Data-driven learning in dynamic pricing
using adaptive optimization

Dimitris Bertsimas

MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139, dbertsim@mit.edu

Phebe Vayanos

MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139, pvayanos@mit.edu

We consider the pricing problem faced by a retailer endowed with a finite inventory of a product offered over a
finite planning horizon in an environment where customers are price-sensitive. The parameters of the product
demand curve are fixed but *unknown* to the seller who only has at his disposal a history of sales data. We
propose an adaptive optimization approach to setting prices that captures the ability of the seller to exploit
information gained as a byproduct of pricing in his quest to maximize revenues. We construct *data-driven*
uncertainty sets that encode the beliefs of the retailer about the demand curve parameters. We capture the
ability of the retailer to *explore* the characteristics of customer behavior by allowing the uncertainty set to
depend on the pricing decisions. We model his capacity to *exploit* the information dynamically acquired by
letting the pricing decisions adapt to the history of observations. These modeling features enable us to unify
optimization and estimation as the uncertainty set is updated "on-the-fly", during optimization. We propose
a hierarchical approximation scheme for the resulting adaptive optimization problem with decision-dependent
uncertainty set which yields a practically tractable mixed-binary conic optimization problem. We discuss
several variants and extensions of our model that illustrate the versatility of the proposed method. We present
computational results that show that the proposed policies: *(a)* yield higher profits compared to commonly
used policies, *(b)* nearly match perfect information results with respect to downside measures such as the
Conditional Value-at-Risk, and *(c)* can be obtained in modest computational time for large-scale problems.

Key words: dynamic pricing, learning-earning, exploration-exploitation, decision rule, adjustable robust
optimization, decision-dependent uncertainty set, generalized semi-infinite programming.

1. Introduction

Dynamic pricing is a business strategy concerned with periodically adjusting the prices of products to reflect changes in circumstances in an environment where customers are price-sensitive and with aim to maximize long-term profitability. Common events that justify an adjustment in selling price of a product include changes in market conditions, increase or depletion of inventory or resources (supply availability), modifications in customer demand behavior (due to e.g., the selling-season, the introduction of complement products), or increase in knowledge about the demand response.

Dynamic pricing policies have been long-employed in the travel, hospitality, and energy sectors (where short-term capacity is rigid) to mitigate imbalances in supply and demand. The main reason for the early adoption of dynamic pricing strategies in these industries was the ability to change prices at low cost and in a centralized fashion (Elmaghraby and Keskinocak (2003)). In contrast, in industries with more flexible supply, such as retail, imbalances between supply and demand have traditionally been moderated by means of dynamic inventory control. While active inventory management is a useful tool for increasing profit, it only serves to decrease costs. Dynamic pricing on the other hand provides a means for also affecting revenues provided the population to which the product is offered is price-sensitive. Unfortunately, high menu costs precluded such industries from regularly adjusting their prices. The emergence of the internet as a sales channel and the increasing use of digital price tags have drastically decreased such costs, while technological advances are by and large permitting the automation of price changes. Dynamic pricing is thus becoming ubiquitous across most industries, enabling retailers to leverage on both sides of the profit equation.

In this paper, we focus on a variant of the dynamic pricing problem, often referred to as *the tactical pricing problem*, whereby a firm is endowed with a finite inventory of a single product available for sale. Thus, our methodology is adequate for products with rigid capacity, short life-cycle (relative to their procurement lead time) (e.g., holiday goods, fashion apparel, products produced oversees), or for products at the end of their life-cycle. We discuss how our framework extends to the *dynamic pricing problem with inventory control*, the *multi-product pricing problem with finite inventories* and the *network revenue management problem*, among others.

The need for modeling demand as an uncertain parameter was recognized very early (Mills (1959)). Nevertheless, a critical assumption made by most academic studies incorporating uncertainty is that the *demand curve* or *demand function*, which maps prices to expected demand, is completely known by the firm. While it has the benefit of reducing the computational complexity of the underlying problems (since the only source of uncertainty is that in the *residual demand*), this assumption of *full-information* is unrealistic in most practical settings. Indeed, knowledge of the characteristics of customer behavior for any given product is typically incomplete or even lacking altogether. This is enhanced by the fact that over the last decades, product life-cycles have decreased (Elmaghraby and Keskinocak (2003)) while new products constantly emerge in the markets. Thus, it is natural to assume that the demand curve is *unknown.*

When the demand curve is unknown, the firm is faced with the trade-off between *exploitation* (pricing to maximize revenue) and *exploration* (demand learning). On the one hand, exploiting available information increases short-term profitability. On the other hand, it may lead to high opportunity costs in the long run. Indeed, in the context of revenue management, demand learning occurs as a byproduct of pricing. Thus, in order to build an accurate model for the demand curve, a retailer must experiment with different prices, invariably deviating from the *myopic* pricing strategy which is optimal based on current information. Rothshield (1974) and McLennan (1984) were the first to highlight the possibility of *incomplete learning* in dynamic pricing (building upon the theory of multi-armed bandits originally proposed by Robbins (1951)). Recognizing that pricing patterns impact their ability to learn the demand curve, retailers increasingly proceed with price experimentation. While frequent price changes may damage the reputation of a firm, they may prove crucial in determining the (asymptotically) optimal selling prices of products.

Beyond the computational complexities associated with the determination of a pricing strategy that optimally balances between learning and earning, even the exploitation of information associated with a simple myopic strategy requires *data.* Historically, this has hampered all but the most technologically advanced firms (that had the ability to gather, store and analyze such data

sets) from learning customer behavior. In recent years, acknowledging that data can help them take more *informed decisions* (and facilitated by a reduction in the costs of information technology and the increased use of electronic transactions), most firms have started accumulating sales data. Thus, it is believed that, in the coming years, most retailers will have the ability to dynamically adjust their prices at low cost and integrating the latest data in their decision-taking.

In this context, firms are thus faced with two questions:

(a) How to build an accurate model of demand curve uncertainty from the available data?

(b) Given this model, how to determine (compute) a pricing strategy that optimally balances between exploitation and exploration?

In this paper, we address both these questions in a unified fashion.

1.1. Literature review

Dynamic pricing in an uncertain environment has received considerable attention by researchers in numerous areas ranging from operations research and management science to economics, computer science and control. The focal point of this review is the recent literature on dynamic pricing with learning in a parametric (stationary) monopolist environment. We also briefly discuss the literature on robust dynamic pricing. For a survey of the research on dynamic pricing (without learning), we refer the interested reader to the books by Talluri and van Ryzin (2004) and Phillips (2005), and the review papers by Bitran and Caldentey (2003) and Elmaghraby and Keskinocak (2003).

Parametric models with Bayesian learning. The leading line of research on dynamic pricing takes a *Bayesian* approach. Thus, a parametric model of the demand is postulated jointly with a prior distribution that reflects the seller's initial knowledge of the model parameters. Demand observations are used to update the prior into a posterior through the application of Bayes' rule.

The vast majority of studies cast in this framework assume that customers arrive according to a homogeneous Poisson process with unknown rate (which can be partially resolved through judicious pricing) and purchase the product with a known price-dependent probability (directly associated with their willingness to pay), see Aviv and Pazgal (2005), Lin (2006), Araman and

Caldentey (2009), Farias and van Roy (2010). Others assume that one of a finite number of models is known to apply, see Harrison et al. (2012). The probability associated with the validity of each model is then updated with additional sales observations. Unfortunately, these formulations do not generally admit an analytic solution and problems of realistic size are computationally intractable.

Thus, researchers have resorted to the design and analysis of heuristics with attractive theoretical properties. For example, Aviv and Pazgal (2005) propose a certainty equivalent heuristic, Araman and Caldentey (2009) propose a greedy pricing heuristic, Farias and van Roy (2010) propose a decay balancing heuristic, while Harrison et al. (2012) propose variants of myopic policies.

A common criticism of the Bayesian modeling paradigm is that, in order to obtain a posterior of the same form as the prior, the prior distribution is artificially constrained to belong to a class of distributions conjugate to the demand process, thus reducing modeling flexibility.

Parametric models and learning with classical estimation. A related line of research proposes to model the demand curve as a fixed parametric function of the price, whose parameters are unknown to the seller, and inference is made by e.g., least-squares or maximum-likelihood estimation. In this context, the optimal pricing strategy can, in principle, be computed by dynamic programming (following the principles of dual control theory, see e.g., Feldbaum (1961)). Unfortunately, this methodology suffers from the *curse of dimensionality* while no closed-form solutions are available.

Thus, several researchers have resorted to approximations. Lobo and Boyd (2003) approximate the value function in the dynamic program by linearizing the inverse of the covariance matrix of the unknown parameters around the myopically optimal policy with "dithering". Carvalho and Puterman (2003, 2005a,b) propose a one-step look ahead policy which, at each step, chooses the price that approximately maximizes the sum of the revenues in the next two periods. Bertsimas and Perakis (2006) reduce the state space of the dynamic program by relying on the principles of least-squares estimation and show that their methodology extends to competitive environments.

Others construct suboptimal policies that are shown to have desirable theoretical properties. Besbes and Zeevi (2009) develop a parametric pricing policy based on maximum likelihood estimation, establish lower bounds on the regret of any policy and show that their policies are close to

Bertsimas and Vayanos: *Data-driven learning in dynamic pricing*

6   Article submitted to *Operations Research*; manuscript no. (Please, provide the manuscript number!)

this lower bound. Broder and Rusmevichientong (2012) propose maximum likelihood based policies that cycle between explicit price experimentation and myopically optimal pricing policies. Harrison et al. (2013) develop variants of a greedy iterated least-squares policies which are shown to be asymptotically optimal. Den Boer (2013c) proposes a variant of a certainty equivalent pricing strategy and demonstrates that this policy yields convergence of the parameter estimates to their true values. Den Boer (2013b) and den Boer and Zwart (2013) enhance the certainty equivalent pricing policy with a "taboo interval" around the average of previously chosen prices and show that with this type of policy, the value of the optimal price will be learned.

Robust approaches. Several authors have proposed to apply the robust optimization paradigm, whereby the uncertain parameters are assumed to lie in an uncertainty set (see Ben-Tal et al. (2009) and Bertsimas et al. (2011a)), to the dynamic pricing problem. The vast majority of this research stream assumes that the demand model is perfectly known, see e.g., Thiele (2006, 2009), Lobel and Perakis (2010), or time-varying and sequentially revealed at each stage, see e.g., Adida and Perakis (2006, 2010). Thus, there is no opportunity for learning. A related line of study takes the distributionally robust approach. In this setting, the seller is immunized against multiple priors in the neighborhood of a given distribution, see e.g., Lim and Shanthikumar (2007) and Lim et al. (2008). While these latter formulations incorporate model uncertainty, they do not capture the learning ability of the retailer and may thus result in overly conservative solutions. An exception in this line of research is the paper by Eren and Maglaras (2010), which assumes that demand is noise-free so that the model will become fully known if the firm experiments with all feasible price points.

For a more in-depth review of the literature on dynamic pricing with learning, we refer to the recent paper by den Boer (2013a).

1.2. Proposed approach and contributions

The goal of this paper is to present a data-driven and distribution-free paradigm for demand learning in dynamic pricing. We postulate a parametric form for the demand curve but do not assume that the residual demand comes from a specific distribution nor that this distribution is

known to the retailer. Instead, we take the robust optimization view-point and merely require that the residual demand be norm-bounded. This modeling paradigm enables us to construct a meaningful *prior uncertainty set* for the demand curve parameters with even few historical data points. We show that this set can naturally be updated into a *posterior uncertainty set*, which progressively "shrinks" around the true demand curve parameters as more price-demand pairs are observed. In particular, all of the sets constructed are *guaranteed* to contain the true parameters, while prior and posterior uncertainty sets are of the same form, an attractive feature as far as tractability is concerned. In order to capture the ability of the retailer to learn the posterior uncertainty sets, we model his pricing decisions as functions of the history of observations. Similarly, we capture his ability to explore the set of demand curve parameters by allowing the set of price-demand pairs to depend on the pricing policy selected. This construction enables us to unify optimization and (dynamic) estimation. The resulting dynamic pricing problem is an adaptive optimization problem with policy-dependent uncertainty set. We propose a hierarchy of approximation schemes inspired from techniques commonly used in robust optimization, which results in a practically tractable formulation.

The main results and contributions of this paper are summarized below:

1. We propose a novel *data-driven* distribution-free paradigm for dynamic learning that *unifies optimization and estimation*. We use techniques inspired from system identification to construct data-driven uncertainty sets that learn the unknown parameters *online* during optimization and compute *adaptive* policies that exploit the information acquired in real time.

2. Our methodological contributions are twofold:

   (a) From a modeling perspective, we propose to capture the ability of the retailer to explore the characteristics of customer behavior by allowing the uncertainty set to depend on the pricing decisions. We model his capacity to exploit the information dynamically acquired by letting the pricing decisions adapt to the history of observations. These modeling features result in an adaptive optimization problem with policy-dependent uncertainty

8

Bertsimas and Vayanos: *Data-driven learning in dynamic pricing*
Article submitted to *Operations Research*; manuscript no. (Please, provide the manuscript number!)

set. To the best of our knowledge, this is the first model of this type proposed in the literature. It naturally captures the trade-off between exploration and exploitation.

(b) From a solution standpoint, we propose a hierarchical inner approximation scheme for adaptive optimization with policy-dependent uncertainty set. We demonstrate that under this approximation, the dynamic pricing problem is equivalent to a mixed-binary conic problem that is practically tractable. Moreover, we suggest numerous strategies that mitigate the loss of optimality of the approximation at low computational overhead. To the best of our knowledge, this is the first solution approach proposed for this problem type.

3. We provide computational evidence that shows that the proposed policies: *(a)* yield higher profits at modest computational expense compared to commonly used pricing strategies, and *(b)* perform nearly as well as perfect information policies with respect to downside measures such as the Conditional Value-at-Risk.

4. We discuss numerous variants and extensions of the basic model that incorporate multiple products and inventory decisions, thus illustrating the versatility of the method. We emphasize that both our proposed modeling paradigm and solution approach remain applicable outside the realm of dynamic pricing.

The paper is organized as follows. The remainder of this section introduces the notation, while Section 2 describes the pricing problem, the data, and the demand model under consideration. The mathematical formulation of the problem is provided in Section 3. Section 4 gives insights into the structure of the optimal pricing policies, and the proposed solution approach is detailed in Section 5. Section 6 describes numerous extensions to the pricing problem to which our solution paradigm remains applicable, while heuristic approaches commonly employed in practice are described in Section 7. Finally, Section 8 reports on numerical results.

Notation. Throughout this paper, vectors (matrices) are denoted by boldface lowercase (uppercase) letters. We let $\mathbf{e}$ denote a vector of all ones; its size will be clear from the context. For any $p \in [1, \infty]$ and $n \in \mathbb{N}$, we denote the standard ℓ_p-norm in $\mathbb{R}^n$ by $\|\cdot\|_p$ and the pth-order cone in $\mathbb{R}^{n+1}$ by $\mathcal{K}_p^{n+1} := \{(\boldsymbol{x}, t) \in \mathbb{R}^{n+1} : \|\boldsymbol{x}\|_p \leq t\}$. For $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^m$, we let $\mathcal{Y}^{\mathcal{X}}$ denote the set of all functions from $\mathcal{X}$ to $\mathcal{Y}$. By convention, we define $\mathcal{Y}^{\emptyset} \equiv \mathcal{Y}$.

2. Problem description, historical data and demand dynamics

In this section, we describe the pricing problem that will be the central focus of the paper, the demand model that will prevail in our exposition, and the mild assumptions on the historical data.

2.1. The (tactical) pricing problem

We consider the pricing problem faced by a firm, henceforth referred to as the *seller*, offering a single product over the finite planning horizon $\mathcal{T} := \{1, \ldots, T\}$ in an environment where customers are price-sensitive. The seller is endowed with a (fixed) finite inventory (capacity) c of the product. At the beginning of each period $t \in \mathcal{T}$, he must choose a price $p_t \in [l, u]$, $0 < l \leq u < \infty$, for his product. We refer to $[l, u]^T$ as the feasible price set. At the end of period t, the seller observes the (price-sensitive) demand d_t generated during period t and resulting in a revenue $d_t p_t$. Cumulative demand generated during the planning horizon is allowed to exceed capacity at a cost $b \in \mathbb{R}_+$ per unit of the product (often termed backlogging or backorder cost in the inventory management literature). Any inventory remaining at the end of the horizon must be held on the premises at cost $h \in \mathbb{R}_+$ per unit of the product (this is usually referred to as holding cost). Thus, for a given price sequence and ensuing demand realization, the profit function of the seller is expressible as

$$\sum_{t \in \mathcal{T}} d_t p_t - \max \left\{ h \left(c - \sum_{t \in \mathcal{T}} d_t \right), \ b \left(\sum_{t \in \mathcal{T}} d_t - c \right) \right\}. \tag{1}$$

We remark that the requirement for the stage-wise feasible price sets $[l, u]$ to be constant over the planning horizon is non-restrictive and is merely introduced in order to simplify notation. In fact, our approach remains applicable in the case when the vector of prices $(p_1, \ldots, p_T)$ is restricted to lie in a polyhedral set (possibly intersected with a discrete set). In particular, we are able to model *mark-up* $(p_{t+1} \geq p_t)$, *mark-down* $(p_{t+1} \leq p_t)$ and *absolute-deviation* $(|p_{t+1} - p_t| \leq \rho)$ constraints that commonly arise in practice.

2.2. Demand dynamics and historical data

We assume that, at the beginning of the planning horizon, the seller has at his disposal H price-demand pair realizations. These may correspond to historical observations, the results of a market

survey, etc. For notational convenience, and independently of how they were generated, we index these observations by non-positive time $t \in \mathcal{H} := \{-H+1, \ldots, 0\}$, thus viewing them as successive historical realizations.

We assume that the demand for the product at time $t \in \overline{\mathcal{T}} := \mathcal{H} \cup \mathcal{T}$ is given by a *fixed* function of the price p_t for that period corrupted by an additive error-term. We refer to $\overline{\mathcal{T}}$ as the *selling horizon* (nuance with the planning horizon $\mathcal{T}$). We focus on *linear* demand functions of the form

$$d_t = \alpha + \beta p_t + \epsilon_t \quad \forall t \in \overline{\mathcal{T}}. \tag{2}$$

Thus, observed demand is the sum of two parts: a deterministic price-dependent part $(\alpha + \beta p_t)$, and a random component (ϵ_t). Inspired from robust optimization, we assume that no statistical knowledge of ϵ_t, $t \in \overline{\mathcal{T}}$, is available and postulate that $\boldsymbol{\epsilon} := (\epsilon_t)_{t \in \overline{\mathcal{T}}}$ is norm-bounded by a known constant $\eta \in (0, +\infty)$, i.e.,

$$\|\boldsymbol{\epsilon}\|_p \le \eta, \tag{3}$$

for some $p \in \{1, 2, +\infty\}$. The intercept $\alpha \in \mathbb{R}$ and slope $\beta \in \mathbb{R}$ of the demand function in (2), hereafter referred to as *parameters* of the demand curve, are chosen by "nature" when the product is devised. Throughout the selling horizon, α and β will likely remain *unknown* to the seller, implying that the residual demand $\epsilon_t \in \mathbb{R}$, $t \in \overline{\mathcal{T}}$, will be *unobservable*. The seller will nevertheless be able to gain information (learn) about α and β as a byproduct of his pricing decisions.

We now formalize our assumptions regarding the model and the data.

(A1) The postulated model (2)–(3) is valid over the feasible price set, with p and η known.

(A2) The set $\{(\alpha, \beta) \in \mathbb{R}^2 \ : \ \|(d_t - \alpha - \beta p_t)_{t \in \mathcal{H}}\|_p \le \eta\}$ is bounded and has non-empty interior, i.e., there exist $(\hat{\alpha}, \hat{\beta}) \in \mathbb{R}^2$ and $\rho > 0$ such that $\|(d_t - \hat{\alpha} - \hat{\beta} p_t)_{t \in \mathcal{H}}\|_p \le \eta - \rho$.

Several comments are in order. First, the restriction to linear demand curves may seem stringent. Indeed, the demand curve for virtually any product is generally nonlinear, converging to zero for large price values, does not take-on negative values and results in revenue functions $((\alpha + \beta p_t) p_t)$ spanning $\mathbb{R}_+$ (see e.g., Talluri and van Ryzin (2004)). Nevertheless, in the context of the pricing

problem set forth above, most firms typically only allow their prices to vary in a moderately sized interval (constructed by taking into account competition, production costs, etc.) over which the demand curve can be well approximated by a linear function. As we are only concerned with the behavior of the demand function over this interval, we argue that a linear demand model is in fact adequate. We note that linear demand models are extremely popular in the revenue management literature due to their simplicity, see e.g., Mas-Colell et al. (1995) and Talluri and van Ryzin (2004). We emphasize that by "model validity" in Assumption **(A1)**, we do not necessarily mean that the model is actually linear with norm-bounded uncertainty, but rather that *no data will be seen during the selling season which invalidates the model*. Assumption **(A1)** implies that the set from Assumption **(A2)** is non-empty. Second, the choice of norm p should be guided both by the data (if sufficient samples are available) and the knowledge of the seller acquired from similar products. As a guideline, we would recommend using a 2-norm for residual errors that resemble a truncated normal and an ∞-norm in other cases. Third, we note that obtaining an accurate bound η may be challenging when little data is available. Nevertheless, we believe that it is easier to select a suitable value for η rather than to estimate an entire distribution for the behavior of $\boldsymbol{\epsilon}$. We note that a large value of η will always guarantee the validity of the model but may slow down learning. Finally, we argue that the requirement that the set in Assumption **(A2)** is bounded is non-restrictive since prices are under the control of the seller: it can always be enforced by experimenting with at least two distinct prices prior to the beginning of the planning horizon. We note that if this set has empty interior, the seller knows the demand curve precisely, and there is no need for exploration.

REMARK 1 (PRIOR INFORMATION). The model discussed above readily extends to the case when the seller has prior (data-independent) information on the possible set of values, $\Theta_{\mathrm{prior}} \subseteq \mathbb{R}^2$, taken on by (α, β), with Θ_{prior} a pth-order cone representable set (see e.g., Ben-Tal et al. (2009) for a definition). In such a case, it suffices to augment the postulated model in assumption **(A1)** with the requirement that $(\alpha, \beta) \in \Theta_{\mathrm{prior}}$ and relax assumption **(A2)** to require that the set $\{(\alpha, \beta) \in \Theta_{\mathrm{prior}} : \|(d_t - \alpha - \beta p_t)_{t \in \mathcal{H}}\|_p \leq \eta\}$ be compact. Thus, if Θ_{prior} is compact, no historical data is

12

Bertsimas and Vayanos: *Data-driven learning in dynamic pricing*
Article submitted to *Operations Research*; manuscript no. (Please, provide the manuscript number!)

needed. Prior information generally available to the seller includes knowledge of the signs of the demand curve parameters e.g., $\alpha > 0$ and/or $\beta \leq 0$.

REMARK 2 (EXTENSIONS). The pricing problem discussed in this section may seem deceptively simple. However, it possesses all the features necessary for us to describe our approach. We emphasize that our methodology remains applicable for a far broader class of pricing problems, including the classical newsvendor with pricing problem and the (multi-product) revenue management problem where demand may depend on current and even past prices of all products sold. It also naturally extends to the case of strategic customers and to a setting where the demand curve is time-varying. This is in sharp contrast to other dynamic pricing approaches which are tailored to a specific problem class. Moreover, we note that the restriction to linear curves can be lifted. An overview of these (and other) variants and extensions is provided in Section 6.

3. Pricing policies, set estimation and problem formulation

In the previous section, we described the problem faced by the seller, the demand dynamics and data available. In this section, we formalize the distribution-free dynamic pricing problem mathematically as an adaptive optimization problem with decision-dependent uncertainty set.

3.1. Information vector and pricing policies

The price for the product at time $t \in \mathcal{T}$ is selected at the beginning of period t *after* the history $(p_\tau, d_\tau)_{\tau=1}^{t-1}$ of price-demand pairs has been observed, but *before* future outcomes $(p_\tau, d_\tau)_{\tau \geq t}$ become available. In order to capture the ability of the retailer to exploit the information available at each stage, we model his pricing decisions as *functions* of the history of observations, and refer to this sequence of functional variables as a *pricing policy*. Formally, a pricing policy corresponds to a non-anticipative sequence $\boldsymbol{\pi} := (\pi_1, \ldots, \pi_T)$, where each π_t, $t \in \mathcal{T}$, is a measurable function from $\mathbb{R}^{2(t-1)}$ to $[l, u]$ that maps historical observations to admissible prices. We henceforth denote by π_t the pricing policy for time t and by p_t the price realization, i.e.,

$$p_t = \pi_t(p_1, d_1, \ldots, p_{t-1}, d_{t-1}) \quad \forall t \in \mathcal{T}.$$

We remark that since data prior to $t = 1$ is known, the pricing policy π_1 is in fact a constant. We define the set

$$\mathcal{N} := \prod_{t \in \mathcal{T}} [l, u]^{\left(\mathbb{R}^{2(t-1)} \right)}$$

that corresponds to all non-anticipative pricing policies taking values in $[l, u]^T$.

REMARK 3 (CLASSICAL ADAPTIVE OPTIMIZATION). In classical adaptive optimization (see e.g., Ben-Tal et al. (2009) and Bertsimas et al. (2011a)), the decisions are modeled as functions of the primitive exogenous uncertainties in the problem. This is either possible because these exogenous uncertainties are directly observable (Ben-Tal et al. (2004)) or made possible by the so-called "purified-output" approach (Ben-Tal et al. (2006)). This latter methodology infers the (information generated by the) primitive uncertainties from an observable output that depends linearly on the controls and the uncertainties. In the context of the dynamic pricing problem under consideration, the residual demand (primitive uncertainty) is not observable and cannot be inferred from the observed demand (since α and β are unknown). This will prevent us from following classical adaptive optimization approaches for solving the dynamic pricing problem with demand learning.

We now describe a methodology for constructing data-driven uncertainty sets for dynamic pricing that adapt to the pricing decisions, thus enabling the seller to explore the demand curve.

3.2. Data-driven adaptive set estimation

Set estimation for the demand curve parameters. The model of uncertainty introduced in Section 2.2 does not assume any statistical knowledge of the error. Under these circumstances, it is natural to define *a set* of possible demand curve parameters. We propose to construct this *prior uncertainty set* as the union of all estimates $(\alpha, \beta) \in \mathbb{R}^2$ that are *compatible* with the model and *unfalsified* by the historical data. This set has maximal "size" when residual errors for the future $(t \geq 1)$ are restricted to be equally zero and is representable as

$$\Theta := \left\{ (\alpha, \beta) \in \mathbb{R}^2 : d_t = \alpha + \beta p_t + \epsilon_t \ \forall t \in \mathcal{H}, \ \|(\epsilon_t)_{t \in \mathcal{H}}\|_p \leq \eta \right\}. \tag{4}$$

14

Bertsimas and Vayanos: *Data-driven learning in dynamic pricing*
Article submitted to *Operations Research*; manuscript no. (Please, provide the manuscript number!)

We emphasize that d_t and p_t in the representation (4) are historical data points, i.e., fixed values. The set (4) thus corresponds to the set of all estimates for (α, β) that are not invalidated from the historical data $(p_t, d_t)_{t \in \mathcal{H}}$, and the noise bound η. We now present two examples of data-driven uncertainty set constructions and discuss their relationships to the popular least-squares estimate.

EXAMPLE 1 (EUCLIDEAN NORM-BOUND ERROR). When $p = 2$, the set in (4) is representable as a single ellipsoid

$$\left\{ (\alpha, \beta) \in \mathbb{R}^2 \; : \; \begin{bmatrix} \alpha - \hat{\alpha} \\ \beta - \hat{\beta} \end{bmatrix}^\top \begin{bmatrix} H & \sum_{t \in \mathcal{H}} p_t \\ \sum_{t \in \mathcal{H}} p_t & \sum_{t \in \mathcal{H}} p_t^2 \end{bmatrix} \begin{bmatrix} \alpha - \hat{\alpha} \\ \beta - \hat{\beta} \end{bmatrix} \leq \eta \right\}$$

centered at the least-square estimate

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} H & \sum_{t \in \mathcal{H}} p_t \\ \sum_{t \in \mathcal{H}} p_t & \sum_{t \in \mathcal{H}} p_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t \in \mathcal{H}} d_t \\ \sum_{t \in \mathcal{H}} p_t d_t \end{bmatrix}$$

for (α, β). Note that since there is some variability in the historical prices (and $H \geq 2$), see Assumption **(A2)**, the matrix

$$\begin{bmatrix} H & \sum_{t \in \mathcal{H}} p_t \\ \sum_{t \in \mathcal{H}} p_t & \sum_{t \in \mathcal{H}} p_t^2 \end{bmatrix}$$

has full rank and is thus invertible. The center and semi-axes of this ellipsoid coincide with those of the ellipsoid arising as a confidence region from least-squares estimation. A concrete example of such data-driven uncertainty set is shown on Figure 1.

EXAMPLE 2 (INFINITY NORM-BOUND ERROR). When $p = \infty$, the set in (4) is expressible as the intersection of finitely many "stripes"

$$\bigcap_{t \in \mathcal{H}} \{ (\alpha, \beta) \in \mathbb{R}^2 \; : \; |d_t - \alpha - \beta p_t| \leq \eta \;\; \forall t \in \mathcal{H} \}.$$

The historical prices chosen determine the orientations of the stripes in (α, β) space, while the realized demands determine the locations of the stripes. The parameter η corresponds to the half-width of the stripes. In accordance with our intuition, the larger the parameter η, the greater the variability of the residual error and the lesser the information provided by each historical measurement.
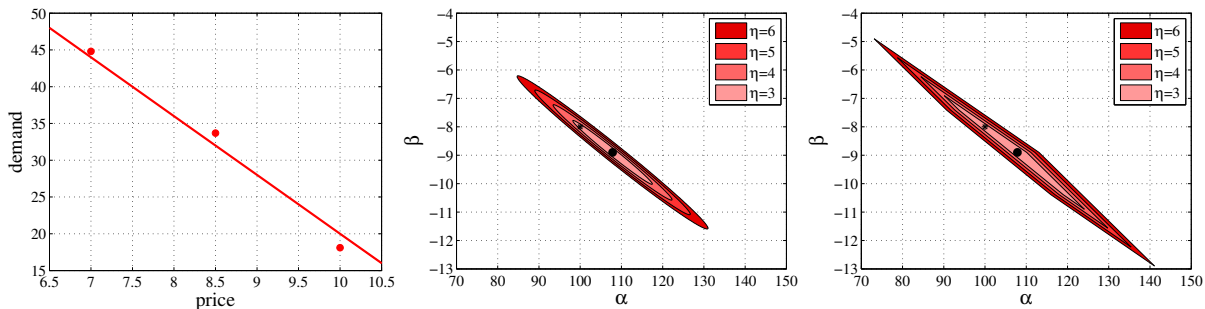
Figure 1   Companion figure for Examples 1 and 2. The figure on the left shows the true demand curve ($\alpha =$ 100, $\beta = -8$) and three historical price-demand pairs obtained by sampling the residual error from the uniform distribution with support $[-2, 2]$. The other two figures show the associated data-driven uncertainty sets for the case when the Euclidean norm (middle) and infinity norm (right) of the error is bounded by η varying from 3 to 6. The latter two figures were obtained using Yalmip, see Löfberg (2004). As the two-norm of the sampled errors is less than 3, none of the sets are empty. The dot and the asterisk on the two figures on the right denote the least-squares estimate of the demand curve parameters and their true value, respectively.

As there is at least some variability in the historical prices (and $H \geq 2$), see Assumption **(A2)**, this set is bounded. We note that the least-square estimate does not necessarily lie in this set. A concrete example of such data-driven uncertainty set is shown on Figure 1.

Adaptive set estimation for the price-demand realizations. In the absence of statistical assumptions on the residual error, we set out to construct *a set* of possible price-demand realizations for the entire planning horizon. As the demand at each stage $t \in \mathcal{T}$ is an unknown function of the price (which itself depends on historical observations), this *uncertainty set* is naturally *policy-dependent*. To the best of our knowledge, uncertainty sets of this type have not previously appeared in the literature. Given a pricing policy $\boldsymbol{\pi} \in \mathcal{N}$ and the uncertainty set (4) for the demand curve parameters, the vector of price-demand realizations must be member of

$$
\mathcal{U}(\boldsymbol{\pi}) := \left\{
\begin{array}{l}
(p_1, \ldots, p_T, d_1, \ldots, d_T) \in \mathbb{R}^{2T} \, : \, d_t = \alpha + \beta p_t + \epsilon_t \;\; \forall t \in \overline{\mathcal{T}}, \\[2mm]
p_t = \pi_t(p_1, d_1, \ldots, p_{t-1}, d_{t-1}) \;\; \forall t \in \mathcal{T}, \, \|\boldsymbol{\epsilon}\|_p \leq \eta
\end{array}
\right\}. \tag{5}
$$

Several comments are in order. First, part of the equality constraints defining the demand ($t \in \mathcal{H}$) in the description of $\mathcal{U}(\boldsymbol{\pi})$ are decision *in*dependent: they only depend on the historical data

16

Bertsimas and Vayanos: *Data-driven learning in dynamic pricing*
Article submitted to *Operations Research*; manuscript no. (Please, provide the manuscript number!)

$(d_t, p_t)_{t \in \mathcal{H}}$. Second, the arguments of the set $\mathcal{U}(\cdot)$ are pricing policies, i.e., functions. Thus, $\mathcal{U}(\boldsymbol{\pi})$ maps adaptive decision variables to possible price-demand paths. By varying the prices chosen, the seller is able to explore various regions of the uncertainty set by pruning out, "on-the-fly", regions that are not compatible with (would falsify) the observations. It is thus apparent that by allowing the uncertainty set to adapt to the pricing decisions, we precisely capture the learning ability of the decision-maker. We note that modeling exploration (adaptive uncertainty set) is only pertinent if we also model exploitation (adaptive decision variables, see Section 3.1): if the seller cannot exploit the information dynamically acquired, there can be no benefit in exploring.

From the representation (5), we observe that, for a fixed pricing policy $\boldsymbol{\pi}$, the *posterior uncertainty set* for the demand curve parameters at stage t can be obtained as the projection of $\mathcal{U}(\boldsymbol{\pi})$ onto the space of (α, β) uncertainties for $(p_\tau, d_\tau)_{\tau=1}^{t-1}$ fixed to their realization. Thus, by observing the price-demand realization up to the end of stage $t-1$, the retailer can fully characterize the posterior uncertainty set. We note that similarly to the prior uncertainty set, the posterior uncertainty set is p^{th}-order cone representable. Moreover, it is always a subset of the prior uncertainty set. Figure 2 illustrates, by means of an example, how the chosen pricing policy affects the information acquired and the "shape" of the posterior uncertainty set for (α, β).

REMARK 4 (RELATIONSHIP TO SET MEMBERSHIP ESTIMATORS). The approach we have proposed for constructing data-driven uncertainty sets for the demand curve parameters is identical to the set membership estimators originally proposed in the late 1960s by Schweppe (1968) and Witsenhausen (1968) in the systems identification literature (see also Milanese and Vicino (1991) and Giarré et al. (1997)). Nevertheless, our proposal for constructing data-driven uncertainty sets for the demand goes a significant step further as it extends the set membership construction into the future (through decision-dependence) to capture the learning (exploration) ability of the decision-maker. To the best of our knowledge, this has not been attempted in the literature.

3.3. Problem formulation: exploration/exploitation trade-off

In the previous section, we showed that, in the absence of distributional information, the exploration ability of the decision-maker can be naturally modeled by allowing the set of possible price-demand
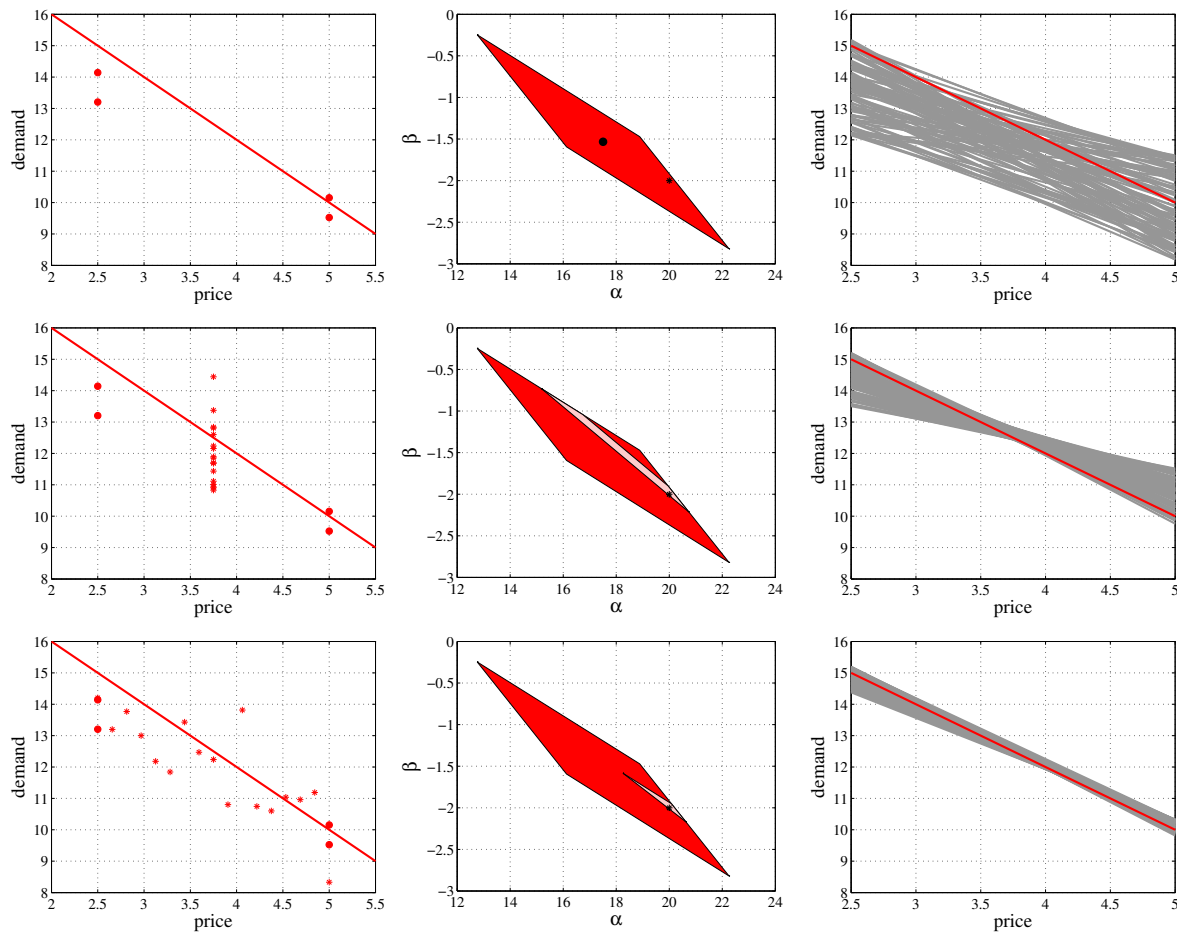
Figure 2 Comparison of the posterior uncertainty sets for (α, β) for two qualitatively different static pricing policies. The figures on the left show the true demand curve ($\alpha = 20$, $\beta = -2$), the historical price-demand pairs (dots) and 15 future price-demand pairs (asterisks) for two different static pricing policies (middle and bottom rows) for the same residual error realizations. The figures in the middle column plot the prior uncertainty set (dark area) and the posterior uncertainty set (light area) built from the price-demand realizations in the corresponding row. Note that the true demand curve parameter pair (asterisk) is always contained in both the prior and posterior uncertainty sets. The figures on the right illustrate the true demand curve (dark line) and one hundred demand curves (light lines) drawn uniformly at random from the prior (top row) and posterior (middle and bottom row) uncertainty sets. The two posterior uncertainty sets are qualitatively very different: In the first case (constant policy at $3.75) the posterior uncertainty set is a thin stripe intersected with the prior set (a single point on the demand curve is well identified); In the second case (uniform policy in the range $[2.5, 5]$), the posterior uncertainty set closely surrounds the true demand curve parameters (the demand curve as a whole is well identified). Thus, depending on the variability of the policy chosen, the type of information acquired and the associated posterior uncertainty sets obtained change substantially.

realizations to *adapt* to the pricing policy. We also discussed that his exploitation capacity can be captured by allowing the decisions to adapt to the history of observations. In this section, we formulate mathematically the problem faced by the seller and show that it intrinsically captures the trade-off between exploration and exploitation.

Since the seller has no statistical knowledge on the residual demand distribution, but only deterministic (bounded norm) information, it is natural that he be immunized against all possible realizations of the price-demand pairs in the uncertainty set (5). We recall that this set corresponds to all possible price-demand realizations given demand curve parameters compliant with the data and a chosen price policy. Thus, the objective of the seller is to maximize the profit function (1) in the worst-case realization of $(\boldsymbol{p}, \boldsymbol{d}) := (p_1, \ldots, p_T, d_1, \ldots, d_T)$, in the policy-dependent set $\mathcal{U}(\boldsymbol{\pi})$. In mathematical terms, we write

$$\underset{\boldsymbol{\pi} \in \mathcal{N}}{\text{maximize}} \quad \underset{(\boldsymbol{p}, \boldsymbol{d}) \in \mathcal{U}(\boldsymbol{\pi})}{\text{infimum}} \quad \boldsymbol{d}^\top \boldsymbol{p} - \max \left\{ h \left(c - \mathbf{e}^\top \boldsymbol{d} \right), b \left(\mathbf{e}^\top \boldsymbol{d} - c \right) \right\}. \tag{$\mathcal{DP}$}$$

Problem $\mathcal{DP}$ inherently captures the trade-off between exploration and exploitation. On the one hand, by choosing a pricing policy that maximizes his immediate revenue (based on available knowledge), the seller may not adequately prune-out regions of the membership set $\mathcal{U}(\boldsymbol{\pi})$, thus incurring large opportunity losses in the future. On the other hand, by choosing prices with sole aim to explore the membership set, significant short-term revenue losses may ensue that are not counterbalanced by the possibility of future income. Note that the optimal objective value of $\mathcal{DP}$ corresponds to the *guaranteed* profit of the seller.

4. Insights: when does the seller benefit from exploration?

In Section 3, we proposed a mathematical formulation for the distribution-free dynamic pricing problem that incorporates learning of the unknown demand curve parameters. In this section, we characterize cases when the seller does not benefit from exploration and show that in the general setting, learning is in fact beneficial. For this purpose, we introduce the following definition.

DEFINITION 1 (STATIC PRICING POLICY). A pricing policy $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_T) \in \mathcal{N}$ is called static if it satisfies

$$\pi_t(\hat{p}_1, \hat{d}_1, \ldots, \hat{p}_{t-1}, \hat{d}_{t-1}) = \pi_t(\tilde{p}_1, \tilde{d}_1, \ldots, \tilde{p}_{t-1}, \tilde{d}_{t-1})$$

for all $(\hat{p}_1, \hat{d}_1, \ldots, \hat{p}_{t-1}, \hat{d}_{t-1})$ and $(\tilde{p}_1, \tilde{d}_1, \ldots, \tilde{p}_{t-1}, \tilde{d}_{t-1})$ in $\mathbb{R}^{2(t-1)}$ and for each $t \in \mathcal{T} \backslash \{1\}$.

From the formulation $\mathcal{DP}$, it becomes apparent that exploration (demand learning) is beneficial (improves the objective value of $\mathcal{DP}$) if and only if there does not exist a static pricing strategy that is optimal in $\mathcal{DP}$. Indeed, if there exists a static pricing strategy that is optimal in $\mathcal{DP}$, exploring regions of the uncertainty set cannot help improve the objective value of $\mathcal{DP}$ since there is no benefit in exploiting that information through adaptation. On the other hand, if there does not exist an optimal static pricing strategy for $\mathcal{DP}$, this implies that pruning out regions of the uncertainty set that are incompatible with additional observations is necessary to improve the objective of $\mathcal{DP}$, so that exploration is imperative.

REMARK 5 (PARETO EFFICIENCY). The existence of an optimal static pricing strategy for $\mathcal{DP}$ does not necessarily mean that there may not be an adaptive pricing strategy that performs better when scenarios other than the worst-case ones materialize (this is closely related to the idea of Pareto efficiency in classical robust optimization, see Iancu and Trichakis (2013)), neither does it imply that there may not be an adaptive pricing strategy that more efficiently prunes-out critical regions of the uncertainty set. It only implies that in order to achieve the highest profit in the worst-case realization of the uncertain parameters, there is no need to explore the uncertainty set.

We are now ready to investigate instances of Problem $\mathcal{DP}$ for which exploration is not beneficial.

4.1. A case when there is no benefit in exploration

The following proposition shows that when $h = b = 0$ (revenue maximization problem), the seller does not benefit from exploration. A proof is provided in the Electronic Companion EC.1.

PROPOSITION 1 (**No benefit in exploration when** $h = b = 0$). *Suppose that the holding and backlogging costs are null (i.e., $h = b = 0$) and $\beta \leq 0 \ \forall (\alpha, \beta) \in \Theta$. Then, there exists a static pricing policy which is optimal in $\mathcal{DP}$. Moreover, such a pricing policy can be obtained by solving a (classical) robust optimization problem.*

4.2. In general, exploration is beneficial

In the previous section, we characterized instances of $\mathcal{DP}$ when exploration is not beneficial. We now show that in the general setting, the seller can significantly benefit from demand learning. A proof of the statement is provided in the Electronic Companion EC.2.

PROPOSITION 2 (**In general, exploration is beneficial**). *If the holding and overbooking costs are not equally zero ($h, b \neq 0$), then there does not necessarily exist an optimal static pricing strategy for the dynamic pricing problem $\mathcal{DP}$, even if $\beta \leq 0 \ \forall (\alpha, \beta) \in \Theta$.*

5. Proposed solution approach

In Sections 2 and 3, we proposed a novel data-driven and distribution-free modeling paradigm for demand learning in dynamic pricing that unifies optimization and estimation. The resulting dynamic pricing problem $\mathcal{DP}$ constitutes an *adaptive generalized semi-infinite optimization problem* that is severely computationally intractable. First, it possesses a large number of time-periods: indeed, despite a persistent reduction of product life-cycles, most products still have shelf lives spanning several months with pricing decisions taken daily. Second, it optimizes over functional decisions. Finally, it possesses, once expressed in epigraph form, an infinite number of constraints enforced over decision-dependent sets. In Section 4, we demonstrated that exploration is generally imperative, implying that it is necessary for the seller to optimize over adaptive policies. In this section, we propose two successive approximations that reduce the sizes of the information and decision spaces while preserving the ability of the seller to adapt, and show that the resulting problem can be formulated as a mixed-binary conic program.

5.1. Approximations

Stage aggregation. Problem $\mathcal{DP}$ presents a large number of time-periods, translating to a large number of observable uncertain parameters. As a first step to achieve computational tractability, we propose to reduce the information space in the problem by drastically decreasing the number of observable uncertainties. To mitigate the loss of optimality incurred by this approximation, we will judiciously construct a new vector of observable uncertainties as a function of the primitive uncertainties $(p_1, d_1, \ldots, p_T, d_T)$ that accurately summarizes the information dynamically acquired.

Concretely speaking, we propose to aggregate the time-periods in the problem, henceforth referred to as *micro-periods*, to fewer *macro-periods*. We denote the set of macro-periods by $\mathcal{M} := \{1, \ldots, M\}$ and let $\mathcal{T}_m$ denote the set of all micro-periods in macro-period $m \in \mathcal{M}$. We require that each $\mathcal{T}_m$ consists of consecutive time-periods and $\cup_{m \in \mathcal{M}} \mathcal{T}_m = \mathcal{T}$. We let $b_m := \min_{t \in \mathcal{T}_m} t$ and $e_m := \max_{t \in \mathcal{T}_m} t$ denote the first and last micro-periods in macro-period m, respectively. We define $\mathcal{T}^m := \cup_{\mu=1}^m \mathcal{T}_\mu$ and $\overline{\mathcal{T}}^m := \mathcal{H} \cup \mathcal{T}^m$. We assume that the price-demand pairs (p_t, d_t) are no longer observed within each micro-period $t \in \mathcal{T}$. Instead, we only observe, at the beginning of macro-period $m \in \mathcal{M}$, certain functions of the history of observations $(p_t, d_t)_{t \in \overline{\mathcal{T}}^{m-1}}$ that concisely summarize the state of the retailer's knowledge. We denote these observable uncertainties by $\boldsymbol{\xi}_m \in \mathbb{R}^{n_\xi}$ and require that they be the *unique* solutions to a system of equations of the form

$$\boldsymbol{R}_m(\boldsymbol{p})\boldsymbol{\xi}_m = \boldsymbol{W}_m(\boldsymbol{p})\boldsymbol{d} + \boldsymbol{r}_m(\boldsymbol{p}),$$

for some matrices $\boldsymbol{R}_m(\boldsymbol{p}) \in \mathbb{R}^{n_\xi \times n_\xi}$, $\boldsymbol{W}_m(\boldsymbol{p}) \in \mathbb{R}^{n_\xi \times T}$ and $\boldsymbol{r}_m(\boldsymbol{p}) \in \mathbb{R}^{n_\xi}$. We refer to $\boldsymbol{\xi}_m$ as the *reduced information vector* for macro period m. In order to ensure that $\boldsymbol{\xi}_m$ be observable at the beginning of macro-period m, we require that $\boldsymbol{R}_m(\boldsymbol{p})$ be invertible for all $\boldsymbol{p} \in [l, u]^T$, that $\boldsymbol{W}_m(\boldsymbol{p})$ have non-zero entries in the first e_{m-1} columns only and that $\boldsymbol{R}_m(\boldsymbol{p})$, $\boldsymbol{W}_m(\boldsymbol{p})$ and $\boldsymbol{r}_m(\boldsymbol{p})$ all be constant in the prices $p_{b_m}, \ldots, p_T$. Finally, we require that $\boldsymbol{R}_m(\boldsymbol{p})$ and $\boldsymbol{W}_m(\boldsymbol{p})$ be quadratic separable in their arguments, thus being expressible as

$$\boldsymbol{R}_m(\boldsymbol{p}) = \boldsymbol{R}_m + \sum_{t=1}^{e_{m-1}} \boldsymbol{R}_{m,t}^1 p_t + \boldsymbol{R}_{m,t}^2 p_t^2, \quad \boldsymbol{W}_m(\boldsymbol{p}) = \boldsymbol{W}_m + \sum_{t=1}^{e_{m-1}} \boldsymbol{W}_{m,t}^1 p_t + \boldsymbol{W}_{m,t}^2 p_t^2$$

22

Bertsimas and Vayanos: *Data-driven learning in dynamic pricing*
Article submitted to *Operations Research*; manuscript no. (Please, provide the manuscript number!)

and

$$\boldsymbol{r}_m(\boldsymbol{p}) = \boldsymbol{r}_m + \sum_{t=1}^{e_{m-1}} \boldsymbol{r}_{m,t}^1 p_t + \boldsymbol{r}_{m,t}^2 p_t^2$$

for some matrices $\boldsymbol{R}_m$, $\boldsymbol{R}_{m,t}^1$, $\boldsymbol{R}_{m,t}^2 \in \mathbb{R}^{n_\xi \times n_\xi}$ and $\boldsymbol{W}_m$, $\boldsymbol{W}_{m,t}^1$, $\boldsymbol{W}_{m,t}^2 \in \mathbb{R}^{n_\xi \times T}$ and vectors $\boldsymbol{r}_m$, $\boldsymbol{r}_{m,t}^1$, $\boldsymbol{r}_{m,t}^2 \in \mathbb{R}^{n_\xi}$, $t \in \mathcal{T}_m$, $m \in \mathcal{M}$.

REMARK 6. As will become clear later on, we can in fact relax this last assumption and let $\boldsymbol{R}_m(\boldsymbol{p})$, $\boldsymbol{W}_m(\boldsymbol{p})$ and $\boldsymbol{r}_m(\boldsymbol{p})$ be polynomial (not necessarily separable) in they arguments. As this assumption does not influence the further derivations and simplifies notation, we enforce it here.

EXAMPLE 3 (LEAST-SQUARES ESTIMATES AND CUMULATIVE DEMAND). A natural choice for $\boldsymbol{\xi}_m$ is to let $(\xi_{m,1}, \xi_{m,2})$ denote the least-square estimates for (α, β) as of the end of the last micro-period in macro-period $m-1$ and define $\xi_{m,3}$ as the cumulative demand incurred up until then. Thus, $\boldsymbol{\xi}_m := (\xi_{m,1}, \xi_{m,2}, \xi_{m,3})$ solves the system of equations

$$\begin{bmatrix} |\overline{\mathcal{T}}^{m-1}| & \sum_{t \in \overline{\mathcal{T}}^{m-1}} p_t & 0 \\ \sum_{t \in \overline{\mathcal{T}}^{m-1}} p_t & \sum_{t \in \overline{\mathcal{T}}^{m-1}} p_t^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \xi_{m,1} \\ \xi_{m,2} \\ \xi_{m,3} \end{bmatrix} = \begin{bmatrix} \sum_{t \in \overline{\mathcal{T}}^{m-1}} d_t \\ \sum_{t \in \overline{\mathcal{T}}^{m-1}} p_t d_t \\ \sum_{t \in \overline{\mathcal{T}}^{m-1}} d_t \end{bmatrix}.$$

We remark that Assumption **(A2)** implies that the matrix on the left above is invertible. This definition of $\boldsymbol{\xi}_m$ thus satisfies all of the observability requirements.

REMARK 7 (IMPACT OF THE REDUCTION IN INFORMATION SPACE). By observing the reduced information vector at the beginning of each macro-period rather than observing the new price-demand realizations each time they are made available, the seller can effectively only compute an outer approximation to the true posterior uncertainty set. This posterior uncertainty set is only updated (inside the optimization) at the beginning of each macro-period rather that at each stage.

There is no incentive to change pricing strategy when no new information is revealed. We thus propose to (adaptively) select, at the beginning of each macro-period, one of finitely many *candidate pricing strategies* $\kappa \in \mathbb{K} := \{1, \ldots, K\}$, each of which takes on the value $\pi_{\kappa,t} \in [l, u]$ at time $t \in \mathcal{T}$. The pricing strategy selected at the beginning of macro-period $m \in \mathcal{M}$ prevails throughout

the macro-period. Concretely, we introduce for each $m \in \mathcal{M}$ an *adaptive* vector of coefficients $\boldsymbol{z}_m$ whose κ^{th} element is 1 if and only if the κth strategy is selected at the beginning of macro-period m. We model $\boldsymbol{z}_m$ as a non-anticipative measurable function from $\mathbb{R}^{Mn_\xi}$ to $\{0,1\}^K$ which maps observable uncertainties $\boldsymbol{\xi} := (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M)$ to choices of pricing policies. We write the non-anticipativity constraints explicitly as

$$\boldsymbol{z}_m(\boldsymbol{\xi}) = \boldsymbol{z}_m(\boldsymbol{\xi}') \quad \forall \boldsymbol{\xi}, \, \boldsymbol{\xi}' : \boldsymbol{O}_m \boldsymbol{\xi} = \boldsymbol{O}_m \boldsymbol{\xi}',$$

where $\boldsymbol{O}_m$ is the projection operator which maps $\boldsymbol{\xi}$ to the portion of uncertain parameters that are observable at the beginning of macro-period m, i.e., $(\boldsymbol{\xi}_\mu)_{\mu=1}^m$. We require that only one candidate strategy be chosen during each macro-period, i.e.,

$$\mathbf{e}^\top \boldsymbol{z}_m(\boldsymbol{\xi}) = 1 \quad \text{for all} \quad \boldsymbol{\xi} \in \mathbb{R}^{Mn_\xi}.$$

The price selected by the seller at time $t \in \mathcal{T}$ is then expressible as

$$\pi_t(p_1, d_1, \ldots, p_{t-1}, d_{t-1}) = \boldsymbol{\pi}_t^\top \boldsymbol{z}_m(\boldsymbol{\xi}) \quad \forall t \in \mathcal{T}_m, \, m \in \mathcal{M},$$

where $\boldsymbol{\pi}_t := (\pi_{t,\kappa})_{\kappa \in \mathbb{K}}$. We remark that the choice of candidate strategy at macro-period $m \in \mathcal{M} \backslash \{1\}$ is *adaptive*, whereas the choice of candidate strategy at macro-period $m = 1$ is a constant.

We now provide two examples of candidate strategies.

EXAMPLE 4 (CONSTANT PRICING STRATEGY). A natural choice for the candidate pricing strategies is to let $\pi_{t,\kappa} = \pi_\kappa$ for some $\pi_\kappa \in [l, u]$ and for all $t \in \mathcal{T}$. Constant strategies have the drawback that they do not favor exploration of the demand curve a whole. Instead, they yield significant information on a specific point of the demand function. In terms of uncertainty set for the demand curve parameters, entering the same price π_κ during multiple micro-periods progressively reduces the set of estimates to a line segment in (α, β) space, with the slope of the line depending on the chosen value π_κ. A significant advantage of constant strategies is that price adjustments become infrequent (they occur at most at the beginning of each macro-period), thus moderating possible customer discontent. They are commonly employed by retailers who periodically perform discounts and must choose the best discount level to optimally balance exploration and exploitation (e.g., 40% discount from full price versus 20% discount from full price).

EXAMPLE 5 (RANDOMIZED PRICING STRATEGY). Another possibility is to allow the seller to choose among randomized pricing strategies whereby, for each $\kappa \in \mathbb{K}$, $\pi_{t,\kappa}$ is chosen at random from a strategy dependent interval $\mathcal{I}_\kappa \subset [l, u]$ (prior to optimization). Such strategies have the benefit of promoting exploration of the uncertainty set for the parameters. Nevertheless, they may elevate customer discontent due to the high frequency of price adjustments. Randomized strategies are particularly suited to cases when each customer cannot view the prices offered to other customers (e.g., interest rate quotes, online travel and hospitality bookings, and online retail).

REMARK 8 (DISCRETE PRICES). In most realistic settings, the admissible price set of the dynamic pricing problem is discrete and finite. The proposed solution approach lends itself particularly well to this case. In fact, in Section 8, we will investigate problems with discrete feasible price sets, as we believe that this is the type of problem of interest to practitioners.

We are now ready to formulate a conservative approximation to $\mathcal{DP}$ following the proposed reduction of the information space. It reads

$$
\begin{aligned}
&\text{maximize} \ \inf_{(\boldsymbol{p},\boldsymbol{d}) \in \mathcal{U}^{\mathrm{a}}(\boldsymbol{z})} \ \boldsymbol{d}^\top \boldsymbol{p} - \max\left\{ h\left(c - \mathbf{e}^\top \boldsymbol{d}\right), b\left(\mathbf{e}^\top \boldsymbol{d} - c\right) \right\} \\
&\text{subject to} \ \boldsymbol{z}_m \in \left(\{0,1\}^K\right)^{\mathbb{R}^{Mn_\xi}} \quad \forall m \in \mathcal{M} \\
&\left. \begin{array}{l}
\boldsymbol{z}_m(\boldsymbol{\xi}) = \boldsymbol{z}_m(\boldsymbol{\xi}') \quad \forall \boldsymbol{\xi}, \boldsymbol{\xi}' \in \mathbb{R}^{Mn_\xi} : \boldsymbol{O}_m \boldsymbol{\xi} = \boldsymbol{O}_m \boldsymbol{\xi}' \\[2mm]
\mathbf{e}^\top \boldsymbol{z}_m(\boldsymbol{\xi}) = 1 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^{Mn_\xi}
\end{array} \right\} \quad \forall m \in \mathcal{M},
\end{aligned}
\tag{$\mathcal{ADP}$}
$$

where

$$
\mathcal{U}^{\mathrm{a}}(\boldsymbol{z}) := \left\{
\begin{array}{l}
(p_1, \ldots, p_T, d_1, \ldots, d_T) \in \mathbb{R}^{2T} : d_t = \alpha + \beta p_t + \epsilon_t \ \forall t \in \overline{\mathcal{T}}, \ \|\boldsymbol{\epsilon}\|_p \le \eta, \\[2mm]
p_t = \boldsymbol{\pi}_t^\top \boldsymbol{z}_m(\boldsymbol{\xi}) \ \forall t \in \mathcal{T}_m, \ \boldsymbol{R}_m(\boldsymbol{p})\boldsymbol{\xi}_m = \boldsymbol{W}_m(\boldsymbol{p})\boldsymbol{d} + \boldsymbol{r}_m(\boldsymbol{p}) \quad \forall m \in \mathcal{M}
\end{array}
\right\}.
$$

The pricing problem $\mathcal{ADP}$ arising from the time-period aggregation corresponds to an adaptive optimization problem with decision-dependent uncertainty set which is solely affected by *binary adaptive* decision variables (the coefficients of the candidate strategies). As we will typically be able to choose $MK \ll T$ and $Mn_\xi \ll 2T$, the stage aggregation significantly reduces both the number

of adaptive decision variables and the number of *observable* uncertain parameters in the problem. Both of these reductions come at a moderate increase in the size of the representation of the uncertainty set. As we will see in the following section, the proposed reduction in the information space is crucial to achieve computational tractability.

Decision rule approximation. The time-period aggregation approximation enabled us to substantially reduce the size of the dynamic pricing problem. The aggregated problem $\mathcal{ADP}$ is still computationally intractable as it optimizes over binary *adaptive* (functional) decision variables and possesses, once expressed in epigraph form, a continuum of constraints enforced over a policy-dependent set. In this section, we propose to approximate these adaptive decision variables by functions that are piecewise constant over a preselected partition of the uncertainty set, in the spirit of Vayanos et al. (2011). In Section 5.3, we will propose several strategies for mitigating the potential loss of optimality incurred from this approximation.

REMARK 9 (FINITE ADAPTABILITY). Adaptive policies for (classical) two-stage robust mixed integer programs that do not necessitate a-priori partitioning of the uncertainty set have been proposed by Bertsimas and Caramanis (2010) and Hanasusanto et al. (2014). These papers investigate the so-called finite adaptability problem, wherein the decision maker pre-commits to a finite number of second-stage policies today and implements the best of these policies once the uncertain parameters are revealed. This type of policies offers a great deal of flexibility as the partitioning of the uncertainty set is left to the optimization and can take on an arbitrary form. On the other hand, it does not lend itself easily to our learning context where the uncertainty set is decision-dependent. We believe that extending the finite adaptability approximation to mixed-integer robust problems with decision-dependent uncertainty set is an interesting topic for future research.

Let $\Xi := \prod_{i=1}^{Mn_\xi} [\underline{\xi}_i, \overline{\xi}_i] \subset \mathbb{R}^{Mn_\xi}$ denote any hyperrectangular set containing the projection of the set $\bigcup_{\boldsymbol{z} \in \{0,1\}^{MK}} \mathcal{U}^{\mathrm{a}}(\boldsymbol{z})$ onto the space of *observable* uncertainties $\boldsymbol{\xi}$, i.e., satisfying

$$
\Xi \supset \left\{
\begin{array}{l}
\boldsymbol{\xi} \in \mathbb{R}^{Mn_\xi} \,:\, d_t = \alpha + \beta p_t + \epsilon_t \;\forall t \in \overline{\mathcal{T}}, \; \|\boldsymbol{\epsilon}\|_p \leq \eta, \\[2mm]
p_t = \boldsymbol{\pi}_t^\top \boldsymbol{z}_m \;\forall t \in \mathcal{T}_m, \; \boldsymbol{R}_m(\boldsymbol{p})\boldsymbol{\xi}_m = \boldsymbol{W}_m(\boldsymbol{p})\boldsymbol{d} + \boldsymbol{r}_m(\boldsymbol{p}) \quad \forall m \in \mathcal{M} \\[2mm]
\boldsymbol{z}_m \in \{0,1\}^K, \; \mathbf{e}^\top \boldsymbol{z}_m = 1 \;\forall m \in \mathcal{M}
\end{array}
\right\}, \tag{6}
$$

26

Bertsimas and Vayanos: *Data-driven learning in dynamic pricing*
Article submitted to *Operations Research*; manuscript no. (Please, provide the manuscript number!)

where $\boldsymbol{z}$ is here treated as a static variable left to the control of "nature".

REMARK 10 (COMPUTING Ξ). The tightest hyperrectangle Ξ satisfying (6) can be obtained by solving $2Mn_\xi$ mixed-binary conic optimization problems (linear programs if $p \in \{1, +\infty\}$). Each of these problems either maximizes or minimizes one of the elements of $\boldsymbol{\xi}$ subject to the constraints defining the set in the right-hand-side of (6). The decision variables of this problem constitute in all the uncertain parameters of the representation (6) (including the $\boldsymbol{z}_m$, $m \in \mathcal{M}$). The bilinear terms in the formulation can be eliminated by replacing the p_t, $t \in \mathcal{T}$, by their definition and subsequently linearizing the products of binary and (bounded) real-valued variables using standard big-M techniques.

We introduce a partition of Ξ into hyperrectangles of the form

$$\Xi_{\boldsymbol{s}} := \left\{ \boldsymbol{\xi} \in \Xi \, : \, w^i_{s_i-1} \le \xi_i < w^i_{s_i}, \, i = 1, \ldots, Mn_\xi \right\}, \tag{7}$$

where $\boldsymbol{s} \in \mathcal{S} := \prod_{i=1}^{Mn_\xi} \{1, \ldots, r_i\}$ and

$$w^i_0 < w^i_1 < \cdots < w^i_{r_i} \quad \text{for } i = 1, \ldots, Mn_\xi$$

represent $r_i + 1$ breakpoints along the axis of the ith observable uncertain parameter ξ_i. We approximate the coefficients of the candidate pricing strategies by functions that are piecewise constant on the subsets $\Xi_{\boldsymbol{s}}$, $\boldsymbol{s} \in \mathcal{S}$, and denote by $\boldsymbol{z}^{\boldsymbol{s}}_m \in \{0,1\}^K$, $m \in \mathcal{M}$, the vector value adopted by these coefficients on the $\boldsymbol{s}^{\text{th}}$ subset. Thus,

$$\boldsymbol{z}_m(\boldsymbol{\xi}) = \sum_{\boldsymbol{s} \in \mathcal{S}} \mathbb{1}_{\Xi_{\boldsymbol{s}}}(\boldsymbol{\xi}) \boldsymbol{z}^{\boldsymbol{s}}_m, \tag{8}$$

and the coefficients $\boldsymbol{z}^{\boldsymbol{s}}_m$ become the new decision variables in the problem. They must satisfy the non-anticipativity constraints

$$\boldsymbol{z}^{\boldsymbol{s}}_m = \boldsymbol{z}^{\boldsymbol{s}'}_m \text{ for all } \boldsymbol{s}, \, \boldsymbol{s}' \in \mathcal{S} \text{ such that } \boldsymbol{O}_m \boldsymbol{s} = \boldsymbol{O}_m \boldsymbol{s}', \, m \in \mathcal{M}. \tag{9}$$

For notational convenience, we define the block-diagonal matrix

$$\boldsymbol{P} := \text{blkdiag}\,(\boldsymbol{P}_1, \ldots, \boldsymbol{P}_M), \text{ with } \boldsymbol{P}_m := \begin{bmatrix} \boldsymbol{\pi}_{b_m} & \boldsymbol{\pi}_{b_m+1} & \cdots & \boldsymbol{\pi}_{e_m} \end{bmatrix}^\top.$$

This enables us to compactly express the price vector on the $\boldsymbol{s}^{\text{th}}$ subset as $\boldsymbol{p}^{\boldsymbol{s}} = \boldsymbol{P}\boldsymbol{z}^{\boldsymbol{s}}$. Following the reduction in the decision space of the problem, a conservative approximation (lower bound) to the aggregated dynamic pricing problem is expressible as

$$\text{maximize} \min_{\boldsymbol{s}\in\mathcal{S}} \inf_{\boldsymbol{d}\in\mathcal{U}_{\boldsymbol{s}}^{\text{a}}(\boldsymbol{z}^{\boldsymbol{s}})} \boldsymbol{d}^{\top}\boldsymbol{P}\boldsymbol{z}^{\boldsymbol{s}} - \max\left\{ h\left(c - \mathbf{e}^{\top}\boldsymbol{d}\right),\, b\left(\mathbf{e}^{\top}\boldsymbol{d} - c\right) \right\}$$

$$\text{subject to } \boldsymbol{z}_m^{\boldsymbol{s}} \in \{0,1\}^K,\ \mathbf{e}^{\top}\boldsymbol{z}_m^{\boldsymbol{s}} = 1 \quad \forall m \in \mathcal{M},\ \boldsymbol{s}\in\mathcal{S} \qquad (\mathcal{LADP})$$

$$\boldsymbol{z}_m^{\boldsymbol{s}} = \boldsymbol{z}_m^{\boldsymbol{s}'} \quad \forall \boldsymbol{s},\, \boldsymbol{s}'\in\mathcal{S} \ :\ \boldsymbol{O}_m\boldsymbol{s} = \boldsymbol{O}_m\boldsymbol{s}',\ m\in\mathcal{M},$$

where the uncertainty set for subset $\boldsymbol{s}\in\mathcal{S}$ is expressible as

$$\mathcal{U}_{\boldsymbol{s}}^{\text{a}}(\boldsymbol{z}) := \left\{ \begin{array}{l} (d_1,\ldots,d_T)\in\mathbb{R}^T \ :\ d_t = \alpha + \beta p_t + \epsilon_t \ \forall t\in\overline{\mathcal{T}},\ \|\boldsymbol{\epsilon}\|_p \leq \eta, \\[2mm] p_t = \boldsymbol{\pi}_t^{\top}\boldsymbol{z}_m \ \forall t\in\mathcal{T}_m,\ \boldsymbol{R}_m(\boldsymbol{p})\boldsymbol{\xi}_m = \boldsymbol{W}_m(\boldsymbol{p})\boldsymbol{d} + \boldsymbol{r}_m(\boldsymbol{p}),\ \boldsymbol{\xi}_m \in \text{cl}(\Xi_{\boldsymbol{s}}) \quad \forall m\in\mathcal{M} \end{array} \right\}.$$

Note that we have here replaced the set $\Xi_{\boldsymbol{s}}$ with its closure. This has left the formulation unchanged since, for any fixed $\boldsymbol{z}^{\boldsymbol{s}} \in \{0,1\}^K$ and $\boldsymbol{s}\in\mathcal{S}$, the function in the objective of $\mathcal{LADP}$ is continuous in $\boldsymbol{d}$. Problem $\mathcal{LADP}$ is a single-stage (static) robust optimization problem with decision-dependent uncertainty set. This type of problem is known as *generalized semi-infinite programming problem* in the literature, see e.g., Still (1999). It is generically severely computationally intractable. In the next section, we propose a methodology for reformulating $\mathcal{LADP}$ as a mixed-binary conic problem.

5.2. Reformulation

Before proceeding with the reformulation of $\mathcal{LADP}$ as a mixed-binary conic program, we make the following observations.

OBSERVATION 1. Fix $\boldsymbol{z} \in \{0,1\}^{MK}$ such that $\mathbf{e}^{\top}\boldsymbol{z}_m = 1 \ \forall m \in \mathcal{M}$. Then, there exists an $\boldsymbol{s}\in\mathcal{S}$ such that $\mathcal{U}_{\boldsymbol{s}}^{\text{a}}(\boldsymbol{z})$ has non-empty relative interior. Moreover, from Assumption **(A2)**, it follows that $\mathcal{U}_{\boldsymbol{s}}^{\text{a}}(\boldsymbol{z})$ is bounded for all $\boldsymbol{s}\in\mathcal{S}$.

OBSERVATION 2. For any fixed $\boldsymbol{z} \in \{0,1\}^{MK}$ such that $\mathbf{e}^{\top}\boldsymbol{z}_m = 1 \ \forall m \in \mathcal{M}$, the matrices $\boldsymbol{R}_m(\boldsymbol{P}\boldsymbol{z})$, $\boldsymbol{W}_m(\boldsymbol{P}\boldsymbol{z})$ and $\boldsymbol{r}_m(\boldsymbol{P}\boldsymbol{z})$ are linear in $\boldsymbol{z}$. Indeed, for any fixed $\mu\in\mathcal{M}$ and $t\in\mathcal{T}_{\mu}$, we have $(\boldsymbol{\pi}_t^{\top}\boldsymbol{z}_{\mu})^2 = (\boldsymbol{\pi}_t \circ \boldsymbol{\pi}_t)^{\top}\boldsymbol{z}_{\mu}$, where $\circ$ denotes the Hadamard operator.

OBSERVATION 3. *If we use the equality $p_t = \boldsymbol{\pi}^\top \boldsymbol{z}$ to eliminate the p_t, $t \in \mathcal{T}$, from the definition of*
$\mathcal{U}_{\boldsymbol{s}}^{\mathrm{a}}$, $\boldsymbol{s} \in \mathcal{S}$, *then the uncertainty set for subset $\boldsymbol{s}$ is expressible as*

$$\mathcal{U}_{\boldsymbol{s}}^{\mathrm{a}}(\boldsymbol{z}) = \left\{ \boldsymbol{d} \in \mathbb{R}^T \ : \ \exists \boldsymbol{\zeta} \in \mathbb{R}^{n_\zeta}, \ \boldsymbol{F}(\boldsymbol{z})\boldsymbol{d} + \boldsymbol{G}(\boldsymbol{z})\boldsymbol{\zeta} \geq \boldsymbol{g}_{\boldsymbol{s}}, \ \boldsymbol{H}\boldsymbol{\zeta} \succeq_{\mathcal{K}_p} \boldsymbol{h} \right\} \tag{10}$$

for some matrices $\boldsymbol{F}(\boldsymbol{z}) \in \mathbb{R}^{m_1 \times n_d}$, $\boldsymbol{G}(\boldsymbol{z}) \in \mathbb{R}^{m_1 \times n_\zeta}$, $\boldsymbol{g}_{\boldsymbol{s}} \in \mathbb{R}^{m_1}$, $\boldsymbol{H} \in \mathbb{R}^{m_2 \times n_\zeta}$ and $\boldsymbol{h} \in \mathbb{R}^{m_2}$, where
$m_1 := 2(H + T + 2Mn_\xi)$, $m_2 := H + T + 1$ *and* $n_\zeta := H + T + 2 + Mn_\xi$. *Moreover, it follows from*
Observation 2 that $\boldsymbol{F}(\boldsymbol{z})$ and $\boldsymbol{G}(\boldsymbol{z})$ are affine in their arguments, i.e., they are representable as
$\boldsymbol{F}(\boldsymbol{z}) = \boldsymbol{F}_0 + \sum_{i \in \mathcal{I}} \boldsymbol{F}_i z_i$ *and* $\boldsymbol{G}(\boldsymbol{z}) = \boldsymbol{G}_0 + \sum_{i \in \mathcal{I}} \boldsymbol{G}_i z_i$, *for some matrices $\boldsymbol{F}_i \in \mathbb{R}^{m_1 \times n_d}$ and $\boldsymbol{G}_i \in$*
$\mathbb{R}^{m_1 \times n_\zeta}$, $i \in \{0\} \cup \mathcal{I}$, $\mathcal{I} := \{1, \dots, MK\}$.

The following proposition shows that $\mathcal{LADP}$ can be reformulated as a mixed-binary conic optimization problem. The proof relies on techniques commonly employed in classical robust optimization (see e.g., Ben-Tal et al. (2009) and Bertsimas et al. (2011a)) combined with well known linearization techniques, see Electronic Companion EC.3. A similar reformulation approach is followed by Hanasusanto et al. (2014) to solve two-stage robust optimization problems under the finite-adaptability approximation approach, see Remark 9.

PROPOSITION 3 (**Conservative approximation to** $\mathcal{DP}$). *Consider the following mixed-binary conic optimization problem*

$$\text{maximize } v$$

subject to $v \in \mathbb{R}$, $\boldsymbol{z}_m^{\boldsymbol{s}} \in \{0,1\}^K$, $\boldsymbol{e}^\top \boldsymbol{z}_m^{\boldsymbol{s}} = 1 \quad \forall m \in \mathcal{M}, \ \boldsymbol{s} \in \mathcal{S}$

$$\left. \begin{aligned}
& \boldsymbol{\mu}_1^{\boldsymbol{s}}, \boldsymbol{\nu}_1^{\boldsymbol{s}} \in \mathbb{R}_+^{m_1}, \ \boldsymbol{\mu}_2^{\boldsymbol{s}}, \boldsymbol{\nu}_2^{\boldsymbol{s}} \in \mathcal{K}_q^{m_2}. \ \boldsymbol{x}_i^{\boldsymbol{s}}, \boldsymbol{y}_i^{\boldsymbol{s}} \in \mathbb{R}_+^{m_1} \quad \forall i \in \mathcal{I} \\
& \boldsymbol{g}_{\boldsymbol{s}}^\top \boldsymbol{\mu}_1^{\boldsymbol{s}} + \boldsymbol{h}^\top \boldsymbol{\mu}_2^{\boldsymbol{s}} \geq v + hc, \ \boldsymbol{g}_{\boldsymbol{s}}^\top \boldsymbol{\nu}_1^{\boldsymbol{s}} + \boldsymbol{h}^\top \boldsymbol{\nu}_2^{\boldsymbol{s}} \geq v - bc \\
& \boldsymbol{F}_0^\top \boldsymbol{\mu}_1^{\boldsymbol{s}} + \sum_{i \in \mathcal{I}} \boldsymbol{F}_i^\top \boldsymbol{x}_i^{\boldsymbol{s}} = \boldsymbol{P}\boldsymbol{z}^{\boldsymbol{s}} + h\boldsymbol{e}, \ \boldsymbol{G}_0^\top \boldsymbol{\mu}_1^{\boldsymbol{s}} + \sum_{i \in \mathcal{I}} \boldsymbol{G}_i^\top \boldsymbol{x}_i^{\boldsymbol{s}} + \boldsymbol{H}^\top \boldsymbol{\mu}_2^{\boldsymbol{s}} = \boldsymbol{0} \\
& \boldsymbol{F}_0^\top \boldsymbol{\nu}_1^{\boldsymbol{s}} + \sum_{i \in \mathcal{I}} \boldsymbol{F}_i^\top \boldsymbol{y}_i^{\boldsymbol{s}} = \boldsymbol{P}\boldsymbol{z}^{\boldsymbol{s}} - b\boldsymbol{e}, \ \boldsymbol{G}_0^\top \boldsymbol{y}_1^{\boldsymbol{s}} + \sum_{i \in \mathcal{I}} \boldsymbol{G}_i^\top \boldsymbol{y}_i^{\boldsymbol{s}} + \boldsymbol{H}^\top \boldsymbol{\nu}_2^{\boldsymbol{s}} = \boldsymbol{0} \\
& \left. \begin{aligned} & \boldsymbol{x}_i^{\boldsymbol{s}} \leq \boldsymbol{\mu}_1^{\boldsymbol{s}}, \ \boldsymbol{x}_i^{\boldsymbol{s}} \leq B z_i^{\boldsymbol{s}} \boldsymbol{e}, \ \boldsymbol{x}_i^{\boldsymbol{s}} \geq \boldsymbol{\mu}_1^{\boldsymbol{s}} - B(1 - z_i^{\boldsymbol{s}})\boldsymbol{e} \\ & \boldsymbol{y}_i^{\boldsymbol{s}} \leq \boldsymbol{\nu}_1^{\boldsymbol{s}}, \ \boldsymbol{y}_i^{\boldsymbol{s}} \leq B z_i^{\boldsymbol{s}} \boldsymbol{e}, \ \boldsymbol{y}_i^{\boldsymbol{s}} \geq \boldsymbol{\nu}_1^{\boldsymbol{s}} - B(1 - z_i^{\boldsymbol{s}})\boldsymbol{e} \end{aligned} \right\} \ \forall i \in \mathcal{I} \\
& \boldsymbol{z}_m^{\boldsymbol{s}} = \boldsymbol{z}_m^{\boldsymbol{s}'} \quad \forall \boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S} \ : \ \boldsymbol{O}_m \boldsymbol{s} = \boldsymbol{O}_m \boldsymbol{s}', \ m \in \mathcal{M},
\end{aligned} \right\} \forall \boldsymbol{s} \in \mathcal{S}$$

$$(\mathcal{LADP}')$$

where $q \in [1, +\infty]$ is such that $1/p + 1/q = 1$ and B corresponds to a suitably chosen "big-M"

constant. Then, $\mathcal{LADP}'$ is always feasible, and

1. *If $p \in [1, +\infty]$, then $\mathcal{LADP}'$ is a conservative approximation (lower bound) to $\mathcal{LADP}$.*

2. *If $p = 1$ or $p = +\infty$, then $\mathcal{LADP}$ and $\mathcal{LADP}'$ are equivalent.*

    Proof.   See Electronic Companion EC.3.

Problem $\mathcal{LADP}'$ is a mixed-binary conic problem that is solvable with off-the-shelf solvers. For

any fixed number of subsets $|\mathcal{S}|$ for the partition, its size remains polynomially bounded with the

number of macro-periods M employed for the time-period aggregation and the number of candidate

strategies K. Nevertheless, if $|\mathcal{S}|$ is not fixed, then the size of the problem becomes exponential

in M. We note that the optimal solutions to problems with fewer subsets can be efficiently used

to warm-start the solution to a problem with finer granularity of the partition. We will follow this

strategy in our numerical experiments, see Section 8.

5.3. Strategies for mitigating the loss of optimality

In the previous section, we proposed to aggregate the periods in the problem and to subsequently,

adaptively select, at the beginning of each macro-period, one of finitely many candidate pricing

strategies. Rather than allowing the coefficients of the candidate strategies to adapt to the entire

history of observations, we restricted them to solely adapt to certain functions of the historical

price-demand pairs that concisely summarize the historical information. We then proposed to

approximate these adaptive coefficients by functions that are piecewise constant over a preselected

partition of the set of observable uncertainties.

On the one hand and as noted in Section 5.1, the size of the resulting problem $\mathcal{LADP}'$ is

exponential in the number of macro-periods if $|\mathcal{S}|$ is not fixed. This implies that it is desirable to

keep $|\mathcal{S}|$ small. On the other hand, keeping $|\mathcal{S}|$ small may result in a significant loss of optimality,

see the proof of Proposition 2 and Section 8. In this section, we propose two strategies that enable

us to select partitions with small $|\mathcal{S}|$ while mitigating the loss of optimality of the approximation.

To the best of our knowledge, neither of these approaches has been proposed in the literature.

Adaptive partition. A natural way to mitigate the conservatism of the proposed approach at moderate computational expense is to model the breakpoints used for the partition (7) as decision variables of the problem. Then, the uncertainty set (10) is not only allowed to depend on the coefficients of the candidate strategies on each subset but also on the breakpoints $w_{s_i}^i \in \mathbb{R}$, $i = 1, \ldots, Mn_\xi$, $s \in \mathcal{S}$. In order to be able to reformulate the resulting problem as a mixed-binary conic problem with the tools proposed in Section 5.2, we approximate these real-valued breakpoints by

$$w_{s_i}^i = \underline{\xi}_i + (\overline{\xi}_i - \underline{\xi}_i) \left\lceil \sum_{b=1}^{\overline{b}} 2^{-b} y_{s_i,b}^i \right\rceil \quad \text{with} \quad y_{s_i,b}^i \in \{0,1\}, \ i = 1, \ldots, Mn_\xi,$$

where $\overline{b} \in \mathbb{N}$ denotes the number of bits used to encode each breakpoint and $y_{s_i,b}^i \in \{0,1\}$ correspond to new (static) binary decision variables that affect the uncertainty set. For fixed $|\mathcal{S}|$, the size of the problem resulting from the methodology proposed in Section 5.2 with the breakpoints treated as decision variables is polynomial in $\overline{b}$.

Lifted formulation. In the proposed decision rule approximation scheme, see Section 5.1, we partitioned the uncertainty set *orthogonally* to the axes of the observable uncertain parameters. This is restrictive and may result in large optimality gaps, even for large values of $|\mathcal{S}|$. We thus naturally propose to augment the vector of observable uncertainties by uncertainties that are expressible as linear functions of $\boldsymbol{\xi}$, and to subsequently partition orthogonally the "lifted" uncertainty set. This is equivalent to partitioning the uncertainty set along arbitrary directions, and adds more flexibility to the decision rule approximation.

6. Variants and Extensions

In this section, we discuss numerous variants and extensions to the basic model introduced in Section 2.2 to which our solution approach and modeling paradigm remain applicable.

6.1. Demand model variants

Demand model with memory of past prices. Probably the most natural extension to the proposed linear demand model is one where customers have memory of the past ℓ prices, whereby the demand is expressible as $d_t = \alpha + \sum_{\tau=0}^{\ell} \beta_\tau p_{t-\tau} + \epsilon_t \quad \forall t \in \overline{\mathcal{T}}$. The solution approach presented in Section 5.1 naturally extends to this demand curve variant.

Polynomial demand curve model. As discussed in Section 2.2, a linear demand curve model may

be inadequate if the feasible price range $[l, u]$ is large. The proposed solution approach is applicable

to the case of polynomial demand curve models of degree $d \in \mathbb{N}$ of the form $d_t = \alpha + \sum_{\alpha=1}^{d} \beta_\alpha p_t^\alpha$,

since $p_t^\alpha = (\boldsymbol{\pi}_t^\top \boldsymbol{z}_m)^\alpha = [(\boldsymbol{\pi}_t \circ \boldsymbol{\pi}_t)^\alpha]^\top \boldsymbol{z}_m$, $t \in \mathcal{T}_m$, is linear in $\boldsymbol{z}_m$, $m \in \mathcal{M}$.

Strategic customers. A possible criticism of the model provided in Section 2 is that it does not

account for the naturally strategic behavior of customers. Indeed, in reality, customers may be

willing to wait before purchasing a product, in anticipation of a drop in price. Our framework

can naturally capture this behavior by allowing the demand to also depend on future prices, i.e.,

$d_t = \alpha + \sum_{\tau=0}^{\ell} \beta_\tau p_{t+\tau} + \epsilon_t \quad \forall t \in \overline{\mathcal{T}}$. In this way, we view customers as being able to anticipate

(forecast in fact) future price movements.

Time-varying demand curve parameters. Finally, our framework easily extends to the case of

a time-varying demand curve, where demand is expressible as $d_t = \alpha_t + \beta_t p_t + \epsilon_t \quad \forall t \in \overline{\mathcal{T}}$, for

some parameters $\alpha_t \in \mathbb{R}$ and $\beta_t \in \mathbb{R}$ that are unknown to the seller. In this context, the variabil-

ity of the parameters may be restricted by e.g., bounding the p-norm of the changes as follows:

$\|(\alpha_t - \alpha_{t-1})_{t \in \overline{\mathcal{T}}}\|_p \leq \eta_\alpha$ and $\|(\beta_t - \beta_{t-1})_{t \in \overline{\mathcal{T}}}\|_p \leq \eta_\beta$, with η_α, $\eta_\beta > 0$. For the stage-aggregation

approximation, one would then employ weighted least-squares estimates with weights decreasing

for data further in the past (thus forgetting information).

6.2. Inventory management with pricing

Newsboy problem with pricing and demand learning. The simplest extension of the dynamic

pricing problem discussed in this paper is the Newsboy problem with pricing. In this variant, the

capacity (inventory) c of the product is decided by the seller at the beginning of the planning

horizon. A production cost $c_p \in \mathbb{R}_+$ is incurred for each unit of the product ordered. Naturally, the

solution approach proposed applies to this variant. We remark that there is no need to encode c

in terms of binary variables as it does not affect the uncertainty set.

Dynamic inventory management with pricing. As discussed in Section 1, it is generally beneficial

for retailers to leverage on both sides of the profit equation by simultaneously controlling inven-

tories and prices to mitigate imbalances in supply and demand. It is thus natural to ask whether

the proposed framework extends to the dynamic pricing problem with inventory management. The

answer is positive. We consider an extension of the inventory planning problem presented in Bert-

simas et al. (2011b) that incorporates pricing decisions and demand learning and briefly discuss

the steps involved in applying our approach.

The model is as follows. At the beginning of each stage $t \in \mathcal{T}$, the retailer chooses the price

$p_t \in [l, u]$ that will prevail during that period. Subsequently, he faces the price-sensitive demand d_t

which is expressible in the form (2). The demand must be satisfied from the on-hand inventory

$x_t \in \mathbb{R}$, which may be replenished by placing orders $u_t \in \mathbb{R}_+$ with a supplier at a cost $r \in \mathbb{R}_+$ per

unit. We assume that ordering decisions take immediate effect. The inventory dynamics are thus

$$x_t = x_{t-1} + u_t - d_{t-1} \quad \forall t \in \mathcal{T}. \tag{11}$$

For a given price-demand path, the profit of the retailer is expressible as

$$\sum_{t \in \mathcal{T}} p_t d_t - r u_t - \max(h x_t, -b x_t).$$

When the parameters of the demand model are unknown, the ordering and inventory decisions (also

the pricing decisions) must be modeled as functions of the history of price-demand observations

$(d_\tau, p_\tau)_{\tau=1}^{t-1}$, see Section 3.1. We let $\chi_t \in (\mathbb{R})^{(\mathbb{R}^{2(t-1)})}$ and $\upsilon_t \in (\mathbb{R}_+)^{(\mathbb{R}^{2(t-1)})}$ denote the inventory and

ordering policies, respectively, i.e., $x_t = \chi_t(p_1, d_1, \ldots, p_{t-1}, d_{t-1})$ and $u_t = \upsilon_t(p_1, d_1, \ldots, p_{t-1}, d_{t-1})$.

In the absence of distributional assumptions, the set of possible realizations for the price-demand

path is representable as in (5).

In order to solve this variant of the pricing problem, we proceed in three steps. First, we eliminate

the χ_t variables using their expression, which follows from (11). Second, we proceed with the stage-

aggregation and decision rule approximations (note that υ_t does not need to be encoded in terms

of binary variables), whereby υ_m (the ordering decisions at the beginning of macro-period m) and

coefficients $\boldsymbol{z}_m$ are modeled as non-anticipative piecewise linear and piecewise constant functions

of $\boldsymbol{\xi}$, respectively. Thus, in the spirit of Vayanos et al. (2011), v_m is expressible as

$$v_m(\boldsymbol{\xi}) = \sum_{\boldsymbol{s} \in \mathcal{S}} \mathbb{1}_{\Xi_{\boldsymbol{s}}}(\boldsymbol{\xi}) (\boldsymbol{v}_m^{\boldsymbol{s}})^{\top} \boldsymbol{O}_m \boldsymbol{\xi},$$

for some $\boldsymbol{v}_m^{\boldsymbol{s}} \in \mathbb{R}^{mn_\xi}$, $m \in \mathcal{M}$, satisfying the non-anticipativity constraints

$$\boldsymbol{v}_m^{\boldsymbol{s}} = \boldsymbol{v}_m^{\boldsymbol{s}'} \quad \text{for all} \quad \boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S} \quad \text{such that} \quad \boldsymbol{O}_m \boldsymbol{s} = \boldsymbol{O}_m \boldsymbol{s}',$$

while $\boldsymbol{z}_m$ satisfies (8) and (9). These approximations result in a linear single-stage robust optimiza-

tion problem with decision-dependent uncertainty-set. The final step consists in reformulating this

problem as a mixed-binary conic problem, which is achieved using the techniques from Section 5.2.

6.3. Multi-product pricing

Network revenue management. In this variant, the seller offers an array of I distinct products,

and is endowed with a finite collection of resources $r \in \mathcal{R} := \{1, \ldots, R\}$, each with (fixed) finite

capacity c_r that are used to produce (or assemble) the products. At the beginning of each period

$t \in \mathcal{T}$, the seller must choose prices $p_{ti} \in [l, u]$ for each product i in his product menu $\mathcal{I} := \{1, \ldots, I\}$.

Each unit of demand for product i consumes a quantity m_{ri} of resource r. At the end of the

planning horizon, the seller incurs holding or backlogging costs for each unit of under- or over-

used resource r. These are denoted by h_r and b_r, respectively. Collecting the costs into vectors

$\boldsymbol{h} := \{h_r\}_{r \in \mathcal{R}}$ and $\boldsymbol{b} := \{b_r\}_{r \in \mathcal{R}}$, the profit function of the seller (for a given price sequence and

ensuing demand realization) is expressible as

$$\sum_{t \in \mathcal{T}} \boldsymbol{d}_t^{\top} \boldsymbol{p}_t - \max \left\{ \boldsymbol{h}^{\top} \left(\boldsymbol{c} - \boldsymbol{M} \sum_{t \in \mathcal{T}} \boldsymbol{d}_t \right), \boldsymbol{b}^{\top} \left(\boldsymbol{M} \sum_{t \in \mathcal{T}} \boldsymbol{d}_t - \boldsymbol{c} \right) \right\},$$

where $\boldsymbol{d}_t := (d_{ti})_{i \in \mathcal{I}}$, $\boldsymbol{p}_t := (p_{ti})_{i \in \mathcal{I}}$, and $\boldsymbol{M} \in \mathbb{R}^{R \times I}$ is the *incidence matrix* which collects the quanti-

ties m_{ri}. A natural extension to our modeling paradigm can be obtained by assuming that a linear

demand model of the form $\boldsymbol{d}_t = \boldsymbol{\alpha} + \boldsymbol{B} \boldsymbol{p}_t + \boldsymbol{\epsilon}_t \quad \forall t \in \overline{\mathcal{T}}$ prevails throughout the selling season, where

$\boldsymbol{\alpha} \in \mathbb{R}^I$ and $\boldsymbol{B} \in \mathbb{R}^{I \times I}$ are the demand curve parameters. In order to capture the ability of the

decision maker to exploit information in this setting, we model his pricing decisions as functions of the historical observations related to all products. Thus, the seller optimizes over a pricing policy $\boldsymbol{\pi} := (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_T)$, where each $\boldsymbol{\pi}_t$, $t \in \mathcal{T}$, is a measurable function from $\mathbb{R}^{2I(t-1)}$ to $[l, u]^I$. Extending the solution approach is simple. We omit the derivation due to space limitations. Note that the multi-product pricing problem with finite inventories is a special case of the network revenue management problem and can thus be accommodated by our modeling and solution paradigms.

7. Heuristic solution approaches

In this section, we present dynamic pricing strategies commonly used in practice. In Section 8, we will benchmark our proposed solution approach against these policies.

7.1. Certainty equivalent policies

The most common pricing strategies are the so-called certainty equivalent or *greedy iterated least-squares* policies. In this framework, the retailer selects at each stage $t \in \mathcal{T}$ the price $p_t \in [l, u]$ that maximizes his profit in the "nominal" realization of the uncertain parameters based on current knowledge. In our distribution-free paradigm, this corresponds to fixing (α, β) to its least-squares estimate $(\hat{\alpha}, \hat{\beta})$ calculated using the historical data $(d_\tau, p_\tau)_{\tau \in \mathcal{H}}$ and setting $\epsilon_t = 0$ for all $t \in \mathcal{T}$. Then, an optimal price p_1 to charge in the first stage can be computed by solving

$$
\begin{aligned}
\text{maximize } & \sum_{t \in \mathcal{T}} (\hat{a} + \hat{\beta} p_t) p_t - \max\{h(c - \sum_{t \in \mathcal{T}} \hat{a} + \hat{\beta} p_t), -b(c - \sum_{t \in \mathcal{T}} \hat{a} + \hat{\beta} p_t)\} \\
\text{subject to } & p_t \in [l, u] \quad \forall t \in \mathcal{T}.
\end{aligned}
\tag{12}
$$

Note that (12) is equivalent to a convex quadratic program if $\hat{\beta} \leq 0$, and to a mixed-binary linear program if the feasible price set is discrete.

7.2. Myopic policies

Another class of policies employed in practice are so-called myopic or *static* pricing strategies, see Section 4.1. In this framework, the seller updates his beliefs about the uncertain parameters and subsequently implements the price that is optimal based on these beliefs, iteratively at each stage.

Contrary to certainty equivalent policies (see Section 7.1), these explicitly account for uncertainty. Nevertheless, they do not account for the fact that the seller will be able to adjust his future prices based on the information dynamically acquired. In the distribution-free framework of this paper, an optimal myopic price p_1 to charge in the first stage can be computed by solving the robust optimization problem

$$\max \min_{(\alpha,\beta,\epsilon)\in\overline{\Theta}} \sum_{t\in\mathcal{T}}(\alpha+\beta p_t+\epsilon_t)p_t - \max\left\{h\left(c-\sum_{t\in\mathcal{T}}\alpha+\beta p_t+\epsilon_t\right),-b\left(c-\sum_{t\in\mathcal{T}}\alpha+\beta p_t+\epsilon_t\right)\right\}$$

$$\text{s.t. } p_t\in[l,u] \quad \forall t\in\mathcal{T},$$

(13)

where

$$\overline{\Theta}:=\{(\alpha,\beta,\boldsymbol{\epsilon})\in\mathbb{R}^{2+T+H} \ : \ d_t=\alpha+\beta p_t+\epsilon_t \ \ \forall t\in\mathcal{H}, \ \|\boldsymbol{\epsilon}\|_p\leq\eta\}.$$

(14)

8. Numerical results

In this section, we benchmark our proposed methodology (adaptive policies) against the heuristics commonly employed in practice (see Section 7) on two synthetic data sets. We begin by describing the data sets under consideration and then address the following questions in turn:

(a) How does the relative *guaranteed* performance of adaptive and myopic policies change as the design parameters of the policies are varied?

(b) What is the computational effort (solver time) required to compute adaptive policies?

(c) What is the relative performance of adaptive, myopic and certainty equivalent policies? How do each of these policies compare to a perfect information (anticipative) policy?

All computational experiments were run on a 2.66GHz Intel Core i7-920 processor machine with 24GB RAM and all optimization problems were solved with CPLEX 12.6.

8.1. Problem parameters and historical data sets

We consider two instances of the the dynamic pricing problem $\mathcal{DP}$, which we denote by $\mathcal{DP}_1$ and $\mathcal{DP}_2$, respectively. The parameters of these instances are provided in Table 1. In both cases, the seller has at his disposal four historical price-demand pairs. These are shown on Figure 3, together with the associated data-driven uncertainty sets for (α,β), constructed as discussed in Section 3.2.

Table 1   Parameter values for the instances $\mathcal{DP}_1$ and $\mathcal{DP}_2$ of the dynamic pricing problem $\mathcal{DP}$. The data in the

first row correspond to the true but unknown demand curve parameter values.

Parameter	$\mathcal{DP}_1$	$\mathcal{DP}_2$
(α, β)	$(20, -2)$	$(21, -3)$
$\mathcal{P}$ (\$)	$\{4, 4.2, 4.4, 4.6, 4.8, 5\}$	$\{2.5, 3, 3.5, 4, 4.5, 5\}$
p	∞	∞
η	1	3.25
T	30	30
c (\$)	320	230
(h, b) (\$)	$(15, 23)$	$(25, 35)$

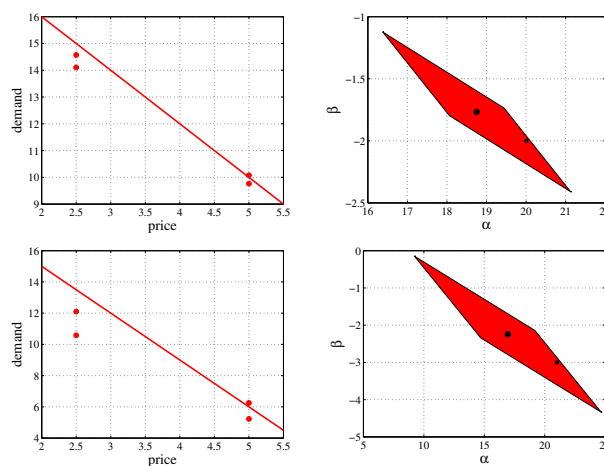**Figure 3**   Historical price-demand observations and true (but unknown) demand curves (left) and prior data-driven uncertainty sets for (α, β) (right) for the problem instances $\mathcal{DP}_1$ (top) and $\mathcal{DP}_2$ (bottom). The dot and the asterisk on the figures on the right denote the least-squares estimate of the demand curve parameters and their true value, respectively

8.2. Performance of optimal adaptive policies in dependence of policy design parameters

We investigate the performance of our proposed policies relative to optimal myopic policies (see Section 7.2) on the instance of the dynamic pricing problem $\mathcal{DP}_1$ presented in Section 8.1. First, we study the *learning gain*, which we define as the increase in guaranteed (worst-case) profits that can be achieved by passing from myopic (i.e., static robust) policies to the proposed adaptive policies. In particular, we investigate the impact of the adaptive policy design parameters (selected by the retailer) on the learning gain. Second, we compare the performance of a computed adaptive policy and an optimal myopic policy out-of-sample.

Adaptive policies. For our numerical experiments, we consider adaptive pricing strategies that depend on the least-squares estimates for (α, β), the cumulative demand, and a linear function of the least-squares estimates at each macro-period, see Example 3 and Section 5.3. Thus, at the beginning of macro-period m, there are $4(m-1)$ observable uncertain parameters and the total number of observable uncertainties is $4(M-1) \ll 2T$. As prices are discrete, see Table 1, we choose to consider only constant candidate pricing strategies, each equal to one of the six feasible prices, i.e., $K = 6$. In other words, at the beginning of each macro-period, we select one of six constant pricing strategies, see Section 5.1.

Impact of policy design parameters on learning gain. For the first set of experiments, we investigate the learning gain in dependence of the policy design parameters (number of macro-periods M and breakpoint configuration $\boldsymbol{r} = (1, \ldots, r_{4(M-1)})$). For each $M \in \{1, \ldots, 5\}$ and for each breakpoint configuration $\boldsymbol{r} = (1, \ldots, r_{4(M-1)})$ such that $|\mathcal{S}| \leq 6$, we compute the optimal objective value of the conservative approximation $\mathcal{LADP}'$ to the dynamic pricing problem, which we denote by $\mathcal{O}_{\boldsymbol{r},M}$. We then record the learning gain $(\mathcal{O}_{\boldsymbol{r},M} - \mathcal{O}_{\mathrm{m}})/\mathcal{O}_{\mathrm{m}}$, where $\mathcal{O}_{\mathrm{m}}$ denotes the optimal objective value of the static robust pricing problem (13), equal to \$412.03 in this instance. We visualize the learning gain in dependence of solver time on Figure 4. From the figures, we observe that the learning gain is negative $(\simeq -10\%)$ when $|\mathcal{S}| = 1$ and $M \leq 2$. This is due to the stage aggregation. Indeed, the myopic strategy adjusts prices at each micro-period $t \in \mathcal{T}$, while the adaptive strategies adjust their prices infrequently $(M \ll T$ times$)$. Nevertheless, with M as small as 3 (i.e., by only adjusting the pricing strategy three times within the planning horizon), we observe that all the points on the efficient frontier achieve a non-negative learning gain. This implies that even by adjusting the prices infrequently, the seller can drastically increase his guaranteed profits by learning the demand curve. In fact, by adjusting the prices four times $(M = 4)$ during the planning horizon, the seller can achieve a learning gain greater than 73%. Finally, we observe that the maximal learning gain achieved (for $|\mathcal{S}| \leq 6$) for $M = 4$ is identical to the maximal learning gain achieved with $M = 5$, indicating convergence of the stage aggregation approximation.
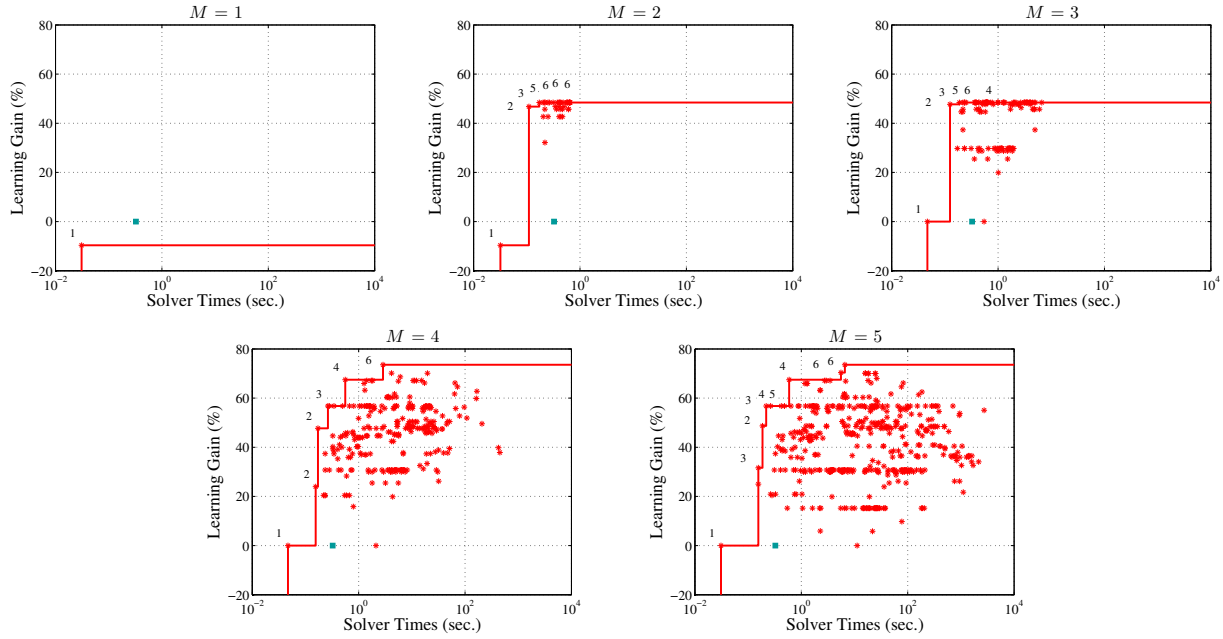
38

Bertsimas and Vayanos: *Data-driven learning in dynamic pricing*
Article submitted to *Operations Research*; manuscript no. (Please, provide the manuscript number!)

Figure 4 Learning gain in dependence of solver time for instance $\mathcal{DP}_1$, for M varying from 1 to 5 and for all breakpoint configurations with $|\mathcal{S}| \leq 6$. The filled line on each figure is the efficient frontier: it connects the markers associated with the problems that achieved the highest learning gain for a given time budget. The square corresponds to the optimal myopic policy. The numbers next to the markers correspond to the cardinality of $\mathcal{S}$ for the problems on the efficient frontier. The problems with $|\mathcal{S}| > 1$ were warm-started with the solution to the associated problem with $|\mathcal{S}| = 1$ (solver times are cumulative).
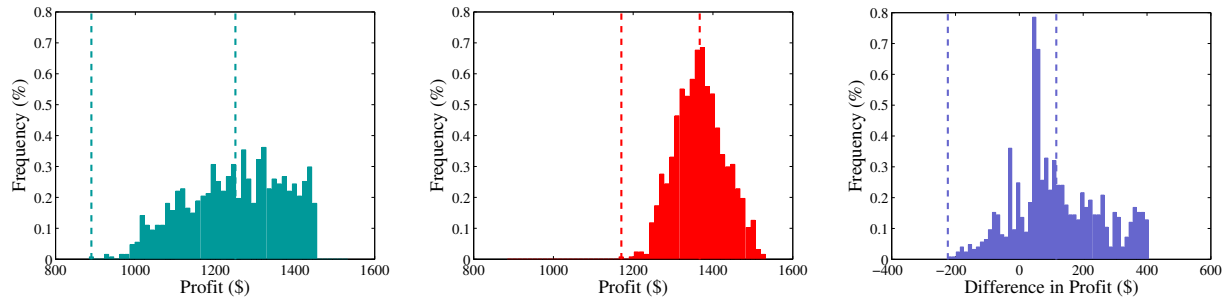

Figure 5 Empirical profit distribution for the myopic policy (left) and for an optimal adaptive policy with $M = 5$ and $|\mathcal{S}| = 16$ (middle) for the instance of the dynamic pricing problem $\mathcal{DP}_1$. The figure on the right depicts the empirical distribution of the difference in profits between the two policies (adaptive less myopic). On each graph, the dotted lines on the left and right correspond to the minimum and mean values of the associated series.

Out-of-sample performance. For the second set of experiments, we compare the empirical profit distributions associated with an optimal myopic policy and an optimal adaptive policy ($M = 5$ and $r = (2, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 1, 1, 1, 1)$). For this purpose, we compute a policy of each type, and draw 1000 samples for (α, β) and ϵ uniformly from the sets of parameters that are compatible with the model and the historical data. For each sample, we implement each policy in turn, and record the profit that the retailer would make by adjusting his prices as dictated by that policy. The profit and loss distributions for these two policies and the distribution of the difference in profits are all shown on Figure 5. By following an adaptive rather than a myopic policy, the retailer is able to increase his profits (over the sample) by over 31.5% in the worst-case and by over 9.2% on average. At the same time, he is able to halve the standard deviation of his profits. Finally, we observe that while the adaptive policy may yield up to \$223 less than the myopic policy, it yields higher profits with probability greater than 81.6%.

8.3. Computational effort in dependence of policy design parameters

In this section, we review the results obtained in the experiments of Section 8.2 from the computational perspective. For this purpose, we recall that Figure 4 depicts the learning gain in dependence of solver time for varying design parameters of the adaptive policies. From the figures, we observe that all problems on the efficient frontiers were solved in less than 10 seconds. In particular, a learning gain of over 73% (maximal learning gain over all experiments) can be achieved in under 3 seconds of solver time ($M = 4$). We note that while the average solver times grow exponentially with $|\mathcal{S}|$, a small $|\mathcal{S}|$ ($= 4$) is sufficient to achieve a learning gain of over 67% ($M = 4$) with associated solver time 0.6 seconds.

8.4. Relative performance of adaptive, myopic and certainty equivalent policies

In this section, we investigate the performance of our proposed adaptive policies relative to pricing strategies commonly employed in practice (see Section 7) in a *folding horizon* setting on the instance of the dynamic pricing problem $\mathcal{DP}_2$ presented in Section 8.1. We also compare each of these approaches to a perfect information (anticipative) policy.

Thus, we draw 500 samples for (α, β) and $\boldsymbol{\epsilon}$ uniformly from the sets of parameters that are compatible with the model and the historical data. For each sample, we compute the optimal price sequence/policy today but implement the "here-and-now" price p_1 only. Subsequently, we observe the resulting (price-dependent) demand faced by the retailer, update the inventory and the retailer's beliefs about the demand curve parameters, and decrement the length of the planning horizon by one. Finally, we re-optimize and repeat this until the end of the planning horizon and for each of myopic, certainty equivalent and adaptive policies (with design parameters $M = 4$ and $\boldsymbol{r} = (1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2)$, and with $\boldsymbol{\xi}$ constructed as in the experiments from Section 8.2). We record the profits earned for each sample and for each policy. For each sample, we also compute and record the profits that would be earned by a perfect information policy (which knows at time $t = 1$ both the true demand curve parameters and the entire sequence of residual error realizations). We emphasize that this latter policy cannot be implemented in reality. Nevertheless, it provides an upper bound on the performance of any non-anticipative pricing strategy. The results are summarized in Table 2 and Figure 6.

From the table and the figures, we observe that the adaptive policy outperforms both the certainty equivalent and myopic policies in terms of all the reported statistics (average and tail performance). The most significant improvements can be seen on the left tail side of the profit and loss distribution. First, the worst-case loss can be decreased by over 76.6% (92.8%) by passing from certainty equivalent (myopic) policies to adaptive strategies. Second, the 1% and 5% Value-at-Risk (VaR) can both be drastically decreased by employing adaptive policies. For example, the 1% VaR drops by over 54% (108%) when passing from certainty equivalent (myopic) to adaptive policies. Finally, the Conditional VaR (CVaR) at levels 1% and 5% can be sharply increased by employing adaptive policies. For example, the 1% CVaR can be increased by over 111.4% by employing adaptive rather than certainty equivalent policies. Moreover, we note that the 1% CVaR for myopic policies is $-\$35.62$ implying that if a retailer employing myopic policies earns less than \$180.19, he will, on average, lose \$35.62. Regarding average performance, we observe that the expected

Table 2 Statistics for the empirical profit distributions associated with a myopic policy, a certainty equivalent policy, and for an optimal adaptive policy with $M = 4$ and $|\mathcal{S}| = 8$ for the instance of the dynamic pricing problem $\mathcal{DP}_2$ solved in a folding horizon fashion. Also, comparison with the performance of a perfect information (anticipative) policy.

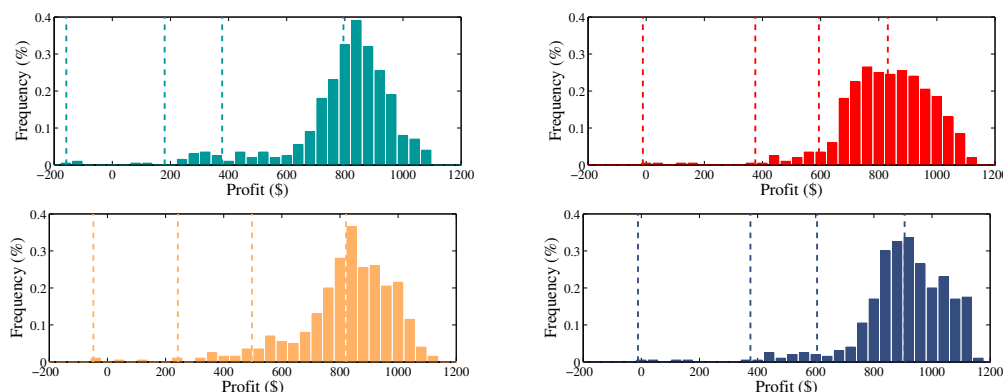|	Myopic	Certainty Equivalent	Adaptive	Perfect Information
Minimum	-158.11	-48.30	-11.29^*	-11.29
Mean	795.26	820.70	831.22^*	905.82
Standard Deviation	181.90	171.07	154.16^*	158.05
5% Value-at-Risk	-378.10	-497.27	-594.33^*	-604.62
1% Value-at-Risk	-180.19	-242.27	-375.51^*	-375.51
5% Conditional VaR	233.19	334.90	433.00^*	435.98
1% Conditional VaR	-35.62	61.15	129.33^*	129.33

Figure 6 Empirical profit distribution for a myopic policy (top left), a certainty equivalent policy (bottom left), an optimal adaptive policy with $M = 4$ and $|\mathcal{S}| = 8$ (top right), and for an optimal anticipative policy (bottom right) for the instance of the dynamic pricing problem $\mathcal{DP}_2$ solved in a folding horizon fashion. On each graph, the dotted lines correspond (from left to right) to the minimum, 1st and 5th percentiles, and mean values of the associated series.

profits (over the sample) can be increased by 1.28% (4.52%) when passing from certainty equivalent (myopic) policies to adaptive policies. Finally, we note that in this instance, the certainty equivalent policy performed significantly better than the myopic policy (both in terms of average and tail performance). This is not surprising: despite the fact that the myopic policy accounts for uncertainty, it optimizes in view of worst-case scenarios and is not guaranteed to outperform the certainty equivalent policy if scenarios other than the worst-case materialize, see Remark 5.

42

Bertsimas and Vayanos: *Data-driven learning in dynamic pricing*
Article submitted to *Operations Research*; manuscript no. (Please, provide the manuscript number!)

From the table and the figures, it becomes apparent that the tail performance of the computed adaptive policy is nearly identical to the tail performance of the perfect information policy. Indeed, worst-case loss, 1% VaR, and 1% CVaR coincide exactly for these two policies. This implies that the performance of the adaptive policy is near-optimal on the left tail (loss) side.

8.5. Conclusions from numerical experiments

We now summarize the insights we obtained from our numerical experiments:

(a) The learning gain (increase in guaranteed profits that can be achieved by employing adaptive rather than myopic policies) can be substantial (over 73% for the instance $\mathcal{DP}_1$). Moreover, adaptive policies perform significantly better than myopic policies out-of-sample (31.5% increase in profits in the worst-case and 9.2% on average for $\mathcal{DP}_1$).

(b) A substantial learning gain can be achieved at modest computational expense (all problems on efficient frontier solved in less than 10 seconds; learning gain of over 67% in under 0.6 seconds of solver time in the case of $\mathcal{DP}_1$).

(c) Adaptive policies with even small M and $|\mathcal{S}|$ can significantly outperform myopic and certainty equivalent policies both on average (expected profit increase of over 1.28% for instance $\mathcal{DP}_2$) and in particular in terms of tail performance (worst-case loss decrease greater than 76.6% and 1% VaR drop greater than 54%). In fact, adaptive policies perform nearly as well as perfect information policies with respect to downside risk measures such as the VaR and CVaR.

References

E. Adida and G. Perakis. A robust optimization approach to dynamic pricing and inventory control with no backorders. *Mathematic Programming*, 107(1):97–129, 2006.

E. Adida and G. Perakis. Dynamic pricing and inventory control: robust vs. stochastic uncertainty models - a computational study. *Annals of Operations Research*, 181:125–157, 2010.

V. F. Araman and R. Caldentey. Dynamic pricing for nonperishable products with demand learning. *Operations Research*, 57:1169–1188, 2009.

Y. Aviv and A. Pazgal. Pricing of short life-cycle products through active learning. Working paper, 2005.

A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

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

A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press, 2009.

D. Bertsimas and C. Caramanis. Finite adaptability for linear optimization. *IEEE Transactions on Automatic Control*, 55(12):2751–2766, 2010.

D. Bertsimas and G. Perakis. Dynamic pricing: a learning approach. In *Mathematic and Computational Models for Congestion Charging*, volume 101 of *Applied Optimization*. Springer, 2006.

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

D. Bertsimas, D. Iancu, and P. Parrilo. A hierarchy of near-optimal policies for multi-stage adaptive optimization. *IEEE Transactions on Automatic Control*, 56(12):2809–2824, 2011b.

O. Besbes and A. Zeevi. Dynamic pricing without knowing the demand function: risk bounds and near-optimal algorithms. *Operations Research*, 57(6):1407–1420, 2009.

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

J. Broder and P. Rusmevichientong. Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980, 2012.

A. X. Carvalho and M. L. Puterman. Dynamic pricing and reinforcement learning. In *Proceedings of the International Joint Conference on Neural Networks*, volume 4, pages 2916–2921, 2003.

A. X. Carvalho and M. L. Puterman. Learning and pricing in an internet environment with binomial demand. *Journal of Revenue and Pricing Management*, 3(4):320–336, 2005a.

A. X. Carvalho and M. L. Puterman. Dynamic optimization and learning: how should a manager set prices when the demand functions is unknown? Technical Report 1117, Instituto de Pesquisa Econômica Aplicada, 2005b.

A. V. den Boer. Dynamic pricing and learning: historical origins, current research, and new directions. Submitted for publication, 2013a.

A. V. den Boer. Dynamic pricing with multiple products and partially specified demand distribution. Submitted for publication, 2013b.

A. V. den Boer. Dynamic pricing and learning with finite inventories. Submitted for publication, 2013c.

A. V. den Boer and B. Zwart. Simultaneously learning and optimizing using controlled variance pricing. Accepted for publication in Management Science, 2013.

W. Elmaghraby and P. Keskinocak. Dynamic pricing: research overview, current practices and future directions. *Manageement Science*, 49:1287–1309, 2003.

S. S. Eren and C. Maglaras. Monopoly pricing with limited demand information. *Journal of Revenue and Pricing Management*, 9:23–48, 2010.

V. F. Farias and B. van Roy. Dynamic pricing with a prior on market response. *Operations Research*, 58(1):16–29, 2010.

A. A. Feldbaum. Dual control theory I-II. *Automation and Remote Control*, 21:874–880, 1033–1039, 1961.

L. Giarré, B.Z. Kacewicz, and M. Milanese. Model quality evaluation in set membership identification. *Automatica*, 33(6):1133—1139, 1997.

G.A. Hanasusanto, D. Kuhn, and W. Wiesemann. Two-stage robust integer programming. Technical report, 2014.

M. J. Harrison, B. N. Keskin, and A Zeevi. Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science*, 58(3):570–586, 2012.

M. J. Harrison, N. B. Keskin, and A. Zeevi. Dynamic pricing with an unknown linear demand model: asymptotic optimal semi-myopic policies. Technical report, The University of Chicago Booth School of Business, 2013.

D. Iancu and N. Trichakis. Pareto efficiency in robust optimization. Technical report, 2013.

A. E. B. Lim and J. G. Shanthikumar. Relative entropy, exponential utility, and robust dynamic pricing. *Operations Research*, 55(2):198–214, 2007.

A. E. B. Lim, J. G. Shanthikumar, and T. Watewai. Robust multi-product pricing. Working Paper, 2008.

K. Y. Lin. Dynamic pricing with real-time demand learning. *European Journal of Operational Research*, 174 (1):522–538, 2006.

R. Lobel and G. Perakis. Dynamic pricing through sampling based optimization. Technical report, Massachusetts Institute of Technology, 2010. Submitted for publication.

M. S. Lobo and S. Boyd. Pricing and learning with uncertain demand. Technical report, Stanford University, 2003.

J. Löfberg. YALMIP: A toolbox for modelling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, 2004.

A. Mas-Colell, M.D. Whinston, and J.R. Green. *Microeconomic Theory*. Oxford student edition. Oxford University Press, 1995.

A. McLennan. Price dispersion and incomplete learning in the long run. *Jounal of Economic Dynamic and Control*, 7(3):331–347, 1984.

M. Milanese and A. Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica*, 27(6):997–1009, 1991.

E. S. Mills. Uncertainty and price theory. *The Quarterly Journal of Economics*, 73(3):116–130, 1959.

R. Phillips. *Pricing and revenue optimization*. Stanford University Press, 2005.

H. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58:527–535, 1951.

M. Rothshield. A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9(2):185–202, 1974.

F.C. Schweppe. Recursive state estimation: Unknown but bounded errors and system input. *IEEE Transactions on Automatic Control*, 13(1):22–28, 1968.

G. Still. Generalized semi-infinite programming: theory and methods. *European Journal of Operational Research*, 119(2):301–313, 1999.

K. T. Talluri and G. J. van Ryzin. *The theory and practice of revenue management.* Kluwer Academic Publishers, 2004.

A. Thiele. Single-product pricing via robust optimization. Working paper, 2006.

A. Thiele. Multi-product pricing via robust optimization. *Journal of Revenue and Pricing Management*, 8: 67–80, 2009.

P. Vayanos, D. Kuhn, and B. Rustem. Decision rules for information discovery in multi-stage stochastic programming. In *Proceedings of the 50th IEEE Conference on Decision and Control*, pages 7368–7373, 2011.

H. Witsenhausen. Sets of possible states of linear systems given perturbed observations. *IEEE Transactions on Automatic Control*, 13(5):556–558, 1968.

Proofs of Statements

EC.1. Proof of Proposition 1

Proof.   A lower bound to the optimal objective value of $\mathcal{DP}$ can be obtained by restricting the pricing policies to be constant. In that case, the demand d_t in the objective of $\mathcal{DP}$ can be replaced by its expression $\alpha + \beta p_t + \epsilon_t$ and the seller's pricing problem (for $h = b = 0$) can be formulated as

$$\text{maximize} \left\{ \underset{(\alpha,\beta,\epsilon)\in\overline{\Theta}}{\text{minimum}} \sum_{t\in\mathcal{T}}(\alpha + \beta p_t + \epsilon_t)p_t \,\middle|\, p_t \in [l,u] \quad \forall t \in \mathcal{T} \right\}, \tag{EC.1}$$

where $\overline{\Theta}$ is defined as in (14). An upper bound to the optimal objective value of $\mathcal{DP}$ can be obtained by inverting the order of the minimization and maximization in (EC.1), yielding

$$\text{minimize} \left\{ \underset{p_t \in [l,u],\, t\in\mathcal{T}}{\text{maximum}} \sum_{t\in\mathcal{T}}(\alpha + \beta p_t + \epsilon_t)p_t \,\middle|\, (\alpha,\beta,\epsilon) \in \overline{\Theta} \right\}. \tag{EC.2}$$

The function $(\boldsymbol{p}, \alpha, \beta, \boldsymbol{\epsilon}) \mapsto \sum_{t\in\mathcal{T}}(\alpha + \beta p_t + \epsilon_t)p_t$ is linear in $(\alpha, \beta, \boldsymbol{\epsilon})$ for each $\boldsymbol{p} \in [l,u]^T$ and concave in $\boldsymbol{p}$ for each $(\alpha, \beta, \boldsymbol{\epsilon}) \in \overline{\Theta}$ (since $\beta \leq 0 \;\forall(\alpha,\beta) \in \Theta$). Moreover, the sets $[l,u]^T$ and $\overline{\Theta}$ are convex. By the minimax theorem, we conclude that the optimal objective values of (EC.1) and (EC.2) coincide. This in turn implies that the optimal objective value of $\mathcal{DP}$ remains unchanged if one optimizes over static pricing strategies only. Moreover, any optimal solution to the robust optimization problem (EC.1) is optimal in $\mathcal{DP}$. This concludes the proof. □

EC.2. Proof of Proposition 2

Proof.   We proceed by means of an example. Consider an instance of the dynamic pricing problem $\mathcal{DP}$ with $T = 2$, $[l,u] = [6,10]$, $c = 20$, $h = 5$ and $b = 15$. Suppose that no historical data is available to the retailer, who nevertheless has prior information (in the form of a box) on the set of possible values for (α, β), which are known to lie in the set

$$\Theta_{\text{prior}} = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, 20 \leq \alpha \leq 30,\, -2 \leq \beta \leq -1\},$$

see Remark 1. Moreover, the retailer knows that the vector of residual errors (ϵ_1, ϵ_2) will be such that $\max(|\epsilon_1|, |\epsilon_2|) \leq 1$. Thus, the uncertainty set for the price-demand realizations is expressible as

$$\mathcal{U}(\pi_1, \pi_2) = \left\{ \begin{array}{l} (p_1, p_2, d_1, d_2) \in \mathbb{R}^4 \;:\; d_t = \alpha + \beta p_t + \epsilon_t \;\; \forall t \in \mathcal{T}, \\[2mm] p_1 = \pi_1, \; p_2 = \pi_2(p_1, d_1), \; \max(|\epsilon_1|, |\epsilon_2|) \leq 1 \end{array} \right\}.$$

An optimal static strategy for this instance of $\mathcal{DP}$ can be found by replacing the demand d_t by its expression $\alpha + \beta p_t + \epsilon_t$ and writing the resulting problem in epigraph form by introducing the auxiliary variable v. This yields

$$
\begin{aligned}
& \text{maximize } v \\[2mm]
& \text{subject to } v \in \mathbb{R}, \; p_t \in [l, u] \;\; \forall t \in \mathcal{T} \\[2mm]
& \qquad \left. \begin{array}{l} v \leq \sum_{t \in \mathcal{T}} p_t(\alpha + \beta p_t + \epsilon_t) - h\{c - \sum_{t \in \mathcal{T}}(\alpha + \beta p_t + \epsilon_t)\} \\[2mm] v \leq \sum_{t \in \mathcal{T}} p_t(\alpha + \beta p_t + \epsilon_t) - b\{\sum_{t \in \mathcal{T}}(\alpha + \beta p_t + \epsilon_t) - c\} \end{array} \right\} \; \forall (\alpha, \beta, \epsilon_1, \epsilon_2) \in \overline{\Theta},
\end{aligned} \tag{EC.3}
$$

where

$$\overline{\Theta} = \{(\alpha, \beta, \epsilon_1, \epsilon_2) \in \mathbb{R}^4 \;:\; (\alpha, \beta) \in \Theta_{\mathrm{prior}}, \; \max(|\epsilon_1|, |\epsilon_2|) \leq 1\}.$$

It can be shown that for any choice of $(p_1, p_2) \in [l, u]^2$, the worst-case in the first constraint of (EC.3) is attained when α, β, ϵ_1 and ϵ_2 are all at their lower bound. Similarly, it can be shown that the worst-case in the second-constraint of (EC.3) is reached for α, β, ϵ_1 and ϵ_2 all fixed to their upper bound. Thus, (EC.3) can be solved as a convex quadratic program (since $\beta \leq 0 \;\; \forall (\alpha, \beta) \in \Theta_{\mathrm{prior}}$). Its optimal objective value is $\tilde{v} = -22$, attained at $\tilde{p}_1 = \tilde{p}_2 = 8$.

We now proceed to construct an adaptive strategy with objective value greater than -22. Define the adaptive strategy $(\pi_1^\star, \pi_2^\star)$ through

$$\pi_1^\star = 8 \quad \text{and} \quad \pi_2^\star(p_1, d_1) = \begin{cases} p_{2,1}^\star := 10 & \text{if } d_1 \geq 13 \\[3mm] p_{2,2}^\star := 6 & \text{else.} \end{cases}$$

We note that the cut-off value 13 corresponds to the value of the demand function at the Chebyshev center of $\overline{\Theta}$ when the chosen price is 8. The objective value of this instance of $\mathcal{DP}$ under the pricing strategy $(\pi_1^\star, \pi_2^\star)$ is given by the greatest value $v \in \mathbb{R}$ satisfying the inequalities

$$\left. \begin{array}{l} v \leq \sum_{t \in \mathcal{T}} p_{t,s}^\star(\alpha + \beta p_{t,s}^\star + \epsilon_t) - h\{c - \sum_{t \in \mathcal{T}}(\alpha + \beta p_{t,s}^\star + \epsilon_t)\} \\ v \leq \sum_{t \in \mathcal{T}} p_{t,s}^\star(\alpha + \beta p_{t,s}^\star + \epsilon_t) - b\{\sum_{t \in \mathcal{T}}(\alpha + \beta p_{t,s}^\star + \epsilon_t) - c\} \end{array} \right\} \forall (\alpha, \beta, \epsilon_1, \epsilon_2) \in \overline{\Theta}_s, \ s \in \{1,2\}$$

where $p_{1,s}^\star = \pi_1^\star$ for $s \in \{1,2\}$,

$$\overline{\Theta}_1 = \{(\alpha, \beta, \epsilon_1, \epsilon_2) \in \overline{\Theta} : \alpha + \pi_1^\star \beta + \epsilon_1 \geq 13\} \quad \text{and} \quad \overline{\Theta}_2 = \{(\alpha, \beta, \epsilon_1, \epsilon_2) \in \overline{\Theta} : \alpha + \pi_1^\star \beta + \epsilon_1 < 13\}.$$

The infimum values of the right-hand-side in the first inequality above are 174 and 16 for $s = 1$ and $s = 2$, respectively. Similarly, the infimum values of the right-hand-side in the second inequality above are 34 and 38 for $s = 1$ and $s = 2$, respectively. Thus, the greatest value of v satisfying these four inequalities is 16, implying that the adaptive pricing strategy $(\pi_1^\star, \pi_2^\star)$ attains an objective value greater than that of any static strategy. Thus, we have provided a simple, two-stage, instance of $\mathcal{DP}$ in which exploration is imperative. In fact, under static pricing strategies, the product appeared not to be profitable in the worst-case (negative objective value), whereas it is profitable under dynamic pricing with learning. This concludes the proof. $\square$

EC.3. Proof of Proposition 3

Proof. We begin the proof by writing $\mathcal{LADP}$ in epigraph form

$$\text{maximize } v$$

$$\text{subject to } v \in \mathbb{R}, \ \boldsymbol{z}_m^{\boldsymbol{s}} \in \{0,1\}^K, \ \mathbf{e}^\top \boldsymbol{z}_m^{\boldsymbol{s}} = 1 \quad \forall m \in \mathcal{M}, \ \boldsymbol{s} \in \mathcal{S}$$

$$\left. \begin{array}{l} v + hc - h\mathbf{e}^\top \boldsymbol{d} \leq \boldsymbol{d}^\top \boldsymbol{P} \boldsymbol{z}^{\boldsymbol{s}} \\ v - bc + b\mathbf{e}^\top \boldsymbol{d} \leq \boldsymbol{d}^\top \boldsymbol{P} \boldsymbol{z}^{\boldsymbol{s}} \end{array} \right\} \ \forall \boldsymbol{d} \in \mathcal{U}_{\boldsymbol{s}}^{\mathrm{a}}(\boldsymbol{z}^{\boldsymbol{s}}), \ \boldsymbol{s} \in \mathcal{S} \qquad (\text{EC.4})$$

$$\boldsymbol{z}_m^{\boldsymbol{s}} = \boldsymbol{z}_m^{\boldsymbol{s}'} \quad \forall \boldsymbol{s}, \ \boldsymbol{s}' \in \mathcal{S} \ : \ \boldsymbol{O}_m \boldsymbol{s} = \boldsymbol{O}_m \boldsymbol{s}', \ m \in \mathcal{M}.$$

We reformulate each generalized semi-infinite constraint in problem (EC.4) in turn. For any fixed $\boldsymbol{s} \in \mathcal{S}$ and $\boldsymbol{z} \in \{0,1\}^K$, it holds that

$$v + hc - h\mathbf{e}^\top \boldsymbol{d} \leq \boldsymbol{d}^\top \boldsymbol{P} \boldsymbol{z} \quad \forall \boldsymbol{d} \in \mathcal{U}_{\boldsymbol{s}}^{\mathrm{a}}(\boldsymbol{z}) \quad \Leftrightarrow \quad \min_{\boldsymbol{d} \in \mathcal{U}_{\boldsymbol{s}}^{\mathrm{a}}(\boldsymbol{z})} (\boldsymbol{P} \boldsymbol{z} + h\mathbf{e})^\top \boldsymbol{d} \geq v + hc \qquad (\text{EC.5})$$

The dual of the inner minimization problem on the right of the equivalence (EC.5) is given by

$$
\begin{aligned}
&\underset{\boldsymbol{\mu}_1,\,\boldsymbol{\mu}_2}{\text{maximize}} \ \ \boldsymbol{g_s}^\top \boldsymbol{\mu}_1 + \boldsymbol{h}^\top \boldsymbol{\mu}_2^{\boldsymbol{s}} \\
&\text{subject to } \boldsymbol{\mu}_1 \in \mathbb{R}_+^{m_1}, \ \boldsymbol{\mu}_2 \in \mathcal{K}_q^{m_2} \\
&\qquad\qquad \boldsymbol{F}(\boldsymbol{z})^\top \boldsymbol{\mu}_1 = \boldsymbol{P}\boldsymbol{z} + h\mathbf{e} \\
&\qquad\qquad \boldsymbol{G}(\boldsymbol{z})^\top \boldsymbol{\mu}_1 + \boldsymbol{H}(\boldsymbol{z})^\top \boldsymbol{\mu}_2 = \boldsymbol{0}.
\end{aligned}
\tag{EC.6}
$$

From Observation 1, it follows that (EC.6) is feasible, since its feasible set is independent of $\boldsymbol{s}$. We now distinguish between the general case $p \in [1, +\infty]$ and the particular setting $p \in \{1, +\infty\}$.

1. If $p \in [1, +\infty]$, then weak duality implies that the optimal objective value of the dual (EC.6) is a lower bound for the optimal objective value of the primal, and

$$
\exists \boldsymbol{\mu}_1 \in \mathbb{R}_+^{m_1}, \ \boldsymbol{\mu}_2 \in \mathcal{K}_q^{m_2} \text{ feasible in (EC.6) with } \boldsymbol{g_s}^\top \boldsymbol{\mu}_1 + \boldsymbol{h}^\top \boldsymbol{\mu}_2^{\boldsymbol{s}} \geq v + hc
$$

$$
\Rightarrow v + hc - h\mathbf{e}^\top \boldsymbol{d} \leq \boldsymbol{d}^\top \boldsymbol{P}\boldsymbol{z} \quad \forall \boldsymbol{d} \in \mathcal{U}_{\boldsymbol{s}}^{\mathrm{a}}(\boldsymbol{z}).
$$

2. If $p \in \{1, +\infty\}$, then strong linear programming duality (which applies since the dual is feasible) implies that

$$
\exists \boldsymbol{\mu}_1 \in \mathbb{R}_+^{m_1}, \ \boldsymbol{\mu}_2 \in \mathcal{K}_q^{m_2} \text{ feasible in (EC.6) with } \boldsymbol{g_s}^\top \boldsymbol{\mu}_1 + \boldsymbol{h}^\top \boldsymbol{\mu}_2^{\boldsymbol{s}} \geq v + hc
$$

$$
\Leftrightarrow v + hc - h\mathbf{e}^\top \boldsymbol{d} \leq \boldsymbol{d}^\top \boldsymbol{P}\boldsymbol{z} \quad \forall \boldsymbol{d} \in \mathcal{U}_{\boldsymbol{s}}^{\mathrm{a}}(\boldsymbol{z}).
$$

Applying the same reasoning to each generalized semi-infinite constraint individually yields the formulation

$$
\begin{aligned}
&\text{maximize } v \\
&\text{subject to } v \in \mathbb{R}, \ \boldsymbol{z}_m^{\boldsymbol{s}} \in \{0,1\}^K, \ \mathbf{e}^\top \boldsymbol{z}_m^{\boldsymbol{s}} = 1 \quad \forall m \in \mathcal{M}, \ \boldsymbol{s} \in \mathcal{S} \\
&\qquad\qquad \boldsymbol{z}_m^{\boldsymbol{s}} = \boldsymbol{z}_m^{\boldsymbol{s}'} \quad \forall \boldsymbol{s}, \ \boldsymbol{s}' \in \mathcal{S} \ : \ \boldsymbol{O}_m \boldsymbol{s} = \boldsymbol{O}_m \boldsymbol{s}', \ m \in \mathcal{M} \\
&\qquad\qquad \left.\begin{aligned}
&\boldsymbol{\mu}_1^{\boldsymbol{s}}, \ \boldsymbol{\nu}_1^{\boldsymbol{s}} \in \mathbb{R}_+^{m_1}, \ \boldsymbol{\mu}_2^{\boldsymbol{s}}, \ \boldsymbol{\nu}_2^{\boldsymbol{s}} \in \mathcal{K}_q^{m_2} \\
&\boldsymbol{g_s}^\top \boldsymbol{\mu}_1^{\boldsymbol{s}} + \boldsymbol{h}^\top \boldsymbol{\mu}_2^{\boldsymbol{s}} \geq v + hc, \ \boldsymbol{g_s}^\top \boldsymbol{\nu}_1^{\boldsymbol{s}} + \boldsymbol{h}^\top \boldsymbol{\nu}_2^{\boldsymbol{s}} \geq v - bc \\
&\boldsymbol{F}(\boldsymbol{z}^{\boldsymbol{s}})^\top \boldsymbol{\mu}_1^{\boldsymbol{s}} = \boldsymbol{P}\boldsymbol{z}^{\boldsymbol{s}} + h\mathbf{e}, \ \boldsymbol{G}(\boldsymbol{z}^{\boldsymbol{s}})^\top \boldsymbol{\mu}_1^{\boldsymbol{s}} + \boldsymbol{H}^\top \boldsymbol{\mu}_2^{\boldsymbol{s}} = \boldsymbol{0} \\
&\boldsymbol{F}(\boldsymbol{z}^{\boldsymbol{s}})^\top \boldsymbol{\nu}_1^{\boldsymbol{s}} = \boldsymbol{P}\boldsymbol{z}^{\boldsymbol{s}} - b\mathbf{e}, \ \boldsymbol{G}(\boldsymbol{z}^{\boldsymbol{s}})^\top \boldsymbol{\nu}_1^{\boldsymbol{s}} + \boldsymbol{H}^\top \boldsymbol{\nu}_2^{\boldsymbol{s}} = \boldsymbol{0}
\end{aligned}\right\} \quad \forall \boldsymbol{s} \in \mathcal{S},
\end{aligned}
\tag{EC.7}
$$

which is equivalent to $\mathcal{LADP}$ if $p \in \{1, +\infty\}$ or constitutes a conservative approximation, otherwise.

Problem (EC.7) is a conic optimization problem involving products of binary and real-valued variables. The remainder of the proof consists in linearizing the bilinear terms using standard big-M techniques. Thus, for each $\boldsymbol{s} \in \mathcal{S}$ and $i \in \{1, \ldots, MK\}$, we introduce the auxiliary variables $\boldsymbol{x}_i^{\boldsymbol{s}}$, $\boldsymbol{y}_i^{\boldsymbol{s}} \in \mathbb{R}_+^{m_1}$ satisfying

$$\boldsymbol{x}_i^{\boldsymbol{s}} = z_i^{\boldsymbol{s}} \boldsymbol{\mu}_1^{\boldsymbol{s}} \quad \Leftrightarrow \quad \boldsymbol{x}_i^{\boldsymbol{s}} \leq \boldsymbol{\mu}_1^{\boldsymbol{s}}, \; \boldsymbol{x}_i^{\boldsymbol{s}} \leq B z_i^{\boldsymbol{s}} \mathbf{e}, \text{ and } \boldsymbol{x}_i^{\boldsymbol{s}} \geq \boldsymbol{\mu}_1^{\boldsymbol{s}} - B(1 - z_i^{\boldsymbol{s}})\mathbf{e}$$

$$\boldsymbol{y}_i^{\boldsymbol{s}} = z_i^{\boldsymbol{s}} \boldsymbol{\nu}_1^{\boldsymbol{s}} \quad \Leftrightarrow \quad \boldsymbol{y}_i^{\boldsymbol{s}} \leq \boldsymbol{\nu}_1^{\boldsymbol{s}}, \; \boldsymbol{y}_i^{\boldsymbol{s}} \leq B z_i^{\boldsymbol{s}} \mathbf{e}, \text{ and } \boldsymbol{y}_i^{\boldsymbol{s}} \geq \boldsymbol{\nu}_1^{\boldsymbol{s}} - B(1 - z_i^{\boldsymbol{s}})\mathbf{e}.$$

This yields the desired formulation and concludes the proof. $\qquad\square$