Robust Pricing and Production with Information Partitioning and Adaptation

Georgia Perakis
Sloan School of Management, Massachusetts Institute of Technology
georgiap@mit.edu

Melvyn Sim
Department of Analytics & Operations, NUS Business School, National University of Singapore
dscsimm@nus.edu.sg

Qinshen Tang
Division of Information Technology & Operations Management, Nanyang Business School, Nanyang Technological University
qinshen.tang@ntu.edu.sg

Peng Xiong
Department of Analytics & Operations, NUS Business School, National University of Singapore
xiongpengnus@gmail.com

We introduce a new distributionally robust optimization model to address a two-period, multi-item joint pricing and production problem, which can be implemented in a data-driven setting using historical demand and side information pertinent to the prediction of demands. Starting from an additive demand model, we introduce a new partitioned-moment-based ambiguity set to characterize its residuals, which also determines how the second-period demand would evolve from the first-period information in a data-driven setting. We investigate the joint pricing and production problem by proposing a cluster-adapted markdown policy and an affine recourse adaptation, which allow us to reformulate the problem as a mixed-integer linear optimization problem that we can solve to optimality using commercial solvers. We also extend our framework to ensemble methods using a set of ambiguity sets constructed from different clustering approaches. Both the numerical experiments and case study demonstrate the benefits of the cluster-adapted markdown policy and the partitioned-moment-based ambiguity set in improving the mean profit over the empirical model—when applied to most out-of-sample tests.

Key words: multi-item, pricing, retail analytics, clustering, distributionally robust optimization
History: April 16, 2022

1. Introduction

As one of the largest sectors in today's economy, the global retail sector is estimated to have revenues of USD 28 trillion in 2019. It represents 31% of the world's GDP and employs billions of people throughout the globe (Research and Markets 2016). Managers, in this sector, usually prepare their products well before the selling season because such external factors as price elasticity, production lead times, capacity constraints, and short selling seasons (see Abernathy et al. 1999, Mantrala and Rao 2001). These factors rule out the possibility of re-orders or quick replenishment of inventory during the selling season (Iyer and Bergen 1997). For example, due to the short selling period of about six to eight weeks, LC Waikiki, the largest apparel retailer in Turkey, manufactures all the products before the selling season (Naderi et al. 2020). Likewise, the fast changing fashion trends and short selling seasons necessitate the Swedish multinational retail-clothing company H&M to stock up their inventories before the selling season to cope with the long delivery lead time from its main suppliers in Asia (Adlarson and Holgersson 2016).

The selling season is often divided into two periods. Zara—one of the world's largest fashion retailer—divides each year into summer and winter "campaigns" (or selling seasons), where each campaign consists of both a regular period and a clearance period (Carboni Borrasé 2009). The commodities or services offered by most producers are sold at one price during the first, regular period but at a *lower* price during the second, clearance period; this pattern is especially typical in the fashion apparel industry as shown before, in which retailers face unpredictable levels of demand within a short selling season. Similar patterns also widely exist in stores selling perishable goods such as confectioneries, bakeries, and flowers. For instance, some bakery stores in Singapore offer up to fifty percent discounts after 9 pm (Goody Feed 2016).

The practice of marking down prices during later periods is observed in many different industries. This phase can be economically significant: Bloomingdale's estimates that about half (around $400 million) of their total sales are due to products at markdown prices. An additional 9% ($72 million) of sales are to salvage retailers, which acquire inventory for pennies on the dollar; in this way, Bloomingdale's avoids excessive "cannibalization" of their new designs (Ferguson and Koenigsberg 2007). Another example comes from the consumer electronics industry, in which producers usually mark down an old model after a new model's release. Shortly after the iPhone 7's release, for instance, Apple reduced the iPhone 6 price from $808 to $700 (Bhardwaj 2017). Markdowns are useful for clearing out any overstock and also attract more buyers, thereby increasing the total number of customers for different commodities. Hence this strategy is widely used by producers, especially those that offer multiple items.

As a result, many firms—ranging from cargo logistics to carpet manufacturing—have created management roles tasked with overseeing joint pricing and production decisions (Cross 2011). Federgruen and Heching (1999) show that optimally integrating price and production can result in 6.5% higher profits for a specialty retailer of high-end women's apparel than under a sequential procedure, whereby a price trajectory is determined first and is then followed by a replenishment policy based on the resulting demand distributions. An industry study sponsored by IBM (Webber et al. 2011) reports that integrating pricing and production can help retailers reduce food waste by enabling them to react more appropriately to uncertain factors, such as unexpected weather.

That study also incorporates a case study that describes how the Dutch grocery retailer Albert Heijn experimented with integrated pricing and inventory control policies in one of its stores.

In light of the magnitude of these implications, in this paper, we examine in particular, a market supplied by a producer who sells multiple items and who, for each item, determines the production quantity and retail price *before* the selling season, which is divided into two periods. Any "leftovers" that remain after the first period's demand realization can then be sold at the *markdown* price, which is quoted at the start of the second period. No production occurs in the second period due to the aforementioned factors such as, among other things, long lead time.

However, there are a number of obstacles that hinder the exploitation of such problem. *First*, as mentioned in Monahan et al. (2004), payoff functions in the joint pricing and production problem are typically not concave. This leads to computational challenges. *Second*, scholars and retailers must account for the intrinsic randomness of demand for a firm's products (Cohen et al. 2018). Thus Petruzzi and Dada (1999) document that, even for the case of a general single period and a single product, the uncertainty of demand entails an exhaustive search to find the optimal pricing and production strategies. *Third*, while the commonly used dynamic programming is a powerful theoretical tool for characterizing the optimal policy in simple systems, it suffers from the "curse of dimensionality" and is ill suited for computing realistic joint pricing and production policies due to the complexity of the underlying recursive equations that increases with the number of state variables involved (Bertsimas and Thiele 2006).

Solving a dynamic programming model requires the modeler to have precise knowledge of the demand distribution, which is often difficult to determine accurately (see, for instance, Chen and Shi 2019). Moreover, we usually do not have enough data in practice to fine tune the distributional model. For instance, if the planning horizon is a day, as in the case of selling perishable goods such as confectioneries, bakeries, and flowers, a ten-year selling period would only yield about 3650 days of demand information. If the planning horizon is a month, then we would have at most 120 sample paths for the ten-year record.

Demand can also be influenced by other exogenous factors or side information such as, *inter alia*, weather and economic outlook. For example, Steele (1951) first shows that weather is correlated with the sales of a department store. For fashion apparel retailers, Martínez-de Albéniz and Belkaid (2021) establish empirically that rain has a large effect on footfall while temperature has a milder effect. As far as we know, side information, despite being pertinent in demand forecasts, has not been exploited in the context of pricing and production decisions.

Conceivably, there are hardly any papers that consider joint pricing and production for multiple items, given the complexity that arises when considering just a single product under uncertainty (Chen and Simchi-Levi 2012). Of course, few firms manufacture only a single product of uniform style and size; nearly all firms produce multiple versions of their products. These different versions are enough to constitute separate products for purposes of pricing. In these situations, the pricing and production decisions must account for the effect of demand interrelations stemming from the substitutability, and/or complementarity among products or product lines. In short, the production manager needs to make multi-dimensional decisions that incorporate a variety of different factors. This complexity arises from the need, when demand is uncertain, to make multiple decisions *jointly* (Ramachandran et al. 2018).

On the other hand, retail analytics' market size is estimated to be over USD 4.3 billion in 2020; it is expected to grow at a compound annual growth rate of over 21.2% over the next few years and is forecasted to be USD 11.1 billion by 2025 (Markets and Markets 2020). Most of the works in retail analytics, however, focus on predictive analytics—it is about sensing what is ahead—but alone does not provide firms the insights that they need (Bradlow et al. 2017) to improve decision making, which is the role of prescriptive analytics. A key challenge moving from predictive to prescriptive analytics is the need to optimize over an exponentially large or infinite number of feasible solutions, which makes it impractical to enumerate all possible solutions and to compare their performances for a given predictive model. The literature on retail analytics typically focuses on estimating salient information on demand such as averages while discarding information on prediction errors. As we will see, in modeling the production and pricing problem, the uncertainty in demand predictions is critical in prescriptive analytics for profit maximization. Apart from the computational tractability, there are also issues of "optimizer's curse" (Smith and Winkler 2006), which may yield inferior solutions when one optimizes over a predictive model without considering uncertainty in the prediction. This paper is a step in this direction; we focus on how to robustly address a multi-item two-period joint pricing and production problem for a commonly used linear predictive demand model by also taking into account the distributional ambiguity of the prediction residuals.

Our main results and contributions

1. We introduce a new distributionally robust optimization model to address a two-period, multi-item joint pricing and production problem, which, to the best of our knowledge, is the first paper that can be implemented in a data-driven setting using historical demand and side information pertinent to the prediction of demands. Our proposed additive demand model is able to capture cross-product effects and also cross-period effects; that is, it can incorporate both substitutability and complementarity among products as well as partial reference effects from the previous period. Using a real-world data set, we demonstrate the cross-product effect at the stock keeping unit (SKU) level and also at the "aggregate over store" level.

2. We introduce a new partitioned-moment-based ambiguity set to characterize uncertain demands in a two-period setting, which is compatible with non-anticipative decision making. Unlike the standard moment-based ambiguity set, we can adjust the level of robustness by varying the number of information clusters from being the most robust as the standard moment-based ambiguity set with one cluster to being the least robust as the empirical distribution.

3. We introduce the cluster-adapted markdown policy, which allows us to reformulate the two-period, multi-item joint pricing and production problem as a mixed-integer linear optimization problem. For greater tractability, we also propose an affine recourse adaptation, which, in our empirical studies, obtains nearly optimal solutions but uses far less computational effort. We also establish the optimality of the approximation when there is only one product. Our framework can also be extended to ensemble methods using a set of ambiguity sets constructed from different clustering approaches.

4. We document the performance of our proposed framework by applying it to real-world data sets including one from Kaggle and another one from a cosmetic company. To the best of our knowledge, we are also the first to consider the multi-item newsvendor problem with side information. We test the model with the Kaggle data and use the oil price as side information, which yields favourable results over the empirical model without side information.

5. As there are a plethora of empirical and machine learning approaches we could experiment and integrate into our proposed framework, we provide in Online Appendix 5 how we can formulate the robust pricing and production problem via the new algebraic modeling language, RSOME (Chen et al. 2020) and attain the solutions using Gurobi. We provide the platform to inspire and facilitate further studies to improve the model.

Literature review

Since the groundbreaking paper of Whitin (1955), the joint pricing and production problem has received sustained attention from researchers in both the operations research and operations management literature. We review related works in *joint inventory and pricing* that stem from various types of demand models, including stochastic, robust, distributionally robust, and those that incorporate side information. For a survey of research on inventory control, interested readers are referred to Simchi-Levi et al. (2005), Axsäter (2015), Shenoy and Rosas (2018), and the references therein. For a survey of research on dynamic pricing, see Bitran and Caldentey (2003), Elmaghraby and Keskinocak (2003), and Talluri and Van Ryzin (2006) as well as related works on markdowns: Yin et al. (2009), Caro and Gallien (2012), and the references therein. In the context of retail analytics, our paper is closely related to data-driven retail operations. Interested readers can refer to Wedel and Kannan (2016) and Bradlow et al. (2017) for excellent commentary on marketing analytics and retail analytics, respectively, and Qi et al. (2020) for survey on data-driven retail operations.

Stochastic optimization models. One of the seminal works in joint inventory and pricing under uncertainty is Federgruen and Heching (1999). These authors consider a single-item, periodic review model and find that a base-stock list price is optimal under most conditions. Their model has been extended to various settings; for comprehensive surveys, see Elmaghraby and Keskinocak (2003), Chan et al. (2004), Yano and Gilbert (2005), Chen and Simchi-Levi (2012), and their references.

Whereas the literature cited above focuses on identifying the model's structural properties, other papers consider one- or two-period models for the purpose of drawing insights; see, for example, Petruzzi and Dada (1999), Agrawal and Seshadri (2000), and Cachon and Kök (2007). Yet most of these papers cannot fully solve the first-period problem without some specific assumptions—for example, that second-period demand is deterministic (Cachon and Kök 2007).

The wealth of literature on pricing and inventory decisions in a single-product setting, as just described, contrasts with the scarcity of papers that consider joint pricing and production for multiple items. This state of affairs is hardly surprising in light of the complexity of even single-product models when uncertainty is involved (Chen and Simchi-Levi 2012). To the best of our knowledge, the only exceptions are Bertsimas and De Boer (2005), Aydin and Porteus (2008), and Song and Xue (2021). However, these three works differ from our paper either by proposing heuristics (Bertsimas and De Boer 2005), considering one-period model (Aydin and Porteus 2008), or by seeking structurally optimal policies (Song and Xue 2021).

Robust optimization models. Almost all the papers mentioned so far assume that producers know both the exact distribution of demand and the market's response to prices. Yet even with abundant past observed data, it is difficult to select the most appropriate functional form and to estimate the distribution of demand uncertainty (Chen et al. 2021)—especially in the multi-item case. Hence, there is an emerging trend of using robust optimization to address inventory and pricing problems. A typical robust optimization model assumes uncertain demand lies within a so-called uncertainty set, which are generally conic representable such as ellipsoidal or polyhedral (see, for instance, Bertsimas and Sim 2004, Ben-Tal et al. 2004). However, related works either consider the inventory control problem (e.g. Bertsimas and Thiele 2006, Yue et al. 2006, Perakis and Roels 2008, Jiang et al. 2011) or the pricing problem (e.g. Lim et al. 2008). Yet there have been only few attempts that address the *joint* pricing and production problem. In a multi-item make-to-order setting, Adida and Perakis (2006) introduce a demand-based fluid model; they (i) show that their robust formulation is of the same order of complexity as the deterministic one and (ii) adaptively use the deterministic solution algorithm to address the robust problem. These authors also extend their model to the case of two firms competing under demand uncertainty (Adida and Perakis

2010). We differ from this stream of literature by adopting the distributionally robust optimization techniques—which utilize more distribution information including the support and moments—to jointly optimize the pricing and production problem.

Distributionally robust optimization models. Since the pioneering work of Scarf (1958), several scholars have considered the distribution-free newsvendor problem; interested readers may refer to Gallego et al. (2001) for a detailed review of the early literature. More recently, and motivated by theoretical work in moment-based distributionally robust optimization (e.g., Delage and Ye 2010), an extensive literature has used similar techniques to address either inventory or pricing problems. There are also works along the lines of statistical-distance-based distributionally robust optimization models (see e.g. Gao and Kleywegt 2016, Mohajerin Esfahani and Kuhn 2017, Hanasusanto and Kuhn 2018). These distributionally robust optimization models have recently been unified under the *robust stochastic optimization* framework proposed in Chen et al. (2020). In fact, part of the generalization framework is chiefly motivated by our model to address the two-period pricing and production problem. To the best of our knowledge, no paper has adopted the distributionally robust optimization approach to address the multi-item joint pricing and production problem in operations management.

Incorporating side information. The use of side information has also been studied in Hannah et al. (2010), Hao et al. (2020), Bertsimas and Kallus (2020), Ban et al. (2019), and Ho and Hanasusanto (2019), to address different optimization problem. Our paper is the first to incorporate information in the joint pricing and production problem.

Notation. We denote by $[I] \triangleq \{1, 2, \ldots, I\}$ the set of positive indices up to I. We use boldface glyphs, such as $\boldsymbol{x} \in \mathbb{R}^I$ and $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ to denote vectors and matrices; we denote by x_i the ith element of vector $\boldsymbol{x}$ and by A_{ij} the element located in the ith row and jth column of matrix $\boldsymbol{A}$. As usual, $|\boldsymbol{x}| = (|x_1|, \ldots, |x_I|)$ signifies the absolute value of $\boldsymbol{x}$. Special vectors of the appropriate dimension include $\mathbf{0}$ and $\mathbf{1}$, which correspond to (respectively) the vector of 0s and the vector of 1s. We use $\tilde{\boldsymbol{v}} \sim \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^I)$ to denote an I-dimensional random variable $\tilde{\boldsymbol{v}}$ governed by a probability distribution $\mathbb{P}$, where $\mathcal{P}_0(\mathbb{R}^I)$ represents the set of all probability distributions in $\mathbb{R}^I$. For a set $\mathcal{V} \subseteq \mathbb{R}^I$, the term $\mathbb{P}[\tilde{\boldsymbol{v}} \in \mathcal{V}]$ represents the probability of $\tilde{\boldsymbol{v}}$ lying in the set $\mathcal{V}$ evaluated on the distribution $\mathbb{P}$. For a probability distribution $\mathbb{P}$, we use $\mathbb{E}_{\mathbb{P}}[\cdot]$ to signify the corresponding expectation. For two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^I$, the expression $\boldsymbol{x} \geq \boldsymbol{y}$ means that $\boldsymbol{x}$ is *component-wise* no less than $\boldsymbol{y}$. Finally, superscripts 1 and 2 always stand for (respectively) period 1 and period 2.

2. The pricing and production problem

We consider the pricing and production problem faced by a manager selling I items in a market with price-sensitive consumers. At the beginning of the first period, the manager manufactures an amount $\boldsymbol{x} \in \mathcal{X} \subseteq [\boldsymbol{0}, \bar{\boldsymbol{x}}]$ of goods at the marginal cost $\boldsymbol{c}$, where $\bar{\boldsymbol{x}}$ is the production capacity; at the same time, this manager also quotes a *retail* price $\boldsymbol{r}^1 \in \mathcal{R}$, where $\mathcal{R}$ is a set of price candidates. The demand in the first period, denoted by $\boldsymbol{z}^1(\boldsymbol{r}^1, \boldsymbol{v}^1)$, is a function of $\boldsymbol{r}^1$ and some realized uncertainty $\boldsymbol{v}^1$ at the end of the first period. With generated revenue for the first period of

$$\pi^1(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{v}^1) \triangleq \sum_{i \in [I]} \Big(r_i^1 \left[\min\{x_i, z_i^1(\boldsymbol{r}^1, \boldsymbol{v}^1)\} \right] \Big),$$

the manager decides on a *markdown* price, $\boldsymbol{r}^2 \leq \boldsymbol{r}^1$ at the beginning of the second period. During this period, the second-period demand, denoted by $\boldsymbol{z}^2(\boldsymbol{r}^1, \boldsymbol{r}^2, \boldsymbol{v}^2)$, is determined by the price decisions, as well as some realized *uncertainty* $\boldsymbol{v}^2$ at the end of the second period. We assume that *no* goods are produced during the second period and that any products left over after fulfilling the second-period demand are salvaged—without any disposal cost—at zero value. This can easily be extended to non-zero salvage cost, but we omitted this for ease of exposition. As such, the revenue in the second period is

$$\pi^2(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}^2, \boldsymbol{v}^1, \boldsymbol{v}^2) \triangleq \sum_{i \in [I]} r_i^2 \min\left\{ \left(x_i - z_i^1(\boldsymbol{r}^1, \boldsymbol{v}^1)\right)^+, z_i^2(\boldsymbol{r}^1, \boldsymbol{r}^2, \boldsymbol{v}^2) \right\},$$

and the total revenue is then expressed as

$$\pi(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}^2, \boldsymbol{v}^1, \boldsymbol{v}^2) = \pi^1(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{v}^1) + \pi^2(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}^2, \boldsymbol{v}^1, \boldsymbol{v}^2).$$

Note that, even though the inventory holding cost is assumed to be zero, this model can easily be extended to settings with nonzero holding costs (see, for example, Cachon and Kök 2007, Chu et al. 2018). We assume that the period boundaries are easy to demarcate. Yet some products even selling in the same store may not have the same two-period boundaries. In this case, we can approximately assign products with similar boundaries into one group and adopt our later proposed distributionally robust optimization model to obtain optimal policies for each group of products, though this may sacrifice certain degree of the model accuracy.

We remark here that retail analytics typically focuses on estimating salient demand information such as averages while discarding information on prediction errors. We emphasize that the uncertainty in demand predictions plays an important role in how well we can optimize the profit. For instance, in the case of the newsvendor problem, a deterministic model would result in an order level exactly at the predicted demand, which would be suboptimal if deviation arises. Hence, quite apart from the retail analytics literature, it is important to also incorporate uncertainty in demand predictions to better address the production and pricing problem.

A stochastic optimization model and it challenges

One commonly used approach to address the joint pricing and production problem is the stochastic dynamic optimization method (see, for example Federgruen and Heching 1999, Cachon and Kök 2007) where the corresponding model is as follows

$$\max_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{r}^1\in\mathcal{R}} -\boldsymbol{c}'\boldsymbol{x}+\mathbb{E}_{\mathbb{P}^\dagger}\left[\pi^1(\boldsymbol{x},\boldsymbol{r}^1,\tilde{\boldsymbol{v}}^1)+\phi(\boldsymbol{x},\boldsymbol{r}^1,\tilde{\boldsymbol{v}}^1)\right], \tag{1}$$

where

$$\phi(\boldsymbol{x},\boldsymbol{r}^1,\boldsymbol{v}^1)=\max_{\boldsymbol{r}^2\in\mathcal{R},\boldsymbol{r}^2\leq\boldsymbol{r}^1}\mathbb{E}_{\mathbb{P}^\dagger}\left[\pi^2(\boldsymbol{x},\boldsymbol{r}^1,\boldsymbol{r}^2,\tilde{\boldsymbol{v}}^1,\tilde{\boldsymbol{v}}^2)\;\middle|\;\tilde{\boldsymbol{v}}^1=\boldsymbol{v}^1\right], \tag{2}$$

for some given probability distribution, $\mathbb{P}^\dagger\in\mathcal{P}_0(\mathbb{R}^I\times\mathbb{R}^I)$, $(\tilde{\boldsymbol{v}}^1,\tilde{\boldsymbol{v}}^2)\sim\mathbb{P}^\dagger$. The stochastic optimization model, however, presents the following challenges.

1. **Distributional ambiguity.** The premise of the stochastic optimization model assumes that the joint demand distribution for all items and across periods is precisely given, which is impractical. Without using the true demand distribution, stochastic optimization models suffer from the "optimizer's curse" and may yield solutions that perform poorly in practice (Smith and Winkler 2006). Although we may fit the demand distribution to a family of distributions whose parameters can be estimated from data, say via Kernel density functions (see e.g. Strijbosch and Heuts 1992, Bertschek and Kaiser 2004), the actual demand distribution is unlikely to be the same. Realistically, we should also recognize the challenges of having sufficient data to obtain reasonably good estimations of these parameters. For instance, for a one-year record, if the planning horizon is a week, then we would only have at most 52 samples to estimate the parameters of the joint demand distribution. The problem is exacerbated if stochastic independence across periods is not assumed; to evaluate Problem (2), one would require to determine the conditional probability distribution of $\tilde{\boldsymbol{v}}^2$ given $\tilde{\boldsymbol{v}}^1=\boldsymbol{v}^1$. This poses a problem in a data-driven setting because, in practice, the data available is only a minuscule fraction of the sample space of $\tilde{\boldsymbol{v}}^1$, which can be exponential in the size or even infinite. Hence, given the difficulties of estimation, we should not ignore the issues associated with having demand ambiguity when solving the pricing and production problem.

2. **Computational challenges.** Even granted the knowledge of the demand distribution, to determine the risk-neutral objective function of the stochastic optimization problem, it entails multi-dimensional integration, which is a computationally excruciating and intractable problem (Nemirovski and Shapiro 2007). Moreover, solving the stochastic optimization model may demand an exhaustive search to obtain the optimal pricing and production strategies, which is exacerbated by the lack of tractable computational structures such as convexity even for the one product case when demand is linearly dependent on retail price (see, for example, Petruzzi and Dada 2001, Mitra 2018).

We propose a new distributionally robust optimization model that would address these issues collectively. Specifically, we first characterize the demand distribution using a new two-period moment-based ambiguity set, for which its parameters can be mapped from data. The ambiguity set partitions the random outcomes in the first period into a small number of clusters, which permits more economical use of data for characterizing distribution ambiguity in the second period. In contrast to stochastic optimization, we can tractably evaluate the worst-case expectation of the objective function over this ambiguity set. Moreover, the distributionally robust optimization approach is also effective in mitigating the effect of "optimizer's curse". In fact, in the out-of-sample evaluation, our numerical studies show that the robust optimization models would perform better than a risk-neutral stochastic optimization model for which the objective under the empirical distribution is minimized. This should allay the fear that the robust optimization model would lead to over-conservative solutions. Finally, to address the computational challenges, we propose effective approximations that enable us to formulate the distributionally robust optimization as a deterministic mixed-integer linear optimization model that we can solve using state-of-the-art solvers the likes of CPLEX and Gurobi.

3. Demand model and ambiguity set

In the field of pricing and production management, the literature has introduced various demand models (see e.g. Federgruen and Heching 1999, Gong et al. 2014). Inspired by the ubiquitous linear regression model, we introduce the *additive demand model* (see e.g. Lu et al. 2014, Bernstein et al. 2015) as follows:

$$\boldsymbol{z}^1(\boldsymbol{r}^1, \tilde{\boldsymbol{v}}^1) \triangleq \boldsymbol{a}^1 + \boldsymbol{B}^1\boldsymbol{r}^1 + \tilde{\boldsymbol{v}}^1,$$
$$\boldsymbol{z}^2(\boldsymbol{r}^1, \boldsymbol{r}^2, \tilde{\boldsymbol{v}}^2) \triangleq \boldsymbol{a}^2 + \boldsymbol{B}^2\boldsymbol{r}^2 + \boldsymbol{B}^{21}\boldsymbol{r}^1 + \tilde{\boldsymbol{v}}^2.$$

Here, $\boldsymbol{a}^1$, $\boldsymbol{B}^1$, $\boldsymbol{a}^2$, $\boldsymbol{B}^2$, and $\boldsymbol{B}^{21}$ are parameters of the additive demand model that we assumed could be reasonably estimated from data. As we will later expound, the residuals in the two periods $(\tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2)$ would play an important role in characterizing the distributional ambiguity of the demands. Observe here that the demand of an item in each period is a bijective function of its residual.

We note that various demand models beyond additive ones could also be found in the literature and refer readers to Talluri and Van Ryzin (2006) for an overview of multiproduct demand functions that are typically used in retail price optimization. We can, in fact, consider polynomial, say, quadratic demand model, but this would necessarily lead to a more complex mathematical optimization problem. In Online Appendix B, we show how the results in the paper can be extended to a more general demand model where each product is concave piecewise linear in its retail price (see for example Hobbs and Pang 2007), and we can do so without significantly increasing the modeling complexity compared to the affine demand model.

In the era of big data, Bradlow et al. advocate the needs to exploit the vast flow of information that avails in retail analytics to improve prediction and decision-making. Consequently, we can extend the additive demand model to incorporate other factors or independent variables, encapsulated in the vector $\boldsymbol{\zeta}$, such as holidays and seasonality, which are known parameters at the beginning of the production period. The generalized demand models would be

$$\boldsymbol{z}^1(\boldsymbol{r}^1, \tilde{\boldsymbol{v}}^1) \triangleq \boldsymbol{a}^1(\boldsymbol{\zeta}) + \boldsymbol{B}^1(\boldsymbol{\zeta})\boldsymbol{r}^1 + \tilde{\boldsymbol{v}}^1,$$
$$\boldsymbol{z}^2(\boldsymbol{r}^1, \boldsymbol{r}^2, \tilde{\boldsymbol{v}}^2) \triangleq \boldsymbol{a}^2(\boldsymbol{\zeta}) + \boldsymbol{B}^2(\boldsymbol{\zeta})\boldsymbol{r}^2 + \boldsymbol{B}^{21}(\boldsymbol{\zeta})\boldsymbol{r}^1 + \tilde{\boldsymbol{v}}^2,$$

where the dependency of parameters on $\boldsymbol{\zeta}$ can be determined by more sophisticated statistical or machine learning techniques. For a given $\boldsymbol{\zeta}$, the demand model retains the affine price dependency, which is necessary for us to obtain an explicit mathematical optimization formulation. Since we can embed this model into our framework without changing the results, for notational convenience, we would leave out the dependency of $\boldsymbol{\zeta}$ in the subsequent presentation.

The addictive demand model is rich enough to capture both cross-product and cross-period effects, as we discuss below.

- *Cross-product effects:* In a multi-item setting, the demands for different products are often correlated because of substitutability and/or complementarity among products (Shen and Su 2007). For products that are substitutes, increasing the price of one product might lead to increased demand for its substitute; the converse holds for complementary products. In the case of product i, for example,

$$z_i^1(\boldsymbol{r}^1, \tilde{\boldsymbol{v}}^1) = a_i^1 + B_{ii}^1 r_i^1 + \sum_{m \neq i} B_{mi}^1 r_m^1 + \tilde{v}_i^1,$$

if $B_{mi}^1 > 0$ then product m is a substitute for product i whereas, if $B_{mi}^1 < 0$, then product m is complementary to product i (see Kuyumcu and Popescu 2006, Schlapp and Fleischmann 2018). In Section 5, we use a real-world data set to demonstrate the presence of cross-product effects at both the SKU level and the aggregate level across stores. One can also find related examples in Anupindi et al. (1998), Russell and Petersen (2000).

- *Cross-period effects:* In a two-period model, the demand for the products in the second period may be affected by prices in the first period (see e.g. Cohen et al. 2017). So for product i we have

$$z_i^2(\boldsymbol{r}^1, \boldsymbol{r}^2, \tilde{\boldsymbol{v}}^2) = a_i^2 + B_{ii}^2 r_i^2 + \sum_{m \neq i} B_{mi}^2 r_m^2 + \sum_{n \in [I]} B_{ni}^{21} r_n^1 + \tilde{v}_i^2,$$

where a larger B_{ni}^{21} indicates a stronger effect of product n's first-period price on the second-period demand for product i (see e.g. Popescu and Wu 2007, Chen et al. 2016). In other words, the cross-period effects are essentially related to a psychological aspect of pricing—reference effects (see for example Greenleaf 1995, Fibich et al. 2003, Kopalle et al. 2012, Hu et al. 2016), where in our model the first-period price is the reference price. We remark that most multi-product demand models in the literature do not explicitly characterize cross-product effects jointly with cross-period effects. We also note that incorporating cross-period effects into the second-period demand model might result in the issue of multicollinearity in the multivariate regression, which may be addressed using, *inter alia*, ridge regression, principal components analysis, or fitting the model using standardized retail prices while subtracting the mean retail prices (Kutner et al. 2005, p431).

Our focus in this paper is to provide a prescriptive analytics framework for profit maximization that is compatible with additive demand models. Nonetheless, we acknowledge the econometric challenges in estimating the parameters of the additive demand model that may arise in retail analytics. Among other things, if only sales data is available, the demand as reflected by the sales would be "censored" by the inventory level. Incidentally, various ways of improving parameters estimations in censored demand have been explored in the literature (see, for example, Wecker 1978, Nahmias 1994, Stefanescu 2009, Jain et al. 2015, Mersereau 2015, Amjad and Shah 2017, Ban 2020). We also refer readers to Section 9.4 in Talluri and Van Ryzin (2006) for an overview of some common methods; some of these ideas are extended in Anupindi et al. (1998), Vulcano et al. (2012), and Musalem et al. (2010) to estimate lost sales for multiple products. Endogeneity is also a common econometric issue, which critically depends on our understanding of the sources of it (Bradlow et al. 2017). We refer readers to Chapter 15-16 of Wooldridge (2016) for a detailed discussion of endogeneity. Incidentally, a popular econometric approach of having instrumental variables would also compatible with the additive demand model.

Information Paritioning

Quite apart from how demand uncertainty is being modeled in the literature, inspired by both machine learning and distributionally robust optimization, we present an information partitioning approach to model the ambiguity of demand distributions. Often, demand may be affected by external factors or side information such as sales data, promotion, and even weather (see, for instance, Steele 1951, Bradlow et al. 2017, Martínez-de Albéniz and Belkaid 2021). Intuitively, different level of side information leads to different demand; For example, demands in rainy days deviate from that in non-rainy days (Martínez-de Albéniz and Belkaid 2021), or demands in holiday are usually higher. As such, the price can be set in the way that it adapts to scenarios such as rainy days and non-rainy days, holidays and non-holidays. Accordingly, one can partition the information via clustering methods in each period into different scenarios. Conceivably, incorporating side information as such to the demand model may help the planner to better predict demand and improve her decision making.

We can partition the information using clustering methods commonly used in demand prediction. For example, Kumar et al. (2002) uses clustering methods to accurate predict demands with seasonality patterns; Ferreira et al. (2016) uses regression tree to predict demand for an online retailer; Thomassey and Fiordaliso (2006) use both K-means clustering and decision tree to predict demand for textile market; see Jain (2010) for a review of clustering methods. In this paper, we propose a two-period information partitioning model that serves two purposes—it reduces the information requirement, which facilitates efficient computation, and determines how the second-period demand uncertainty would evolve from the given first-period information partition, which can be applied to non-anticipative decision making.

Our premise stems from a data-driven setting where the producer has access to a set of H historical samples, each comprising two-period realized demands alongside the information pertinent to the prediction of demands. Specifically, each sample $h \in [H]$ comprises of two periods of retail prices, $(\boldsymbol{r}_h^1, \boldsymbol{r}_h^2)$, realized demands, $(\boldsymbol{z}_h^1, \boldsymbol{z}_h^2)$, and realized information, $(\boldsymbol{w}_h^1, \boldsymbol{w}_h^2)$ that are associated with demand forecasts. As a generalization, in the h sample, $(\boldsymbol{w}_h^1, \boldsymbol{w}_h^2)$ may include the realized demand residuals of $(\tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2)$, which we denote by $(\boldsymbol{v}_h^1, \boldsymbol{v}_h^2)$. We first partition the historical first-period information, $\boldsymbol{w}_h^1$, $h \in [H]$, into S clusters, each is associated with a polyhedral support set, $\mathcal{W}_s^1 \subseteq \mathbb{R}^J$, $s \in [S]$, which is not necessarily closed and does not intersect with other support sets so that sample h is associated with the sth cluster, if and only if $\boldsymbol{w}_h^1 \in \mathcal{W}_s^1$. The union of the support sets also satisfies $\bigcup_{s\in[S]} \mathcal{W}_s^1 = \mathcal{W}^1$, where $\mathcal{W}^1$ is the support of the random variable $\tilde{\boldsymbol{w}}^1$. Subsequently, for all the samples associated with a first-period cluster, $s \in [S]$, *i.e.*,

$$\mathcal{H}_s \triangleq \left\{ h \in [H] \mid \boldsymbol{w}_h^1 \in \mathcal{W}_s^1 \right\},$$

we cluster the historical second-period information $\boldsymbol{w}_h^2$, $h \in \mathcal{H}_s$ into T clusters, with each being associated with a non-overlapping polyhedral support set, $\mathcal{W}_{st}^2 \subseteq \mathbb{R}^J$, $t \in [T]$. Similarly, we also define

$$\mathcal{H}_{s,t} \triangleq \left\{ h \in [H] \mid \boldsymbol{w}_h^1 \in \mathcal{W}_s^1, \boldsymbol{w}_h^2 \in \mathcal{W}_{st}^2 \right\}.$$

Hence, in total, there will be $S \times T$ clusters in the second period, which is upper bounded by H to ensure each support set $\mathcal{W}_{st}^2$ contains at least one empirical sample. We let $\mathcal{W}_s^2 \triangleq \bigcup_{t\in[T]} \mathcal{W}_{st}^2$ and denote $\mathcal{W}^2 \triangleq \bigcup_{s\in[S]} \mathcal{W}_s^2$ as the support of the random variable $\tilde{\boldsymbol{w}}^2$. As we will further show, the number of first-period clusters, S also relates to how adaptive the recourse actions would be in response to uncertain outcomes in the first period, with each recourse action being associated with the cluster that contains the realized first-period information. Moreover, observe that conditional on the realization of the first-period information in the sth cluster, $s \in [S]$, the sample size associated with the second-period information would satisfy

$$\min_{s\in[S]} |\mathcal{H}_s| \leq H/S.$$

Hence, the larger the number of first-period clusters, S, the less empirical information we may have to determine the distribution of the second-period uncertainty. In our computational studies, we also observe the diminishing benefits as the number of clusters, S increases. In practice, S and T may not need to be large to achieve reasonable performance that we can obtain the solutions within acceptable computational limits. Note that while we recognize the challenges of obtaining the optimal number of clusters, S and T, it is worthwhile to consider cross-validation methods for this purpose (see, for instance, Section 7.10 of Hastie et al. 2009). We note that there are different clustering approaches (see, *e.g.,* Jain 2010), and for simplicity, we will focus on K-means clustering and regression tree.

Example 1 (K-means clustering). K-means clustering aims to partition the historical data into k clusters and assigns each observation in the data to the cluster with the nearest mean. This is achieved via minimizing the squared Euclidean distances. Specifically, to implement the K-means clustering, we first normalize the first-period historical information $\{\boldsymbol{w}_h^1\}_{h\in[H]}$, noting that the demand residual $\boldsymbol{v}_h^1$ may be embedded within $\boldsymbol{w}_h^1$ and it may have different scale with other side information embedded within $\boldsymbol{w}_h^1$ as well; As a special case, in the absence of side information, we would have $\boldsymbol{w}_h^1 = \boldsymbol{v}_h^1$—As we will demonstrate in the computational studies, even in the absence of side information, with the information $\tilde{\boldsymbol{w}}^\tau$ being $\tilde{\boldsymbol{v}}^\tau$, $\tau \in \{1,2\}$, we can also obtain improvement in solutions when addressing the two-period pricing and production problem. Subsequently, the K-means clustering heuristics partitions the data into S labeled clusters, where each cluster s is associated with the subset of samples $\mathcal{H}_s$. The K-means clustering heuristics aims to minimize the sum of squared errors:

$$\min_{(\mathcal{H}_1,\ldots,\mathcal{H}_S)\in\Pi(S)} \sum_{s\in[S]} \sum_{h\in\mathcal{H}_s} \left\| \boldsymbol{\Sigma}\Big(\boldsymbol{w}_h^1 - \frac{1}{|\mathcal{H}_s|} \sum_{h\in\mathcal{H}_s} \boldsymbol{w}_h^1 \Big) \right\|_2^2,$$

where $\boldsymbol{\Sigma}$ is a normalization matrix and $\Pi(S)$ is a set of all possible partitions of $\{1,2,\ldots,H\}$ with S clusters. Likewise, given each cluster s in the first period, we further partition the normalized $(\boldsymbol{w}_h^2)$, $h\in\mathcal{H}_s$ into T labeled clusters. We illustrate how we use K-means clustering for a two-product, two-period problem with $S=3$ and $T=2$ in Figure 1.

Example 2 (Regression tree). One can also use the regression tree for the case of supervised learning to obtain the partition. Regression tree is a widely used non-parametric method to obtain partitions, see Breiman et al. (1984), De′Ath and Fabricius (2000) for detailed introduction of univariate regression tree, and De′Ath (2002) for multivariate regression tree. To implement the regression tree, we first specify the tree structure by a set of splitting rules, usually hyperplanes, then partition the space of the information depending on how well the demand residuals can be classified. Specifically, we recursively split $\mathcal{W}^1$ and its subsets into two further subsets until we

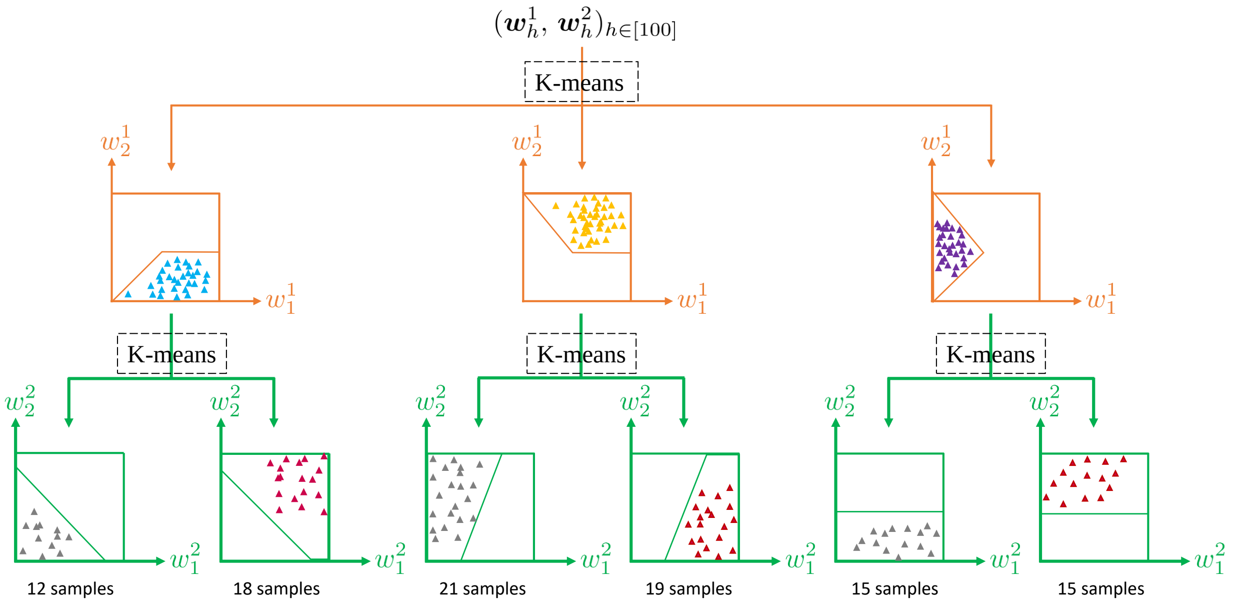


Figure 1 The historical data $(\boldsymbol{w}_h^1,\ \boldsymbol{w}_h^2)_{h\in[100]}$ is partitioned into three clusters based on the K-means clustering of $\boldsymbol{w}^1$. Given each cluster in the first period, we further partition the corresponding second-period data into two clusters and display the support of the second-period information respectively.

obtain a tree with S leaf nodes. This tree is determined by minimizing the sum of squared errors of the information vectors:

$$\min_{(\mathcal{H}_1,\ldots,\mathcal{H}_S)\in\text{tree}(S)} \sum_{s\in[S]}\sum_{h\in\mathcal{H}_s}\left\|\boldsymbol{w}_h^1-\frac{1}{|\mathcal{H}_s|}\sum_{h\in\mathcal{H}_s}\boldsymbol{w}_h^1\right\|_2^2,$$

where tree(S) is a set of all trees with S leaf nodes and different tree corresponds with different $\mathcal{H}_s, s\in[S]$.

Each leaf node represents a subset, which together form a partition of $\mathcal{W}^1$. The historical data set, that is, $(\boldsymbol{w}_h^1,\boldsymbol{w}_h^2)_{h\in[H]}$, then falls into S labeled clusters according to the tree, where each cluster s is associated with the subset of samples $\mathcal{H}_s$. Similarly, given each cluster s in the first period, we further split $\mathcal{W}_s^2$ so that $\boldsymbol{w}_h^2,\ h\in\mathcal{H}_s$ falls into T labeled clusters.

Partitioned-moment-based ambiguity set.

Given the partition of the demand residuals, that is, $(\boldsymbol{v}_h^1,\boldsymbol{v}_h^2)_{h\in\mathcal{H}_s}, s\in[S]$, for each cluster, we define a moment-based ambiguity set that matches the descriptive statistics such as the support, mean and deviation of the empirical cluster-wise demand residuals. The empirical cluster-wise residuals means associated with the cluster are determined as

$$\boldsymbol{\mu}_s^1=\frac{1}{|\mathcal{H}_s|}\sum_{h\in\mathcal{H}_s}\boldsymbol{v}_h^1,\quad \boldsymbol{\mu}_{st}^2=\frac{1}{|\mathcal{H}_{st}|}\sum_{h\in\mathcal{H}_{st}}\boldsymbol{v}_h^2.$$

Correspondingly, we also characterize the mean absolute deviations of the empirical cluster-wise residuals associated as follows:

$$\boldsymbol{\sigma}_s^1 = \frac{1}{|\mathcal{H}_s|} \sum_{h \in \mathcal{H}_s} |\boldsymbol{v}_h^1 - \boldsymbol{\mu}_s^1|, \quad \boldsymbol{\sigma}_{st}^2 = \frac{1}{|\mathcal{H}_{st}|} \sum_{h \in \mathcal{H}_{st}} |\boldsymbol{v}_h^2 - \boldsymbol{\mu}_{st}^2|.$$

To preserve linearity, the support sets of $\tilde{\boldsymbol{v}}^1$ and $\tilde{\boldsymbol{v}}^2$, given respectively as $\mathcal{V}_s^1$ and $\mathcal{V}_{st}^2$, are restricted to polyhedra.

Consequently, we characterize the demand residuals via the following *partitioned-moment-based ambiguity set:*

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^I \times \mathbb{R}^I \times \mathbb{R}^J \times \mathbb{R}^J) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2, \tilde{\boldsymbol{w}}^1, \tilde{\boldsymbol{w}}^2) \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{v}}^1 \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}_s^1] = \boldsymbol{\mu}_s^1 & \forall s \in [S] \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{v}}^2 \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}_s^1, \tilde{\boldsymbol{w}}^2 \in \mathcal{W}_{st}^2] = \boldsymbol{\mu}_{st}^2 & \forall (s,t) \in [S] \times [T] \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{v}}^1 - \boldsymbol{\mu}_s^1| \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}_s^1] \le \boldsymbol{\sigma}_s^1 & \forall s \in [S] \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{v}}^2 - \boldsymbol{\mu}_{st}^2| \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}_s^1, \tilde{\boldsymbol{w}}^2 \in \mathcal{W}_{st}^2] \le \boldsymbol{\sigma}_{st}^2 & \forall (s,t) \in [S] \times [T] \\ \mathbb{P}[\tilde{\boldsymbol{v}}^1 \in \mathcal{V}_s^1 \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}_s^1] = 1 & \forall s \in [S] \\ \mathbb{P}[\tilde{\boldsymbol{v}}^2 \in \mathcal{V}_{st}^2 \mid \tilde{\boldsymbol{w}}^2 \in \mathcal{W}_{st}^2] = 1 & \forall (s,t) \in [S] \times [T] \\ \mathbb{P}[\tilde{\boldsymbol{w}}^1 \in \mathcal{W}_s^1, \tilde{\boldsymbol{w}}^2 \in \mathcal{W}_{st}^2] = q_{st} & \forall (s,t) \in [S] \times [T] \end{array} \right\},$$

where $q_{st} = |\mathcal{H}_{st}|/H$. The first two sets of moment equalities anchor the mean values of the demand residuals within each cluster of the information partition, and likewise the subsequent two sets of moments inequalities provide upper bounds to the mean absolute deviations of the cluster-wise demand residuals. The third and fourth sets of probability constraints specify the support sets of the cluster-wise demand residuals, while the last set of probability constraints define each cluster's probability. It is worth noting that the partitioned-moment-based ambiguity set does not aim to predict demand but to characterize its demand ambiguity. Together with the cluster-adapted markdown policy proposed later, it can facilitate efficient computation and determine how the second-period price adapting to the first-period information partition. Besides, the ambiguity set generalizes traditional moment-based ambiguity set and the set only contains the empirical distribution, as shown in the following proposition.

Proposition 1. *Suppose $S = 1$ and $T = 1$, the ambiguity set $\mathcal{F}$ projected on the residuals $\tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2$ is a moment-based ambiguity set given by*

$$\prod_{\tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2} \mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^I \times \mathbb{R}^I) \;\middle|\; \begin{array}{l} (\tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2) \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{v}}^1] = \boldsymbol{\mu}_1^1 \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{v}}^2] = \boldsymbol{\mu}_{11}^2 \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{v}}^1 - \boldsymbol{\mu}_1^1|] \le \boldsymbol{\sigma}_1^1 \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{v}}^2 - \boldsymbol{\mu}_{11}^2|] \le \boldsymbol{\sigma}_{11}^2 \\ \mathbb{P}[\tilde{\boldsymbol{v}}^1 \in \mathcal{V}_1^1, \tilde{\boldsymbol{v}}^2 \in \mathcal{V}_{11}^2] = 1 \end{array} \right\}.$$

Suppose $S = H$ *and* $T = 1$*, then the ambiguity set* $\mathcal{F}$ *projected on the residuals only contains the empirical distribution, that is,* $\prod_{\tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2} \mathcal{F} = \{\hat{\mathbb{P}}\}$*, where*

$$\hat{\mathbb{P}}[\tilde{\boldsymbol{v}}^1 = \boldsymbol{v}_h^1, \tilde{\boldsymbol{v}}^2 = \boldsymbol{v}_h^2] = 1/H \qquad \forall h \in [H].$$

Proof. The case for $S = 1$, $T = 1$ follows trivially. For the case with $S = H$ and $T = 1$, observe that since each cluster contains exactly one sample, we have $\boldsymbol{\mu}_s^1 = \boldsymbol{v}_s^1$, $\boldsymbol{\mu}_{s1}^2 = \boldsymbol{v}_s^2$, $\boldsymbol{\sigma}_s^1 = \boldsymbol{\sigma}_{s1}^2 = \mathbf{0}$ and $q_{s1} = 1/H$ for all $s \in [H]$. Hence, the corresponding partitioned-moment-based ambiguity set only contains the empirical distribution. ■

When there is only one cluster in both periods, the partitioned-moment-based ambiguity set becomes the moment-based ambiguity set (see, for example, Wiesemann et al. 2014, Bertsimas et al. 2019). With $S = H$ and $T = 1$, then each sample will be embedded in one cluster and consequently, only the empirical distribution is characterized by the partitioned-moment-based ambiguity set. Hence, from Proposition 1, we see that this ambiguity set allows us to model a rich variety of structural information about random demand, and to adjust the level of robustness that varies depending on the number of clusters, from the standard moment-based ambiguity set with one cluster to the empirical distribution being the least robust. The partitioned-moment-based ambiguity set also captures demand correlations across periods—unlike most stochastic and dynamic optimization models, which often impose the assumption of independence (see, for example, Federgruen and Heching 1999, Yang et al. 2014).

We remark here that in the absence of side information, one can also cluster over the demand residuals via K-means, which would be suitable to characterize an underlying demand model with mixture distribution, as observed in Vaagen and Wallace (2008) and Hanasusanto et al. (2015a). The support sets associated with these clusters are polyhedra described formally as

$$\mathcal{V}_s^1 = \left\{ \boldsymbol{v}^1 \mid 2(\boldsymbol{v}^1)'(\boldsymbol{\mu}_l^1 - \boldsymbol{\mu}_s^1) \leq (\boldsymbol{\mu}_l^1)'\boldsymbol{\mu}_l^1 - (\boldsymbol{\mu}_s^1)'\boldsymbol{\mu}_s^1 \ \forall l \in [S] \right\}$$

and

$$\mathcal{V}_{st}^2 = \left\{ \boldsymbol{v}^2 \mid 2(\boldsymbol{v}^2)'(\boldsymbol{\mu}_{sl}^2 - \boldsymbol{\mu}_{st}^2) \leq (\boldsymbol{\mu}_{sl}^2)'\boldsymbol{\mu}_{sl}^2 - (\boldsymbol{\mu}_{st}^2)'\boldsymbol{\mu}_{st}^2 \ \forall l \in [T] \right\},$$

where $\boldsymbol{\mu}_s^1$ is the centroid of cluster $s \in [S]$ and $\boldsymbol{\mu}_{st}^2$ is the centroid of cluster $(s,t) \in [S] \times [T]$.

Observe that this is a special case when the information contains only the demand residuals, that is, $\tilde{\boldsymbol{w}}^\tau = \tilde{\boldsymbol{v}}^\tau, \tau \in \{1,2\}$ and $\mathcal{W}_s^1 = \mathcal{V}_s^1, \mathcal{W}_{st}^2 = \mathcal{V}_{st}^2$, for all $(s,t) \in [S] \times [T]$, and the partitioned-moment-based ambiguity set would be simplified as

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^I \times \mathbb{R}^I) \left| \begin{array}{ll} (\tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2) \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{v}}^1 \mid \tilde{\boldsymbol{v}}^1 \in \mathcal{V}_s^1] = \boldsymbol{\mu}_s^1 & \forall s \in [S] \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{v}}^2 \mid \tilde{\boldsymbol{v}}^1 \in \mathcal{V}_s^1, \tilde{\boldsymbol{v}}^2 \in \mathcal{V}_{st}^2] = \boldsymbol{\mu}_{st}^2 & \forall (s,t) \in [S] \times [T] \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{v}}^1 - \boldsymbol{\mu}_s^1| \mid \tilde{\boldsymbol{v}}^1 \in \mathcal{V}_s^1] \le \boldsymbol{\sigma}_s^1 & \forall s \in [S] \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{v}}^2 - \boldsymbol{\mu}_{st}^2| \mid \tilde{\boldsymbol{v}}^1 \in \mathcal{V}_s^1, \tilde{\boldsymbol{v}}^2 \in \mathcal{V}_{st}^2] \le \boldsymbol{\sigma}_{st}^2 & \forall (s,t) \in [S] \times [T] \\ \mathbb{P}[\tilde{\boldsymbol{v}}^1 \in \mathcal{V}_s^1, \tilde{\boldsymbol{v}}^2 \in \mathcal{V}_{st}^2] = q_{st} & \forall (s,t) \in [S] \times [T] \end{array} \right. \right\}.$$

4. Adaptive distributionally robust optimization model

To circumvent the modeling and optimization challenges, we introduce a new distributionally robust optimization model to address the two-period, multi-item joint pricing and production problem, which can be implemented in a data-driven setting using historical demand and side information pertinent to the prediction of demands. To further facilitate the modeling of the problem while keeping it practicable, we first incorporate two widely used business rules into the demand model; in particular, we assume discrete prices and the existence of price markdowns. These features of the model, which are described next, have ramifications that enable us to reformulate the problem as a deterministic mixed-integer linear optimization problem.

Discrete price ladder It is a common practice in business to choose each product's price from a set of "admissible" prices (see e.g. Gallego and Van Ryzin 1994, Carboni Borrasé 2009, Cohen et al. 2017). Thus, for each product i in period τ, the price r_i^τ must be chosen from the set $\mathcal{R}_i = \{p_{i1}, \ldots, p_{iN}\}$; here $p_{i1} < \cdots < p_{iN}$ and $i \in [I]$. We shall put $\mathcal{R} = \prod_{i \in [I]} \mathcal{R}_i$.

Price markdown Another widely observed phenomenon in retail businesses is the price markdown—as applied by, among many others, Ann Taylor, Macy's, H&M, World Co., and Mango (Caro and Gallien 2012). Under this business "rule", if there are two selling periods then, for each product i, the second-period price (r_i^2) cannot exceed the first-period one (r_i^1). Hence, we may write $r_i^1 \ge r_i^2$ for $i \in [I]$. This business rule, as shown below in Proposition 2, will help us to reformulate the total revenue so as to obtain the exact model (5). Incidentally, allowing price markup would lead to the violation of the concavity of the revenue function with respect to the production quantity, leading to a significant escalation in the complexity of the optimization problem. Hence, we limit our attention to the price markdown case.

Proposition 2. *Under a markdown policy, the total revenue can be written as*

$$\pi(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}^2, \boldsymbol{v}^1, \boldsymbol{v}^2) = \sum_{i \in [I]} \min\{r_i^1 x_i,\ r_i^2 x_i + (r_i^1 - r_i^2) z_i^1(\boldsymbol{r}^1, v_i^1),\ r_i^1 z_i^1(\boldsymbol{r}^1, v_i^1) + r_i^2 z_i^2(\boldsymbol{r}^1, \boldsymbol{r}^2, v_i^2)\}.$$

Proof. Given the demand in both periods, the revenue derived from the sales of product-i is separable into three cases:

$$\begin{cases} r_i^1 x_i, & \text{if } 0 \leq x_i < z_i^1(\boldsymbol{r}^1, v_i^1) \\ r_i^2 x_i + (r_i^1 - r_i^2) z_i^1(\boldsymbol{r}^1, v_i^1), & \text{if } z_i^1(\boldsymbol{r}^1, v_i^1) \leq x_i < z_i^1(\boldsymbol{r}^1, v_i^1) + z_i^2(\boldsymbol{r}^1, \boldsymbol{r}^2, v_i^2) \\ r_i^1 z_i^1(\boldsymbol{r}^1, v_i^1) + r_i^2 z_i^2(\boldsymbol{r}^1, \boldsymbol{r}^2, v_i^2), & \text{otherwise.} \end{cases}$$

Observe that, regarding x_i, the gradients for the three cases are (respectively) r_i^1, r_i^2, and 0 such that $r_i^1 \geq r_i^2 \geq 0$. This is due to the markdown policy, where $r_i^1 - r_i^2 \geq 0$. Thus we can see that the revenue of product $i \in [I]$ is a concave and piecewise linear function on x_i. ■

Cluster-adapted markdown policy

A markdown policy, in the general setting, is a function of the information realized in the first period given by $\hat{\boldsymbol{r}}^2(\boldsymbol{w}^1) : \mathcal{W}^1 \mapsto \mathcal{R}(\boldsymbol{r}^1)$, where we define

$$\mathcal{R}(\boldsymbol{r}^1) \triangleq \left\{ \boldsymbol{r}^2 \in \mathcal{R} \mid \boldsymbol{r}^2 \leq \boldsymbol{r}^1 \right\},$$

as the set of feasible markdown prices. Recall that since $\tilde{\boldsymbol{v}}^1$ is embedded within the information $\tilde{\boldsymbol{w}}^1$, for notational simplicity, we do not need to include the realized residual $\boldsymbol{v}^1$ in the markdown policy. Based on our partitioned-moment-based ambiguity set, we proposed a more tractable markdown policy $\mathcal{M}$ that adapts to the realized information cluster in the first-period as follows:

$$\mathcal{M}(\boldsymbol{r}^1) \triangleq \left\{ \hat{\boldsymbol{r}}^2 : \mathcal{W}^1 \mapsto \mathcal{R}(\boldsymbol{r}^1) \,\middle|\, \begin{array}{ll} \hat{\boldsymbol{r}}^2(\boldsymbol{w}^1) = \boldsymbol{r}_s^2 \text{ if } \boldsymbol{w}^1 \in \mathcal{W}_s^1 & s \in [S] \\ \text{for some } \boldsymbol{r}_s^2 \in \mathcal{R}(\boldsymbol{r}^1) & \forall s \in [S] \end{array} \right\}.$$

Under the cluster-adapted markdown policy, the manager solves the following distributionally robust optimization problem:

$$Z^* = \max_{\boldsymbol{x} \in \mathcal{X},\, \hat{\boldsymbol{r}}^2 \in \mathcal{M}(\boldsymbol{r}^1),\, \boldsymbol{r}^1 \in \mathcal{R}} -\boldsymbol{c}'\boldsymbol{x} + \inf_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}} \left[\pi(\boldsymbol{x}, \boldsymbol{r}^1, \hat{\boldsymbol{r}}^2(\tilde{\boldsymbol{w}}^1), \tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2) \right]. \tag{3}$$

In Problem (3), the objective is to find the optimal production quantity and retail prices in order to maximize the worst-case expected profit over all possible distributions in ambiguity set $\mathcal{F}$. The partitioned-moment-based ambiguity set characterizes the moments of the cluster-wise demand residuals, while the cluster-adapted markdown policy embraces the dynamics across the two periods. As a result, quite apart from the dynamic programming approach as we have discussed in Section 2, which requires solving the problem sequentially, Problem (3) provides all the decisions simultaneously, and these decisions are directly implementable. We remark here that the multi-item newsvendor problem is a special case if $\boldsymbol{r}^1$ is exogenous and when we do not consider the second-period problem. We will test this special case in Section 5 to explore the impact of information partitioning.

Before we derive a practicable optimization model for Problem (3), in what follows, we first draw an equivalent nonlinear robust optimization model.

Proposition 3. *With respect to the cluster-adapted markdown policy, we can formulate Problem (3) as the following nonlinear robust optimization model:*

$$
\begin{array}{lll}
\max & -\boldsymbol{c}'\boldsymbol{x} + \displaystyle\sum_{(s,t)\in[S]\times[T]} \Big(\alpha_{st} + q_{st}(\boldsymbol{\beta}_s^1)'\boldsymbol{\mu}_s^1 + (\boldsymbol{\beta}_{st}^2)'\boldsymbol{\mu}_{st}^2 + q_{st}\ (\boldsymbol{\gamma}_s^1)'\boldsymbol{\sigma}_s^1 + (\boldsymbol{\gamma}_{st}^2)'\boldsymbol{\sigma}_{st}^2\Big) & \\
\text{s.t.} & \alpha_{st} + q_{st}(\boldsymbol{\beta}_s^1)'\boldsymbol{v}^1 + (\boldsymbol{\beta}_{st}^2)'\boldsymbol{v}^2 + q_{st}(\boldsymbol{\gamma}_s^1)'\boldsymbol{u}^1 + (\boldsymbol{\gamma}_{st}^2)'\boldsymbol{u}^2 & \\
 & \qquad \le q_{st}\pi(\boldsymbol{x},\boldsymbol{r}^1,\boldsymbol{r}_s^2,\boldsymbol{v}^1,\boldsymbol{v}^2) & \forall(\boldsymbol{v}^1,\boldsymbol{v}^2,\boldsymbol{u}^1,\boldsymbol{u}^2)\in\mathcal{L}_{st}, \\
 & & \quad (s,t)\in[S]\times[T]; \\
 & \boldsymbol{r}^1 \ge \boldsymbol{r}_s^2 & \forall s\in[S]; \\
 & \boldsymbol{\gamma}_s^1, \boldsymbol{\gamma}_{st}^2 \le 0 & \forall(s,t)\in[S]\times[T]; \\
 & \boldsymbol{r}^1, \boldsymbol{r}_s^2 \in \mathcal{R} & \forall s\in[S]; \\
 & \alpha_{st}\in\mathbb{R},\ \boldsymbol{\beta}_s^1, \boldsymbol{\beta}_{st}^2, \boldsymbol{\gamma}_s^1, \boldsymbol{\gamma}_{st}^2 \in \mathbb{R}^I & \forall(s,t)\in[S]\times[T]; \\
 & \boldsymbol{x}\in\mathcal{X}. &
\end{array}
\tag{4}
$$

Note that $\mathcal{L}_{st}$ *is the lifted support set defined as*

$$
\mathcal{L}_{st} = \left\{ (\boldsymbol{v}^1,\boldsymbol{v}^2,\boldsymbol{u}^1,\boldsymbol{u}^2) \in \mathbb{R}^I\times\mathbb{R}^I\times\mathbb{R}^I\times\mathbb{R}^I \;\middle|\; \begin{array}{l} \boldsymbol{v}^1\in\mathcal{V}_s^1,\ \boldsymbol{v}^2\in\mathcal{V}_{st}^2 \\ \boldsymbol{u}^1 \ge |\boldsymbol{v}^1-\boldsymbol{\mu}_s^1|,\ \boldsymbol{u}^2 \ge |\boldsymbol{v}^2-\boldsymbol{\mu}_{st}^2| \end{array} \right\}.
$$

Proof. See Online Appendix A.1. ■

The nonlinearity of $\pi(\boldsymbol{x},\boldsymbol{r}^1,\boldsymbol{r}_s^2,\boldsymbol{v}^1,\boldsymbol{v}^2)$ in Model (4) prevents us from obtaining the robust counterpart. Nevertheless, the following theorem gives an equivalent linear reformulation, which we refer to as the *exact model* (EXACT).

Theorem 1. *Problem (3) is equivalent the following robust optimization model:*

$$
\begin{array}{lll}
\max & -\boldsymbol{c}'\boldsymbol{x} + \displaystyle\sum_{(s,t)\in[S]\times[T]} \Big(\alpha_{st} + (\boldsymbol{\beta}_s^1)'\boldsymbol{\mu}_s^1 + (\boldsymbol{\beta}_{st}^2)'\boldsymbol{\mu}_{st}^2 + (\boldsymbol{\gamma}_s^1)'\boldsymbol{\sigma}_s^1 + (\boldsymbol{\gamma}_{st}^2)'\boldsymbol{\sigma}_{st}^2\Big) & \\
\text{s.t.} & \alpha_{st} + q_{st}(\boldsymbol{\beta}_s^1)'\boldsymbol{v}^1 + (\boldsymbol{\beta}_{st}^2)'\boldsymbol{v}^2 + q_{st}(\boldsymbol{\gamma}_s^1)'\boldsymbol{u}^1 + (\boldsymbol{\gamma}_{st}^2)'\boldsymbol{u}^2 & \\
 & \qquad \le q_{st}\phi(\mathcal{I}_1,\mathcal{I}_2,\mathcal{I}_3,\boldsymbol{x},\boldsymbol{r}^1,\boldsymbol{r}_s^2,\boldsymbol{v}^1,\boldsymbol{v}^2) & \forall(\boldsymbol{v}^1,\boldsymbol{v}^2,\boldsymbol{u}^1,\boldsymbol{u}^2)\in\mathcal{L}_{st}, \\
 & & \quad \mathcal{I}_1,\mathcal{I}_2,\mathcal{I}_3\in\mathcal{B}, \\
 & & \quad (s,t)\in[S]\times[T]; \\
 & \boldsymbol{r}^1 \ge \boldsymbol{r}_s^2 & \forall s\in[S]; \\
 & \boldsymbol{\gamma}_s^1, \boldsymbol{\gamma}_{st}^2 \le 0 & \forall(s,t)\in[S]\times[T]; \\
 & \boldsymbol{r}^1, \boldsymbol{r}_s^2 \in \mathcal{R} & \forall s\in[S]; \\
 & \alpha_{st}\in\mathbb{R},\ \boldsymbol{\beta}_s^1, \boldsymbol{\beta}_{st}^2, \boldsymbol{\gamma}_s^1, \boldsymbol{\gamma}_{st}^2 \in \mathbb{R}^I & \forall(s,t)\in[S]\times[T]; \\
 & \boldsymbol{x}\in\mathcal{X}. &
\end{array}
\tag{5}
$$

Note that

$$
\phi\left(\mathcal{I}_1,\mathcal{I}_2,\mathcal{I}_3,\boldsymbol{x},\boldsymbol{r}^1,\boldsymbol{r}_s^2,\boldsymbol{v}^1,\boldsymbol{v}^2\right)
$$

$$
\begin{aligned}
&\triangleq \sum_{i\in\mathcal{I}_1} r_i^1 x_i + \sum_{i\in\mathcal{I}_2}\Big(r_{si}^2 x_i + \left(r_i^1 - r_{si}^2\right)\left(a_i^1 + \boldsymbol{e}_i'\boldsymbol{B}^1\boldsymbol{r}^1 + v_i^1\right)\Big)\\
&\quad + \sum_{i\in\mathcal{I}_3}\Big(r_i^1\left(a_i^1 + \boldsymbol{e}_i'\boldsymbol{B}^1\boldsymbol{r}^1 + v_i^1\right) + r_{si}^2\left(a_i^2 + \boldsymbol{e}_i'\boldsymbol{B}^2\boldsymbol{r}_s^2 + \boldsymbol{e}_i'\boldsymbol{B}^{21}\boldsymbol{r}^1 + v_i^2\right)\Big),
\end{aligned}
$$

and

$$
\mathcal{B} \triangleq \left\{\mathcal{I}_1,\mathcal{I}_2,\mathcal{I}_3 \subseteq [I] \,\middle|\, \begin{array}{l}\mathcal{I}_1\cup\mathcal{I}_2\cup\mathcal{I}_3 = [I]\\ \mathcal{I}_i\cap\mathcal{I}_j = \emptyset \ \forall (i,j)\in\{1,2,3\},\, i\neq j\end{array}\right\}.
$$

Proof. See Online Appendix A.2. ■

In Theorem 1, the revenue π is linearized and is denoted by ϕ with respect to different partitions of set $[I]$. The first term of ϕ is the total revenue of those products whose realized demand is less than the order quantity. Similar explanations apply to the second and third terms. It is easy to see that there is an exponential number of elements in the partition of set $[I]$. Hence, given the number of clusters S and T, the size of Problem (5) grows exponentially with the number of items, as a result, this is not a computationally scalable approach.

Recall that Problem (3) can be expressed as

$$
\max_{\boldsymbol{x}\in\mathcal{X},\,\hat{\boldsymbol{r}}^2\in\bar{\mathcal{M}}(\boldsymbol{r}^1),\,\boldsymbol{r}^1\in\mathcal{R}} -\boldsymbol{c}'\boldsymbol{x} + \inf_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\Big[\pi(\boldsymbol{x},\boldsymbol{r}^1,\hat{\boldsymbol{r}}^2(\tilde{\boldsymbol{w}}^1),\,\tilde{\boldsymbol{v}}^1,\,\tilde{\boldsymbol{v}}^2)\Big], \tag{6}
$$

where

$$
\bar{\mathcal{M}}(\boldsymbol{r}^1) \triangleq \left\{\hat{\boldsymbol{r}}^2 : \mathcal{W}^1 \mapsto \mathcal{R}(\boldsymbol{r}^1)\right\}
$$

is the set of all possible markdown policies adapting to the information $\tilde{\boldsymbol{w}}^1$. Observe that the stochastic dynamic optimization model (1) is a special case when $\mathcal{F} = \{\mathbb{P}^\dagger\}$ and classical robust optimization is also a special case when the ambiguity set $\mathcal{F}$ only characterizes the support of the random variables $\tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2, \tilde{\boldsymbol{w}}^1, \tilde{\boldsymbol{w}}^2$. A key challenge of addressing the joint pricing and production problem is how we could approximate the markdown policies that would enable us to evaluate the worst-case expected objective, or provide a tractable approximation. The popular affine recourse adaptation does not apply in this setting because of the discrete nature of the recourse action. Even if we were to relax the recourse to a continuous function and apply affine recourse adaptation, it would lead to a complex nonconvex optimization problem that is difficult to linearize and solve. The simplest approach to ensure tractability is to assume a static policy for discrete recourse, which has been proposed in Bertsimas and Thiele (2006), Adida and Perakis (2010), and conceivably, such non-adaptive policy may not perform as well as adaptive ones.

To obtain a computationally tractable model, we adopt the lifted affine recourse adaptation to obtain a scalable model that serves as an approximation of Model (5).

Affine recourse adaptation

Using the lifted affine recourse adaptation (see Bertsimas et al. 2019, Chen et al. 2020), we consider the following family of functions:

$$\mathcal{A} \triangleq \left\{ f\colon \mathbb{R}^I \times \mathbb{R}^I \times \mathbb{R}^I \times \mathbb{R}^I \mapsto \mathbb{R} \;\middle|\; \begin{array}{l} \exists\, f^0 \in \mathbb{R},\, \boldsymbol{f}^1, \boldsymbol{f}^2, \boldsymbol{f}^3, \boldsymbol{f}^4 \in \mathbb{R}^I : \\ f(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) = f^0 + (\boldsymbol{f}^1)'\boldsymbol{v}^1 + (\boldsymbol{f}^2)'\boldsymbol{v}^2 + (\boldsymbol{f}^3)'\boldsymbol{u}^1 + (\boldsymbol{f}^4)'\boldsymbol{u}^2 \end{array} \right\}.$$

We can then obtain a lower bound for Problem (3), which we refer to as the *affine recourse adaptation model* (ARA), as shown in our next theorem. The key idea is to approximate the revenue of each product at each cluster by an affine function.

THEOREM 2. *Consider the following robust optimization problem under an affine recourse adaptation, where the optimal objective value is denoted Z_R:*

$$\begin{array}{lll}
\max & -\boldsymbol{c}'\boldsymbol{x} + \displaystyle\sum_{(s,t)\in[S]\times[T]} \Big(\alpha_{st} + q_{st}(\boldsymbol{\beta}^1_s)'\boldsymbol{\mu}^1_s + (\boldsymbol{\beta}^2_{st})'\boldsymbol{\mu}^2_{st} + q_{st}(\boldsymbol{\gamma}^1_s)'\boldsymbol{\sigma}^1_s + (\boldsymbol{\gamma}^2_{st})'\boldsymbol{\sigma}^2_{st} \Big) & \\
\text{s.t.} & \left.\begin{aligned}
&\alpha_{st} + q_{st}(\boldsymbol{\beta}^1_s)'\boldsymbol{v}^1 + (\boldsymbol{\beta}^2_{st})'\boldsymbol{v}^2 + q_{st}(\boldsymbol{\gamma}^1_s)'\boldsymbol{u}^1 + (\boldsymbol{\gamma}^2_{st})'\boldsymbol{u}^2 \\
&\qquad \le q_{st} \sum_{i\in[I]} f_{sti}(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2), \\
&f_{sti}(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \le \omega^1_i, \\
&f_{sti}(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \le \omega^2_{si} + (\boldsymbol{p}_i'\boldsymbol{y}^1_i - \boldsymbol{p}_i'\boldsymbol{y}^2_{si})(a^1_i + v^1_i) + \rho^{11}_i + \rho^{12}_{si}, \\
&f_{sti}(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \\
&\qquad \le \boldsymbol{p}_i'\boldsymbol{y}^1_i(a^1_i + v^1_i) + \rho^{11}_{si} + \boldsymbol{p}_i'\boldsymbol{y}^2_{si}(a^2_i + v^2_i) + \rho^{22}_{si} + \rho^{21}_{si}
\end{aligned}\right\} & \begin{array}{l} \forall (\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \in \mathcal{L}_{st}, \\ \quad i \in [I],\, (s,t) \in [S]\times[T]; \end{array} \\
& \left.\begin{aligned}
&\omega^1_i \le p_{in}x_i + p_{iN}\bar{x}_i(1 - y^1_{in}), \\
&\rho^{11}_i \le p_{in}\boldsymbol{e}_i'\boldsymbol{B}^1(\boldsymbol{p}_1'\boldsymbol{y}^1_1, \ldots, \boldsymbol{p}_I'\boldsymbol{y}^1_I)' + p_{iN}|\boldsymbol{e}_i'\boldsymbol{B}^1(\boldsymbol{p}_1'\boldsymbol{y}^1_1, \ldots, \boldsymbol{p}_I'\boldsymbol{y}^1_I)'|(1 - y^1_{in})
\end{aligned}\right\} & \forall i \in [I],\, s \in [S],\, n \in [N]; \\
& \left.\begin{aligned}
&\omega^2_{si} \le p_{in}x_i + p_{iN}\bar{x}_i(1 - y^2_{sin}), \\
&\rho^{12}_{si} \le -p_{in}\boldsymbol{e}_i'\boldsymbol{B}^1(\boldsymbol{p}_1'\boldsymbol{y}^1_1, \ldots, \boldsymbol{p}_I'\boldsymbol{y}^1_I)' + p_{iN}|\boldsymbol{e}_i'\boldsymbol{B}^1(\boldsymbol{p}_1'\boldsymbol{y}^1_1, \ldots, \boldsymbol{p}_I'\boldsymbol{y}^1_I)'|(1 - y^2_{sin}), \\
&\rho^{21}_{si} \le p_{in}\boldsymbol{e}_i'\boldsymbol{B}^{21}(\boldsymbol{p}_1'\boldsymbol{y}^1_1, \ldots, \boldsymbol{p}_I'\boldsymbol{y}^1_I)' + p_{iN}|\boldsymbol{e}_i'\boldsymbol{B}^{21}(\boldsymbol{p}_1'\boldsymbol{y}^1_1, \ldots, \boldsymbol{p}_I'\boldsymbol{y}^1_I)'|(1 - y^2_{sin}), \\
&\rho^{22}_{si} \le p_{in}\boldsymbol{e}_i'\boldsymbol{B}^2(\boldsymbol{p}_1'\boldsymbol{y}^2_{s1}, \ldots, \boldsymbol{p}_I'\boldsymbol{y}^2_{sI})' + p_{iN}|\boldsymbol{e}_i'\boldsymbol{B}^2(\boldsymbol{p}_1'\boldsymbol{y}^1_1, \ldots, \boldsymbol{p}_I'\boldsymbol{y}^1_I)'|(1 - y^2_{sin})
\end{aligned}\right\} & \forall i \in [I],\, s \in [S],\, n \in [N]; \\
& \boldsymbol{p}_i'\boldsymbol{y}^1_i \ge \boldsymbol{p}_i'\boldsymbol{y}^2_{si}, \mathbf{1}'\boldsymbol{y}^1_i = 1,\, \mathbf{1}'\boldsymbol{y}^2_{si} = 1, \boldsymbol{y}^1_i, \boldsymbol{y}^2_{si} \in \{0,1\}^N & \forall i \in [I],\, s \in [S]; \\
& f_{sti} \in \mathcal{A}, \boldsymbol{\gamma}^1_s, \boldsymbol{\gamma}^2_{st} \le 0 & \forall i \in [I],\, (s,t) \in [S]\times[T]; \\
& \alpha_{st} \in \mathbb{R}, \boldsymbol{\beta}^1_s, \boldsymbol{\beta}^2_{st}, \boldsymbol{\gamma}^1_s, \boldsymbol{\gamma}^2_{st}, \boldsymbol{\omega}^1, \boldsymbol{\omega}^2_s, \boldsymbol{\rho}^{11}, \boldsymbol{\rho}^{12}_s, \boldsymbol{\rho}^{21}_s, \boldsymbol{\rho}^{22}_s \in \mathbb{R}^I & \forall (s,t) \in [S]\times[T]; \\
& \boldsymbol{x} \in \mathcal{X}. &
\end{array} \tag{7}$$

We then have $Z_R \le Z^*$. Equality holds when $I = 1$.

Proof. See Online Appendix A.3. ■

Recall that this approach differs from extant research in that the affine recourse adaptation adapts itself to each cluster within the partitioned-moment-based ambiguity set; otherwise, we would be unable to obtain Theorem 2's "tightness" result for $I = 1$. Although that result does not hold for $I > 1$, our computation study (see Section 5) reveals that the approximation yields solutions that are nearly identical to those under the exact model.

Solving Model (7) requires that we reformulate its robust counterpart, as described next.

PROPOSITION 4. *Assuming that $\mathcal{V}_s^1$ and $\mathcal{V}_{st}^2$ are polyhedra respectively defined as $\mathcal{V}_s^1 = \{\boldsymbol{v}^1 : \boldsymbol{A}_s^1\boldsymbol{v}^1 \le \boldsymbol{b}_s^1\}$ and $\mathcal{V}_{st}^2 = \{\boldsymbol{v}^2 : \boldsymbol{A}_{st}^2\boldsymbol{v}^2 \le \boldsymbol{b}_{st}^2\}$, then the feasibility of $(\boldsymbol{p}^1, \boldsymbol{p}^2, \boldsymbol{q}^1, \boldsymbol{q}^2, g)$ in*

$$(\boldsymbol{p}^1)'\boldsymbol{v}^1 + (\boldsymbol{q}^1)'\boldsymbol{u}^1 + (\boldsymbol{p}^2)'\boldsymbol{v}^2 + (\boldsymbol{q}^2)'\boldsymbol{u}^2 \le g \quad \forall(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \in \mathcal{L}_{st}$$

is equivalent to its feasibility under the linear constraints

$$\begin{aligned}
&(\boldsymbol{\mu}_s^1)'(-\boldsymbol{\tau}_s^{11} + \boldsymbol{\tau}_s^{12}) - (\boldsymbol{\mu}_{st}^2)'(\boldsymbol{\tau}_{st}^{21} - \boldsymbol{\tau}_{st}^{22}) + (\boldsymbol{\kappa}_s^1)'\boldsymbol{b}_s^1 + (\boldsymbol{\kappa}_{st}^2)'\boldsymbol{b}_{st}^2 \le g\\
&-\boldsymbol{\tau}_s^{11} + \boldsymbol{\tau}_s^{12} + \boldsymbol{A}_s^1\kappa_s^1 = \boldsymbol{p}^1;\\
&-\boldsymbol{\tau}_{st}^{21} + \boldsymbol{\tau}_{st}^{22} + \boldsymbol{A}_{st}^2\kappa_{st}^2 = \boldsymbol{p}^2;\\
&\boldsymbol{\tau}_s^{11} + \boldsymbol{\tau}_s^{12} = \boldsymbol{q}^1;\\
&\boldsymbol{\tau}_{st}^{21} + \boldsymbol{\tau}_{st}^{22} = \boldsymbol{q}^2;\\
&\boldsymbol{\kappa}_s^1, \boldsymbol{\kappa}_{st}^2 \ge 0;\\
&\boldsymbol{\tau}_s^{11}, \boldsymbol{\tau}_s^{12}, \boldsymbol{\tau}_{st}^{21}, \boldsymbol{\tau}_{st}^{22} \le 0,
\end{aligned} \tag{8}$$

for some $\boldsymbol{\kappa}_s^1, \boldsymbol{\kappa}_{st}^2, \boldsymbol{\tau}_s^{11}, \boldsymbol{\tau}_s^{12}, \boldsymbol{\tau}_{st}^{21}, \boldsymbol{\tau}_{st}^{22} \in \mathbb{R}^I$.

Proof. See Online Appendix A.4. ■

So now, since the support sets $\mathcal{V}_s^1$, and $\mathcal{V}_{st}^2$ ($s \in [S]$, $t \in [T]$) are polyhedra, the robust optimization problem can be recast as a mixed-integer linear optimization problem. For this problem, there are in total $O(STI^2)$ of variables, and $O(STI^2 + SIN)$ constraints. We can see that, given the cluster numbers, both the number of variables and the number of constraints of the resulted mixed-integer linear optimization problem increases quadratically with the number of items. In contrast, the mixed-integer linear optimization problem associated with the exact model would lead to exponential number of constraints and decision variables.

The information partitioning approach allows us to take the adaption a step further to consider price markdown in the second period, which, under the cluster-adapted markdown policy, enables us to formulate the two-period pricing and production problem as a practically solvable mixed-integer optimization problem. This would be a significant challenge and has not been attempted in the literature. While there are proposals of K-adaptability to address some classes of adaptive robust optimization problems (see, for example, Bertsimas and Caramanis 2010, Hanasusanto et al.

2015b, Subramanyam et al. 2019), which optimally partition an uncertainty set to K regions, and for each region, an optimal recourse decision is applied, we do not know how to extend this framework to more complex problems such as the two-period distributionally robust joint pricing and production problem.

An ensemble method

It has well been known that an ensemble of clustering methods, *i.e.*, reconciling clustering information from multiple clustering algorithms, can result in robust clusters (see for example Strehl and Ghosh 2002, Gionis et al. 2007). In the same vein, we can also consider an ensemble method comprising L partitioned-moment-based ambiguity sets, where each ambiguity set, $\mathcal{F}_\ell$, $\ell \in [L]$ can be associated with a different clustering method as follows,

$$\mathcal{F}_\ell = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^I \times \mathbb{R}^I \times \mathbb{R}^J \times \mathbb{R}^J) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2, \tilde{\boldsymbol{w}}^1, \tilde{\boldsymbol{w}}^2) \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{v}}^1 \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}^1_{\ell s}] = \boldsymbol{\mu}^1_{\ell s} & \forall s \in [S_\ell] \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{v}}^2 \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}^1_{\ell s}, \tilde{\boldsymbol{w}}^2 \in \mathcal{W}^2_{\ell st}] = \boldsymbol{\mu}^2_{\ell st} & \forall (s,t) \in [S_\ell] \times [T_\ell] \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{v}}^1 - \boldsymbol{\mu}^1_{\ell s}| \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}^1_{\ell s}] \le \boldsymbol{\sigma}^1_{\ell s} & \forall s \in [S_\ell] \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{v}}^2 - \boldsymbol{\mu}^2_{\ell st}| \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}^1_{\ell s}, \tilde{\boldsymbol{w}}^2 \in \mathcal{W}^2_{\ell st}] \le \boldsymbol{\sigma}^2_{\ell st} & \forall (s,t) \in [S_\ell] \times [T_\ell] \\ \mathbb{P}[\tilde{\boldsymbol{v}}^1 \in \mathcal{V}^1_{\ell s} \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}^1_{\ell s}] = 1 & \forall s \in [S_\ell] \\ \mathbb{P}[\tilde{\boldsymbol{v}}^2 \in \mathcal{V}^2_{\ell st} \mid \tilde{\boldsymbol{w}}^2 \in \mathcal{W}^2_{\ell st}] = 1 & \forall (s,t) \in [S_\ell] \times [T_\ell] \\ \mathbb{P}[\tilde{\boldsymbol{w}}^1 \in \mathcal{W}^1_{\ell s}, \tilde{\boldsymbol{w}}^2 \in \mathcal{W}^2_{\ell st}] = q_{\ell st} & \forall (s,t) \in [S_\ell] \times [T_\ell] \end{array} \right\}.$$

Note that the markdown policy adapts to a set of S first-period clusters, associated with the information and residuals support sets, $\mathcal{W}^1_s, \mathcal{V}^1_s$, $s \in [S]$, which can be different from those obtained from the L clustering methods. Nonetheless, we are free to choose $S = S_\ell$, $\mathcal{V}^1_s = \mathcal{V}^1_{\ell s}$, $\mathcal{W}^1_s = \mathcal{W}^1_{\ell s}$, for all $s \in [S]$ so that the markdown policy would adapt to the first-period clusters associated with the ℓth clustering method.

In the ensemble method, we will solve the following problem

$$Z^* = \max_{\boldsymbol{x} \in \mathcal{X},\, \hat{\boldsymbol{r}}^2 \in \mathcal{M}(\boldsymbol{r}^1),\, \boldsymbol{r}^1 \in \mathcal{R}} -\boldsymbol{c}'\boldsymbol{x} + \sum_{\ell \in [L]} \theta_\ell \inf_{\mathbb{P} \in \mathcal{F}_\ell} \mathbb{E}_{\mathbb{P}}\Big[\pi(\boldsymbol{x}, \boldsymbol{r}^1, \hat{\boldsymbol{r}}^2(\tilde{\boldsymbol{w}}^1), \tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2)\Big], \tag{9}$$

where $\theta_\ell > 0$ are the normalized weights (*i.e.* $\sum_{\ell\in[L]} \theta_\ell = 1$). We can extend Proposition 3 and formulate Problem (9) as the following nonlinear robust optimization model:

$$\begin{array}{lll}
\max & -\boldsymbol{c}'\boldsymbol{x} + \displaystyle\sum_{\ell\in[L]} \sum_{(s,t)\in[S_\ell]\times[T_\ell]} \theta_\ell \Big(\alpha_{\ell st} + q_{\ell st} (\boldsymbol{\beta}^1_{\ell s})'\boldsymbol{\mu}^1_{\ell s} + (\boldsymbol{\beta}^2_{\ell st})'\boldsymbol{\mu}^2_{\ell st} + q_{\ell st}\ (\boldsymbol{\gamma}^1_{\ell s})'\boldsymbol{\sigma}^1_{\ell s} + (\boldsymbol{\gamma}^2_{\ell st})'\boldsymbol{\sigma}^2_{\ell st} \Big) & \\
\text{s.t.} & \alpha_{\ell st} + q_{\ell st}(\boldsymbol{\beta}^1_{\ell s})'\boldsymbol{v}^1 + (\boldsymbol{\beta}^2_{\ell st})'\boldsymbol{v}^2 + q_{\ell st}(\boldsymbol{\gamma}^1_{\ell s})'\boldsymbol{u}^1 + (\boldsymbol{\gamma}^2_{\ell st})'\boldsymbol{u}^2 & \\
& \qquad \leq q_{\ell st}\pi(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}^2_{\bar{s}}, \boldsymbol{v}^1, \boldsymbol{v}^2) & \forall (\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \in \mathcal{L}_{\ell\bar{s}st}, \\
& & \quad (\bar{s}, s, t) \in \mathcal{S}_\ell, \ell \in [L]; \\
& \boldsymbol{r}^1 \geq \boldsymbol{r}^2_{\bar{s}} & \forall \bar{s} \in [S]; \\
& \boldsymbol{\gamma}^1_{\ell s}, \boldsymbol{\gamma}^2_{\ell st} \leq 0 & \forall (s,t) \in [S_\ell]\times[T_\ell], \ell \in [L]; \\
& \boldsymbol{r}^1, \boldsymbol{r}^2_{\bar{s}} \in \mathcal{R} & \forall \bar{s} \in [S]; \\
& \alpha_{\ell st} \in \mathbb{R}, \boldsymbol{\beta}^1_{\ell s}, \boldsymbol{\beta}^2_{\ell st}, \boldsymbol{\gamma}^1_{\ell s}, \boldsymbol{\gamma}^2_{st} \in \mathbb{R}^I & \forall (s,t) \in [S_\ell]\times[T_\ell], \ell \in [L]; \\
& \boldsymbol{x} \in \mathcal{X},
\end{array} \tag{10}$$

where the lifted support set is

$$\mathcal{L}_{\ell\bar{s}st} = \left\{ (\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \in \mathbb{R}^I \times \mathbb{R}^I \times \mathbb{R}^I \times \mathbb{R}^I \;\middle|\; \begin{array}{l} \boldsymbol{v}^1 \in \mathcal{V}^1_{\bar{s}} \cap \mathcal{V}^1_{\ell s}, \, \boldsymbol{v}^2 \in \mathcal{V}^2_{\ell st} \\ \boldsymbol{u}^1 \geq |\boldsymbol{v}^1 - \boldsymbol{\mu}^1_{\ell s}|, \, \boldsymbol{u}^2 \geq |\boldsymbol{v}^2 - \boldsymbol{\mu}^2_{\ell st}| \end{array} \right\},$$

and

$$\mathcal{S}_\ell = \left\{ (\bar{s}, s, t) \in [S]\times[S_\ell]\times[T_\ell] \;\mid\; \mathcal{V}^1_{\bar{s}} \cap \mathcal{V}^1_{\ell s} \neq \emptyset \right\}.$$

Likewise, we can also apply suitable affine recourse adaptation and extend Theorem 2 for the ensemble method. We note that, although Problem (9) has the same computational structure as Problem (3), it would be more tedious to formulate and solve as L increases. Hence, part of the design challenge in the ensemble method would be how to suitably determine a computational attractive ensemble-wise markdown policy that would also perform well in practice. In the spirit of machine learning, this would require extensive experimentation using a large collection of real data sets for testing and evaluation. Our aim is to introduce a suitable framework for this purpose.

5. Numerical experiments

We conduct three different numerical experiments. In the first experiment, we explore the effect of information partitioning by considering the multi-item newsvendor problem with ambiguous demand distributions characterized using the partitioned-moment-based ambiguity set. The second experiment is a two-period production and markdown price problem, which serves to demonstrate (a) the effect of a cluster-wise markdown policy and (b) the computational efficiency of our affine recourse adaptation (ARA) model. Finally, to complement this simulation study, we apply our approach using data from an international cosmetics company to illustrate the benefits of information partitioning compared to an empirical model (EM) without price adaptation.

All of our numerical experiments were conducted in RSOME (a MATLAB-based algebraic modeling package, see in https://www.rsomerso.com/), while using the GUROBI solver on a Windows-OS computer with 64 GB of RAM and a 3.9-GHz CPU. We attach the code of Model (7) in Online Appendix F.

Effects of information partitioning

We test the effects of information partitioning on the multi-item newsvendor problem with real data. This problem is a special case of Model (3), in which $\boldsymbol{r}^1$ is a given parameter, assuming that all the leftovers at the end of the first period will be disposed. The manager solves the following problem

$$\max_{\boldsymbol{x}\in\mathcal{X}} -\boldsymbol{c}'\boldsymbol{x} + \inf_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\Big[\sum_{i\in[I]} \big(r_i^1\left[\min\{x_i,\tilde{z}_i^1)\}\right]\big)\Big], \tag{11}$$

for the following ambiguity set,

$$\mathcal{F} = \left\{ \mathbb{P}\in\mathcal{P}_0(\mathbb{R}^I\times\mathbb{R}^J) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{z}},\tilde{\boldsymbol{w}})\sim\mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}\mid\tilde{\boldsymbol{w}}\in\mathcal{W}_s] = \boldsymbol{\mu}_s & \forall s\in[S] \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{z}}-\boldsymbol{\mu}_s|\mid\tilde{\boldsymbol{w}}\in\mathcal{W}_s] \le \boldsymbol{\sigma}_s & \forall s\in[S] \\ \mathbb{P}[\tilde{\boldsymbol{z}}\in\mathcal{V}_s|\tilde{\boldsymbol{w}}\in\mathcal{W}_s] = 1 & \forall s\in[S] \\ \mathbb{P}[\tilde{\boldsymbol{w}}\in\mathcal{W}_s] = q_s & \forall s\in[S] \end{array} \right\}.$$

To test this model, we use a data set from Kaggle (https://www.kaggle.com/c/favorita-grocery-sales-forecasting/overview) of Corporación Favorita, a large Ecuadorian-based grocery retailer, operating hundreds of supermarkets. The data contains sales of grocery or perishable items across 54 stores over a period of 4.5 years (January 2013 to July 2017). It also includes the oil price during that period. We aggregate the sales over all stores and select 11 items according to the sales quantity and their degree of correlation with the oil price. Hence, each item we selected is correlated with oil price. For each item, there are 1679 demand observations. We divide the 11 items into 5 groups; each group has five products. Specifically, the 1st group comprises item 1 through item 5, the 2nd group comprises item 3 through item 7, and so forth, until the last group with items $9, 10, 11$ and $1, 2$. We further partition the data for each group into training data (70%) and testing data (30%). We use the oil price as information because Ecuador is an oil-dependent country and it is economical health is highly vulnerable to shocks in oil prices. As shown in Online Appendix D, demands for most of the products are highly correlated with oil price, with some positively correlated and others negatively correlated.

In our experimental study, we use the empirical model as a benchmark as follows,

$$\max_{\boldsymbol{x}\in\mathcal{X}} -\boldsymbol{c}'\boldsymbol{x} + \frac{1}{H}\sum_{h\in[H]}\sum_{i\in[I]} r_i \min\{x_i, z_{ih}\}.$$

We refer to Model (11) as the information model (or "w" for simplicity), and we refer to its counterpart without using the oil price as the without information model (or "w/o" for brevity). We implement the information partitioning by using K-means[1]. Prior to the K-means partitioning, we translate and normalize the information so that the empirical means and variances are zeros and ones respectively. Details of the experimental study are relegated to Online Appendix D.

Table 1 displays the mean profit of the three models, where the rows correspond to the groups and columns correspond to the number of clusters used. We report only the corresponding percentage improvement of the model with information model and the model without information over the EM in the out-of-sample test; that percentages are calculated as (w - EM)/EM × 100% and (w/o - EM)/EM × 100%, respectively. From Table 1, we can see that 16 out of 20 cases, the model with information outperforms the empirical model, while there are only 7 out of 20 cases, the model without information outputs a higher mean profit over the empirical model. Hence, side information can significantly strengthen the model; the improvement can be as much as 7.3% over the empirical model. Another observation is that with only a small number of clusters, Model (11) is better off than its empirical counterpart. Yet the out-of-sample performance does not necessarily improve with larger number of clusters. As we increase the number of clusters, the ambiguity set converges to the empirical distributions, which could result in over-fitting. In practice, we can adjust the level of robustness by varying the number of clusters, and we can determine the appropriate level via cross validation techniques.

Table 1 **Experimental study of multi-item newsvendor with information: K-means method**

Group	EM	2		3		4		5	
		w/o	w	w/o	w	w/o	w	w/o	w
1	27423	-2.95%	-1.35%	-0.40%	1.53%	-0.44%	0.22%	-2.12%	1.21%
2	53784	2.86%	2.05%	-2.71%	3.15%	-2.81%	2.62%	-4.18%	-1.40%
3	54585	-5.36%	0.12%	-0.94%	5.18%	2.12%	-0.40%	1.88%	2.97%
4	30469	2.96%	2.54%	0.83%	7.30%	-0.22%	4.93%	-6.16%	3.26%
5	31619	0.47%	2.24%	-1.89%	4.84%	1.17%	3.10%	-0.86%	-3.56%

In subsequent numerical studies, we would not consider side information and the information is solely based on the demand residual, that is, $\tilde{\boldsymbol{w}}^\tau = \tilde{\boldsymbol{v}}^\tau, \tau \in \{1,2\}$.

---

[1] We also implement the information partitioning by using regression tree, but the improvement on mean profit is not as good as the K-means method, as shown in Table D.2 of Online Appendix D.

Simulated data

Before embarking on the full application of our model to the case study, we use simulated data and investigate the EXACT, ARA, and the EM without a first-period pricing decision. So, for both EXACT and ARA, a manager solves the following distributionally robust optimization model:

$$\max_{\boldsymbol{x}\in\mathcal{X},\, \hat{\boldsymbol{r}}^2\in\mathcal{M}(\boldsymbol{r}^1)} -\boldsymbol{c}'\boldsymbol{x} + \inf_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[\pi(\boldsymbol{x}, \boldsymbol{r}^1, \hat{\boldsymbol{r}}^2(\tilde{\boldsymbol{v}}^1), \tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2)], \tag{12}$$

where $\boldsymbol{r}^1$ is the given retail price. Its corresponding empirical model (EM) is

$$\max_{\boldsymbol{x}\in\mathcal{X}, \boldsymbol{r}^2\in\mathcal{R}} -\boldsymbol{c}'\boldsymbol{x} + \frac{1}{H}\sum_{h\in[H]}\sum_{i\in[I]} \pi(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}^2, v^1_{ih}, v^2_{ih}),$$

which can be easily linearized as a mixed-integer linear optimization problem as shown in Online Appendix C.

For a fair comparison with the EM model, we also solve the ARA in the absence of a cluster-adapted markdown policy; thus, we solve Model (12) with $\hat{\boldsymbol{r}}^2(\tilde{\boldsymbol{v}}^1) = \boldsymbol{r}^2 \in \mathcal{R}$. We refer to this model as the non-adaptive affine recourse approximation model (NA-ARA).

We consider instances with $I \in \{1, \ldots, 5\}$, $S \in \{1, 2, 3, 4, 5\}$, and $T \in \{1, 2, 3, 4\}$. For each instance, we generate 2,000 sample paths in which (i) the underlying probability distribution of the first-period residual is a three-peak mixture distribution and (ii) the second-period residual is generated so as to be dependent on those first-period peaks. We also generate 100,000 sample paths—using the same distribution—to obtain the mean profit of the corresponding solutions. To streamline the presentation, in what follows, we report results only for the case $I = 5$; results for other instances of I are similar and thus omitted for the sake of brevity. The parameters for all demand models are set by following (in part) the numerical study of den Boer (2014). Specifically, we set $\boldsymbol{a}^1 = (56.32, 59.57, 57.10, 57.70, 58.04)'$, $\boldsymbol{a}^2 = (51.21, 54.32, 52.25, 52.63, 53.68)'$, and $\boldsymbol{c} = (2.53, 1.87, 2.62, 1.84, 1.82)'$ which is half of the optimal price in den Boer (2014). The price elasticity matrix $\boldsymbol{B}^1$ is quoted as follows:

$$\boldsymbol{B}^1 = \begin{pmatrix} -3.10 & 0.10 & 0.09 & 0.19 & 0.11 \\ 0.11 & -3.40 & 0.04 & 0.10 & 0.02 \\ 0.03 & 0.09 & -2.49 & 0.18 & 0.07 \\ 0.10 & 0.02 & 0.10 & -2.37 & 0.17 \\ 0.04 & 0.03 & 0.10 & 0.11 & -2.22 \end{pmatrix}$$

and $\boldsymbol{B}^2 = \mathrm{diag}(-2.50, -2.30, -3.50, -2.60, -4.10)$. For each product, the admissible set with cardinality of 6 is generated by using the cost and maximum admissible price 15.00. For instance, the admissible price set for product 1 is $\boldsymbol{p}_1 = \{2.53, 5.02, 7.52, 10.01, 12.51, 15.00\}$. We also set a different admissible price set for each product of the second period with $\boldsymbol{p}^2 = 0.9\boldsymbol{p}^1$. The first-period price is given as $\boldsymbol{r}^1 = (12.51, 12.37, 10.05, 12.37, 9.73)'$.

Table 2 Simulation study: Objective value

	1		2		3		4	
	Z^*	Z_R	Z^*	Z_R	Z^*	Z_R	Z^*	Z_R
1	2056	<0.001%	2087	<0.001%	2110	<0.001%	2111	<0.001%
2	2097	<0.001%	2120	<0.001%	2122	<0.001%	2123	<0.001%
3	2136	<0.001%	2137	<0.001%	2137	0.003%	2137	0.008%
4	2135	<0.001%	2138	0.005%	2138	0.004%	2138	0.010%
5	2137	<0.001%	2139	0.005%	2139	0.002%	2138	0.006%

Table 3 Simulation study: Computation time (s)

	3			4			5		
	ARA	NA-ARA	EXACT	ARA	NA-ARA	EXACT	ARA	NA-ARA	EXACT
1	0.038	0.035	0.077	0.043	0.040	0.075	0.111	0.048	0.107
2	0.081	0.106	0.289	0.106	0.104	0.411	0.210	0.134	0.547
3	0.203	0.206	2.170	0.274	0.207	6.238	0.412	0.237	32.06
4	0.515	0.333	57.84	0.844	0.410	395.9	1.573	0.495	277.5
5	0.635	0.512	1237	1.407	0.629	3013	1.692	0.696	4728

Table 2 displays the objective value of EXACT and ARA. In this table, the rows correspond to the number of first-period clusters $(1,\ldots,5)$ while the columns correspond to the number of second-period clusters $(1,\ldots,4)$. For the ARA's objective value, we report only the gap between the two models, which is calculated as $(\text{ARA} - \text{EXACT})/\text{EXACT} \times 100$. Table 3 shows the computation time of the ARA, NA-ARA, and EXACT. In this table, the rows correspond to the number of products $(1,\ldots,5)$ while the columns correspond to the number of first-period clusters $(3,\ldots,5)$. We fix the number of second-period clusters to four.

It is clear from Table 3 that, as the number of products increases, the EXACT computational time increases exponentially, whereas the time for ARA and NA-ARA usually increases by only a few seconds. Nonetheless, Table 2 shows that the EXACT and ARA objective values coincide; that is, the affine recourse adaptation usually yields the same solution as the exact model. So even though the later is not computationally scalable, ARA does an excellent job of approximating the exact solution; we therefore, exclude the exact model from our case study.

We plot the mean profit in Figure 2, where the horizontal axis corresponds to the number of first-period clusters and where markers of different colors and shapes (see the figure's Key) represent

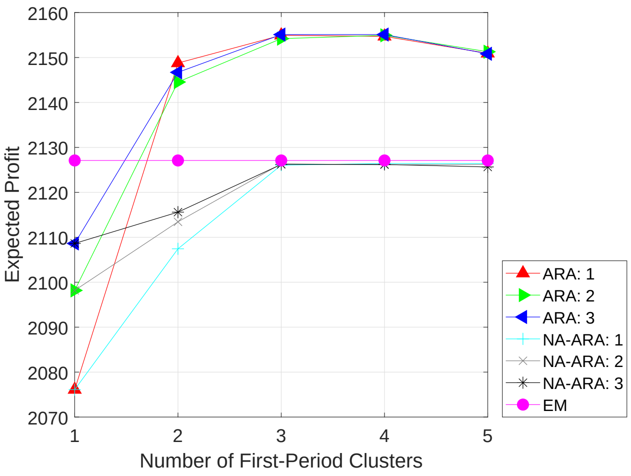


Figure 2 **Simulation study: Mean profit**

the mean profit under different models with different numbers of second-period clusters. Visual inspection of this figure allows us to have the following observations.

1. The EM always performs better than the NA-ARA. Yet when there are only a few clusters, the mean profit of the NA-ARA is remarkably close to that of the EM.
2. Models incorporating the cluster-adapted markdown policy yield higher mean profit than their non-adaptive counterparts.
3. The ARA outperforms the EM when there are 2–5 first-period clusters.

On cross-product effect. We now investigate the impact of the cross-product effect. Specifically, we first examine how the substitution and complementary affects the expected profit. This is achieved by letting $\boldsymbol{B}^1 = \text{diag}(-3.10, -3.40, -2.49, -2.37, -2.22) + \vartheta\big(\mathbf{1} - \boldsymbol{I}\big)$, where $\vartheta \in \{-0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3\}$, $\boldsymbol{I}$ being the identity matrix. When ϑ is negative (positive), there only exists complementary (substitution) effects. We then study the impact of item combination, *i.e.*, how the proportion of complementary products, denoted as ϖ, will affect the expected profit. We conduct this by setting $\vartheta = 0.1$ and changing $J \in \{0, 4, 8, 12, 16, 20\}$ of the non-diagonal elements to be -0.1 row by row so that $\varpi \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$; for example, when $J = 8(\varpi = 0.4)$, we should have

$$\boldsymbol{B}^1 = \begin{pmatrix} -3.10 & -0.10 & -0.10 & -0.10 & -0.10 \\ -0.10 & -3.40 & -0.10 & -0.10 & -0.10 \\ 0.10 & 0.10 & -2.49 & 0.20 & 0.10 \\ 0.10 & 0.10 & 0.10 & -2.37 & 0.10 \\ 0.10 & 0.10 & 0.10 & 0.20 & -2.22 \end{pmatrix}.$$

We plot the expected profit of ARA averaging over all clusters with respect to ϑ and ϖ in left and middle panel of Figure 3, from which we see that strong substitution effects benefit the firm but strong complementary effects are detrimental to the firm. As such, it is better to sell products that compose more substitutes.

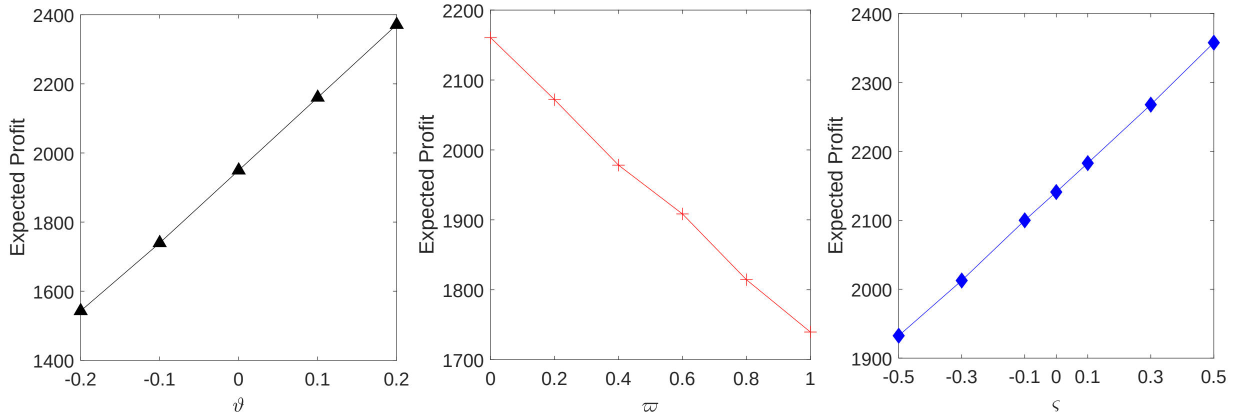


Figure 3 **Simulation study: Impact of cross-product effect (left and middle) and cross-period effect (right)**

On cross-period effect. We also explore the impact of cross-period effects. For simplicity, we let $\boldsymbol{B}^{21} = \varsigma \boldsymbol{I}$, where $\varsigma \in \{-0.5, -0.3, -0.1, 0, 0.1, 0.3, 0.5\}$, so that only the price of product i in the first period will affect its demand in the second period, and the higher the first-period price, the more likely it will have a higher demand in the second period. The manager would expect higher profit as ς increases, which is the case as observed in the right panel of Figure 3.

Sensitivity analysis on the production capacity. We also conduct sensitive analysis on the production capacity $\bar{\boldsymbol{x}}$. We first obtain the maximum possible production quantity $\hat{\boldsymbol{x}}$ and determine the total maximum demand in both periods. Subsequently, we let the production capacity to be $\bar{\boldsymbol{x}} = (1+\delta)\hat{\boldsymbol{x}}$, where δ is a scale parameter ranging from -0.45 to 0.1 with step of 0.05. The out-of-sample average profits are shown in Table 4, where the rows correspond to different values of δ and columns correspond to the out-of-sample mean profits of EM and ARA under different number of first-period clusters, averaging over second-period clusters (1,...,4). As before, we report only the corresponding percentage improvement of ARA over EM under different production capacities. We see that a lower production capacity results in lower mean profit. However, when the manager has limited production capacity, the ARA still outperforms the EM even with a modest number of clusters.

The simulation demonstrates that (1) the affine recourse adaptation is near optimal but far more tractable; (2) distributionally robust optimization models are not conservative; (3) the cluster-adapted markdown policy helps to improve the mean profit; (4) substitution products and cross-period effects may improve the firm's profit, while complementary products may reduce it; (5) even if the production capacity is limited, the ARA with a modest number of clusters can still outperforms the EM.

Case study

We quantify the value of our information partitioning framework by applying it to the two-period, multi-item joint pricing and production problem through a real-world data set. For both implementations (ARA and EM), we report out-of-sample results in terms of the mean of profit.

Table 4 Sensitivity analysis on production capacity

δ	EM	ARA				
		1	2	3	4	5
-0.45	2052.03	-0.31%	0.34%	0.31%	0.31%	0.31%
-0.40	2089.69	-0.26%	0.98%	1.09%	1.09%	0.96%
-0.35	2114.76	-0.50%	1.10%	1.32%	1.33%	1.16%
-0.30	2126.73	-1.01%	0.98%	1.36%	1.37%	1.18%
-0.25	2127.09	-1.23%	0.91%	1.30%	1.31%	1.13%
-0.20	2127.09	-1.37%	0.91%	1.30%	1.31%	1.13%
-0.15	2127.09	-1.37%	0.91%	1.30%	1.31%	1.13%
-0.10	2127.09	-1.37%	0.91%	1.30%	1.31%	1.13%
-0.05	2127.09	-1.37%	0.91%	1.30%	1.31%	1.13%
0.00	2127.09	-1.37%	0.91%	1.30%	1.31%	1.13%
0.05	2127.09	-1.37%	0.91%	1.30%	1.31%	1.13%
0.10	2127.09	-1.37%	0.91%	1.30%	1.31%	1.13%

Data description The dataset was acquired from an international cosmetics company. It spans nearly a decade (from August 2007 to September 2016) and includes sales and inventory records of 123 products carried by 341 stores in Indonesia. This data set has 18 variables features; examples include daily volume of products sold, regular price, discount percentage, and so forth. We focus on the skin care product category. For the purpose of demand estimation and informationing, we aggregate the demand for each SKU over all stores and then select twelve products for the joint pricing and production problem (see Online Appendix E for details of the selection procedure). Although we have sales data (*i.e.*, the minimum of on-hand inventory and realized demand), actual demand information is censored. With regard to techniques for recovering an approximation of the real demand, see, Heien and Wesseils (1990) and Huh et al. (2011). Yet because our focus is not on the prediction of demands but rather on the effectiveness of our information-partition framework, sales data are used to proxy for real demand. For ease of interpretation (and to maintain confidentiality of the prices), we normalize the regular prices—commonly adopted when analyzing real data (see, e.g., Cohen et al. 2017).

Testing for the cross-product effect. We provide statistical evidence that there does exist a cross-product effect. This demonstration is based on running a multivariate linear regressions on the demand model of the form:

$$z_{ih} = a_{ih} + B_{ii} r_{ih} + \sum_{m \neq i} B_{mi} r_{mh} + \boldsymbol{t}_i \textbf{Year} + v_{ih},$$

for $i \in [I]$ and $h \in [H]$. Here $\boldsymbol{t}_i\textbf{Year}$ is added to remove the time trend. We report the regression results for item 6 in Table 5, and we summarize the linear regression results for twelve selected items in Online Appendix E.

Table 5 **Linear regression results for product 6: Aggregate level**

Coefficient	Estimate	S.E.	t value	Pr($>\|t\|$)	Signif.
a_1	77.83404	5.47115	14.226	$<$ 2e-16	***
B_{11}	-2.33564	0.30376	-7.689	1.94e-14	***
B_{21}	0.39584	0.21878	1.809	0.07049	.
B_{31}	-3.04536	0.38548	-7.9	3.76e-15	***
B_{41}	-1.08372	0.26601	-4.074	4.73e-05	***
B_{51}	0.87208	0.20802	4.192	2.83e-05	***
B_{61}	-4.01455	0.21308	-18.84	$<$ 2e-16	***
B_{71}	0.5943	0.23673	2.51	0.01211	*
B_{81}	0.13071	0.13202	0.99	0.32223	
B_{91}	0.38688	0.19348	2	0.04563	*
$B_{10,1}$	0.03709	0.2369	0.157	0.87558	
$B_{11,1}$	0.49349	0.24065	2.051	0.04038	*
$B_{12,1}$	0.13789	0.17825	0.774	0.43926	
$t_{2008,1}$	9.18822	1.90737	4.817	1.52e-06	***
$t_{2009,1}$	4.80143	2.05358	2.338	0.01944	*
$t_{2010,1}$	5.32402	2.02286	2.632	0.00853	**
$t_{2011,1}$	19.07968	2.04695	9.321	$<$ 2e-16	***
$t_{2012,1}$	21.52176	2.04476	10.525	$<$ 2e-16	***
$t_{2013,1}$	36.41536	2.20425	16.52	$<$ 2e-16	***
$t_{2014,1}$	40.20539	3.12288	12.874	$<$ 2e-16	***
$t_{2015,1}$	40.55375	3.26036	12.438	$<$ 2e-16	***
$t_{2016,1}$	32.08677	3.58782	8.943	$<$ 2e-16	***

Signif. codes: '***', 0.001; '**', 0.01; '*', 0.05; '.', 0.1; ' ', 1
Residual standard error, 16.66 on 3,291 degrees of freedom (df)
Multiple R^2, 0.3624; adjusted R^2, 0.3584
F-statistic, 89.09 on 21 and 3,291 df; p-value, <2.2e-16

Table 5 clearly shows that the demand for product 6 declines as its price increases. With respect to this product, products 5, 7, 9, and 11 are substitutes whereas products 1, 3, and 4 are complements. This table also reveals a significant time effect.

We perform the same test at the SKU level: instead of aggregating demand over stores, we identify the store with the highest sales/demand and then rerun the linear regression model. These regression results for item 6 are reported in Table 6. At the SKU level, products 2, 5 and 12 are substitutes for product 6 while products 1, 4, 7 and 10 are complements to product 6.

Experiments on joint pricing and production problem We use sales data on twelve selected products and sequentially partition them into five groups. As we have discussed before, each group contains four products, and the subsequent group differs by two products, which emulates the random selection of four products in each group. For each group, the data are sorted into training data (60% of the data) and testing data (40% of the data), hence for each group, there will be two sub-datasets. If we denote each sub-dataset as $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_{2H}\}$, we then select $\{\boldsymbol{z}_1, \boldsymbol{z}_3, \ldots, \boldsymbol{z}_{2H-1}\}$

Table 6 Linear regression results for product 6: SKU level

Coefficient	Estimate	S.E.	t value	Pr($>\|t\|$)	Signif.
a_1	6.2547	0.65758	9.512	$<$ 2e-16	***
B_{11}	-0.13708	0.0361	-3.797	0.000149	***
B_{21}	0.15503	0.04538	3.417	0.000642	***
B_{31}	-0.03455	0.02516	-1.373	0.169865	
B_{41}	-0.07098	0.02131	-3.33	0.000877	***
B_{51}	0.0667	0.02866	2.328	0.019996	*
B_{61}	-0.22722	0.02808	-8.093	8.12e-16	***
B_{71}	-0.08116	0.02693	-3.014	0.002599	**
B_{81}	0.02207	0.03239	0.681	0.495698	
B_{91}	-0.01966	0.02483	-0.792	0.428647	
$B_{10,1}$	-0.0695	0.02822	-2.463	0.013844	*
$B_{11,1}$	-0.01473	0.01538	-0.957	0.338407	
$B_{12,1}$	0.09395	0.02315	4.058	5.07e-05	***
$t_{2008,1}$	0.33086	0.23449	1.411	0.15835	
$t_{2009,1}$	0.50595	0.25202	2.008	0.044771	*
$t_{2010,1}$	1.42326	0.24814	5.736	1.06e-08	***
$t_{2011,1}$	0.67061	0.25096	2.672	0.007573	**
$t_{2012,1}$	1.04016	0.25052	4.152	3.38e-05	***
$t_{2013,1}$	1.44387	0.26964	5.355	9.16e-08	***
$t_{2014,1}$	2.02587	0.37874	5.349	9.45e-08	***
$t_{2015,1}$	2.9283	0.39541	7.406	1.65e-13	***
$t_{2016,1}$	3.01194	0.43371	6.945	4.55e-12	***

Signif. codes: '***', 0.001; '**', 0.01; '*', 0.05; '.', 0.1; ' ', 1
Residual standard error, 2.054 on 3,291 degrees of freedom (df)
Multiple R^2, 0.1272; adjusted R^2, 0.1216
F-statistic, 23.83 on 21 and 3,291 df; p-value, <2.2e-16

to represent the first-period demand, and the rest elements— $\{\boldsymbol{z}_2, \boldsymbol{z}_4, \ldots, \boldsymbol{z}_{2H}\}$—represent second-period demand. Applying multivariate linear regression to the training data yields, as output, the demand model's coefficients. Details on deriving these coefficients (and on the other parameters) are provided in Online Appendix E. The admissible price set is defined such that the optimal price is an interior solution, although in practice it can be defined to comply with any particular business rule(s). We then use the estimated demand model to obtain perturbation samples for both the training data and the test data.

Leveraging these perturbation samples, we solve the affine recourse adaptation model (7) with a cluster-adapted markdown policy (ARA) and also without such a policy (NA-ARA). As shown in our simulation study, a small number of clusters results in good performance in terms of the mean profit. In this experiment, we set the number of clusters at 4 and 5 for the first period and 3 and 4 for the second period. We thus obtain the corresponding optimal policy, and apply it to the test data, and compare the test results with those under the EM approach.

As before, we focus on the performance metrics of mean profit. For each metric, we report only the corresponding percentage improvement of ARA (or of NA-ARA) over EM; that percentage

Table 7 Case study: profit in out-of-sample test

Group	EM	(4,3)		(4,4)		(5,3)		(5,4)	
		ARA	NA-ARA	ARA	NA-ARA	ARA	NA-ARA	ARA	NA-ARA
1	6663	0.56%	0.02%	0.50%	-0.05%	-0.06%	-0.34%	0.08%	-0.23%
2	2699	1.44%	-0.50%	1.42%	-0.46%	1.91%	-0.30%	1.56%	-0.29%
3	2973	1.18%	-0.15%	1.40%	-0.27%	1.33%	-0.04%	1.41%	-0.27%
4	3008	0.21%	-0.24%	0.33%	-0.28%	-0.13%	-0.06%	0.23%	-0.03%
5	1876	2.22%	0.09%	1.07%	0.13%	1.61%	-0.18%	0.88%	0.43%

is calculated as $(\text{ARA} - \text{EM})/\text{EM} \times 100$. For the out-of-sample tests, the yielded profit is shown in Table 7. In this table, the rows correspond to the group number $(1, \ldots, 5)$ while the columns correspond to different combinations of clusters in the two periods. For instance, $(4, 3)$ represents what the number of clusters in the first and second period are, that is, 4 and 3, respectively. For the ARA and NA-ARA's profit, we report only the corresponding improvement over EM.

We have four observations from Table 7. First, the NA-ARA is near optimal similarly to EM; the gap is 0.15% over all instances. It demonstrates that employing the partitioned-moment-based ambiguity set can reduce the conservativeness of distributionally robust models. Second, adopting the cluster-adapted markdown policy usually results in a higher profit than does the EM. For instance, in case $(4, 3)$, across all 5 groups, the profit improves by 1.12% (on average) and by as much as 2.22% as compared to the EM. Third, it is not always a recommendation to adopt the markdown policy; it depends on product combinations. It is beneficial to apply the policy on group 2, 3, and 5. Forth, a comparison between ARA and NA-ARA reveals that, in general, the former's adaptivity results in higher profits. For example, in case $(4, 3)$, it is 1.28% higher on average in terms of profit.

6. Conclusions

We introduce a new distributionally robust optimization model to address a two-period, multi-item joint pricing and production problem, which can be implemented in a data-driven setting using historical demand and side information pertinent to the prediction of demands. Our empirical study shows that we can improve the out-of-sample profit by incorporating side information and information partitioning using only a small number of clusters. Our simulation study also establishes the efficacy of the affine recourse adaptation, which typically achieves almost the same outcomes as does the exact model, and doing so with modest computational effort. Using a real-world data set, we also show favorable results in the pricing and production problem against an empirical model. While we have only explored the basic clustering method, it may well be possible to optimize over the clusters to achieve better performance, which we will pursue in future research. In practice, there could be multiple markdown periods. The associated problem can be myopically solved by implementing our two-period model in a rolling horizon manner.

Another possible direction is to consider the censored demand setting, which has never been addressed in distributionally robust optimization framework before. Censored demand is an important problem in inventory management. There are at least 20 papers in the literature in this area and most of the papers focus on a single demand problem. Shi et al. (2016) is perhaps the only work to propose a non-parametric algorithm to address the multi-item inventory system. There are also several papers proposing different heuristics to recover an approximation of the real demand distribution (see, for example, Stefanescu 2009, Mersereau 2015). It would be worthwhile to incorporate loss sales information within an ambiguity set.

Acknowledgments

The authors would like to thank the editors and the reviewers for the constructive comments, which have greatly improved the paper. In particular, the ensemble method is proposed by one of the anonymous reviewers. The authors gratefully acknowledge Mr. Herman Tan for sharing the data with us. The research was conducted while Qinshen Tang was visiting in Sloan School of Management at the Massachusetts Institute of Technology and was partly financed by NUS Business School, FY2018 Ph.D. Exchange Fellowship. This paper was supported by Nanyang Technological University [Start-Up Grant 020022-00001], and the Ministry of Education, Singapore, under its 2019 Academic Research Fund Tier 3 grant call (Grant MOE-2019-T3-1-010).

References

Abernathy, Frederick H, John T Dunlop, Janice H Hammond, David Weil. 1999. *A stitch in time: Lean retailing and the transformation of manufacturing–lessons from the apparel and textile industries*. Oxford University Press.

Adida, Elodie, Georgia Perakis. 2006. A robust optimization approach to dynamic pricing and inventory control with no backorders. *Mathematical Programming* **107**(1-2) 97–129.

Adida, Elodie, Georgia Perakis. 2010. Dynamic pricing and inventory control: Uncertainty and competition. *Operations Research* **58**(2) 289–302.

Adlarson, Carl, Marcus Holgersson. 2016. Pre-and in-season stock allocation at H&M online. Master's thesis, Chalmers University of Technology.

Agrawal, Vipul, Sridhar Seshadri. 2000. Impact of uncertainty and risk aversion on price and order quantity in the newsvendor problem. *Manufacturing & Service Operations Management* **2**(4) 410–423.

Amjad, Muhammad J, Devavrat Shah. 2017. Censored demand estimation in retail. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* **1**(2) 1–28.

Anupindi, Ravi, Maqbool Dada, Sachin Gupta. 1998. Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science* **17**(4) 406–423.

Axsäter, Sven. 2015. *Inventory control*, vol. 225. Springer.

Aydin, Goker, Evan L Porteus. 2008. Joint inventory and pricing decisions for an assortment. *Operations Research* **56**(5) 1247–1255.

Ban, Gah-Yi. 2020. Confidence intervals for data-driven inventory policies with demand censoring. *Operations Research* **68**(2) 309–326.

Ban, Gah-Yi, Jérémie Gallien, Adam J Mersereau. 2019. Dynamic procurement of new products with covariate information: The residual tree method. *Manufacturing & Service Operations Management* **21**(4) 798–815.

Ben-Tal, Aharon, Alexander Goryashko, Elana Guslitzer, Arkadi Nemirovski. 2004. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming* **99**(2) 351–376.

Bernstein, Fernando, Yang Li, Kevin Shang. 2015. A simple heuristic for joint inventory and pricing models with lead time and backorders. *Management Science* **62**(8) 2358–2373.

Bertschek, Irene, Ulrich Kaiser. 2004. Productivity effects of organizational change: Microeconometric evidence. *Management Science* **50**(3) 394–404.

Bertsimas, Dimitris, Constantine Caramanis. 2010. Finite adaptability in multistage linear optimization. *IEEE Transactions on Automatic Control* **55**(12) 2751–2766.

Bertsimas, Dimitris, Sanne De Boer. 2005. Special issue papers: dynamic pricing and inventory control for multiple products. *Journal of Revenue and Pricing Management* **3**(4) 303–319.

Bertsimas, Dimitris, Nathan Kallus. 2020. From predictive to prescriptive analytics. *Management Science* **66**(3) 1025–1044.

Bertsimas, Dimitris, Melvyn Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.

Bertsimas, Dimitris, Melvyn Sim, Meilin Zhang. 2019. Adaptive distributionally robust optimization. *Management Science* **65**(2) 604–618.

Bertsimas, Dimitris, Aurélie Thiele. 2006. A robust optimization approach to inventory theory. *Operations Research* **54**(1) 150–168.

Bhardwaj, Prachi. 2017. When to start looking for iphone 7 price cuts. Accessed Jan. 01, 2018. Available at: `https://www.consumerreports.org/smartphones/iphone-price-cuts/`.

Bitran, Gabriel, René Caldentey. 2003. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management* **5**(3) 203–229.

Bradlow, Eric T, Manish Gangwar, Praveen Kopalle, Sudhir Voleti. 2017. The role of big data and predictive analytics in retailing. *Journal of Retailing* **93**(1) 79–95.

Breiman, Leo, Jerome H Friedman, Richard A Olshen, Charles J Stone. 1984. *Classification and regression trees*. The Wadsworth Statistics and Probability Series, Wadsworth International Group, Belmont California.

Cachon, Gérard P, A Gürhan Kök. 2007. Implementation of the newsvendor model with clearance pricing: How to (and how not to) estimate a salvage value. *Manufacturing & Service Operations Management* **9**(3) 276–290.

Carboni Borrasé, Rodolfo. 2009. Clearance pricing optimization at Zara. Ph.D. thesis, Massachusetts Institute of Technology.

Caro, Felipe, Jérémie Gallien. 2012. Clearance pricing optimization for a fast-fashion retailer. *Operations Research* **60**(6) 1404–1422.

Chan, Lap MA, ZJ Max Shen, David Simchi-Levi, Julie L Swann. 2004. Coordination of pricing and inventory decisions: A survey and classification. *Handbook of quantitative supply chain analysis*. Springer, 335–392.

Chen, Boxiao, Xiuli Chao, Cong Shi. 2021. Nonparametric algorithms for joint pricing and inventory control with lost-sales and censored demand. *Mathematics of Operations Research.* **46**(2) 726–756.

Chen, Xin, Peng Hu, Zhenyu Hu. 2016. Efficient algorithms for the dynamic pricing problem with reference price effect. *Management Science* **63**(12) 4389–4408.

Chen, Xin, David Simchi-Levi. 2012. *Pricing and inventory management*. Citeseer.

Chen, Yiwei, Cong Shi. 2019. Joint pricing and inventory management with strategic customers. *Operations Research* **67**(6) 1610–1627.

Chen, Zhi, Melvyn Sim, Peng Xiong. 2020. Robust stochastic optimization made easy with RSOME. *Management Science* **66**(8) 3295–3798.

Chu, Leon Yang, Guang Li, Paat Rusmevichientong. 2018. Optimal pricing and inventory planning with charitable donations. *Manufacturing & Service Operations Management* **20**(4) 687–703.

Cohen, Maxime C, Ngai-Hang Zachary Leung, Kiran Panchamgam, Georgia Perakis, Anthony Smith. 2017. The impact of linear optimization on promotion planning. *Operations Research* **65**(2) 446–468.

Cohen, Maxime C, Ruben Lobel, Georgia Perakis. 2018. Dynamic pricing through data sampling. *Production and Operations Management* **27**(6) 1074–1088.

Cross, Robert G. 2011. *Revenue management: Hard-core tactics for market domination*. Crown Business.

De′Ath, Glenn. 2002. Multivariate regression trees: a new technique for modeling species–environment relationships. *Ecology* **83**(4) 1105–1117.

De′Ath, Glenn, Katharina E Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* **81**(11) 3178–3192.

Delage, Erick, Yinyu Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* **58**(3) 595–612.

den Boer, Arnoud V. 2014. Dynamic pricing with multiple products and partially specified demand distribution. *Mathematics of Operations Research* **39**(3) 863–888.

Elmaghraby, Wedad, Pınar Keskinocak. 2003. Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Science* **49**(10) 1287–1309.

Federgruen, A W I, Aliza Heching. 1999. Combined pricing and inventory control under uncertainty. *Operations Research* **47**(3) 454–475.

Ferguson, Mark E, Oded Koenigsberg. 2007. How should a firm manage deteriorating inventory? *Production and Operations Management* **16**(3) 306–321.

Ferreira, Kris Johnson, Bin Hong Alex Lee, David Simchi-Levi. 2016. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management* **18**(1) 69–88.

Fibich, Gadi, Arieh Gavious, Oded Lowengart. 2003. Explicit solutions of optimization models and differential games with nonsmooth (asymmetric) reference-price effects. *Operations Research* **51**(5) 721–734.

Gallego, Guillermo, Jennifer K Ryan, David Simchi-Levi. 2001. Minimax analysis for finite-horizon inventory models. *IIE Transactions* **33**(10) 861–874.

Gallego, Guillermo, Garrett Van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science* **40**(8) 999–1020.

Gao, Rui, Anton J Kleywegt. 2016. Distributionally robust stochastic optimization with Wasserstein distance. Available at arXiv:1604.02199.

Gionis, Aristides, Heikki Mannila, Panayiotis Tsaparas. 2007. Clustering aggregation. *ACM Transactions on Knowledge Discovery from Data* **1**(1) 1217299–1217303.

Gong, Xiting, Xiuli Chao, Shaohui Zheng. 2014. Dynamic pricing and inventory management with dual suppliers of different lead times and disruption risks. *Production and Operations Management* **23**(12) 2058–2074.

Goody Feed. 2016. 6 food that are sold out at almost half price after 9:00 p.m. in singapore. Accessed Jan. 01, 2018. Available at: `https://goodyfeed.com/6-foods-that-are-sold-at-almost-half-price-after-900-p-m-in-spore/`.

Greenleaf, Eric A. 1995. The impact of reference price effects on the profitability of price promotions. *Marketing Science* **14**(1) 82–104.

Hanasusanto, Grani A, Daniel Kuhn. 2018. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Operations Research* **66**(3) 849–869.

Hanasusanto, Grani A, Daniel Kuhn, Stein W Wallace, Steve Zymler. 2015a. Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming* **152**(1-2) 1–32.

Hanasusanto, Grani A, Daniel Kuhn, Wolfram Wiesemann. 2015b. K-adaptability in two-stage robust binary programming. *Operations Research* **63**(4) 877–891.

Hannah, Lauren, Warren Powell, David M Blei. 2010. Nonparametric density estimation for stochastic optimization with an observable state variable. *Advances in Neural Information Processing Systems*. 820–828.

Hao, Zhaowei, Long He, Zhenyu Hu, Jun Jiang. 2020. Robust vehicle pre-allocation with uncertain covariates. *Production and Operations Management* **29**(4) 955–972.

Hastie, Trevor, Robert Tibshirani, Jerome Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Heien, Dale, Cathy Roheim Wesseils. 1990. Demand systems estimation with microdata: a censored regression approach. *Journal of Business & Economic Statistics* **8**(3) 365–371.

Ho, Chin Pang, Grani A Hanasusanto. 2019. On data-driven prescriptive analytics with side information: A regularized Nadaraya-Watson approach. Tech. rep., Technical report, March.

Hobbs, Benjamin F, Jong-Shi Pang. 2007. Nash-cournot equilibria in electric power markets with piecewise linear demand functions and joint constraints. *Operations Research* **55**(1) 113–127.

Hu, Zhenyu, Xin Chen, Peng Hu. 2016. Dynamic pricing with gain-seeking reference price effects. *Operations Research* **64**(1) 150–157.

Huh, Woonghee Tim, Retsef Levi, Paat Rusmevichientong, James B Orlin. 2011. Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. *Operations Research* **59**(4) 929–941.

Iyer, Ananth V, Mark E Bergen. 1997. Quick response in manufacturer-retailer channels. *Management Science* **43**(4) 559–570.

Jain, Aditya, Nils Rudi, Tong Wang. 2015. Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need. *Operations Research* **63**(1) 134–150.

Jain, Anil K. 2010. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* **31**(8) 651–666.

Jiang, Houyuan, Serguei Netessine, Sergei Savin. 2011. Robust newsvendor competition under asymmetric information. *Operations Research* **59**(1) 254–261.

Kopalle, Praveen K, PK Kannan, Lin Bao Boldt, Neeraj Arora. 2012. The impact of household level heterogeneity in reference price effects on optimal retailer pricing policies. *Journal of Retailing* **88**(1) 102–114.

Kumar, Mahesh, Nitin R Patel, Jonathan Woo. 2002. Clustering seasonality patterns in the presence of errors. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. 557–563.

Kutner, Michael H, Christopher J Nachtsheim, John Neter, William Li, et al. 2005. *Applied linear statistical models*, vol. 5. McGraw-Hill Irwin New York.

Kuyumcu, H Ahmet, Ioana Popescu. 2006. Deterministic price–inventory management for substitutable products. *Journal of Revenue and Pricing Management* **4**(4) 354–366.

Lim, Andrew, J George Shanthikumar, Thaisiri Watewai. 2008. Robust multi-product pricing. Available at SSRN 1078012.

Lu, Ye, Youhua Chen, Miao Song, Xiaoming Yan. 2014. Optimal pricing and inventory control policy with quantity-based price differentiation. *Operations Research* **62**(3) 512–523.

Mantrala, Murali K, Surya Rao. 2001. A decision-support system that helps retailers decide order quantities and markdowns for fashion goods. *Interfaces* **31**(3_supplement) S146–S165.

Markets and Markets. 2020. Retail analytics market. Accessed Feb. 01, 2021. Available at: `https://www.marketsandmarkets.com/Market-Reports/retail-analytics-market-123460609.html`.

Martínez-de Albéniz, Victor, Abdel Belkaid. 2021. Here comes the sun: Fashion goods retailing under weather fluctuations. *European Journal of Operational Research* **294**(3) 820–830.

Mersereau, Adam J. 2015. Demand estimation from censored observations with inventory record inaccuracy. *Manufacturing & Service Operations Management* **17**(3) 335–349.

Mitra, Subrata. 2018. Newsvendor problem with clearance pricing. *European Journal of Operational Research* **268**(1) 193–202.

Mohajerin Esfahani, Peyman, Daniel Kuhn. 2017. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 1–52.

Monahan, George E, Nicholas C Petruzzi, Wen Zhao. 2004. The dynamic pricing problem from a newsvendor's perspective. *Manufacturing & Service Operations Management* **6**(1) 73–91.

Musalem, Andrés, Marcelo Olivares, Eric T Bradlow, Christian Terwiesch, Daniel Corsten. 2010. Structural estimation of the effect of out-of-stocks. *Management Science* **56**(7) 1180–1197.

Naderi, Siamak, Kemal Kilic, Abdullah Dasci. 2020. A deterministic model for the transshipment problem of a fast fashion retailer under capacity constraints. *International Journal of Production Economics* **227** 107687.

Nahmias, Steven. 1994. Demand estimation in lost sales inventory systems. *Naval Research Logistics (NRL)* **41**(6) 739–757.

Nemirovski, Arkadi, Alexander Shapiro. 2007. Convex approximations of chance constrained programs. *SIAM Journal on Optimization* **17**(4) 969–996.

Perakis, Georgia, Guillaume Roels. 2008. Regret in the newsvendor model with partial information. *Operations Research* **56**(1) 188–203.

Petruzzi, Nicholas C, Maqbool Dada. 1999. Pricing and the newsvendor problem: A review with extensions. *Operations Research* **47**(2) 183–194.

Petruzzi, Nicholas C, Maqbool Dada. 2001. Information and inventory recourse for a two-market, price-setting retailer. *Manufacturing & Service Operations Management* **3**(3) 242–263.

Popescu, Ioana, Yaozhong Wu. 2007. Dynamic pricing strategies with reference effects. *Operations Research* **55**(3) 413–429.

Qi, Meng, Ho-Yin Mak, Zuo-Jun Max Shen. 2020. Data-driven research in retail operations—a review. *Naval Research Logistics (NRL)* **67**(8) 595–616.

Ramachandran, Karthik, Necati Tereyağoğlu, Yusen Xia. 2018. Multidimensional decision making in operations: An experimental investigation of joint pricing and quantity decisions. *Management Science* **64**(12) 5544–5558.

Research and Markets. 2016. Overview & evolution of the global retail industry. Accessed Jan. 01, 2019. Available at: `https://www.businesswire.com/news/home/20160630005551/en/Global-Retail-Industry-Worth-USD-28-Trillion`.

Russell, Gary J, Ann Petersen. 2000. Analysis of cross category dependence in market basket selection. *Journal of Retailing* **76**(3) 367–392.

Scarf, Herbert. 1958. A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production* .

Schlapp, Jochen, Moritz Fleischmann. 2018. Multiproduct inventory management under customer substitution and capacity restrictions. *Operations Research* **66**(3) 740–747.

Shen, Zuo-Jun Max, Xuanming Su. 2007. Customer behavior modeling in revenue management and auctions: A review and new research opportunities. *Production and Operations Management* **16**(6) 713–728.

Shenoy, Dinesh, Roberto Rosas. 2018. *Problems & Solutions in Inventory Management*. Springer.

Shi, Cong, Weidong Chen, Izak Duenyas. 2016. Nonparametric data-driven algorithms for multiproduct inventory systems with censored demand. *Operations Research* **64**(2) 362–370.

Simchi-Levi, David, Xin Chen, Julien Bramel. 2005. *The Logic of Logistics*. Springer.

Smith, James E, Robert L Winkler. 2006. The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science* **52**(3) 311–322.

Song, Jing-Sheng, Zhengliang Xue. 2021. Demand shaping through bundling and product configuration: A dynamic multiproduct inventory-pricing model. *Operations Research* **69**(2) 525–544.

Steele, AT. 1951. Weather's effect on the sales of a department store. *Journal of Marketing* **15**(4) 436–443.

Stefanescu, Catalina. 2009. Multivariate customer demand: modeling and estimation from censored sales. *Available at SSRN 1334353* .

Strehl, Alexander, Joydeep Ghosh. 2002. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research* **3**(Dec) 583–617.

Strijbosch, Leonardus Wilhelmus Gerardus, Rudolf Maria Jozef Heuts. 1992. Modelling (s, Q) inventory systems: Parametric versus non-parametric approximations for the lead time demand distribution. *European Journal of Operational Research* **63**(1) 86–101.

Subramanyam, Anirudh, Chrysanthos E Gounaris, Wolfram Wiesemann. 2019. K-adaptability in two-stage mixed-integer robust optimization. *Mathematical Programming Computation* 1–32.

Talluri, Kalyan T, Garrett J Van Ryzin. 2006. *The theory and practice of revenue management*, vol. 68. Springer Science & Business Media.

Thomassey, Sébastien, Antonio Fiordaliso. 2006. A hybrid sales forecasting system based on clustering and decision trees. *Decision Support Systems* **42**(1) 408–421.

Vaagen, Hajnalka, Stein W Wallace. 2008. Product variety arising from hedging in the fashion supply chains. *International Journal of Production Economics* **114**(2) 431–455.

Vulcano, Gustavo, Garrett Van Ryzin, Richard Ratliff. 2012. Estimating primary demand for substitutable products from sales transaction data. *Operations Research* **60**(2) 313–334.

Webber, Bjŏrn, Sarah Herrlein, Greg Hodge. 2011. The challenge of food waste. Accessed Jan. 01, 2019. Available at: `https://www.gs1.org/sites/default/files/gs1_uk_the_challenge_of_food_waste.pdf`.

Wecker, William E. 1978. Predicting demand from sales data in the presence of stockouts. *Management Science* **24**(10) 1043–1054.

Wedel, Michel, PK Kannan. 2016. Marketing analytics for data-rich environments. *Journal of Marketing* **80**(6) 97–121.

Whitin, Thomson M. 1955. Inventory control and price theory. *Management Science* **2**(1) 61–68.

Wiesemann, Wolfram, Daniel Kuhn, Melvyn Sim. 2014. Distributionally robust convex optimization. *Operations Research* **62**(6) 1358–1376.

Wooldridge, Jeffrey M. 2016. *Introductory econometrics: A modern approach*. Nelson Education.

Yang, Yi, Youhua Chen, Yun Zhou. 2014. Coordinating inventory control and pricing strategies under batch ordering. *Operations Research* **62**(1) 25–37.

Yano, Candace A, Stephen M Gilbert. 2005. Coordinated pricing and production/procurement decisions: A review. *Managing Business Interfaces*. Springer, 65–103.

Yin, Rui, Yossi Aviv, Amit Pazgal, Christopher S Tang. 2009. Optimal markdown pricing: Implications of inventory display formats in the presence of strategic customers. *Management Science* **55**(8) 1391–1408.

Yue, Jinfeng, Bintong Chen, Min-Chiang Wang. 2006. Expected value of distribution information for the newsvendor problem. *Operations Research* **54**(6) 1128–1136.

Online Appendices of "Robust Pricing and Production with Information Partitioning and Adaptation"

Georgia Perakis, Melvyn Sim, Qinshen Tang, Peng Xiong

Online Appendix A: Proofs of statements

A.1. Proof of Proposition 3

We provided a sketch of the proof that is similar to Bertsimas et al. (2019). Note that because of the linear optimization framework, we do not require Slater's condition for strong duality to hold. Given $\boldsymbol{x}, \boldsymbol{r}^1 \in \mathcal{R}, \hat{\boldsymbol{r}}^2 \in \mathcal{M}(\boldsymbol{r}^1)$, the worst-case total revenue can be calculated as the following infinite dimensional linear program,

$$
\begin{array}{lll}
\inf & \displaystyle\sum_{(s,t)\in[S]\times[T]} q_{st}\mathbb{E}_{\mathbb{P}}\left[\pi\left(\boldsymbol{x}, \boldsymbol{r}^1, \hat{\boldsymbol{r}}^2(\tilde{\boldsymbol{w}}^1), \boldsymbol{v}^1, \boldsymbol{v}^2\right) \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}^1_s, \tilde{\boldsymbol{w}}^2 \in \mathcal{W}^2_{st}\right] & \\
\text{s.t.} & \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{v}}^1 \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}^1_s\right] = \boldsymbol{\mu}^1_s & \forall s \in [S]; \\
& \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{v}}^2 \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}^1_s, \tilde{\boldsymbol{w}}^2 \in \mathcal{W}^2_{st}\right] = \boldsymbol{\mu}^2_{st} & \forall (s,t) \in [S]\times[T]; \\
& \mathbb{E}_{\mathbb{P}}\left[|\tilde{\boldsymbol{v}}^1 - \boldsymbol{\mu}^1_s \mid |\tilde{\boldsymbol{w}}^1 \in \mathcal{W}^1_s\right] \le \boldsymbol{\sigma}^1_s & \forall s \in [S]; \\
& \mathbb{E}_{\mathbb{P}}\left[|\tilde{\boldsymbol{v}}^2 - \boldsymbol{\mu}^2_{st} \mid |\tilde{\boldsymbol{w}}^1 \in \mathcal{W}^1_s, \tilde{\boldsymbol{w}}^2 \in \mathcal{W}^2_{st}\right] \le \boldsymbol{\sigma}^2_{st} & \forall (s,t) \in [S]\times[T]; \\
& \mathbb{P}[\tilde{\boldsymbol{v}}^1 \in \mathcal{V}^1_s \mid \tilde{\boldsymbol{w}}^1 \in \mathcal{W}^1_s] = 1 & \forall s \in [S] \\
& \mathbb{P}[\tilde{\boldsymbol{v}}^2 \in \mathcal{V}^2_{st} \mid \tilde{\boldsymbol{w}}^2 \in \mathcal{W}^2_{st}] = 1 & \forall (s,t) \in [S]\times[T] \\
& \mathbb{P}\left[\tilde{\boldsymbol{w}}^1 \in \mathcal{W}^1_s, \tilde{\boldsymbol{w}}^2 \in \mathcal{W}^2_{st}\right] = q_{st} & \forall (s,t) \in [S]\times[T].
\end{array}
$$

Note that the last three set of probability constraints is equivalent to $\mathbb{P}\left[\tilde{\boldsymbol{v}}^1 \in \mathcal{V}^1_s, \tilde{\boldsymbol{v}}^2 \in \mathcal{V}^2_{st}\right] = q_{st}$ and $\mathbb{P}[\tilde{\boldsymbol{v}}^1 \in \mathcal{V}^1_s] = \sum_{t\in[T]} q_{st}, \forall (s,t) \in [S]\times[T]$. As a result, the above optimization problem is equivalent to

$$
\begin{array}{llll}
\inf & \displaystyle\sum_{(s,t)\in[S]\times[T]} q_{st}\mathbb{E}_{\mathbb{P}_{st}}\left[\pi\left(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}^2_s, \tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2\right)\right] & & \\
\text{s.t.} & \displaystyle\sum_{t\in[T]} q_{st}\mathbb{E}_{\mathbb{P}_{st}}\left[\tilde{\boldsymbol{v}}^1\right] = \sum_{t\in[T]} q_{st}\boldsymbol{\mu}^1_s, & \forall s \in [S] & : \ \boldsymbol{\beta}^1_s \\
& \mathbb{E}_{\mathbb{P}_{st}}\left[\tilde{\boldsymbol{v}}^2\right] = q_{st}\boldsymbol{\mu}^2_{st}, & \forall (s,t) \in [S]\times[T] & : \ \boldsymbol{\beta}^2_{st} \\
& \displaystyle\sum_{t\in[T]} q_{st}\mathbb{E}_{\mathbb{P}_{st}}\left[|\tilde{\boldsymbol{v}}^1 - \boldsymbol{\mu}^1_s|\right] \le \sum_{t\in[T]} q_{st}\boldsymbol{\sigma}^1_s, & \forall s \in [S] & : \ \boldsymbol{\gamma}^1_s \\
& \mathbb{E}_{\mathbb{P}_{st}}\left[|\tilde{\boldsymbol{v}}^2 - \boldsymbol{\mu}^2_{st}|\right] \le q_{st}\boldsymbol{\sigma}^2_{st}, & \forall (s,t) \in [S]\times[T] & : \ \boldsymbol{\gamma}^2_{st} \\
& \mathbb{P}_{st}\left[\tilde{\boldsymbol{v}}^1 \in \mathcal{V}^1_s, \tilde{\boldsymbol{v}}^2 \in \mathcal{V}^2_{st}\right] = 1. & \forall (s,t) \in [S]\times[T] & : \ \alpha_{st}
\end{array}
$$

where $\boldsymbol{r}^2_s$ is a new variable satisfying $\boldsymbol{r}^2_s \le \boldsymbol{r}^1$ and depends on the cluster, and $\mathbb{E}_{\mathbb{P}_{st}}[\cdot] \triangleq \mathbb{E}_{\mathbb{P}}\left[\cdot \mid \boldsymbol{v}^1 \in \mathcal{V}^1_s, \boldsymbol{v}^2 \in \mathcal{V}^2_{st}\right]$. Assigning $\boldsymbol{\beta}^1_s, \boldsymbol{\beta}^2_{st}, \boldsymbol{\gamma}^1_s, \boldsymbol{\gamma}^2_{st}$, and α_{st} as the dual variable(s) to the five type of constraints, respectively, we can obtain the dual problem as follows

$$\begin{aligned} \sup \quad & \sum_{(s,t)\in[S]\times[T]} \Big(\alpha_{st} + q_{st}\left(\boldsymbol{\beta}_s^1\right)'\boldsymbol{\mu}_s^1 + \left(\boldsymbol{\beta}_{st}^2\right)'\boldsymbol{\mu}_{st}^2 + q_{st}\left(\boldsymbol{\gamma}_s^1\right)'\boldsymbol{\sigma}_s^1 + \left(\boldsymbol{\gamma}_{st}^2\right)'\boldsymbol{\sigma}_{st}^2\Big) && \\ \text{s.t.} \quad & \alpha_{st} + q_{st}\left(\boldsymbol{\beta}_s^1\right)'\boldsymbol{v}^1 + \left(\boldsymbol{\beta}_{st}^2\right)'\boldsymbol{v}^2 + q_{st}\left(\boldsymbol{\gamma}_s^1\right)'\left|\boldsymbol{v}^1 - \boldsymbol{\mu}_s^1\right| + \left(\boldsymbol{\gamma}_{st}^2\right)'\left|\boldsymbol{v}^2 - \boldsymbol{\mu}_{st}^2\right| && \\ & \qquad\quad \le q_{st}\pi\left(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}_s^2, \boldsymbol{v}^1, \boldsymbol{v}^2\right), && \forall \boldsymbol{v}^1 \in \mathcal{V}_s^1, \boldsymbol{v}^2 \in \mathcal{V}_{st}^2, \\ & && \quad s \in [S], t \in [T] \\ & \boldsymbol{\gamma}_s^1, \boldsymbol{\gamma}_{st}^2 \le 0, && \forall (s,t) \in [S]\times[T] \\ & \alpha_s \in \mathbb{R}, \boldsymbol{\beta}_s^1, \boldsymbol{\beta}_{st}^2, \boldsymbol{\gamma}_s^1, \boldsymbol{\gamma}_{st}^2 \in \mathbb{R}^I, && \forall (s,t) \in [S]\times[T], \end{aligned}$$

which can be rewritten as follows,

$$\begin{aligned} \sup \quad & \sum_{(s,t)\in[S]\times[T]} \Big(\alpha_{st} + q_{st}\left(\boldsymbol{\beta}_s^1\right)'\boldsymbol{\mu}_s^1 + \left(\boldsymbol{\beta}_{st}^2\right)'\boldsymbol{\mu}_{st}^2 + q_{st}\left(\boldsymbol{\gamma}_s^1\right)'\boldsymbol{\sigma}_s^1 + \left(\boldsymbol{\gamma}_{st}^2\right)'\boldsymbol{\sigma}_{st}^2\Big) && \\ \text{s.t.} \quad & \alpha_{st} + q_{st}\left(\boldsymbol{\beta}_s^1\right)'\boldsymbol{v}^1 + \left(\boldsymbol{\beta}_{st}^2\right)'\boldsymbol{v}^2 + q_{st}\left(\boldsymbol{\gamma}_s^1\right)'\boldsymbol{u}^1 + \left(\boldsymbol{\gamma}_{st}^2\right)'\boldsymbol{u}^2 && \\ & \qquad\quad \le q_{st}\pi\left(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}_s^2, \boldsymbol{v}^1, \boldsymbol{v}^2\right), && \forall \left(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2\right) \in \mathcal{L}_{st}, \\ & && \quad s \in [S], t \in [T] \\ & \boldsymbol{\gamma}_s^1, \boldsymbol{\gamma}_{st}^2 \le 0, && \forall (s,t) \in [S]\times[T] \\ & \alpha_s \in \mathbb{R}, \boldsymbol{\beta}_s^1, \boldsymbol{\beta}_{st}^2, \boldsymbol{\gamma}_s^1, \boldsymbol{\gamma}_{st}^2 \in \mathbb{R}^I, && \forall (s,t) \in [S]\times[T], \end{aligned}$$

where $\mathcal{L}_{st}$ is the lifted support set defined as

$$\mathcal{L}_{st} = \left\{ \left(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2\right) \in \mathbb{R}^I \times \mathbb{R}^I \times \mathbb{R}^I \times \mathbb{R}^I \,\middle|\, \begin{array}{l} \boldsymbol{v}^1 \in \mathcal{V}_s^1,\ \boldsymbol{v}^2 \in \mathcal{V}_{st}^2 \\ \boldsymbol{u}^1 \ge \left|\boldsymbol{v}^1 - \boldsymbol{\mu}_s^1\right|,\ \boldsymbol{u}^2 \ge \left|\boldsymbol{v}^2 - \boldsymbol{\mu}_{st}^2\right| \end{array} \right\}.$$

$\blacksquare$

A.2. Proof of Theorem 1

Based on Proposition 2, given any $\left(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2\right) \in \mathcal{L}_{st}, (s,t) \in [S]\times[T]$, the total revenue can be reformulated as follows,

$$\pi\left(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}_s^2, \boldsymbol{v}^1, \boldsymbol{v}^2\right) = \sum_{i\in[I]} \min\left\{ r_i^1 x_i,\ r_{si}^2 x_i + \left(r_i^1 - r_{si}^2\right) z^1(\boldsymbol{r}^1, v_i^1),\ r_i^1 z^1(\boldsymbol{r}^1, v_i^1) + r_{si}^2 z^2(\boldsymbol{r}^1, \boldsymbol{r}_s^2, v_i^2) \right\}.$$

Let $\mathcal{I}_j, j \in \{1,2,3\}$ constitute a partition for set $[I]$, *i.e.*,

$$\mathcal{I}_1 \cup \mathcal{I}_2 \cup \mathcal{I}_3 = [I],\ \mathcal{I}_i \cap \mathcal{I}_j = \emptyset,\ \forall i,j \in \{1,2,3\}, i \ne j.$$

Denote all partitions of set $[I]$ as

$$\mathcal{B} \triangleq \left\{ \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3 \subseteq [I] \,\middle|\, \begin{array}{l} \mathcal{I}_1 \cup \mathcal{I}_2 \cup \mathcal{I}_3 = [I] \\ \mathcal{I}_i \cap \mathcal{I}_j = \emptyset \quad \forall i,j \in \{1,2,3\}, i \ne j \end{array} \right\}.$$

and let

$$\begin{aligned} & \phi\left(\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}_s^2, \boldsymbol{v}^1, \boldsymbol{v}^2\right) \\ & \triangleq \sum_{i\in\mathcal{I}_1} r_i^1 x_i + \sum_{i\in\mathcal{I}_2} \left(r_{si}^2 x_i + \left(r_i^1 - r_{si}^2\right) z^1(\boldsymbol{r}^1, v_i^1) \right) + \sum_{i\in\mathcal{I}_3} \left(r_i^1 z^1(\boldsymbol{r}^1, v_i^1) + r_{si}^2 z^2(\boldsymbol{r}^1, \boldsymbol{r}_s^2, v_i^2) \right), \end{aligned}$$

for all $\left(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2\right) \in \mathcal{L}_{st}, s \in [S], t \in [T]$, then the total revenue can be rewritten as

$$\pi\left(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}_s^2, \boldsymbol{v}^1, \boldsymbol{v}^2\right) = \min_{\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3 \in \mathcal{B}} \phi\left(\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}_s^2, \boldsymbol{v}^1, \boldsymbol{v}^2\right).$$

Hence,

$$\alpha_{st} + q_{st}\left(\boldsymbol{\beta}_s^1\right)'\boldsymbol{v}^1 + \left(\boldsymbol{\beta}_{st}^2\right)'\boldsymbol{v}^2 + q_{st}\left(\boldsymbol{\gamma}_s^1\right)'\boldsymbol{u}^1 + \left(\boldsymbol{\gamma}_{st}^2\right)'\boldsymbol{u}^2 \le q_{st}\min_{\mathcal{I}_1,\mathcal{I}_2,\mathcal{I}_3\in\mathcal{B}}\phi\left(\mathcal{I}_1,\mathcal{I}_2,\mathcal{I}_3,\boldsymbol{x},\boldsymbol{r}^1,\boldsymbol{r}^2,\boldsymbol{v}^1,\boldsymbol{v}^2\right),$$

is equivalent to

$$\alpha_{st} + q_{st}\left(\boldsymbol{\beta}_s^1\right)'\boldsymbol{v}^1 + \left(\boldsymbol{\beta}_{st}^2\right)'\boldsymbol{v}^2 + q_{st}\left(\boldsymbol{\gamma}_s^1\right)'\boldsymbol{u}^1 + \left(\boldsymbol{\gamma}_{st}^2\right)'\boldsymbol{u}^2 \le q_{st}\phi\left(\mathcal{I}_1,\mathcal{I}_2,\mathcal{I}_3,\boldsymbol{x},\boldsymbol{r}^1,\boldsymbol{r}^2,\boldsymbol{v}^1,\boldsymbol{v}^2\right), \forall\mathcal{I}_1,\mathcal{I}_2,\mathcal{I}_3\in\mathcal{B}.$$

■

A.3. Proof of Theorem 2

Following from Proposition 2, suppose $f_{sti}\in\mathcal{A}, \forall i\in[I]$, $(s,t)\in[S]\times[T]$ satisfies

$$\left.\begin{aligned}
&f_{sti}(\boldsymbol{v}^1,\boldsymbol{v}^2,\boldsymbol{u}^1,\boldsymbol{u}^2)\le r_i^1 x_i,\\
&f_{sti}(\boldsymbol{v}^1,\boldsymbol{v}^2,\boldsymbol{u}^1,\boldsymbol{u}^2)\le r_{si}^2 x_i + (r_i^1 - r_{si}^2)\left(a_i^1 + \boldsymbol{e}_i'\boldsymbol{B}^1\boldsymbol{r}^1 + v_i^1\right)\\
&f_{sti}(\boldsymbol{v}^1,\boldsymbol{v}^2,\boldsymbol{u}^1,\boldsymbol{u}^2)\le r_i^1\left(a_i^1 + \boldsymbol{e}_i'\boldsymbol{B}^1\boldsymbol{r}^1 + v_i^1\right)\\
&\qquad\qquad + r_{si}^2\left(a_i^2 + \boldsymbol{e}_i'\boldsymbol{B}^2\boldsymbol{r}_s^2 + \boldsymbol{e}_i'\boldsymbol{B}^{21}\boldsymbol{r}^1 + v_i^2\right)
\end{aligned}\right\}\forall\left(\boldsymbol{v}^1,\boldsymbol{v}^2,\boldsymbol{u}^1,\boldsymbol{u}^2\right)\in\mathcal{L}_{st},$$

we then have

$$\sum_{i\in[I]} f_{sti}(\boldsymbol{v}^1,\boldsymbol{v}^2,\boldsymbol{u}^1,\boldsymbol{u}^2)\le\pi\left(\boldsymbol{x},\boldsymbol{r}^1,\boldsymbol{r}_s^2,\boldsymbol{v}^1,\boldsymbol{v}^2\right)\qquad\forall\left(\boldsymbol{v}^1,\boldsymbol{v}^2,\boldsymbol{u}^1,\boldsymbol{u}^2\right)\in\mathcal{L}_{st},$$

and the formulation in Problem (7) ensues with appropriate linearization for some nonlinear terms in $\boldsymbol{r}^1$, $\boldsymbol{x}$, and $\boldsymbol{r}_s^2$. Specifically, for each $i\in[I]$, we let $\boldsymbol{p}_i=(p_{i1},\ldots,p_{iN})$ and introduce binary decision variables $\boldsymbol{y}_i^1,\boldsymbol{y}_{si}^2\in\{0,1\}^N$, satisfying $\mathbf{1}'\boldsymbol{y}_i^1=\mathbf{1}'\boldsymbol{y}_{si}^2=1$ so that $r_i^1=\boldsymbol{p}_i'\boldsymbol{y}_i^1$, $r_{si}^2=\boldsymbol{p}_i'\boldsymbol{y}_{si}^2$, $s\in[S]$. The constraint $\cdots\le r_i^1 x_i$ can be linearized by $\cdots\le\omega_i^1$, for some decision variable ω_i^1 satisfying

$$\omega_i^1\le p_{in}x_i + p_{iN}\bar{x}_i(1-y_{in}^1)\quad\forall n\in[N].$$

The same technique applies to the constraint $\cdots\le r_{si}^2 x_i$. Similarly, the constraint $\cdots\le r_i^1\boldsymbol{e}_i'\boldsymbol{B}^1\boldsymbol{r}^1$ can be linearized by $\cdots\le\rho_i^{11}$, for some decision variable ρ_i^{11} satisfying

$$\rho_i^{11}\le p_{in}\boldsymbol{e}_i'\boldsymbol{B}^1\left(\boldsymbol{p}_1'\boldsymbol{y}_1^1,\ldots,\boldsymbol{p}_I'\boldsymbol{y}_I^1\right)' + p_{iN}|\boldsymbol{e}_i'\boldsymbol{B}^1\left(\boldsymbol{p}_1'\boldsymbol{y}_1^1,\ldots,\boldsymbol{p}_I'\boldsymbol{y}_I^1\right)'|(1-y_{in}^1)\quad\forall n\in[N].$$

Same technique applies to $\cdots\le r_{si}^2\boldsymbol{e}_i'\boldsymbol{B}^1\boldsymbol{r}^1$, $\cdots\le r_{si}^2\boldsymbol{e}_i'\boldsymbol{B}^2\boldsymbol{r}_s^2$, and $\cdots\le r_{si}^2\boldsymbol{e}_i'\boldsymbol{B}^{21}\boldsymbol{r}^1$.

On the other hand, any optimal solution for Problem (7) is feasible for Problem (4). Hence, $Z_R\le Z^*$. We next prove the optimality of the affine recourse adaptation when $I=1$. For simplicity, we don't apply the linearization techniques during the proof. In this case, Z^* is expressed as follows

$$\begin{array}{lll}
\max & -cx + \displaystyle\sum_{(s,t)\in[S]\times[T]}\left(\alpha_{st} + q_{st}\beta_s^1\mu_s^1 + \beta_{st}^2\mu_{st}^2 + q_{st}\gamma_s^1\sigma_s^1 + \gamma_{st}^2\sigma_{st}^2\right) & \\
\text{s.t.} & \alpha_{st} + q_{st}\beta_s^1 v^1 + \beta_{st}^2 v^2 + q_{st}\gamma_s^1 u^1 + \gamma_{st}^2 u^2 \le q_{st}\pi(x,r^1,r_s^2,v^1,v^2), & \forall\,(v^1,v^2,u^1,u^2)\in\mathcal{L}_{st},\\
& & (s,t)\in[S]\times[T]\\
& \gamma_s^1,\gamma_{st}^2\le 0, & \forall(s,t)\in[S]\times[T]\\
& r^1\ge r_s^2, & \forall s\in[S]\\
& \alpha_{st},\beta_s^1,\beta_{st}^2,\gamma_s^1,\gamma_{st}^2\in\mathbb{R}, & \forall(s,t)\in[S]\times[T]\\
& r^1,r_s^2\in\mathcal{R} & \forall s\in[S]\\
& x\in\mathcal{X}. &
\end{array}\tag{A.1}$$

while Z_R is

$$
\begin{array}{lll}
\max & -cx + \displaystyle\sum_{(s,t)\in[S]\times[T]} \left(\alpha_{st} + q_{st}\beta_s^1\mu_s^1 + \beta_{st}^2\mu_{st}^2 + q_{st}\gamma_s^1\sigma_s^1 + \gamma_{st}^2\sigma_{st}^2\right) & \\
\text{s.t.} & \left.\begin{array}{l} \alpha_{st} + q_{st}\beta_s^1 v^1 + \beta_{st}^2 v^2 + q_{st}\gamma_s^1 u^1 + \gamma_{st}^2 u^2 \\ \qquad\qquad \le f_{st}^0 + f_s^1 v^1 + f_{st}^2 v^2 + f_s^3 u^1 + f_{st}^4 u^2, \\ f_{st}^0 + f_s^1 v^1 + f_{st}^2 v^2 + f_s^3 u^1 + f_{st}^4 u^2 \le q_{st}\pi(x, r^1, r_s^2, v^1, v^2), \end{array}\right\} & \begin{array}{l} \forall\,(v^1,v^2,u^1,u^2)\in\mathcal{L}_{st} \\ (s,t)\in[S]\times[T] \end{array} \\
& \gamma_s^1, \gamma_{st}^2 \le 0, & \forall (s,t)\in[S]\times[T] \\
& r^1 \ge r_s^2, & \forall s\in[S] \\
& \alpha_{st}, \beta_s^1, \beta_{st}^2, \gamma_s^1, \gamma_{st}^2, f_{st}^0, f_s^1, f_{st}^2, f_s^3, f_{st}^4 \in \mathbb{R}, & \forall (s,t)\in[S]\times[T] \\
& r^1, r_s^2 \in \mathcal{R} & \forall s\in[S] \\
& x \in \mathcal{X}, &
\end{array}
\tag{A.2}
$$

Let $\alpha_{st}, q_{st}\beta_s^1, \beta_{st}^2, q_{st}\gamma_s^1, \gamma_{st}^2$, respectively equal to $f_{st}^0, f_s^1, f_{st}^2, f_s^3, f_{st}^4$ associated with $f_{st}, s\in[S], t\in[T]$, then it is clear that

$$\alpha_{st} + q_{st}\beta_s^1 v^1 + \beta_{st}^2 v^2 + q_{st}\gamma_s^1 u^1 + \gamma_{st}^2 u^2 = f_{st}(v^1, v^2, u^1, u^2), \forall\left(v^1, v^2, u^1, u^2\right) \in \mathcal{L}_{st},$$

which ensures that for any optimal solution in Problem (A.1), we can always construct it as a feasible solution for Problem (A.2). Hence, $Z_R \le Z^* \le Z_R$, *i.e.*, $Z_R = Z^*$. ■

A.4. Proof of Proposition 4

For any $(s,t)\in[S]\times[T]$, the inequality $(\boldsymbol{p}^1)'\boldsymbol{v}^1 + (\boldsymbol{q}^1)'\boldsymbol{u}^1 + (\boldsymbol{p}^2)'\boldsymbol{v}^2 + (\boldsymbol{q}^2)'\boldsymbol{u}^2 \le g, (\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \in \mathcal{L}_{st}$ is equivalent to $RC \le g$, where

$$
\begin{array}{rlll}
RC = & \max & (\boldsymbol{p}^1)'\boldsymbol{v}^1 + (\boldsymbol{q}^1)'\boldsymbol{u}^1 + (\boldsymbol{p}^2)'\boldsymbol{v}^2 + (\boldsymbol{q}^2)'\boldsymbol{u}^2 & \\
& s.t. & \boldsymbol{A}_s^1\boldsymbol{v}^1 \le \boldsymbol{b}_s^1 & : \boldsymbol{\kappa}_s^1 \\
& & \boldsymbol{A}_{st}^2\boldsymbol{v}^2 \le \boldsymbol{b}_{st}^2 & : \boldsymbol{\kappa}_{st}^2 \\
& & \boldsymbol{u}^1 - \boldsymbol{v}^1 \ge -\boldsymbol{\mu}_s^1, & : \boldsymbol{\tau}_s^{11} \\
& & \boldsymbol{u}^1 + \boldsymbol{v}^1 \ge \boldsymbol{\mu}_s^1, & : \boldsymbol{\tau}_s^{12} \\
& & \boldsymbol{u}^2 - \boldsymbol{v}^2 \ge -\boldsymbol{\mu}_{st}^2, & : \boldsymbol{\tau}_{st}^{21} \\
& & \boldsymbol{u}^2 + \boldsymbol{v}^2 \ge \boldsymbol{\mu}_{st}^2, & : \boldsymbol{\tau}_{st}^{22}
\end{array}
\tag{A.3}
$$

The dual of the above linear programming is as follows

$$
\begin{array}{l}
(\boldsymbol{\mu}_s^1)'(-\boldsymbol{\tau}_s^{11} + \boldsymbol{\tau}_s^{12}) - (\boldsymbol{\mu}_{st}^2)'(\boldsymbol{\tau}_{st}^{21} - \boldsymbol{\tau}_{st}^{22}) + (\boldsymbol{\kappa}_s^1)'\boldsymbol{b}_s^1 + (\boldsymbol{\kappa}_{st}^2)'\boldsymbol{b}_{st}^2 \le g \\
-\boldsymbol{\tau}_s^{11} + \boldsymbol{\tau}_s^{12} + \boldsymbol{A}_s^1\kappa_s^1 = \boldsymbol{p}^1; \\
-\boldsymbol{\tau}_{st}^{21} + \boldsymbol{\tau}_{st}^{22} + \boldsymbol{A}_{st}^2\kappa_{st}^2 = \boldsymbol{p}^2; \\
\boldsymbol{\tau}_s^{11} + \boldsymbol{\tau}_s^{12} = \boldsymbol{q}^1; \\
\boldsymbol{\tau}_{st}^{21} + \boldsymbol{\tau}_{st}^{22} = \boldsymbol{q}^2; \\
\boldsymbol{\kappa}_s^1, \boldsymbol{\kappa}_{st}^2 \ge 0; \\
\boldsymbol{\tau}_s^{11}, \boldsymbol{\tau}_s^{12}, \boldsymbol{\tau}_{st}^{21}, \boldsymbol{\tau}_{st}^{22} \le 0.
\end{array}
\tag{A.4}
$$

■

Online Appendix B: Concave piecewise linear demand function

In this Appendix, we show that all our main results still hold if the demand for each product is concave piecewise linear in its retail price. Specifically, we now have

$$z_i^1(\boldsymbol{r}^1, \tilde{v}_i^1) = y_i^1(\boldsymbol{r}^1) + \tilde{v}_i^1, i \in [I], s \in [S] \tag{B.5}$$

$$z_i^2(\boldsymbol{r}^1, \boldsymbol{r}^2, \tilde{v}_i^2) = y_i^2(\boldsymbol{r}^1, \boldsymbol{r}^2) + \tilde{v}_i^2, i \in [I] \tag{B.6}$$

where $y_i^1(\boldsymbol{r}^1) = \min_{p_i^1 \in [P_i^1]} \left\{ a_{p_i^1 i} + B^1_{p_i^1 ii} r_i^1 \right\} + \sum_{m \neq i} B^1_{mi} r^1_m$, $y_i^2(\boldsymbol{r}^1, \boldsymbol{r}^2) = \min_{p_i^2 \in [P_i^2]} \left\{ a^2_{p_i^2 i} + B^2_{p_i^2 ii} r_i^2 \right\} + \sum_{m \neq i} B^2_{mi} r^2_m + \sum_{n \in [I]} B^{21}_{ni} r^1_m$, and P_i^t is the given number of pieces for product $i \in [I]$ at period $\tau \in \{1, 2\}$.

It is obvious that Proposition 2 and Proposition 3 remain the same, except that we need to substitute $z_i^1(\boldsymbol{r}^1, \tilde{v}_i^1)$ and $z_i^2(\boldsymbol{r}^1, \boldsymbol{r}^2, \tilde{v}_i^2)$ by using equation (B.5) and (B.6) respectively. However, Theorem 1 does not hold because it is difficult to find a way to enumerate all the possible cases of the total revenue π.

As for Theorem 2, Proposition 2 still holds, now suppose $f_{sti} \in \mathcal{A}, \forall i \in [I]$, $(s,t) \in [S] \times [T]$ and satisfies

$$\left.\begin{aligned} & f_{sti}(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \le r_i^1 x_i, \\ & f_{sti}(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \le r_{si}^2 x_i + (r_i^1 - r_{si}^2)\left(y_i^1(\boldsymbol{r}^1) + v_i^1\right) \\ & f_{sti}(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \le r_i^1 \left(y_i^1(\boldsymbol{r}^1) + \boldsymbol{v}_i^1\right) + r_{si}^2 \left(y_i^1(\boldsymbol{r}^1, \boldsymbol{r}^2) + v_i^2\right) \end{aligned}\right\} \forall \left(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2\right) \in \mathcal{L}_{st},$$

we still have

$$\sum_{i \in [I]} f_{sti}(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \le \pi\left(\boldsymbol{x}, \boldsymbol{r}^1, \boldsymbol{r}_s^2, \boldsymbol{v}^1, \boldsymbol{v}^2\right) \qquad \forall \left(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2\right) \in \mathcal{L}_{st}.$$

The only problem is how we can linearize these constraints. The first set of constraints $f_{sti}(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \le r_i^1 x_i$ remains the same, while for the second set of constraints

$$f_{sti}(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \le r_{si}^2 x_i + (r_i^1 - r_{si}^2)\left(y_i^1(\boldsymbol{r}^1) + v_i^1\right),$$

it is equivalent to

$$\begin{aligned} f_{sti}(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) &\le r_{si}^2 x_i + (r_i^1 - r_{si}^2)\big(\min_{p_i^1 \in [P_i^1]} \left\{ a_{p_i^1 i} + B^1_{p_i^1 ii} r_i^1 \right\} + \sum_{m \neq i} B^1_{mi} r^1_m + v_i^1\big) \\ &= \min_{p_i^1 \in [P_i^1]} \left\{ r_{si}^2 x_i + (r_i^1 - r_{si}^2)\big(a_{p_i^1 i} + B^1_{p_i^1 ii} r_i^1 + \sum_{m \neq i} B^1_{mi} r^1_m + v_i^1\big) \right\} \\ &= r_{si}^2 x_i + (r_i^1 - r_{si}^2)\big(a_{p_i^1 i} + B^1_{p_i^1 ii} r_i^1 + \sum_{m \neq i} B^1_{mi} r^1_m + v_i^1\big), \quad \forall p_i^1 \in [P_i^1]. \end{aligned}$$

Its robust counterpart can be easily derived accoring to Proposisition 4.

Now let us look at the third set of constraints

$$f_{sti}(\boldsymbol{v}^1, \boldsymbol{v}^2, \boldsymbol{u}^1, \boldsymbol{u}^2) \le r_i^1 \left(y_i^1(\boldsymbol{r}^1) + v_i^1\right) + r_{si}^2 \left(y_i^1(\boldsymbol{r}^1, \boldsymbol{r}^2) + v_i^2\right).$$

Note that for the right hand side, there are $P_i^1 P_i^2$ possible cases. One way of reformulation is to enumerate all the cases out so that we can linearize this set of constraints. Another way is to apply the affine recourse adaptation again for the two inner min functions for each product. Therefore, when the two-period demands are expressed by equations (B.5) and (B.6), respectively, we can still reformulate the Model (3) into a mixed integer linear optimization problem.

For the "tightness" result when $I = 1$, the proof is similar, we thus omit it here.

Online Appendix C: Linearization of the empirical model

The full empirical model is

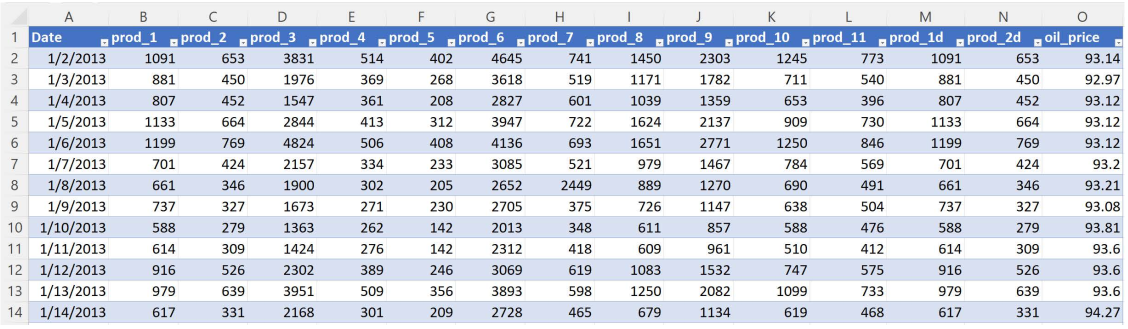
$$\max_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{r}^1,\boldsymbol{r}^2\in\mathcal{R},\boldsymbol{r}^2\le\boldsymbol{r}^1} -\boldsymbol{c}'\boldsymbol{x}+\frac{1}{H}\sum_{h\in[H]}\sum_{i\in[I]}\pi(x_i,r_i^1,r_i^2,v_{ih}^1,v_{ih}^2).$$

By using similar linearization techniques as in the proof of Theorem 2, we can obtain the below linearized empirical model:

$$\begin{array}{rlll}
Z_{\text{EM}}= & \max & -\boldsymbol{c}'\boldsymbol{x}+\dfrac{1}{H}\displaystyle\sum_{h\in[H]}\sum_{i\in[I]} f_{ih} & \\
& \text{s.t.} & f_{ih}\le\omega_i^1 & \forall h\in[H]; \\
& & f_{ih}\le\omega_i^2+(\boldsymbol{p}_i'\boldsymbol{y}_i^1-\boldsymbol{p}_i'\boldsymbol{y}_i^2)(a_i^1+v_{ih}^1)+\rho_i^{11}+\rho_i^{12} & \forall h\in[H]; \\
& & f_{ih}\le\boldsymbol{p}_i'\boldsymbol{y}_i^1(a_i^1+v_{ih}^1)+\rho_i^{11}+\boldsymbol{p}_i'\boldsymbol{y}_i^2(a_i^2+v_{ih}^2)+\rho_i^{22}+\rho_i^{21} & \forall h\in[H]; \\
& & \omega_i^1\le p_{in}x_i+M(1-y_{in}^1) & \forall i\in[I],\, n\in[N]; \\
& & \omega_i^2\le p_{in}x_i+M(1-y_{in}^2) & \forall i\in[I],\, n\in[N]; \\
& & \rho_i^{11}\le p_{in}\boldsymbol{e}_i'\boldsymbol{B}^1(\boldsymbol{p}_1'\boldsymbol{y}_1^1,\ldots,\boldsymbol{p}_I'\boldsymbol{y}_I^1)'+M(1-y_{in}^1) & \forall i\in[I],\, n\in[N]; \\
& & \rho_i^{12}\le -p_{in}\boldsymbol{e}_i'\boldsymbol{B}^1(\boldsymbol{p}_1'\boldsymbol{y}_1^1,\ldots,\boldsymbol{p}_I'\boldsymbol{y}_I^1)'+M(1-y_{in}^2) & \forall i\in[I],\, n\in[N]; \\
& & \rho_i^{21}\le p_{in}\boldsymbol{e}_i'\boldsymbol{B}^{21}(\boldsymbol{p}_1'\boldsymbol{y}_1^1,\ldots,\boldsymbol{p}_I'\boldsymbol{y}_I^1)'+M(1-y_{in}^2) & \forall i\in[I],\, n\in[N]; \\
& & \rho_i^{22}\le p_{in}\boldsymbol{e}_i'\boldsymbol{B}^2(\boldsymbol{p}_1'\boldsymbol{y}_1^2,\ldots,\boldsymbol{p}_I'\boldsymbol{y}_I^2)'+M(1-y_{in}^2) & \forall i\in[I],\, n\in[N]; \\
& & \boldsymbol{p}_i'\boldsymbol{y}_i^1\ge\boldsymbol{p}_i'\boldsymbol{y}_i^2 & \forall i\in[I]; \\
& & \boldsymbol{1}'\boldsymbol{y}_i^1=1,\ \boldsymbol{1}'\boldsymbol{y}_i^2=1 & \forall i\in[I]; \\
& & \boldsymbol{y}_i^1,\boldsymbol{y}_i^2\in\{0,1\}^N & \forall i\in[I]; \\
& & f_{ih}\in\mathbb{R},\ \boldsymbol{\omega}^1,\boldsymbol{\omega}^2,\boldsymbol{\rho}^{11},\boldsymbol{\rho}^{12},\boldsymbol{\rho}^{21},\boldsymbol{\rho}^{22}\in\mathbb{R}^I & \forall i\in[I],\, h\in[H]; \\
& & \boldsymbol{x}\in\mathcal{X}. &
\end{array}$$

Online Appendix D: Suppliment to multi-item newsvendor experimental study

A screenshot of the data obtained from Kaggle after processing is shown in Figure D.1. We present the regression results of demand of each product on the oil price in Table D.1.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Date	prod_1	prod_2	prod_3	prod_4	prod_5	prod_6	prod_7	prod_8	prod_9	prod_10	prod_11	prod_1d	prod_2d	oil_price
2	1/2/2013	1091	653	3831	514	402	4645	741	1450	2303	1245	773	1091	653	93.14
3	1/3/2013	881	450	1976	369	268	3618	519	1171	1782	711	540	881	450	92.97
4	1/4/2013	807	452	1547	361	208	2827	601	1039	1359	653	396	807	452	93.12
5	1/5/2013	1133	664	2844	413	312	3947	722	1624	2137	909	730	1133	664	93.12
6	1/6/2013	1199	769	4824	506	408	4136	693	1651	2771	1250	846	1199	769	93.12
7	1/7/2013	701	424	2157	334	233	3085	521	979	1467	784	569	701	424	93.2
8	1/8/2013	661	346	1900	302	205	2652	2449	889	1270	690	491	661	346	93.21
9	1/9/2013	737	327	1673	271	230	2705	375	726	1147	638	504	737	327	93.08
10	1/10/2013	588	279	1363	262	142	2013	348	611	857	588	476	588	279	93.81
11	1/11/2013	614	309	1424	276	142	2312	418	609	961	510	412	614	309	93.6
12	1/12/2013	916	526	2302	389	246	3069	619	1083	1532	747	575	916	526	93.6
13	1/13/2013	979	639	3951	509	356	3893	598	1250	2082	1099	733	979	639	93.6
14	1/14/2013	617	331	2168	301	209	2728	465	679	1134	619	468	617	331	94.27

Figure D.1 A screenshot of the dataset from Kaggle

Table D.1 **Regression results of demand on oil price**

	Intercept	oil_price	Pr($>\|t\|$)
prod_1	545.77	1.85	<2e-16 ***
prod_2	766.36	-2.69	<2e-16 ***
prod_3	742.70	9.71	<2e-16 ***
prod_4	358.72	1.71	<2e-16 ***
prod_5	469.39	-1.52	<2e-16 ***
prod_6	2038.55	8.62	<2e-16 ***
prod_7	1151.18	-5.92	<2e-16 ***
prod_8	466.99	3.70	<2e-16 ***
prod_9	750.99	6.80	<2e-16 ***
prod_10	350.42	3.79	<2e-16 ***
prod_11	1082.57	-4.60	<2e-16 ***

	A	B	C	D	E	F	G	H	I	J	K	L
1	Date	Year	Demand_1	Price_1	Demand_2	Price_2	Demand_3	Price_3	Demand_4	Price_4	Demand_5	Price_5
2	8/21/2007	2007	40	10.202	80	13.374	39	7.755	61	21.132	49	18.889
3	8/22/2007	2007	33	11.153	69	13.438	40	8.088	63	21.109	56	18.715
4	8/23/2007	2007	39	10.685	91	13.478	52	7.910	56	20.983	47	18.052
5	8/24/2007	2007	45	10.759	97	13.320	26	7.742	70	21.288	61	18.420
6	8/25/2007	2007	69	10.217	141	13.331	82	7.664	95	21.005	73	18.319
7	8/26/2007	2007	67	10.488	137	13.391	65	7.784	90	21.238	94	18.489
8	8/27/2007	2007	40	10.928	54	13.292	24	7.717	36	21.462	34	18.200
9	8/28/2007	2007	55	9.378	79	13.206	22	7.776	41	20.749	48	18.449
10	8/29/2007	2007	48	10.461	56	13.276	42	7.799	40	20.761	40	18.472
11	8/30/2007	2007	45	9.879	63	13.463	35	7.503	42	20.963	43	18.495
12	8/31/2007	2007	37	8.075	64	13.337	14	7.989	36	20.963	47	17.295
13	9/1/2007	2007	47	10.498	95	13.409	24	7.677	78	20.826	66	18.392
14	9/2/2007	2007	77	9.533	79	13.241	27	7.684	80	20.865	97	18.299
15	9/3/2007	2007	12	11.809	26	14.798	5	8.221	9	23.767	13	20.777

Figure E.2 **A screenshot of the dataset**

The dataset does not include the price and cost information. We use some prices from the case study in Section 5 as follows,

$$\boldsymbol{r} = (16.538, 21.916, 11.180, 26.013, 23.231, 26.538, 11.916, 21.180, 16.0134, 20.508, 14.126, 14.281, 17.012)'.$$

We set the cost $\boldsymbol{c}$ to be proportional to $\boldsymbol{r}$. In our numerical study, we let $\boldsymbol{c} = 0.35\boldsymbol{r}$.

We present the results using the regression tree method in Table D.2.

Table D.2 **Experimental study of multi-item newsvendor with information: regression tree method**

Group	EM	2	3	4	5	6	7
1	27423	-0.57%	0.36%	0.47%	-0.12%	-0.51%	-0.97%
2	53784	0.92%	0.80%	0.78%	0.81%	0.83%	0.65%
3	54585	0.47%	1.34%	1.48%	1.48%	1.48%	1.12%
4	30469	1.11%	1.07%	1.59%	1.34%	1.63%	2.24%
5	27925	1.41%	1.68%	1.31%	1.53%	1.85%	2.33%

Online Appendix E: Data processing to obtain additive demand parameters

We first depict the data structure with a screenshot in Figure E.2.

Step 1. Select the top twelve products according to the total sales quantity across 187 stores.

Step 2. Filter stores that do not sell all of these products. In total, 89 stores are selected.

Step 3. Aggregate demand over these 89 stores for each SKU.

Step 4. Normalize the retail price.

Step 5. Chronologically partition the data into training data (60% of the data) and testing data (40% of the data).

Step 6. Construct sample paths from each data set: take the sales for the first day as the demand for the first period, and the subsequent sales as the demand for the second period, so on so forth. In total, there are 916 and 611 sample paths (respectively) for the training data and testing data.

Step 7. Deal with outliers: denote $q(x)$ as the xth % quantile, we cap those observations outside the lower limit with the value of $q(5)$ and those that lie above the upper limit, with the value of $q(95)$. The lower and upper limits are defined as:lower $\text{limit} = \text{q}(10) - 2\text{IQR}$, upper $\text{limit} = \text{q}(90) + 2\text{IQR}$. Here $\text{IQR} = \text{q}(75) - \text{q}(25)$.

Test for the cross-product effect

To test for the cross-product effect in the aggregate over store level, we run the linear regression on the unpartitioned data for $i \in \{1, \ldots, 12\}$ on the model of the form:

$$z_{ih} = a_{ih} + B_{ii} r_{ih} + \sum_{m \neq i} B_{mi} r_{mh} + \boldsymbol{t}_i \mathbf{Year} + v_{ih}.$$

A summary the results for all the twelve products is in Table E.4. We also test the cross-product effect on the SKU level in the store with the most sales. One such example is for product 6 in Table 6.

Obtaining demand parameters

We sequentially partition the twelve products into five groups with each group having four products. For example, product 1 to 4 are in group 1, and product 3 to 6 are in group 2. We then run the linear regression model for each product within each group and implement model selection for each model. For those estimators that are not significant after the model selection, we set it to be 0. For example, for group 1, the parameters for the first-period demand model are shown in Table E.3.

Table E.3 Group 1: demand parameters

	z_1^1	z_2^1	z_3^1	z_4^1
intercept	123.8494	209.3457	98.9150	183.5499
r_1^1	-7.6904	-4.0066	0	-2.9678
r_2^1	0	-2.8125	0	-1.4824
r_3^1	0	0	-5.6138	-1.3409
r_4^1	0	-3.1783	0	-3.6942
$\mathbb{1}(Year_{2008})$	16.4519	10.7582	-1.0468	7.6760
$\mathbb{1}(Year_{2009})$	24.2390	17.1783	0.8701	12.5888
$\mathbb{1}(Year_{2010})$	33.2558	22.9417	-46.9984	14.9437
$\mathbb{1}(Year_{2011})$	27.7001	25.1247	-49.4063	14.3893
$\mathbb{1}(Year_{2012})$	51.0636	35.8613	-34.6589	15.9042

Table E.4 A summary of linear regression results

	Dependent variables:											
	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
r_1	-9.083***	-5.121***	0.017	-3.089***	-2.517***	-2.336***	-1.584***	-2.822***	-1.593***	-2.844***	-1.837***	-1.296***
	(0.612)	(0.644)	(0.537)	(0.332)	(0.817)	(0.304)	(0.184)	(0.417)	(0.228)	(0.523)	(0.232)	(0.131)
r_2	-0.070	-7.208***	0.058	-1.140***	1.443**	0.396*	-0.258*	0.879***	-0.034	0.074	0.260	0.284***
	(0.441)	(0.464)	(0.387)	(0.239)	(0.589)	(0.219)	(0.132)	(0.300)	(0.164)	(0.377)	(0.167)	(0.094)
r_3	-0.045	4.804***	-3.878***	2.528***	2.571**	-3.045***	0.799***	1.436***	1.238***	3.271***	1.247***	1.475***
	(0.777)	(0.817)	(0.682)	(0.421)	(1.037)	(0.385)	(0.233)	(0.529)	(0.289)	(0.664)	(0.295)	(0.166)
r_4	1.358**	-0.986*	-0.230	-2.395***	0.935	-1.084***	0.004	-1.339***	-0.254	0.192	0.264	-0.132
	(0.536)	(0.564)	(0.470)	(0.291)	(0.716)	(0.266)	(0.161)	(0.365)	(0.200)	(0.458)	(0.203)	(0.115)
r_5	1.670***	-0.639	0.722**	-0.783***	-8.775***	0.872***	-0.158	-0.213	-0.438***	-1.308***	-0.132	-0.069
	(0.419)	(0.441)	(0.368)	(0.227)	(0.560)	(0.208)	(0.126)	(0.285)	(0.156)	(0.358)	(0.159)	(0.090)
r_6	-0.724*	0.214	0.047	0.047	-1.127**	-4.015***	-0.299**	0.213	-0.209	-1.226***	-0.615***	0.241***
	(0.429)	(0.452)	(0.377)	(0.233)	(0.573)	(0.213)	(0.129)	(0.292)	(0.160)	(0.367)	(0.163)	(0.092)
r_7	2.040***	0.863*	-0.208	-0.677***	0.241	0.594**	-2.035***	-0.902***	0.312*	-0.400	-0.598***	0.038
	(0.477)	(0.502)	(0.419)	(0.259)	(0.637)	(0.237)	(0.143)	(0.325)	(0.178)	(0.408)	(0.181)	(0.102)
r_8	-0.372	-0.824***	-1.208***	-0.595***	-0.868**	0.131	-0.160**	-3.505***	-0.268***	-1.079***	-0.622***	-0.274***
	(0.266)	(0.280)	(0.233)	(0.144)	(0.355)	(0.132)	(0.080)	(0.181)	(0.099)	(0.227)	(0.101)	(0.057)
r_9	0.995**	-0.709*	-0.025	-0.170	-0.332	0.387**	0.204*	-0.124	-2.290***	1.190***	-0.074	-0.375***
	(0.390)	(0.410)	(0.342)	(0.212)	(0.521)	(0.193)	(0.117)	(0.265)	(0.145)	(0.333)	(0.148)	(0.083)
r_{10}	-0.123	0.218	-0.581	-0.164	-2.641***	0.037	0.392***	-0.801**	0.546***	-4.914***	-0.061	-0.170*
	(0.477)	(0.502)	(0.419)	(0.259)	(0.637)	(0.237)	(0.143)	(0.325)	(0.178)	(0.408)	(0.181)	(0.102)
r_{11}	-0.210	2.009***	0.578	0.078	1.445**	0.493**	0.247*	-0.352	-0.164	0.534	-1.503***	-0.217**
	(0.485)	(0.510)	(0.426)	(0.263)	(0.647)	(0.241)	(0.146)	(0.330)	(0.181)	(0.414)	(0.184)	(0.104)
r_{12}	-0.777**	-0.775**	0.345	-0.457**	-1.564***	0.138	-0.181*	0.120	-0.164	-1.126***	-0.353***	-0.954***
	(0.359)	(0.378)	(0.315)	(0.195)	(0.480)	(0.178)	(0.108)	(0.244)	(0.134)	(0.307)	(0.136)	(0.077)
Intercept	83.707***	205.260***	104.498***	170.184***	187.422***	77.834***	60.323***	181.707***	67.346***	143.758***	76.819***	47.485***
	(11.026)	(11.602)	(9.674)	(5.982)	(14.720)	(5.471)	(3.310)	(7.502)	(4.104)	(9.420)	(4.183)	(2.357)

Online Appendix F: RSOME code of Model (9)

We present the code of the main model (7) below to show that our model can be easily implemented by the algebraic modeling package RSOME via using Gurobi as solver.

```
function [model, prod, pr1, pr2] = RPPIPA(const, part, S, T, adapt)
% RPPIPA solves the robust pricing and production problem with information
    partitioning and adaptation.

%% Adaptation
if nargin < 5
    adapt = true;
end

%% Create the model and retrieve parameters
model = rsome;
model.Param.solver='gurobi';
model.Param.mipgap = 1e-3;

BigM = 1e4;                         % Big-M constant for linearization
c = const.c;                        % Costs of products
p1 = const.p1;                      % Admissible price for products in period
    1
p2 = const.p2;                      % Admissible price for products in period
    2

I = length(c);                      % The number of products
k = length(part);                   % The total number of partitions
N1 = size(p1, 2);                   % The number of price options for period 1
N2 = size(p2, 2);                   % The number of price options for period 2

%% Random variables and the ambiguity set
e1 = model.random(I);               % Random errors of period 1
e2 = model.random(I);               % Random errors of period 2
u1 = model.random(I);
u2 = model.random(I);

F = model.ambiguity(k);             % An ambiguity set with k scenarios
Pr = [part.Pr]';                    % Probabilities for all partition
for s = 1: S
    for t = 1: T
        h = (s-1)*T + t;
        MAD1 = part(h).MAD1';
        MAD2 = part(h).MAD2';
        Bound1 = part(h).Bounds1';
        Bound2 = part(h).Bounds2';
        F(h).suppset(e1 >= Bound1(:, 1), e1 <= Bound1(:, 2), ...
                     e2 >= Bound2(:, 1), e2 <= Bound2(:, 2), ...
                     abs(e1) <= u1, ...
                     abs(e2) <= u2);         % Define the support
        F(h).exptset(expect(e2) == 0, ...
                     expect(u2) == MAD2);    % Define the expectation set
    end
    h1 = (s-1)*T + (1:T);
    F(h1).exptset(expect(e1) == 0, ...
                  expect(u1) == MAD1);
```

```
end
pr = F.prob;                                % Vector of partition
   probabilities
F.probset(pr == Pr);                        % Values of partition
   probabilities
model.with(F);                              % Associate F with the model

%% Decisions and adaptive rules
x = model.decision(I);                % Production decision x
y1 = model.decision(I, N1, 'B');      % Price policy for period 1
y2 = model.decision(I, N2, 'B');      % Price policy for period 2
omega1 = model.decision(I);           % sum(w1, 2) = r1*x
omega2 = model.decision(I);           % sum(w2, 2) = r2*x

rho12 = model.decision(I);            %-r2.*(B1*r1)
rho21 = model.decision(I);            % r2.*(B21*r1)
rho11 = model.decision(I);            % r1.*(B1*r1)
rho22 = model.decision(I);            % r2.*(B2*r2)

f = model.decision(I);                % Piecewise expression for each
   partition
f.affadapt(e1);
f.affadapt(u1);
f.affadapt(e2);
f.affadapt(u2);

for s = 1: S
    h1 = (s-1)*T + (1:T);             % Indices of partitions in period 1
    if adapt
        y2.evtadapt(h1);
        omega2.evtadapt(h1);
    end
    rho12.evtadapt(h1);
end
for h = 1: k                          % pi adapts to all partitions
    rho21.evtadapt(h);
    rho22.evtadapt(h);
    f.evtadapt(h);
end

%% Objective function and constraints
model.max(expect(-c'*x + sum(f)));

model.append(x >= 0);                 % Productions are non-negative
model.append(y1 * ones(N1,1) == 1);   % Select only one price option
model.append(y2 * ones(N2,1) == 1);   % Select only one price option


r1 = sum(p1 .* y1, 2);
r2 = sum(p2 .* y2, 2);                % All prices for period 2

model.append(omega1 * ones(1, N1) <= p1 .* (x * ones(1, N1)) + (1 - y1)*
   BigM);
model.append(omega2 * ones(1, N2) <= p2 .* (x * ones(1, N2)) + (1 - y2)*
   BigM);
```

```
dv1 = (r1 - r2) .* e1;
dv2 = r1 .* e1 + r2 .*e2;
model.append(f <= omega1);
for h = 1: S*T
    a1 = part(h).Coeff1.a1;
    a2 = part(h).Coeff2.a2;
    B11 = part(h).Coeff1.B11;
    B21 = part(h).Coeff2.B21;
    B22 = part(h).Coeff2.B22;
    model.append(f <= omega2 + (r1 - r2).*a1 + rho11 + dv1 + rho12, h);
    model.append(f <= r1.*a1 + r2.*a2 + rho11 + rho22 + dv2 + rho21, h);
    model.append(rho12*ones(1, N1) <= -(B11*r1*ones(1, N1)).*p2 + BigM*(1
       - y2), h);
    model.append(rho21*ones(1, N2) <= (B21*r1*ones(1, N1)).*p2 + BigM*(1 -
        y2), h);
    model.append(rho22*ones(1, N2) <= (B22*r2*ones(1, N2)).*p2 + BigM*(1 -
        y2), h);
    model.append(rho11*ones(1, N1) <= (B11*r1*ones(1, N1)).*p1 + BigM * (1
        - y1));

    model.append(a2 + B21*r2 + B22*r1 + e2 >= 0, h);
end
model.append(r1 >= r2);

model.solve();

%% Solutions
prod = x.get;
pr1 = sum(p1 .* y1.get, 2);
yTable = y2.get;
if adapt
    %pr2 = table();
    %for name = yTable.Properties.VariableNames
    pr2_mat = zeros(I, S);
    for i = 1:S
        %column = array2table(sum(p2 .* table2array(yTable(:, name)), 2),
            ...
        %                     'VariableNames',name);
        %pr2 = [pr2, column];
        pr2_mat(:, i) = sum(p2 .* table2array(yTable(:, i)), 2);
    end
    pr2 = array2table(pr2_mat, 'VariableNames', yTable.Properties.
       VariableNames);
else
    pr2 = sum(p2 .* y2.get, 2);
end

end
```