mathbf{p}^G)_n)\right)$, then we have $\mathbf{u}(\mathbf{z}(\mathbf{p}^G)) = \mathbf{u}'(\mathbf{p}^G)$

$$
\begin{aligned}
\frac{\partial \mathcal{W}(\mathbf{z}(\mathbf{p}^G))}{\partial p_n^G} &= z(\mathbf{p}^G)_n + \sum_{j\in[n]} \frac{p_j^G \partial z(\mathbf{p}^G)_j}{\partial p_n^G} \\
&= z(\mathbf{p}^G)_n + \sum_{j\in[N]} p_j^G \frac{\partial z(\mathbf{p}^G)_j}{\partial u(\mathbf{z}(\mathbf{p}^G))_j} \frac{\partial u(\mathbf{z}(\mathbf{p}^G))_j}{\partial p_n^G} \\
&= z(\mathbf{p}^G)_n + \sum_{j\in[N]} p_j^G \frac{\partial u'(\mathbf{p}^G)_j}{\partial p_n^G}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2 \mathcal{W}(\mathbf{z}(\mathbf{p}^G))}{\partial p_n^G \partial p_k^G} &= \frac{\partial z(\mathbf{p}^G)_n}{\partial p_k^G} + \sum_{j\in[N]} p_j^G \frac{\partial^2 u'(\mathbf{p}^G)_j}{\partial p_n^G \partial p_k^G} \\
&= \frac{\partial z(\mathbf{p}^G)_n}{\partial u(\mathbf{z}(\mathbf{p}^G))_n} \frac{\partial u(\mathbf{z}(\mathbf{p}^G))_n}{\partial p_k^G} + \sum_{j\in[N]} p_j^G \frac{\partial^2 u'(\mathbf{p}^G)_j}{\partial p_n^G \partial p_k^G} \\
&= \frac{\partial u'(\mathbf{p}^G)_n}{\partial p_k^G} + \sum_{j\in[N]} p_j^G \frac{\partial^2 u'(\mathbf{p}^G)_j}{\partial p_n^G \partial p_k^G}
\end{aligned}
$$

Moreover, if we denote $\mathcal{W}'(\mathbf{p}^G) = \sum_{n\in[N]} u'(\mathbf{p}^G)_n p_n^G$, we also have

$$
\frac{\partial^2 \mathcal{W}'(\mathbf{p}^G)}{\partial p_n^G \partial p_k^G} = \frac{\partial u'(\mathbf{p}^G)_n}{\partial p_k^G} + \sum_{j\in[N]} p_j^G \frac{\partial^2 u'(\mathbf{p}^G)_j}{\partial p_n^G \partial p_k^G}, \ \forall n,k \in [N].
$$

So, $\nabla^2 \mathcal{W}(\mathbf{z}(\mathbf{p}^G)) = \nabla^2 \mathcal{W}'(\mathbf{p}^G)$. We also see that $\mathcal{W}'(\mathbf{p}^G)$ is the expected revenue function (as a function of the purchase probabilities $\mathbf{p}^G$) where there are N products, the choice model is MNL, the PSP are $\widetilde{\mathbf{b}}$ and the utility vector is $-u'(\mathbf{p}^G) \circ \widetilde{\mathbf{b}}$ ($\circ$ is the *dot product*). So we know that $\nabla^2 \mathcal{W}'(\mathbf{p}^G)$ is positive definite (Zhang et al. 2018), so $\mathcal{W}'(\mathbf{z}(\mathbf{p}^G))$ is strictly concave in $\mathbf{p}^G$. □

With all of the above results, we can formulate (27) as

$$
\begin{aligned}
\max_{\mathbf{p}^G, \mathbf{y}} \quad & \mathcal{W}(\mathbf{z}(\mathbf{p}^G)) - \sum_{t=1}^T \lambda_t y_t \\
\text{subject to} \quad & (\mathbf{d}^t)^{\mathrm{T}} \mathbf{p}^G - y_t \leq r_t \\
& \sum_{n\in[N]} p_n^G \leq 1 \\
& \mathbf{p}^G, \mathbf{y} \geq 0.
\end{aligned}
$$

which is a convex optimization problem, as $\mathcal{W}(\mathbf{z}(\mathbf{p}^G))$ is strictly concave.

Appendix E: Detailed numerical results

	Average				Max				Worst				Percentile rank of RO worst-case		
	DET	SA10	SA50	RO	DET	SA10	SA50	RO	DET	SA10	SA50	RO	DET	SA10	SA50
0.5	22.6	22.6	22.4	22.4	23.5	23.4	22.9	22.9	21.3	21.5	21.7	21.7	8	4	2
1.0	22.4	22.1	22.1	21.7	24.0	22.8	22.8	22.3	19.1	20.6	20.6	20.7	13	2	2
1.5	21.9	21.9	21.9	21.0	24.3	23.7	22.9	21.4	16.2	17.7	19.0	19.8	19	12	5
2.0	21.7	21.7	21.6	20.2	24.4	23.4	22.9	20.6	12.5	15.7	16.8	18.9	20	12	8
2.5	20.8	20.9	21.0	19.4	24.5	24.0	23.5	19.7	8.8	10.4	12.3	18.1	27	22	17
3.0	19.9	20.5	20.7	18.7	24.5	22.8	22.9	18.9	5.7	10.9	10.6	17.3	32	17	15
3.5	19.2	20.2	20.0	17.9	24.6	22.6	22.8	18.1	3.6	8.5	7.9	16.6	33	15	17
4.0	18.1	19.5	19.3	17.3	24.6	22.9	23.4	17.4	2.0	4.8	3.7	16.0	37	20	26
4.5	17.5	18.4	18.6	16.6	24.6	23.5	22.9	16.7	1.1	2.3	3.3	15.3	38	31	27
5.0	17.5	17.5	18.2	15.9	24.6	23.7	23.1	16.1	0.6	1.1	1.6	14.6	35	34	28
5.5	16.5	17.5	17.0	15.3	24.6	23.4	23.2	15.4	0.4	0.7	0.9	14.1	39	31	35
6.0	16.0	17.0	16.7	14.7	24.6	23.7	23.3	14.8	0.2	0.4	0.5	13.5	40	34	35
6.5	15.7	16.5	16.0	14.2	24.6	23.0	23.3	14.3	0.1	0.3	0.3	13.1	41	33	37
7.0	15.3	15.5	15.8	13.7	24.6	23.2	23.3	13.7	0.1	0.2	0.1	12.4	41	37	37
7.5	14.8	15.1	15.7	13.1	24.6	23.5	23.1	13.2	0.0	0.1	0.1	12.0	43	40	36
8.0	15.4	14.9	15.2	12.7	24.6	23.5	22.9	12.7	0.0	0.0	0.1	11.5	39	39	37
8.5	14.6	14.4	15.0	12.2	24.6	23.4	23.3	12.3	0.0	0.0	0.0	11.0	42	41	38
9.0	13.5	15.1	14.5	11.7	24.6	23.7	22.7	11.8	0.0	0.0	0.0	10.6	46	38	39
9.5	14.5	14.0	14.5	11.3	24.6	23.9	22.8	11.4	0.0	0.0	0.0	10.2	42	43	38
10.0	13.8	14.1	13.5	10.9	24.6	23.8	23.4	11.0	0.0	0.0	0.0	9.8	45	41	43
Average													34	27.3	26.1

Table 1 Comparison results for unconstrained robust (RO), deterministic (DET), and sampling-based (SA10 and SA50) pricing under the MNL logit model with homogeneous PSP.

	Average				Max				Worst				Percentile rank of RO worst-case		
	DET	SA10	SA50	RO	DET	SA10	SA50	RO	DET	SA10	SA50	RO	DET	SA10	SA50
0.5	23.5	23.3	23.3	23.3	24.5	23.8	23.8	23.8	22.2	22.6	22.6	22.6	18	4	4
1.0	23.2	23.0	22.7	22.8	24.9	23.8	23.2	23.3	20.0	21.4	21.6	21.6	31	5	2
1.5	22.7	22.6	22.5	21.9	25.2	23.6	23.2	22.3	16.7	20.0	20.3	20.6	40	8	6
2.0	22.3	22.3	22.0	21.1	25.3	23.5	22.9	21.5	12.8	18.0	18.9	19.7	44	14	9
2.5	21.3	21.8	21.6	20.4	25.4	23.3	22.5	20.7	9.0	15.7	17.4	18.9	60	19	11
3.0	20.2	21.6	21.3	19.5	25.5	23.6	22.6	19.7	5.7	11.6	14.5	18.1	75	26	15
3.5	19.9	20.8	20.8	18.8	25.5	23.2	23.3	19.0	3.4	9.5	8.9	17.4	66	29	32
4.0	18.8	20.1	20.2	18.1	25.5	23.6	22.7	18.2	2.0	5.6	7.8	16.6	75	46	32
4.5	18.2	19.6	19.5	17.3	25.5	23.7	22.9	17.5	1.1	3.1	4.5	16.0	74	48	41
5.0	17.3	19.3	19.0	16.7	25.5	22.7	22.9	16.8	0.6	3.2	2.8	15.3	79	39	45
5.5	16.8	18.5	18.2	16.1	25.5	22.9	23.1	16.2	0.3	1.7	1.6	14.7	81	46	52
6.0	16.2	17.5	17.7	15.5	25.5	24.0	23.1	15.6	0.2	0.5	0.9	14.2	81	66	58
6.5	16.5	16.9	16.9	14.9	25.5	23.2	23.5	15.0	0.1	0.4	0.4	13.6	77	65	69
7.0	16.1	16.6	16.7	14.3	25.5	23.3	22.8	14.4	0.0	0.2	0.4	13.1	80	69	65
7.5	15.5	16.2	16.4	13.8	25.5	22.8	23.3	13.9	0.0	0.2	0.1	12.5	83	66	66
8.0	15.0	15.6	15.8	13.2	25.5	24.1	23.5	13.3	0.0	0.0	0.1	12.1	87	75	70
8.5	14.9	15.2	15.3	12.7	25.5	23.3	23.1	12.8	0.0	0.0	0.0	11.6	86	76	75
9.0	14.9	15.2	15.2	12.4	25.5	24.1	23.4	12.4	0.0	0.0	0.0	11.2	85	77	76
9.5	15.4	14.6	14.7	11.9	25.5	23.9	22.9	12.0	0.0	0.0	0.0	10.8	83	81	76
10.0	14.8	14.4	14.7	11.5	25.5	22.8	22.5	11.5	0.0	0.0	0.0	10.3	85	77	72
Average													69.5	46.8	43.8

Table 2 Comparison results for unconstrained robust (RO), deterministic (DET), and sampling-based (SA10 and SA50) pricing under the nested logit model with homogeneous PSP.

	Average				Max				Worst				Percentile rank of RO worst-case		
	DET	SA10	SA50	RO	DET	SA10	SA50	RO	DET	SA10	SA50	RO	DET	SA10	SA50
0.2	38.5	38.5	38.5	38.2	39.8	39.7	39.6	38.9	36.7	36.9	37.0	37.2	9	6	4
0.4	38.3	38.2	38.2	37.3	40.5	39.6	39.7	38.0	33.6	35.1	35.0	35.6	10	3	4
0.6	38.1	38.1	38.0	36.1	40.9	40.0	39.5	36.7	29.7	32.7	33.3	34.3	18	6	5
0.8	37.5	37.7	37.3	34.8	41.2	40.4	38.8	35.3	25.9	27.2	31.3	33.2	27	15	5
1.0	37.2	37.2	36.9	33.6	41.3	38.9	38.5	33.9	18.4	26.3	28.7	31.9	30	8	8
1.2	36.7	36.7	36.6	32.3	41.5	39.2	38.5	32.5	14.1	23.9	24.0	30.8	30	13	8
1.4	36.0	36.1	35.6	31.0	41.5	38.2	37.1	31.2	10.9	22.5	21.9	29.6	34	11	8
1.6	35.2	35.3	35.3	29.8	41.5	36.7	36.6	29.9	8.2	21.6	22.4	28.6	40	7	5
1.8	35.4	35.2	34.7	28.6	41.5	37.0	36.0	28.7	7.4	16.7	16.9	27.3	37	8	5
2.0	34.9	34.7	34.4	27.4	41.6	36.5	35.8	27.5	3.8	14.0	14.7	26.5	38	8	8
2.2	34.4	34.7	33.8	26.3	41.6	38.4	35.0	26.4	2.7	7.6	17.4	25.5	38	23	5
2.4	34.4	34.5	33.8	25.2	41.6	36.7	35.2	25.3	1.1	7.8	14.4	24.6	35	11	8
2.6	33.7	33.4	33.7	24.1	41.6	34.7	35.2	24.2	1.0	10.2	9.5	23.5	40	6	7
2.8	33.8	33.5	33.5	23.1	41.6	35.6	35.5	23.2	0.7	6.5	7.4	22.6	36	10	9
3.0	33.4	31.7	33.3	22.1	41.6	32.5	35.3	22.2	0.6	15.2	5.8	21.6	37	4	10
3.2	33.7	34.0	33.0	21.2	41.6	36.7	34.5	21.3	0.3	3.0	7.3	20.4	35	12	6
3.4	34.1	33.1	30.8	20.2	41.6	35.0	31.6	20.3	0.3	4.5	16.0	19.5	32	8	3
3.6	33.7	32.0	31.7	19.3	41.6	33.2	32.8	19.5	0.2	1.5	10.4	18.6	33	6	5
3.8	33.0	32.6	32.0	18.4	41.6	42.6	33.4	18.6	0.1	0.1	5.3	17.8	36	40	5
4.0	34.2	31.8	31.4	17.6	41.6	32.8	32.4	17.8	0.0	7.1	6.3	17.0	29	4	4
4.2	33.6	32.8	30.8	16.8	41.6	35.1	31.8	16.9	0.1	0.9	4.8	16.1	30	9	4
4.4	34.0	29.2	30.6	16.0	41.6	29.7	31.6	16.2	0.0	5.8	4.2	15.4	29	3	4
4.6	34.4	33.6	30.7	15.3	41.6	36.9	31.7	15.4	0.1	0.1	1.9	14.6	26	12	4
4.8	34.0	33.8	31.0	14.5	41.6	41.6	32.1	14.7	0.0	0.0	4.6	13.9	28	31	4
5.0	33.9	32.3	29.6	13.8	41.6	34.3	30.3	14.0	0.0	1.5	4.0	13.2	27	8	3
Acerage													30.6	10.9	5.6

Table 3 Comparison results for unconstrained robust (RO), deterministic (DET), and sampling-based (SA10 and SA50) pricing under a nested logit model with partition-wise homogeneous PSP.

		Average				Max				Worst				Percentile rank of RO worst-case		
λ	ϵ	DET-CON	DET-PEN	SA10	RO	DET-CON	DET-PEN	SA10	RO	DET-CON	DET-PEN	SA10	RO	DET-CON	DET-PEN	SA10
0.2	0.5	19.9	20.9	20.7	20.1	21.4	21.9	21.3	20.4	17.4	18.7	19.2	19.3	47	3	2
	1.0	19.9	20.9	20.2	18.7	22.8	22.4	20.8	18.9	13.6	16.0	17.9	17.9	21	6	2
	1.5	19.9	20.9	21.0	17.1	23.6	22.5	22.5	17.2	8.5	12.7	12.3	16.3	17	6	6
	2.0	19.9	20.9	20.5	15.6	24.4	22.6	21.5	15.7	5.2	8.0	10.9	15.0	16	8	4
	2.5	20.0	20.8	20.3	14.2	24.9	22.6	21.3	14.2	2.9	5.8	5.0	13.5	14	8	4
	3.0	20.0	20.7	19.7	12.8	25.8	22.6	20.5	12.9	1.5	3.2	6.4	12.1	12	8	3
0.4	0.5	19.8	20.0	19.9	19.2	21.0	21.0	21.0	19.5	17.4	17.8	17.9	18.4	10	3	4
	1.0	19.8	20.0	19.6	17.7	21.6	21.5	20.4	17.9	14.1	13.8	16.4	16.7	8	4	2
	1.5	19.7	20.0	19.1	16.2	22.2	21.6	20.1	16.3	8.6	10.5	13.7	15.3	10	6	3
	2.0	19.6	19.9	19.9	14.6	22.6	21.6	21.6	14.8	6.3	6.8	8.2	13.5	10	6	6
	2.5	19.5	19.8	18.1	13.3	22.9	21.7	18.8	13.4	2.3	4.9	9.0	11.9	9	7	3
	3.0	19.6	19.8	19.7	11.9	23.1	21.7	21.7	12.0	1.5	2.2	1.2	10.3	9	7	7
0.6	0.5	19.7	19.7	19.6	19.0	20.9	20.9	20.5	19.4	17.4	17.3	17.5	18.1	6	5	3
	1.0	19.6	19.6	19.6	17.3	21.3	21.3	21.3	17.8	14.0	13.8	14.7	16.1	5	5	4
	1.5	19.5	19.5	19.5	15.9	21.4	21.4	21.4	16.2	9.6	10.2	9.1	14.6	7	7	7
	2.0	19.4	19.3	18.6	14.2	21.5	21.5	19.9	14.7	6.9	4.5	9.3	12.9	8	9	4
	2.5	19.3	19.3	18.6	12.7	21.5	21.5	20.2	13.3	4.2	2.9	5.6	11.4	8	8	5
	3.0	19.1	19.0	18.8	10.5	21.5	21.5	21.5	11.9	1.5	2.3	2.2	3.2	2	2	5
0.8	0.5	19.6	19.6	19.6	18.8	20.8	20.8	20.8	19.4	17.3	17.4	17.2	17.8	3	3	3
	1.0	19.5	19.5	19.5	17.2	21.3	21.3	21.3	17.8	13.7	13.6	12.7	15.6	4	3	3
	1.5	19.3	19.3	19.3	15.5	21.5	21.4	21.5	16.2	9.6	9.1	8.9	13.6	5	4	4
	2.0	19.0	19.1	17.1	13.7	21.5	21.5	18.1	14.7	6.0	6.0	13.2	13.3	5	4	5
	2.5	18.8	18.9	18.9	11.4	21.5	21.5	21.5	13.2	3.4	3.5	2.4	4.1	2	2	6
	3.0	18.7	18.7	18.7	10.6	21.5	21.5	21.5	12.0	1.9	2.5	2.0	7.4	5	5	5
Average														10.1	5.4	4.2

Table 4 Comparison results for deterministic constrained pricing (DET-CON), deterministic pricing with over-expected-sale penalties (DET-PEN), sampling-based pricing with over-expected-sale penalties (SA10), and robust pricing with over-expected-sale penalties (RO).