

Joint Pricing and Production

A Fusion of Machine Learning and Robust Optimization

Georgia Perakis

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

Melvyn Sim, Qinshen Tang, Peng Xiong

Department of Analytics & Operations, NUS Business School, National University of Singapore
dscsim@nus.edu.sg, tang@u.nus.edu, xiongpengnus@gmail.com

We integrate machine learning with distributionally robust optimization to address a two-period problem for the joint pricing and production of multiple items. First, we generalize the additive demand model to capture both cross-product and cross-period effects as well as the demand dependence across periods. Next, we apply K-means clustering to the demand residual mapping based on historical data and then construct a K-means ambiguity set on that residual while specifying only the mean, the support, and the mean absolute deviation. Finally, we investigate the joint pricing and production problem by proposing a K-means adaptive markdown policy and an affine recourse approximation; the latter allows us to reformulate the problem as an approximate but more tractable mixed-integer linear programming problem. Both the case study and our simulation demonstrate that, with only a few clusters, the K-means adaptive markdown policy and ambiguity set can increase expected profits by 1.11% on average and by as much as 2.22%—as compared with the empirical model—when applied to most out-of-sample tests.

Key words: multi-item, pricing, inventory control, K-means clustering, distributionally robust optimization

History: January 1, 2019

1. Introduction

Managers, especially those in the retail or fashion industry, usually prepare their inventories well before the products are launched because of such 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](#)) and hence, increase the importance of matching demand and supply, especially since managers often overstock or understock the product. According to a study by the IHL Group (a global research and advisory firm), more than \$1.1 trillion of costs due to such “inventory distortion” were observed in the retail industry in 2015—an amount that was expected to increase by nearly \$100 billion annually worldwide ([Buzek 2015](#)). The Swedish fashion retailer H&M has recently struggled with a mounting stack of unsold inventory valued at some \$4.3 billion

(Paton 2018). According to the U.S. Census Bureau, inventory held by retailers increased from \$240 billion in 1992 to \$640 billion in 2018 (U.S. Bureau of the Census 2018). All these statements indicate the need for efficient and effective production management.

Production management traditionally involves making optimal decisions about production levels. However, the uncertainty of demand renders this decision insufficient, in itself, for matching supply (Van Mieghem 1998, Phillips 2005). For that purpose, an effective approach in common use is to adjust the selling price. For example, Zara increased its revenue by \$69.3 million in 2008 after implementing a different pricing model (Carboni Borrásé 2009). Pricing decisions affect demand and thereby alter the demand forecasts used by production systems. In turn, the availability of inventory affects pricing decisions. The interaction between pricing and production is widely recognized. As a result, many firms—ranging from cargo logistics to carpet manufacturing—have created management roles tasked with overseeing joint production and pricing decisions (Cross 2011). Federgreen and Heching (1997) show that optimally integrating price and production can result in 6.5% greater 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, it is hardly surprising that this integration problem has received considerable research attention. With the greater availability of data that is enabled by new technologies, such as RFID tags (Chen et al. 2014), producers can now track not only recent sales and product flows but also consumer behavior; examples include point-of-sale data, online orders, the tracking of pre-purchase online searches, and so forth (Dong et al. 2009).

However, there are a number of obstacles that hinder the exploitation of such data. *First*, as mentioned in Monahan et al. (2004), payoff functions in the joint production and pricing 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 single period and a single product, the uncertainty of demand entails an exhaustive search to find the optimal production and pricing strategies. *Third*, dynamic programming is a powerful theoretical tool for characterizing the optimal policy in simple systems but 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); this is the so-called curse of dimensionality.

Fourth, dynamic programming also requires full knowledge of the entire distribution—yet that can be difficult to obtain accordingly in practice despite the increased availability of historical data. Indeed, even with past observed demand data it is difficult to identify the most appropriate functional form or to estimate the distribution of demand uncertainty (Chen and Shi 2017).

These reasons explain why 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 decision 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).

Motivated by these issues, we aim in this paper to provide a tractable and robust formulation for determining optimal policies for the two-period, multi-item joint pricing and production problem. Towards that end, we combine machine learning and distributionally robust optimization via a two-period, K-means ambiguity set and the technique of affine recourse approximation. We examine in particular, a market supplied by a manager who sells multiple items and who, for each item, determines the production quantity and 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.

Our choice of modeling these decisions as a two-period problem not only improves tractability but also reflects current practice, since producers often divide the selling season into two periods that may feature different prices. Thus, for instance, 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 Borrásé 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, in which retailers face unpredictable levels of demand. 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 ahead of 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.

Our approach and contributions

We adopt a data-driven and distribution-free paradigm for this two-period, multi-item joint pricing and production problem. In contrast to traditional stochastic models, which assume the existence of a known joint distribution for the residual demand, we apply a distributionally robust optimization approach that does not require the distribution to be specified. This approach relies on integrating robust optimization with K-means clustering.

Below we summarize our main results and contributions.

1. We introduce a two-period, K-means ambiguity set based on historical data. This set requires only information on the mean, the support, and the mean absolute deviation of each cluster. This approach is capable of incorporating demand dependence across periods. We provide easily estimated expressions for all of the relevant parameters.

2. We propose a K-means adaptive markdown policy to solve the two-period, multi-item joint pricing and production problem. This policy helps to (approximately) reformulate the problem as a tractable, mixed-integer linear programming (MILP) problem via an affine recourse approximation, which provides a lower bound to the problem. We establish the optimality of the affine recourse approximation when there is only one product, whereas for multiple products, we numerically show that the approximation is near optimal.

3. We introduce a generalized additive demand model that also can be derived from historical data. This 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.

4. We demonstrate that, in all our computational experiments, the affine recourse approximation is nearly optimal. We document the performance of our proposed framework by applying it to a real-world data set. The result is an improvement of 1.11% in profit on average and by as much as 2.22% (i.e., in comparison with an empirical optimization 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. In reviewing the various streams of research related to our paper, we focus on the literature that takes a data-driven approach to addressing the multi-item joint pricing and production problem. More specifically, we review related work in joint inventory and pricing that incorporates stochastic models, robust inventory and pricing, and data-driven robust optimization. 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.

Stochastic models on joint inventory and pricing. One of the seminal works in joint inventory and pricing under uncertainty is [Federgruen and Heching \(1997\)](#). 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 useful 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 Song \(2018\)](#). However, these three works differ from our paper either by proposing heuristics ([Bertsimas and De Boer 2005](#)) or by seeking structurally optimal policies ([Song and Song 2018](#)).

Robust models on joint inventory and pricing. 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. 2017a](#))—especially in the

multi-item case. Hence there is an emergent trend of using robust optimization to address inventory and pricing problems. 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 (distributionally) robust optimization (e.g., Ben-Tal et al. 2004, Bertsimas and Sim 2004, Delage and Ye 2010), an extensive literature has used similar techniques to address inventory and pricing problems. For related inventory control problems, see for example, Bertsimas and Thiele (2006), Perakis and Roels (2008), Hanasusanto et al. (2015a), Xin and Goldberg (2015), Mamani et al. (2016); for related pricing problems, see Lim et al. (2008), Chen and Farias (2015), and Bertsimas et al. (2018a).

These works consider either the inventory control problem or the pricing problem by adopting robust optimization techniques. 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). Unlike nearly all of the papers mentioned, which employ robust optimization to consider either the pricing problem or the inventory control problem, we consider the multi-item joint pricing and production problem in a two-period setting. For this purpose, we adopt the distributionally robust optimization technique and K-means ambiguity set introduced in Chen et al. (2017b) and then adopt an affine recourse approximation to reformulate the problem as an MILP problem.

Data-driven robust optimization. Our paper is also related to the data-driven robust optimization literature, in which the critical step is to construct the uncertainty set or ambiguity set. Examples are given by Ben-Tal and Nemirovski (2000), Delage and Ye (2010), Klabjan et al. (2013), Bertsimas et al. (2018a), and Jiang and Guan (2018). Our paper differs in that we (i) extend the scenario-wise ambiguity set in Chen et al. (2017b) to the two-period setting and (ii) construct the ambiguity set by using K-means clustering from historical data (as in, e.g., MacQueen 1967, Jain and Dubes 1988).

The rest of this paper is organized as follows. After concluding this section with a summary of our notation, in Section 2 we discuss the procedures we introduce for deriving a demand model and constructing an ambiguity set from the data. Then, in Section 3, we recast the multi-item joint pricing and production problem as an MILP problem. Extensive computational studies are presented in Section 4, and the paper concludes in Section 5 with some suggestions for future research.

Notation. We denote by $[I] \triangleq \{1, 2, \dots, I\}$ the set of positive indices up to I . We use boldface glyphs, such as $\mathbf{x} \in \mathbb{R}^I$ and $\mathbf{A} \in \mathbb{R}^{M \times N}$ to denote vectors and matrices; we denote by x_i the i th element of vector \mathbf{x} and by A_{ij} the element located in the i th row and j th column of matrix \mathbf{A} . As usual, $|\mathbf{x}| = (|x_1|, \dots, |x_I|)$ signifies the absolute value of \mathbf{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 $\tilde{\varepsilon} \sim \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^I)$ to denote an I -dimensional random variable $\tilde{\varepsilon}$ 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 $\mathcal{S} \subseteq \mathbb{R}^I$, the term $\mathbb{P}[\tilde{\varepsilon} \in \mathcal{S}]$ represents the probability of $\tilde{\varepsilon}$ lying in the set \mathcal{S} 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 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^I$, the expression $\mathbf{x} \geq \mathbf{y}$ means that \mathbf{x} is *component-wise* no less than \mathbf{y} . The dot product $\mathbf{x} \cdot \mathbf{y} = \sum_{i \in [I]} x_i y_i$. Finally, superscript 1 and 2 always stand for (respectively) period 1 and period 2.

2. Data, demand model, and ambiguity set

In this section we introduce the procedure used to estimate the demand model from data and then characterize its residuals by constructing the ambiguity set via K-means clustering. We assume that the producer has data consisting of H sample paths. Each sample path $h \in [H]$ comprises of two periods of retail prices, \mathbf{r}_h^t , and realized demands, \mathbf{z}_h^t , for $t \in \{1, 2\}$.

From data to demand model

In the field of pricing and production management, the literature has introduced various demand models (see e.g. Federgruen and Heching 1997, Gong et al. 2014). To simplify the exposition, we consider an additive demand model (Lu et al. 2014, Bernstein et al. 2015) as follows:

$$\mathbf{z}^1(\mathbf{r}^1, \tilde{\varepsilon}^1) \triangleq \mathbf{a}^1 + \mathbf{B}^1 \mathbf{r}^1 + \tilde{\varepsilon}^1, \quad (1)$$

$$\mathbf{z}^2(\mathbf{r}^1, \mathbf{r}^2, \tilde{\varepsilon}^2) \triangleq \mathbf{a}^2 + \mathbf{B}^2 \mathbf{r}^2 + \mathbf{B}^{21} \mathbf{r}^1 + \tilde{\varepsilon}^2. \quad (2)$$

Here $\tilde{\varepsilon}^t$ is the residual in period t , and the parameters \mathbf{a}^1 , \mathbf{B}^1 , \mathbf{a}^2 , \mathbf{B}^2 , and \mathbf{B}^{21} can be estimated by using multivariate linear regression or maximum likelihood estimation techniques (see e.g. Genesove and Mullin 1998, Baardman et al. 2018). This 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 demand for different products are often correlated because of substitutability and/or complementarity among products. 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,

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

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, Cohen-Hillel et al. 2018). In Session 4, 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.

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

$$z_i^2(\mathbf{r}^1, \mathbf{r}^2, \tilde{\boldsymbol{\varepsilon}}_i) = a_i^2 + B_{ii}^2 r_i^2 + \sum_{n \in [I]} B_{ni}^{21} r_n^1 + \tilde{\varepsilon}_i,$$

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). We remark that most multi-product demand models in the literature do not explicitly characterize cross-product effects jointly with cross-period effects.

Our demand model incorporates two widely used business rules; 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 MILP problem.

Discrete price ladder It is common practice in business to choose each product's price from a set of "admissible" prices (see e.g. Gallego and Van Ryzin 1994, Rusmevichientong et al. 2006, Carboni Borrásé 2009, Cohen et al. 2017). Thus, for each product i in period t , the price r_i^t must be chosen from the set $\mathcal{R}_i = \{p_{i1}, \dots, p_{iN}\}$; here $p_{i1} < \dots < p_{iN}$ 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 price (r_i^1). Hence we may write $r_i^1 \geq r_i^2$ for $i \in [I]$.

Two-period K-means ambiguity set

Having set up the demand model, we can use the data to determine—for each sample path $h \in [H]$ —the corresponding residuals as follows:

$$\boldsymbol{\varepsilon}_h^1 = \mathbf{z}_h^1 - \mathbf{a}^1 - \mathbf{B}^1 \mathbf{r}_h^1, \quad \boldsymbol{\varepsilon}_h^2 = \mathbf{z}_h^2 - \mathbf{a}^2 - \mathbf{B}^2 \mathbf{r}_h^2 - \mathbf{B}^{21} \mathbf{r}_h^1.$$

Then, for each realized demand, there is a one-to-one mapping to its residual. Instead of assuming a probability distribution over the demand model's residuals, $\tilde{\boldsymbol{\varepsilon}}^t$, we adopt a machine learning approach to characterize an *ambiguity set* of probability distributions that matches those residuals. Thus we use K-means clustering to construct our proposed two-period, K-means ambiguity set.

To implement the K-means clustering, we first partition the support set of $\{\varepsilon_1^1, \dots, \varepsilon_H^1\}$ into K^1 non-overlapping, full-dimensional clusters. For each first-period cluster j , the corresponding second-period residuals are classified into K^2 clusters. An illustration for a two-product, two-period problem with $K^1 = 3$ and $K^2 = 4$ is illustrated in Figure 1, which plots the K-means clusters of the demand residuals.

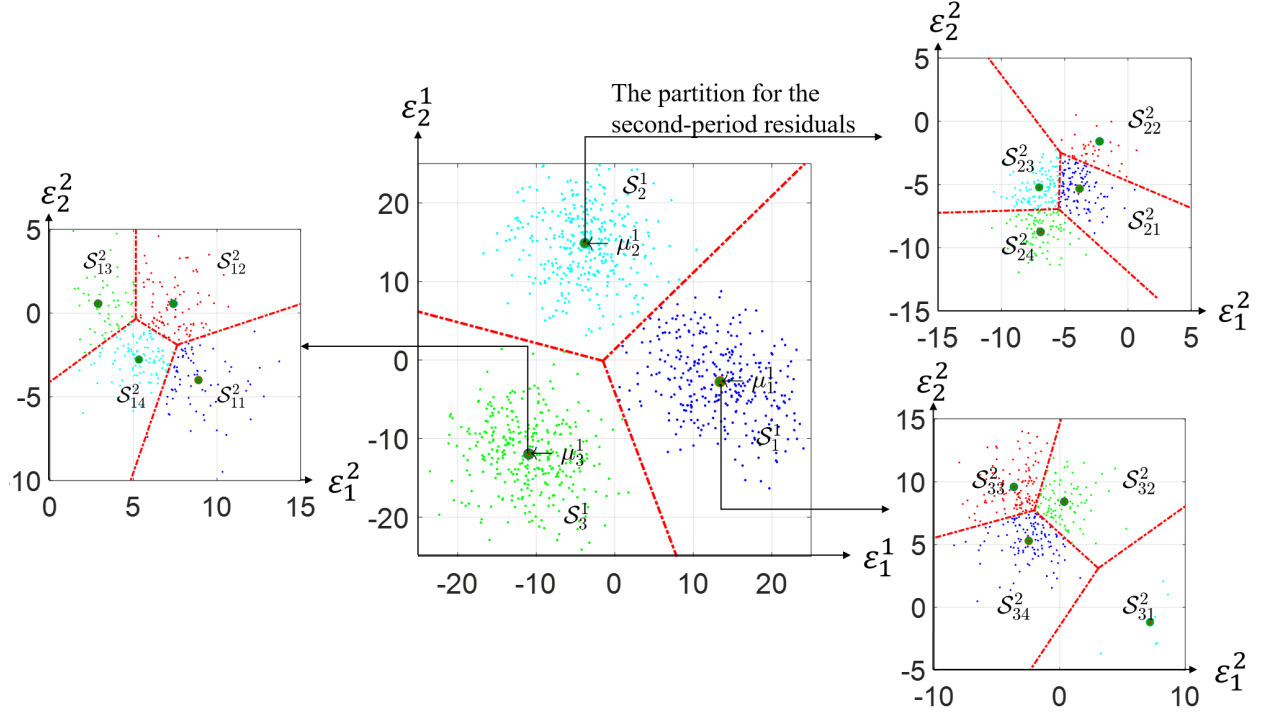


Figure 1 Two-period, K-means clustering (dark circles mark the cluster centroids)

The K-means clustering algorithm outputs the centroids for both periods, which we denote as μ_j^1 and μ_{jk}^2 for $j \in [K^1]$ and $k \in [K^2]$. The support sets associated with these clusters are polyhedrons that are described formally as

$$\mathcal{S}_j^1 = \{\varepsilon^1 \in [\underline{\varepsilon}^1, \bar{\varepsilon}^1] \mid 2(\varepsilon^1)'(\mu_l^1 - \mu_j^1) \leq (\mu_l^1)' \mu_l^1 - (\mu_j^1)' \mu_j^1 \forall l \in [K^1]\} \quad (3)$$

and

$$\mathcal{S}_{jk}^2 = \{\varepsilon^2 \in [\underline{\varepsilon}_j^2, \bar{\varepsilon}_j^2] \mid 2(\varepsilon^2)'(\mu_{jl}^2 - \mu_{jk}^2) \leq (\mu_{jl}^2)' \mu_{jl}^2 - (\mu_{jk}^2)' \mu_{jk}^2 \forall l \in [K^2]\},$$

where $\underline{\varepsilon}^1$ and $\bar{\varepsilon}^1$ are (respectively) the upper and lower bounds of ε^1 such that $\mathbf{z}^1(\mathbf{r}^1, \underline{\varepsilon}^1) \geq 0$ for all $\mathbf{r}^1 \in \mathcal{R}$. Given that the first-period residual is represented by the j th cluster, $\underline{\varepsilon}_j^2$ and $\bar{\varepsilon}_j^2$ are similarly the upper and lower bounds of ε^2 such that $\mathbf{z}^2(\mathbf{r}^1, \mathbf{r}^2, \underline{\varepsilon}_j^2) \geq 0$ for all admissible $\mathbf{r}^1, \mathbf{r}^2 \in \mathcal{R}$.

We can now determine the weight, $q_{jk} = \mathbb{P}[\tilde{\varepsilon}^1 \in \mathcal{S}_j^1, \tilde{\varepsilon}^2 \in \mathcal{S}_{jk}^2]$ as

$$q_{jk} = \frac{1}{H} \sum_{h \in [H]} \mathbb{1}(\varepsilon_h^1 \in \mathcal{S}_j^1, \varepsilon_h^2 \in \mathcal{S}_{jk}^2);$$

where $\mathbb{1}(\cdot)$ is the indicator function. We improve this characterization by also determining the mean absolute deviation associated with each cluster:

$$\sigma_j^1 = \frac{1}{q_j H} \sum_{h \in [H]} \mathbb{1}(\varepsilon_h^1 \in \mathcal{S}_j^1) |\varepsilon_h^1 - \mu_j^1|,$$

where $q_j = \sum_{k \in [K^2]} q_{jk}$; and

$$\sigma_{jk}^2 = \frac{1}{q_{jk} H} \sum_{h \in [H]} \mathbb{1}(\varepsilon_h^1 \in \mathcal{S}_j^1, \varepsilon_h^2 \in \mathcal{S}_{jk}^2) |\varepsilon_h^2 - \mu_{jk}^2|.$$

This allows us to obtain a *two-period, K-means ambiguity set* as follows:

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^I \times \mathbb{R}^I) \left\{ \begin{array}{ll} (\tilde{\varepsilon}^1, \tilde{\varepsilon}^2) \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}[\tilde{\varepsilon}^1 \mid \tilde{\varepsilon}^1 \in \mathcal{S}_j^1] = \mu_j^1 & \forall j \in [K^1] \\ \mathbb{E}_{\mathbb{P}}[\tilde{\varepsilon}^2 \mid \tilde{\varepsilon}^1 \in \mathcal{S}_j^1, \tilde{\varepsilon}^2 \in \mathcal{S}_{jk}^2] = \mu_{jk}^2 & \forall j \in [K^1], k \in [K^2] \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\varepsilon}^1 - \mu_j^1| \mid \tilde{\varepsilon}^1 \in \mathcal{S}_j^1] \leq \sigma_j^1 & \forall j \in [K^1] \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\varepsilon}^2 - \mu_{jk}^2| \mid \tilde{\varepsilon}^1 \in \mathcal{S}_j^1, \tilde{\varepsilon}^2 \in \mathcal{S}_{jk}^2] \leq \sigma_{jk}^2 & \forall j \in [K^1], k \in [K^2] \\ \mathbb{P}[\tilde{\varepsilon}^1 \in \mathcal{S}_j^1, \tilde{\varepsilon}^2 \in \mathcal{S}_{jk}^2] = q_{jk} & \forall j \in [K^1], k \in [K^2] \end{array} \right. \right\}.$$

This ambiguity set allows us to model a rich variety of structural information about random demand, and in resolutions that vary depending on the number of clusters. The K-means 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 1997](#), [Yang et al. 2014](#)).

3. Adaptive distributionally robust optimization model

We consider the joint 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 $\mathbf{x} \in \mathcal{X} \subseteq [\mathbf{0}, \bar{\mathbf{x}}]$ of goods at the marginal cost \mathbf{c} , where $\bar{\mathbf{x}}$ is the upper bound of \mathbf{x} ; at the same time, this manager also sets a *retail* price $\mathbf{r}^1 \in \mathcal{R}$. After the realization of the first-period demand residual, $\varepsilon^1 \in [\underline{\varepsilon}^1, \bar{\varepsilon}^1]$, the manager decides on a *markdown* price, $\mathbf{r}^2 \leq \mathbf{r}^1$. We assume that no goods are produced during the second period and that any product left over after second-period demand is satisfied, $\mathbf{z}^2(\mathbf{r}^1, \mathbf{r}^2, \varepsilon^2)$, is 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. The total revenue is then expressed as

$$\pi(\mathbf{x}, \mathbf{r}^1, \mathbf{r}^2, \varepsilon^1, \varepsilon^2) = \sum_{i \in [I]} (r_i^1 [\min\{x_i, z_i^1(\mathbf{r}^1, \varepsilon_i^1)\}] + r_i^2 [\min\{(x_i - z_i^1(\mathbf{r}^1, \varepsilon_i^1))^+, z_i^2(\mathbf{r}^1, \mathbf{r}^2, \varepsilon_i^2)\}]),$$

where the first (resp. second) term within large parentheses represents the first-period (resp. second-period) revenue. 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).

PROPOSITION 1. *Under a markdown policy, the total revenue can be written as*

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

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

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

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 due to product $i \in [I]$ is a concave and piecewise linear function on x_i . ■

K-means adaptive markdown policy

It is in general difficult to optimize over a recourse function with discrete outcomes. One example of such a function is the markdown policy, $\hat{\mathbf{r}}^2 : [\underline{\boldsymbol{\varepsilon}}^1, \bar{\boldsymbol{\varepsilon}}^1] \mapsto \mathcal{R}$. In contrast with linear decision rules, which are fairly tractable and therefore ubiquitous in distributionally robust optimization, using a K-means ambiguity set enables us to adopt a markdown policy \mathcal{M} to the finite residual clusters associated with the first-period demand:

$$\mathcal{M}(\mathbf{r}^1) \triangleq \left\{ \hat{\mathbf{r}}^2 : \mathcal{S}^1 \mapsto \mathcal{R} \left| \begin{array}{l} \hat{\mathbf{r}}^2(\boldsymbol{\varepsilon}^1) = \mathbf{r}_j^2 \text{ if } \boldsymbol{\varepsilon}^1 \in \mathcal{S}_j^1 \quad j \in [K^1] \\ \text{for some } \mathbf{r}_j^2 \in \mathcal{R}, \mathbf{r}_j^2 \leq \mathbf{r}^1 \quad \forall j \in [K^1] \end{array} \right. \right\}.$$

REMARK 1. The K-means adaptive markdown policy described in this section is related to the K-adaptive policy presented by Bertsimas and Caramanis (2010), Hanasusanto et al. (2015b, 2016), and Subramanyam et al. (2017). Although both policies allow for discrete recourse, the clusters in K-adaptive models are determined via optimization whereas the clusters in K-means adaptive models are predetermined by clustering on data. Hence the former models are more computationally demanding to solve than the latter.

Under the K-means adaptive markdown policy, the manager solves the following distributionally robust optimization problem:

$$Z^* = \max_{\mathbf{x} \in \mathcal{X}, \hat{\mathbf{r}}^2 \in \mathcal{M}(\mathbf{r}^1), \mathbf{r}^1 \in \mathcal{R}} -\mathbf{c}'\mathbf{x} + \inf_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[\pi(\mathbf{x}, \mathbf{r}^1, \hat{\mathbf{r}}^2(\bar{\boldsymbol{\varepsilon}}^1), \bar{\boldsymbol{\varepsilon}}^1, \bar{\boldsymbol{\varepsilon}}^2)]. \quad (4)$$

In Problem (4), 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} . Note that even for a multi-product newsvendor problem, the evaluation of worst-case expectation would generally be intractable (see, for example, Hanasusanto et al. 2015a). Before we derive tractable models, in what follows, we first draw an equivalent nonlinear robust optimization model.

PROPOSITION 2. *With respect to the K -means adaptive markdown policy, we can formulate Problem (4) as the following nonlinear robust optimization model:*

$$\begin{aligned}
\max \quad & -\mathbf{c}'\mathbf{x} + \sum_{j \in [K^1], k \in [K^2]} (\alpha_{jk} + (\beta_j^1)'\boldsymbol{\mu}_j^1 + (\beta_{jk}^2)'\boldsymbol{\mu}_{jk}^2 + (\gamma_j^1)'\boldsymbol{\sigma}_j^1 + (\gamma_{jk}^2)'\boldsymbol{\sigma}_{jk}^2) \\
\text{s.t.} \quad & \alpha_{jk} + (\beta_j^1)'\boldsymbol{\varepsilon}^1 + (\beta_{jk}^2)'\boldsymbol{\varepsilon}^2 + (\gamma_j^1)'\mathbf{u}^1 + (\gamma_{jk}^2)'\mathbf{u}^2 \\
& \leq q_{jk}\pi(\mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2) & \forall (\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \in \mathcal{W}_{jk}, \\
& & j \in [K^1], k \in [K^2]; \\
& \mathbf{r}^1 \geq \mathbf{r}_j^2 & \forall j \in [K^1]; \\
& \gamma_j^1, \gamma_{jk}^2 \leq 0 & \forall j \in [K^1], k \in [K^2]; \\
& \mathbf{r}^1, \mathbf{r}_j^2 \in \mathcal{R} & \forall j \in [K^1]; \\
& \alpha_{jk} \in \mathbb{R}, \beta_j^1, \beta_{jk}^2, \gamma_j^1, \gamma_{jk}^2 \in \mathbb{R}^I & \forall j \in [K^1], k \in [K^2]; \\
& \mathbf{x} \in \mathcal{X}.
\end{aligned} \tag{5}$$

Note that \mathcal{W}_{jk} is the lifted support set defined as

$$\mathcal{W}_{jk} = \left\{ (\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \in \mathbb{R}^I \times \mathbb{R}^I \times \mathbb{R}^I \times \mathbb{R}^I \mid \begin{array}{l} \boldsymbol{\varepsilon}^1 \in \mathcal{S}_j^1, \boldsymbol{\varepsilon}^2 \in \mathcal{S}_{jk}^2 \\ \mathbf{u}^1 \geq |\boldsymbol{\varepsilon}^1 - \boldsymbol{\mu}_j^1|, \mathbf{u}^2 \geq |\boldsymbol{\varepsilon}^2 - \boldsymbol{\mu}_{jk}^2| \end{array} \right\}.$$

Proof. See Appendix A.1. ■

The nonlinearity of $\pi(\mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2)$ in Model (5) prevents us from obtaining the robust counterpart. Nevertheless, the following theorem gives an equivalent linear reformulation, which we refer to as the *exact model*.

THEOREM 1. *Problem (4) is equivalent the following robust optimization model:*

$$\begin{aligned}
\max \quad & -\mathbf{c}'\mathbf{x} + \sum_{j \in [K^1], k \in [K^2]} (\alpha_{jk} + (\beta_j^1)'\boldsymbol{\mu}_j^1 + (\beta_{jk}^2)'\boldsymbol{\mu}_{jk}^2 + (\gamma_j^1)'\boldsymbol{\sigma}_j^1 + (\gamma_{jk}^2)'\boldsymbol{\sigma}_{jk}^2) \\
\text{s.t.} \quad & \alpha_{jk} + (\beta_j^1)'\boldsymbol{\varepsilon}^1 + (\beta_{jk}^2)'\boldsymbol{\varepsilon}^2 + (\gamma_j^1)'\mathbf{u}^1 + (\gamma_{jk}^2)'\mathbf{u}^2 \\
& \leq q_{jk}\phi(\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2) & \forall (\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \in \mathcal{W}_{jk}, \\
& & \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3 \in \mathcal{B}, \\
& & j \in [K^1], k \in [K^2]; \\
& \mathbf{r}^1 \geq \mathbf{r}_j^2 & \forall j \in [K^1]; \\
& \gamma_j^1, \gamma_{jk}^2 \leq 0 & \forall j \in [K^1], k \in [K^2]; \\
& \mathbf{r}^1, \mathbf{r}_j^2 \in \mathcal{R} & \forall j \in [K^1]; \\
& \alpha_{jk} \in \mathbb{R}, \beta_j^1, \beta_{jk}^2, \gamma_j^1, \gamma_{jk}^2 \in \mathbb{R}^I & \forall j \in [K^1], k \in [K^2]; \\
& \mathbf{x} \in \mathcal{X}.
\end{aligned} \tag{6}$$

Note that

$$\phi(\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2)$$

$$\begin{aligned} &\triangleq \sum_{i \in \mathcal{I}_1} r_i^1 x_i + \sum_{i \in \mathcal{I}_2} (r_{ji}^2 x_i + (r_i^1 - r_{ji}^2) (a_i^1 + \mathbf{e}'_i \mathbf{B}^1 \mathbf{r}^1 + \varepsilon_i^1)) \\ &+ \sum_{i \in \mathcal{I}_3} (r_i^1 (a_i^1 + \mathbf{e}'_i \mathbf{B}^1 \mathbf{r}^1 + \varepsilon_i^1) + r_{ji}^2 (a_i^2 + \mathbf{e}'_i \mathbf{B}^2 \mathbf{r}_j^2 + \mathbf{e}'_i \mathbf{B}^{21} \mathbf{r}^1 + \varepsilon_i^2)), \end{aligned}$$

and

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

Proof. See 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, the size of Problem (6) grows exponentially with the number of items, as a result, this is not a computationally scalable approach.

Affine recourse approximation

Using the lifted affine recourse approximation proposed by Bertsimas et al. (2018b), and used also by Chen et al. (2017b), we consider the following family of functions:

$$\mathcal{L} \triangleq \left\{ v: \mathbb{R}^I \times \mathbb{R}^I \times \mathbb{R}^I \times \mathbb{R}^I \mapsto \mathbb{R} \mid \begin{array}{l} \exists v^0 \in \mathbb{R}, \mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3, \mathbf{v}^4 \in \mathbb{R}^I : \\ v(\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) = v^0 + (\mathbf{v}^1)' \boldsymbol{\varepsilon}^1 + (\mathbf{v}^2)' \boldsymbol{\varepsilon}^2 + (\mathbf{v}^3)' \mathbf{u}^1 + (\mathbf{v}^4)' \mathbf{u}^2 \end{array} \right\}.$$

We can then obtain a lower bound for Problem (4), 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 approximation, where the optimal objective value is denoted Z_R :

$$\begin{aligned}
& \max -\mathbf{c}'\mathbf{x} + \sum_{j \in [K^1], k \in [K^2]} (\alpha_{jk} + (\beta_j^1)' \boldsymbol{\mu}_j^1 + (\beta_{jk}^2)' \boldsymbol{\mu}_{jk}^2 + (\gamma_j^1)' \boldsymbol{\sigma}_j^1 + (\gamma_{jk}^2)' \boldsymbol{\sigma}_{jk}^2) \\
& \left. \begin{aligned}
& \alpha_{jk} + (\beta_j^1)' \boldsymbol{\varepsilon}^1 + (\beta_{jk}^2)' \boldsymbol{\varepsilon}^2 + (\gamma_j^1)' \mathbf{u}^1 + (\gamma_{jk}^2)' \mathbf{u}^2 \\
& \leq q_{jk} \sum_{i \in [I]} v_{ijk}(\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2), \\
s.t. \quad & v_{ijk}(\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \leq w_i^1, \\
& v_{ijk}(\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \leq w_i^2 + (\mathbf{p}'_i \mathbf{y}_i^1 - \mathbf{p}'_i \mathbf{y}_{ji}^2)(a_i^1 + \varepsilon_i^1) + \rho_i^{11} + \rho_{ji}^{12}, \\
& v_{ijk}(\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \\
& \leq \mathbf{p}'_i \mathbf{y}_i^1 (a_i^1 + \varepsilon_i^1) + \rho_i^{11} + \mathbf{p}'_i \mathbf{y}_{ji}^2 (a_i^2 + \varepsilon_i^2) + \rho_{ji}^{22} + \rho_{ji}^{21} \\
& w_i^1 \leq p_{in} x_i + M(1 - y_{in}^1), \\
& \rho_i^{11} \leq p_{in} \mathbf{e}'_i \mathbf{B}^1(\mathbf{p}'_1 \mathbf{y}_1^1, \dots, \mathbf{p}'_I \mathbf{y}_I^1)' + M(1 - y_{in}^1) \\
& w_{ji}^2 \leq p_{in} x_i + M(1 - y_{jin}^2), \\
& \rho_{ji}^{12} \leq -p_{in} \mathbf{e}'_i \mathbf{B}^1(\mathbf{p}'_1 \mathbf{y}_1^1, \dots, \mathbf{p}'_I \mathbf{y}_I^1)' + M(1 - y_{jin}^2), \\
& \rho_{ji}^{21} \leq p_{in} \mathbf{e}'_i \mathbf{B}^{21}(\mathbf{p}'_1 \mathbf{y}_1^1, \dots, \mathbf{p}'_I \mathbf{y}_I^1)' + M(1 - y_{jin}^2), \\
& \rho_{ji}^{22} \leq p_{in} \mathbf{e}'_i \mathbf{B}^2(\mathbf{p}'_1 \mathbf{y}_{j1}^2, \dots, \mathbf{p}'_I \mathbf{y}_{jI}^2)' + M(1 - y_{jin}^2) \\
& \mathbf{p}'_i \mathbf{y}_i^1 \geq \mathbf{p}'_i \mathbf{y}_{ji}^2, \mathbf{1}' \mathbf{y}_i^1 = 1, \mathbf{1}' \mathbf{y}_{ji}^2 = 1, \mathbf{y}_i^1, \mathbf{y}_{ji}^2 \in \{0, 1\}^N \\
& v_{ijk} \in \mathcal{L}, \gamma_j^1, \gamma_{jk}^2 \leq 0 \\
& \alpha_{jk} \in \mathbb{R}, \beta_j^1, \beta_{jk}^2, \gamma_j^1, \gamma_{jk}^2, \mathbf{w}^1, \mathbf{w}_j^2, \rho^{11}, \rho_j^{12}, \rho_j^{21}, \rho_j^{22} \in \mathbb{R}^I \\
& \mathbf{x} \in \mathcal{X}.
\end{aligned} \right\} \begin{aligned}
& \forall (\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \in \mathcal{W}_{jk}, \\
& i \in [I], j \in [K^1], k \in [K^2]; \\
& \forall i \in [I], n \in [N]; \\
& \forall i \in [I], j \in [K^1], n \in [N]; \\
& \forall i \in [I], j \in [K^1]; \\
& \forall i \in [I], j \in [K^1], k \in [K^2]; \\
& \forall j \in [K^1], k \in [K^2];
\end{aligned}
\end{aligned} \tag{7}$$

Note that when M is a sufficiently large positive number, then $Z_R \leq Z^*$. Equality holds when $I = 1$.

Proof. See Appendix A.3. ■

Recall that this approach differs from extant research in that the affine recourse approximation adapts itself to each cluster of the 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 4) 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 3. *The feasibility of $(\mathbf{f}^1, \mathbf{f}^2, \mathbf{h}^1, \mathbf{h}^2, g)$ in*

$$(\mathbf{f}^1)' \boldsymbol{\varepsilon}^1 + (\mathbf{h}^1)' \mathbf{u}^1 + (\mathbf{f}^2)' \boldsymbol{\varepsilon}^2 + (\mathbf{h}^2)' \mathbf{u}^2 \leq g \quad \forall (\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \in \mathcal{W}$$

is equivalent to its feasibility under the linear constraints

$$\begin{aligned}
& \sum_{\ell \in [K^1]} \kappa_{j\ell}^1 ((\boldsymbol{\mu}_\ell^1)' \boldsymbol{\mu}_\ell^1 - (\boldsymbol{\mu}_j^1)' \boldsymbol{\mu}_j^1) + \sum_{\ell \in [K^2]} \kappa_{jk\ell}^2 ((\boldsymbol{\mu}_{j\ell}^2)' \boldsymbol{\mu}_{j\ell}^2 - (\boldsymbol{\mu}_{jk}^2)' \boldsymbol{\mu}_{jk}^2) \\
& \quad - (\boldsymbol{\tau}_j^{11})' \boldsymbol{\mu}_j^1 + (\boldsymbol{\tau}_j^{12})' \boldsymbol{\mu}_j^1 - (\boldsymbol{\tau}_{jk}^{21})' \boldsymbol{\mu}_{jk}^2 \\
& \quad + (\boldsymbol{\tau}_{jk}^{22})' \boldsymbol{\mu}_{jk}^2 + (\boldsymbol{\omega}^{11})' \underline{\boldsymbol{\epsilon}}^1 + (\boldsymbol{\omega}^{12})' \bar{\boldsymbol{\epsilon}}^1 + (\boldsymbol{\omega}^{21})' \underline{\boldsymbol{\epsilon}}_j^2 + (\boldsymbol{\omega}^{22})' \bar{\boldsymbol{\epsilon}}_j^2 \leq g; \\
& 2\kappa_{jl}^1 (\boldsymbol{\mu}_l^1 - \boldsymbol{\mu}_j^1) - \boldsymbol{\tau}_j^{11} + \boldsymbol{\tau}_j^{12} + \boldsymbol{\omega}^{11} + \boldsymbol{\omega}^{12} = \mathbf{f}^1 \quad \forall l \in [K^1]; \\
& 2\kappa_{jkm}^2 (\boldsymbol{\mu}_{jk}^2 - \boldsymbol{\mu}_{jk}^2) - \boldsymbol{\tau}_{jk}^{21} + \boldsymbol{\tau}_{jk}^{22} + \boldsymbol{\omega}_j^{21} + \boldsymbol{\omega}_j^{22} = \mathbf{f}^2 \quad \forall k \in [K^2]; \\
& \boldsymbol{\tau}_j^{11} + \boldsymbol{\tau}_j^{12} = \mathbf{h}^1; \\
& \boldsymbol{\tau}_{jk}^{21} + \boldsymbol{\tau}_{jk}^{22} = \mathbf{h}^2; \\
& \kappa_j^1, \kappa_{jk}^2, \boldsymbol{\omega}^{12}, \boldsymbol{\omega}_j^{22} \geq 0; \\
& \boldsymbol{\tau}_j^{11}, \boldsymbol{\tau}_j^{12}, \boldsymbol{\tau}_{jk}^{21}, \boldsymbol{\omega}^{11}, \boldsymbol{\omega}_j^{21}, \boldsymbol{\omega}_j^{22} \leq 0,
\end{aligned} \tag{8}$$

for some $\kappa_j^1, \kappa_{jk}^2, \boldsymbol{\tau}_j^{11}, \boldsymbol{\tau}_j^{12}, \boldsymbol{\tau}_{jk}^{21}, \boldsymbol{\omega}^{11}, \boldsymbol{\omega}^{12}, \boldsymbol{\omega}_j^{21}, \boldsymbol{\omega}_j^{22} \in \mathbb{R}^I$.

Proof. See Appendix A.4. ■

So now, since the support sets \mathcal{S}_j^1 , and \mathcal{S}_{jk}^2 ($j \in [K^1]$, $k \in [K^2]$) are polyhedral sets, the robust optimization problem can be recast as a mixed-integer linear programming.

4. Numerical experiments

We conduct numerical experiments for the case in which the first-period price is known in order to demonstrate (a) the effect of a K-means adaptive markdown policy and (b) the computational efficiency of our affine recourse approximation (ARA). To complement this simulation study, we apply our approach using data from an international cosmetics company. We use the operational data of that company to illustrate the solution quality of our cluster-based framework as compared to an empirical model (EM) that optimizes over the distribution constructed through H sample

paths. For the latter model:

$$\begin{aligned}
Z_{\text{EM}} = \max \quad & -\mathbf{c}'\mathbf{x} + \frac{1}{H} \sum_{h \in [H]} \sum_{i \in [I]} v_{ih} \\
\text{s.t.} \quad & v_{ih} \leq w_i^1 \quad \forall h \in [H]; \\
& v_{ih} \leq w_i^2 + (\mathbf{p}'_i \mathbf{y}_i^1 - \mathbf{p}'_i \mathbf{y}_i^2)(a_i^1 + \varepsilon_{ih}^1) + \rho_i^{11} + \rho_i^{12} \quad \forall h \in [H]; \\
& v_{ih} \leq \mathbf{p}'_i \mathbf{y}_i^1 (a_i^1 + \varepsilon_{ih}^1) + \rho_i^{11} + \mathbf{p}'_i \mathbf{y}_i^2 (a_i^2 + \varepsilon_{ih}^2) + \rho_i^{22} + \rho_i^{21} \quad \forall h \in [H]; \\
& w_i^1 \leq p_{in} x_i + M(1 - y_{in}^1) \quad \forall i \in [I], n \in [N]; \\
& w_i^2 \leq p_{in} x_i + M(1 - y_{in}^2) \quad \forall i \in [I], n \in [N]; \\
& \rho_i^{11} \leq p_{in} \mathbf{e}'_i \mathbf{B}^1 (\mathbf{p}'_1 \mathbf{y}_1^1, \dots, \mathbf{p}'_I \mathbf{y}_I^1)' + M(1 - y_{in}^1) \quad \forall i \in [I], n \in [N]; \\
& \rho_i^{12} \leq -p_{in} \mathbf{e}'_i \mathbf{B}^1 (\mathbf{p}'_1 \mathbf{y}_1^1, \dots, \mathbf{p}'_I \mathbf{y}_I^1)' + M(1 - y_{in}^2) \quad \forall i \in [I], n \in [N]; \\
& \rho_i^{21} \leq p_{in} \mathbf{e}'_i \mathbf{B}^{21} (\mathbf{p}'_1 \mathbf{y}_1^1, \dots, \mathbf{p}'_I \mathbf{y}_I^1)' + M(1 - y_{in}^2) \quad \forall i \in [I], n \in [N]; \\
& \rho_i^{22} \leq p_{in} \mathbf{e}'_i \mathbf{B}^2 (\mathbf{p}'_1 \mathbf{y}_1^2, \dots, \mathbf{p}'_I \mathbf{y}_I^2)' + M(1 - y_{in}^2) \quad \forall i \in [I], n \in [N]; \\
& \mathbf{p}'_i \mathbf{y}_i^1 \geq \mathbf{p}'_i \mathbf{y}_i^2 \quad \forall i \in [I]; \\
& \mathbf{1}' \mathbf{y}_i^1 = 1, \mathbf{1}' \mathbf{y}_i^2 = 1 \quad \forall i \in [I]; \\
& \mathbf{y}_i^1, \mathbf{y}_i^2 \in \{0, 1\}^N \quad \forall i \in [I]; \\
& v_{ih} \in \mathbb{R}, \mathbf{w}^1, \mathbf{w}^2, \boldsymbol{\rho}^{11}, \boldsymbol{\rho}^{12}, \boldsymbol{\rho}^{21}, \boldsymbol{\rho}^{22} \in \mathbb{R}^I \quad \forall i \in [I], h \in [H]; \\
& \mathbf{x} \in \mathcal{X}.
\end{aligned}$$

All of our numerical experiments were conducted in AROMA (a MATLAB-based algebraic modeling package), while using the GUROBI solver on a Windows-OS computer with 64 GB of RAM and a 2.1-GHz CPU.

Simulated data

Before embarking on the full application of our model to the case study, we use simulated data and investigate the exact model (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:

$$\max_{\mathbf{x} \in \mathcal{X}, \hat{\mathbf{r}}^2 \in \mathcal{M}(\mathbf{r}^1)} -\mathbf{c}'\mathbf{x} + \inf_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[\pi(\mathbf{x}, \hat{\mathbf{r}}^2(\tilde{\boldsymbol{\varepsilon}}^1), \tilde{\boldsymbol{\varepsilon}}^1, \tilde{\boldsymbol{\varepsilon}}^2)], \quad (9)$$

where \mathbf{r}^1 is the given retail price. We also solve the ARA in the absence of a K-means adaptive markdown policy; thus, we solve Model (9) with $\hat{\mathbf{r}}^2 = \mathbf{r}^2 \in \mathcal{R}$, $\forall j \in [K^1]$. We refer to this model as the non-adaptive affine recourse approximation model (NA-ARA).

We consider instances with $I \in \{1, \dots, 5\}$, $K^1 \in \{1, 2, 3, 4, 5\}$, and $K^2 \in \{1, 2, 3\}$. 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

Table 1 Simulation study: Objective value

	1		2		3		4	
	Z^*	Z_R	Z^*	Z_R	Z^*	Z_R	Z^*	Z_R
1	2053.135	0.000	2085.265	0.000	2106.571	0.000	2072.761	0.000
2	2093.398	0.000	2116.613	0.000	2117.954	0.000	2146.819	0.000
3	2134.066	0.000	2134.866	0.000	2135.124	0.000	2152.971	0.000
4	2134.736	0.000	2134.955	0.000	2136.161	0.000	2152.691	0.000
5	2134.185	0.000	2135.688	0.000	2136.444	0.000	2143.372	0.000

Table 2 Simulation study: Computation time (s)

	3			4			5		
	ARA	NA-ARA	EXACT	ARA	NA-ARA	EXACT	ARA	NA-ARA	EXACT
1	0.063	0.047	0.094	0.093	0.063	0.094	0.094	0.078	0.126
2	0.094	0.125	0.328	0.171	0.140	0.406	0.250	0.156	0.625
3	0.225	0.188	3.358	0.293	0.253	15.867	0.778	0.293	41.724
4	0.364	0.501	89.165	0.748	0.672	154.206	1.135	0.866	367.042
5	0.717	0.853	1827.608	1.567	1.083	2399.554	2.188	1.626	4998.541

so as to be dependent on those first-period peaks. We also generate 100,000 sample paths—using the same distribution—to obtain the expected 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 \(2014\)](#); see Appendix C for details.

Table 1 displays the objective value of EXACT and ARA. In this table, the rows correspond to the number of first-period clusters ($1, \dots, 5$) while the columns correspond to the number of second-period clusters ($1 \dots 4$). 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 2 shows the computation time of the ARA, NA-ARA, and EXACT. In this table, the rows correspond to the number of products ($1, \dots, 5$) while the columns correspond to the number of first-period clusters ($3 \dots 5$). We fix the number of second-period clusters to be 4.

It is clear from Table 2 that, as the number of clusters or 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 1 shows that the EXACT and ARA objective values coincide; that is, the affine recourse approximation usually yields the same solution

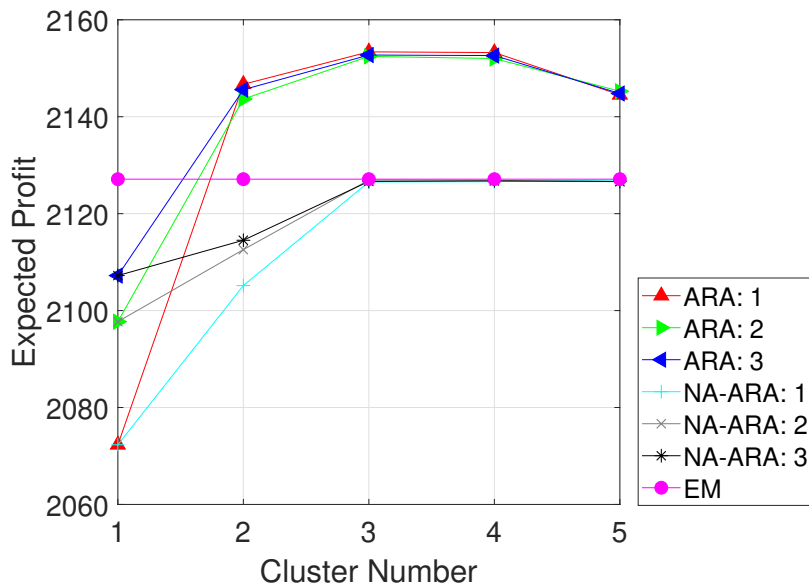


Figure 2 Simulation study: Expected profit

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

The simulation demonstrates that (1) the affine recourse approximation is near optimal but far more tractable; (2) distributionally robust optimization models are not conservative.

Case study

We quantify the value of our cluster-based 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 two performance metrics: the mean and the standard deviation of profit.

Data description The dataset was acquired from an international cosmetics company. The data span nearly a decade (from August 2007 to December 2015) and include sales and inventory

records of 123 products carried by 341 stores in Indonesia. This data set has 18 variables features; examples include store key, 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 clustering, we aggregate the demand for each SKU over all stores and then select ten products for the joint pricing and production problem (see Appendix B 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 demand but rather on the effectiveness of our cluster-based 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:

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

for $i \in [I]$ and $h \in [H]$. Here $\mathbf{t}_i \mathbf{Year}$ is added to remove the time trend. We report the regression results for item 1 in Table 3, and we summarize the linear regression results for ten selected items in Appendix B.

Table 3 clearly shows that the demand for product 1 declines as its price increases. With respect to this product, products 4, 9, and 10 are substitutes whereas products 2, 5, 7, and 8 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 1 are reported in Table 4. At the SKU level, products 4 and 10 are substitutes for product 1 while products 2 and 3 are complements to product 1.

Experiments on joint pricing and production problem We use sales data on ten selected products and sequentially partition them into five groups; each group contains four products, among which two are from the previous group and the other two are newly added in. For each group, the data are sorted into training data (60% of the data) and testing data (40% of the data). Let each of these data sets be denoted by $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{2H}\}$; then $\{\mathbf{z}_1, \mathbf{z}_3, \dots, \mathbf{z}_{2H-1}\}$ is selected to represent the first-period demand while the other elements represent second-period demand. Applying multivariate linear regression to the training data yields, as output, the demand model's

Table 3 Linear regression results for product 1: Aggregate level

Coefficient	Estimate	S.E.	t value	$\Pr(> t)$
a_1	117.9362	9.566	12.329	<2e-16 ***
B_{11}	-5.3663	0.4604	-11.655	<2e-16 ***
B_{21}	-3.5273	0.5722	-6.164	7.94e-10 ***
B_{31}	-0.138	0.4103	-0.336	0.73665
B_{41}	3.4183	0.7512	4.551	5.54e-06 ***
B_{51}	-1.4075	0.5199	-2.707	0.00682 **
B_{61}	0.1414	0.3767	0.375	0.70735
B_{71}	-1.6579	0.4098	-4.045	5.34e-05 ***
B_{81}	-1.1625	0.4133	-2.813	0.00494 **
B_{91}	2.172	0.5172	4.2	2.74e-05 ***
$B_{10,1}$	1.0401	0.3795	2.741	0.00616 **
$t_{2008,1}$	7.449	3.7404	1.992	0.04650 *
$t_{2009,1}$	10.7901	3.9832	2.709	0.00679 **
$t_{2010,1}$	24.3605	3.8957	6.253	4.54e-10 ***
$t_{2011,1}$	25.4127	3.9935	6.363	2.25e-10 ***
$t_{2012,1}$	38.0537	3.9085	9.736	<2e-16 ***
$t_{2013,1}$	67.1254	4.2879	15.655	<2e-16 ***
$t_{2014,1}$	75.7112	5.9573	12.709	<2e-16 ***
$t_{2015,1}$	102.5329	6.2516	16.401	<2e-16 ***
$t_{2016,1}$	92.434	6.9011	13.394	<2e-16 ***

Signif. codes: '***', 0.001; '**', 0.01; '*', 0.05; '.', 0.1; ' ', 1

Residual standard error, 32.58 on 3,293 degrees of freedom (df)

Multiple R^2 , 0.4246; adjusted R^2 , 0.4213

F -statistic, 127.9 on 19 and 3,293 df; p -value, <2.2e-16

coefficients. Details on deriving these coefficients (and on the other parameters) are provided in Appendix C. 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 approximation model (7) with a K-means adaptive 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 is calculated as $(\text{ARA} - \text{EM})/\text{EM} \times 100$. For the out-of-sample tests, the yielded profit is shown in Table 5. In this table, the rows correspond to the group number $(1, \dots, 5)$ while the columns correspond to different combinations of clusters in the two periods. For instance, $(4, 3)$ represents

Table 4 Linear regression results for product 1: SKU level

Coefficient	Estimate	S.E.	<i>t</i> value	Pr(> <i>t</i>)
a_1	5.2153	0.5935	8.7870	<2e-16 ***
B_{11}	-0.2220	0.0279	-7.9510	2.53e-15 ***
B_{21}	-0.1507	0.0351	-4.289	1.84e-05 ***
B_{31}	-0.0733	0.0261	-2.812	0.00495 **
B_{41}	0.1390	0.0454	3.063	0.00221 **
B_{51}	-0.0096	0.0324	-0.295	0.76785
B_{61}	-0.0171	0.0230	-0.743	0.45733
B_{71}	-0.0224	0.0250	-0.895	0.37071
B_{81}	-0.0282	0.0250	-1.124	0.26104
B_{91}	0.0093	0.0319	0.291	0.77086
$B_{10,1}$	0.0952	0.0232	4.101	4.21e-05 ***
$t_{2008,1}$	0.4774	0.2358	2.025	0.04292 *
$t_{2009,1}$	0.5304	0.2511	2.112	0.03472 *
$t_{2010,1}$	1.4662	0.2454	5.975	2.55e-09 ***
$t_{2011,1}$	0.7755	0.2516	3.082	0.00207 **
$t_{2012,1}$	1.1669	0.2464	4.737	2.27e-06 ***
$t_{2013,1}$	1.6995	0.2694	6.308	3.21e-10 ***
$t_{2014,1}$	2.2205	0.3712	5.982	2.45e-09 ***
$t_{2015,1}$	3.0049	0.3892	7.72	1.54e-14 ***
$t_{2016,1}$	2.9596	0.4273	6.926	5.19e-12 ***

Signif. codes: '***', 0.001; '**', 0.01; '*', 0.05; '.', 0.1; ' ', 1

Residual standard error, 2.06 on 3,293 degrees of freedom (df)

Multiple R^2 , 0.1213; adjusted R^2 , 0.1162

F -statistic, 23.92 on 19 and 3,293 df; p -value, <2.2e-16

Table 5 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	6662.528	0.56%	0.02%	0.50%	-0.05%	-0.06%	-0.34%	0.08%	-0.23%
2	2699.402	1.44%	-0.50%	1.42%	-0.46%	1.91%	-0.30%	1.56%	-0.29%
3	2972.640	1.18%	-0.15%	1.40%	-0.27%	1.33%	-0.04%	1.41%	-0.27%
4	3008.000	0.21%	-0.24%	0.33%	-0.28%	-0.13%	-0.06%	0.23%	-0.03%
5	1875.554	2.22%	0.09%	1.07%	0.13%	1.61%	-0.18%	0.88%	0.43%

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 5. First, the NA-ARA is near optimal similarly to EM; the gap is 0.15% over all instances. It demonstrates that employing the cluster-based, two-period ambiguity set can reduce the conservativeness of distributionally robust models. Second, adopting the K-means adaptive 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.

5. Conclusions

We integrate distributionally robust optimization and machine learning in order to identify a two-period, K-means ambiguity set. We then use an affine recourse approximation—via a K-means adaptive markdown policy—to devise an approximate reformulation of the two-period, multi-item joint pricing and production problem as a more tractable problem that can be addressed using mixed-integer linear programming. We then construct both our demand model and the ambiguity set using a real-world data set provided by an international cosmetics company. Our demand model is able to explain both cross-product and cross-period effects in a multi-item, multi-period system. The affine recourse approximation enables us to obtain a lower bound for the exact model. We establish the theoretical optimality of such approximation for solving the one-product, two-period problem.

To demonstrate the effectiveness and computational efficiency of our cluster-based framework, we first consider a simulation study. Our numerical experiments establish that, with just a small number of clusters, the affine recourse approximation typically achieves almost the same outcomes as does the exact model. We complement the simulation study with experimental studies—performed using a real-world data set—on the joint pricing and production problem. We find that, on the out-of-sample test, ARA performs better than the empirical model in terms of mean profit. Our model, on average, yields a 1.11% (as much as 2.22%.) improvement in profit.

The optimization framework we introduce relies on considering information only about the mean, the support, and the mean absolute deviation. However, this model could easily be extended to incorporate other factors, such as the second-moment information about the demand model’s randomness. Note also that we explored only the basic clustering method and did not seek any improvements in the clustering algorithm’s efficiency. Future research could employ other clustering methods and/or the dynamic clustering of realized demand via a “support vector machine” approach among others.

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.
- 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.
- 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.
- Baardman, Lennart, Maxime C Cohen, Kiran Panchamgam, Georgia Perakis, Danny Segev. 2018. Scheduling promotion vehicles to boost profits. *Management Science* .
- Ben-Tal, Aharon, Alexander Goryashko, Elana Guslitzer, Arkadi Nemirovski. 2004. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming* **99**(2) 351–376.
- Ben-Tal, Aharon, Arkadi Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical programming* **88**(3) 411–424.
- 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.
- 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, Vishal Gupta, Nathan Kallus. 2018a. Data-driven robust optimization. *Mathematical Programming* **167**(2) 235–292.
- Bertsimas, Dimitris, Melvyn Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.
- Bertsimas, Dimitris, Melvyn Sim, Meilin Zhang. 2018b. Adaptive distributionally robust optimization. *Management Science* .
- 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. URL <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.
- Buzek, Greg. 2015. Retail’s \$1.1 trillion inventory distortion problem. URL <https://www.ihlservices.com/wp-content/uploads/2015/06/WeLostAustralia-Outline.pdf>.
- 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 Borrásé, 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. 2017a. Nonparametric algorithms for joint pricing and inventory control with lost-sales and censored demand .
- 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, Zhan Pang, Limeng Pan. 2014. Coordinating inventory control and pricing strategies for perishable products. *Operations Research* **62**(2) 284–300.
- Chen, Xin, David Simchi-Levi. 2012. Pricing and inventory management.
- Chen, Yiwei, Vivek F Farias. 2015. Robust dynamic pricing with strategic customers. .
- Chen, Yiwei, Cong Shi. 2017. Joint pricing and inventory management with strategic customers .
- Chen, Zhi, Melvyn Sim, Peng Xiong. 2017b. Tractable distributionally robust optimization with data. http://www.optimization-online.org/DB_FILE/2017/06/6055.pdf .
- Chu, Leon Yang, Guang Li, Paat Rusmevichientong. 2018. Optimal pricing and inventory planning with charitable donations. *Manufacturing & Service Operations Management* .
- 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* .
- Cohen-Hillel, Tamar, Kiran Panchamgam, Georgia Perakis. 2018. High-low promotion policies for peak end demand models .

- Cross, Robert G. 2011. *Revenue management: Hard-core tactics for market domination*. Crown Business.
- 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.
- Dong, Lingxiu, Panos Kouvelis, Zhongjun Tian. 2009. Dynamic pricing and inventory control of substitute products. *Manufacturing & Service Operations Management* **11**(2) 317–339.
- Elmaghraby, Wedad, Pinar 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. 1997. Combined pricing and inventory control under uncertainty **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.
- 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.
- Genesove, David, Wallace P Mullin. 1998. Testing static oligopoly models: conduct and cost in the sugar industry, 1890-1914. *The RAND Journal of Economics* 355–377.
- 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.
- 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.
- Hanasusanto, Grani A, Daniel Kuhn, Wolfram Wiesemann. 2016. K-adaptability in two-stage distributionally robust binary programming. *Operations Research Letters* **44**(1) 6–11.
- Heien, Dale, Cathy Roheim Wesseils. 1990. Demand systems estimation with microdata: a censored regression approach. *Journal of Business & Economic Statistics* **8**(3) 365–371.
- 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, Anil K, Richard C Dubes. 1988. Algorithms for clustering data .
- Jiang, Ruiwei, Yongpei Guan. 2018. Risk-averse two-stage stochastic program with distributional ambiguity. *Operations Research* .
- Klabjan, Diego, David Simchi-Levi, Miao Song. 2013. Robust stochastic lot-sizing by means of histograms. *Production and Operations Management* **22**(3) 691–710.
- 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 Watwai. 2008. Robust multi-product pricing .
- 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.
- MacQueen, James. 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1. Oakland, CA, USA, 281–297.
- Mamani, Hamed, Shima Nassiri, Michael R Wagner. 2016. Closed-form solutions for robust inventory management. *Management Science* **63**(5) 1625–1643.
- 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.
- 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.
- Paton, Elizabeth. 2018. H&m, a fashion giant, has a problem: \$4.3 billion in unsold clothes. URL <https://www.nytimes.com/2018/03/27/business/hm-clothes-stock-sales.html>.
- 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.
- Phillips, Robert Lewis. 2005. *Pricing and revenue optimization*. Stanford University Press.
- Popescu, Ioana, Yaozhong Wu. 2007. Dynamic pricing strategies with reference effects. *Operations Research* **55**(3) 413–429.
- Ramachandran, Karthik, Necati Tereyağoğlu, Yusen Xia. 2018. Multidimensional decision making in operations: An experimental investigation of joint pricing and quantity decisions. *Management Science* .

- Rusmevichientong, Paat, Benjamin Van Roy, Peter W Glynn. 2006. A nonparametric approach to multi-product pricing. *Operations Research* **54**(1) 82–98.
- 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. Technical note—multiproduct inventory management under customer substitution and capacity restrictions. *Operations Research* **66**(3) 740–747.
- Shenoy, Dinesh, Roberto Rosas. 2018. *Problems & Solutions in Inventory Management*. Springer.
- Simchi-Levi, David, Xin Chen, Julien Bramel. 2005. *The Logic of Logistics*. Springer.
- Song, Jing-Sheng Jeannette, Zhengliang Xue Song. 2018. Demand shaping through bundling and product configuration: A dynamic multiproduct inventory-pricing model .
- Subramanyam, Anirudh, Chrysanthos E Gounaris, Wolfram Wiesemann. 2017. K-adaptability in two-stage mixed-integer robust optimization. *arXiv preprint arXiv:1706.07097* .
- Talluri, Kalyan T, Garrett J Van Ryzin. 2006. *The theory and practice of revenue management*, vol. 68. Springer Science & Business Media.
- U.S. Bureau of the Census. 2018. Retailers inventories [retailimsa], retrieved from fred, federal reserve bank of st. louis. URL <https://fred.stlouisfed.org/series/RETAILIMSA,October1,2018> .
- Van Mieghem, Jan A. 1998. Investment strategies for flexible resources. *Management Science* **44**(8) 1071–1078.
- Webber, Björn, Sarah Herrlein, Greg Hodge. 2011. The challenge of food waste. URL https://www.gs1.org/sites/default/files/gs1_uk_the_challenge_of_food_waste.pdf.
- Whitin, Thomson M. 1955. Inventory control and price theory. *Management science* **2**(1) 61–68.
- Xin, Linwei, David A Goldberg. 2015. Distributionally robust inventory control when demand is a martingale. *arXiv preprint arXiv:1511.09437* .
- 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.

Appendix A: Proofs of statements

A.1. Proof of Proposition 2

We provide a sketch of the proof that is similar to Bertsimas et al. (2018b). Note that because of the linear optimization framework, we do not require Slater's condition for strong duality to hold. Given $\mathbf{x}, \mathbf{r}^1 \in \mathcal{R}, \hat{\mathbf{r}}^2 \in \mathcal{M}(\mathbf{r}^1)$, the worst-case total revenue can be calculated as the following infinite dimensional linear program,

$$\begin{aligned} & \inf \sum_{j \in [K^1], k \in [K^2]} q_{jk} \mathbb{E}_{\mathbb{P}} [\pi(\mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2) \mid \tilde{\boldsymbol{\varepsilon}}^1 \in \mathcal{S}_j^1, \tilde{\boldsymbol{\varepsilon}}^2 \in \mathcal{S}_{jk}^2] \\ \text{s.t. } & \mathbb{E}_{\mathbb{P}} [\tilde{\boldsymbol{\varepsilon}}^1 \mid \tilde{\boldsymbol{\varepsilon}}^1 \in \mathcal{S}_j^1] = \boldsymbol{\mu}_j^1 & \forall j \in [K^1]; \\ & \mathbb{E}_{\mathbb{P}} [\tilde{\boldsymbol{\varepsilon}}^2 \mid \tilde{\boldsymbol{\varepsilon}}^1 \in \mathcal{S}_j^1, \tilde{\boldsymbol{\varepsilon}}^2 \in \mathcal{S}_{jk}^2] = \boldsymbol{\mu}_{jk}^2 & \forall j \in [K^1], k \in [K^2]; \\ & \mathbb{E}_{\mathbb{P}} [|\tilde{\boldsymbol{\varepsilon}}^1 - \boldsymbol{\mu}_j^1| \mid \tilde{\boldsymbol{\varepsilon}}^1 \in \mathcal{S}_j^1] \leq \boldsymbol{\sigma}_j^1 & \forall j \in [K^1]; \\ & \mathbb{E}_{\mathbb{P}} [|\tilde{\boldsymbol{\varepsilon}}^2 - \boldsymbol{\mu}_{jk}^2| \mid \tilde{\boldsymbol{\varepsilon}}^1 \in \mathcal{S}_j^1, \tilde{\boldsymbol{\varepsilon}}^2 \in \mathcal{S}_{jk}^2] \leq \boldsymbol{\sigma}_{jk}^2 & \forall j \in [K^1], k \in [K^2]; \\ & \mathbb{P} [\tilde{\boldsymbol{\varepsilon}}^1 \in \mathcal{S}_j^1, \tilde{\boldsymbol{\varepsilon}}^2 \in \mathcal{S}_{jk}^2] = q_{jk} & \forall j \in [K^1], k \in [K^2]; \end{aligned}$$

Here \mathbf{r}_j^2 is a new variable satisfying $\mathbf{r}_j^2 \leq \mathbf{r}^1$. Assigning $\beta_j^1, \beta_{jk}^2, \gamma_j^1, \gamma_{jk}^2$, and α_{jk} as the dual variable(s) to the five type of constraints, respectively, we can obtain the dual problem as follows

$$\begin{aligned} & \sup \sum_{j \in [K^1], k \in [K^2]} \left(\alpha_{jk} + (\beta_j^1)' \boldsymbol{\mu}_j^1 + (\beta_{jk}^2)' \boldsymbol{\mu}_{jk}^2 + (\gamma_j^1)' \boldsymbol{\sigma}_j^1 + (\gamma_{jk}^2)' \boldsymbol{\sigma}_{jk}^2 \right) \\ \text{s.t. } & \alpha_{jk} + (\beta_j^1)' \boldsymbol{\varepsilon}^1 + (\beta_{jk}^2)' \boldsymbol{\varepsilon}^2 + (\gamma_j^1)' |\boldsymbol{\varepsilon}^1 - \boldsymbol{\mu}_j^1| + (\gamma_{jk}^2)' |\boldsymbol{\varepsilon}^2 - \boldsymbol{\mu}_{jk}^2| \\ & \leq q_{jk} \pi(\mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2) & \forall \boldsymbol{\varepsilon}^1 \in \mathcal{S}_j^1, \boldsymbol{\varepsilon}^2 \in \mathcal{S}_{jk}^2, j \in [K^1], k \in [K^2]; \\ & \gamma_j^1, \gamma_{jk}^2 \leq 0 & \forall j \in [K^1], k \in [K^2]; \\ & \alpha_j, \beta_j^1, \beta_{jk}^2, \gamma_j^1, \gamma_{jk}^2 \in \mathbb{R} & \forall j \in [K^1], k \in [K^2], \end{aligned}$$

which can be rewritten as follows,

$$\begin{aligned} & \sup \sum_{j \in [K^1], k \in [K^2]} \left(\alpha_{jk} + (\beta_j^1)' \boldsymbol{\mu}_j^1 + (\beta_{jk}^2)' \boldsymbol{\mu}_{jk}^2 + (\gamma_j^1)' \boldsymbol{\sigma}_j^1 + (\gamma_{jk}^2)' \boldsymbol{\sigma}_{jk}^2 \right) \\ \text{s.t. } & \alpha_{jk} + (\beta_j^1)' \boldsymbol{\varepsilon}^1 + (\beta_{jk}^2)' \boldsymbol{\varepsilon}^2 + (\gamma_j^1)' \mathbf{u}^1 + (\gamma_{jk}^2)' \mathbf{u}^2 \\ & \leq q_{jk} \pi(\mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2) & \forall (\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \in \mathcal{W}_{jk}, j \in [K^1], k \in [K^2]; \\ & \gamma_j^1, \gamma_{jk}^2 \leq 0 & \forall j \in [K^1], k \in [K^2]; \\ & \alpha_j, \beta_j^1, \beta_{jk}^2, \gamma_j^1, \gamma_{jk}^2 \in \mathbb{R} & \forall j \in [K^1], k \in [K^2]. \end{aligned}$$

■

A.2. Proof of Theorem 1

Based on Proposition 1, given any $(\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \in \mathcal{W}_{jk}, j \in [K^1], k \in [K^2]$, the total revenue can be reformulated as follows,

$$\pi(\mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2) = \sum_{i \in [I]} \min \{ r_i^1 x_i, r_{ji}^2 x_i + (r_i^1 - r_{ji}^2) z_i^1(\mathbf{r}^1, \boldsymbol{\varepsilon}_i^1), r_i^1 z_i^1 + r_{ji}^2 z_i^2(\mathbf{r}_j^2, \boldsymbol{\varepsilon}_i^2) \}.$$

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

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

Denote all partitions of set $[I]$ as

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

and let

$$\begin{aligned} & \phi(\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2) \\ & \triangleq \sum_{i \in \mathcal{I}_1} r_i^1 x_i + \sum_{i \in \mathcal{I}_2} (r_{ji}^2 x_i + (r_i^1 - r_{ji}^2) z_i^1(\mathbf{r}^1, \boldsymbol{\varepsilon}_i^1)) + \sum_{i \in \mathcal{I}_3} (r_i^1 z_i^1(\mathbf{r}^1, \boldsymbol{\varepsilon}_i^1) + r_{ji}^2 z_i^2(\mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}_i^2)), \end{aligned}$$

for all $(\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \in \mathcal{W}_{jk}, j \in [K^1], k \in [K^2]$, then the total revenue can be rewritten as

$$\pi(\mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2) = \min_{\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3 \in \mathcal{B}} \phi(\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2).$$

Hence,

$$\alpha_{jk} + (\boldsymbol{\beta}_j^1)' \boldsymbol{\varepsilon}^1 + (\boldsymbol{\beta}_{jk}^2)' \boldsymbol{\varepsilon}^2 + (\boldsymbol{\gamma}_j^1)' \mathbf{u}^1 + (\boldsymbol{\gamma}_{jk}^2)' \mathbf{u}^2 \leq q_{jk} \min_{\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3 \in \mathcal{B}} \phi(\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2),$$

is equivalent to

$$\alpha_{jk} + (\boldsymbol{\beta}_j^1)' \boldsymbol{\varepsilon}^1 + (\boldsymbol{\beta}_{jk}^2)' \boldsymbol{\varepsilon}^2 + (\boldsymbol{\gamma}_j^1)' \mathbf{u}^1 + (\boldsymbol{\gamma}_{jk}^2)' \mathbf{u}^2 \leq q_{jk} \phi(\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2), \forall \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3 \in \mathcal{B}. \quad \blacksquare$$

A.3. Proof of Theorem 2

Following from Proposition 1, suppose $v_{ijk} \in \mathcal{L}, \forall i \in [I], j \in [K^1], k \in [K^2]$ satisfies

$$\left. \begin{aligned} v_{ijk}(\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) & \leq r_i^1 x_i, \\ v_{ijk}(\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) & \leq r_{ji}^2 x_i + (r_i^1 - r_{ji}^2) (a_i^1 + \mathbf{e}'_i \mathbf{B}^1 \mathbf{r}^1 + \boldsymbol{\varepsilon}_i^1) \\ v_{ijk}(\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) & \leq r_i^1 (a_i^1 + \mathbf{e}'_i \mathbf{B}^1 \mathbf{r}^1 + \boldsymbol{\varepsilon}_i^1) \\ & \quad + r_{ji}^2 (a_i^2 + \mathbf{e}'_i \mathbf{B}^2 \mathbf{r}_j^2 + \mathbf{e}'_i \mathbf{B}^{21} \mathbf{r}^1 + \boldsymbol{\varepsilon}_i^2) \end{aligned} \right\} \forall (\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \in \mathcal{W}_{jk},$$

we then have

$$\sum_{i \in [I]} v_{ijk}(\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \leq \pi(\mathbf{x}, \mathbf{r}^1, \mathbf{r}_j^2, \boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2) \quad \forall (\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \in \mathcal{W}_{jk}$$

and the formulation in Problem (7) ensues with appropriate linearization for some nonlinear terms in \mathbf{r}^1 , \mathbf{x} , and \mathbf{r}_j^2 . Specifically, for each $i \in [I]$, we let $\mathbf{p}_i = (p_{i1}, \dots, p_{iN})$ and introduce binary decision variables $\mathbf{y}_i^1, \mathbf{y}_{i1}^2, \dots, \mathbf{y}_{iJ}^2 \in \{0, 1\}^N$, satisfying $\mathbf{1}' \mathbf{y}_i^1 = \mathbf{1}' \mathbf{y}_{i1}^2 = \dots = \mathbf{1}' \mathbf{y}_{iJ}^2 = 1$ so that $r_i^1 = \mathbf{p}'_i \mathbf{y}_i^1$, $r_{ji}^2 = \mathbf{p}'_i \mathbf{y}_{ji}^2$, $j \in [K^1]$. The constraint $\dots \leq r_i^1 x_i$ can be linearized by $\dots \leq w_i^1$, for some decision variable w_i^1 satisfying

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

Same technique applies to the constraint $\dots \leq r_{ji}^2 x_i$. Similarly, the constraint $\dots \leq r_i^1 \mathbf{e}'_i \mathbf{B}^1 \mathbf{r}^1$ can be linearized by $\dots \leq \rho_i^{11}$, for some decision variable ρ_i^{11} satisfying

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

Same technique applies to $\dots \leq r_{ji}^2 \mathbf{e}'_i \mathbf{B}^1 \mathbf{r}^1$, $\dots \leq r_{ji}^2 \mathbf{e}'_i \mathbf{B}^2 \mathbf{r}_j^2$, and $\dots \leq r_{ji}^2 \mathbf{e}'_i \mathbf{B}^{21} \mathbf{r}^1$.

A.4. Proof of Proposition 3

For any $j \in [K^1], k \in [K^2]$, the inequality $(\mathbf{f}^1)' \boldsymbol{\varepsilon}^1 + (\mathbf{h}^1)' \mathbf{u}^1 + (\mathbf{f}^2)' \boldsymbol{\varepsilon}^2 + (\mathbf{h}^2)' \mathbf{u}^2 \leq d, (\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \mathbf{u}^1, \mathbf{u}^2) \in \mathcal{W}_{jk}$ is equivalent to $RC \leq d$, where

$$\begin{aligned}
RC &= \max (\mathbf{f}^1)' \boldsymbol{\varepsilon}^1 + (\mathbf{h}^1)' \mathbf{u}^1 + (\mathbf{f}^2)' \boldsymbol{\varepsilon}^2 + (\mathbf{h}^2)' \mathbf{u}^2 \\
s.t. \quad & 2(\boldsymbol{\varepsilon}^1)' (\boldsymbol{\mu}_l^1 - \boldsymbol{\mu}_j^1) \leq (\boldsymbol{\mu}_l^1)' \boldsymbol{\mu}_l^1 - (\boldsymbol{\mu}_j^1)' \boldsymbol{\mu}_j^1 \quad \forall l \in [K^1]; \quad : \kappa_{jl}^1 \\
& 2(\boldsymbol{\varepsilon}^2)' (\boldsymbol{\mu}_{jm}^2 - \boldsymbol{\mu}_{jk}^2) \leq (\boldsymbol{\mu}_{jm}^2)' \boldsymbol{\mu}_{jm}^2 - (\boldsymbol{\mu}_{jk}^2)' \boldsymbol{\mu}_{jk}^2 \quad \forall m \in [K^2]; \quad : \kappa_{jkm}^2 \\
& \mathbf{u}^1 - \boldsymbol{\varepsilon}^1 \geq -\boldsymbol{\mu}_j^1; \quad : \boldsymbol{\tau}_j^{11} \\
& \mathbf{u}^1 + \boldsymbol{\varepsilon}^1 \geq \boldsymbol{\mu}_j^1; \quad : \boldsymbol{\tau}_j^{12} \\
& \mathbf{u}^2 - \boldsymbol{\varepsilon}^2 \geq -\boldsymbol{\mu}_{jk}^2; \quad : \boldsymbol{\tau}_{jk}^{21} \\
& \mathbf{u}^2 + \boldsymbol{\varepsilon}^2 \geq \boldsymbol{\mu}_{jk}^2; \quad : \boldsymbol{\tau}_{jk}^{22} \\
& \boldsymbol{\varepsilon}^1 \geq \underline{\boldsymbol{\varepsilon}}^1; \quad : \boldsymbol{\omega}^{11} \\
& \boldsymbol{\varepsilon}^1 \leq \bar{\boldsymbol{\varepsilon}}^1; \quad : \boldsymbol{\omega}^{12} \\
& \boldsymbol{\varepsilon}^2 \geq \underline{\boldsymbol{\varepsilon}}_j^2; \quad : \boldsymbol{\omega}_j^{21} \\
& \boldsymbol{\varepsilon}^2 \leq \bar{\boldsymbol{\varepsilon}}_j^2. \quad : \boldsymbol{\omega}_j^{22}
\end{aligned} \tag{12}$$

The dual of the above linear programming is as follows

$$\begin{aligned}
\min \quad & \sum_{l \in [K^1]} \kappa_{jl}^1 \left((\boldsymbol{\mu}_l^1)' \boldsymbol{\mu}_l^1 - (\boldsymbol{\mu}_j^1)' \boldsymbol{\mu}_j^1 \right) + \sum_{m \in [K^2]} \left((\boldsymbol{\mu}_{jm}^2)' \boldsymbol{\mu}_{jm}^2 - (\boldsymbol{\mu}_{jk}^2)' \boldsymbol{\mu}_{jk}^2 \right) - (\boldsymbol{\tau}_j^{11})' \boldsymbol{\mu}_j^1 \\
& + (\boldsymbol{\tau}_j^{12})' \boldsymbol{\mu}_j^1 - (\boldsymbol{\tau}_{jk}^{21})' \boldsymbol{\mu}_{jk}^2 + (\boldsymbol{\tau}_{jk}^{22})' \boldsymbol{\mu}_{jk}^2 + (\boldsymbol{\omega}^{11})' \underline{\boldsymbol{\varepsilon}}^1 + (\boldsymbol{\omega}^{12})' \bar{\boldsymbol{\varepsilon}}^1 + (\boldsymbol{\omega}_j^{21})' \underline{\boldsymbol{\varepsilon}}_j^2 + (\boldsymbol{\omega}_j^{22})' \bar{\boldsymbol{\varepsilon}}_j^2 \\
s.t. \quad & 2\kappa_{jl}^1 (\boldsymbol{\mu}_l^1 - \boldsymbol{\mu}_j^1) - \boldsymbol{\tau}_j^{11} + \boldsymbol{\tau}_j^{12} + \boldsymbol{\omega}^{11} + \boldsymbol{\omega}^{12} = \mathbf{f}^1 \quad \forall l \in [K^1]; \\
& 2\kappa_{jkm}^2 (\boldsymbol{\mu}_{jm}^2 - \boldsymbol{\mu}_{jk}^2) - \boldsymbol{\tau}_{jk}^{21} + \boldsymbol{\tau}_{jk}^{22} + \boldsymbol{\omega}_j^{21} + \boldsymbol{\omega}_j^{22} = \mathbf{f}^2 \quad \forall k \in [K^2]; \\
& \boldsymbol{\tau}_j^{11} + \boldsymbol{\tau}_j^{12} = \mathbf{h}^1; \\
& \boldsymbol{\tau}_{jk}^{21} + \boldsymbol{\tau}_{jk}^{22} = \mathbf{h}^2; \\
& \kappa_{jl}^1, \kappa_{jkm}^2, \boldsymbol{\omega}^{12}, \boldsymbol{\omega}_j^{22} \geq 0; \\
& \boldsymbol{\tau}_j^{11}, \boldsymbol{\tau}_j^{12}, \boldsymbol{\tau}_{jk}^{21}, \boldsymbol{\omega}^{11}, \boldsymbol{\omega}_j^{21}, \boldsymbol{\omega}_j^{22} \leq 0.
\end{aligned} \tag{13}$$

■

Appendix B: Procedure to process the data and get demand coefficients

We first depict the data structure with a screenshot in Figure 3.

- Step 1. Select the top 10 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 x th % 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 limit = $q(10) - 2\text{IQR}$, upper limit = $q(90) + 2\text{IQR}$. Here $\text{IQR} = q(75) - q(25)$.

	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	21/8/2007	2007	40	10.202	80	13.374	39	7.755	61	21.132	49	18.889
3	22/8/2007	2007	33	11.153	69	13.438	40	8.088	63	21.109	56	18.715
4	23/8/2007	2007	39	10.685	91	13.478	52	7.910	56	20.983	47	18.052
5	24/8/2007	2007	45	10.759	97	13.320	26	7.742	70	21.288	61	18.420
6	25/8/2007	2007	69	10.217	141	13.331	82	7.664	95	21.005	73	18.319
7	26/8/2007	2007	67	10.488	137	13.391	65	7.784	90	21.238	94	18.489
8	27/8/2007	2007	40	10.928	54	13.292	24	7.717	36	21.462	34	18.200
9	28/8/2007	2007	55	9.378	79	13.206	22	7.776	41	20.749	48	18.449
10	29/8/2007	2007	48	10.461	56	13.276	42	7.799	40	20.761	40	18.472
11	30/8/2007	2007	45	9.879	63	13.463	35	7.503	42	20.963	43	18.495
12	31/8/2007	2007	37	8.075	64	13.337	14	7.989	36	20.963	47	17.295
13	1/9/2007	2007	47	10.498	95	13.409	24	7.677	78	20.826	66	18.392
14	2/9/2007	2007	77	9.533	79	13.241	27	7.684	80	20.865	97	18.299
15	3/9/2007	2007	12	11.809	26	14.798	5	8.221	9	23.767	13	20.777
16	4/9/2007	2007	20	11.809	23	14.798	2	8.221	27	23.767	28	20.777
17	5/9/2007	2007	16	11.809	16	14.798	5	8.221	14	23.767	16	20.777
18	6/9/2007	2007	21	11.809	19	14.548	2	8.221	13	23.767	16	20.777
19	7/9/2007	2007	21	11.786	13	14.798	5	8.221	9	23.767	14	20.777
20	8/9/2007	2007	28	13.303	29	14.798	7	8.221	30	23.767	31	20.777
21	9/9/2007	2007	16	13.303	28	14.798	1	8.221	20	23.457	28	20.777
22	10/9/2007	2007	13	13.303	17	14.798	21	8.221	11	23.767	15	20.777
23	11/9/2007	2007	13	13.303	24	14.798	54	7.880	18	23.767	13	20.777
24	12/9/2007	2007	13	13.303	25	14.798	59	8.221	16	23.767	23	20.777
25	13/9/2007	2007	9	13.303	38	14.798	91	8.221	13	23.767	16	20.777
26	14/9/2007	2007	16	13.303	61	14.350	73	8.221	29	22.314	9	20.777
27	15/9/2007	2007	24	13.303	49	14.798	140	8.183	26	23.767	24	20.500
28	16/9/2007	2007	38	13.303	69	14.798	130	8.185	41	23.767	16	20.777
29	17/9/2007	2007	11	13.303	30	14.798	43	8.132	13	23.767	18	20.777
30	18/9/2007	2007	11	13.303	30	14.798	62	7.455	27	23.767	15	20.777
31	19/9/2007	2007	13	13.303	30	14.798	72	8.147	18	23.767	9	20.777

Figure 3 A screenshot of the dataset

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 $i \in \{1, \dots, 10\}$ one the model of the form:

$$z_{ih} = a_{ih} + B_{ii}r_{ih} + \sum_{m \neq i} B_{mi}r_{mh} + t_i \text{Year} + \varepsilon_{ih}.$$

A summary the linear regression results for all the 10 products is in Table 6. We also test the cross-product effect on the SKU level in the store with the most sales. One such example is for product 1 in Table 4.

Obtaining demand parameters

We sequentially partition the 10 products into four 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 7.

Table 6 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}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
r_1	-9.963*** (0.671)	-4.406*** (0.681)	-1.345** (0.547)	-2.988*** (0.348)	-4.211*** (0.344)	-3.381*** (0.895)	-2.235*** (0.306)	-1.853*** (0.286)	-1.649*** (0.241)	-3.527*** (0.572)
r_2	-0.130 (0.481)	-8.328*** (0.489)	-0.772** (0.392)	-1.716*** (0.249)	-0.125 (0.247)	1.381** (0.642)	0.564** (0.220)	-0.414** (0.205)	-0.064 (0.173)	-0.138 (0.410)
r_3	0.353 (0.881)	5.258*** (0.895)	-4.422*** (0.718)	2.679*** (0.457)	0.434 (0.452)	2.497** (1.174)	-3.128*** (0.402)	1.247*** (0.376)	1.388*** (0.317)	3.418*** (0.751)
r_4	1.537** (0.610)	-2.203*** (0.619)	-1.187** (0.497)	-3.351*** (0.316)	-1.326*** (0.313)	-1.524* (0.813)	-0.956*** (0.278)	-0.220 (0.260)	-0.540** (0.219)	-1.407*** (0.520)
r_5	-0.072 (0.442)	2.503*** (0.449)	1.862*** (0.360)	0.659*** (0.229)	-0.150 (0.226)	1.980*** (0.589)	-0.128 (0.202)	-0.119 (0.189)	-0.044 (0.159)	0.141 (0.377)
r_6	1.748*** (0.481)	-0.756 (0.488)	0.700* (0.392)	-0.681*** (0.249)	-0.594** (0.246)	-10.218*** (0.641)	0.788*** (0.219)	-0.259 (0.205)	-0.549*** (0.173)	-1.658*** (0.410)
r_7	-0.745 (0.485)	0.545 (0.492)	0.048 (0.395)	0.157 (0.251)	-1.100*** (0.248)	-0.982 (0.646)	-4.011*** (0.221)	-0.571*** (0.207)	-0.169 (0.174)	-1.163*** (0.413)
r_8	2.633*** (0.606)	-0.399 (0.616)	1.539*** (0.494)	-0.811*** (0.314)	0.928*** (0.311)	3.774*** (0.809)	0.911*** (0.277)	-2.023*** (0.259)	0.516** (0.218)	2.172*** (0.517)
r_9	0.657 (0.445)	-0.320 (0.452)	0.055 (0.363)	0.064 (0.231)	-0.202 (0.228)	-1.347** (0.593)	0.284 (0.203)	0.570*** (0.190)	-2.568*** (0.160)	1.040*** (0.379)
r_{10}	-0.088 (0.540)	0.329 (0.548)	-0.755* (0.440)	-0.351 (0.280)	-0.367 (0.277)	-1.970*** (0.720)	0.122 (0.247)	0.488** (0.230)	0.641*** (0.194)	-5.366*** (0.460)
Intercept	57.228*** (11.218)	191.285*** (11.393)	86.032*** (9.141)	158.637*** (5.814)	123.751*** (5.750)	152.754*** (14.957)	86.124*** (5.123)	66.403*** (4.788)	62.581*** (4.034)	117.936*** (9.566)

Table 7 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
$\mathbf{1}(Year_{2008})$	16.4519	10.7582	-1.0468	7.6760
$\mathbf{1}(Year_{2009})$	24.2390	17.1783	0.8701	12.5888
$\mathbf{1}(Year_{2010})$	33.2558	22.9417	-46.9984	14.9437
$\mathbf{1}(Year_{2011})$	27.7001	25.1247	-49.4063	14.3893
$\mathbf{1}(Year_{2012})$	51.0636	35.8613	-34.6589	15.9042

Appendix C: Computation study parameters

Parameter in the the simulation study are set by following (in part) the numerical study of den Boer (2014). Specifically, we set $\mathbf{a}^1 = (56.32, 59.57, 57.10, 57.70, 58.04)'$, $\mathbf{a}^1 = (51.21, 54.32, 52.25, 52.63, 53.68)'$, and $\mathbf{c} = (2.53, 1.87, 2.62, 1.84, 1.82)'$ which is half of the optimal price in den Boer (2014). The price elasticity matrix \mathbf{B}^1 is quoted as follows:

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

and $\mathbf{B}^2 = \text{diag}((-2.50, -2.30, -3.50, -2.60, -4.10))$, $\mathbf{B}^{21} = \mathbf{0}$. 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 $\mathbf{p}_1 = \{2.53, 5.02, 7.52, 10.01, 12.51, 15.00\}$. We also set a different admissible price set for each product of the second period with $\mathbf{p}^2 = 0.9\mathbf{p}^1$. The first period price is given as $\mathbf{r}^1 = (12.51, 12.37, 10.05, 12.37, 9.73)'$.