DATA-DRIVEN DISTRIBUTIONALLY ROBUST MULTIPRODUCT PRICING PROBLEMS UNDER PURE CHARACTERISTICS DEMAND MODELS*

JIE JIANG[†], HAILIN SUN[‡], AND XIAOJUN CHEN[§]

Abstract. This paper considers a multiproduct pricing problem under pure characteristics demand models when the probability distribution of the random parameter in the problem is uncertain. We formulate this problem as a distributionally robust optimization (DRO) problem based on a constructive approach to estimating pure characteristics demand models with pricing by Pang, Su and Lee. In this model, the consumers' purchase decision is to maximize their utility. We show that the DRO problem is well-defined, and the objective function is upper semicontinuous by using an equivalent hierarchical form. We also use the data-driven approach to analyze the DRO problem when the ambiguity set, i.e., a set of probability distributions that contains some exact information of the underlying probability distribution, is given by a general moment-based case. We give convergence results as the data size tends to infinity and analyze the quantitative statistical robustness in view of the possible contamination of driven data. Furthermore, we use the Lagrange duality to reformulate the DRO problem as a mathematical program with complementarity constraints, and give a numerical procedure for finding a global solution of the DRO problem under certain specific settings. Finally, we report numerical results that validate the effectiveness and scalability of our approach for the distributionally robust multiproduct pricing problem.

Key words. pure characteristics demand model, stochastic optimization, distributional robustness, data-driven, mathematical program with complementarity constraints

MSC codes. 90C15, 90C33

1. Introduction.

The utility theory has been widely adopted to describe the behavior of individual consumers in economics and finance, since the seminal work on games and economic behavior by Von Neumann and Morgenstern [36]. In a pure characteristics demand model, utility functions of consumers are functions of product characteristics including the price, which are used to obtain the market share equations [3]. Such utility functions are discontinuous and lead to computationally intractable estimation of the demand model. To overcome the computational difficulty, in [29], Pang et al. gave a novel and constructive reformulation, in which the consumers' purchase decision problems were formulated by a system of linear complementarity constraints. Such formulation allows us to estimate the consumers' pure characteristics demand model by a quadratic program with linear complementarity constraints, which is numerically tractable by using some existing methodology [33]. Motivated by the work in [29], Chen et al. considered in [4] a regularized sample average approximation (SAA) of a class of optimization problems involving set-valued stochastic equilibrium constraints that includes the estimation problem with exogenous price proposed in [29], and established graphical convergence results. Recently, Jiang

*Submitted to the editors DATE.

Funding: This work is supported by The Hong Kong Polytechnic University Post-doctoral Fellow Scheme, The Hong Kong Grant Council Grant (PolyU15300322 and N_PolyU507/22) and the National Natural Science Foundation of China (12201084, 12122108 and 12261160365).

[†]College of Mathematics and Statistics, Chongqing University, Chongqing, China (jiangjiecq@163.com).

[‡]Key Laboratory of NSLSCS, Ministry of Education, Jiangsu International Joint Laboratory of BDMCA, School of Mathematical Sciences, Nanjing Normal University, Nanjing, China (hlsun@njnu.edu.cn).

[§]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China (maxjchen@polyu.edu.hk).

and Chen employed the distributionally robust approach to estimate the parameters in a pure characteristics demand model with the fixed price when the probability distribution is uncertain in [21]. It is worth pointing out that the aforementioned works [4, 21, 29] estimated the parameters in utility functions of pure characteristics demand models when the characteristics of products are given.

The price is an important factor for consumers when they determine their purchase decisions. When the parameters in the pure characteristics demand model are known, multiproduct pricing models are established based on the pure characteristics demand model and the observed product characteristics to obtain the optimal prices in [29, 34]. It is noteworthy that a set of finite numbers of random samples was used in [29], while continuous random variables and a regularized SAA approach were employed in [34] under the assumption that the true probability distribution of random parameters in the model is known. However, in practical applications, the true probability distribution cannot be detected exactly. In this paper, we consider the multipruduct pricing problem when the true probability distribution of the consumers' preference random parameter is unknown. We will apply the distributionally robust optimization (DRO) approach (see, e.g., [7, 10, 28]) to deal with the unknown information by accessing a set of probability distributions that includes the true one.

To present our DRO approach, we first introduce some basic settings. Consider a market with $T\,(T > 1)$ firms and $m\,(m > 1)$ products indexed by $t = 1, \dots, T$ and $j = 1, \dots, m$ respectively, where each product can only been produced by one firm. The target firm is the first firm which produces products $1, \dots, K$ with $K < m$. We assume that the target firm will produce product i rather than product j for any $1 \le i < j \le K$ when products i and j have the same net profit. Namely, these products are indexed in rank order according to the firm's individual preference. Each product j is characterized by a vector of observed characteristics $x_j \in \mathbb{R}^\ell$ and price $p_j > 0$. Suppose that the consumers in the market are heterogenous. The $\mathbb{R}^s$-valued random vector ξ with support set being $\Xi \subseteq \mathbb{R}^s$ is used to estimate heterogeneous consumers' preferences or tastes over the observed product characteristics and price in the differentiated product setting.

For fixed product characteristics, we use $u_j(p_j, \xi)$ to denote a consumer's utility with preference ξ purchasing product j at price p_j for $j = 1, \dots, K$. In [29], the utility for a consumer purchasing product j with preference ξ is given by

$$u_j(p_j, \xi) = \beta_j(\xi)^\top x_j - \alpha_j(\xi) p_j + \eta_j(\xi), \qquad j = 1, \dots, K, \tag{1.1}$$

where $\beta_j(\xi) \in \mathbb{R}^\ell_+$ and $\alpha_j(\xi) \in \mathbb{R}_+$ model the consumer's preference regarding the observed product j's characteristics x_j and price p_j, respectively, and $\eta_j(\xi) \in \mathbb{R}$ is the product characteristic or demand shock that is observed by the firms and consumers but is not available in the data. We use $u_j(\xi)$ to denote a consumer's utility with preference ξ purchasing product j at fixed price p_j for $j = K+1, \dots, m$. Let $\mathcal{P}$ be a convex and compact set in $\mathbb{R}^K_{++}$. We assume that the utility function $u : \mathcal{P} \times \Xi \to \mathbb{R}^m$ with

$$u(p, \xi) := (u_1(p_1, \xi), \dots, u_K(p_K, \xi), u_{K+1}(\xi), \dots, u_m(\xi))^\top$$

is continuous with respect to (w.r.t.) the tuple (p, ξ).

To estimate the consumer's purchasing strategies with preference ξ, Pang et al. [29, (7)] proposed to maximize the consumer's utility with preference ξ by the following maximization problem

$$\begin{aligned} \max_y \quad & y^\top u(p, \xi) \\ \text{s.t.} \quad & e^\top y \le 1, \ y \ge 0, \end{aligned} \tag{1.2}$$

where y is an m-dimensional decision variable with the ith ($1 \leq i \leq m$) component denoting the purchase weight of product i and $e \in \mathbb{R}^m$ is a vector with each element being one. The KKT condition of the linear program (1.2) is necessary and sufficient for the optimality, that is, y^* is a solution of (1.2) if and only if there is $\gamma^* \in \mathbb{R}_+$ such that

$$0 \leq \begin{pmatrix} y^* \\ \gamma^* \end{pmatrix} \perp \begin{pmatrix} 0 & e \\ -e^\top & 0 \end{pmatrix} \begin{pmatrix} y^* \\ \gamma^* \end{pmatrix} + \begin{pmatrix} -u(p,\xi) \\ 1 \end{pmatrix} \geq 0.$$

Pang et al. in [29] formulated the target firm's pricing problem as a mathematical program with linear complementarity constraints (see monographs [6, 9, 27]):

$$\begin{aligned} \max_{p \in \mathcal{P}} \quad & \mathbb{E}\left[y_{[K]}(\xi)^\top (p-c)\right] \\ \text{s.t.} \quad & 0 \leq \begin{pmatrix} y(\xi) \\ \gamma(\xi) \end{pmatrix} \perp \begin{pmatrix} 0 & e \\ -e^\top & 0 \end{pmatrix} \begin{pmatrix} y(\xi) \\ \gamma(\xi) \end{pmatrix} + \begin{pmatrix} -u(p,\xi) \\ 1 \end{pmatrix} \geq 0, \end{aligned}$$

where $c \in \mathbb{R}^K_+$ is a vector whose entry c_j denotes the marginal cost of product j for $j = 1, \ldots, K$, $y_{[K]}(\xi)$ is a K-dimensional vector consisting of the first K components of $y(\xi)$ such that the objective function is well-defined.

For fixed (p, ξ), let $\mathcal{S}(p,\xi)$ be the optimal solution set of problem (1.2). The target firm's pricing problem can be equivalently written as follows (see [29, (23)] and [34, (2) and (4)]):

$$\begin{aligned} \max_{p \in \mathcal{P}} \quad & \mathbb{E}\left[y_{[K]}(\xi)^\top (p-c)\right] \\ \text{s.t.} \quad & y(\xi) \in \mathcal{S}(p,\xi), \end{aligned} \tag{1.3}$$

where $y(\xi)$ is a measurable selection selected from $S(p,\xi)$ that makes the objective function $\mathbb{E}\left[y_{[K]}(\xi)^\top (p-c)\right]$ achieve a maximum. $\mathcal{S}(p,\xi)$ is generally set-valued and we cannot find a continuous single-valued function $y(p,\xi) \in \mathcal{S}(p,\xi)$ w.r.t. p for almost every ξ. Consider a simple example as in [4]: $u(p,\xi) := (\xi_1 - p, \xi_2) \in \mathbb{R}^2$, where $\xi = (\xi_1, \xi_2)^\top$ with $\xi_1 \in \mathbb{R}$ and $\xi_2 > 0$. Then the solution set has the form:

$$\mathcal{S}(p,\xi) = \begin{cases} (1,0)^\top, & p < \xi_1 - \xi_2; \\ \left\{(\alpha, 1-\alpha)^\top : \alpha \in [0,1]\right\}, & p = \xi_1 - \xi_2; \\ (0,1)^\top, & p > \xi_1 - \xi_2, \end{cases}$$

and we can not find a continuous single-valued function $y(p,\xi) \in S(p,\xi)$ w.r.t. p. The standard optimization method and SAA scheme in the literature become intractable for solving problem (1.3).

We consider the following extended multiproduct pricing problem as a two-stage stochastic optimization problem:

$$\max_{p \in \mathcal{P}} \mathbb{E}\left[Q(p,\xi)\right], \tag{1.4}$$

where $Q(p,\xi) := H(p,\xi) - h(p,\xi)$, and $H(p,\xi)$ is the second stage optimal value function, i.e.,

$$\begin{aligned} H(p,\xi) := \quad \max_{y(\xi)} \quad & g\left(y_{[K]}(\xi)^\top (p-c)\right) \\ \text{s.t.} \quad & y(\xi) \in \mathcal{S}(p,\xi). \end{aligned} \tag{1.5}$$

Here $g : \mathbb{R} \to \mathbb{R}$ is a strictly increasing and continuous function, which can be viewed as a utility function of the profit, and $h : \mathbb{R}^K \times \Xi \to \mathbb{R}_+$ is continuous w.r.t. p for almost every $\xi \in \Xi$ and measurable w.r.t. ξ for all $p \in \mathbb{R}^K$. This term $h(p,\xi)$ can be viewed as a regularization term or a budget term, which is used to ensure some additional properties of the pricing model, such as boundedness, sparsity, etc. When $h(p,\xi) \equiv 0$ and $g\left(y_{[K]}(\xi)^\top (p-c)\right) = y_{[K]}(\xi)^\top (p-c)$, problem (1.4) is equivalent to problem (1.3). Also, from the viewpoint of two-stage stochastic optimization, the term $-\mathbb{E}[h(p,\xi)]$ can be viewed as a first stage profit. When $\mathcal{S}(p,\xi)$ is not a singleton, problem (1.5) tacitly assumes that the firm will take the best selection of a vector from $\mathcal{S}(p,\xi)$ to achieve its goal. In fact, such selection determines an optimistic attitude of the firm. Therefore, it can be viewed as an optimistic version. Correspondingly, the pessimistic type can be defined.

In practice, it is usually argued that the true probability distribution of ξ in (1.4) cannot be captured exactly. To obtain the true probability distribution, it requires that the size of the empirical data tends to infinity, which is usually impracticable and costly. In most real applications, only limiting finite empirical data (i.e., partial information) are available. DRO is a popular approach to settle this dilemma (see [7, 28]). In view of this, we further consider the distributionally robust counterpart of the extended multiproduct pricing problem (1.4) as follows:

$$\text{(P)} \qquad \max_{p\in\mathcal{P}} \inf_{F\in\mathcal{F}} \mathbb{E}_F\left[Q(p,\xi)\right],$$

where $\mathcal{F}$ is the ambiguity set.

The main contributions of this paper are summarized as follows.

- We establish interesting properties of the extended multiproduct pricing problem (1.4) and its distributionally robust counterpart (P) in a hierarchical form on the measurability and semicontinuity of the second stage optimal value function with a closed form sparse solution. We prove the existence of solutions of the discontinuous and nonconvex optimization problems (1.4) and (P).
- Problem (P) is analyzed from a data-driven viewpoint when the ambiguity set is given by a general moment-based form. We derive convergence results when the data-driven moment information converges almost surely to the true one as data size tends to infinity. It is worth pointing out that our data-driven analysis differs from the existing ones [7, 28] regarding the ambiguity sets. Additionally, we give a quantitative statistical robustness assertion under moderate conditions when the data-driven moment information is contaminated. The data-driven analysis ensures that the data-driven model is reliable when the data size is sufficiently large or even if the data are contaminated slightly.
- We reformulate problem (P) with a general moment ambiguity set as a mathematical program with complementarity constraints (MPCC) by using the Lagrange duality. We propose a numerical procedure to find a global solution for problem (P) with finite elements in Ξ. This procedure is based on the MPCC reformulation and the closed-form expression of the second stage optimal value function. We report some numerical results using this procedure, which preliminarily illustrate the necessariness of the distributionally robust approach and data-driven analysis for multiproduct pricing problems.

The reminder of the paper is organized as follows. In Section 2, we present some useful properties, including measurability, semicontinuity, etc. In Section 3, the data-driven analysis is studied. In Section 4, the equivalent MPCC reformulation of problem (P) is discussed. In Section 5, numerical procedures are given and some numerical results are reported. Finally, we give concluding remarks in Section 6.

Notations. For some integer $n \geq 1$, $\mathbb{R}^n_+$ denotes the nonnegative part of $\mathbb{R}^n$, and $\mathbb{R}^n_{++}$ denotes the set of positive vectors (in the componentwise sense) in $\mathbb{R}^n$. $\|\cdot\|$ and $\|\cdot\|_\infty$ denote the Euclidean norm and the infinity norm, respectively. $(\cdot)_+ := \max\{0, \cdot\}$. For $x \in \mathbb{R}^n$ and $X, Y \subseteq \mathbb{R}^n$, $\mathrm{d}(x, Y) := \inf_{y \in Y} \|x - y\|$ and $\mathrm{d}(X, Y) := \sup_{x \in X} \inf_{y \in Y} \|x - y\|$. We use $\mathbb{D}$ with some subscripts to denote probability metrics, such as $\mathbb{D}_{\mathcal{G}}(\cdot, \cdot)$ denotes the ζ-structure probability metric induced by a set of measurable functions $\mathcal{G}$, $\mathbb{D}_{TV}(\cdot, \cdot)$ denotes the total variational metric, $\mathbb{D}_W(\cdot, \cdot)$ denotes the Kantorovich metric, etc. $\mathbb{B}$ denotes the closed unit ball in the corresponding space.

2. Properties.

In this section, we will explore several useful properties of our models. Specifically, we will investigate the semicontinuity of the second stage optimal value function $H(p, \xi)$, as well as the existence of solutions for problem (1.4) and problem (P). We first establish the measurability of these problems. To this end, we first recall some concepts, which can be found in [31, Definitions 14.1 and 14.27]. Let $(\Xi, \mathcal{A})$ be a measurable space with Ξ being the nonempty support set of ξ and $\mathcal{A}$ being some σ-field of subsets of Ξ. A mapping $\varphi : \Xi \to \mathbb{R}^n$ is measurable if for every open set $O \subseteq \mathbb{R}^n$ the set $\varphi^{-1}(O) := \{\xi \in \Xi : \varphi(\xi) \in O\} \in \mathcal{A}$. A set-valued mapping $S : \Xi \rightrightarrows \mathbb{R}^n$ is measurable if for every open set $O \subseteq \mathbb{R}^n$ the set $S^{-1}(O) := \{\xi \in \Xi : S(\xi) \cap O \neq \emptyset\} \in \mathcal{A}$. A function $f : \mathbb{R}^n \times \Xi \to \overline{\mathbb{R}} := \{\mathbb{R} \cup \{\pm\infty\}\}$ is called a normal integrand if its epigraphical mapping $S_f : \Xi \rightrightarrows \mathbb{R}^n \times \mathbb{R}$, i.e. $S_f(\xi) := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : f(x, \xi) \leq \alpha\}$, is closed-valued and measurable.

PROPOSITION 2.1. *For any fixed $p \in \mathcal{P}$, the optimal solution set $\mathcal{S}(p, \cdot)$ of problem (1.2) is closed-valued and measurable.*

Proof. Consider $Y := \{y \in \mathbb{R}^m_+ : e^\top y \leq 1\}$ and $\ell(y, \xi) := -y^\top u(p, \xi) + \delta_Y(y)$, where $\delta_Y(\cdot)$ is the indicator function regarding to Y, i.e., $\delta_Y(y) = 0$ for $y \in Y$ and $\delta_Y(y) = +\infty$ otherwise. Then we have

$$\mathcal{S}(p, \xi) = \arg\min_{y} \ell(y, \xi).$$

Since Y is a closed set, it is not difficult to verify that $\delta_Y(y)$ is lower semicontinuous (lsc) (see [31, Definition 1.5]) on $\mathbb{R}^m$. Due to the continuity of u, we know that $\ell(y, \xi)$ is lsc w.r.t. (y, ξ). Then, we have that $\mathcal{S}(p, \cdot)$ is closed-valued. Further, we know from [31, Example 14.31] that $\ell(y, \xi)$ is a normal integrand. Finally, based on [31, Theorem 14.37], we have that $\mathcal{S}(p, \cdot)$ is measurable. ◻

PROPOSITION 2.2. *For any fixed $p \in \mathcal{P}$, $Q(p, \cdot)$ in problem (1.4) is finite and measurable.*

Proof. Due to the nonemptiness and boundedness of $\mathcal{S}(p, \xi)$, $Q(p, \cdot)$ is finite obviously. In what follows, we focus on the measurability of $Q(p, \cdot)$.

Consider problem (1.5). Since g is continuous and strictly increasing, we know from [31, Example 14.51] and Proposition 2.1 that $H(p, \cdot)$ is measurable. Moreover, since h is continuous, we have that $Q(p, \xi) = H(p, \xi) - h(p, \xi)$ is also measurable. ◻

For given p, denote the inner infimum of problem (P) by $\vartheta(p)$, i.e.,

$$\vartheta(p) := \inf_{F \in \mathcal{F}} \mathbb{E}_F [Q(p, \xi)] \tag{2.1}$$

and for given p and ξ, denote the index set

$$\mathcal{I}(p, \xi) := \left\{ s : u_s(p_s, \xi) = \|(u(p, \xi))_+\|_\infty, s \in \{1, \ldots, K\} \right\}.$$

To investigate the semicontinuity of Q, we need the following concept named the sparse solution.

DEFINITION 2.3 (the sparse solution, [34, Definition 2]). *For given $p \in \mathcal{P}$ and $\xi \in \Xi$, the sparse solution of problem (1.5) denoted by $y(p,\xi)$, is defined as*

(i) if $\mathcal{I}(p,\xi) \neq \emptyset$, then $y_s(p,\xi) = 1$ and $y_i(p,\xi) = 0$ for $i = 1,\ldots,m$ and $i \neq s$, where $s := \min\{j : (p-c)_j = \max_{i\in\mathcal{I}(p,\xi)}(p-c)_i\}$;

(ii) if $\mathcal{I}(p,\xi) = \emptyset$ and $\|(u(p,\xi))_+\|_\infty > 0$, then $y_s(p,\xi) = 1$ and $y_i(p,\xi) = 0$ for $i = 1,\ldots,m$ and $i \neq s$, where $s := \min\{j : u_j(p_j,\xi) = \|u(p,\xi)\|_\infty\}$;

(iii) if $\mathcal{I}(p,\xi) = \emptyset$ and $\|(u(p,\xi))_+\|_\infty = 0$, then $y(p,\xi) = 0$.

Based on Definition 2.3, we know that for any given $p \in \mathcal{P}$ and $\xi \in \Xi$, there always exists a unique corresponding sparse solution $y(p,\xi)$. To facilitate understanding of the sparse solution, we provide the following example.

Example 2.4. Assume that there are three products in the market, indexed by $1,2,3$, two firms with the target firm producing the products 1 and 2 and the rival firm producing product 3, two kinds of consumers' tastes, i.e., $\Xi = \{\xi_1,\xi_2\}$. Let $c = (0.5,2.5)^\top$ and $\mathcal{P} = [1,3]\times[2,4]$. Further, let $u_1(p_1,\xi_1) = 3-p_1$, $u_1(p_1,\xi_2) = 6-2p_1$, $u_2(p_2,\xi_1) = 3-2p_2$, $u_2(p_2,\xi_2) = 7-p_2$, $u_3(\xi_1) = 3$ and $u_3(\xi_2) = 2$. Now consider the sparse solution for $p = (1,3)^\top \in \mathcal{P}$ and $\xi = \xi_1,\xi_2$.

As for consumers with taste ξ_1, we have

$$u_1(p_1,\xi_1) = 2 < 3 = u_3(\xi_1) \text{ and } u_2(p_2,\xi_1) = -3 < 3 = u_3(\xi_1),$$

which implies that the consumers with taste ξ_1 would prefer to product 3.

As for consumers with taste ξ_2, we have

$$u_1(p_1,\xi_2) = u_2(p_2,\xi_2) = 4 > 2 = u_3(\xi_2).$$

Based on Definition 2.3, we have that the sparse solutions for $p = (1,3)^\top$ and $\xi = \xi_1,\xi_2$ are $y(p,\xi_1) = (0,0,1)^\top$ and $y(p,\xi_2) = (1,0,0)^\top$, respectively.

Note that products $1,2,3$ are indexed in rank order according to the target firm's individual preference. The sparse solution implies not only the preference of consumers, but also the preference of the target firm. That is, both the target firm and consumers with taste ξ_2 would like to choose the sparse solution $y(p,\xi_2) = (1,0,0)^\top$.

With the aid of the sparse solution, we can give the closed-form expression of H.

PROPOSITION 2.5. *For given $p \in \mathcal{P}$ and $\xi \in \Xi$, $H(p,\xi) = g\left(y_{[K]}(p,\xi)^\top(p-c)\right)$, where $y_{[K]}(p,\xi)$ is the first K components of the sparse solution $y(p,\xi)$.*

Proof. We give the proof by considering the following two cases.

Case 1: $\mathcal{I}(p,\xi) \neq \emptyset$. In this case, there exists some $i \in \{1,\ldots,K\}$ such that $u_i(p_i,\xi) = \|(u(p,\xi))_+\|_\infty$. Let $y(p,\xi)$ be the sparse solution and s be the smallest index such that $(p-c)_s = \max_{i\in\mathcal{I}(p,\xi)}(p-c)_i$. Then $s \in \{1,\ldots,K\}$, $y_s(p,\xi) = 1$ and $y_i(p,\xi) = 0$ for all $i \neq s$. Obviously, $y(p,\xi) \in \mathcal{S}(p,\xi)$ with

$$\mathcal{S}(p,\xi) = \left\{y : e^\top y \le 1, y \ge 0, \text{ and } y_i = 0 \text{ if } u_i(p_i,\xi) < \|(u(p,\xi))_+\|_\infty\right\}.$$

Since $(p-c)_s$ is one of the largest component of $p-c$, $y_{[K]}(p,\xi)^\top(p-c) \ge \bar{y}_{[K]}^\top(p-c)$ for all $\bar{y} \in \mathcal{S}(p,\xi)$, where $\bar{y}_{[K]}$ is the first K components of $\bar{y}$. Due to the monotonicity of g, we have $g(y_{[K]}(p,\xi)^\top(p-c)) \ge g(\bar{y}_{[K]}^\top(p-c))$, which verifies that $H(p,\xi) = g(y_{[K]}(p,\xi)^\top(p-c))$.

Case 2: $\mathcal{I}(p,\xi)=\emptyset$. In this case, by the definition of $\mathcal{S}(p,\xi)$ and $y(p,\xi)$, for all $y\in\mathcal{S}(p,\xi)$, $y_t=0, t=1,\dots,K$ and thus $H(p,\xi)=g(0)=g(y_{[K]}(p,\xi)^\top(p-c))$.

By summarizing the above two cases, the proof is complete. ∎

In general, H is not continuous. To see this, we give a simple example as follows.

Example 2.6. Assume that there are products $1,2$ in the market. The target firm produces product 1 and the rival produces product 2. Let $g(t)=t$, $h\equiv 0$, $u_1(p,\xi)=\xi_1-\xi_2 p$ and $u_2(\xi)=\xi_3$, where $\xi_i\sim U(0,1)$ for $i=1,2,3$ are independent with each other. Let $\xi=(\xi_1,\xi_2,\xi_3)^\top$. In this case, we have

$$H(p,\xi)=\begin{cases}0, & \xi_1-\xi_2 p<\xi_3,\\ p-c, & \xi_1-\xi_2 p\ge\xi_3,\end{cases}$$

which is discontinuous w.r.t. p for given ξ in general.

Despite the discontinuity of $H(\cdot,\xi)$, we have the following upper semicontinuity property.

PROPOSITION 2.7. *For fixed $\xi\in\Xi$, $H(\cdot,\xi)$ is upper semicontinuous over $\mathcal{P}$, i.e.,*

$$\limsup_{p'\to p} H(p',\xi)\le H(p,\xi) \tag{2.2}$$

for any $p\in\mathcal{P}$. Moreover, $\vartheta(\cdot)$, defined in (2.1)*, is also upper semicontinuous.*

Proof. We prove the upper semicontinuity of $H(\cdot,\xi)$ by considering two cases.

Case 1: $\mathcal{I}(p,\xi)\neq\emptyset$. Based on the definition of sparse solution $y(p,\xi)$, we know that there exists an $s\in\{1,\dots,K\}$ such that the sth component of $y(p,\xi)$ equals to 1, i.e., $y_s(p,\xi)=1$. Moreover, for any index $i\in\{1,\dots,K\}$, we have one of the following three cases holds:

(1) $u_i(p_i,\xi)=u_s(p_s,\xi)$ and $(p-c)_s>(p-c)_i$ for $i\neq s$;

(2) $u_i(p_i,\xi)<u_s(p_s,\xi)$ for $i\neq s$;

(3) $u_i(p_i,\xi)=u_s(p_s,\xi)$ and $(p-c)_s=(p-c)_i$ for $i\ge s$.

We use notations $\mathcal{I}_1$, $\mathcal{I}_2$ and $\mathcal{I}_3$ to represent the sets of indexes satisfying above three cases, respectively. Obviously, we have

$$\cup_{i=1}^3\mathcal{I}_i=\{1,\dots,K\} \text{ and } \mathcal{I}_k\cap\mathcal{I}_j=\emptyset \text{ for } k\neq j \text{ and } k,j=1,2,3.$$

Consider $p':=(p'_1,\dots,p'_K)^\top\in\mathbb{R}^K$ that is sufficiently closed to p.

For $i\in\mathcal{I}_1$, there are two possible cases: (1a) $u_i(p'_i,\xi)=\|(u(p',\xi))_+\|_\infty$; (1b) $u_i(p'_i,\xi)<\|(u(p',\xi))_+\|_\infty$. If case (1a) holds, we know from $(p-c)_s>(p-c)_i$ that $y_i(p',\xi)(p'-c)_i=0$ or $0<y_i(p',\xi)(p'-c)_i\le y_s(p,\xi)(p-c)_s$; if case (1b) holds, we have $y_i(p',\xi)=0$ and thus $y_i(p',\xi)(p'-c)_i=0$.

For $i\in\mathcal{I}_2$, we know from the continuity of $u(\cdot,\cdot)$ that $u_i(p'_i,\xi)<u_s(p'_s,\xi)$, and then $y_i(p',\xi)=0$. Thus, $y_i(p',\xi)(p'-c)_i=0$ for $i\in\mathcal{I}_2$.

For $i\in\mathcal{I}_3$ and any sequence $\{p^k\}_{k\ge1}$ with $p^k\to p$ as $k\to\infty$, we have that either $y_i(p^k,\xi)=0$ (and thus $y_i(p^k,\xi)(p'-c)_i=0$) or $y_i(p^k,\xi)(p^k-c)_i\to y_s(p,\xi)(p-c)_s$ as $k\to\infty$.

To summarize the above three cases, we obtain that

$$\begin{aligned}\limsup_{p'\to p} H(p',\xi) &= \limsup_{p'\to p}\left(g\left(y_{[K]}(p',\xi)^\top(p'-c)\right)\right)=\lim_{k\to\infty} g\left(y_{[K]}(p^k,\xi)^\top(p^k-c)\right)\\ &=\lim_{k\to\infty} g\left(y_{s^k}(p^k,\xi)(p^k-c)_{s^k}\right)\le g\left(y_{[K]}(p,\xi)^\top(p-c)\right)=H(p,\xi),\end{aligned}$$

where $\{p^k\}_{k\geq 1}$ is a sequence such that $p^k \to p$ as $k \to \infty$ and

$$\limsup_{p'\to p} g\left(y_{[K]}(p',\xi)^\top (p'-c)\right) = \lim_{k\to\infty} g\left(y_{[K]}(p^k,\xi)^\top (p^k-c)\right),$$

s^k is the index with $y_{s^k}(p^k,\xi)=1$, if $\mathcal{I}(p^k,\xi)\neq\emptyset$; s^k is any index in $\{1,\ldots,K\}$, if $\mathcal{I}(p^k,\xi)=\emptyset$.

Case 2: $\mathcal{I}(p,\xi)=\emptyset$. We have $y_{[K]}(p,\xi)=0\in\mathbb{R}^K$ and $\max_{1\leq i\leq K} u_i(p_i,\xi) < \|(u(p,\xi))_+\|_\infty$. According to the continuity of $u(\cdot,\cdot)$, for p' being sufficiently closed to p, we know that $\max_{1\leq i\leq K} u_i(p_i',\xi) < \|(u(p',\xi))_+\|_\infty$, which indicates $y_{[K]}(p',\xi) = 0\in\mathbb{R}^K$, and thus $H(p',\xi)=H(p,\xi)=0$, which indicates that $\limsup_{p'\to p} H(p',\xi) = 0 = H(p,\xi)$. To sum up, we verified (2.2).

Next, we focus on the upper semicontinuity of $\vartheta(\cdot)$ on the basis of (2.2). By using Fatou's lemma, we have, for any $F\in\mathcal{F}$, that

$$\begin{aligned}\limsup_{p'\to p}\mathbb{E}_F[H(p',\xi)] &= \limsup_{p'\to p}\int_\Xi H(p',\xi)F(\mathrm{d}\xi) \leq \int_\Xi \limsup_{p'\to p} H(p',\xi)F(\mathrm{d}\xi)\\ &\leq \mathbb{E}_F[H(p,\xi)],\end{aligned} \tag{2.3}$$

where the last inequality follows from the upper semicontinuity of $H(\cdot,\xi)$ for each fixed ξ. Note that

$$\begin{aligned}\limsup_{p'\to p}\vartheta(p') &= \limsup_{p'\to p}\inf_{F\in\mathcal{F}}\mathbb{E}_F[H(p',\xi)+h(p',\xi)]\\ &\leq \inf_{F\in\mathcal{F}}\limsup_{p'\to p}\mathbb{E}_F[H(p',\xi)+h(p',\xi)]\\ &\leq \inf_{F\in\mathcal{F}}\mathbb{E}_F[H(p,\xi)+h(p,\xi)]\\ &= \vartheta(p),\end{aligned}$$

where the last inequality follows from (2.3). ∎

The upper semicontinuity of $\vartheta(\cdot)$ is an important property for a maximization problem. Immediately, we have the following proposition.

PROPOSITION 2.8. *Problem* (P) *has an optimal solution* $p^*\in\mathcal{P}$ *with an optimal solution of the second stage problem* (1.5) *being the corresponding sparse solution.*

Proof. By Proposition 2.7 (i.e., the upper semicontinuity of $\vartheta(\cdot)$) and the compactness of $\mathcal{P}$, we know that an optimal p^* is attained for problem (P). Plugging p^* into problem (1.5), we can always select the sparse solution $y(p^*,\cdot)$ such that problem (1.5) attains the maximum (Proposition 2.5). According to Proposition 2.2, $Q(p^*,\cdot)$ is measurable. Therefore, p^* is a solution of problem (P) with the corresponding second stage sparse solution $y(p^*,\cdot)$. ∎

3. Data-driven analysis. To proceed the study in this section, we need to define the ambiguity set $\mathcal{F}$ in the distributionally robust multiproduct pricing problem (P). Generally speaking, there are mainly two types of ambiguity sets. One is the moment-based type (see e.g. [7]); the other one is the distance-based type (see e.g. [28]). Of particular interest of this paper, we consider the general moment-based ambiguity set, which can be written as

$$\mathcal{F}(\eta) = \{F\in\mathcal{M}(\Xi) : \mathbb{E}_F\left[\Psi(\eta,\xi)\right]\in\mathcal{K}\}, \tag{3.1}$$

where $\mathcal{M}(\Xi)$ denotes the collection of all probability measures supported on Ξ, Ψ is a mapping consisting of vectors and/or matrices with measurable components, η is some nominal moment information, the mathematical expectation of Ψ is taken w.r.t. each component of Ψ and $\mathcal{K}$ is a closed convex cone in the Cartesian product of some finite dimensional vector and/or matrix spaces.

We give two examples to validate the general moment ambiguity set (3.1).

Example 3.1 (Delage and Ye [7]). Consider the following ambiguity set with the first- and second-order moment information:

$$\mathcal{F} = \left\{ F \in \mathcal{M}(\Xi) : \begin{array}{l} (\mathbb{E}_F[\xi] - \mu)^\top \Sigma^{-1} (\mathbb{E}_F[\xi] - \mu) \le \gamma_1 \\ \mathbb{E}_F\left[(\xi - \mu)(\xi - \mu)^\top\right] \preceq \gamma_2 \Sigma \end{array} \right\}, \tag{3.2}$$

where $\mu \in \mathbb{R}^s$ and $\Sigma \in \mathbb{R}^{s \times s}$ denote the perceived mean vector and positive definite covariance matrix of the nominal probability distribution, respectively, and $\gamma_1 > 0$ and $\gamma_2 \ge 1$ are two constants quantifying decision-maker's confidence in μ and Σ. By using the well-known Schur complement, we can rewrite (3.2) as (3.1) with

$$\Psi(\eta, \xi) = \begin{pmatrix} \begin{bmatrix} -\Sigma & \mu - \xi \\ (\mu - \xi)^\top & -\gamma_1 \end{bmatrix} \\ (\xi - \mu)(\xi - \mu)^\top - \gamma_2 \Sigma \end{pmatrix} \text{ and } \mathcal{K} = \mathbb{S}_-^{s+1} \times \mathbb{S}_-^{s},$$

where $\eta = (\mu, \Sigma)$ and $\mathbb{S}_-^{s+1}$ and $\mathbb{S}_-^{s}$ denote the cones of $(s+1) \times (s+1)$ and $s \times s$ negative semidefinite symmetric matrices, respectively.

Example 3.2 (Guo et al. [13]). The second example of (3.1) is the so-called piecewise uniform approximation of ambiguity set based on moment condition. Let Ψ be a continuous vector-valued function. Consider, for example, that

$$\Psi(\eta, \xi) := \begin{pmatrix} \xi - \mu - \gamma_1 e \\ (\xi - \mu)^\top \Sigma^{-1} (\xi - \mu) - \gamma_2 \end{pmatrix} \text{ and } \mathcal{K} = \mathbb{R}_-^{s+1},$$

where $\eta = (\mu, \Sigma)$, μ and Σ denote the perceived mean vector and positive definite covariance matrix of the nominal probability distribution respectively, and γ_1 and γ_2 are corresponding confidence parameters.

To measure the distance between two probability measures, we give the definition of a class of probability metrics, which is known as ζ-structure probability metrics.

DEFINITION 3.3 (ζ-structure probability metrics). *Let $\mathcal{G}$ be a set of measurable functions from Ξ to $\mathbb{R}$. For $F', F \in \mathcal{M}(\Xi)$, we say*

$$\mathbb{D}_{\mathcal{G}}(F', F) := \sup_{\hbar \in \mathcal{G}} |\mathbb{E}_{F'}[\hbar(\xi)] - \mathbb{E}_F[\hbar(\xi)]|$$

a ζ-structure metric between F' and F induced by $\mathcal{G}$.

In what follows, for $F \in \mathcal{M}(\Xi)$ and $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{M}(\Xi)$, we use the following notations

$$\mathbb{D}_{\mathcal{G}}(F, \mathcal{F}_1) := \inf_{F' \in \mathcal{F}_1} \mathbb{D}_{\mathcal{G}}(F, F'), \quad \mathbb{D}_{\mathcal{G}}(\mathcal{F}_1, \mathcal{F}_2) := \sup_{F \in \mathcal{F}_1} \inf_{F' \in \mathcal{F}_2} \mathbb{D}_{\mathcal{G}}(F, F') \tag{3.3}$$

and

$$\mathbb{H}_{\mathcal{G}}(\mathcal{F}_1, \mathcal{F}_2) := \max\left\{\mathbb{D}_{\mathcal{G}}(\mathcal{F}_1, \mathcal{F}_2), \mathbb{D}_{\mathcal{G}}(\mathcal{F}_2, \mathcal{F}_1)\right\} \tag{3.4}$$

to denote the distance between F and $\mathcal{F}_1$, the deviation between $\mathcal{F}_1$ and $\mathcal{F}_2$, the Hausdorff distance between $\mathcal{F}_1$ and $\mathcal{F}_2$ induced by $\mathbb{D}_{\mathcal{G}}$, respectively.

Since the ζ-structure metric $\mathbb{D}_{\mathcal{G}}(\cdot,\cdot)$ is defined by $\mathcal{G}$, $\mathcal{G}$ is also called the generator of $\mathbb{D}_{\mathcal{G}}(\cdot,\cdot)$. With different generators, probability metrics with ζ-structure include many commonly-used probability metrics, such as Fortet-Mourier metric, total variation metric and Kantorovich metric, etc [30]. Specifically, we give definitions of the total variation metric and the Kantorovich metric.

Let

$$\mathcal{G}_{TV} := \left\{ \hbar : \Xi \to \mathbb{R} : \hbar \text{ is measurable and } \sup_{\xi\in\Xi} |\hbar(\xi)| \le 1 \right\}.$$

The total variation metric between $F', F \in \mathcal{M}(\Xi)$ is defined as

$$\mathbb{D}_{TV}(F', F) := \sup_{\hbar\in\mathcal{G}_{TV}} \left| \mathbb{E}_{F'}[\hbar(\xi)] - \mathbb{E}_F[\hbar(\xi)] \right|.$$

Similar to (3.3) and (3.4), for $F \in \mathcal{M}(\Xi)$ and $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{M}(\Xi)$, let

$$\mathbb{D}_{TV}(F, \mathcal{F}_1) := \inf_{F'\in\mathcal{F}_1} \mathbb{D}_{TV}(F, F'), \quad \mathbb{D}_{TV}(\mathcal{F}_1, \mathcal{F}_2) := \sup_{F\in\mathcal{F}_1} \inf_{F'\in\mathcal{F}_2} \mathbb{D}_{TV}(F, F')$$

and the Hausdorff distance $\mathbb{H}_{TV}(\mathcal{F}_1, \mathcal{F}_2) := \max\left\{\mathbb{D}_{TV}(\mathcal{F}_1, \mathcal{F}_2), \mathbb{D}_{TV}(\mathcal{F}_2, \mathcal{F}_1)\right\}$.

Let $\mathcal{G}_W := \{\hbar : \Xi \to \mathbb{R} : |\hbar(\xi) - \hbar(\xi')| \le \|\xi - \xi'\|\}$. The Kantorovich metric between $F', F \in \mathcal{M}(\Xi)$ is defined as $\mathbb{D}_W(F, F') = \sup_{\hbar\in\mathcal{G}_W} \left|\mathbb{E}_F[\hbar(\xi)] - \mathbb{E}_{F'}[\hbar(\xi)]\right|$. It is worth pointing out that the Kantorovich metric is also known as the first Wasserstein metric (see [35, Theorem 5.10]), which is defined as

$$\mathbb{D}_W(F', F) = \inf_{\pi\in\Pi(F',F)} \int_{\Xi\times\Xi} \|\xi' - \xi\| \, \mathrm{d}\pi(\xi', \xi),$$

where $\Pi(F', F)$ denotes the set of all probability distributions supported on $\Xi \times \Xi$ with marginal distributions being F' and F, respectively.

In practice, it is more likely that the decision maker can only have in hand some data, which can be used to deduce the information of η, for example, N independent identically distributed (iid) samples of ξ. Based on these data, we can then construct the data-driven counterpart of η, denoted by $\hat{\eta}_N$. Thus, the data-driven counterpart of the general moment-based ambiguity set (3.1) reads

$$\mathcal{F}(\hat{\eta}_N) := \left\{F \in \mathcal{M}(\Xi) : \mathbb{E}_F\left[\Psi(\hat{\eta}_N, \xi)\right] \in \mathcal{K}\right\}. \tag{3.5}$$

In what follows, to simplify the notation, without any confusion, we use $\mathcal{F}$ and $\widehat{\mathcal{F}}_N$ to represent $\mathcal{F}(\eta)$ and $\mathcal{F}(\hat{\eta}_N)$, respectively.

On the basis of the data-driven ambiguity set (3.5), we obtain the following data-driven counterpart of the DRO problem (P) as follows:

$$\max_{p\in\mathcal{P}} \inf_{F\in\widehat{\mathcal{F}}_N} \mathbb{E}_F\left[Q(p,\xi)\right]. \tag{3.6}$$

Analogous to $\vartheta(p)$ in (2.1), we denote $\hat{\vartheta}_N(p) := \inf_{F\in\widehat{\mathcal{F}}_N} \mathbb{E}_F\left[Q(p,\xi)\right]$. Then, in this section, we will concentrate on the relationship between the following two problems:

$$\max_{p\in\mathcal{P}} \vartheta(p) \tag{3.7}$$

and

$$\max_{p\in\mathcal{P}} \hat{\vartheta}_N(p), \tag{3.8}$$

which, in fact, are problems (P) and (3.6), respectively.

To facilitate the forthcoming discussion, we denote optimal values and optimal solution sets of problems (3.7) and (3.8) by v^*, $\mathcal{P}^*$ and $\hat{v}_N$, $\hat{\mathcal{P}}_N$, respectively.

In what follows, we focus on discussing the relationship between problems (3.7) and (3.8). First, we assume that the data-driven moment information $\hat{\eta}_N \to \eta$ with probability 1 (w.p.1) as $N \to \infty$, and the convergence assertions are established as the data size N tends to infinity. After that, in view of the fact that the driven data may contain noises, we investigate the statistical robustness quantitatively.

3.1. Convergence analysis. First, we have the following lemma in which an upper bound of the discrepancy between optimal values of problems (3.7) and (3.8) is given on the basis of the total variation metric.

LEMMA 3.4. *Assume that there exists an $L > 0$ such that $|Q(p, \xi)| \leq L$ for any $p \in \mathcal{P}$ and $\xi \in \Xi$. Then*

$$|\hat{v}_N - v^*| \leq L\mathbb{H}_{TV}(\widehat{\mathcal{F}}_N, \mathcal{F}).$$

Proof. Note the following derivation:

$$\begin{aligned}
\hat{v}_N - v^* &= \max_{p\in\mathcal{P}} \hat{\vartheta}_N(p) - \max_{p\in\mathcal{P}} \vartheta(p) \leq \max_{p\in\mathcal{P}} \left(\hat{\vartheta}_N(p) - \vartheta(p)\right) \\
&= \max_{p\in\mathcal{P}} \left(\inf_{F'\in\widehat{\mathcal{F}}_N} \mathbb{E}_{F'}\left[Q(p,\xi)\right] - \inf_{F\in\mathcal{F}} \mathbb{E}_F\left[Q(p,\xi)\right] \right) \\
&= \max_{p\in\mathcal{P}} \left(\inf_{F'\in\widehat{\mathcal{F}}_N} \sup_{F\in\mathcal{F}} \left(\mathbb{E}_{F'}\left[Q(p,\xi)\right] - \mathbb{E}_F\left[Q(p,\xi)\right]\right) \right) \\
&\leq \max_{p\in\mathcal{P}} \inf_{F'\in\widehat{\mathcal{F}}_N} \sup_{F\in\mathcal{F}} \left|\mathbb{E}_{F'}\left[Q(p,\xi)\right] - \mathbb{E}_F\left[Q(p,\xi)\right]\right| \\
&\overset{(a)}{\leq} L \inf_{F'\in\widehat{\mathcal{F}}_N} \sup_{F\in\mathcal{F}} \mathbb{D}_{TV}(F', F) \\
&= L\mathbb{D}_{TV}(\widehat{\mathcal{F}}_N, \mathcal{F}),
\end{aligned}$$

where (a) follows from the boundedness property $|Q(p, \xi)| \leq L$, the measurability of $Q(p, \cdot)$ (see Proposition 2.2) and the definition of the total variation metric.

A similar procedure can be applied to the case $v^* - \hat{v}_N$, and we can obtain that $v^* - \hat{v}_N \leq L\mathbb{D}_{TV}(\mathcal{F}, \widehat{\mathcal{F}}_N)$. Thus, we obtain $|\hat{v}_N - v^*| \leq L\mathbb{H}_{TV}(\widehat{\mathcal{F}}_N, \mathcal{F})$. ∎

Remark 3.5. In Lemma 3.4, the uniform boundedness of $|Q(p, \xi)|$ over $\mathcal{P} \times \Xi$ is required. This assumption can be satisfied trivially under certain specific conditions. For instance, if Ξ is bounded, we know from the boundedness of $\mathcal{P}$ and the continuity of g and h in (1.5) that the uniform boundedness property holds.

To derive the convergence assertion, we investigate the convergence $\mathbb{H}_{TV}(\widehat{\mathcal{F}}_N, \mathcal{F})$ to zero as N tends to infinity. Then we make the following standard assumption.

Assumption 3.6 (Slater condition). There exist an $F_0 \in \mathcal{M}(\Xi)$ and a positive constant $\gamma > 0$ such that $\mathbb{E}_{F_0}[\Psi(\eta, \xi)] + \gamma\mathbb{B} \subseteq \mathcal{K}$ holds.

We give the following lemma which can be found in [26, Corollary 6].

LEMMA 3.7. *Let Assumption 3.6 hold and $\mathcal{F}(\eta)$ be defined in* (3.1). *Suppose: (i) there exist a $\lambda_0 > 0$ and a measurable function $\kappa(\xi)$ such that $\|\Psi(\eta_1, \xi) - \Psi(\eta_2, \xi)\| \leq \kappa(\xi)\|\eta_1 - \eta_2\|$ for all η_1, η_2 with $\|\eta_i\| \leq \lambda_0$, $i = 1, 2$; (ii) there exists a $C > 0$ such* *that* $\mathbb{E}_F[\kappa(\xi)] \le C$ *for all* $F \in \cup_{\bar\eta\in\{\eta':\|\eta'\|\le\lambda_0\}}\mathcal{F}(\bar\eta)$. *Then*

$$\mathbb{H}_{\mathcal{G}}(\mathcal{F}(\eta_1), \mathcal{F}(\eta_2)) \le \frac{2C\Delta}{\gamma}\|\eta_1 - \eta_2\|$$

for all η_1, η_2 *with* $\|\eta_i\| \le \lambda_0$, $i = 1, 2$, *where* $\Delta := \max_{F\in\mathcal{M}(\Xi)} \mathbb{D}_{\mathcal{G}}(F, F_0)$ *and the generator* $\mathcal{G}$, γ *and* F_0 *are defined in Assumption 3.6.*

Then we are ready to present the main result of this subsection.

THEOREM 3.8. *Let Assumption 3.6 hold and* $\mathcal{F}(\eta)$ *be defined in* (3.1). *Suppose that: (i) there exists an* $L > 0$ *such that* $|Q(p,\xi)| \le L$ *for any* $p \in \mathcal{P}$ *and* $\xi \in \Xi$; *(ii) there exist a* $\lambda_0 > 0$ *and a measurable function* $\kappa(\xi)$ *such that* $\|\Psi(\eta_1,\xi) - \Psi(\eta_1,\xi)\| \le \kappa(\xi)\|\eta_1 - \eta_2\|$ *for all* η_1, η_2 *with* $\|\eta_i - \eta\| \le \lambda_0$, $i = 1, 2$; *(iii) there exists a* $C > 0$ *such that* $\mathbb{E}_F[\kappa(\xi)] \le C$ *for all* $F \in \cup_{\bar\eta\in\{\eta':\|\eta'-\eta\|\le\lambda_0\}}\mathcal{F}(\bar\eta)$. *If* $\hat\eta_N \to \eta$ *w.p.1 as* $N \to \infty$, *then we have* $\hat v_N \to v^*$ *w.p.1 as* $N \to \infty$. *Furthermore,* $\mathrm{d}(\hat{\mathcal{P}}_N, \mathcal{P}^*) \to 0$ *w.p.1 as* $N \to \infty$.

Proof. By invoking Lemma 3.7, we know from $\max_{F\in\mathcal{M}(\Xi)} \mathbb{D}_{TV}(F, F_0) \le 2$ (based on the definition of the total variational metric) that: for any η_1, η_2 with $\|\eta_i - \eta\| \le \lambda_0$ for $i = 1, 2$, $\mathbb{H}_{TV}(\mathcal{F}(\eta_1), \mathcal{F}(\eta_2)) \le 4C\|\eta_1 - \eta_2\|/\gamma$. Since $\hat\eta_N \to \eta$ w.p.1 as $N \to \infty$, we obtain $\|\hat\eta_N - \eta\| \le \lambda_0$ w.p.1 for sufficiently large N. Thus,

$$\mathbb{H}_{TV}(\widehat{\mathcal{F}}_N, \mathcal{F}) \le \frac{4C}{\gamma}\|\hat\eta_N - \eta\|$$

holds w.p.1 for sufficiently large N. According to Lemma 3.4, we obtain

$$\limsup_{N\to\infty} |\hat v_N - v^*| \le L \limsup_{N\to\infty} \mathbb{H}_{TV}(\widehat{\mathcal{F}}_N, \mathcal{F}) \le \frac{4LC}{\gamma} \limsup_{N\to\infty} \|\hat\eta_N - \eta\| \to 0$$

w.p.1, which implies that $\hat v_N \to v^*$ w.p.1 as $N \to \infty$.

Note from the proof procedure of Lemma 3.4 that

$$\sup_{p\in\mathcal{P}} \left|\hat\vartheta_N(p) - \vartheta(p)\right| \le L\mathbb{H}_{TV}(\widehat{\mathcal{F}}_N, \mathcal{F}) \to 0 \text{ w.p.1 as } N \to \infty.$$

With this observation, by using Proposition 2.7 and [20, Lemma C.1], we know that

$$\mathrm{d}(\hat{\mathcal{P}}_N, \mathcal{P}^*) \to 0 \text{ w.p.1 as } N \to \infty.$$

The proof is complete. ∎

Remark 3.9. All assumptions in Lemma 3.7 are routine. Specifically, the convergence $\hat\eta_N \to \eta$ w.p.1 as $N \to \infty$ can be ensured by the celebrated law of large numbers (LLN) if the driven data $\xi^1, \ldots, \xi^N$ are iid samples of ξ. The other assumptions can also be found in [26].

3.2. Quantitative statistical robustness. The concept of statistical robustness aims at allowing for arbitrarily small variation of the concentrated statistical estimator when a sufficiently small perturbation is introduced into the underlying empirical probability distribution. This idea primarily stems from the pioneering work of Hampel [15], and a comprehensive summary of statistical robustness is provided by Huber in [18]. Significant research has been conducted on both qualitative statistical robustness [5, 23, 24, 25] and quantitative statistical robustness [12, 37, 14].

In this subsection, we consider the quantitative statistical robustness of the data-driven problem (3.6). To this end, we assume that the driven data are perturbed or contaminated, denoted by $\tilde{\xi}^1, \ldots, \tilde{\xi}^N$, which follow from another probability distribution, denoted by $\tilde{F}$. The moment information of the contaminated data $\tilde{\xi}^1, \ldots, \tilde{\xi}^N$ is denoted by $\tilde{\eta}_N$. Analogously, we denote the following contaminated data-driven ambiguity set

$$\mathcal{F}(\tilde{\eta}_N) := \{F \in \mathcal{M}(\Xi) : \mathbb{E}_F [\Psi(\tilde{\eta}_N, \xi)] \in \mathcal{K}\},$$

which is simply written as $\widetilde{\mathcal{F}}_N$. Then we obtain the following contaminated data-driven problem

$$\max_{p\in\mathcal{P}} \inf_{F\in\widetilde{\mathcal{F}}_N} \mathbb{E}_F [Q(p,\xi)]. \tag{3.9}$$

Denote $\tilde{\vartheta}_N(p) := \inf_{F\in\widetilde{\mathcal{F}}_N} \mathbb{E}_F [Q(p,\xi)]$ and thus problem (3.9) can be recast as

$$\max_{p\in\mathcal{P}} \tilde{\vartheta}_N(p). \tag{3.10}$$

In what follows, we estimate the quantitative relationship between problems (3.8) and (3.10). We first give the following Lipschitz continuity property of the optimal value function.

LEMMA 3.10. *Under the conditions of Lemmas 3.4 and 3.7, there exists a positive constant C, independent of N, such that*

$$\left|v(\eta_N^1) - v(\eta_N^2)\right| \le C \left\|\eta_N^1 - \eta_N^2\right\|$$

for any $\left\|\eta_N^i\right\| \le \lambda_0, i = 1, 2$, where $\lambda_0 > 0$ is defined in Lemma 3.7 and $v(\eta_N^i)$ is the optimal value of problem $\max_{p\in\mathcal{P}} \inf_{F\in\mathcal{F}(\eta_N^i)} \mathbb{E}_F [Q(p,\xi)]$ for $i = 1, 2$.

Proof. Similar to Lemma 3.4, we have

$$\begin{aligned}
v(\eta_N^1) - v(\eta_N^2) &= \max_{p\in\mathcal{P}} \inf_{F\in\mathcal{F}(\eta_N^1)} \mathbb{E}_F [Q(p,\xi)] - \max_{p\in\mathcal{P}} \inf_{F\in\mathcal{F}(\eta_N^2)} \mathbb{E}_F [Q(p,\xi)] \\
&\le \max_{p\in\mathcal{P}} \left(\inf_{F\in\mathcal{F}(\eta_N^1)} \mathbb{E}_F [Q(p,\xi)] - \inf_{F\in\mathcal{F}(\eta_N^2)} \mathbb{E}_F [Q(p,\xi)] \right) \\
&= \max_{p\in\mathcal{P}} \left(\inf_{F'\in\mathcal{F}(\eta_N^1)} \sup_{F\in\mathcal{F}(\eta_N^2)} \left(\mathbb{E}_{F'} [Q(p,\xi)] - \mathbb{E}_F [Q(p,\xi)]\right) \right) \\
&\le \max_{p\in\mathcal{P}} \inf_{F'\in\mathcal{F}(\eta_N^1)} \sup_{F\in\mathcal{F}(\eta_N^2)} \left|\mathbb{E}_{F'} [Q(p,\xi)] - \mathbb{E}_F [Q(p,\xi)]\right| \\
&\le C_1 \inf_{F'\in\mathcal{F}(\eta_N^1)} \sup_{F\in\mathcal{F}(\eta_N^2)} \mathbb{D}_{TV}(F', F) = C_1 \mathbb{D}_{TV}(\mathcal{F}(\eta_N^1), \mathcal{F}(\eta_N^2)),
\end{aligned}$$

where C_1 is some positive constant. The other side $v(\eta_N^2) - v(\eta_N^1)$ can be estimated analogously. Finally, we obtain $\left|v(\eta_N^1) - v(\eta_N^2)\right| \le C_1 \mathbb{H}_{TV}(\mathcal{F}(\eta_N^1), \mathcal{F}(\eta_N^2))$. Then, by using Lemma 3.7 and replacing $\mathbb{H}_{\mathcal{G}}$ with $\mathbb{H}_{TV}$, we complete the proof. ∎

We need the following assumption, which specifies how the moment information relies on the driven data.

Assumption 3.11. There exists an $L > 0$ such that moment information parameters η_N^j from $\xi_j^1, \ldots, \xi_j^N$, $j = 1, 2$ satisfy $\left\|\eta_N^1 - \eta_N^2\right\| \le \frac{L}{N} \sum_{i=1}^N \left\|\xi_1^i - \xi_2^i\right\|$.

It is noteworthy that some similar assumptions can be found in [12, Lemma 1] and [37]. The following example shows Assumption 3.11 holds when Ξ is bounded.

Example 3.12. Let Ξ be bounded. Assume that the moment information η is consist of mean vector and covariance matrix (see, e.g., [7]), i.e., $\eta = (\mu, \Sigma)$. Then, for $j = 1, 2$, we have $\eta_N^j = \left(\bar{\mu}_N^j, \overline{\Sigma}_N^j\right)$, where

$$\bar{\mu}_N^j = \frac{1}{N}\sum_{i=1}^N \xi_j^i \quad \text{and} \quad \overline{\Sigma}_N^j = \frac{1}{N}\sum_{i=1}^N (\xi_j^i - \bar{\mu}_N^j)(\xi_j^i - \bar{\mu}_N^j)^\top .$$

Immediately, we have

$$\left\|\bar{\mu}_N^1 - \bar{\mu}_N^2\right\| = \left\| \frac{1}{N}\sum_{i=1}^N \xi_1^i - \frac{1}{N}\sum_{i=1}^N \xi_2^i \right\| \leq \frac{1}{N}\sum_{i=1}^N \left\|\xi_1^i - \xi_2^i\right\|$$

and

$$\begin{aligned}
(3.11)\quad \left\|\overline{\Sigma}_N^1 - \overline{\Sigma}_N^2\right\| &= \left\| \frac{1}{N}\sum_{i=1}^N (\xi_1^i - \bar{\mu}_N^1)(\xi_1^i - \bar{\mu}_N^1)^\top - \frac{1}{N}\sum_{i=1}^N (\xi_2^i - \bar{\mu}_N^2)(\xi_2^i - \bar{\mu}_N^2)^\top \right\| \\
&\leq \frac{1}{N}\sum_{i=1}^N \left\|(\xi_1^i - \bar{\mu}_N^1)(\xi_1^i - \bar{\mu}_N^1)^\top - (\xi_2^i - \bar{\mu}_N^2)(\xi_2^i - \bar{\mu}_N^2)^\top\right\| .
\end{aligned}$$

Note that, for $i = 1, \ldots, N$,

(3.12)

$$\begin{aligned}
&\left\|(\xi_1^i - \bar{\mu}_N^1)(\xi_1^i - \bar{\mu}_N^1)^\top - (\xi_2^i - \bar{\mu}_N^2)(\xi_2^i - \bar{\mu}_N^2)^\top\right\| \\
&= \left\|(\xi_1^i - \bar{\mu}_N^1)\left((\xi_1^i - \bar{\mu}_N^1) - (\xi_2^i - \bar{\mu}_N^2) + (\xi_2^i - \bar{\mu}_N^2)\right)^\top - (\xi_2^i - \bar{\mu}_N^2)(\xi_2^i - \bar{\mu}_N^2)^\top\right\| \\
&= \left\|(\xi_1^i - \bar{\mu}_N^1)\left((\xi_1^i - \bar{\mu}_N^1) - (\xi_2^i - \bar{\mu}_N^2)\right)^\top + (\xi_1^i - \bar{\mu}_N^1)(\xi_2^i - \bar{\mu}_N^2)^\top - (\xi_2^i - \bar{\mu}_N^2)(\xi_2^i - \bar{\mu}_N^2)^\top\right\| \\
&= \left\|(\xi_1^i - \bar{\mu}_N^1)\left(\xi_1^i - \bar{\mu}_N^1 - \xi_2^i + \bar{\mu}_N^2\right)^\top + (\xi_1^i - \bar{\mu}_N^1 - \xi_2^i + \bar{\mu}_N^2)(\xi_2^i - \bar{\mu}_N^2)^\top\right\| \\
&\leq \left\|\xi_1^i - \bar{\mu}_N^1\right\| \left\|\xi_1^i - \bar{\mu}_N^1 - \xi_2^i + \bar{\mu}_N^2\right\| + \left\|\xi_1^i - \bar{\mu}_N^1 - \xi_2^i + \bar{\mu}_N^2\right\| \left\|\xi_2^i - \bar{\mu}_N^2\right\| \\
&= \left(\left\|\xi_1^i - \bar{\mu}_N^1\right\| + \left\|\xi_2^i - \bar{\mu}_N^2\right\|\right) \left\|\xi_1^i - \bar{\mu}_N^1 - \xi_2^i + \bar{\mu}_N^2\right\| \\
&\leq \left(\left\|\xi_1^i - \bar{\mu}_N^1\right\| + \left\|\xi_2^i - \bar{\mu}_N^2\right\|\right) \left(\left\|\xi_1^i - \xi_2^i\right\| + \left\|\bar{\mu}_N^1 - \bar{\mu}_N^2\right\|\right) \\
&\leq C\left(\left\|\xi_1^i - \xi_2^i\right\| + \frac{1}{N}\sum_{j=1}^N \left\|\xi_1^j - \xi_2^j\right\|\right),
\end{aligned}$$

where $C > 0$ depends only on the diameter of the support set Ξ. By substituting (3.12) into (3.11), we obtain

$$\left\|\overline{\Sigma}_N^1 - \overline{\Sigma}_N^2\right\| \leq \frac{C}{N}\sum_{i=1}^N \left(\left\|\xi_1^i - \xi_2^i\right\| + \frac{1}{N}\sum_{j=1}^N \left\|\xi_1^j - \xi_2^j\right\|\right) = \frac{2C}{N}\sum_{i=1}^N \left\|\xi_1^i - \xi_2^i\right\| .$$

In this case, by letting $L = 2C$, we know that Assumption 3.11 holds. $\blacksquare$

Finally, we give the following quantitative statistical robustness result.

THEOREM 3.13. *Let Assumption 3.11 hold. Suppose that: (i) conditions in Lemmas 3.4 and 3.7 hold; (ii)* $F, \tilde{F} \in \mathcal{M}_1(\Xi) := \{F' \in \mathcal{M}(\Xi) : \mathbb{E}_{F'}[\|\xi\|] < \infty\}$. *Then*

$$\mathbb{D}_W \left(F^{\otimes N} \circ \hat{v}_N^{-1}, \tilde{F}^{\otimes N} \circ \hat{v}_N^{-1} \right) \leq L\mathbb{D}_W(F, \tilde{F}),$$

for all $N \in \mathbb{N}$, *where* $F^{\otimes N} \circ \hat{v}_N^{-1}$ *and* $\tilde{F}^{\otimes N} \circ \hat{v}_N^{-1}$ *are probability distributions over* $\mathbb{R}$ *induced by the optimal value* $\hat{v}_N$ *of problem (3.8),* $F^{\otimes N}$ *(or* $\tilde{F}^{\otimes N}$*) denotes the probability distribution over* $\Xi^{\otimes N}$ *with marginal being* F *(or* $\tilde{F}$*),* $\Xi^{\otimes N}$ *denotes the Cartesian product* $\underbrace{\Xi \times \ldots \times \Xi}_{N}$ *and* L *is defined in Assumption 3.11.*

The proof of Theorem 3.13 is similar to that in [12, 22, 37], which is mainly based on the definition of Kantorovich metric, and thus we omit it here.

4. MPEC reformulation. In this section, we consider the reformulation of the distributionally robust multiproduct pricing problem (P), which paves the way for solving problem (P) numerically.

For fixed $p \in \mathcal{P}$, we consider the inner minimization problem of (P) under the ambiguity set (3.1) as follows:

$$\begin{aligned} \inf_{F \in \mathcal{M}(\Xi)} \quad & \mathbb{E}_F[Q(p, \xi)] \\ \text{s.t.} \quad & \mathbb{E}_F[\Psi(\xi)] \in \mathcal{K}. \end{aligned} \tag{4.1}$$

The Lagrangian function of the minimization problem (4.1) is

$$\mathcal{L}(F, \Lambda) := \mathbb{E}_F[Q(p, \xi)] + \langle \Lambda, \mathbb{E}_F[\Psi(\xi)] \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in the space of $\mathcal{K}$, $\Lambda \in \mathcal{K}^*$ and $\mathcal{K}^*$ denotes the polar cone of $\mathcal{K}$, i.e., $\mathcal{K}^* := \{\Lambda : \langle \Lambda, \Gamma \rangle \leq 0, \ \forall \, \Gamma \in \mathcal{K}\}$, which is also a closed convex cone since $\mathcal{K}$ is a closed convex cone.

Then the Lagrangian dual problem of (4.1) can be written as

$$\sup_{\Lambda \in \mathcal{K}^*} \inf_{F \in \mathcal{M}(\Xi)} \mathcal{L}(F, \Lambda). \tag{4.2}$$

Consider the inner minimization problem of (4.2)

$$\inf_{F \in \mathcal{M}(\Xi)} \left(\mathbb{E}_F[Q(p, \xi)] + \langle \Lambda, \mathbb{E}_F[\Psi(\xi)] \rangle \right) = \inf_{F \in \mathcal{M}(\Xi)} \mathbb{E}_F\left[Q(p, \xi) + \langle \Lambda, \Psi(\xi) \rangle \right],$$

where the equality is due to the definition of inner product in $\mathcal{K}$ (in the sense of componentwise). Obviously, its optimal value, denoted by $\varphi(p, \Lambda)$, is

$$\varphi(p, \Lambda) := \inf_{\xi \in \Xi} \left(Q(p, \xi) + \langle \Lambda, \Psi(\xi) \rangle \right) \tag{4.3}$$

due to the definition of probability distribution, that is, F will take a single point probability distribution (or Dirac probability measure) to attain the minimum.

Therefore, the Lagrangian dual problem (4.2) can be further written as

$$\sup_{\Lambda \in \mathcal{K}^*} \varphi(p, \Lambda). \tag{4.4}$$

Finally, we obtain the reformulation of problem (P) as follows:

$$\max_{p \in \mathcal{P}, \Lambda \in \mathcal{K}^*} \varphi(p, \Lambda). \tag{4.5}$$

The following assertions follow from [32, Proposition 3.4], which asserts the dual gap between problem (P) and its dual problem (4.5).

PROPOSITION 4.1. *Let $p \in \mathcal{P}$ be fixed. If the Slater-type constraint qualification*

$$\alpha\mathbb{B} \subseteq -\{\mathbb{E}_F[\Psi(\xi)] : F \in \mathcal{M}(\Xi)\} + \mathcal{K} \tag{4.6}$$

holds for some $\alpha > 0$, then there is no dual gap between the primal problem (4.1) *and the Lagrangian dual problem* (4.4) *(i.e., the optimal values of problems* (4.1) *and* (4.4) *are consistent). If, in addition, these optimal values are finite, then the optimal solution set of* (4.4) *is nonempty and bounded.*

Conversely, if the optimal value of problem (4.4) *is finite and the optimal solution set of problem* (4.4) *is nonempty and bounded, then Slater-type condition* (4.6) *holds.*

In general, $\varphi(p, \Lambda)$ in (4.3) cannot be computed trivially if the support set Ξ contains infinite elements. In view of this, we consider its discrete approximation $\Xi^\nu = \{\xi^1, \ldots, \xi^\nu\}$, where samples $\xi^1, \ldots, \xi^\nu$ are obtained by some random or deterministic way (see also [29]). It can also be viewed as that all consumers in the market have ν preferences or tastes. Then we denote

$$\varphi_\nu(p, \Lambda) := \inf_{\xi \in \Xi^\nu} \left(Q(p, \xi) + \langle \Lambda, \Psi(\xi) \rangle\right) = \min_{1 \le i \le \nu} \left(Q(p, \xi^i) + \left\langle \Lambda, \Psi(\xi^i) \right\rangle\right).$$

Thus, we obtain the approximation of problem (4.5) as follows:

$$\max_{p \in \mathcal{P}, \Lambda \in \mathcal{K}^*} \varphi_\nu(p, \Lambda). \tag{4.7}$$

In fact, based on the definition of $Q(p, \xi)$ in (1.5), problem (4.7) can be recast as a large-scale constrained optimization problem as follows:

$$\begin{array}{lll} \max\limits_{p \in \mathcal{P}, \Lambda \in \mathcal{K}^*} & \min\limits_{1 \le i \le \nu} & \left(\left\langle \Lambda, \Psi(\xi^i) \right\rangle - h(p, \xi^i) + \max\limits_{y^i, \gamma^i} g\left(\left(y^i_{[K]} \right)^\top (p - c) \right) \right) \\ \text{s.t.} & & 0 \le \begin{pmatrix} y^i \\ \gamma^i \end{pmatrix} \perp \begin{pmatrix} 0 & e \\ -e^\top & 0 \end{pmatrix} \begin{pmatrix} y^i \\ \gamma^i \end{pmatrix} + \begin{pmatrix} -u(p, \xi^i) \\ 1 \end{pmatrix} \ge 0, \ 1 \le i \le \nu. \end{array} \tag{4.8}$$

In what follows, we will adopt some routine approaches in robust optimization [2] to equivalently reformulate problem (4.8).

For given $p \in \mathcal{P}$ and $\Lambda \in \mathcal{K}^*$, the inner min-max problem of (4.8), i.e.,

$$\begin{array}{ll} \min\limits_{1 \le i \le \nu} & \left(\left\langle \Lambda, \Psi(\xi^i) \right\rangle - h(p, \xi^i) + \max\limits_{y^i, \gamma^i} g\left(\left(y^i_{[K]} \right)^\top (p - c) \right) \right) \\ \text{s.t.} & 0 \le \begin{pmatrix} y^i \\ \gamma^i \end{pmatrix} \perp \begin{pmatrix} 0 & e \\ -e^\top & 0 \end{pmatrix} \begin{pmatrix} y^i \\ \gamma^i \end{pmatrix} + \begin{pmatrix} -u(p, \xi^i) \\ 1 \end{pmatrix} \ge 0, \ 1 \le i \le \nu \end{array} \tag{4.9}$$

is equivalent to a max-min problem as below:

$$\begin{array}{lll} \max\limits_{\{(y^i, \gamma^i)\}_{i=1}^\nu} & \min\limits_{1 \le i \le \nu} & \left(\left\langle \Lambda, \Psi(\xi^i) \right\rangle - h(p, \xi^i) + g\left(\left(y^i_{[K]} \right)^\top (p - c) \right) \right) \\ \text{s.t.} & & 0 \le \begin{pmatrix} y^i \\ \gamma^i \end{pmatrix} \perp \begin{pmatrix} 0 & e \\ -e^\top & 0 \end{pmatrix} \begin{pmatrix} y^i \\ \gamma^i \end{pmatrix} + \begin{pmatrix} -u(p, \xi^i) \\ 1 \end{pmatrix} \ge 0, \ 1 \le i \le \nu. \end{array} \tag{4.10}$$

In fact, it is known that the optimal value of problem (4.9) is always larger than or equal to that of problem (4.10). Then we only need to verify that it holds vice versa. For any given $1 \le i \le \nu$, denote $(y^{i,*}, \gamma^{i,*})$ an arbitrary optimal solution of the inner maximization problem of (4.9). Then $\{(y^{i,*}, \gamma^{i,*})\}_{i=1}^\nu$ is a feasible solution of the outer maximization problem of (4.10). By letting $(y^i, \gamma^i) = (y^{i,*}, \gamma^{i,*})$ for $i = 1, \ldots, \nu$

in problem (4.10), we obtain a lower bound of the optimal value of problem (4.10) as below:

$$\min_{1\le i\le \nu} \ \left\langle \Lambda, \Psi(\xi^i)\right\rangle - h(p,\xi^i) + g\left(\left(y_{[K]}^{i,*}\right)^\top (p-c)\right),$$

which actually equals to the optimal value of problem (4.9). Thus, we have shown that the optimal values of problems (4.9) and (4.10) are equal. Then, by using (4.10), we can rewrite problem (4.8) as

(4.11)
$$\begin{array}{cl} \max\limits_{p\in\mathcal{P},\Lambda\in\mathcal{K}^*,\{(y^i,\gamma^i)\}_{i=1}^{\nu}} & \left(\min\limits_{1\le i\le\nu} \left\langle \Lambda, \Psi(\xi^i)\right\rangle - h(p,\xi^i) + g\left(\left(y_{[K]}^{i}\right)^\top (p-c)\right)\right) \\ \text{s.t.} & 0 \le \begin{pmatrix} y^i \\ \gamma^i\end{pmatrix} \perp \begin{pmatrix} 0 & e \\ -e^\top & 0\end{pmatrix}\begin{pmatrix} y^i \\ \gamma^i\end{pmatrix} + \begin{pmatrix} -u(p,\xi^i) \\ 1\end{pmatrix} \ge 0, \ 1\le i\le \nu. \end{array}$$

We then summarize the above discussion and obtain the following proposition.

Proposition 4.2. *Suppose that: (i) the support set $\Xi = \{\xi^1,\ldots,\xi^\nu\}$; (ii) the Slater-type constraint qualification* (4.6) *holds. Then, the optimal value of problem* (P) *is equal to that of problem* (4.11). *Moreover, p is an optimal solution of problem* (P) *if and only if there exist $\Lambda, \{(y^i,\gamma^i)\}_{i=1}^{\nu}$ such that p together with them is an optimal solution of problem* (4.11).

Problem (4.11) is a typical MPCC that has been extensively studied (see monograph [27]). Numerous papers (e.g., [1, 17, 19, 11]) have contributed to solving (4.11) for various types of stationary points. Furthermore, we observe that the objective function of problem (4.11) is concave w.r.t. p and Λ. The observation and the closed-form expression of the sparse solution $y_{[K]}$ can help us to develop numerical procedures to a global optima of problem (P) with a support set Ξ containing a finite number of elements.

5. Numerical experiments.

In this section, by employing the MPCC reformulation (4.11) and the sparse solution (see Definition 2.3), we give numerical procedures to find a global optima of problem (P) in some specific cases. Moreover, we illustrate our approach by three numerical examples.

5.1. Numerical procedures for problems (1.4) and (P).

In this subsection, we consider some numerical procedures for problems (1.4) and (P) when the support set is finite. To this end, we assume that the support set $\Xi = \{\xi^1,\ldots,\xi^\nu\}$ for some $\nu\in\mathbb{N}$ and the probability for $\xi=\xi^i$ is π_i for $i=1,\ldots,\nu$. Denote $\pi=(\pi_1,\ldots,\pi_\nu)^\top$. Surely, we have $\pi\ge 0$ and $e^\top\pi = 1$.

First of all, we consider the numerical procedures of problem (1.4), that is,

$$\max_{p\in\mathcal{P}} \quad \textstyle\sum_{i=1}^{\nu} \pi_i Q(p,\xi^i), \tag{5.1}$$

where $Q(p,\xi^i) = H(p,\xi^i) - h(p,\xi^i)$ and

$$\begin{array}{rcl} H(p,\xi^i) = & \max\limits_{y^i} & g\left((y_{[K]}^i)^\top (p-c)\right) \\ & \text{s.t.} & y^i\in\mathcal{S}(p,\xi^i), \ i=1,\ldots,\nu. \end{array} \tag{5.2}$$

Denote $\mathcal{P}_j^i := \{p\in\mathcal{P} : y_j(p,\xi^i)=1\}$ for $i=1,\ldots,\nu$ and $j=1,\ldots,K$, where $y_j(p,\xi^i)$ denotes the value of the jth component of the sparse solution for given p and

ξ^i (see Definition 2.3). For fixed i, denote $\mathcal{P}^i_{K+1} := \mathcal{P} \backslash (\cup_{j=1}^K \mathcal{P}^i_j)$. It is worth pointing out that $\mathcal{P}^i_j$ might be empty for some $i \in \{1, \ldots, \nu\}$ and $j \in \{1, \ldots, K\}$. Furthermore, if the utility function $u(p, \xi)$ is given by a linear case (i.e., (1.1)) and $\mathcal{P}$ is convex, then $\mathcal{P}^i_j$ is convex for $i = 1, \ldots, \nu$ and $j = 1, \ldots, K$. To see this, consider the feasible set of (1.2) and let Π be the set of vertices of the feasible set. Then for any $\hat{y} \in \Pi$,

$$\{u \in \mathbb{R}^m : \hat{y} \in \arg\max_y \; y^\top u \;\; \text{s.t. } e^\top y \leq 1, \; y \geq 0\}$$

is a convex set formed by the convex combination of edges emanating from this vertex. Since affine mappings carry convex sets to convex sets, and $\mathcal{P}$ is convex, $\mathcal{P}^i_j$ is also convex.

Let $J := \{\{j_i\}_{i=1}^\nu : j_i \in \{1, \ldots, K+1\}, i = 1, \ldots, \nu\}$. Since, for each $p \in \mathcal{P}$ and $i \in \{1, \ldots, \nu\}$, there exists a j_i such that $p \in \mathcal{P}^i_{j_i}$, we have $\mathcal{P} = \cup_{\{j_i\}_{i=1}^\nu \in J} \left(\cap_{i=1}^\nu \mathcal{P}^i_{j_i}\right)$. Moreover, due to the uniqueness of the sparse solution, for different $\{j_i\}_{i=1}^\nu, \{\tilde{j}_i\}_{i=1}^\nu \in J$, $\left(\cap_{i=1}^\nu \mathcal{P}^i_{j_i}\right) \cap \left(\cap_{i=1}^\nu \mathcal{P}^i_{\tilde{j}_i}\right) = \emptyset$. Then there exists a partition of $\mathcal{P}$ induced by J such that there exist at most $(K+1)^\nu$ blocks in the partition and each block corresponding to a subproblem as follows:

$$\begin{array}{ll} \max\limits_p & \sum_{i=1}^\nu \pi_i g\left(y_{[K]}(p, \xi^i)^\top (p - c)\right) - \sum_{i=1}^\nu \pi_i h(p, \xi^i) \\ \text{s.t.} & p \in \cap_{i=1}^\nu \mathcal{P}^i_{j_i}, \end{array} \tag{5.3}$$

where $y_{[K]}(p, \xi^i)$ denotes the first K components of the sparse solution of the second stage problem (5.2) for given p and ξ^i. Note that for each $p \in \mathcal{P}^i_{j_i}$, $y_{j_i}(p, \xi^i) = 1$ and $y_k(p, \xi^i) = 0$ for $k \neq j_i$, which implies $H(p, \xi^i) = g(p_{j_i} - c_{j_i})$. Therefore, problem (5.3) can be further recast as

$$\begin{array}{ll} \max\limits_p & \sum_{i=1}^\nu \pi_i g\left(p_{j_i} - c_{j_i}\right) - \sum_{i=1}^\nu \pi_i h(p, \xi^i) \\ \text{s.t.} & p \in \cap_{i=1}^\nu \mathcal{P}^i_{j_i}. \end{array} \tag{5.4}$$

Specially, when $\mathcal{P}^i_{j_i}$ is convex and closed, $g(\cdot)$ is concave and $h(\cdot, \xi^i)$ is convex for $i = 1, \ldots, \nu$, problem (5.4) is convex, which can be solved effectively.

To summarize the aforementioned statements, we have the following procedures to compute a global solution of problem (1.4).

- S1 Compute partitions $\cap_{i=1}^\nu \mathcal{P}^i_{j_i}$, $\{j_i\}_{i=1}^\nu \in J$.
- S2 For each given $\{j_i\}_{i=1}^\nu$ with $j_i \in \{1, \ldots, K+1\}$, $i = 1, \ldots, \nu$, calculate a global solution of subproblem (5.4).
- S3 Choose one of the largest objectives among these subproblems, and output its optimal value and optimal solution.

Next, we consider problem (P), i.e., the distributionally robust counterpart of problem (1.4), as follows:

$$\max_{p \in \mathcal{P}} \inf_{\pi \in \mathcal{F}} \sum_{i=1}^\nu \pi_i Q(p, \xi^i), \tag{5.5}$$

where $Q(p, \xi^i)$ is the same as that in (5.1). By using the dual reformulation in Section 4 and the ν partitions of $\mathcal{P}$ in (5.3), (5.5) can be divided into at most $(K+1)^\nu$ subproblems as follows:

$$\begin{array}{ll} \max\limits_{p, \Lambda \in \mathcal{K}^*} & \left(\min\limits_{1 \leq i \leq \nu} \langle \Lambda, \Psi(\xi^i) \rangle - h(p, \xi^i) + g\left(y_{[K]}(p, \xi^i)^\top (p - c)\right)\right) \\ \text{s.t.} & p \in \cap_{i=1}^\nu \mathcal{P}^i_{j_i}, \end{array} \tag{5.6}$$

where $y_{[K]}(p,\xi^i)$ denotes the first K components of the sparse solution of problem (5.2) for given p and ξ^i. Similarly, problem (5.6) is equivalent to the following problem:

$$\begin{array}{ll}\max\limits_{p,\Lambda\in\mathcal{K}^*} & \left(\min\limits_{1\le i\le\nu}\left\langle \Lambda,\Psi(\xi^i)\right\rangle - h(p,\xi^i) + g\left(p_{j_i}-c_{j_i}\right)\right)\\ \text{s.t.} & p\in\cap_{i=1}^{\nu}\mathcal{P}^i_{j_i}.\end{array}\tag{5.7}$$

To solve problem (5.5), we only need to replace S2 by S2′ as follows.

S2′ For each given $\{j_i\}_{i=1}^{\nu}$ with $j_i\in\{1,\ldots,K+1\}$, $i=1,\ldots,\nu$, compute a global solution of (5.7).

Since J induces a partition of $\mathcal{P}$, we have the following assertions.

PROPOSITION 5.1. *Procedures S1, S2 and S3 output the globally optimal value and a globally optimal solution of problem* (5.1). *Procedures S1, S2′ and S3 output the globally optimal value and a globally optimal solution of problem* (5.5).

5.2. Numerical results. In this subsection, we provide three numerical examples to illustrate our models and approaches. First, we consider the stress test (see, e.g., [8, 16]) using a simple example where the random vector has three possible realizations. The second example is performed with one pricing product and some larger sample sizes. Based on the second example, the last example considers a general case with multiple pricing products and larger sample sizes. All codes were implemented in MATLAB R2018b on a laptop with the 13th Gen Intel(R) Core(TM) i9-13900H (2.60 GHz) and 32GB RAM.

First of all, we do the stress test, which shows the reasonability and necessariness of the distributionally robust multiproduct pricing problem (P).

Example 5.2. Let $K=2$ and $m=4$, i.e., there are total four products in the market and the target firm produces two products. The utility of a consumer with preference $\xi=(\xi_1,\xi_2,\xi_3)^\top$ for purchasing product j $(j=1,2,3,4)$ is defined as $u_j(p_j,\xi)=\xi_1+\xi_2x_j-\xi_3p_j$. Set $x=(x_1,x_2,x_3,x_4)^\top=(5,2,3,1)^\top$, $p_3=3$, $p_4=0.5$ and $c=(c_1,c_2)^\top$ with $c_1=5$, $c_2=3$. Then, the target firm aims to determine the price $p=(p_1,p_2)^\top$.

Let the probability distribution of random vector ξ be

$$\xi=\begin{cases}\xi^1=(3,3,1)^\top \text{ with probability } \pi_1=\frac{3}{4},\\ \xi^2=(2,2,1)^\top \text{ with probability } \pi_2=\frac{1}{8},\\ \xi^3=(1,1,2)^\top \text{ with probability } \pi_3=\frac{1}{8}.\end{cases}\tag{5.8}$$

Set $\mathcal{P}=[1,9]\times[1,9]$, $g\left(y_{[K]}(\xi)^\top(p-c)\right)=y_{[K]}(\xi)^\top(p-c)$ and $h(p,\xi)=\frac{\|p-\bar{p}\|^2}{64}$, where $\bar{p}=(5,4)^\top$ is a predetermined price vector.

It is highly probable that the estimated probability distribution of the random vector ξ is not the true distribution. To account for this uncertainty, we construct an ambiguity set defined as

$$\mathcal{F}:=\left\{\pi=(\pi_1,\pi_2,\pi_3)^\top\in\mathbb{R}^3_+ : \pi_1\xi^1+\pi_2\xi^2+\pi_3\xi^3-\mu-0.5e\le 0,\ e^\top\pi=1\right\},$$

where μ is the nominal mean vector of ξ, $e\in\mathbb{R}^3$ be a vector with all elements equal to 1, and $\mathcal{F}$ includes the discrete probability distribution in (5.8).

Analysis of Example 5.2: Immediately, an ambiguity-neutral target firm will make a decision according to the stochastic programming problem (1.4), that is

$$\max_{p\in\mathcal{P}}\sum_{i=1}^{3}\pi_i\left(y_1^i(p)(p_1-c_1)+y_2^i(p)(p_2-c_2)\right)-\frac{\|p-\bar{p}\|^2}{64}\tag{5.9}$$

where π_1, π_2, π_3 are defined in (5.8) and $y^i(p) = (y^i_1(p), y^i_2(p), y^i_3(p), y^i_4(p))^\top$ is the sparse solution of the corresponding second stage problem with price p and ξ^i for $i = 1, 2, 3$.

An ambiguity-averse target firm hedges against the possibility, and would like to make a decision according to problem (P), that is the following DRO problem

$$\max_{p\in\mathcal{P}} \inf_{\pi\in\mathcal{F}} \sum_{i=1}^{3} \pi_i \left(y^i_1(p)(p_1 - c_1) + y^i_2(p)(p_2 - c_2) \right) - \frac{\|p - \bar{p}\|^2}{64}. \tag{5.10}$$

To solve problem (5.9), we employ procedures S1, S2 and S3 to find an optimal solution. Note that in this case, $i = 1, 2, 3$ and $j_i \in \{1, 2, 3\}$. Then we can find the partition $\cap_{i=1}^{\nu} \mathcal{P}^i_{j_i}$, $\{j_i\}_{i=1}^{\nu} \in J$ of $\mathcal{P}$ as in S_1 as follows: $\mathcal{P}^1_1 = [1, 9] \times [1, 9]$, $\mathcal{P}^2_1 = [1, 7] \times [1, 9]$, $\mathcal{P}^2_3 = [7, 9] \times [1, 9]$, $\mathcal{P}^3_1 = [1, 2.5] \times [1, 9]$, $\mathcal{P}^3_3 = [2.5, 9] \times [1, 9]$ and $\mathcal{P}^i_{j_i} = \emptyset$ for the rest (i, j_i). The corresponding sparse solution reads: $y^1_{[2]}(p) = (1, 0)^\top$, $p \in [1, 9] \times [1, 9]$,

$$y^2_{[2]}(p) = \begin{cases} (1, 0)^\top & p \in [1, 7] \times [1, 9] \\ (0, 0)^\top & \text{otherwise,} \end{cases} \quad \text{and} \quad y^3_{[2]}(p) = \begin{cases} (1, 0)^\top, & p \in [1, 2.5] \times [1, 9], \\ (0, 0)^\top, & \text{otherwise.} \end{cases}$$

Therefore, by procedure $S2$, problem (5.9) can be solved via the following three subproblems:

$$\max_{p\in\mathcal{P}^1_1\cap\mathcal{P}^2_1\cap\mathcal{P}^3_1} \quad \frac{3}{4}(p_1 - 5) + \frac{1}{8}(p_1 - 5) + \frac{1}{8}(p_1 - 5) - \frac{\|p - \bar{p}\|^2}{64}, \tag{5.11}$$

$$\max_{p\in\mathcal{P}^1_1\cap\mathcal{P}^2_1\cap\mathcal{P}^3_3} \quad \frac{3}{4}(p_1 - 5) + \frac{1}{8}(p_1 - 5) - \frac{\|p - \bar{p}\|^2}{64}, \tag{5.12}$$

$$\max_{p\in\mathcal{P}^1_1\cap\mathcal{P}^2_3\cap\mathcal{P}^3_3} \quad \frac{3}{4}(p_1 - 5) - \frac{\|p - \bar{p}\|^2}{64}. \tag{5.13}$$

The optimal solutions for problems (5.11), (5.12), and (5.13) are $(2.5, 4)^\top$, $(7, 4)^\top$, and $(9, 4)^\top$, with optimal values of $-\frac{665}{256}$, $\frac{27}{16}$, and $\frac{11}{4}$, respectively. Therefore, $(9, 4)^\top$ and $\frac{11}{4}$ are the optimal solution and optimal value of problem (5.9), respectively.

In what follows, we calculate an optimal solution and the optimal value of problem (5.10). According to (5.6), we consider the following problem

$$\max_{p\in\mathcal{P}^1_{j_1}\cap\mathcal{P}^2_{j_2}\cap\mathcal{P}^3_{j_3},\Lambda\in\mathcal{K}^*} \left(\min_{1\le i\le\nu} \left(\langle \Lambda, \Psi(\xi^i) \rangle + y^i_{[K]}(p)^\top (p - c) \right) - \frac{\|p - \bar{p}\|^2}{64} \right) \tag{5.14}$$

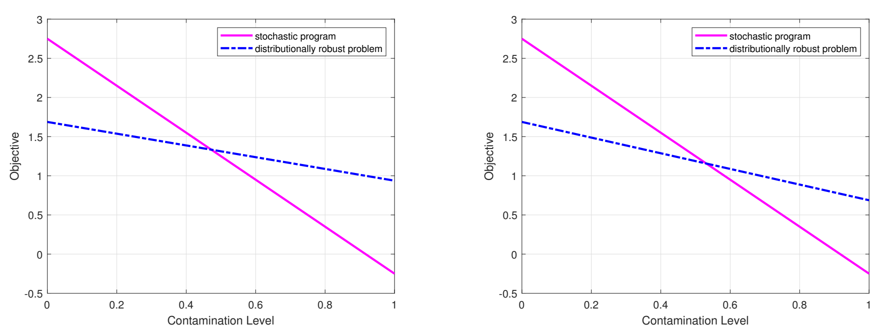
with $(j_1, j_2, j_3) = (1, 1, 1), (1, 1, 3)$ or $(1, 3, 3)$, where $\Psi(\xi) = \xi - \mu - 0.5e$ and $\mathcal{K}^* = \mathbb{R}^3_+$. It is noteworthy that for different $\{j_i\}_{i=1}^3$, $y^i_{[K]}(p)$, $i = 1, 2, 3$, are given above, then problem (5.14) is convex w.r.t. (p, Λ), which can be solved effectively.

When we take $\mu = (2.2, 2.2, 1)^\top$ in the ambiguity set $\mathcal{F}$, the optimal solution of problem (5.10) is $(p, \Lambda) = (7, 4, 0, 0, 0)^\top$, achieving an optimal value $\frac{15}{16}$. By setting $p = (7, 4)^\top$ in (5.10), we can obtain the worst-case probability distribution $\pi = (0, 0.5, 0.5)^\top$ for problem (5.10). Similarly, when $\mu = (2.625, 2.625, 1.125)^\top$ is set in the ambiguity set $\mathcal{F}$, the optimal solution to problem (5.10) is $p = (7, 4)^\top$, achieving an optimal value of $\frac{11}{16}$. The worst-case probability distribution in this case is $\pi = (0, 0.375, 0.625)^\top$.

To make a stress test, consider contaminations of the discrete probability distribution in (5.8) with the worst probability distribution from (5.10) under $\mu = (2.2, 2.2, 1)^\top$ and $\mu = (2.625, 2.625, 1.125)^\top$, respectively, that is

$$(1-\alpha)\begin{pmatrix}0.75\\0.125\\0.125\end{pmatrix} + \alpha\begin{pmatrix}0\\0.5\\0.5\end{pmatrix} \text{ and } (1-\alpha)\begin{pmatrix}0.75\\0.125\\0.125\end{pmatrix} + \alpha\begin{pmatrix}0\\0.375\\0.625\end{pmatrix},$$

where $\alpha \in [0, 1]$ denotes the contamination level. Under different contamination levels, we plot objectives when $p = (9, 4)^\top$ (an optimal solution of the ambiguity-neutral target firm) and $p = (7, 4)^\top$ (an optimal solution of the ambiguity-averse target firm) in Figure 1. Figure 1 shows that around $\alpha = 0.477$ (or $\alpha = 0.533$), the optimal solution of the ambiguity-averse target firm begins to perform better than the optimal solution of the ambiguity-neutral target firm. This means that if the perceptive probability distribution in (5.8) is contaminated (e.g., $\alpha > 0.477$ for $\mu = (2.2, 2.2, 1)^\top$ and $\alpha > 0.533$ for $\mu = (2.625, 2.625, 1.125)^\top$), the ambiguity-neutral target firm might make a worse decision than the ambiguity-averse one. Additionally, the fact that the objective value for the ambiguity-neutral target firm changes more steeply than that for the ambiguity-averse target firm suggests that the distributionally robust multiproduct pricing model is more resilient to contaminated data. In practice, it is often difficult to know the true distribution exactly, which highlights the reasonableness and necessariness of our distributionally robust multiproduct pricing model.



(a) Stress test for $\mu = (2.2, 2.2, 1)^\top$. (b) Stress test for $\mu = (2.625, 2.625, 1.125)^\top$.

FIG. 1. *Objectives of stochastic and distributionally robust models under different levels of contamination.*

In the next example, we apply the same methodology to a larger sample size case.

Example 5.3. Let $K = 1$ and $m = 3$. Assume that ξ is a random vector supported over $\mathbb{R}^3_+$, i.e., $\xi = (\xi_1, \xi_2, \xi_3)^\top$; the utility of the consumer with preference ξ purchasing product j ($j = 1, 2, 3$) is defined as $u_j(p_j, \xi) = \xi_1 + \xi_2 x_j - \xi_3 p_j$, where $x = (x_1, x_2, x_3) = (5, 1, 3)$, $p_2 = 2$, $p_3 = 4$ and $c_1 = 2$; g is an identity mapping, i.e., $g(t) = t$, and $h(p_1, \xi) = \|p_1 - 3\|^2/81$; the feasible set of the price is $\mathcal{P} = [1, 9]$. The ambiguity set $\mathcal{F}(\eta)$ is defined as (see Example 3.2):

$$\mathcal{F}(\eta) := \left\{F \in \mathcal{M}(\Xi) : \mathbb{E}_F\left[\begin{pmatrix}\xi - \mu - \gamma_1 e\\(\xi-\mu)^\top\Sigma^{-1}(\xi-\mu) - \gamma_2\end{pmatrix}\right] \in \mathbb{R}^4_-\right\}, \tag{5.15}$$

where $\eta = (\mu, \Sigma) \in \mathbb{R}^3 \times \mathbb{R}^{3\times 3}$ with Σ being positive definite, $\gamma_1, \gamma_2 \in \mathbb{R}$ are two scalars.

To generate the discrete samples $\{\xi^i\}_{i=1}^{\nu}$, we adopt the uniform probability distribution over $[1, 7]$. Specifically, we generate $\{\xi_1^i\}_{i=1}^{\nu}$, $\{\xi_2^i\}_{i=1}^{\nu}$ and $\{\xi_3^i\}_{i=1}^{\nu}$ independently, and each of them are iid and follow the uniform probability distribution over $[1, 7]$. Based on (4.11) and ambiguity set (5.15), the DRO problem can be written as

$$\text{(5.16)} \quad \begin{array}{cl} \max\limits_{\substack{p_1 \in \mathcal{P}, \Lambda \in \mathcal{K}^*,\\ \{(y^i, \gamma^i)\}_{i=1}^{\nu}}} & \left(\min\limits_{1 \le i \le \nu} \left\langle \Lambda, \Psi(\xi^i) \right\rangle - \frac{(p_1-3)^2}{81} + y_1^i (p_1 - c_1) \right) \\ \text{s.t.} & 0 \le \begin{pmatrix} y^i \\ \gamma^i \end{pmatrix} \perp \begin{pmatrix} 0 & e \\ -e^\top & 0 \end{pmatrix} \begin{pmatrix} y^i \\ \gamma^i \end{pmatrix} + \begin{pmatrix} -u(p, \xi^i) \\ 1 \end{pmatrix} \ge 0, \ 1 \le i \le \nu, \end{array}$$

where $\Psi(\xi) = \begin{pmatrix} \xi - \mu - \gamma_1 e \\ (\xi - \mu)^\top \Sigma^{-1} (\xi - \mu) - \gamma_2 \end{pmatrix}$.

Analysis of Example 5.3: First, for $\nu = 20, 50, 100, 200, 400, 1000, 2000, 5000$, we compute the optimal solutions and the optimal values of problem (5.16). In problem (5.16), we set $\gamma_1 = \gamma_2 = 1$, $\mu = (4, 4, 4)^\top$, and $\Sigma = \text{diag}(3, 3, 3)$. The numerical results are presented in Table 1.

Table 1
Optimal solutions and optimal values of (5.16) *for* $\nu = 20, 50, 100, 200, 400, 1000, 2000, 5000$.

sample size ν	20	50	100	200	400	1000	2000	5000
optimal solutions	6.57	4.13	3.81	3.46	3.41	3.45	3.53	3.39
optimal values	3.53	1.31	0.93	0.86	0.79	0.71	0.68	0.63
CPU times (s)	2.14	4.51	10.10	19.37	38.12	117.18	215.98	836.25

Second, we show the convergence tendency of the objective of DRO problem (5.16) when η is approximated. We set $\nu = 100, 200, 400, 1000, 2000, 5000$ and fix $\gamma_1 = \gamma_2 = 1$, $\eta = (\mu, \Sigma)$ with $\mu = (4, 4, 4)^\top$, and $\Sigma = \text{diag}(3, 3, 3)$. To perturb η, we set $\eta_\epsilon = (\mu + \epsilon_1 e, \Sigma + \epsilon_2 I)$, where I is an identity matrix with a proper dimension, $\epsilon = (\epsilon_1, \epsilon_2)$ are chosen from

$$\{(0.4, 4), (0.3, 3), (0.2, 2), (0.1, 1), (0.05, 0.5), (0.02, 0.2), (0.01, 0.1), (0, 0)\}.$$

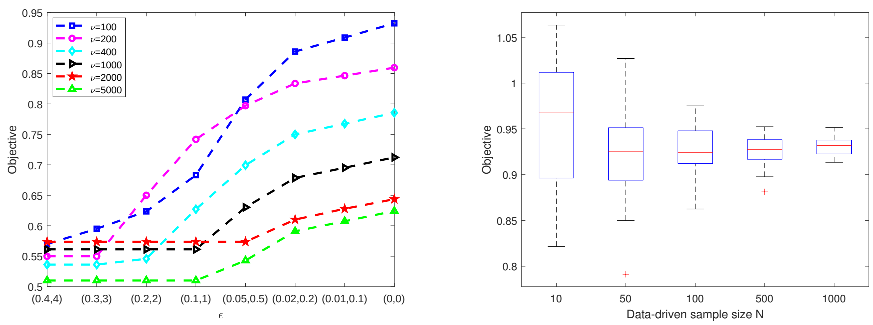
For fixed ν, we plot in Figure 2 (a) the objective of the DRO problem (5.16) regarding to ϵ. We can clearly observe from Figure 2 that the objective gradually converges to the true one, i.e., $\epsilon = (0, 0)$.

Moreover, we generate $\{\omega_j^i\}_{i=1}^N$, $j = 1, 2, 3$ independently, using the uniform probability distribution over $[1, 7]$. Then we define the data-driven moment information of (μ, Σ) by $(\hat{\mu}_N, \hat{\Sigma}_N)$ with

$$\hat{\mu}_N = \frac{1}{N} \left(\sum_{i=1}^N \omega_1^i, \sum_{i=1}^N \omega_2^i, \sum_{i=1}^N \omega_3^i \right)^\top \text{ and } \hat{\Sigma}_N = \frac{1}{N} \text{diag} \left(\sum_{i=1}^N \tau_1^i, \sum_{i=1}^N \tau_2^i, \sum_{i=1}^N \tau_3^i \right),$$

where $\tau_j^i = (\omega_j^i - \sum_{i=1}^N \omega_j^i)^2$. For each sample size $N = 10, 50, 100, 500, 1000$, we generate the data-driven moment information $(\hat{\mu}_N, \hat{\Sigma}_N)$ 20 times and compute the optimal value of problem (5.16) when $\nu = 100$. The convergence behavior of the optimal value as the sample size grows is shown in the boxplot in Figure 2(b).

In the last example, we consider a multiproduct case with larger sample sizes.



(a) Convergence for $\nu = 100$, 200, 400, 1000, 2000, 5000.

(b) Boxplots for $\nu = 100$ with different data-driven sample sizes.

FIG. 2. *Convergence of the DRO problem* (5.16).

Example 5.4. Let $m = 11$ and $K = 10$. Similarly, we assume that $\xi = (\xi_1, \xi_2, \xi_3)^\top$ and the utility of the consumer with preference ξ purchasing product j ($j = 1, \ldots, m$) is defined as $u_j(p_j, \xi) = \xi_1 + \xi_2 x_j - \xi_3 p_j$, where $x = (x_1, \ldots, x_m)^\top$, p_m and $c = (c_1, \ldots, c_K)^\top$ are given. Again, we assume that $g(t) = t$ and $h(p, \xi) = \|p - c\|^2 /81$. The feasible set of the price p is $\mathcal{P} = \underbrace{[1,9] \times \ldots \times [1,9]}_{K}$. The ambiguity sets $\mathcal{F}(\eta)$ and $\mathcal{F}_\nu(\eta)$ are the same as those in Example 5.3.

Analysis of Example 5.4: First of all, we randomly generate $x = (x_1, \ldots, x_m)^\top$, p_m and $c = (c_1, \ldots, c_K)^\top$. By (4.11), the DRO problem for an ambiguity-averse target firm reads

(5.17)

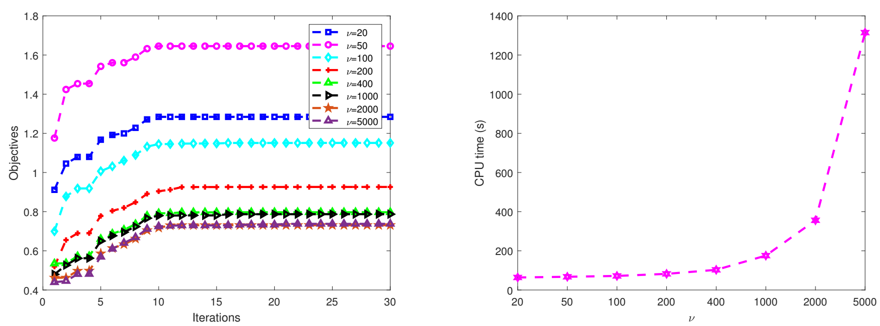
$$\begin{array}{cl} \max\limits_{p\in\mathcal{P},\Lambda\in\mathcal{K}^*,\{(y^i,\gamma^i)\}_{i=1}^{\nu}} & \left(\min\limits_{1\le i\le \nu} \left\langle \Lambda, \Psi(\xi^i) \right\rangle - \frac{\|p-c\|^2}{81} + \left(y^i_{[K]}\right)^\top (p_{[K]} - c_{[K]}) \right) \\ \text{s.t.} & 0 \le \begin{pmatrix} y^i \\ \gamma^i \end{pmatrix} \perp \begin{pmatrix} 0 & e \\ -e^\top & 0 \end{pmatrix} \begin{pmatrix} y^i \\ \gamma^i \end{pmatrix} + \begin{pmatrix} -u(p, \xi^i) \\ 1 \end{pmatrix} \ge 0, \ 1 \le i \le \nu. \end{array}$$

Since there are multiple products in this example, using the numerical procedures in subsection 5.1 directly may lead to the curse of dimensionality. This motivates us to price each product alternately using an alternate pricing method. Specifically, we first randomly assign an initial price to the K products, and then, for i from 1 to K, we price product i while keeping the prices of the other products fixed. We repeat this process until the prices converge. In fact, the pricing problem for a single product is the same as that in Example 5.3. To generate samples, we set $\nu = 20, 50, 100, 200, 400, 1000, 2000, 5000$, and independently generate $\{\xi_1^i\}_{i=1}^{\nu}$, $\{\xi_2^i\}_{i=1}^{\nu}$, and $\{\xi_3^i\}_{i=1}^{\nu}$, each of which are i.i.d. samples according to the uniform probability distribution over the interval $[1, 7]$. We set the parameters in $\mathcal{F}_\nu(\eta)$ as follows: $\gamma_1 = 0.5$, $\gamma_2 = 1$, $\mu = (4, 4, 4)^\top$, and $\Sigma = \mathrm{diag}(3, 3, 3)$.

The numerical results for problem (5.17) are presented in Table 2 with CPU times, which show that the scalability of the solution procedure presented in subsection 5.1 is acceptable. Furthermore, we show the objectives of problem (5.17) during the alternate iteration process in Figure 3. As it can be observed from Figure 3, the objective values increase with the number of iterations and eventually become stable, which illustrates the effectiveness of the alternate method. In addition, as the sample size increases, the final objective values decrease. This observation is consistent with the fact that the ambiguity set $\mathcal{F}_\nu(\eta)$ in problem (5.17) enlarges as the sample size increases. Also, the objective values tend to converge as the sample size increases, which indicates the empirical convergence between problems (4.7) and (4.5) as ν tends to infinity.

TABLE 2
Optimal solutions p^* of (5.17) for $\nu = 20, 50, 100, 200, 400, 1000, 2000, 5000$.

ν	p_1^*	p_2^*	p_3^*	p_4^*	p_5^*	p_6^*	p_7^*	p_8^*	p_9^*	p_{10}^*
20	3.62	2.57	2.09	3.54	2.70	2.27	3.11	3.67	2.54	3.42
50	3.98	2.90	2.40	3.54	2.70	2.58	3.33	3.67	2.54	3.42
100	3.53	2.60	1.98	3.54	2.70	2.19	3.06	3.67	2.54	3.42
200	3.35	2.42	1.91	3.54	2.70	2.08	2.84	3.67	2.54	3.42
400	3.06	2.25	1.79	3.54	2.70	1.95	2.64	3.67	2.54	3.42
1000	3.17	2.32	1.81	3.54	2.70	1.97	2.73	3.67	2.54	3.42
2000	3.08	2.26	1.79	3.54	2.70	1.94	2.65	3.67	2.54	3.42
5000	3.10	2.27	1.81	3.54	2.70	1.95	2.67	3.67	2.54	3.42



(a) Objective values for ν = 20, 50, 100, 200, 400, 1000, 2000, 5000.

(b) CPU times for ν = 20, 50, 100, 200, 400, 1000, 2000, 5000

FIG. 3. *Numerical results of problem (5.17) for ν = 20, 50, 100, 200, 400, 1000, 2000, 5000.*

6. Conclusions. In this paper, we consider the distributionally robust multiproduct pricing problem (P) in a hierarchical form. We establish measurability and semicontinuity by using a sparse solution of the second stage optimization problem (1.5) of problem (P). Moreover, we conduct the data-driven analysis of problem (P) when the ambiguity set is given by a general moment-based case. Specifically, we investigate the convergence properties when the moment information is exactly approximated by true data, and the quantitative statistical robustness when the moment information is approximated by noisy data. Finally, we propose a numerical procedure to compute a solution of the distributionally robust multiproduct pricing problem (P) based on a MPCC reformulation (4.11) and the sparse solution of problem (1.5). Preliminary numerical results are reported to illustrate the effectiveness of our models and approaches.

Acknowledgments. We would like to thank the associate editor and two referees for their very helpful comments.

REFERENCES

[1] R. Andreani, L. D. Secchin, and P. J. Silva, *Convergence properties of a second order augmented Lagrangian method for mathematical programs with complementarity constraints*, SIAM J. Optim., 28 (2018), pp. 2574–2600.

[2] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*, vol. 28, Princeton University Press, Princeton, 2009.

[3] S. Berry and A. Pakes, *The pure characteristics demand model*, Internat. Econom. Rev., 48 (2007), pp. 1193–1225.

[4] X. Chen, H. Sun, and R. J.-B. Wets, *Regularized mathematical programs with stochastic equilibrium constraints: Estimating structural demand models*, SIAM J. Optim., 25 (2015), pp. 53–75.

[5] R. Cont, R. Deguest, and G. Scandolo, *Robustness and sensitivity analysis of risk measurement procedures*, Quant. Finance, 10 (2010), pp. 593–606.

[6] R. W. Cottle, J.-S. Pang, and R. E. Stone, *The Linear Complementarity Problem*, SIAM, Philadelphia, 2009.

[7] E. Delage and Y. Ye, *Distributionally robust optimization under moment uncertainty with application to data-driven problems*, Oper. Res., 58 (2010), pp. 595–612.

[8] J. Dupačová, *Stress testing via contamination*, in Marti, K., et al. (eds.) Coping with Uncertainty: Modeling and Policy Issues, Lecture Notes in Economics and Mathematical Systems, Berlin, 2006, Springer, pp. 29–46.

[9] F. Facchinei and J.-S. Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.

[10] R. Gao and A. Kleywegt, *Distributionally robust stochastic optimization with Wasserstein distance*, Math. Oper. Res., 48 (2023), pp. 603–655.

[11] L. Guo and Z. Deng, *A new augmented Lagrangian method for MPCCs—theoretical and numerical comparison with existing augmented Lagrangian methods*, Math. Oper. Res., 47 (2022), pp. 1229–1246.

[12] S. Guo and H. Xu, *Statistical robustness in utility preference robust optimization models*, Math. Program., 190 (2021), pp. 679–720.

[13] S. Guo, H. Xu, and L. Zhang, *Convergence analysis for mathematical programs with distributionally robust chance constraint*, SIAM J. Optim., 27 (2017), pp. 784–816.

[14] S. Guo, H. Xu, and L. Zhang, *Statistical robustness of empirical risks in machine learning*, J. Mach. Learn. Res., 24 (2023), pp. 1–38.

[15] F. R. Hampel, *A general qualitative definition of robustness*, Ann. Math. Stat., 42 (1971), pp. 1887–1896.

[16] G. A. Hanasusanto, D. Kuhn, S. W. Wallace, and S. Zymler, *Distributionally robust multi-item newsvendor problems with multimodal demand distributions*, Math. Program., 152 (2015), pp. 1–32.

[17] X. Huang, X. Yang, and K. L. Teo, *Partial augmented Lagrangian method and mathematical programs with complementarity constraints*, J. Global Optim., 35 (2006), pp. 235–254.

[18] P. J. Huber, *Robust Statistics*, John Wiley & Sons, New York, 1981.

[19] A. F. Izmailov, M. V. Solodov, and E. Uskov, *Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints*, SIAM J. Optim., 22 (2012), pp. 1579–1606.

[20] J. Jiang and X. Chen, *Optimality conditions for nonsmooth nonconvex-nonconcave min-max problems and generative adversarial networks*, SIAM J. Math. Data Sci., 5 (2023), pp. 693–722.

[21] J. Jiang and X. Chen, *Pure characteristics demand models and distributionally robust mathematical programs with stochastic complementarity constraints*, Math. Program., 198 (2023), pp. 1449–1484.

[22] J. Jiang and S. Li, *Statistical robustness of two-stage stochastic variational inequalities*, Optim. Lett., 16 (2022), pp. 2591–2605.

[23] V. Krätschmer, A. Schied, and H. Zähle, *Qualitative and infinitesimal robustness of tail-dependent statistical functionals*, J. Multivariate Anal., 103 (2012), pp. 35–47.

[24] V. Krätschmer, A. Schied, and H. Zähle, *Comparative and qualitative robustness for law-invariant risk measures*, Finance Stoch., 18 (2014), pp. 271–295.

[25] V. Krätschmer, A. Schied, and H. Zähle, *Domains of weak continuity of statistical func-*

tionals with a view toward robust statistics, J. Multivariate Anal., 158 (2017), pp. 1–19.
[26] Y. Liu, A. Pichler, and H. Xu, *Discrete approximation and quantification in distributionally robust optimization*, Math. Oper. Res., 44 (2019), pp. 19–37.
[27] Z.-Q. Luo, J.-S. Pang, and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, 1996.
[28] P. Mohajerin Esfahani and D. Kuhn, *Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations*, Math. Program., 171 (2018), pp. 115–166.
[29] J.-S. Pang, C.-L. Su, and Y.-C. Lee, *A constructive approach to estimating pure characteristics demand models with pricing*, Oper. Res., 63 (2015), pp. 639–659.
[30] S. T. Rachev, L. B. Klebanov, S. V. Stoyanov, and F. Fabozzi, *The Methods of Distances in the Theory of Probability and Statistics*, vol. 10, Springer, New York, 2013.
[31] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, vol. 317, Springer Science & Business Media, New York, 2009.
[32] A. Shapiro, *On duality theory of conic linear problems*, in Semi-infinite Programming, Springer, 2001, pp. 135–165.
[33] C.-L. Su, *Equilibrium problems with equilibrium constraints: Stationarities, algorithms, and applications*, PhD thesis, Stanford University, 2005.
[34] H. Sun, C.-L. Su, and X. Chen, *SAA-regularized methods for multiproduct price optimization under the pure characteristics demand model*, Math. Program., 165 (2017), pp. 361–389.
[35] C. Villani et al., *Optimal Transport: Old and New*, vol. 338, Springer, Berlin, 2009.
[36] J. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, 1947.
[37] H. Xu and S. Zhang, *Quantitative statistical robustness in distributionally robust optimization models*, Pacific J. Optim., 19 (2023), pp. 335–361.