Data-driven Pricing and Inventory Management with Applications in Fashion Retail

by

Mila Nambiar

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2019

Signature redacted

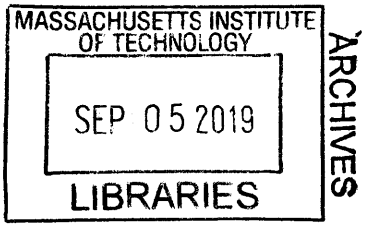
Author .....
.....
Sloan School of Management
August 9, 2019

Signature redacted

Certified by.....
.....
David Simchi-Levi
Professor
Thesis Supervisor

Signature redacted

Accepted by .....
.....
Patrick Jaillet
Professor
Co-director, Operations Research Center



Data-driven Pricing and Inventory Management with Applications in Fashion Retail

by

Mila Nambiar

Submitted to the Sloan School of Management
on July 10, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

Fashion retail is typically characterized by (1) high demand uncertainty and products with short life cycles, which complicates demand forecasting, and (2) low salvage values and long supply lead times, which penalizes for inaccurate demand forecasting. In this thesis, we are interested in the design of algorithms that leverage fashion retail data to improve demand forecasting, and that make revenue-maximizing or cost-minimizing pricing and inventory management decisions.

First, we study a multi-period dynamic pricing problem with feature information. We are especially interested in demand model misspecification, and show that it can lead to price endogeneity, and hence inconsistent price elasticity estimates and sub-optimal pricing decisions. We propose a "random price shock" (RPS) algorithm that combines instrumental variables, well known in econometrics, with online learning, in order to simultaneously estimate demand and optimize revenue. We demonstrate strong theoretical guarantees on the regret of RPS for both IID and non IID features, and numerically validate the algorithm's performance on synthetic data.

Next, we present a case study in collaboration with Oracle Retail. We extend RPS to incorporate common business constraints such as markdown pricing and inventory constraints. We then conduct a counterfactual analysis where we simulate the algorithm's performance using fashion retail data. Our analysis estimates that the RPS algorithm will increase by 2-7% relative to current practice.

Finally, we study an inventory allocation problem in a single-warehouse multiple-retailer setting with lost sales. We show that under general conditions this problem is convex, and that a Lagrangian relaxation-based approach can be applied to solve it in a computationally tractable, and near-optimal way. This analysis allows us to prove structural results that give insights into how the allocation policy should depend on factors such as the retailer demand distributions, and demand learning.

Thesis Supervisor: David Simchi-Levi
Title: Professor

Acknowledgments

First and foremost, I would like to thank my advisor, David Simchi-Levi. David has always inspired me with his energy, and with his vision for how research ideas can grow and take shape as compelling stories. At the same time, he has constantly challenged me to become a clearer and more effective thinker and communicator. This thesis, and my growth as a researcher, would not have been possible without his support.

I would also like to thank the other members of my thesis committee, John Tsitsiklis and Steven Graves, for taking the time to ask detailed questions, and to share suggestions and feedback that have shaped this thesis.

Next, a big thank you must go to the co-authors of the works in this thesis. Chapters 2-4 are joint work with my advisor, David Simchi-Levi, and He Wang. I would like to thank He Wang for working with me throughout these years, and for graciously hosting me at Georgia Tech earlier this year to work on the material in Chapter 4. I have learned a lot from his knowledge and experience, and from his creativity in and intuition for problem solving. I would also like to thank Su-Ming Wu and Setareh Borjian from Oracle Retail Global Business Unit, who are co-authors on the work in Chapter 3. I am grateful to them for taking the time to share their experience with revenue management practice with me. Their suggestions and feedback have been invaluable in designing the numerical experiments in this chapter.

Finally, I would like to thank some friends who have been especially supportive during this process: Ying Qiao Hee, Nathan Watson and Delphine Watson, for coming all the way from LA just to attend my defense, Rushabh Shah, Peter Yun Zhang, Charmaine Chia, and Xuewei Loy, for always checking in on me, and of course, my best friend and partner Ben Simon, for his unconditional support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis, we are interested in pricing and inventory management problems that arise in fashion retail. Among retail applications, fashion retail is unique for a number of reasons. Most distinctive is the **seasonality** of fashion retail: Every season, new product lines are introduced at stores, sold for only around 10 weeks, then removed from stores (Fisher and Raman, 1996).

Also, the amount of **inventory** available during each season tends to be **fixed**. This is because items have to be shipped across long distances, from the countries where they are manufactured, to reach stores located in the US, Europe, etc. Since supply lead times are long relative to the product lifecycles, replenishments during the short selling seasons are thus either not possible, or are limited in number. For example, at the Spain-based fast fashion retailer Zara, which is one of the largest fashion retailers today, initial shipments from the production to distribution centers before the start of each season make up around half of the total volume of available inventory during each season (Gallien et al., 2017).

A third defining characteristic of fashion retail is that items that are not sold to customers by the end of each season are resold at very **low salvage values**. These salvage values tend to be much lower than the full prices of the items, and may even be lower than the cost of production (Fisher and Raman, 1996).

Given these factors, fashion retailers face two big challenges: For one, it is difficult to forecast demand accurately for products with short lifecycles, as only a limited

amount of information is available on the sales of each product. At the same time, since inventory is fixed and salvage values are low, inaccurate demand forecasting, or inadequate planning to meet demand, can lead to costly lost sales opportunities.

The three parts of this thesis address these various challenges in fashion retail by proposing data-driven algorithms that make use of the kinds of data that are available to fashion retailers - such as transaction and item feature data - to perform more accurate demand forecasting, and to consequently make better pricing and inventory management decisions. We are particularly interested in algorithms that "learn and earn" on the fly, meaning that learning and optimization take place simultaneously rather than in separate stages. As additional sales or demand data is observed, it is used to update the retailer's demand forecasts, and to then make pricing or inventory management decisions that generate demand, and so on and so forth. Since product lifecycles are so short in fashion retail, incorporating the most recent data in demand forecasting is essential to making more accurate predictions.

Another strategy that fashion retailers can use to improve their demand forecasting for products with short lifecycles is feature-based pricing. Rather than considering each item in isolation, fashion retailers can look at the sales of products with similar characteristics, i.e. products that are in the same category, and that share similar features such as color, brand, design pattern etc, and that are therefore likely to experience similar demands. These kinds of item feature data can be combined with transaction data in order to perform demand estimation and make pricing decisions. Large fashion retailers, such as Zara and its closest competitors H&M and Forever 21, where the number of new articles of clothing released every year are in the thousands (Caro and Gallien, 2012), especially stand to benefit from feature-based pricing.

Chapters 2 and 3 of this thesis study feature-based pricing, the former from a theoretical perspective, and the latter from the point of view of the company Oracle Retail, which is a business unit of Oracle and a leading provider of software and IT solutions to retailers. In Chapter 2, Dynamic learning and pricing with model misspecification, we show that one of the pitfalls of designing a dynamic pricing policy when demand depends on feature information lies in correctly estimating the causal

relationship between demand and price. For example, if the policy assumes a misspecified demand model, either because the decision maker is unsure of how demand is affected by the features, or of how to model such a dependence, demand noise can become correlated with the price (price endogeneity). Price endogeneity then leads to biased estimates of the demand-price relationship, and results in suboptimal pricing decisions. Another factor that can lead to price endogeneity is that pricing managers at companies may choose admissible price sets based on their own beliefs of future demand, and in response to demand factors such as product quality, trendiness, etc., that are unobservable to the algorithm.

In Chapter 2, we thus propose a "random price shock" (RPS) algorithm that manages these pitfalls by dynamically generating randomized price shocks to estimate price elasticity while maximizing revenue. We show that the RPS algorithm has strong theoretical performance guarantees, that it is robust to model misspecification, and that it can be adapted to a number of business settings, including (1) when the feasible price set is a price ladder, and (2) when the contextual information is not IID. We also perform numerical experiments on synthetic data to gauge the performance of RPS, and find that it significantly outperforms competing algorithms that do not account for price endogeneity.

In Chapter 3, Feature-based dynamic pricing with for fashion retail: A case study, we adapt the RPS algorithm to be broadly applicable to fashion retail settings. We keep the price experimentation structure of the RPS algorithm proposed in Chapter 2, but modify it to incorporate business constraints faced by many fashion retailers, such as fixed inventory and markdown pricing constraints. The modified algorithm makes use of intuitive and computationally tractable approximations to optimize the retailer's total expected revenue subject to these constraints. To gauge its performance, we have run a number of offline numerical experiments using retail data from one of Oracle Retail's clients. The heuristic exhibits revenue gains of around 2-7% over current practice, and seems robust to different retailer parameter settings such as the length of the markdown and no-touch periods.

Finally, in Chapter 4, we turn our attention from pricing to inventory alloca-

tion. We study an inventory allocation problem in a two-echelon (single-warehouse multiple-retailer) setting with lost sales. At the start of a finite selling season, a fixed amount of inventory is available at the warehouse, and can be allocated to the retailers over the course of the selling horizon with the objective of minimizing total expected lost sales costs and holding costs. We allow each retailer to experience correlated demands, and show how this framework can capture learning in the sense of demand forecasting (e.g. ARMA) model, as well as a Bayesian learning mdoel.

Then, we pose the questions of (1) how to solve the inventory allocation problem under demand learning in a computationally tractable way, and (2) how demand learning impacts effective inventory allocation policies. To address the first question, we adapt the Lagrangian relaxation-based technique proposed by Marklund and Rosling (2012) for a backordering, no-learning setting. We show under general assumptions that the resulting heuristic remains near-optimal in our setting, compared to the original dynamic program. Finally, we use this analysis to investigate the relationship between demand learning and early allocation decisions. Through a combination of theoretical and numerical analysis, we show our main result: Demand learning has a similar effect as risk pooling on inventory allocation policies, as it provides an incentive for the decision maker to withhold inventory at the warehouse rather than allocating it in earlier periods.

Chapter 2

Dynamic Learning and Pricing with Model Misspecification

We study a multi-period dynamic pricing problem with contextual information where the seller uses a misspecified demand model. The seller sequentially observes past demand, updates model parameters, and then chooses the price for the next period based on time-varying features. We show that model misspecification leads to correlation between price and prediction error of demand per period, which in turn leads to inconsistent price elasticity estimate and hence suboptimal pricing decisions. We propose a "random price shock" (RPS) algorithm that dynamically generates randomized price shocks to estimate price elasticity while maximizing revenue. We show that the RPS algorithm has strong theoretical performance guarantees, that it is robust to model misspecification, and that it can be adapted to a number of business settings, including (1) when the feasible price set is a price ladder, and (2) when the contextual information is not IID. We also perform numerical experiments gauging the performance of RPS on synthetic data, and find that it significantly outperforms competing algorithms that do not account for price endogeneity.

2.1 Introduction

Motivated by the growing availability of data in many revenue management applications, we consider a dynamic pricing problem for a data-rich environment. In such an environment, a firm (i.e., seller) observes some time-varying *contextual information* or *features* that encode external information. The firm estimates demand as a function of both price and features, and chooses price to maximize revenue. By including features into demand models, the firm can potentially obtain more accurate demand forecasts and achieve higher revenues.

In this work, we are especially interested in the consequences of *model misspecification*, namely, when the firm assumes an incorrect demand function on features. In practice, features may contain various kinds of information about demand such as product characteristics, customer types, and economic conditions of the market. A mixed set of heterogeneous features can affect demand in a complex way. The seller may assume an incorrect demand model either because it is unsure how demand is affected by features, or because it prefers a simple model for analytical tractability. In fact, several recent works on dynamic pricing with features make the assumption that demand is a *linear* or *generalized linear* function of features (Cohen et al., 2016; Qiang and Bayati, 2016; Javanmard and Nazerzadeh, 2016; Ban and Keskin, 2017).

We observe that when the demand model is misspecified, model parameters estimated from demand data may become biased and inconsistent. This phenomenon is illustrated in Fig. 2-1 below. In this figure, the inner oval represents a parametric family of demand models assumed by the seller. The white "x" mark represents the seller's initial parameter estimation. The triangle mark represents the true model, which lies outside the oval region, since the model assumed by the seller is misspecified. Over time, as the seller collects past demand data and updates demand model parameters, one would expect that the updated parameters would converge to the best approximation of the true model (denoted by a solid "x" dot on the boundary, i.e., the projection of the triangle mark to the oval region). Under some assumptions, the model with the best approximation is also the one associated with the highest

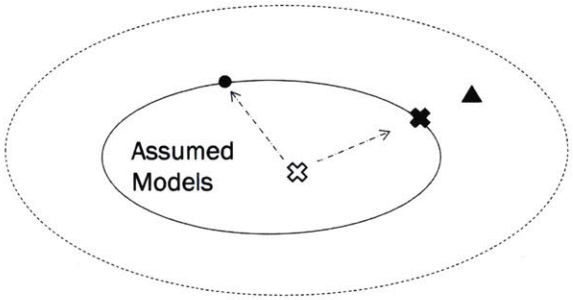


Figure 2-1: The dynamics of parameter estimates under model misspecification.

revenue performance within the assumed model family (see Sometimes, the updated model parameter (circle dot) may even have worse revenue performance than the initial model estimation (white "x")!

The reason why the estimated parameters are inconsistent is that model misspecification can cause correlation between the price and demand prediction error. We refer to this correlation effect as *price endogeneity*. If an estimation method ignores the endogeneity effect and naively treats the assumed model as the true model, it would produce biased estimates. Note that we use the word "endogeneity" here in a pure statistical sense, indicating that an independent variable is correlated with the error term in a linear regression model (Greene, 2003).

In this work, we will mainly focus on the price endogeneity effect caused by model misspecification; however, a discussion of all possible factors that may cause price endogeneity is beyond our scope.

2.1.1 Overview

To illustrate the price endogeneity effect caused by model misspecification, we specifically consider a dynamic pricing setting where the true demand model is *quasi-linear*, in that the expected demand is linear in price but nonlinear with respect to features. The seller does not know the underlying demand function, and incorrectly assumes that the demand function is linear in both price and features.

To address the issue of model misspecification, we propose a "random price shock" (RPS) algorithm that is able to obtain unbiased and consistent estimates of the model parameters while controlling for the price endogeneity effect. The idea of the RPS algorithm is to add random price perturbations to "greedy prices" recommended by some price optimization model using biased parameter estimates. The variances of these price perturbations are carefully controlled by the algorithm to balance the so-called *exploration-exploitation tradeoff*. Intuitively, using a larger variance can help explore and learn the demand function, while using a smaller variance can generate a price that is closer to greedy prices, which can exploit current parameter estimates to maximize revenue.

The RPS algorithm is related to three types of methods in econometrics and operations management for demand estimation. First, the RPS algorithm is in some sense similar to the randomized controlled trials (RCT) method, which offers randomly generated prices to eliminate selection bias. For example, Fisher et al. (2017) applies RCT in a field experiment to estimate an online retailer's demand model. However, it is important to note a key difference between RPS algorithm and the RCT method: the price offered by RPS algorithm is not completely random, because it is the sum of a greedy price, which is endogenous, and a small perturbation. As a result, the sum of the two prices is also endogenous; therefore, standard analysis for randomized control trials cannot be applied to the RPS algorithm. Moreover, Fisher et al. (2017) implemented RCT in two phases: a first phase where random prices are offered, and a second phase where optimized prices are tested. In contrast, the RPS algorithm does not have these two phases, and it estimates the demand model

while optimizing price. The benefit of estimating demand and optimizing price concurrently is discussed in Besbes and Zeevi (2009); Wang et al. (2014). The analysis of the RPS algorithm is also significantly different from that of RCT, and the proof idea for the RPS algorithm is built on the analysis of the least squares method in nonlinear models (Hsu et al., 2014).

The second type of method that is related to the RPS algorithm is the instrumental variables (IV) method. Instrumental variables is a widely used econometric method to obtain unbiased estimates of coefficients of endogenous variables. It aims at finding the so-called instrumental variables that are correlated with endogenous variables but are uncorrelated with prediction error. In the RPS algorithm, the randomly generated price perturbation serves as an instrumental variable, because it is correlated with an endogenous variable, i.e., the actual price offered by the firm (recall that the actual price is the sum of a greedy price and a perturbation), but is obviously uncorrelated with prediction error since it is randomly generated by a computer. This connection to instrumental variables allows us to use econometrics tools in the design of RPS algorithm, more specifically the two-stage least squares (2SLS) method.

Lastly, the RPS algorithm is related to the family of "semi-myopic" pricing policies that has been studied in the revenue management literature more recently (Keskin and Zeevi, 2014; den Boer and Zwart, 2013; Besbes and Zeevi, 2015). A semi-myopic pricing policy keeps track of whether there has been sufficient variations in historical prices; if not, an adjustment is made such that the actual price offered would deviate from greedy or myopic price. Our proposed RPS algorithm belongs to the family of semi-myopic policies. However, it is important to note that most existing semi-myopic algorithms make *deterministic* price adjustments to the greedy prices, whereas the RPS algorithm makes *randomized* price adjustments. This is a major difference since our ability to perform unbiased parameter estimation in the presence of price endogeneity heavily relies on the fact that those price perturbations are randomized.

In Section 2.3, we show that the RPS algorithm accurately identifies the "best" linear approximation to the true quasi-linear model in the presence of model misspecification. The algorithm achieves an expected regret of $O((1+m)\sqrt{T})$ compared to

a clairvoyant who knows the best linear approximation, where m is the dimension of features and T is the number of periods. Our regret bound matches the best possible lower bound of $\Omega(\sqrt{T})$ that *any* non-anticipating algorithm can possibly achieve. Moreover, RPS improves the $O(\sqrt{T}\log T)$ regret bound proven by Keskin and Zeevi (2014) for a special case of linear models without features (i.e., $m = 0$).

Two extensions of the RPS algorithm are considered. In the first extension, we consider the case where prices must be chosen from a discrete set. We establish a $O(T^{2/3})$ regret bound for this generalized setting. In the second extension, we remove the assumption that feature vectors are drawn IID, and allow them to be sampled from an arbitrary distribution. Again, a $O(T^{2/3})$ regret bound is shown.

Finally, we test the numerical performance of the RPS algorithm using synthetic data in Section 2.4. These experiments demonstrate that the RPS algorithm obtains unbiased estimation in the presence of price endogeneity, and shows that it outperforms other pricing algorithms proposed in the literature, which do not account for price endogeneity.

2.1.2 Background and Literature Review

Demand model misspecification is a common problem faced by managers in revenue management practice (Kuhlmann, 2004). Cooper et al. (2006) have discussed several reasons why model misspecification can arise, including revenue managers' lack of understanding of the pricing problem, or their preference for simplified models for the sake of analytical tractability.

Several previous papers study the consequences of model misspecification in dynamic pricing. Cooper et al. (2006) study a problem where an airline revenue manager updates seat protection levels sequentially using historical booking data. The revenue manager incorrectly assumes that customer demand is exogenous and independent, but because the true demands for different fare-classes are substitutable, the booking data is affected by the manager's own control policy. Cooper et al. show that when an incorrect demand model is assumed, the firm's revenues would systematically decrease over time to the worst possible values for a broad class of statistical learning

methods, resulting in a so-called "spiral down effect." On a high level, the spiral-down effect discovered by Cooper et al. (2006) is analogous to the phenomenon we illustrated in Fig. 2-1: As more data is collected, ignoring model misspecification in the estimation process increases bias in parameter estimates over time, and the seller's revenues deteriorates. Besbes and Zeevi (2015) consider a single product dynamic pricing problem in which the seller uses a linear demand function to approximate the unknown, nonlinear true demand function. The authors have proposed a learning algorithm that would converge to the optimal price of the true model. Cooper et al. (2015) consider an oligopoly pricing setting where firms face competition from each other, but their demand models do not explicitly incorporate other firm's decisions. The authors have studied conditions under which the firms' decisions would converge to Nash equilibria. We note that in these three papers, demand function is assumed to be stationary. Instead, we consider a setting where demand function is affected by features, which are changing over time.

The effect of model misspecification on decision making has also been studied in other operations management applications. For example, Dana Jr. and Petruzzi (2001) study a newsvendor problem where the customer demand distribution depends on the inventory stock level chosen by an inventory manager, but the manager incorrectly assumes that demand distribution is exogenous. Cachon and Kök (2007) consider a newsvendor model where the salvage value is endogenously determined by remaining inventory, while the inventory manager assumes the salvage value is exogenous.

We consider a setting where the demand model contains unknown parameters that are being estimated dynamically from sales data. In such a setting, the firm faces an *exploration-exploitation tradeoff*: towards the beginning of the selling season, it may test different prices to learn the unknown parameters; over time, the firm can exploit the parameter estimations to set a price that maximizes revenue. Our problem setting is closely related to the one considered by Keskin and Zeevi (2014). They study a linear demand model without features, and consider a class of semi-myopic algorithms that introduce appropriately chosen deviations to the greedy price

in order to maximize revenue. Keskin and Zeevi show that this class of algorithms has the optimal regret rate, i.e., no other pricing policy can earn higher expected revenue asymptotically (up to a logarithmic factor). Another related paper is den Boer and Zwart (2013), which proposes a quasi-maximum-likelihood-based pricing policy that dynamically controls the empirical variances of the price. Besbes and Zeevi (2009) and Wang et al. (2014) consider dynamic pricing for a single problem under an unknown nonparametric demand model. Besbes and Zeevi (2012) extend the previous result to a setting with multiple products and multiple resources under an unknown nonparametric demand model. For an overview of some of the other problem settings and solution techniques used in dynamic learning and pricing, we refer readers to the recent survey by den Boer (2015).

We are particularly focused on a dynamic learning and pricing problem that contains contextual information (i.e. features). In related work on dynamic pricing with features, Qiang and Bayati (2016) extend the linear demand model in Keskin and Zeevi (2014) to incorporate features, and apply a greedy least squares method to estimate model parameters. Cohen et al. (2016) propose a feature-based pricing algorithm to estimate model parameters when demand is binary. Javanmard and Nazerzadeh (2016) and Ban and Keskin (2017) study pricing problems where feature vector is high dimensional and the demand parameter has some sparsity structure. We note that all these papers assume that demand models are correctly specified. In contrast, we study a feature-based pricing problem where the model is *misspecified*, and focuses on the impact of model misspecification on the seller's revenue. Among these papers, Qiang and Bayati (2016) and Ban and Keskin (2017) are closest to our work as they both consider linear demand models with features. Nevertheless, due to the differences in model assumptions, the regret bounds in Qiang and Bayati (2016) ($O(\log T)$), Ban and Keskin (2017) ($O(\sqrt{T}\log T)$) and this paper ($O(\sqrt{T})$) cannot be directly compared. In particular, Qiang and Bayati (2016) make an "incumbent price" assumption, which gives the firm more information initially and allows the firm to achieve a much lower regret bound of $O(\log T)$ rather than $O(\sqrt{T})$.

We note that a few recent papers apply nonparametric statistical learning ap-

proaches to pricing with features in a batch learning setting where historical data are given as input (Chen et al., 2015; Bertsimas and Kallus, 2016). These works differ from ours in that we focus on a dynamic, multi-period setting. As stated in Van Ryzin and McGill (2000) and Cooper et al. (2006), in revenue management practice, there is usually a repeated process where controls (e.g., booking limits or prices) are enacted, new data are observed, and parameter estimates are updated. In this work, we are specifically interested in the case where historical data is dynamically generated the seller's pricing decisions. In addition, although nonparametric approaches avoid model misspecification, parametric models are widely used in revenue management practice (Kuhlmann, 2004; Cooper et al., 2006; Besbes and Zeevi, 2015), so the consequence of model misspecification remains highly relevant to revenue management practice.

As mentioned earlier, model misspecification can cause price endogeneity, because the demand prediction error and the seller's pricing decisions are both determined endogenously by the feature vector. More generally, the phenomenon of price endogeneity are extensively studied in economics, marketing, and operations management. Empirical studies have found that price endogeneity exists and has a significant impact on price elasticity estimation in many real-world business settings (Bijmolt et al., 2005). The econometrics literature has proposed various methods to identify model parameters with endogeneity effect (e.g. Greene, 2003; Angrist and Pischke, 2008); Talluri and Van Ryzin (2005) also provides an overview of these methods with revenue management applications. The price endogeneity effect has been studied in settings with consumer choice (Berry et al., 1995), consumer strategic behavior (Li et al., 2014), and competition (Berry et al., 1995; Li et al., 2016); these factors are beyond the scope of this paper. We note that empirical revenue management studies often take the perspective of an econometrician who is outside the firm and does not observe all the information that revenue managers can observe, such as cost, product characteristics, consumer features, etc. (e.g. Phillips et al., 2015). However, we take the perspective of a revenue manager within the firm who makes pricing decisions, much like in Cooper et al. (2006) and Besbes and Zeevi (2015). We show that even if a

decision maker observes all the past pricing decisions, untruncated historical demand and contextual information, price endogeneity can still arise when the seller assumes an incorrect model.

Notation

For two sequences $\{a_n\}$ and $\{b_n\}$ $(n = 1, 2, \ldots)$, we write $a_n = O(b_n)$ if there exists a constant C such that $a_n \leq Cb_n$ for all n; we write $a_n = \Omega(b_n)$ is there exists a constant c such that $a_n \geq cb_n$ for all n. All vectors in this chapter are understood to be column vectors. For any vector $\mathbf{x} \in \mathbb{R}^k$, we denote its transpose by $\mathbf{x}^\mathsf{T}$ and denote its Euclidean norm by $\|\mathbf{x}\| := \sqrt{\mathbf{x}^\mathsf{T}\mathbf{x}}$. We let $\|\mathbf{x}\|_1$ be the ℓ_1 norm of $\mathbf{x}$, defined as $\|\mathbf{x}\|_1 = \sum_i |\mathbf{x}_i|$. We let $\|\mathbf{x}\|_\infty$ be the ℓ_∞ norm, defined as $\|\mathbf{x}\|_\infty = \max_i |\mathbf{x}_t|$. For any square matrix $M \in \mathbb{R}^{k \times k}$, we denote its transpose by M^T, its inverse by M^{-1} and its trace by $\mathrm{tr}(M)$; if M is also symmetric $(M = M^\mathsf{T})$, we denote its largest eigenvalue by $\lambda_{\max}(M)$ and its smallest eigenvalue by $\lambda_{\min}(M)$. We let $\|M\|_2$ be the spectral norm of matrix M, defined by $\|M\|_2 = \sqrt{\lambda_{\max}(M^\mathsf{T}M)}$. We denote the Frobenius norm of M by $\|M\|_F$, namely $\|M\|_F = \sqrt{\mathrm{tr}(M^\mathsf{T}M)}$.

2.2 Model

We consider a firm (seller) selling a single product over finite horizon. At the beginning of each time period $(t = 1, 2, \ldots, T)$, the seller observes a feature vector, $\mathbf{x_t} \in \mathbb{R}^m$, which represents exogenous information that may affect demand in the current period. We assume that feature vectors $\mathbf{x_t}$ are sampled independently for $t = 1, 2, \ldots, T$ from a fixed but unknown distribution with bounded support. (In Section 2.3.5, we will relax the IID assumption on the features and assume an arbitrary sequence of random feature vectors.) Without loss of generality, we assume $\mathbf{x_t} \in [-1, 1]^m$ after appropriate scaling. Moreover, we assume that the matrix

$$M = \mathrm{E}\left[\begin{bmatrix} 1 & \mathbf{x_t}^\mathsf{T} \\ \mathbf{x_t} & \mathbf{x_t}\mathbf{x_t}^\mathsf{T} \end{bmatrix}\right]$$

is positive definite.[1]

Given the feature vector $\mathbf{x_t}$, customer demand for period t as a function of price p is given by

$$D_t(p) = bp + f(\mathbf{x_t}) + \epsilon_t, \quad \forall p \in [\underline{p}_t, \bar{p}_t]. \tag{2.1}$$

Here, parameter b is a constant representing price sensitivity of customer demand, and $f : \mathbb{R}^m \to \mathbb{R}$ is a function that measures the effect of features on customer demand. Both b and f are *unknown* to the seller. We assume that the demand function is strictly decreasing in price p (i.e. $b < 0$), and $f(\mathbf{x_t})$ is bounded for all $\mathbf{x_t}$ such that $|f(\mathbf{x_t})| \leq \bar{f}$. The latter assumption would follow immediately from the fact that the set of all features $\mathbf{x_t}$ is compact if f were continuous. The last term ϵ_t in Eq (2.1) represents a demand noise. Without loss of generality, we assume ϵ_t has zero mean conditional of $\mathbf{x_t}$: $\mathrm{E}[\epsilon_t \mid \mathbf{x_t}] = 0$; otherwise, the conditional mean $\mathrm{E}[\epsilon_t \mid \mathbf{x_t}]$ can be shifted into function $f(\mathbf{x_t})$. We assume that ϵ_t has bounded second moment ($\mathrm{E}[\epsilon_t^2] \leq \sigma^2, \forall t$), and is independent of historical data $(\mathbf{x_s}, \epsilon_s)$ for all $1 \leq s \leq t-1$. However, the distribution of ϵ_t is allowed to vary over time. We refer to Eq (2.1) as a *quasi-linear demand model*, since the demand function is linear with respect to price, but is possibly nonlinear with respect to features.

We denote the admissible price range in period t, i.e. the range of prices from which the price p must be chosen, by $[\underline{p}_t, \bar{p}_t]$. In particular, we allow the admissible price interval to vary over time. We assume that $\underline{p}_t$ and $\bar{p}_t$ are inputs to the seller's decision problem, while they may be arbitrarily correlated with features $\mathbf{x_t}$ and demand noise ϵ_t. We also assume there exist constants $\delta > 0$ and $p_{\max}$ such that $\bar{p}_t \leq p_{\max}$ and $\bar{p}_t - \underline{p}_t \geq \delta$ for all t. Given features x_t, we denote the optimal price for the true demand model (as a function of x_t) by $\tilde{p}_t(\mathbf{x_t}) = -\frac{f(\mathbf{x_t})}{2b}$. We assume that the optimal price $\tilde{p}_t(\mathbf{x_t}) \in [\underline{p}_t, \bar{p}_t]$ for all t.

[1]This assumption is equivalent to the condition of "no perfect collinearity," i.e., no variable in the feature vector can be expressed as an affine function of the other variables. If matrix M is not positive definite, the dimension of feature vector can be reduced by replacing certain variable as a combination of other variables.

2.2.1 Applications of the Model

The above model has applications in several business settings that involve feature-based dynamic pricing. One example is dynamic pricing for fashion retail, which will be discussed in more detail in Chapter 3. In the fashion retail setting, a retail manager dynamically sets prices for fashion items throughout a selling season, while the demand is highly uncertain when the selling season begins. The feature vectors represent the characteristics of fashion items, such as color and design pattern, as well as seasonality variables. Throughout the season, the retail manager may learn from sales data about how customer demand varies for different product features, and adjust prices accordingly to maximize revenue. Another example of an application of feature-based pricing is personalized financial services. Phillips et al. (2015) describe a setting in the auto loan context, where the price (interest rate for a loan) is adjusted based on features such as credit score of the buyer, the amount and term of the loan, the type of vehicle purchased, etc. They find that using a centralized, data-driven pricing algorithm could improve profits significantly over the current practice, where local salespeople are granted discretion to negotiate price. In our model, the admissible price interval $[\underline{p}_t, \bar{p}_t]$ is allowed to vary for different periods. For example, in the auto loan context, the price interval represents the range of admissible interest rates set by the financial headquarters, which varies based on the amount and term of the loan offered. As time-varying bounds may depend on features and demand noise, our model makes no assumption of the distribution of price range $[\underline{p}_t, \bar{p}_t]$, and allows the price bounds to be arbitrarily correlated with past prices, feature vector $\mathbf{x_t}$, and noise ϵ_t. If such a correlation is present, it will lead to price endogeneity (in addition to the price endogeneity caused by model misspecification) and will be accounted for in our pricing algorithm.

2.2.2 Model Misspecification and Non-anticipating Pricing Policies

We consider a seller who is either unaware that the true demand function has a nonlinear dependence on features, or is unsure how to model such dependence. As a result, the seller uses a misspecified *linear* demand function to approximate the true quasi-linear demand function given by Eq (2.1). The seller assumes a linear demand model as

$$D_t(p) = a + bp + \mathbf{c}^\mathsf{T}\mathbf{x_t} + \nu_t, \quad \forall p \in [\underline{p}_t, \bar{p}_t], \tag{2.2}$$

where $a \in \mathbb{R}$ and $\mathbf{c} \in \mathbb{R}^m$ are constants and ν_t is an error term.

We focus on the linear demand model, because the linear model and its variations are widely used in revenue management practice and in the demand learning literature (Qiang and Bayati, 2016; Ban and Keskin, 2017); in addition, the model can capture nonlinear factors in the feature vector by including higher order terms in the feature vector.

The parameters (a, b, c) are *unknown* to the seller at the beginning of the selling season. We assume that the seller knows that the parameters a and $\mathbf{c}$ are bounded, and that there exist $\bar{a}$, $\bar{c}$ such that $|a| \leq \bar{a}$ and $\|\mathbf{c}\|_1 \leq \bar{c}$, but do not assume that the seller knows the values of $\bar{a}$, $\bar{c}$. As for the price sensitivity parameter b, we assume that the seller knows not only that the parameter b is bounded, but also the *range* within which b lies, $0 < \underline{b} \leq |b| \leq \bar{b}$. The assumption that the range of b is known to the seller is strong, and is indeed a limitation of our model. However, there are applications for which it may be reasonable to assume that the seller has some knowledge about this range, perhaps from her prior experience with the sales of similar items during previous selling seasons. For example, in our case study in Chapter 3, where we apply our demand model to a fashion retail setting, our estimates of b for different categories of fashion items were found to be of the same order of magnitude, lying in the range $[-1, -0.1]$. Thus the seller could assume that b lies in the range $[-1, -0.1]$ for future selling seasons. More generally, the economics and marketing literature finds that price elasticity, a quantity related to our price sensitivity parameter, tends

to fall within finite ranges across markets and products. Bijmolt et al. (2005), for example, analyze 1851 price elasticities from 81 different publications between 1961 and 2004, across different products, markets and countries. They observe a mean price elasticity of -2.62 and find that the distribution is strongly peaked, with 50 percent of the observations between -1 and -3.

The seller must select a price $p_t \in [\underline{p}_t, \bar{p}_t]$ for each period $t = 1, 2, \ldots, T$ sequentially while estimating the values of $(a, b, \mathbf{c})$ using realized demand data. The seller's objective is to maximize her total expected revenue over T periods.

We denote the realized demand given p_t by d_t, defined as

$$d_t = D_t(p_t) = bp_t + f(x_t) + \epsilon_t.$$

Note that the realized demand is generated from the true model, i.e., the quasi-linear model Eq (2.1).

The history up to the end of period $t-1$ is defined as

$$\mathcal{H}_{t-1} = (x_1, p_1, \epsilon_1, \ldots, x_{t-1}, p_{t-1}, \epsilon_{t-1}).$$

We say that π is a non-anticipating pricing policy if for any t, price p_t is a measurable function with respect to $\mathcal{H}_{t-1}$ and the current feature vector and the feasible price range: $p_t = \pi(\mathcal{H}_{t-1}, x_t, \underline{p}_t, \bar{p}_t)$. The seller cannot foresee the future and is restricted to using non-anticipating pricing policies.

2.2.3 Price Endogeneity Caused by Model Misspecification and Other Factors

By comparing the true quasi-linear demand model (cf. Eq (2.1)) and the misspecified linear model (cf. Eq (2.2)), it is easily verified that the error term in the misspecified linear model is equal to $\nu_t = f(\mathbf{x_t}) - (a + \mathbf{c}^\top \mathbf{x_t}) + \epsilon_t$. This error term ν_t is composed of an approximation error, $f(\mathbf{x_t}) - (a + \mathbf{c}^\top \mathbf{x_t})$, which is correlated with features $\mathbf{x_t}$, and a random noise ϵ_t, which is uncorrelated with the features. When the model

is misspecified, we have $f(\mathbf{x_t}) - (a + \mathbf{c}^\mathsf{T}\mathbf{x_t}) \neq 0$, so the error term ν_t is not mean independent of feature $\mathbf{x_t}$, namely $\mathrm{E}[\nu_t \mid \mathbf{x_t}] \neq 0$.

The fact that the error term is not mean independent of the features could cause bias in the seller's demand estimates if the estimation procedure is not designed properly. Suppose the seller uses a non-anticipating pricing policy π such that

$$p_t = \pi(\mathcal{H}_{t-1}, \mathbf{x_t}, \underline{p}_t, \bar{p}_t). \tag{2.3}$$

Because the error term ν_t is correlated with the features $\mathbf{x_t}$ while price p_t is a function of $\mathbf{x_t}$, the seller's pricing decision causes a correlation between ν_t and p_t. More specifically, we have $\mathrm{E}[\nu_t p_t] \neq 0$ since $\mathrm{E}[\nu_t p_t \mid \mathbf{x_t}] = \mathrm{E}[\nu_t \cdot \pi(\mathcal{H}_{t-1}, \mathbf{x_t}, \underline{p}_t, \bar{p}_t) \mid \mathbf{x_t}] \neq 0$. We refer to the correlation between p_t and the error term ν_t as the price endogeneity effect, and refer to p_t as the endogenous variable. Throughout this chapter, the word "endogeneity" is used in a pure econometric sense to indicate the correlation between p_t and ν_t.

It is well known that in a linear regression model

$$d_t = a + bp_t + \mathbf{c}^\mathsf{T}\mathbf{x_t} + \nu_t, \tag{2.4}$$

when the regressor p_t is endogenous, naive estimation methods such as ordinary least squares (OLS) would give biased and inconsistent estimates of parameters $(a, b, \mathbf{c})$.

Biased estimates of model parameters then lead to suboptimal pricing decisions. Moreover, the seller cannot test whether price p_t and error ν_t are correlated using historical data, since she does *not* observe the error term ν_t directly; even if the seller has complete historical data, without knowing the values of $(a, b, \mathbf{c})$, the term ν_t cannot be computed.

In addition to model misspecification, other factors can also cause the price endogeneity effect. If a manager believes she has expert knowledge about future demand, she may set the price range $[\underline{p}_t, \bar{p}_t]$ in anticipation of future demand, so the price bounds $\underline{p}_t, \bar{p}_t$ are endogenous. Because our pricing algorithm chooses price p_t in $[\underline{p}_t, \bar{p}_t]$, the price p_t also becomes endogenous. In our algorithm proposed in Sec-

tion 2.3, we account for such endogeneity by allowing the price bounds $\underline{p}_t, \bar{p}_t$ to be correlated with noise ϵ_t.

The endogeneity problem has been extensively studied in the econometrics literature (Greene, 2003; Angrist and Pischke, 2008). There are a few key differences between the pricing model considered in this work and typical research problems studied in econometrics. First, we study a pricing problem from the perspective of a *firm* that wants to maximize its revenue, whereas econometricians often take the perspective of a researcher who is *outside the firm* and wants to estimate causal effects of model parameters. The second key difference is that econometrics and empirical studies often consider *batch* data, whereas our pricing model considers *sequential* data generated from dynamic pricing decisions. Analyzing these two types of data usually requires different statistical methods and correspondingly different performance metrics.

Although there are differences between the problem considered in this work and those in the econometrics and empirical literature, a common challenge is in studying a regression model with endogenous independent variables. In fact, the dynamic pricing algorithm that we introduce in the next section is inspired by statistical tools in econometrics such as instrumental variables and two-stage least squares.

2.3 Random Price Shock Algorithm

In this section, we propose a dynamic pricing algorithm which we call the *random price shock* (RPS) algorithm. The idea behind the RPS algorithm is that the seller can add a random price shock to the greedy price obtained from the current parameter estimates. As the number of periods (T) grows, the parameters estimated by the RPS algorithm are guaranteed to converge to the "best" parameters within the linear demand model family, which we will define shortly in Section 2.3.1. Therefore, the prices chosen by the RPS algorithm will also converge to the optimal prices under the misspecified linear demand model.

We present the RPS algorithm below (Algorithm 1). The RPS algorithm starts

Algorithm 1 Random Price Shock (RPS) algorithm.

input: parameter bound on b, $B = [-\bar{b}, -\underline{b}]$
initialize: set $\hat{a}_1 = 0, \hat{b}_1 = -\bar{b}, \hat{\mathbf{c}}_1 = 0$
for $t = 1, \ldots, T$ **do**
  set $\delta_t \leftarrow \frac{\delta}{2} t^{-\frac{1}{4}}$
  given x_t, set unconstrained greedy price: $p^u_{g,t} \leftarrow -\frac{\hat{a}_t + \hat{\mathbf{c}}_t^\intercal \mathbf{x}_t}{2\hat{b}_t}$
  project greedy price: $p_{g,t} \leftarrow \text{Proj}(p^u_{g,t}, [\underline{p}_t + \delta_t, \bar{p}_t - \delta_t])$
  generate an independent random variable $\Delta p_t \leftarrow \delta_t$ w.p. $\frac{1}{2}$ and $\Delta p_t \leftarrow -\delta_t$ w.p. $\frac{1}{2}$
  set price $p_t \leftarrow p_{g,t} + \Delta p_t$
  choose an arbitrary price $p_t \in [\underline{p}_t, \bar{p}_t]$
  observe demand $d_t = D_t(p_t)$
  set $\hat{b}_{t+1} \leftarrow \text{Proj}(\frac{\sum_{s=1}^t \Delta p_s d_s}{\sum_{s=1}^t \Delta p_s^2}, B)$
  set $(\hat{a}_{t+1}, \hat{\mathbf{c}}_{t+1}) \leftarrow \arg\min \sum_{s=1}^t (d_s - \hat{b}_{t+1} p_s - \alpha - \gamma^\intercal \mathbf{x}_s)^2$
end for

each period by choosing a perturbation factor δ_t. The algorithm computes the greedy price, $p_{g,t}$, and adds it to a random price shock, Δp_t. Note that the greedy price is projected to the interval $[\underline{p}_t + \delta_t, \bar{p}_t - \delta_t]$, so that the sum of greedy price and price shock is always in the feasible price range $[\underline{p}_t, \bar{p}_t]$. (We denote the projection of a point $\mathbf{x}$ to a set S by $\text{Proj}(\mathbf{x}, S) = \arg\min_{\mathbf{x}' \in S} \|\mathbf{x} - \mathbf{x}'\|$.) The interval $[\underline{p}_t + \delta_t, \bar{p}_t - \delta_t]$ is non-empty, since $\bar{p}_t - \underline{p}_t - 2\delta_t \geq \delta(1 - t^{-1/4}) \geq 0$. The price shock is generated independently of the feature vector and the demand noise (e.g., it can be a random number generated by a computer).

After the demand in period t is observed, the algorithm updates parameter estimations by a *two-stage least squares* procedure. First, the price parameter b is estimated by applying linear regression for d_t against Δp_t. It is important to note that we cannot estimate b by regressing d_t against the actual price p_t, since p_t may be endogenous and correlated with demand noise. Since the random price shock Δp_t is correlated with the actual price p_t but uncorrelated with demand noise, we can view it as an *instrumental variable*. Therefore, this step allows an unbiased estimate of parameter b. The second stage estimates the remaining parameters, a and $\mathbf{c}$.

In the RPS algorithm, the variance of the price shock introduced at each time period (Δp_t) is an important tuning parameter. Intuitively, choosing a large variance

of Δp_t generates large price perturbations, which can help the seller learn demand more quickly; choosing a small variance means that the actual price offered would be closer to the greedy price, which allows the seller to earn more revenue if the greedy price is close to the optimal price. The tradeoff between choosing a large price perturbation versus a small price perturbation illustrates the classical "exploration-exploitation" tradeoff faced by many dynamic learning problems. In Algorithm 1, the variances of the price shocks are set as $O(t^{-\frac{1}{2}})$ to balance the exploration-exploitation tradeoff and control the performance of the algorithm.

We would like to make two remarks about the RPS algorithm. First, the idea of adding time-dependent price perturbations to greedy prices has also been used in Besbes and Zeevi (2015). However, there is a fundamental difference between the price shocks introduced in RPS algorithm and the price perturbations in Besbes and Zeevi (2015), which assumes a fixed (unknown) demand function. The algorithm proposed in Besbes and Zeevi (2015) separates the time horizon into cycles and requires testing *two* prices in each cycle: a greedy price (say, p_g) and a perturbed price (say, $p_g + \Delta p$). Observed demand under the two prices is then used to estimate price elasticity. This strategy of testing two prices is not applicable when demand function depends on feature vectors, because demand is constantly changing as features are randomly sampled. As a result, the RPS algorithm can only test *one* price for each realized demand function, since the demand functions in future periods may vary. That is, the RPS algorithm only observes demand under price $p_g + \Delta p$, but not p_g.

Second, one may ask why the RPS algorithm is concerned with the correlation between the price p_t and the error term ν_t, but ignores the correlation between feature vector $\mathbf{x_t}$ and error term ν_t. Indeed, the error term ν_t contains an approximation part $f(\mathbf{x_t}) - (a + \mathbf{c}^\top \mathbf{x_t})$ due to model misspecification, so the least squares parameters a and $\mathbf{c}$ will be biased if $\mathbf{x_t}$ and ν_t are correlated. The reason why we can ignore the correlation between $\mathbf{x_t}$ and ν_t in pricing decisions is that computing an optimal price for the linear model, namely $-(a + \mathbf{c}^\top \mathbf{x_t})/(2b)$, only requires an unbiased estimate of the *aggregated* effect of the feature vector on demand, which is measured by the numerator $a + \mathbf{c}^\top \mathbf{x_t}$ rather than the treatment effect of each individual component

of $\mathbf{x_t}$. Unbiased estimates of $a + \mathbf{c}^\top \mathbf{x_t}$ and b can be provided by the RPS algorithm. However, when $\mathbf{x_t}$ is endogenous, the RPS cannot guarantee that the estimates of $a, \mathbf{c}$ are unbiased component-wise.

2.3.1 Performance Metric and Regret Bound

To analyze the performance of Algorithm 1, let us first define *regret* as the performance metric. Recall that the true demand function is given by

$$D_t(p) = bp + f(\mathbf{x_t}) + \epsilon_t, \quad \forall t = 1, \ldots, T \tag{2.5}$$

where both b and $f(\cdot)$ are unknown to the seller. We would like to compare the performance of our algorithm to that of a clairvoyant who knows the true model a priori. However, it can be shown that the optimal revenue of the true model cannot be achieved when the seller is restricted to use *linear* demand models, because the optimal price $\tilde{p}_t(\mathbf{x_t}) = -\frac{f(\mathbf{x_t})}{2b}$ cannot be expressed as an affine function of $\mathbf{x_t}$. In Appendix A.1, we show that if the sequence of p_t for $t = 1, \ldots, T$ is chosen based on linear demand models, the model misspecification error is quantified by

$$\mathrm{E}\left[\sum_{t=1}^{T} \tilde{p}_t(\mathbf{x_t}) D(\tilde{p}_t(\mathbf{x_t})) - \sum_{t=1}^{T} p_t D(p_t)\right] = \Omega(T).$$

Therefore, the optimal revenue of the true model is not an informative benchmark, since no algorithm can achieve a sublinear ($o(T)$) regret rate with a misspecified model.

If the $\Omega(T)$ misspecification error is large, a first order concern would of course be to find a better demand model family than the current linear model in order to reduce the misspecification error. However, even if the seller uses other parametric models, the revenue gap to the true model can always grow as $\Omega(T)$, as the same argument in Appendix A.1 applies to any parametric model because it is always possible that the parametric model is misspecified.

We then consider the revenue of a linear model that is the "projection" of the true

model to the linear model family. Ideally, if the true model is well-approximated by a linear model, the seller will be able to achieve near optimal revenue even though it uses a misspecified model. Given a nonlinear function f, we define the following linear demand model:

$$D_t(p) = a + bp + \mathbf{c}^\top \mathbf{x_t} + \nu_t, \quad \forall p \in [\underline{p}_t, \bar{p}_t], \tag{2.6}$$

where a and $\mathbf{c}$ are population least squares estimates of $f(\mathbf{x_t})$:

$$a, \mathbf{c} = \arg\min_{\alpha,\gamma} \mathrm{E}[\|f(\mathbf{x_t}) - (\alpha + \gamma^\top \mathbf{x_t})\|^2].$$

It can be shown by solving first order conditions that $a, \mathbf{c}$ are given by the closed form expression:

$$\begin{bmatrix} a \\ \mathbf{c} \end{bmatrix} = \left(\mathrm{E}\left[\begin{bmatrix} 1 & \mathbf{x_t}^\top \\ \mathbf{x_t} & \mathbf{x_t}\mathbf{x_t}^\top \end{bmatrix} \right] \right)^{-1} \mathrm{E}\left[\begin{bmatrix} f(\mathbf{x_t}) \\ f(\mathbf{x_t})\mathbf{x_t} \end{bmatrix} \right]. \tag{2.7}$$

One can view linear model (2.6) as the projection of the true quasi-linear model (2.1) to the linear model family (see Fig. 2-1). Let $p_t^*(\mathbf{x_t}) = -\frac{a+\mathbf{c}^\top \mathbf{x_t}}{2b}$ be the optimal price under the best linear model given by Eq (2.6). The proposition below shows that the linear demand function (2.6) gives the highest revenue among all linear demand functions. Therefore, we will call it the *best linear model.*

Proposition 1. *For any period t, consider a price $p_t' = -\frac{\alpha+\gamma^\top \mathbf{x_t}}{2\beta}$ that is affine in features $\mathbf{x_t}$, where α, β, γ are measurable with respect to history $\mathcal{H}_{t-1}$. Then, the revenue under price p_t' is upper bounded by the revenue under $p_t^*(\mathbf{x_t})$, namely*

$$\mathrm{E}[p_t^*(\mathbf{x_t}) D_t(p_t^*(\mathbf{x_t}))] - \mathrm{E}[p_t' D_t(p_t')] = -b\mathrm{E}[(p_t^*(\mathbf{x_t}) - p_t')^2] \geq 0.$$

By Proposition 1, if the seller uses a linear demand model $D_t'(p) = \alpha + \beta p + \gamma^\top \mathbf{x_t}$ for period t, the expected revenue of its optimal price $p_t' = -\frac{\alpha+\gamma^\top \mathbf{x_t}}{2\beta}$ is maximized when $p_t' = p_t^*$.

We now define the seller's *regret* as the difference in the cumulative expected revenue of a clairvoyant who uses the best linear model and the expected revenue achieved by an admission pricing policy, namely

$$\text{Regret}(T) = \sum_{t=1}^{T} \mathrm{E}[p_t^*(\mathbf{x_t}) D(p_t^*(\mathbf{x_t}))] - \sum_{t=1}^{T} \mathrm{E}[p_t D(p_t)], \tag{2.8}$$

where the expectation is taken over all random quantities including features x_t, price ranges $[\underline{p}_t, \bar{p}_t]$, demand noise ϵ_t, and possibly external randomization used in the pricing policy.

To reiterate, in the definition of regret in Eq (2.8), we use the optimal price of the best linear model (2.6), $p_t^*(\mathbf{x_t})$, instead of the absolute optimal price for the true quasi-linear model (2.5), $\tilde{p}_t(\mathbf{x_t})$. The reason is that the optimal price $p_t^*(\mathbf{x_t})$ of model (2.6) gives the highest *achievable* revenue if the seller is restricted to making pricing decisions using linear demand models. Should we replace $p_t^*(\mathbf{x_t})$ by $\tilde{p}_t(\mathbf{x_t})$ in the definition of regret in (2.8), the benchmark would be too strong to be achieved by any linear model, and the regret would grow linearly in T no matter which pricing policy is used.

2.3.2 A Upper Bound of Regret

We now prove the following regret upper bound for the RPS algorithm.

Theorem 1. *Under the quasi-linear demand model in Eq* (2.1)*, the regret of Algorithm 1 over a horizon of length T is $O(\frac{m+1}{\lambda_{\min}(M)}\sqrt{T})$.*

Theorem 1 expresses the upper bound on regret in terms of the horizon length T, the dimension of features m, and the minimum eigenvalue of the design matrix $M = \mathrm{E}[(1, \mathbf{x_t})(1, \mathbf{x_t})^\intercal]$, while the constant factor within the big O notation only depends on model parameters $\underline{b}, \bar{b}, \sigma^2$ and $p_{\max}$. We note that the constant factor does not depend on the unknown values of a, b, c, or the unknown distribution of $\mathbf{x_t}$ except through the parameter $\lambda_{\min}(M)$. The proof of Theorem 1, which is deferred to Appendix A.3.2, shows explicitly how the regret depends on these parameters.

The main idea behind the proof of Theorem 1 is to decompose the regret into the loss in revenue due to adding random price shocks, and the loss in revenue due to parameter estimation errors. Since the randomized price shocks have variance $O(t^{-1/2})$ at period t, the former part is bounded by $O(\sqrt{T})$. The latter part can be bounded in terms of the expected difference between the true parameters a, b, c and the estimated parameters. We then modify results on linear regression in the random design case (Hsu et al., 2014) to prove that the estimated parameters converge sufficiently quickly to their true values.

Theorem 1 shows that the RPS algorithm is robust to model misspecification: Even if the true demand model is nonlinear in features, the RPS algorithm is guaranteed to converge to the best linear demand model (2.6), which gives the highest expected revenue among all linear models. The RPS algorithm achieves such robustness because it correctly addresses the price endogeneity effect introduced by model misspecification.

Theorem 1 immediately invites comparison with the upper bound in Keskin and Zeevi (2014)). Keskin and Zeevi (2014) consider a linear demand model without features and fixed price bounds. They propose a family of "semi-myopic" pricing policies that ensure the price selected at any period is both sufficiently deviated from the historical average of prices and sufficiently close to the greedy price. They show that such policies attain a worst case regret of at most $O(\sqrt{T}\log T)$. Since the model in Keskin and Zeevi (2014) is a special case of demand model (2.1) with $f(x_t) = 0$, the result for the RPS algorithm in Theorem 1 thus improves the upper bound in Keskin and Zeevi (2014) by a factor of $\log T$. In addition, as we have already noted, the RPS algorithm can be applied to a broader setting with features and price endogeneity.

2.3.3 A Lower Bound on Regret

The upper bound on the regret of the RPS algorithm scales with $O(\sqrt{T})$ as the number of period T grows. We can prove a corresponding lower bound on the regret of any admissible pricing policy.

Theorem 2. *The regret of any non-anticipating pricing policy over a selling horizon of length T is $\Omega(\sqrt{T})$.*

The proof of Theorem 2 is given in Appendix A.3.3. This theorem relies on a Van Trees inequality-based proof technique (Gill and Levit, 1995), and is related to the lower bound of $\Omega(\sqrt{T})$ described by Keskin and Zeevi (2014) on the regret of any non-anticipating pricing policy in the special case of our model where $m = 0$ (i.e., there are no features) and the demand model is linear (i.e., model is correctly specified). Theorem 2 extends the result of Keskin and Zeevi (2014) to the case where $m > 0$, showing that the regret lower bound does not change in terms of T even in the presence of features. Further, Theorem 2 shows that the regret of the RPS algorithm is optimal in terms of T.

Note that the lower bound in Theorem 2 does not depend on the dimension of the feature vector m. The upper bound in Theorem 1, however, grows with m, and our numerical experiments show the regret usually increase with m (see Appendix A.2.2). We conjecture that the RPS algorithm's dependence on m is due to the two-stage least squares procedure needed to obtain an unbiased estimate of the price coefficient b. We leave it as future work to close the gap between upper and lower bounds.

2.3.4 Price Ladder

A common business constraint faced by retailers is that prices must be selected from a *price ladder* rather than from a continuous price interval. A price ladder consists of a discrete set of prices that are typically fairly evenly spaced apart. For example, a firm may use prices such as \$9.99, \$19.99, \$29.99, etc., because these prices are familiar to customers and easy to understand. In this section, we show how the RPS algorithm and theoretical results can be adapted to the setting where prices are drawn from a price ladder rather than from price intervals.

We model this setting as follows. Suppose that the seller is interested in selecting prices from the price ladder $\{q_1, \ldots, q_N\}$ where $N \geq 2$ and $q_1 < \cdots < q_N$. Assume that for the purposes of price experimentation, she is also allowed to use two additional

prices q_0, q_{N+1} such that $0 < q_0 < q_1$ and $p_{\max} > q_{N+1} > q_N$. Then at each time period t, the selected price satisfies $p_t \in \{q_0, q_1 \ldots, q_N, q_{N+1}\}$ where $N \geq 2$ and $q_0 < q_1 < \cdots < q_{N+1}$. Analogous to our assumption in Section 2.2 on the width of the price intervals, we assume here that $\underline{\delta} \leq q_{i+1} - q_i \leq \bar{\delta}$ for some positive constants $\underline{\delta}, \bar{\delta}$ and all $i = 0, \ldots, N$. The remaining assumptions on features x_t, demand noise ϵ_t and function f are as stated in Section 2.2.

We benchmark the performance of admissible pricing algorithms against a clairvoyant who knows the "best" linear demand model given by (2.6), and selects price

$$p_t^* = \text{Proj}(-(a + \mathbf{c}^\mathsf{T}\mathbf{x_t})/(2b), \{q_1, q_2, \ldots, q_N\})$$

upon observing feature $\mathbf{x_t}$. Then the expected regret, as before, is given by

$$\text{Regret}(T) = \sum_{t=1}^{T} \text{E}[p_t^* D(p_t^*)] - \sum_{t=1}^{T} \text{E}[p_t D(p_t)], \tag{2.9}$$

where the expectation is taken over all random quantities including features x_t and the demand noise ϵ_t.

The RPS algorithm as designed for the price interval setting (Algorithm 1) cannot be directly applied to the case where prices must be drawn from a price ladder. In the experimentation structure of Algorithm 1, price shocks of decreasing magnitude are selected along the selling horizon, violating the price ladder constraint. We thus adapt the RPS algorithm to the price ladder setting by modifying the price experimentation step. Suppose at time period t the estimated greedy price $p_{g,t}$ is $p_{g,t} = q_i$ for some $1 \leq i \leq N$. We perform price experimentation by selecting the price p_t from the set $\{q_{i-1}, q_i, q_{i+1}\}$ with probabilities set to ensure $\Delta p_t = p_t - p_{g,t}$ satisfies $\text{E}[\Delta p_t] = 0$ and $\text{Var}[\Delta p_t]$ is a decreasing function of t. While Algorithm 1 sets Δp_t such that $\text{Var}[\Delta p_t] \propto \frac{1}{\sqrt{t}}$, our modified RPS algorithm sets Δp_t such that $\text{Var}[\Delta p_t] \propto \frac{1}{t^{1/3}}$. This shifts the balance between exploitation and exploration, allowing our modified RPS algorithm to reduce its regret. The full statement of the Random Price Shock (RPS) algorithm for the price ladder setting is given in Algorithm 2.

Algorithm 2 Random Price Shock (RPS) algorithm with price ladder.

input: parameter bound on b, $B = [-\bar{b}, -\underline{b}]$
initialize: choose $\hat{a}_1 = 0, \hat{b}_1 = -\bar{b}, \hat{\mathbf{c}}_\mathbf{1} = \mathbf{0}$
for $t = 1, \ldots, T$ **do**
  given x_t, set unconstrained greedy price: $p^u_{g,t} \leftarrow -\frac{\hat{a}_t + \hat{\mathbf{c}}_\mathbf{t}^T x_t}{2\hat{b}_t}$
  find $i_t = \arg\min_{j \in \{1,\ldots,N\}} |q_j - p^u_{g,t}|$ and set constrained greedy price: $p_{g,t} \leftarrow q_i$
  generate an independent random variable

$$\Delta p_t \leftarrow \begin{cases} q_{i_t} - q_{i_t - 1} \text{ w.p. } \frac{q_{i_t+1} - q_{i_t}}{(q_{i_t+1} - q_{i_t-1})t^{1/3}} \\ q_{i_t+1} - q_i \text{ w.p. } \frac{q_{i_t} - q_{i_t-1}}{(q_{i_t+1} - q_{i_t-1})t^{1/3}} \\ 0 \text{ w.p. } 1 - t^{-1/3} \end{cases}$$

  set price $p_t \leftarrow p_{g,t} + \Delta p_t$
  observe demand $d_t = D_t(p_t)$
  set $\hat{b}_{t+1} \leftarrow \text{Proj}(\frac{\sum_{s=1}^t \Delta p_s d_s}{\sum_{s=1}^t \frac{(q_{i_s} - q_{i_s-1})(q_{i_s+1} - q_{i_s})}{\sqrt{s}}}, B)$
  set $(\hat{a}_{t+1}, \hat{\mathbf{c}}_\mathbf{t+1}) \leftarrow \arg\min \sum_{s=1}^t (d_s - \hat{b}_{t+1} p_s - \alpha - \gamma^\intercal x_s)^2$
end for

We prove the following regret bound for the RPS algorithm with price ladder. As in the previous section, we assume the regret is benchmarked against a linear clairvoyant who uses the optimal price for the best linear approximation given by Eq (2.6).

Theorem 3. *The regret of Algorithm 2 over a selling horizon of length T is $O(\sqrt{\frac{m+1}{\lambda_{\min}(M)}} \cdot T^{2/3})$.*

The proof of Theorem 3 is given in Appendix A.3.4. We can see that when price intervals are replaced with a price ladder, our bound on the regret of the RPS algorithm worsens in terms of T. The intuition is that the clairvoyant's optimal prices $p_t^* = \text{Proj}(-\frac{a + \mathbf{c}^\intercal \mathbf{x_t}}{2b}, \{q_1, q_2, \ldots, q_N\})$ do not satisfy the first-order optimality condition, $\nabla R_t(p_t^*) = 0$, in the price ladder setting. Deviations from the clairvoyant's price are thus more costly, worsening the regret bound.

2.3.5 Non-IID features

Previously, we assumed that the features $\{\mathbf{x_t}\}_{t=1}^T$ are drawn from an IID distribution. This assumption is too strong for some scenarios. For example, when the features

include seasonal variables such as day of the week, day of the month, or month or the year etc., the distribution of x_t is correlated over t and is not IID. In this section, we relax the IID assumption and allow the sequence $\{\mathbf{x_t}\}_{t=1}^T$ to be sampled from an arbitrary distribution on $[-1,1]^m$ (after appropriate scaling). The assumptions on the demand noises ϵ_t, the function f and the price sensitivity parameter b are the same as in Section 2.2.

Since the sequence $\{\mathbf{x_t}\}$ is non-IID, we redefine the regret benchmark as the following linear model:

$$\hat{D}_t(p) = a_x + bp + \mathbf{c_x}^\mathsf{T}\mathbf{x_t}, \quad \forall p \in [\underline{p}_t, \bar{p}_t], \tag{2.10}$$

where a_x and $\mathbf{c_x}$ are defined for an arbitrary sequence of features, $\{\mathbf{x_t}\}_{t=1}^T$, as

$$\begin{bmatrix} a_x \\ \mathbf{c_x} \end{bmatrix} = \arg\min_{a',\mathbf{c}'} \sum_{t=1}^{T} \|f(\mathbf{x_t}) - (a' + \mathbf{c}'^\mathsf{T}\mathbf{x_t})\|^2.$$

It can be shown by solving first order conditions that $a_x, \mathbf{c_x}$ are given by the closed form expression:

$$\begin{bmatrix} a_x \\ \mathbf{c_x} \end{bmatrix} = \left(\sum_{t=1}^{T} \begin{bmatrix} 1 & \mathbf{x_t}^\mathsf{T} \\ \mathbf{x_t} & \mathbf{x_t}\mathbf{x_t}^\mathsf{T} \end{bmatrix} \right)^{-1} \sum_{t=1}^{T} \begin{bmatrix} f(\mathbf{x_t}) \\ f(\mathbf{x_t})\mathbf{x_t} \end{bmatrix}. \tag{2.11}$$

Notice that Eq (2.10) describes the linear model that best approximates $f(x_t)$ under the empirical distribution given $\{\mathbf{x_t}\}_{t=1}^T$.

We assume that the parameters $(a_x, b, \mathbf{c_x})$ are bounded as follows: $|a_x| \leq \bar{a}$, $\underline{b} \leq |b| \leq \bar{b}$, $\|\mathbf{c_x}\|_1 \leq \bar{c}$. The seller is assumed to know the bounds on b, $\underline{b}$ and $\bar{b}$, but not the bounds on a_x and $\mathbf{c_x}$. The regret of any admissible pricing policy over a selling horizon of length T can now be defined as the difference in the expected revenue of a clairvoyant who uses a linear demand model with parameters $a_x, b, \mathbf{c_x}$, and the expected revenue achieved by that pricing policy.[2] We note here that although the

[2]Again, we could define regret relative to the "true clairvoyant," who knows the *true* demand model and sets price $\tilde{p}_t = -\frac{f(\mathbf{x_t})}{2b}$ at each time period. But this definition can re-

clairvoyant has full knowledge of the realization $\{\mathbf{x_t}\}_{t=1}^T$ at the start of the selling horizon, any admissible pricing policy does not know this realization, and can only observe the history $\mathcal{H}_{t-1} = \{p_1, \mathbf{x_1}, d_1, \ldots, p_{t-1}, \mathbf{x_{t-1}}, d_{t-1}\}$. The expected regret is given by

$$\text{Regret}(T) = \sum_{t=1}^{T} \mathrm{E}[p_t^* D(p_t^*)] - \sum_{t=1}^{T} \mathrm{E}[p_t D(p_t)], \tag{2.12}$$

where $p_t^* = -\frac{a_x + \mathbf{c_x}^\top \mathbf{x_t}}{2b}$ are the price chosen by the clairvoyant upon observing feature $\mathbf{x_t}$. The expectation in Eq (2.12) is taken over all random quantities, including the features $\mathbf{x_t}$, price ranges $[\underline{p}_t, \bar{p}_t]$, and demand noise ϵ_t.

To validate that the linear clairvoyant is indeed an upper bound of any pricing policy using linear demand models, let the prices chosen by our linear clairvoyant be $p_t^*(\mathbf{x}) = -\frac{a_x + \mathbf{c_x}^\top \mathbf{x}}{2b}$ for all t and for any features $\mathbf{x}$. Analogous to Proposition 1, Proposition 2 below shows that the linear demand function (2.10) gives the highest revenue among all linear demand functions, justifying our choice of regret benchmark.

Proposition 2. *Given a particular realization* $\{\mathbf{x_t}\}_{t=1}^T$ *of the features, consider price* $p_t' = -\frac{\alpha + \gamma^\top \mathbf{x_t}}{2\beta}$ *where* α, β, γ *are measurable with respect to history*

$$\mathcal{H}_{t-1} = \{p_1, d_1, \ldots, p_{t-1}, d_{t-1}\}.$$

Then, we have

$$\sum_{t=1}^{T} \mathrm{E}[p_t^* D_t(p_t^*) | \mathbf{x_1}, \ldots, \mathbf{x_T}] \geq \sum_{t=1}^{T} \mathrm{E}[p_t' D_t(p_t') | \mathbf{x_1}, \ldots, \mathbf{x_T}].$$

Algorithm. We adapt the RPS algorithm to the non-IID setting by introducing two main modifications: Firstly, we modify the second regression step in the two-stage regression performed by RPS. Instead of using b_{t+1} as an estimate for b and regressing $d_s - b_{t+1} p_s$ against previously observed feature vectors $\mathbf{x_s}$, we use b_s as an estimate

sult in a linear regret (see details in Appendix A.1). Namely if $\{\mathbf{x_t}\}$ happens to be IID, $\mathrm{E}\left[\sum_{t=1}^T \tilde{p}_t(\mathbf{x_t}) D(\tilde{p}_t(\mathbf{x_t})) - \sum_{t=1}^T p_t D(p_t)\right] = \Omega(T)$. Therefore, the optimal revenue of the true model is not a particularly informative benchmark, since no algorithm can achieve a sublinear regret rate with misspecified model.

Algorithm 3 Random Price Shock (RPS) algorithm for the non-IID setting.

input: parameter bound on b, $B = [-\bar{b}, -\underline{b}]$
initialize: choose $\hat{a}_1 = 0$, $\hat{b}_1 = -\bar{b}$, $\hat{c}_1 = 0$
for $t = 1, \ldots, T$ **do**
  set $\delta_t \leftarrow \frac{\delta}{2} t^{-\frac{1}{6}}$
  given $\mathbf{x_t}$, set unconstrained greedy price: $p^u_{g,t} \leftarrow -\frac{\hat{a}_t + \hat{\mathbf{c}}_\mathbf{t}^\mathsf{T} \mathbf{x_t}}{2\hat{b}_t}$
  project greedy price: $p_{g,t} \leftarrow \text{Proj}(p^u_{g,t}, [\underline{p}_t + \delta_t, \bar{p}_t - \delta_t])$
  generate an independent random variable $\Delta p_t \leftarrow \delta_t$ w.p. $\frac{1}{2}$ and $-\delta_t$ w.p. $\frac{1}{2}$
  set price $p_t \leftarrow p_{g,t} + \Delta p_t$
  choose an arbitrary price $p_t \in [\underline{p}_t, \bar{p}_t]$
  observe demand $d_t = D_t(p_t)$
  set $\hat{b}_{t+1} \leftarrow \text{Proj}(\frac{\sum_{s=1}^t \Delta p_s d_s}{\sum_{s=1}^t \Delta p_s^2}, B)$
  set $(\hat{a}_{t+1}, \hat{\mathbf{c}}_{\mathbf{t+1}}) \leftarrow \arg\min \sum_{s=1}^t (d_s - \hat{b}_s p_s - \alpha - \gamma^\mathsf{T} \mathbf{x_s})^2 + \| \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} \|^2 + (\alpha + \gamma^\mathsf{T} \mathbf{x_{t+1}})^2$
end for

for b at period s and then use Vovk-Azoury-Warmuth (VAW) estimator (Azoury and Warmuth, 2001) to regress $d_s - b_s p_s$ against past $\mathbf{x_s}$. This modification allows us to extend the analysis to an arbitrary sequence of features $\{\mathbf{x_t}\}_{t=1}^T$. Secondly, the magnitude of the price shock at each period t is increased from $t^{-\frac{1}{4}}$ to $t^{-\frac{1}{6}}$. This changes the balance of exploration and exploitation, putting more emphasis on exploration and allowing the modified RPS algorithm to learn the parameters more accurately regardless of the distribution of features. The full description of our modified algorithm is given in Algorithm 3.

We can prove the following upper bound on the regret of the RPS algorithm in the non-IID setting.

Theorem 4. *The regret obtained by the RPS algorithm for the non-IID setting is* $O(T^{2/3})$.

The proof of Theorem 4, given in Appendix A.3.6, relies on the properties of the VAW estimator, a variant of the ridge regression forecaster (Cesa-Bianchi and Lugosi, 2006, Ch 11.8). Our analysis follows the analysis in Cesa-Bianchi and Lugosi (2006), which studies the prediction of sequences in the presence of feature information. In their setup, a sequence $\{(\mathbf{y_1}, g(\mathbf{y_1})), (\mathbf{y_2}, g(\mathbf{y_2})), \ldots\}$ is observed, where the $\mathbf{y_n}$s are d-dimensional feature vectors and the function g determining the outcome variable

$g(\mathbf{y_n})$ is potentially nonlinear. The goal is to predict the outcomes $g(\mathbf{y_n})$ for each n based on the observations $\{(\mathbf{y_i}, g(\mathbf{y_i})), i = 1 \ldots n-1\}$. Cesa-Bianchi and Lugosi (2006) show that if the VAW estimator is used to predict outcomes, the regret for the square loss relative to the best offline estimator that can observe the entire sequence can be shown to be logarithmic in terms of n. They show that this bound is optimal in n. Since our linear clairvoyant functions as the best offline estimator, we can bound the regret of Algorithm 3 by expressing it in terms of the square loss regret in Cesa-Bianchi and Lugosi (2006).

Theorem 4 shows that even when the features $\{\mathbf{x_t}\}$ are generated from a non-IID distribution, it is possible to achieve a non trivial, sublinear regret in terms of the length of the selling horizon T as long as the features and the component $f(\mathbf{x})$ of demand are bounded. Nevertheless, it is not clear whether this upper bound on the regret is asymptotically optimal as we do not have a matching lower bound in the order of $\Omega(T^{2/3})$. Noting that Proposition 2 implies that in the special case that the features are IID, the expected revenue of the non-IID linear clairvoyant is *at least* as much as the expected revenue of the IID linear clairvoyant, we see that Theorem 2 also serves as a lower bound in this setting. Thus there is a mismatch between our lower bound $\Omega(\sqrt{T})$ from Theorem 2 and upper bound $O(T^{2/3})$ from Theorem 4. We leave the problem of determining the asymptotic optimality of Algorithm 3 to future work. Finally, the constants in our upper bound are given in our proof of the theorem in Appendix A.3.6.

2.4 Numerical Results

In this section, we add to the analysis in the previous section with numerical simulations that empirically gauge the performance of the RPS algorithm. These are conducted on synthetic data, and investigate the dependence of the regret of the RPS algorithm on the length of the selling horizon T for the IID, price ladder and non-IID settings. They show that the regret growth matches our theoretical guarantees from the previous section, thus validating our theoretical analysis. We also bench-

mark the RPS algorithm against competing algorithms that do not account for price endogeneity, and show that the RPS algorithm alone learns the correct parameters of the demand function over the selling horizon, and thus outperforming competing algorithms in terms of the revenue earned over the course of the selling horizon.

Each simulation is run over a selling horizon of length 5000 periods and repeated 200 times. The three competing algorithms that we benchmark the RPS algorithm against are as follows:

- *Greedy algorithm*: The greedy algorithm (Algorithm 4) operates by estimating the demand parameters at each time period using linear regression, then setting the price to the optimal price assuming that the estimated parameters are the true parameters. This algorithm has been shown to be asymptotically optimal by Qiang and Bayati (2016) in a linear demand model setting with features, and with the availability of an incumbent price, but in general is known to suffer from *incomplete learning*, i.e., insufficient exploration in price Keskin and Zeevi (2014).

- *One-stage regression*: This algorithm introduces randomized price shocks to force price exploration, but uses a one-stage regression instead of a two-step regression as in RPS to learn the parameters. A full description of the one-stage regression algorithm (Algorithm 5) is given below. The one-stage regression algorithm is analogous to the class of semi-myopic algorithms introduced by Keskin and Zeevi (2014), which use (deterministic) price perturbations to guarantee sufficient exploration. However, Algorithm 5 does not consider the price endogeneity effect caused by model misspecification in the estimation process.

- *No feature clairvoyant*: As a benchmark, the performance of RPS is compared with the performance of a no feature clairvoyant. This clairvoyant knows the values of the parameters a and b but considers the features x, which will be drawn from a zero-mean distribution, to be part of the demand noise. Hence this clairvoyant will set prices to be $-\frac{a}{2b}$ at each time period. Such a pricing policy would be optimal in the absence of features but would evidently incur

regret linear in T when $m > 0$. This highlights the importance of considering demand features in dynamic pricing.

Algorithm 4 Greedy algorithm.

input: parameter bounds $B = [-\bar{b}, -\underline{b}]$
initialize: choose $\hat{a}_1 = 0$, $\hat{b}_1 = -\bar{b}$, $\hat{c}_1 = 0$
for $t = 1, \ldots, T$ **do**
  given x_t, set unconstrained greedy price: $p^u_{g,t} \leftarrow -\frac{\hat{a}_t + \hat{c}_t^\intercal x_t}{2\hat{b}_t}$
  if admissible price set is a price ladder **then**
    project greedy price onto price ladder: $p_{g,t} \leftarrow \text{Proj}(p^u_{g,t}, [q_1, \ldots, q_N])$
  else
    project greedy price onto price interval: $p_{g,t} \leftarrow \text{Proj}(p^u_{g,t}, [\underline{p}_t, \bar{p}_t])$
  end if
  set price $p_t \leftarrow p_{g,t}$
  observe demand $d_t := D_t(p_t)$
  set $(\hat{a}_{t+1}, \hat{b}_{t+1}, \hat{c}_{t+1}) \leftarrow \arg\min_{\alpha,\beta \in B,\gamma} \sum_{s=1}^{\intercal} (d_s - \alpha - \beta p_s - \gamma^\intercal x_s)^2$
end for

IID Setting

The first simulation example considers the case where the features x_t are independently distributed, prices are chosen from continuous price intervals, and the source of endogeneity is a misspecified demand function. In this set up, demand is given by the quasi-linear function

$$D_t(p) = \frac{1}{2(x_t + 1.03)} + 1 - 0.9p + \epsilon_t,$$

where x_t is a one-dimensional random variable uniformly distributed between $[-1, 1]$ and the noise ϵ_t is normally distributed with mean 0 and standard deviation 0.1. Using the closed-form expression in Eq (2.7), it can be seen that the linear demand model approximated by least squares is given by

$$\hat{D}_t(p) \approx 2.05 - 0.90p - 1.76x_t,$$

where all coefficients are expressed to 2 decimal places. The price range at period t is lower bounded by $\underline{p}_t = \$0.69$ and upper bounded by $\bar{p}_t = \$9.81$. The retailer assumes

Algorithm 5 One step regression.

input: parameter bounds $B = [-\bar{b}, -\underline{b}]$
initialize: choose $\hat{a}_1 = 0$, $\hat{b}_1 = -\bar{b}$, $\hat{c}_1 = 0$
for $t = 1, \ldots, T$ **do**
  given x_t, set unconstrained greedy price: $p^u_{g,t} \leftarrow -\frac{\hat{a}_t + \hat{c}_t^\intercal x_t}{2\hat{b}_t}$
  if admissible price set is a price ladder **then**
    find $i = \arg\min_{j \in \{1,\ldots,N\}} |q_j - p^u_{g,t}|$ and set constrained greedy price: $p_{g,t} \leftarrow q_i$
    generate an independent random variable

$$\Delta p_t \leftarrow \begin{cases} q_i - q_{i-1} \text{ w.p. } \frac{q_{i+1}-q_i}{2(q_{i+1}-q_{i-1})t^{1/3}} \\ q_{i+1} - q_i \text{ w.p. } \frac{q_i - q_{i-1}}{2(q_{i+1}-q_{i-1})t^{1/3}} \\ 0 \text{ w.p. } 1 - t^{-1/3} \end{cases}$$

  else

$$\text{set } \delta_t \leftarrow \begin{cases} \frac{\delta}{2} t^{-\frac{1}{4}} \text{ if } \{x_t\} \text{ is IID} \\ \frac{\delta}{2} t^{-\frac{1}{6}} \text{ otherwise.} \end{cases}$$

    project greedy price: $p_{g,t} \leftarrow \text{Proj}(p^u_{g,t}, [\underline{p}_t + \delta_t, \bar{p}_t - \delta_t])$
    generate an independent random variable $\Delta p_t \leftarrow \delta_t$ w.p. $\frac{1}{2}$
    and $\Delta p_t \leftarrow -\delta_t$ w.p. $\frac{1}{2}$
    set price $p_t \leftarrow p_{g,t} + \Delta p_t$
    choose an arbitrary price $p_t \in [\underline{p}_t, \bar{p}_t]$
  end if
  observe demand $d_t := D_t(p_t)$
  set $(\hat{a}_{t+1}, \hat{b}_{t+1}, \hat{c}_{t+1}) \leftarrow \arg\min_{\alpha,\beta \in B, \gamma} \sum_{s=1}^{\intercal} (d_s - \alpha - \beta p_s - \gamma^\intercal x_s)^2$
end for

that a lies in the interval $[1.5,\ 2.5]$, b lies in the interval $[-1.2,\ -0.5]$ and c lies in the interval $[-2.2,\ -1.2]$.

Results. Fig. 2-2a shows that in this numerical example, the regret of the greedy algorithm, the one-stage regression algorithm, and the clairvoyant who ignores features, grow linearly with t, and in all cases the regrets are higher than that of RPS after around 1000 iterations. Fig. 2-2b confirms that the regret of the RPS algorithm is $O(\sqrt{T})$. Finally, Table 2.1, which provides summary statistics of the parameter estimates of all the pricing algorithms except the clairvoyant at the end of the selling horizon, shows that the RPS algorithm produces close estimates of all the parameters. However, for the greedy and one step regression algorithms, the parameter estimates are actually moving away from the least squares true value, and converge to a point on the boundary of the feature parameter set. This demonstrates that parameter estimates may be significantly biased when the endogeneity effect caused by model

misspecification is not handled properly.

In Appendix A.2.1, we include additional numerical experiments for sensitivity analysis. We consider a family of quasi-linear demand functions of the form

$$D_t(p) = \frac{1}{2(x_t + \gamma)} + 1 - 0.9p + \epsilon_t,$$

where γ ranges from 1.02 to 2. As γ decreases and approaches to 1, the function $f(x_t) = 1/2(x_t + \gamma)$ becomes more nonlinear for $x_t \in [-1, 1]$, and the fit of the closest linear approximation of demand function deteriorates. Since model misspecification worsens as γ approaches 1, we would expect that the endogeneity effect is more significant for demand models with smaller values of γ. The simulation results confirm that the regret gap between the RPS algorithm and the one-stage regression algorithm increases as γ decreases. Moreover, we find that the RPS algorithm produces unbiased parameter estimates for all γ, while the estimates from the one-stage regression algorithm are biased especially when γ is close to 1.

We also analyze how the regret of the RPS algorithm changes with the dimension of the feature vectors, m. The detailed simulation results are included in Appendix A.2.2 . We find that the regret of RPS tends to increase with m, and that the growth rate of regret appears to match Theorem 1's theoretical bound of $O((m+1)\sqrt{T})$ in terms of m.

Price ladder setting We now consider the same set up as in the IID setting, but replace the price range [\$0.69, \$9.81] with a price ladder [\$0.50, \$0.70, . . . , \$9.70, \$9.90]. where the features x_t are independently distributed, prices are chosen from continuous price intervals, and the source of endogeneity is a misspecified demand function.

Results. As in the previous subsection, the regret of the Greedy algorithm, the One Step Regression algorithm, and the clairvoyant who ignores features, grow linearly with T (Fig. 2-2c) while the regret of the RPS algorithm (Algorithm 2) is $O(T^{2/3})$ (Fig. 2-2d). The summary statistics of the parameter estimates of the competing algorithm (Table 2.2) again show that the RPS algorithm produces close estimates of all the parameters, while we once more observe that the greedy and

one-step regression produce biased estimates.

Non IID setting Finally, we consider the case where prices are chosen from continuous price intervals but the features x_t are not independently distributed. In this set up, the demand function is given by the quasilinear function

$$D_t(p) = -0.9p + f(x_t) + \epsilon_t,$$

with

$$f(x) = \frac{1}{2(x + 1.1)} + 1.5.$$

We assume that x_t is one dimensional (i.e. $m = 1$), $x_t = -1 + \frac{2}{\sqrt{t}}$ for $t = 1, \ldots, 5000$ (note that $x_t \in [-1,\ 1]\ \forall t$) and the noise ϵ_t is normally distributed with mean 0 and standard deviation 0.1.

Recall from the definition of the cumulative expected regret in Section 2.3.5 that in the non-IID setting, Regret(T) is expressed relative to a clairvoyant who bases pricing decisions on the realized sequence of feature vectors, $\{x_1, \ldots, x_T\}$. Thus, to estimate Regret(t) for $t = 1, \ldots, 5000$, we define a separate clairvoyant for each time period t; we calculate the regret by comparing the cumulative revenue of our pricing policies at time t with the cumulative revenue of a clairvoyant who bases pricing decisions on $\{x_1, \ldots, x_t\}$. Denote the demand model parameters assumed by the clairvoyant at time t as $(a(t), b, c(t))$.

The remaining parameter settings are as follows: At period t, the admissible price range is set to

$$[\underline{p}_t, \bar{p}_t] = [-\frac{f(1)}{2b}, -\frac{f(-1)}{2b}] = [\$0.97,\ \$3.61].$$

We assume that the retailer knows that a lies in the interval

$$[\min_t\{a(t)\} - 0.5, \max_t\{a(t)\}] = [1.9,\ 2.6],$$

that b lies in the interval $[-1.2,\ -0.1]$ and that c lies in the interval

$$[\min_t\{c(t)\} - 0.5, \max_t\{c(t)\}] = [-7.3,\ 0.3].$$

Results Fig. 2-2e plots the average regret of the RPS algorithm (Algorithm 3), as well as the competing Greedy and One-step regression algorithms. The regret incurred by the RPS algorithm is for $t > 1000$ lower than the regret of the other three algorithms, and its regret is $O(T^{2/3})$ as shown by Fig. 2-2f. Table 2.3 shows that the RPS algorithm accurately estimates the parameters $a(5000), b, c(5000)$ while the Greedy and One Step Regression algorithms do not.

Table 2.1: End of selling horizon parameter estimates in the IID setting

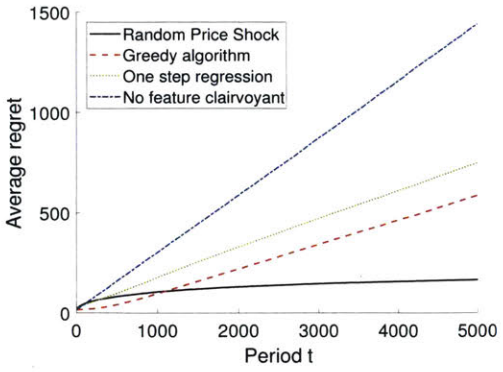
	True value	RPS algo.	Greedy algo.	One step reg.
Mean ($\hat{a}_T$)	2.05	2.04	1.50	1.50
Median ($\hat{a}_T$)	2.05	2.04	1.50	1.50
Mean ($\hat{b}_T$)	-0.90	-0.91	-0.50	-0.50
Median ($\hat{b}_T$)	-0.90	-0.89	-0.50	-0.50
Mean ($\hat{c}_T$)	-1.76	-1.74	-1.20	-1.20
Median ($\hat{c}_T$)	-1.76	-1.75	-1.20	-1.20

Table 2.2: End of selling horizon parameter estimates in the price ladder setting

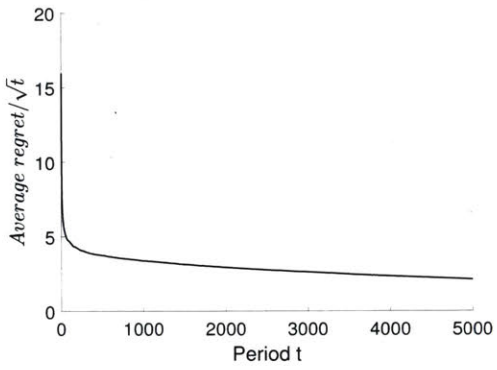
	True value	RPS algo.	Greedy algo.	One step reg.
Mean ($\hat{a}_T$)	2.05	2.16	1.50	1.50
Median ($\hat{a}_T$)	2.05	2.31	1.50	1.50
Mean ($\hat{b}_T$)	-0.90	-1.01	-0.50	-0.50
Median ($\hat{b}_T$)	-0.90	-1.11	-0.50	-0.50
Mean ($\hat{c}_T$)	-1.76	-1.81	-1.20	-1.20
Median ($\hat{c}_T$)	-1.76	-1.88	-1.20	-1.20

Table 2.3: End of selling horizon parameter estimates in the non IID setting

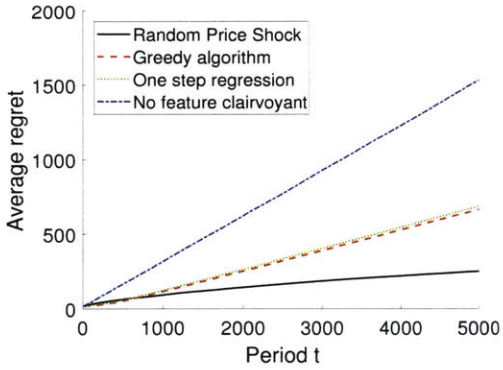
	True value	RPS algo.	Greedy algo.	One step reg.
Mean ($\hat{a}_T$)	-1.38	-1.35	-1.49	-0.87
Median ($\hat{a}_T$)	-1.38	-1.37	-1.50	-0.88
Mean ($\hat{b}_T$)	-0.90	-0.91	-0.16	-0.40
Median ($\hat{b}_T$)	-0.90	-0.91	-0.16	-0.40
Mean ($\hat{c}_T$)	-6.63	-6.60	-3.95	-4.40
Median ($\hat{c}_T$)	-6.63	-6.66	-3.97	-4.40



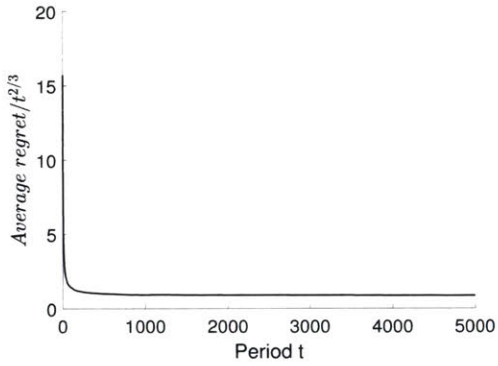
(a) IID setting – average regret



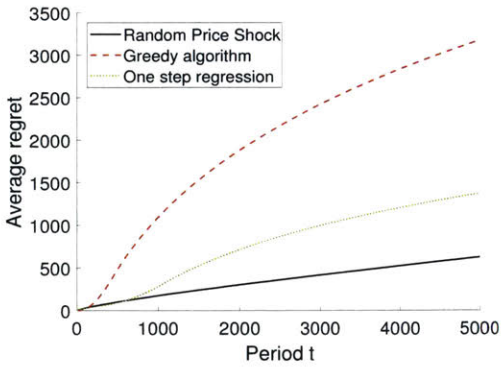
(b) IID setting – avg regret scaled by $1/\sqrt{t}$



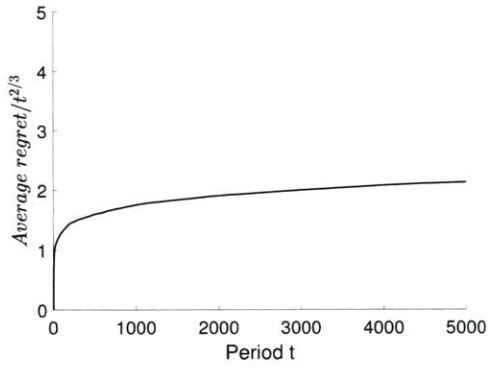
(c) Price ladder setting – average regret



(d) Price ladder – avg regret scaled by $t^{-2/3}$



(e) Non-IID setting – average regret



(f) Non-IID setting – avg regret scaled by $t^{-2/3}$

Figure 2-2: Average regret and scaled regret in IID, price ladder and non IID settings

2.5 Conclusion

We have shown that in dynamic pricing with contextual information, model misspecification can give rise to price endogeneity. We have proposed a "random price shock" (RPS) algorithm, which employs a combination of randomly generated price shocks and a two-stage regression procedure in order to produce unbiased estimates of price elasticity. This allows the RPS algorithm to maximize its revenue despite the presence of endogeneity. Our analysis shows that RPS does indeed exhibit strong numerical and theoretical performance; Our upper bound on the expected regret, $O((m+1)\sqrt{T})$, is optimal in T.

We have also shown that the RPS algorithm is versatile and can be adapted to a number of common business settings, where the feasible price set is a price ladder, and where the contextual information is not IID. We have introduced simple modifications to the RPS algorithm to adapt it to these settings and proved corresponding theoretical guarantees; the regret of the modified RPS algorithm is $O(\sqrt{(m+1)}T^{2/3})$ in the price ladder setting, and $O(T^{2/3})$ in the non IID setting.

We end by noting that in this paper, we are primarily interested in model misspecification, and have addressed the problem of price endogeneity in dynamic pricing specifically as caused by model misspecification. A natural question is whether our model and analysis can be generalized to include other sources of endogeneity potentially faced by a retailer, such as competition and strategic customers. These are beyond the scope of this paper, and we leave such extensions to future work.

Chapter 3

Feature-based Dynamic Pricing for Fashion Retail: A Case Study

In collaboration with Oracle Retail, we design and test a pricing heuristic, called the "random price shock" heuristic, that is broadly applicable to fashion retail settings. This heuristic is an online learning algorithm that does not assume any knowledge of the underlying demand distribution, but instead learns this distribution by dynamically generating randomized price shocks to accurately estimate price elasticity. In addition, the RPS heuristic incorporates business constraints faced by many fashion retailers, such as fixed inventory and markdown pricing constraints. It makes use of intuitive and computationally tractable approximations to optimize the retailer's total expected revenue subject to these constraints. To gauge the performance of the RPS heuristic, we have run a number of offline numerical experiments using retail data from one of Oracle Retail's clients. The heuristic exhibits revenue gains of around 2-7% over current practice, and seems robust to different retailer parameter settings such as the length of the markdown and no-touch periods.

3.1 Introduction

Pricing in fashion retail has historically been the province of retail managers - executives who, through a combination of expertise and experience, anticipate how demand

will be influenced by factors such as product characteristics, seasonality and geography, and set prices accordingly. With the successful application of data-driven and algorithmic approaches to price optimization in fashion retail (Ferreira et al., 2015; Caro and Gallien, 2012), a growing number of retailers have become interested in pairing their price managers' expertise with more rigorous analytics-based solutions. Companies such as Oracle Retail, SAP SE, and IBM Retail fill this niche by offering software and IT solutions that help retailers to understand and exploit their data without the need for in-house data scientists.

In this work, we collaborate with Oracle Retail, a business unit of Oracle and one of the leading providers of software and IT solutions to retailers, to design and test a feature-based dynamic pricing algorithm that is broadly applicable to a variety of fashion retail settings. Our aim is to eventually integrate our algorithm into Oracle Retail's suite of price optimization products. Currently, these products forecast demand by using static estimates of the demand-price relationship: Prior to the start of each selling season, the demand-price relationship is estimated using historical data, and is not re-estimated until the start of the next selling season. Instead, we design and test a dynamic learning and pricing algorithm. Every season, this algorithm takes a batch of new, never-before-seen line of products, and updates the prices of these products on a weekly basis without assuming any prior knowledge of the underlying demand distribution. This demand distribution is learned through a combination of price experimentation and exploitation, by using historical transaction and feature data to update demand model estimates and prices in an online fashion.

As we have discussed in Chapter 2, one of the pitfalls of designing a dynamic pricing policy when demand depends on feature information lies in correctly estimating the causal relationship between demand and price. For example, if the policy assumes a misspecified demand model, either because the decision maker is unsure of how demand is affected by the features, or of how to model such a dependence, demand noise can become correlated with the price (in other words, we will have price endogeneity). Another factor that can complicate demand estimation in this context, and which was only briefly discussed in Chapter 2, is that retailers seldom

give data driven pricing algorithms free rein in setting prices. Instead, prices must fall within some price interval or set recommended by pricing managers. These bounds could be determined by managers in response to signals observed by them but not by the algorithm, such as product quality, the cost of manufacturing the product, or competitors' prices, behavior, and market shares. Since these signals could be correlated with demand noise, the prices set by any policy could in turn be correlated with demand noise, thus inducing price endogeneity. Our first order of business is hence to propose a feature-based pricing algorithm that corrects for this endogeneity effect, and that sets prices based on a correctly estimated demand model.

Recall that in Chapter 2, we proposed a "Random Price Shock" (RPS) algorithm that corrects for price endogeneity by adding randomized price perturbations to the greedy prices proposed by some price optimization model. RPS tailors the variances of these price perturbations to balance the exploration-exploitation tradeoff inherent in this class of learning and optimization problems. It uses larger variances at the start of the season to explore and learn the demand function, and uses smaller variances towards the end of the selling horizon to exploit current parameter estimates and maximize revenue. While this idea of introducing randomized price shocks is applicable to our setting, none of the variants of RPS proposed in Chapter 2 (Algorithms 1, 2 and 3) can be directly applied to a fashion retail setting. Firstly, these algorithms assume that a single product, rather than a batch of products, is sold throughout the selling horizon. More importantly, they do not take into account business constraints commonly practiced by retailers. For example, fashion retailers universally practice markdown pricing towards the end of each sales cycle, where unsold items are repriced at dramatic discounts. Retailers also face fixed inventory constraints, where each store only has access to a limited amount of inventory of each item. In this work, we show how the RPS algorithm can be adapted to incorporate these constraints in a natural and computationally tractable way.

To gauge how the RPS heuristic might perform in a real world setting, Oracle Retail has shared with us sales data from an anonymous brick-and-mortar company. This data consists of a customer transaction dataset as well as an item feature dataset

describing transitions spanning from August 2012 to July 2015. Using these datasets, we designed experiments that estimate the revenue that *would have been* earned by the retailer if the prices of the items in the dataset had been chosen by the RPS algorithm. This was a two-stage process: First, we used predictive modeling to build a counterfactual model of weekly demand based on historical data. Then, using our predictive model as a "ground truth" model, or a stand-in for the true demand, we simulated the performance of the RPS heuristic over the selling horizon, allowing it to price the items in the historical dataset based on their feature information.

The rest of the chapter is organized as follows: The rest of this introductory section gives an overview of the relevant empirical literature on demand estimation and data-driven optimization in fashion retail. In Section 3.2, we introduce the pricing problem faced by Oracle Retail by describing in full detail the kinds of business constraints that are commonly faced by their clients in fashion retail. Section 3.3 then discusses the current solution approach used by Oracle Retail in their price optimization products for fashion retailers, followed by our solution approach, which adapts the RPS algorithm developed in Chapter 2 in order to satisfy business constraints on pricing and inventory clearance. Finally, Section 3.4 presents numerical experiments that use historical data from a single fashion retailer in order to validate our algorithms. These experiments allow us to compare RPS' performance with current practice.

3.1.1 Literature Review

This work is very closely related to the academic literature that applies price optimization to revenue management practice in retail. For detailed overviews of this literature, we refer the reader to Talluri and Van Ryzin (2005), Elmaghraby and Keskinocak (2003) and Bitran and Rene (2003), and highlight here several papers that relate specifically to fashion retail. These include Caro and Gallien (2012), Smith and Achabal (1998), and Mantrala and Rao (2001), who develop and field test markdown pricing decision support tools for retailers, and Ferreira et al. (2015), who develop and also field test a feature-based pricing tool for the flash sales retailer Rue La La. All of

these decision support tools operate in two main stages: First, a demand forecasting model is calibrated, using historical data as in Caro and Gallien (2012), Smith and Achabal (1998) and Ferreira et al. (2015), or the predictions of human experts as in Mantrala and Rao (2001). Then, a price optimization problem is solved based on this demand model. Although the pricing decision tool developed in this work also separates demand model estimation and markdown price optimization into two different stages, a key difference is that we treat the pre-markdown phase as an opportunity for price optimization. Prior to the markdown phase, learning and earning takes place on the fly.

Our work is thus also closely related to the literature on dynamic learning and pricing, particularly when inventory is limited. Notably, Besbes and Zeevi (2012) and Wang et al. (2011) look at the setting where demand is Poisson, with parameters that are unknown and have to be learned over time. Meanwhile, Badanidiyuru et al. (2014) and Ferreira et al. (2017) look at Multi-armed Bandit formulations of the pricing problem, where prices are selected from some discrete set rather than from continuous intervals. The algorithms proposed in these works have price experimentation structures that are designed to balance between the exploration-exploitation tradeoff inherent in this class of dynamic learning and pricing problems. Further, they tackle the computational complexity caused by the fixed inventory constraint by using certainty equivalent approximations of demand. Our pricing algorithm adopts both of these features, but unlike these papers, which gauge the performance of their algorithms through theoretical regret bounds, our work is practice-driven and validates our algorithm through numerical simulations on real world data.

The literature on structural demand estimation is also relevant to the methodology of our offline numerical experiments. From this literature, Berry et al. (1995) estimates a multinomial choice model using automobile sales data, Phillips et al. (2015) estimates a probit choice model using data from the autolending industry, Veeraraghavan and Vaidyanathan (2011) uses ordinal regression to estimate how the perceived seat values in stadiums and theaters depend on their locations within the venue, Vulcano et al. (2010) estimate a multinomial choice model using airline sales

data, and Fisher et al. (2018) use retail data to estimate a consumer choice model that captures substitution effects among competing products from different retailers. Vulcano et al. (2010) and Fisher et al. (2018) also use their calibrated demand models to perform price optimization based on counterfactual revenues predicted by these models. Of these works, Berry et al. (1995), Phillips et al. (2015) and Fisher et al. (2018) are especially relevant to our work as they use instrumental variables to correct for endogeneity in their datasets. Phillips et al. (2015) in particular also use a Hausman-type variable as an instrumental variable, where the endogenous price is averaged over groups of products to average out the effect of omitted demand variables. However, a main difference between these works and ours is that the ground truth demand estimation in our offline simulations combines econometrics methods with machine learning methods. This allows us to learn the causal relationship between price and demand while also obtaining a demand model with enhanced predictive power.

3.2 Objectives and Assumptions

The pricing problem faced by Oracle Retail's clients in fashion retail is as follows: Every week, an assortment of products with limited inventory is sold at each store. Each product has its own *life cycle*, which is defined as the period of time between the first week in which the product is sold, and the final week in which it can be sold (after which all remaining units of inventory are taken off the market). The length of the life cycle is known as the product lifetime, and the life cycles of different products can be staggered in the sense that product 2's life cycle could begin either before or after product 1's lifecycle.

For each product, its lifecycle begins with a no-touch period, during which price changes must be kept to a minimum. This is followed by an exploitation period, during which the algorithm is allowed to change prices for the purposes of learning the underlying demand model, and optimizing revenue. Finally, the lifecycle ends with a markdown phase, during which the product must be priced at a discount.

The lengths of the no-touch, exploitation and markdown phases are determined by retailers according to their own marketing strategies and cannot be optimized by Oracle Retail. Typically, the markdown phase is a substantial portion of the product lifecycle, and can last from between a third to a half of the full lifecycle.

Retailers also faces inventory constraints. The main constraint is that only a fixed amount of inventory of each product is available at each store, and that replenishment is not possible even if demands for the product turn out to be higher than anticipated. Below we give a full list of all the business constraints, including pricing and inventory constraints, that are faced by the retailer.

1. (Price ladder constraint) Price must at all time periods be selected from a price ladder, assumed for simplicity to have prices in \$1 increments, e.g. \$3.99, \$5.99, \$7.99.

2. (No-touch constraint) The selling horizon for each product begins with a no-touch period, during which price experimentation must be kept to a minimum. The retailer will provide the algorithm with some recommended price p_{rec}, and may stipulate that there can be no price experimentation at all, or allow price experimentation to within 10% (or the maximum of this value and \$1) of the recommended price.

3. (Price bound constraint) The no-touch period is followed by an exploitation period that starts at week T_e. This exploitation period is of length roughly one-third the selling horizon, during which prices must be kept within 20% (or the maximum of this value and \$1) of the recommended price.

4. (Inventory constraint 1) Only a fixed and finite amount of inventory is available for each product at the start of the selling horizon, with no replenishments.

5. (Inventory constraint 2) All available inventory must be *cleared* by the end of the selling horizon.

6. (Markdown constraint 1) The selling horizon for each product ends with a markdown period. The markdown period begins at a pre-specified week number T_m

and cannot be optimized by the algorithm.

7. (Markdown constraint 2) During the markdown period, only a single price can be charged. This price must always be less than the recommended price, and at least 30% of the recommended price.

Note that all of these constraints are always feasible except for the second inventory constraint. Since inventory is selected by the retailer (i.e. it is an input and not a decision variable in our algorithm), it is possible that this constraint may not always be feasible, e.g. if if the retailer orders excess inventory expecting demand to be larger than its realized values. We thus treat Inventory constraint 2 as a guideline rather than a strict constraint, and use the amount of inventory cleared as one of the metrics for the success of our proposed algorithm.

3.3 Solution Approach: The Random Price Shock Algorithm

We now present the algorithm that we have designed to tackle the pricing problem outlined in the previous section. First, we describe the methodology used by Oracle Retail's current price optimization products, and explain the advantages of our dynamic pricing approach. Then we describe our approach, which involves modifying the Random Price Shock algorithm from Chapter 2 to handle the various business constraints faced by fashion retailers.

3.3.1 Legacy Pricing Process

The current implementation of Oracle Retail's price optimization products for fashion retailers mainly focuses on the markdown pricing component of the pricing problem described in the previous section. To select the markdown prices of different items, a demand forecasting model is calibrated, and the markdown prices that maximize counterfactual revenues under this model are selected.

The current approach does not dynamically re-update the demand model as new sales data is observed during the selling season. Instead, it uses historical data prior to the selling season, along with a standard correction for endogeneity, to compute a static estimate of the demand-price relationship prior to the start of each selling season. During the selling season, markdown prices are selected as a function of this static estimate of the demand-price relationship, and as well as of the remaining inventory levels.

There are a number of reasons as to why a dynamic pricing approach is preferable to a static one. First, the static approach used by Oracle Retail ignores data from the current season. However, fashion retail is highly trend sensitive, and demand patterns can change significantly from season to season. Performing demand forecasting solely based on less relevant data from previous seasons is likely at the cost of accuracy. A second disadvantage to the static approach is that it consumes significant computational resources; Since the price optimization products offered by Oracle Retail are cloud-based, and large amounts of historical data have to be stored to perform demand forecasting using the current method, these products can be costly to retailers. A dynamic pricing approach, on the other hand, can be run using only data from the current season. The reduction in necessary storage or computational resources would lead to immediate savings for retailers, making the pricing product more marketable.

In the rest of this section, we adapt the dynamic learning and pricing algorithm developed in Chapter 2 to a fashion retail context. Unlike Oracle Retail's current static approach, RPS re-estimates the demand-price relationship on a weekly basis as more demand data is observed. Although we do not compare RPS with Oracle Retail's current price optimization strategy, we perform a case study using a single retailer's data to design and run offline numerical simulations that compare RPS with the retailer's current practice. We leave comparisons between the RPS algorithm and Oracle Retail's current price optimization strategy to future work.

3.3.2 The Random Price Shock Algorithm

In the RPS algorithm and its variants presented in Chapter 2 (Algorithms 1, 2 and 3), price optimization was straightforward; we simply solved the unconstrained convex optimization problem $\max p_t \mathrm{E}[D_t(p_t)]$ and projected the solution to some feasible interval (in the IID and non IID settings) or feasible set (in the price ladder setting). In our fashion retail setting, however, the fixed inventory and inventory clearance constraints (inventory constraints 1 and 2 in Section 3.2) couple the pricing decisions across time. Even when the demand distribution is known, finding the optimal prices involves solving a dynamic program that is computationally complex and does not have an explicit solution.

Instead of solving this DP, we have proposed a heuristic that looks at a certainty equivalent approximation of the pricing problem. The idea behind this approximation is to replace future demand realizations with the expected demands, thus relaxing the requirement that the inventory constraints are satisfied almost surely to the case where they are only satisfied in expectation. This relaxation technique was proposed by Gallego and van Ryzin (1994), who showed in their setting that the value of the deterministic program is an upper bound on the original DP, and further, that the solution to this deterministic program is asymptotically optimal in the sense that the revenue loss as a fraction of the optimal revenue goes to zero as the number of time periods goes to infinity.

In our setting, we apply the certainty equivalent approximation technique as follows: During the exploitation phase, we focus on not exceeding the available inventory, and choose prices by maximizing revenue subject to the constraint that the sum of deterministic demands until the end of the selling horizon is *at most* the available inventory. During the markdown phase, we focus on clearing the remaining inventory, and choose prices by maximizing revenue subject to the constraint that the sum of deterministic demands until the end of the selling horizon is *at least* the remaining inventory. Denoting the price at period t by p_t, the features at period t by $\mathbf{x_t}$, the inventory at period t by $\mathrm{Inv}(t)$ and the markdown period length by M, the resulting

optimization problems are given by (3.1) and (3.2).

Exploitation Phase Optimization Problem:

$$\begin{aligned} \max_{\{p_s\}} \quad & \sum_{s=t}^{T} p_s(bp_s + \mathbf{c}^\mathsf{T}\mathbf{x_s}) \\ \text{subject to} \quad & \sum_{s=t}^{T} bp_s + \mathbf{c}^\mathsf{T}\mathbf{x_s} \leq \text{Inv}(t) \end{aligned} \tag{3.1}$$

Markdown Phase Optimization Problem:

$$\begin{aligned} \max_{p} \quad & \sum_{s=t}^{t+M-1} p(bp_s + \mathbf{c}^\mathsf{T}\mathbf{x_s}) \\ \text{subject to} \quad & \sum_{s=t}^{t+M-1} bp_s + \mathbf{c}^\mathsf{T}\mathbf{x_s} \geq \text{Inv}(t) \end{aligned} \tag{3.2}$$

The solution to (3.2) is given by $p^* = \min\{-\frac{\mathbf{c}^\mathsf{T}\mathbf{x_t}}{2b}, \frac{I_m - \sum_{t=T_m}^{T} \mathbf{c}^\mathsf{T}\mathbf{x_t}}{b(T-T_m+1)}\}$. This p^* can then be projected to the interval $[0\ p_{\text{rec}}]$ and projected again to the price ladder to obtain a feasible markdown price. It is easy to see that an explicit solution to (3.1) can be found as well by rewriting this optimization problem in terms of new variables. Let $\delta_t = p_t - p^*_{u,t}$, where $p^*_{u,t} = -\frac{\mathbf{c}^\mathsf{T}\mathbf{x_t}}{2b}$ denotes the optimal price in the absence of inventory constraints. For any p_t, we can check that $p_t(-bp_t + \mathbf{c}^\mathsf{T}\mathbf{x_t}) - p^*_{u,t}(-bp^*_{u,t} + \mathbf{c}^\mathsf{T}\mathbf{x_t}) = -b\delta_t^2$. Then, rewriting the objective function of (3.1), this gives (3.3). By the concavity and the symmetry of the objective function of (3.3), we see that the optimal δ_ts must be the same for all t. Then $\delta_t^* = \max\{0, \frac{I_{T_e} - \sum_{t=T_e}^{T} \mathbf{c}^\mathsf{T}\mathbf{x_t}}{b(T-T_e+1)}\}$, which gives $p_t^* = p^*_{u,t} + \max\{0, \frac{I_{T_e} - \sum_{t=T_e}^{T} \mathbf{c}^\mathsf{T}\mathbf{x_t}}{b(T-T_e+1)}\}$. This p_t^* can then be projected to the interval $[0.8p_{\text{rec}}\ \ 1.2p_{\text{rec}}]$ and projected again to the price ladder to obtain a feasible price.

Markdown Phase Reformulated Optimization Problem:

$$\begin{aligned} \max_{\{\delta_s\}} \quad & -b\sum_{s=t}^{T} \delta_s^2 \\ \text{subject to} \quad & b\sum_{s=t}^{T} \delta_t \leq \text{Inv}(t) - \sum_{s=t}^{T} \mathbf{c}^\mathsf{T}\mathbf{x_s} \end{aligned} \tag{3.3}$$

The discussion here is summarized in the statement of Algorithm 6. At each time

period t, B_t items are sold. Since the lifecycles of different items may start at different points during the season, their exploration and markdown periods may also begin at different times. Period(j,t) denotes the current period in the sales cycle of item j, with Period$(j,t) = 1$ corresponding to the no-touch period, Period$(j,t) = 2$ corresponding to the exploration period, and Period$(j,t) = 3$ corresponding to the markdown period. During the no-touch period, price is simply set to the recommended price. During the exploration period, the heuristic computes the greedy price by solving (3.3) and projecting the price to the intersection of the admissible price range and the price ladder PL. Since the price ladder is spaced at \$1 intervals, a random price shock is selected from the set $\{-\$1, \$0, \$1\}$ and added to the greedy price for the purposes of price experimentation. As with the algorithms proposed in Chapter 2, this price shock has mean 0 and variance decreasing with time, allowing the price to grow closer to the greedy price - and hence the optimal price - as the estimates of the demand coefficients improve with time. Finally, at the start of the markdown period, the markdown price is computed by solving (3.2), and this price is charged until the end of Product j's life cycle.

3.4 Experimental Design

To validate Algorithm 6, we conducted numerical experiments based on data from a single retailer, which was shared with us by Oracle Retail as an example of the kinds of sales data that are generated by its clients in fashion retail. This retailer has a chain of brick-and-mortar stores across the US, and the data shared with us comes from transactions at different stores from August 2012 to July 2015. Purchased items are in the categories of fashion, furniture and housewares, but for the purposes of this study, we restrict our attention to fashion items.

The data consists of two main types of datasets:

1. Customer transaction data – This dataset consists of customer transactions Information on the time of each transaction, the location (i.e. the store, district and region) and the prices and IDs of the items purchased is included.

Algorithm 6 Random Price Shock (RPS) heuristic.

input: parameter bounds $B = [-\bar{b}, -\underline{b}]$
initialize: choose $\hat{a}_1 = 0$, $\hat{b}_1 = -\bar{b}$, $\hat{c}_1 = 0$
for $t = 1, \ldots, T$ **do**
  for items $j = 1, \ldots, B_t$ **do**
    if Period$(j,t) = 1$ **then**
      set $p_{j,t} \leftarrow p_{j,\text{rec}}$
    else
      if Period$(j,t) = 2$ **then**
        set $t_2 \leftarrow t_2 + 1$
        set price $p_{g,t} \leftarrow \frac{-\mathbf{c_t}^\top \mathbf{x_{j,t}}}{2b_t} + \max\{0, \frac{\text{Inv}(j,t) - \sum_{s=t}^{T_j} \mathbf{c_t}^\top \mathbf{x_{j,s}}}{b(T_j - t + 1)}\}$
        set price $p_{g,t} \leftarrow \text{Proj}(p_{g,t}, \text{PL} \cap [0.8p_{j,\text{rec}}\ 1.2p_{j,\text{rec}}])$
        generate an independent random variable
        $$\Delta p_t \leftarrow \begin{cases} 1 \text{ w.p. } \frac{1}{2t_2^{1/3}} \\ -1 \text{ w.p. } \frac{1}{2t_2^{1/3}} \\ 0 \text{ w.p. } 1 - t_2^{-1/3} \end{cases}$$
        set price $p_t \leftarrow p_{g,t} + \Delta p_t$
      else
        if Markdown week $= 1$ **then**
          set $p^m_{j,t} \leftarrow \min\{-\frac{\mathbf{c_t}^\top \mathbf{x_{j,t}}}{2b}, \frac{\text{Inv}(j,t) - \sum_{s=t}^{t+M-1} \mathbf{c_t}^\top \mathbf{x_{j,t}}}{b_t(M)}\}$
          set $p^m_{j,t} \leftarrow \text{Proj}(p^m_{j,t}, \text{PL} \cap [0\ p_{j,\text{rec}} - 1])$
        end if
        set $p_{j,t} \leftarrow p^m_{j,t}$
      end if
    end if
    observe demand $d_t = D_t(p_t)$
  end for
  set $\hat{b}_{t+1} \leftarrow \text{Proj}(\frac{\sum_{s=1}^{t} \sum_{j=1}^{B_t} I\{\text{Period}(j,t)==2\} \Delta p_{j,t} d_{j,t}}{\sum_{s=1}^{t} I\{\text{Period}(j,t)==2\} \Delta p_{j,t}^2}, B)$
  set $(\hat{a}_{t+1}, \hat{\mathbf{c}}_{\mathbf{t+1}}) \leftarrow \arg\min_{a', \mathbf{c}'} \sum_{s=1}^{t} \sum_{j=1}^{B_t} (a' + \mathbf{c}'^\top \mathbf{x_{j,s}} - (d_s - \hat{b}_s p_{j,s}))^2 + \| \begin{bmatrix} a' \\ \mathbf{c}' \end{bmatrix} \|^2 + (a' + \mathbf{c}'^\top \mathbf{x_{j,t+1}})^2$
end for

2. Item feature data – To supplement the transaction data, we had datasets providing information on each item, such as its class, subclass, and feature information. For fashion items, classes include categories of products such as shorts, t-shirts and dresses. Examples of product features include brand, color, pattern, neckline and sleeve length. A total of 51 features were included in the dataset, though not all features had been filled in - either because they were irrelevant to the class of items, or because of inconsistencies in data entry by the retailer.

In our numerical experiments, we used these historical datasets to first estimate a ground truth demand model. Then we used our ground truth model to run offline simulations that compare RPS' performance with current practice. The design of our offline simulations is described below.

3.4.1 Data Processing

We processed the raw data by first merging the customer transaction data and the item feature data. Next, we aggregated the sales at the week-district-item grandparent level, where an item *grandparent* combines store keeping units (SKUs) of the same design, regardless of color or sizing. This method of aggregation is valid as for the vast majority of the week-district-item grandparent groupings, only one price is offered for all SKUs, at all stores and on all days within the group. Week-district-item grandparent groupings for which more than one price was offered were removed from the dataset.

We then employed several cleaning steps suggested by our collaborators at Oracle Retail, including removing the first 5% and last 5% of sales for each item grandparent-region pair to avoid long tail ends in sales. We expanded the feature vector with additional information, mainly relating to seasonality. In our dataset, the level of sales seasonality is very significant. Fig. 3-1 shows the aggregate sales for a selected class of products, normalized from 0 to 1 within each year. Thus we added to the feature vector a variable recording the month, and indicator variables for holidays such as Christmas and Black Friday. We also added a variable indicating the number

of weeks that had elapsed since the first sale of the item grandparent within the district. Finally, we converted our categorical features into binary features using the standard method of one-hot encoding.

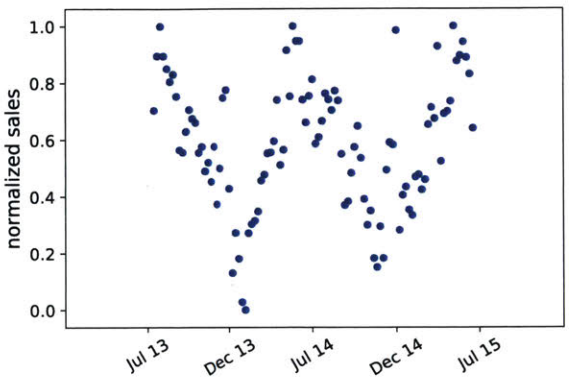


Figure 3-1: Seasonality of demand

3.4.2 Demand Model

For any subclass S of products, because products in the same subclass are similar to each other, we made a simplifying assumption that they have the same price sensitivity parameter b_S. The heterogeneity of product items is modeled using item-specific feature. A single demand function was thus used to describe the demand for all item grandparent i in subclass S, at district d and week w:

$$D^S_{i,d,w} = b_S p_{i,d,w} + f_S(\mathbf{x_{i,d,w}}) + \epsilon^S_{i,d,w}. \tag{3.4}$$

Here, $p_{i,d,w}$ represents the price of item i offered in district d and week w, and feature vector $\mathbf{x_{i,d,w}}$ represents the item-specific features and seasonal information. This function is linear in price and possibly nonlinear in the features $\mathbf{x_{i,d,w}}$, and is analogous to the single product demand function we defined in Eq (2.1).

3.4.3 Estimation and Endogeneity.

The dataset contains items belonging to 57 classes and 122 subclasses. Throughout the rest of this section, we focus on four subclasses. Before we discuss the estimation procedure for the demand model, we introduce the following standard metrics that we used to measure the accuracy of the estimated model:

1. Mean Absolute Percentage Error (MAPE), given by $(1/n) \cdot \sum_{i=1}^{n} |\hat{d}_i - d_i|/|d_i|$, where $d_1, \ldots, d_n$ are the true values and $\hat{d}_1, \ldots, \hat{d}_n$ are the predicted values.

2. Median Absolute Percentage Error (MDAPE), which is the median of the set $\{|\hat{d}_i - d_i|/|d_i|, i = 1, \ldots, n\}$.

We then used the following framework to train and test our demand model: We randomly split the dataset into a training set and a testing set in the ratio 70:30. This split was *not* chronological, because our objective was to build a ground-truth demand model that reflects the true demand process as accurately as possible. Note that the algorithm that we will design in the next chapter is an online algorithm that has no knowledge of the parameters of the ground-truth model. The ground-truth model, which is fit to the data in the training set, and then validated on the testing set data, is simply used to predict the counterfactual demand corresponding to the prices selected by the algorithm.

We began the demand estimation process by estimating the parameter b_S in (3.4) for each subclass S. Our initial approach was to simply apply ordinary linear regression (OLS) on the training data. We used standard variable selection techniques and measured the accuracy of the estimated model on the testing set. However, the first column of Table 3.1 shows that the coefficients of price in the baseline model were estimated to be either very close to 0, or positive in the case of Subclass 4. These results are unrealistic as they imply that demand barely depends on price, or increases with price. We note that there are certain luxury goods (known as Veblen goods) for which demand is usually observed to increase with price. These luxury goods include jewelry and designer fashion items. However, since the seller in the dataset is an off-price retailer, it seems that a more likely explanation of the baseline model price

coefficient estimates is price endogeneity caused by unobserved attributes. Namely, prices were set manually by the retailer based on items attributes such as costs of production to which we did not have access. Demand could also depend on these unobserved attributes (for example, demand could depend on quality, which is correlated with the cost of production), causing our baseline OLS model to obtain biased estimates of price coefficient.

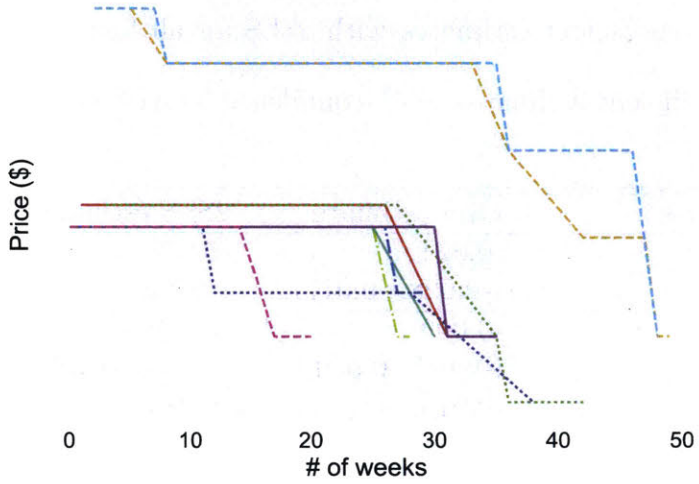


Figure 3-2: Markdown pricing: each trajectory represents the price of one item from the category.

We thus attempted to correct for endogeneity by using the two-stage least squares (2SLS) method. Typically, cost-side variables are used as instruments to control for endogeneity - or, in absence of such information, Hausman-type variables, where the average of price is taken over other regions, or lagged prices. Since we suspect that the omitted variables inducing price endogeneity relate to quality and other unobserved product characteristics, Hausman-type variables rather than lagged prices are the most appropriate instrument. For each item grandparent-district-week tuple, we computed the average price of all item grandparents sold in other districts during the same week, and set this as our instrumental variable. This method of averaging over prices was used in conjunction with a control-function approach (Phillips et al., 2015; Petrin and Train, 2010) to correct respectively for endogeneity in data from the

auto lending industry, and on households' choices of television reception options. By averaging over prices, we expect to also average out unobserved characteristics, causing the instrumental variable to be uncorrelated with the demand noise. The average price is also correlated with the price of each item sold in that week (thus meeting the second criteria of an instrumental variable), since we have observed that markdown pricing causes the prices of many items sold within the same season to decrease in sync with time (see Fig. 3-2). The covariance between our instrumental variable and the price were $0.23, 0.14, 0.23, 0.17$ for Subclasses 1–4, confirming this assumption. The corrected price coefficient estimates with 2SLS for all four subclasses are given

Table 3.1: Price coefficient estimates (95% confidence interval estimates in parentheses)

Subclass	OLS estimate	2SLS estimate
1	-0.022	-0.278
	(-0.028, -0.017)	(-0.308, -0.248)
2	-0.009	-0.280
	(-0.017, 0.000)	(-0.365, -0.195)
3	-0.018	-0.383
	(-0.028, -0.008)	(-0.599, -0.166)
4	0.028	-0.383
	(0.020, 0.037)	(-0.634, -0.132)

in the second column of Table 3.1 along with 95% confidence intervals. Running the Wu-Hausman test gave a p-value of less than 0.05 for all four subclasses, thus rejecting the hypothesis that there is no correlation between the price and demand noise, and supporting our claim that price endogeneity was present in the data for all four subclasses.

Next, we estimated the function $f_S(\cdot)$ in Eq (3.4). Substituting our 2SLS estimates of the demand elasticity b_S from Table 3.1 into Eq (3.4), we trained a function f_S to predict the remaining component of demand. We tested several ways to estimate $f_S(\cdot)$, including modeling it as a linear function, a regression tree, and a random forest. Table 3.2 compares the demand prediction errors (MAPE and MDAPE) when f_S is modeled as a linear function and as a random forest. We found that using random forest to predict demand with features gave the best prediction errors.

Table 3.2: Demand prediction errors using different demand models

	(MAPE)		(MDAPE)	
Subclass	Linear	Random Forest	Linear	Random Forest
1	61.5%	57.2%	49.3%	41.9%
2	55.0%	46.2%	45.8%	33.8%
3	56.4%	46.3%	47.5%	31.7%
4	68.2%	52.1%	55.5%	38.7%

3.4.4 Alternative Demand Models.

Recall that we made two key assumptions on our demand model: firstly, within any given subclass, demand for all products share the same price coefficient; secondly, demand for each item is independent of the prices of other items. To evaluate the robustness of these assumptions, we also considered the following candidates for demand models:

(M1) Demand for item grandparent i has its own price sensitivity parameter and its own demand function $D_i = a_i + b_i p_i + \epsilon_i$. This model relaxes the assumption that items in the same subclass share the same price coefficient, but ignore item-specific features.

(M2) The same demand function describes all item grandparent-district-week tuples within the same subclass, but each tuple has its own price elasticity: $D^S_{i,d,w} = a_S + (b_S \mathbf{x_{i,d,w}}) p_{i,d,w} + c_S^\intercal \mathbf{x_{i,d,w}} + \epsilon_{i,d,w}$. This is demand model is analogous to the one studied in Ban and Keskin (2017).

(M3) The demand for each item grandparent-district-week tuple depends on the prices of other products sold within that week: $D^S_{i,d,w} = a^S_{i,d,w} + b^S_1 p_{i,d,w} + b^S_2 \bar{p}_w + c_S^\intercal \mathbf{x_{i,d,w}} + \epsilon_{i,d,w}$, where $\bar{p}_w$ is the average price of all item grandparent-district tuples sold within the week. This model relaxes the assumption that demand between different items are independent, but ignores nonlinear effect of features.

These alternative demand models were evaluated and compared with the baseline model defined in Eq (3.4) where the function f_S is our random forest estimator. In Table 3.3, we show the prediction errors of the baseline model and the three

alternatives M1–M3 for products in Subclass 1. The results indicate that using these alternative models does not significantly reduce prediction errors. Therefore, we use the baseline model (3.4) as the counterfactual demand model in our simulations.

Table 3.3: Test set errors of alternate models on Subclass 1

	Baseline	M1	M2	M3
MAPE	57.2%	56.2%	55.2%	55.4%
MDAPE	41.9%	50.3%	48.4%	48.6%

3.4.5 Selecting Markdown Period Lengths and other Parameters

Before we could run our numerical experiments using our algorithm and ground truth demand model, we had to select appropriate values for parameters such as the starting inventory and markdown period lengths. Replicating the retail environment was essential to being able to draw a fair comparison between the RPS heuristic and current practice. However, a major difficulty we faced in doing so was that the retailer has only shared transaction and item feature data with Oracle Retail, and not timing information on the starts of markdown periods, or even timing information on the start of the lifecycles of the different products. Information on stock out timings and the total amount of inventory available for each product are also missing from these datasets, meaning that we do not know whether the existing pricing policy was always successful in clearing inventory.

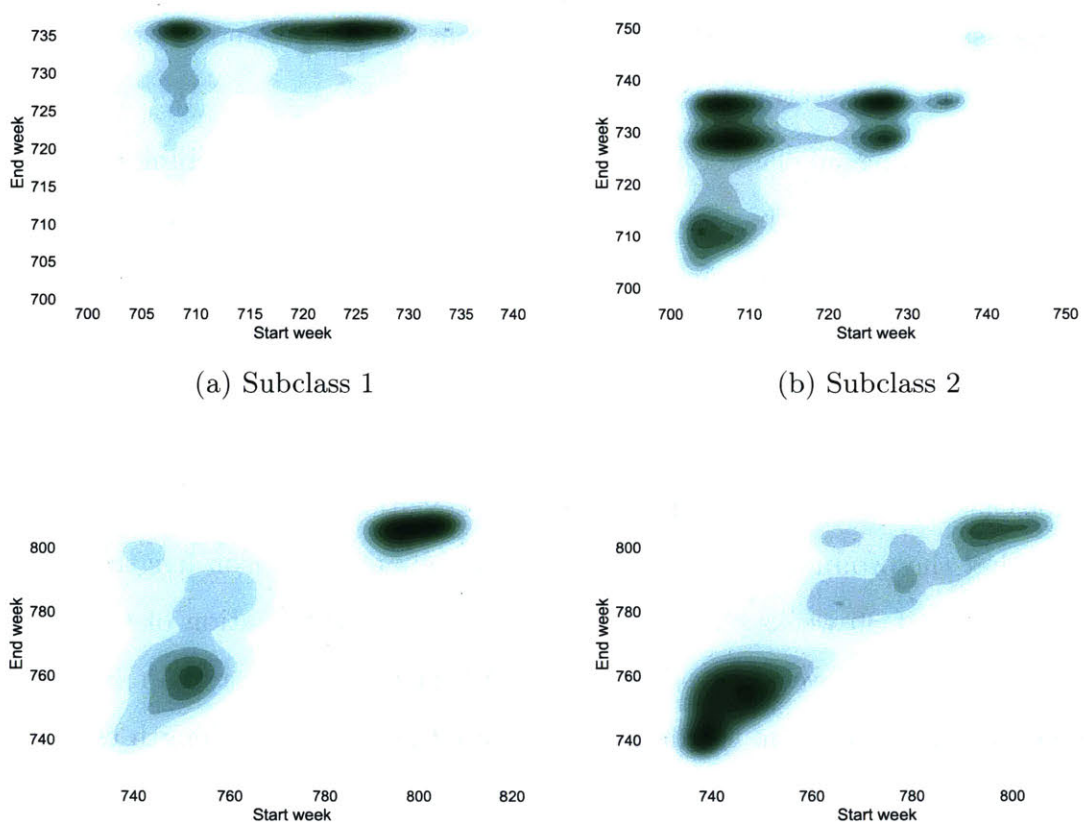
To address these challenges, we have worked closely with Oracle Retail to make a number of simplifying assumptions in our simulations. We have assumed that for all products, inventory was cleared by the end of the selling horizon, meaning that the total amount of inventory available for each product is assumed to be the sum of sales quantities in the historical dataset. Note that this assumption causes us to underestimate the revenues of the RPS heuristic, as its revenue is increasing in the amount of available inventory.

Estimating the actual starts and ends of the different products' life cycles was

trickier; Figures 3-3a-3-3d, which plot the distributions of the first and last weeks of the observed sales of different products, show that even though products tend to fall within distinct clusters, there is high variance in the start week within each cluster, and in the lifetime (defined as the number of weeks between the first and last weeks of observed sales) from cluster to cluster. In Subclass 1, the two main product clusters have start weeks that differ by about 20 weeks but similar end weeks. This lack of correlation between start and end week is unexpected, and suggests a problem with the dataset. We have thus discarded data from Subclass 1 and restricted our experiments in the next section to Subclasses 2-4.

For these remaining subclasses, we have demarcated seasons by batching products and assuming that products in each batch have the same life cycles, and share identical markdown periods and season end weeks. Products are batched by the month of the first observed sale, which is consistent with Figures 3-3a-3-3d, since these show that the difference in start weeks from cluster to cluster is typically at least 4 weeks. We then estimated the life cycles by looking at product lifetimes. The mean product lifetime for Class 303 is 21 weeks, with a standard deviation of 6 weeks. Since stockouts imply that the product lifetime is an underestimate of the season length, we set the season length at 30 weeks - roughly within one standard deviation of the mean product lifetime. Then, we removed products with lifetimes shorter than 10 weeks and longer than 30 weeks from the dataset to ensure that our season length estimate is reasonable for products included in the simulations.

Finally, as the exact markdown rules and other parameter settings used by the retailer are unknown to us and can only be estimated, we have run our simulations with a variety of inputs: (1) With a no-touch period of length 10 and price experimentation of at most 10% of the recommended price during this period, (2) with a no-touch period of length 3 and no price experimentation during this period, i.e. price is set to the recommended price, (3) with a markdown length of 10, and (4) with a markdown length of 15. Since it is common practice in fashion retail to set a markdown period length of around a third to a half of the total season length and a shorter no-touch period, it is likely that the true inputs lie somewhere within these



(a) Subclass 1

(b) Subclass 2

(c) Subclass 3

(d) Subclass 4

Figure 3-3: Kernel density estimation plots showing the distributions of the sales start and end weeks for different products across all four subclasses

ranges. Repeating our experiments for a variety of inputs within this range thus guarantees some measure of robustness.

3.5 Simulation Results

With the set up described above and for all possible combinations of the no-touch period and markdown settings, we ran 10 iterations of the RPS heuristic and compared the cumulative revenue earned by the heuristic over the time horizon with the cumulative revenue earned by the retailer' existing pricing policy under the ground truth demand model. Table 3.4 gives the average revenue gains of the RPS heuristic over the 10 iterations, with 95% confidence interval estimates in parentheses. Across the subclasses and parameter settings, revenue gains are on average between 2-7% – considerably lower than in the unlimited inventory setting studied in Phase 1, as is to be expected, but still significant. Also as expected, the revenue gains decrease as the markdown length is increased, and also as the amount of price experimentation allowed during the no-touch period decreases. However, with the exception of Subclass 4 with markdown length 15 and no-touch length 3, this decrease in revenue gains tends to be slight, suggesting that the RPS heuristic is robust to different retailer settings.

Next, we looked at the inventory clearance of the RPS heuristic across the different retailer settings. Table 3.5 reports the mean percentage of inventory left at the end of the selling horizon, where the average is taken across all products over 10 iterations, with the 95th percentile and max values of inventory left in parentheses. For all three subclasses in our simulations, we see that the RPS heuristic successfully clears all the available inventory for more than 95% of products by the end of the selling horizon. However, there are outliers for which a significant proportion of the inventory remains unsold at the end of the selling horizon. We inspected the simulated and actual sales of these outliers to find out why this was the case, and found that all of these products are also outliers in the sense that the ground truth demand model significantly underestimates the actual demand observed in the historical dataset.

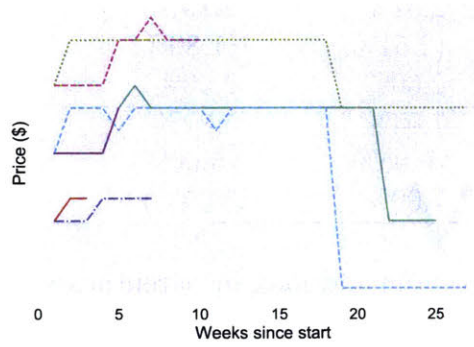
Then, since inventory levels are based on historical demand and simulated demands are based on our ground truth model, the inventory clearance constraint is infeasible for these products. As this is a feature of our demand model rather than of our algorithm, we believe that the 95 percentile values in Table 3.5 are a better indication of how effective the RPS heuristic is at clearing inventory.

We also looked at the price trajectories selected by the RPS heuristic across different retailer settings. Fig 3-4 gives these price trajectories for a sample of products from Subclass 2 for all four parameter settings. These plots confirm that the heuristic adheres to the pricing constraint and markdown constraints described in Section 3.3. Interestingly, we see that the prices selected by the RPS heuristic tend to increase rather than decrease during the exploitation period. This suggests that the prices currently set by the retailer are below the optimal prices.

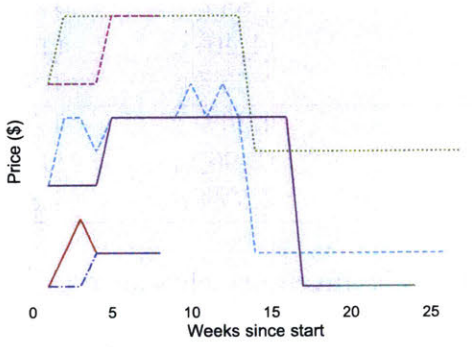
Finally, recall that in our numerical experiments with synthetic data in Chapter 2 (Section 2.4), we benchmarked the RPS algorithm against other dynamic pricing algorithms proposed in the literature that do not control for price endogeneity. We found that the RPS algorithm significantly performs the other algorithms in terms of its estimated cumulative revenues, as well as in terms of its parameter estimates of the underlying demand model. A natural question is whether RPS also exhibits superior performance on our fashion retail dataset, when the demand model is as described in Section 3.4. We address this question in Appendix B, and show that RPS does indeed outperform competing dynamic pricing algorithms that do not account for endogeneity on our fashion retail dataset.

3.6 Conclusion

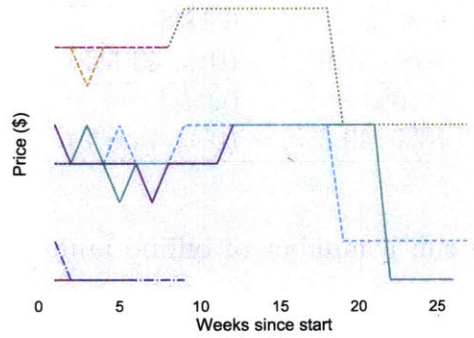
We have proposed a pricing heuristic that does not assume any knowledge of the underlying demand distribution and learns this distribution through a combination of price experimentation and exploitation. Our heuristic is intuitive, computationally inexpensive to implement, and is applicable to many fashion retail settings as it incorporates common business constraints such as inventory and markdown pricing



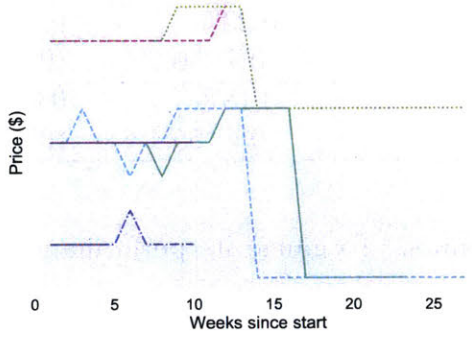
(a) No touch period length = 3
markdown period length = 10



(b) No touch period length = 3
markdown period length = 15



(c) No touch period length = 10
markdown period length = 10



(d) No touch period length = 10
markdown period length = 15

Figure 3-4: Price trajectories of the RPS heuristic for a sample of 10 products from Subclass 2

Table 3.4: Revenue gains relative to current practice (Mean, 95% confidence interval estimates in parentheses)

	No-touch length = 3		No-touch length = 10	
Subclass	Markdown length = 10	Markdown length = 15	Markdown length = 10	Markdown length = 15
2	6.74% (6.68%, 6.85%)	6.20% (6.00%, 6.37%)	7.61% (7.52%, 7.69%)	6.89% (6.53%, 7.13%)
3	1.55% (1.40%, 1.65%)	1.51% (1.05%, 1.80%)	2.57% (2.51%, 2.60%)	2.19% (1.89%, 2.50%)
4	4.26% (3.96%, 4.55%)	1.51% (1.15%, 1.87%)	7.04% (6.95%, 7.09%)	5.02% (5.00%, 5.05%)

Table 3.5: Inventory clearance (Mean, 95 percentile and max in parentheses)

	No-touch length = 3		No-touch length = 10	
Subclass	Markdown length = 10	Markdown length = 15	Markdown length = 10	Markdown length = 15
2	0.34% (0%, 49.6%)	0.25% (0%, 38.0%)	0.43% (0%, 57.2%)	0.26% (0%, 43.6%)
3	0.23% (0%, 50.7%)	0.14% (0%, 47.0%)	0.30% (0%, 57.9%)	0.18% (0%, 43.5%)
4	0.18% (0%, 55.4%)	0.05% (0%, 40.8%)	0.16% (0%, 49.3%)	0.04% (0%, 44.6%)

constraints. To gauge its performance, we have run a number of offline numerical experiments using retail data. The RPS algorithm exhibits revenue gains of around 2-7% over the retailer' existing pricing policy, and seems robust to different retailer settings such as the length of the markdown and no-touch periods.

Given the promising performance of RPS, Oracle Retail is currently building into their price optimization products the ability for retailers to use the RPS algorithm at selected stores for markdown optimization. Retailers will thus be able to easily pilot the algorithm at particular stores, and to compare the results with the current approach using A/B testing.

Another business impact of our work is that the offline simulator we have developed, which builds a ground truth demand model and simulates each selling season

for a range of business parameters using this ground truth model, will also become part of Oracle Retail's pricing software. This offline simulator will allow retailers to estimate the possible improvement in revenues from using the RPS algorithm before taking the risk of performing a pilot. We anticipate that this will make retailers using the pricing software more comfortable with performing a pilot, and thus speed up the adoption of the RPS algorithm.

Chapter 4

Inventory Allocation with Demand Learning for Seasonal Goods

We study an inventory allocation problem in a two-echelon (single-warehouse multiple-retailer) setting with lost sales. At the start of a finite selling season, a fixed amount of inventory is available at the warehouse, and can be allocated to the retailers over the course of the selling horizon with the objective of minimizing total expected lost sales costs and holding costs. We are particularly interested in demand learning in this context, where the decision maker can use historical demand observations to predict future demand, and consequently make allocation decisions. We thus model demand in order to capture learning, and show that our model can describe both demand forecasting (e.g. ARMA) frameworks, as well as a Bayesian framework. Then, we pose the questions of (1) how to solve the inventory allocation problem under demand learning in a computationally tractable way, and (2) how demand learning impacts effective inventory allocation policies. To address the first question, we adapt the Lagrangian relaxation-based technique proposed by Marklund and Rosling (2012) for a backordering, no-learning setting. We show under general assumptions that the resulting heuristic remains near-optimal in our setting, compared to the original dynamic program. Finally, we use this analysis to investigate the relationship between demand learning and early allocation decisions. We show through a combination of theoretical and numerical analysis the following intuitive result: Demand learning

provides an incentive for the decision maker to withhold inventory at the warehouse rather than allocating it in earlier periods.

4.1 Introduction

Motivated by a real example faced by a large fashion retailer based in the U.S, we study an inventory allocation problem for seasonal goods in a two echelon, or one-warehouse multiple-retailer, setting. At the start of each selling season, a fixed amount of inventory is available at the warehouse, and the decision maker must decide how to allocate this inventory from the warehouse to the retailers so as to minimize the total expected lost sales costs and holding costs. It has been widely observed in the literature on inventory allocation in two echelon systems (Eppen and Schrage, 1981; Jackson, 1988; Jackson and Muckstadt, 1989) that the warehouse has a strong incentive to make multiple allocations throughout the selling season, rather than allocate all its inventory to the retailers at the start of the selling season: If some retailers experience high demands in earlier periods, reserving inventory at the warehouse allows the retailer to rebalance inventory levels by allocating more inventory to these retailers (and less inventory to retailers who experience low demands) in later periods. This ability to mitigate demand fluctuations by storing inventory in a central location is known as the "risk-pooling" effect.

In this work, our chief contribution is to show that there is a second motive for the decision maker to delay inventory allocations to later periods: Namely, demand learning. Demand learning generally allows the decision maker to improve her demand forecasts with time, by observing historical demand and updating her demand beliefs in response to these observations. Intuitively, we would expect that by reserving inventory at the warehouse, the decision maker can make more informed allocation decisions later on in the time horizon, and allocate more inventory to retailers who have been observed to experience high demands, and less inventory to retailers who have been observed to experience low demands. In this work, we provide evidence to support this intuition by showing that we can expect the decision maker's first

period allocations from the warehouse to the retailers to *decrease* with the extent of learning.

To prove this property, we first address the question of how the decision maker can allocate inventory from the warehouse to the retailers in a computationally tractable way. Since the allocation problem becomes computationally intractable as soon as either the number of retailers or time periods grows large, and is also difficult to study analytically, we propose a heuristic method. Of the many heuristics proposed in the literature (Federgruen and Zipkin, 1984; Jackson, 1988; McGavin et al., 1993) for the two echelon inventory allocation problem, we base our work on the heuristic developed by Marklund and Rosling (2012) for a backordering, independent demand setting, since optimality bounds exist for this heuristic. The idea behind their method is that the source of the computational complexity in the allocation problem is that the fixed warehouse inventory couples the allocation decisions across the retailers. If each retailer had to pay some ordering cost for each unit of inventory, instead of satisfying the fixed warehouse inventory constraint exactly, the allocation problem would decouple, and each retailer could solve its inventory ordering problem separately.

Thus, following Marklund and Rosling (2012), we propose to relax the fixed inventory constraint to allow for this problem decomposition. We then prove the same optimality gap as in Marklund and Rosling (2012) between the expected costs of applying the heuristic, and the optimal value of the original allocation problem. However, a key difference between our result and theirs, besides the fact that they study a backordering setting and we study a lost sales setting, is that our result applies to a setting with correlated demands, which includes demand forecasting settings as well as settings with Bayesian learning. In Marklund and Rosling (2012), however, demands at each retailer are assumed to be IID across time.

Then, using the heuristic as a proxy for the exactly optimal solution, we investigate how the decision maker's allocation policy depends on demand learning. We formulate a simple, two period demand forecasting model with identical retailers that parametrizes the extent of demand uncertainty in the second period. We show analytically, by further approximating our lost sales setting with a perishable inventory

setting (where no remaining inventory at the retailers can be carried over to the next period), that the dependence of the first period allocation on the level of uncertainty is consistent with the property that early allocations should decrease as the amount of demand learning increases. This allows us to conclude that demand learning complements risk pooling in incentivizing the decision maker to reserve inventory at the warehouse, and to delay inventory allocations to later periods.

We also separately use the heuristic to prove an additional structural result on the allocation decisions when the retailers are non-identical. In particular, we look at a setting with independent but non-identical retailers experiencing truncated normal demands, and who share the same demand means but different variances. We ask how the retailer should prioritize among these retailers when allocating inventory. We show that the decision maker's strategy should on the amount of available warehouse inventory: When the warehouse inventory is small, the decision maker should favor a conservative policy, and allocate more inventory to retailers with lower demand variances, since these retailers have a lower chance of experiencing low demands. On the other hand, when the warehouse inventory is large, the decision maker should take a risk on retailers with higher demand variances, and allocate more inventory to these retailers, since they have a higher chance of experiencing high demands.

The rest of the chapter is organized as follows: The rest of this section gives a review of the related literature on inventory allocation in two echelon systems, both in settings with and without demand learning. Section 4.2 then describes our model and assumptions, and shows how our correlated demand model can capture demand learning in the sense of both a demand forecasting setting, such as when retailer demands are generated by ARMA or ARIMA processes, as well as a Bayesian setting, where retailer demand is given by some parameterized family of distributions, and where the unknown parameters are distributed according to priors held by the decision maker.

In Section 4.3, we present our heuristic, adapted from Marklund and Rosling (2012), that solves the inventory allocation problem in our lost sales, demand learning setting. We show that the asymptotic optimality bound from Marklund and Rosling

(2012) continues to hold in our setting under some general assumptions on the prices and holding costs at the retailers. Then, in Section 4.4, we analyze this heuristic to understand the structure of the near-optimal allocation policy. We look at how the decision maker's allocation decisions are affected by demand learning as well as by non-identical retailers. Finally, in Section 4.5, we present numerical experiments that validate the structural results in Section 4.4.

4.1.1 Literature Review

The study of inventory allocation in two echelon systems began with Clark and Scarf (1960), who observed that the complexity of this problem relative to the single retailer inventory ordering problem stems from the fact that that the retailers' inventory positions cannot be lowered through transshipments or returns to the warehouse, causing the optimal allocations at each time to depend not only on the total system stock, but on the inventory positions at *all* the retailers. Since the inventory allocation problem becomes computationally intractable as soon as either the number of retailers or time periods grows large, a number of papers have proposed effective but computationally tractable policies that rely on approximations of the original problem. Jackson (1988) considers a very specific class of policies, called "order-up-to-S" policies, where the warehouse stocks each retailer (all of whom are assumed to experience IID demand) up to S every period until it runs out of inventory. He proposes that when the warehouse runs out of inventory to allocate all the retailers' inventory levels up to S, it should solve a "run out allocation" problem, and develops approximations to efficiently solve this optimization problem. Jackson and Muckstadt (1989) study a two period model with backordering. They approximate the cost function by analyzing the case where the number of retailers tends to infinity, and use this approximation to develop an efficient optimization procedure. McGavin et al. (1993) and McGavin et al. (1997), like us, study the lost sales setting, but with only two periods and identical retailers. They show that the optimal policy takes the form of balancing policy, and also propose a heuristic, known as the infinite retailer heuristic, which estimates the first period optimal allocations and second period order-up-to levels by

approximating their set up with a deterministic setting with infinitely many retailers.

The above papers demonstrate the effectiveness of their proposed heuristics through numerical simulations. Marklund and Rosling (2012) is the first that we know of that proves a bound on the optimality gap between the expected cost of their heuristic and that of the exactly optimal solution. They study a backordering setting with N non identical retailers, each of whom experiences demand that is IID across time. They propose relaxing the constraint that the total allocations to the retailers across the selling horizon is at most the warehouse inventory: Instead of requiring that this constraint is satisfied almost surely, they relax it so that it only has to be satisfied in expectation. Then, they show that dualizing this relaxed constraint gives rise to N separate inventory ordering problems, each of which can be solved in a computationally tractable way. They prove that the ratio of the optimality gap between the heuristic and the original problem to the value of the optimal problem is bounded by $O(\sqrt{N})$, implying that this optimality gap goes to 0 as the number of retailers grows large. In this work, we adapt the result in Marklund and Rosling (2012) to our lost sales, correlated demand setting to get the same optimality gap bound of $O(\sqrt{N})$. The main technical contribution we have made in adapting their proof lies in showing the convexity of the relaxed optimization problem in the lost sales setting. We give a sufficient condition on the relationship between prices and holding costs that guarantees the convexity of this optimization problem, which ensures that their result holds in our setting.

A closely related stream of literature has also studied the risk pooling phenomenon, where storing inventory at a centralized location can help mitigate demand shocks that cause imbalances among retailer inventory positions. Eppen and Schrage (1981) study a backordering system, and find that backordering costs are lower if the depot acts as centralized ordering facility. Jackson (1988), mentioned above, finds through numerical simulations that the cost savings with risk pooling can be up to 70% compared to when inventory is not centralized. Jackson and Muckstadt (1989) analytically derive results that shed light on two different aspects of risk pooling: First, risk pooling causes the distribution of inventory to be more balanced across retailers.

Second, the centralization of inventory guarantees that the order up to levels at the retailers will be close to being constants.

Another stream of literature that our paper is closely related to is the literature on inventory allocation in a two echelon setting with demand learning. Many of these papers are motivated by real examples of allocation decision problems faced by fashion retail companies, where demand in later periods of a season can only be learned by observing early sales. Fisher and Raman (1996) study a two period model where the first period allocation is unconstrained but the total second period allocations is limited. They approximate the decision maker's optimization problem in order to solve it in a more computationally tractable way, and test their algorithm on data from the fashion retail company Sport Obermeyer. In addition, they look at the special case that demand is bivariate normal, and give a closed form solution of the optimal first period allocations to the retailers in terms of their demand means and variances. This is related to our result in Section 4.4, where we compare the allocations to retailers with different demand variances. However, the result in Fisher and Raman (1996) does not show how the warehouse should prioritize among retailers with low and high variances. In this work, we show that the warehouse's strategy should depend on the amount of available inventory at the warehouse.

Besides Fisher and Raman (1996), Fisher and Rajaram (2000) study the problem of merchandise testing, i.e. of allocating small amounts of inventory to a small number of selected retailers before the season starts in order to learn demand. They develop an algorithm to determine which retailers testing inventory should be allocated to in order to maximize learning during the testing period, and test their algorithm on data from a real fashion retailer. Gallien et al. (2017) work with Zara to study the problem of determining inventory allocations to retailers early in the season. They propose an algorithm that approximately solves this problem, prove an asymptotic optimality bound on the proposed algorithm, and run field experiments to validate its performance. We note that all three of these demand learning papers study two period models. This shows that many real world settings faced by fashion retailers can be formulated as two period models, and suggests that the model used in our

structural analysis in Section 4.4 to a two period setting is not only simpler to analyze, but also practical.

It is also worth comparing our work with several papers on demand learning for the single retailer setting, where the focus is also on structural properties of the optimal allocation. From this stream of literature, Ding et al. (2002) study a two period newsvendor model where demand learning can take place through Bayesian updates of unknown parameters based on observed sales data. They compare the optimal first period allocations when demand is censored (i.e. demand at each retailer can only be observed up to its inventory position) with the optimal allocations when demand is fully observable. They analytically derive the intuitive result that when demand is censored, the decision maker should allocate more inventory in the first period so as to obtain more accurate demand information. Azoury (1988) and Azoury and Miller (1984) also study a two period model, though with backordering, and like us, they compare the optimal first period allocations with and without learning. In their setting, however, demand learning takes place through Bayesian updates of the unknown parameters, whereas our structural analysis in Section 4.4 is based on a demand forecasting model. They prove that under specific assumptions on demand (such as the fact that it belongs to a family of distributions that satisfies what is known as the single crossing property), the first period allocations are greater without learning (when the parameter is updated) than when it is updated. The intuition behind this result is that in the learning setting, a parameter update could reveal that demand is on average lower than anticipated; Then, allocating too much inventory in the first period puts the decision maker at risk of having a higher inventory position than is optimal. In this work, we derive a similar structural result that says that the first period allocations are decreasing with the extent of learning. However, because we assume a multiple retailer setting with fixed inventory (whereas in Azoury (1988) and Azoury and Miller (1984) inventory is unlimited), the interpretation of our result is different. In our case, allocating less inventory when there is learning has to do with saving inventory for the second period, when demand uncertainty is lower, and it is clearer which retailers will experience high demands, and which retailers will

experience low demands.

4.1.2 Notation

Let $\mathbf{1_n}$ denote a vector of all ones of length n, and let $\mathbf{0_n}$ denote a vector of all zeros of length n. We will omit the subscript n when the dimension of these vectors is evident. For any m-dimensional vector $\mathbf{x}$, let $[\mathbf{x}]^+ := \max\{0, \mathbf{x}\}$ where max is the elementwise maximum.

4.2 Model

We study an inventory allocation problem for seasonal goods in a one-warehouse multiple-retailer setting. At the start of the selling season of length T, a fixed amount of inventory w_0 is available at the warehouse, while the N retailers have no starting inventory. At the start of each time period $t = 1, 2, \ldots, T$, the warehouse can choose to allocate some amount of inventory $a_{i,t} \geq 0$ to each retailer i. Let $\mathbf{a_t}$ denote the N dimensional vector with i^{th} entry $a_{i,t}$. Transshipments (moving inventory between retailers) or returning inventory from the retailers to the warehouse are not allowed, and we assume that there is no additional replenishment to the warehouse during the season. This implies that the total allocation to all retailers over the selling horizon satisfies $\sum_{i=1}^{N} \sum_{t=1}^{T} a_{i,t} \leq w_0$ almost surely.

We denote the starting (pre-allocation) inventory position at the warehouse at time t by w_t, and the starting inventory positions at the retailers by the N-dimensional vector $\mathbf{x_t}$. The post-allocation inventory positions at the retailers are denoted by the N-dimensional vector $\mathbf{y_t}$, where $\mathbf{y_t} = \mathbf{x_t} + \mathbf{a_t}$. After the warehouse allocates inventory to the retailers, each retailer i then sells the product at price $p_{i,t}$ and correspondingly observes demand $D_{i,t}$. This is an uncensored observation, i.e. the decision maker's knowledge of demand is not limited by the available inventory at the retailer. We assume that for each i and t, $D_{i,t}$ is a random variable that is discrete and bounded by some constant $D_{\max}$. We also assume that demand is independent from retailer to retailer. However, for each retailer, we allow the demands $D_{i,t}$ to

be correlated across time. In particular, the distribution of $D_{i,t}$ can depend on the history $\mathcal{H}_{i,t} = \{D_{i,1}, \ldots, D_{i,t-1}\}$ as well as other exogenous variables.

We study a lost sales setting with no leadtimes. Any demand that is not met is lost and not backordered, giving a starting inventory position at the retailers at time t of $\mathbf{x_{t+1}} = [\mathbf{y_t} - \mathbf{D_t}]^+$. Any demand that is not met by retailer i incurs a per-unit lost sales cost of $p_{i,t}$, where $p_{i,t}$ is the price of the product and $p_{i,t} \geq 0$. Any demand that remains at retailer i at time t incurs a per-unit holding cost of $h_{i,t}$, with $h_{i,t} \geq 0$. We ignore transportation costs.

The decision maker's objective is to minimize the expected discounted costs (for a given discount factor is α) incurred by the different retailers over the course of the selling horizon. At time t, given that retailer i has a post allocation inventory position of $y_{i,t}$, the cost incurred by this retailer is this is the sum of a lost sales component of $p_{i,t}[D_{i,t} - y_{i,t}]^+$ and a holding cost component of $h_{i,t}[y_{i,t} - D_{i,t}]^+$. We define the cost

$$L_{i,t}(y_{i,t}) := p_{i,t}[D_{i,t} - y_{i,t}]^+ + h_{i,t}[y_{i,t} - D_{i,t}]^+.$$

The decision maker then determines allocations to the retailers at each time period based on the warehouse inventory, the history of demand observations, and its belief of future demand. At each time t, we assume she has full knowledge of the history of past demand realizations $\{D_{i,s}, i = 1, \ldots, N, s = 1, \ldots, t-1\}$, as well as full knowledge of the distribution of $D_{i,t}$ for $i = 1, \ldots, N$ conditional on these past demand realizations. However, she does not know the future demand realizations beforehand.

Using the notation defined above, the warehouse's optimization problem at time t can be formulated with the following dynamic program:

$$V_t(w_t, \mathbf{x_t}) = \min_{\substack{\mathbf{y} \geq \mathbf{x_t} \\ \mathbf{1}^\mathsf{T}\mathbf{y} \leq \mathbf{1}^\mathsf{T}\mathbf{x_t} + w_t}} G_t(\mathbf{y}, w_t, \mathbf{x_t}) \tag{4.1}$$

$$G_t(\mathbf{y}, w_t, \mathbf{x_t}) = \sum_{i=1}^{N} \mathrm{E}[L_{i,t}(\mathbf{y})] + \mathrm{E}[V_{t+1}(w_t - \mathbf{1}^\mathsf{T}(\mathbf{y} - \mathbf{x_t}), [\mathbf{y} - \mathbf{D_t}]^+)] \tag{4.2}$$

$$V_{T+1}(w_{0,T+1}, \mathbf{x_{T+1}}) = 0. \tag{4.3}$$

4.2.1 Demand Models with Learning/Forecasting

Our demand model allows the demands at each retailer to be correlated across time. Two types of correlated demand models that this framework can capture are (1) demand forecasting models, such as ARMA or ARIMA models, as well as (2) Bayesian models, where the decision maker updates her beliefs on the unknown demand model parameter distributions with time. Examples of these kinds of models are described below.

Demand forecasting Consider the following AR(1) model: $D_{i,t} = \alpha D_{i,t-1} + \epsilon_{i,t}$ for $t = 2, \ldots, T, i = 1, \ldots, N$, where the demand at each retailer and time period is the sum of the previous period demand at that retailer, as well as some demand noise. Then if the decision maker knows the parameter α, as well as the distributions of the demand noises and the first period demands at the different retailers, the distribution at any time period conditional on the previous period demands is fully known, and the model fits our assumptions on demand. In this vein, we can capture other demand forecasting models as long as the parameters are known to the decision maker.

A Bayesian model Consider the following Bayesian model, where the demands at each retailer i are known to be IID across time, and are Poisson distributed with mean θ_i. If the decision maker has priors on the parameters $\{\theta_i, i = 1, \ldots, N\}$, she can update the posteriors on these parameters based on the demand realization history, and thus always knows the distribution of demand at any retailer and time period conditional on this history. Note that since we have assumed that the retailers are independent, the parameters $\{\theta_i, i = 1, \ldots, N\}$ have to be non identical across retailers.

In Section 4.4, we use a demand forecasting model to investigate the impact of demand learning on effective inventory allocation policies. Although this model is different from machine learning and statistical learning models in that the decision maker is not learning the underlying demand distributions with time, what it has in common with learning models is that the decision maker can use historical demand observations to make more accurate predictions of future demand, and consequently

make more informed allocation decisions.

4.3 Heuristic

The dynamic program (4.1)-(4.3) is difficult to solve explicitly, as it suffers from the curse of dimensionality, where its state space grows exponentially in N and T. In this section, we consider a heuristic that is suboptimal but computationally much less expensive to implement. This heuristic is based on the observation that (4.1)-(4.3) is weakly coupled, meaning that if the constraint $\mathbf{1}^\top \mathbf{y} \leq \mathbf{1}^\top \mathbf{x_t} + w_t$ is relaxed, the problem decouples into N separate dynamic programs, reducing the computational complexity to linear in N. The idea is to achieve this problem decomposition by approximating the original problem with its Lagrangian relaxation. We dualize the coupling constraint and add an associated Lagrangian term to the objective function, thus decomposing the problem into N separable optimization problems. This Lagrangian relaxation technique has been used to approximate weakly coupled optimization problems for a variety of applications. A survey is given in Adelman and Mersereau (2008). In the literature on inventory allocation and risk pooling, Marklund and Rosling (2012), who also study an inventory allocation problem for single warehouse multiple retailer setting, but with backordering instead of lost sales, propose a heuristic based on this technique. They show that this relaxation technique gives a lower bound on the optimal cost-to-go function, and that the performance of the heuristic converges to the lower bound as N goes to infinity, meaning that the heuristic is asymptotically optimal.

Below, we show how Marklund and Rosling (2012)'s approach can be adapted to our lost sales setting. Our demand model, like theirs, also assumes that the retailers are independent. However, while Marklund and Rosling (2012) assume that each retailer's demands are IID across time, we allow the demands experienced by each retailer to be correlated across time, such as in the demand forecasting and Bayesian models described in the previous section.

4.3.1 Algorithm

We start by rewriting (4.1)-(4.3) as a stochastic program as follows:

$$\begin{aligned}
\min \quad & \sum_{t=1}^{T}\sum_{i=1}^{N} \mathrm{E}[\mathrm{E}[L_{i,t}(y_{i,t})|\mathcal{H}_{i,t}]] \\
\text{subject to} \quad & y_{i,t} = x_{i,t} + a_{i,t} \ \forall i,t,\mathcal{H}_{i,t} \\
& x_{i,t+1} = [y_{i,t} - D_{i,t}]^{+}, \ \forall i,t,\mathcal{H}_{i,t} \\
& \sum_{i=1}^{N}\sum_{t=1}^{T} a_{i,t} \leq w_0 \text{ a.s.} \\
& a_{i,t} \geq 0 \ \forall i,t,\mathcal{H}_{i,t}\,.
\end{aligned} \tag{P1}$$

Here, the variables $a_{i,t}, x_{i,t}, y_{i,t}$ are non-anticipative and are functions of the demand history $\mathcal{H}_{i,t}$.

Now, to achieve the desired problem decomposition, we will relax the inventory constraint $\sum_{i=1}^{N}\sum_{t=1}^{T} a_{i,t} \leq w_0$ by requiring that it is only satisfied in expectation over all demand realizations, rather than almost surely. This gives the constraint $\sum_{i=1}^{N}\sum_{t=1}^{T} \mathrm{E}[a_{i,t}] \leq w_0$. We will also eliminate the state variables $x_{i,t}$ and $y_{i,t}$ from the formulation and express the problem in terms of the decision variables $a_{i,t}$. This can be achieved by using the equations

$$\begin{aligned}
x_{i,t} &= \max\{0, \sum_{s=u}^{t-1} a_{i,s} - D_{i,s}, \ \forall 1 \leq u \leq t-1\}\,, \\
y_{i,t} &= x_{i,t} + a_{i,t}\,.
\end{aligned}$$

This gives

$$\begin{aligned}
\min \quad & \sum_{t=1}^{T}\sum_{i=1}^{N} \mathrm{E}[\mathrm{E}[\alpha^{t-1} L_{i,t}(a_{i,t} + \max\{0, \sum_{s=u}^{t-1} a_{i,s} - D_{i,s}, \ \forall 1 \leq u \leq t-1\})|\mathcal{H}_{i,t}]]) \\
\text{subject to} \quad & \sum_{i=1}^{N}\sum_{t=1}^{T} \mathrm{E}[a_{i,t}] \leq w_0 \\
& a_{i,t} \geq 0 \ \forall i,t,\mathcal{H}_{i,t}\,.
\end{aligned} \tag{P2}$$

The relaxed constraint $\sum_{i=1}^{N}\sum_{t=1}^{T} \mathrm{E}[a_{i,t}] \leq w_0$ can then be dualized. This gives the following decomposition:

$$\begin{aligned} \max \quad & -\lambda w_0 + \mathrm{SUB}_i(\lambda) \\ \text{subject to} \quad & \lambda \geq 0\,, \end{aligned} \tag{D1}$$

where $\mathrm{SUB}_i(\lambda)$ is the solution to the single retailer inventory ordering problem

$$\begin{aligned} \min \quad & \sum_{t=1}^{T}\sum_{i=1}^{N} \mathrm{E}[\mathrm{E}[\alpha^{t-1} L_{i,t}(a_{i,t} + \max\{0, \sum_{s=u}^{t-1} a_{i,s} - D_{i,s},\ \forall 1 \leq u \leq t-1\}) | \mathcal{H}_{i,t}]] \\ & + \lambda \alpha^{t-1} \sum_{t=1}^{T}\sum_{i=1}^{N} \mathrm{E}[a_{i,t}] \\ \text{subject to} \quad & a_{i,t} \geq 0 \ \forall i, t, \mathcal{H}_{i,t}\,. \end{aligned} \tag{D2}$$

(D2) is a classical inventory ordering problem, with ordering costs λ. When only a limited amount of inventory is available at the warehouse, λ is large, i.e. the retailers have to pay a large penalty to order inventory. On the other hand, when the warehouse inventory is large, λ is closer to 0. For a given λ, the inventory ordering problem for each retailer can be solved recursively, by searching a (finite) set of possible allocations. The complexity of this problem grows exponentially in terms of the length of the selling horizon T. However, since the retailers' subproblems can be solved independently, it is far more tractable to solve these subproblems in parallel for a given λ than the original allocation problem (P1).

Before we show how to find the optimal λ^* that maximizes (D1), we discuss the convexity of the allocation problem (P2). We can show that under some conditions on the prices and holding costs at the retailers, namely that the price at each retailer and time period is at most the sum of the price and holding cost at that retailer in the previous time period, the relaxed optimization problem (P2) is convex. An intuitive interpretation of this condition is that there is no incentive for any retailer to withhold inventory for the next period: That is, if a retailer knows in advance that there will be at least a unit of demand in a particular time period, it is never profitable for

that retailer to withhold inventory, leave that demand unsatisfied, and instead satisfy demand in the next period. This convexity result is stated in Lemma 1 and proven in Appendix C.1.1. We note that a sufficient condition for $p_{i,t} + h_{i,t} \geq \alpha p_{i,t+1}$ is that the prices at each retailer are non-increasing with time. We thus expect that this condition, and thus the convexity of (P2), holds for many fashion retail applications, where the practice of markdown pricing means that prices tend to decrease with time.

Lemma 1. *Assume that for all $t < T$, $p_{i,t} + h_{i,t} \geq \alpha p_{i,t+1}$. Then* (P2) *is a convex optimization problem.*

Given the convexity result in Lemma 1, we can solve for the optimal λ, λ^*, that maximizes (D1) using the complementary slackness condition that either $\lambda^* = 0$ or $\sum_{t=1}^{T} \sum_{i=1}^{N} \mathrm{E}[a^*_{i,t}(\lambda^*)] = 0$. We know that λ^* must always always fall within the range $[0 \ \max_{i=1,\ldots,N,t=1,\ldots,T} p_{i,t}]$, since for all λ such that $\lambda \geq \max_{i=1,\ldots,N,t=1,\ldots,T} p_{i,t}$, the optimal allocation policy is simply to not allocate any inventory to any of the retailers. Then the optimal λ, λ^*, can be found by searching this interval, through a technique such as bisection, for the λ that satisfies the complementary slackness conditions. We can calculate this λ^* prior to the start of the selling horizon, and allocate inventory to the retailers according to this policy, without needing to resolve for λ, until the warehouse runs out of inventory. These steps are summarized in the following inventory allocation heuristic:

1. Before $t = 1$, find λ^* that maximizes (D1). One possible approach is to use bisection to search the interval $[0 \ \max_{i=1,\ldots,N,t=1,\ldots,T} p_{i,t}]$ until a λ that satisfies the complementary slackness conditions, $\lambda = 0$ or $\sum_{t=1}^{T} \sum_{i=1}^{N} \mathrm{E}[a_{i,t}] = 0$, is found.

2. For each retailer, solve (D2) with λ set to λ^*, and calculate the corresponding optimal allocations $a^*_{i,t}(\lambda^*), i = 1, \ldots, N, t = 1, \ldots, T$.

3. For each $t = 1, \ldots, T$, and retailer $i = 1, \ldots, N$, allocate the minimum of $a^*_{i,t}(\lambda^*)$ and the remaining inventory at the warehouse.

4.3.2 An Optimality Bound

In the backordering setting, and assuming that demand is IID across time, Marklund and Rosling (2012) show that the cost of implementing the heuristic described in the previous section is at most $O(\sqrt{N})$ greater than the optimal value of their original dynamic program, where N is the number of retailers. We will show that given the assumptions on demand laid out in Section 4.2, their argument can be adapted to our lost sales, correlated demand setting to prove an optimality bound of the same order on our inventory allocation heuristic. This result is stated in Theorem 5 below.

Theorem 5. *Assume that the condition for convexity, $p_{i,t} + h_{i,t} \geq \alpha p_{i,t+1}$, is satisfied for all t such that $t < T$. Denote the expected cost of our heuristic by UB, and denote the value of* (P1) *by OPT. We have the following relationship between UB and OPT: $UB - OPT$ is $O(\sqrt{NT})$.*

If we assume further that the value of each retailer subproblem (D2) is always lower bounded by a constant for $\lambda = \lambda^*$, it is easy to see that OPT is lower bounded by N times this constant. Then, Theorem 5 implies that $\frac{UB-OPT}{OPT}$ is $O(1 + \frac{1}{\sqrt{N}})$.[1] The heuristic would thus be asymptotically optimal in the number of retailers N, in that its expected costs converge to the value of the original dynamic program (P1) as N grows large. However, the heuristic is not necessarily asymptotically optimal in the length of the selling horizon T.

The proof of Theorem 5 is deferred to Appendix C.1.2. The idea behind the proof is that the value of (D1) is a lower bound on (P1). This follows from the fact that the convexity of (P2) implies strong duality. Thus the relaxed optimization problem (P2), whose value is a lower bound on (P1), is equal to its dual (D1). Now the cost of the heuristic is clearly an upper bound on the optimal value of the problem. Since the value of (D1) is a lower bound on (P1), the cost of applying the dual policy is greater than the optimal value of the problem exactly when the sum of the recommended allocations is greater than its expected value. Since the retailers are independent, the

[1] The value of each retailer subproblem (D2) is not always lower bounded by a constant for the optimal λ^*. For example, if the holding costs $h_{i,t} = 0$ and $w_0 \geq D_{\max} NT$, $\lambda^* = 0$, and the value of each (D2) would be 0.

difference between the sum of the recommended allocations and its expected values is of order $\sqrt{N}$ rather than N, thus allowing us to bound the optimality gap as a factor of $\sqrt{N}$.

We would like to make two remarks comparing the analysis of Theorem 5 with the analysis in Marklund and Rosling (2012). Firstly, in the backordering setting studied in Marklund and Rosling (2012), the convexity of the relaxed inventory allocation problem is guaranteed regardless of the assumptions on prices and holding costs. However, in our lost sales setting, we prove a sufficient condition on prices and holding costs to guarantee the convexity of (P2). Secondly, our heuristic and optimality bound, unlike Marklund and Rosling (2012), can be applied to settings with correlated demand, and hence with demand learning in the forecasting and Bayesian senses described in Section 4.2.1. However, it is important to note that the optimality bound 5 is not a regret bound (unlike the optimality bounds on the online learning algorithms presented in Chapter 2), and that it says nothing about the rate of learning any underlying demand model parameters.

4.4 Structural Results

The heuristic presented in Section 4.3 allows us to solve the two echelon inventory allocation problem in a computationally tractable way by expressing this problem in terms of the simpler single retailer inventory ordering problem. In this section, we will also use this connection between the two problems to shed light on the structure of effective inventory management policies. Using our heuristic as a proxy for the exactly optimal solution, we investigate how the decision maker's inventory management policies depend on (1) demand learning, and (2) different levels of demand uncertainty among the retailers.

To address these questions, we limit our analysis in the rest of this work to a two period setting ($T = 2$), but continue to allow N to be arbitrary. As discussed in the literature review in Section 4.1.1, two period models are widely used in the literature on inventory allocation in two echelon systems with demand learning, including in

papers that study real settings faced by fashion retailers (Fisher and Raman, 1996; Fisher and Rajaram, 2000; Gallien et al., 2017). Thus we expect the two period setting studied in this section to be not only simpler to analyze, but to also be of practical relevance.

4.4.1 Demand learning

We first investigate the impact of demand learning on the optimal allocation policy. We assume that the N retailers are identical in that they experience identically distributed demands, and experience price p and holding cost h in both time periods. We then model demand learning using the following demand forecasting model: In the first period, each retailer i experiences demand D_i, where $\{D_i, i = 1, 2\}$ are IID. At the end of the first period, the decision maker observes uncensored demands $\{D_i, i = 1, \ldots, N\}$. The demand experienced by retailer i in the second period is then given by

$$D_i + \rho\epsilon_i \text{ for some } \rho > 0 \tag{4.4}$$

where the demand noises $\{\epsilon_i, i = 1, \ldots, N\}$ are IID and mean 0. The decision maker thus knows a component of the second period demand beforehand, and can allocate inventory to the retailers in the second period based on the first period demand observations. The component of demand that is not learned, $\rho\epsilon_i$, has variance proportional to ρ^2. ρ is thus a measure of the amount of learning - as ρ increases, the second period demand forecast accuracy decreases. We are interested in how the first period allocations to the retailers depend on this parameter ρ, or, equivalently, on the extent of learning.

Unfortunately, it is difficult to analyze the exactly optimal allocation policy, or even the heuristic allocations, in our lost sales setting. We thus first approximate the inventory allocation problem (P1) with (P2), which relaxes the fixed inventory constraint, then make a further approximation that assumes that no inventory that is left over at the retailers after demand is observed can be carried over from the first period to the second period. We then analyze the corresponding optimal solution.

Although Theorem 5 shows that (P2) is a good approximation of (P1) when the number of retailers N grows large, we have no theoretical bounds on the quality of the second approximation. Intuitively, however, we expect that it is more reasonable to assume that inventory at the retailers cannot be carried over when the warehouse inventory is limited compared to the demands experienced by the retailers. If the warehouse inventory level is limited, then the inventory positions at the retailers after demand is observed would be close to zero.

The dynamic program representing this approximation is given below in (P3). Similar to the analysis in Section 4.3, we can show that the inventory constraint can be dualized, causing the problem to separate into N independent single-retailer inventory ordering problems. The dual problem is given below in (D3).

$$\begin{aligned} \min \quad & \sum_{t=1}^{T}\sum_{i=1}^{N} \mathrm{E}[\mathrm{E}[\alpha^{t-1} L_{i,t}(a_{i,t} + \max\{0, \sum_{s=u}^{t-1} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t-1\})|\mathcal{H}_{i,t}]] \\ \text{subject to} \quad & \sum_{i=1}^{N}\sum_{t=1}^{T} \mathrm{E}[a_{i,t}] \le w_0 \\ & a_{i,t} \ge 0\ \forall i, t, \mathcal{H}_{i,t}. \end{aligned} \tag{P3}$$

$$\begin{aligned} \max \quad & -\lambda w_0 + \mathrm{SUB}_i(\lambda) \\ \text{subject to} \quad & \lambda \ge 0, \end{aligned} \tag{D3}$$

where $\mathrm{SUB}_i(\lambda)$ is the solution to the single retailer inventory ordering problem

$$\begin{aligned} \min \quad & \sum_{t=1}^{T}\sum_{i=1}^{N} \mathrm{E}[\mathrm{E}[\alpha^{t-1} L_{i,t}(a_{i,t})|\mathcal{H}_{i,t}]] \\ & + \lambda\alpha^{t-1} \sum_{t=1}^{T}\sum_{i=1}^{N} \mathrm{E}[a_{i,t}] \\ \text{subject to} \quad & a_{i,t} \ge 0\ \forall i, t, \mathcal{H}_{i,t}. \end{aligned} \tag{D2}$$

Since the retailers have identically distributed demands, the first period allocations that optimize (D3) are by symmetry the same across all retailers. This allocation $a^*_{i,1}(\rho)$ depends on ρ, the standard deviation of the demand forecasting error in the

second period. In fact, we can show that for sufficiently small warehouse inventory levels w_0, $a^*_{i,1}(\rho)$ is strictly increasing in the parameter ρ, i.e. the first period allocations are strictly decreasing with the extent of learning. This result is stated in Theorem 6 below and is proven in Appendix C.1.3.

Theorem 6. *For the perishable inventory approximation described above, there exists some maximum warehouse inventory level $w_{\max}$, such that for $w_0 < w_{\max}$, the optimal first period allocation to each retailer, $a^*_{i,1}(\rho)$, is strictly increasing in the parameter ρ.*

An intuitive explanation of the result in Theorem 6 is that when the decision maker is able to forecast the second period demand more accurately, she should save more of the available warehouse inventory for the second period, as she will derive more value from deploying this inventory in the second period rather than in the first period. Theorem 6 thus suggests that demand learning has the same effect as risk pooling in incentivizing the decision maker to reserve inventory at the warehouse, and to delay inventory allocations to later periods.

4.4.2 Nonidentical retailers

We now look at a different setting with non-identical retailers, and investigate the impact of different levels of demand uncertainty among the retailers on the inventory allocation decisions. If the mean demand varies from retailer to retailer, intuition says that it is generally optimal to allocate more inventory to retailers with higher demand means. However, the structure of the optimal policy is less clear when retailers have the same means but different variances. On the one hand, the decision maker could allocate more inventory to retailers with lower variance demands (i.e. lower demand uncertainty), as they have a lower chance of experiencing lower demands. On the other hand, the decision maker could take a risk on retailers with high variance demands, even though these retailers have a greater chance of experiencing lower demands, as these retailers also have a greater chance of experiencing higher demands.

To understand how the decision maker should balance between these tradeoffs,

we analyze a two period setting where the prices and holding costs are the same across retailers and time periods. Denote this price by p, and this holding cost by h, and assume further that $p > h$ (this is true of many retail settings, where items are sold at prices that are high compared to other costs). We model the demands at the retailers using truncated normal demand distributions that are symmetrical about their means. For each retailer i, we let demand $D_{i,t}$ have mean μ and range $[\mu - b \ \mu + b]$, i.e. the means and ranges are kept constant across retailers. Since we are interested in isolating the impact of different levels of demand uncertainties among the retailers, only the variances $\{\sigma_i, i = 1, \ldots, N\}$ corresponding to demands $\{D_{i,t}, i = 1, \ldots, N\}$ may be non-identical across retailers.

We once again analyze the allocations chosen by the heuristic (i.e. the optimal solution to (D1)) rather than the exactly optimal solution. We find that the decision maker's allocations under the heuristic depend on the starting inventory at the warehouse. If this is large, then the decision maker should favor a risk taking policy, and should allocate more inventory to retailers with higher demand variances. However, if the starting inventory at the warehouse is small, the decision maker should be more cautious, and should allocate more inventory to retailers with lower demand variances. This result is stated in Theorem 7, and is proven in Appendix C.1.4

Theorem 7. *There exists a warehouse starting inventory level $w_{\min}$ such that for $w_{\min} \leq w_{0,1}$, the heuristic's first period allocations, denoted by $\{a^*_{i,1}, i = 1, \ldots, N\}$ satisfy $a^*_{i,1} < a^*_{j,1}$ whenever $\sigma_i < \sigma_j$. Similarly, the second period order-up-to levels, denoted by $\{x^*_{i,2}, i = 1, \ldots, N\}$ also satisfy $x^*_{i,1} < x^*_{j,1}$ whenever $\sigma_i < \sigma_j$.*

There also exists a warehouse starting inventory level $w_{\max}$, $w_{\max} > 0$, such that for $w_{0,1}$ satisfying $0 \leq w_{0,1} \leq w_{\max}$, the heuristic's first period allocations, denoted by $\{a^*_{i,1}, i = 1, \ldots, N\}$ satisfy $a^*_{i,1} > a^*_{j,1}$ whenever $\sigma_i < \sigma_j$. Similarly, the second period order-up-to levels, denoted by $\{x^*_{i,2}, i = 1, \ldots, N\}$ also satisfy $x^*_{i,1} > x^*_{j,1}$ whenever $\sigma_i < \sigma_j$.

4.5 Numerical Experiments

To verify the structural properties proven in Section 4.4, we ran numerical simulations on synthetic data. These simulations consider simple settings with two periods, two to three retailers, and demands that are drawn from either discrete uniform distributions, or from discretized truncated normal distributions. The exactly optimal first period allocations (i.e. the allocations that solve (P1)), as well as the heuristic first period allocations (i.e. the allocations solving (D2) for a given value of the dual variable λ) can be easily solved for these settings by first determining the optimal second period allocation policy, then recursively computing the expected costs of the first period allocations and searching over the finite set of demands to determine the optimal allocation.

4.5.1 Demand learning

We first ran a set of simulations to empirically investigate the impact of demand learning on the first period allocations. As in Section 4.4.1, we model learning using the demand forecasting model (4.4). While it is difficult to theoretically analyze this model for our lost sales setting, we were able to compute the exactly optimal first period allocations (i.e. the allocations that solve (P1)) using the recursive procedure described above. This then allowed us to numerically investigate the relationship between the exactly optimal allocations and the parameter ρ, which is proportional to the standard deviation of the second period demand forecasting error, and is therefore inversely related to the extent of learning.

For these simulations, we fixed the prices at \$1, the holding costs at \$0.20, and the starting warehouse inventory at $w_{0,1} = 12$, but varied the distributions of demand at the retailers. First we simulated a setting where for all retailers i, period 1 demand $D_{i,1}$ is drawn from a truncated normal distribution with parameters μ, σ, a, b, i.e. $D_{i,1}$ is normally distributed according to $\mathcal{N}(\mu, \sigma)$ conditional on $D_{i,1}$ belonging to the interval $[a\ b]$. We discretized this distribution by discretizing the interval $[a\ b]$ with stepsize 0.01, and forcing demand to take values from this discretized set. We then

set $\sigma = 1$, varied the mean demand μ in the set $\{2, 2.5, 3, 3.5, 4\}$, and for each μ set $a = \mu - 1, b = \mu + 1$.

Next, we simulated a setting where each retailer i experiences first period demand that is drawn from a discrete uniform distribution. For a given mean demand μ, period 1 demand $D_{i,1}$ is drawn from the set $\{\mu - 1, \mu - 1 + 0.01, \ldots, \mu + 1 - 0.01, \mu + 1\}$ with equal probability. Again, we varied the mean demand μ within the set $\{2, 2.5, 3, 3.5, 4\}$. For both sets of simulations, the period 2 demand noise $\epsilon_{i,1}$ is drawn from a truncated normal distribution with parameters $\mu = 0, \sigma = 1, a = -1, b = 1$.

For each choice of demand distribution, we solved for the optimal allocations given $\rho \in \{0, 0.2, 0.4, 0.6, 1\}$. We solved the retailers' optimal allocations as follows: First, we computed the optimal second period allocations for each retailer corresponding to each possible tuple of the first period allocation y, and realized demands $D_{i,1}$ and $D_{i,2}$. Then, using the property that the optimal first period allocations must by symmetry be the same for both retailers, we recursively computed the expected cost of each possible first period allocation from the set $\{0, 0.01, 0.02, \ldots, 6\}$, and selected the allocation minimizing this cost.

Figures 4-1a and 4-1b plot these exactly optimal first period allocations to the retailers for different values of the learning parameter ρ. Figure 4-1a corresponds to the setting where the first period demand is drawn from a discretized truncated normal distribution, and Figure 4-1b corresponds to the setting where first period demand is uniformly distributed. For both these settings, and for all values of the mean demand, we see that the optimal first period allocations are indeed always increasing in ρ. This agrees with our finding in Theorem 6 that the first period allocations are decreasing as the extent of learning increases, which is in turn consistent with the property that demand learning, like the risk pooling effect, incentivizes reserving inventory at the warehouse for later periods. However, unlike Theorem 6, which makes several approximations of the inventory allocation problem (P1) and analyzes the allocation decisions corresponding to the approximate optimization problem, Figures 4-1a and 4-1b give the allocations that exactly solve the inventory allocation problem in its original form. These results thus suggest that the demand learning effect is a property

of our original allocation problem, and not just of the approximate problem studied in Theorem 6.

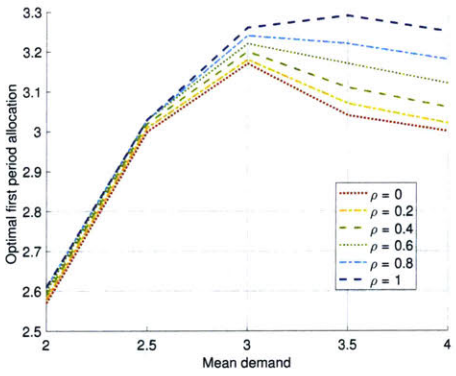
To gauge the quality of the approximation proposed in Section 4.4.1, we also looked at the first period allocations under the approximation that inventory that is left over at the retailers at the end of the first period cannot be carried over to the next period.

Finally, it is also interesting to note that our plots of the exactly optimal solutions in Figures 4-1a and 4-1b show that for our demand forecasting model, the optimal first period allocations are *not* necessarily monotonically increasing in the mean demand at the retailers. This is contrary to what we would observe if the demands at each retailer were IID across time periods, and indeed, this effect becomes less pronounced as ρ increases from 0 to 1 (i.e. as the periods 1 and 2 demands become less strongly correlated). Although this result may seem counterintuitive, we interpret it as being related to the property that demand learning can incentivize saving inventory for the second period: As the demand means increase, and the warehouse inventory becomes more limited with respect to demand, the decision maker derives more value from deploying this limited inventory in the second period, when an improved demand forecast is available.

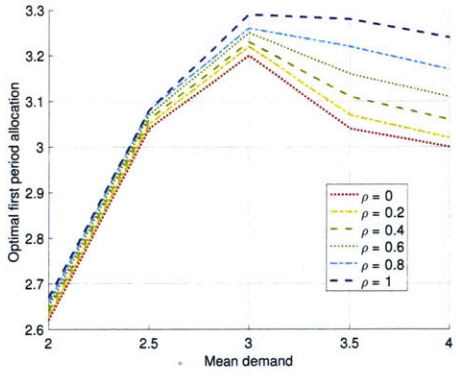
4.5.2 Non-identical retailers

We also conducted a second set of simulations on a setting with non-identical retailers in order to verify Theorem 7. For these simulations, we considered a set up with three retailers, all of whom experience demand drawn from truncated normal distributions with parameters μ, σ_i, a, b, i.e. retailer i's demands $D_{i,t}$ are normally distributed according to $\mathcal{N}(\mu, \sigma^2)$, conditional on $D_{i,t}$ belonging to the range $[a\ b]$. As in the simulations on demand learning, we discretized these truncated normal distributions by discretizing the interval $[a\ b]$ with stepsize 0.01, and forcing demand to take values from the discretized set.

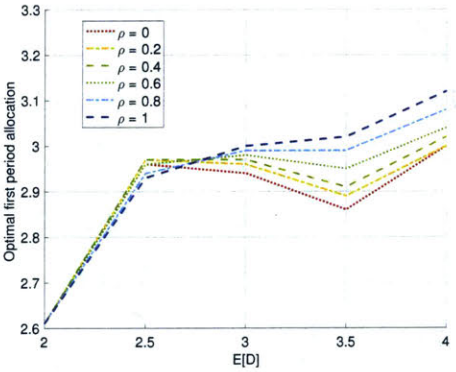
For the three different retailers, we kept the demand means and ranges fixed at $\mu = 2, a = 1, b = 3$, but varied the demand variances. For Retailer 1, or the 'low



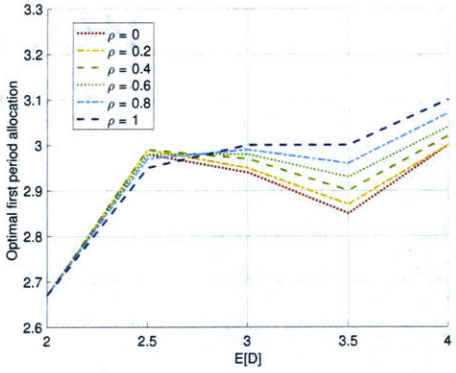
(a) Optimal first period allocations when retailers experience truncated normal demand in period 1 with parameters μ, σ, a, b (i.e. each $D_{i,1}$ is normally distributed according to $\mathcal{N}(\mu, \sigma)$ conditional on $D_{i,1}$ belonging to the interval $[a\ b]$). $\sigma = 1$, mean demand μ is drawn from $\{2, 2.5, 3, 3.5, 4\}$, and for each μ, $a = \mu - 1, b = \mu + 1$.



(b) Optimal first period allocations when retailers experience uniformly distributed demand in period 1, $D_{i,1} \sim \mathcal{U}[\mu-1,\ \mu+1]$. Demand mean μ is drawn from the set $\{2, 2.5, 3, 3.5, 4\}$.



(c) Optimal first period allocations when retailers experience truncated normal demand in period 1 with parameters μ, σ, a, b (i.e. each $D_{i,1}$ is normally distributed according to $\mathcal{N}(\mu, \sigma)$ conditional on $D_{i,1}$ belonging to the interval $[a\ b]$). $\sigma = 1$, mean demand μ is drawn from $\{2, 2.5, 3, 3.5, 4\}$, and for each μ, $a = \mu - 1, b = \mu + 1$.



(d) Optimal first period allocations when retailers experience uniformly distributed demand in period 1, $D_{i,1} \sim \mathcal{U}[\mu-1,\ \mu+1]$. Demand mean μ is drawn from the set $\{2, 2.5, 3, 3.5, 4\}$.

Figure 4-1: First period optimal and approximate allocations for fixed starting warehouse inventory, $w_{0,1} = 12$, prices = \$1, holding costs = \$0.20, and $\rho = 0, 0.2, 0.4, 0.6, 0.8, 1$. In both the truncated normal and uniform demand settings, period 2 demand noise $\epsilon_{i,1}$ is drawn from a truncated normal distribution with parameters $\mu = 0, \sigma = 1, a = -1, b = 1$

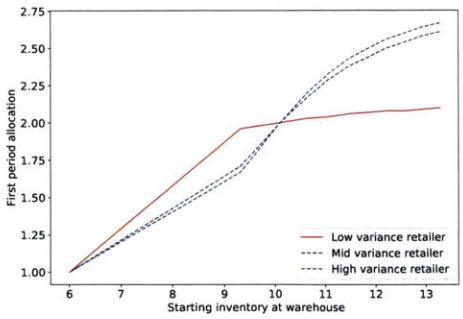
variance retailer', we set $\sigma = 0.1$. For Retailer 2, or the 'mid variance retailer', we set $\sigma = 1$. Finally, for Retailer 3, or the 'high variance retailer', we set $\sigma = 5$.

We then computed the first period heuristic allocations to the different retailers. Instead of solving for the λ that optimizes the dual problem (D1) for some given level of the starting warehouse inventory w_0, we varied λ between 0 and price p (since, as we have observed in Section 4.3, the optimal λ always lies within this range), and solved for the first period allocations that optimize (D2) given this λ. This was done recursively: First, using the well known result that the optimal allocation policy for the single retailer inventory ordering problem is an order-up-to policy, we computed the second period order-up-to levels by discretizing the interval [1 3] with a stepsize of 0.01 (i.e. for each retailer, we searched the set $[1, 1.001, 1.002, \ldots, 3]$ for the second period newsvendor levels). We then recursively computed the optimal first period allocation by once again discretizing the interval [1 3], and selecting the allocation from this set with the lowest expected cost. By also recursively computing the total expected allocations across all retailers and time periods, we were able to obtain the warehouse inventory w_0 corresponding to our chosen λ.

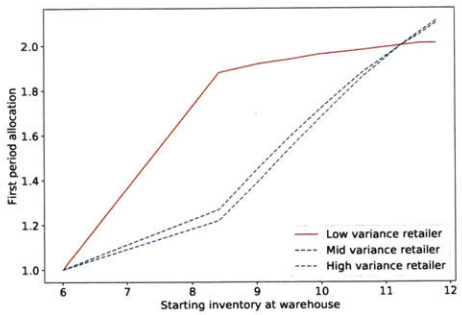
Figure 4-2a plots the heuristic's first period allocations to the three retailers when prices at all retailers are set to \$1 and holding costs are set to \$0.20, and Figure 4-2b plots the heuristic's first period allocations to the three retailers when prices at all retailers are set to \$1 and holding costs are set to \$0.80. For both configurations of the price and holding cost, we see that the results directly verify Theorem 7: For sufficiently small starting warehouse inventory, the first period heuristic allocations are decreasing in the variance of the retailers. However, when the starting warehouse inventory is sufficiently large, the first period heuristic allocations are increasing in the variance of the retailers.

4.6 Conclusion

We have studied a two echelon inventory allocation problem for a lost sales, correlated demand setting. We have shown that the heuristic developed in Marklund and Rosling



(a) Heuristic allocations with prices = \$1, Holding costs = \$0.20



(b) Heuristic allocations with prices = \$1, Holding costs = \$0.80

Figure 4-2: First period allocations under the heuristic in a setting with 3 retailers, all off whom experience demand drawn from truncated normal distributions with parameters μ, σ_i, a, b (retailer i's demands $D_{i,t}$ are normally distributed according to $\mathcal{N}(\mu, \sigma^2)$, conditional on $D_{i,t}$ belonging to the range $[a\ b]$.) For Retailer 1, the 'low variance retailer', $\mu = 2, \sigma = 0.1, a = 1, b = 3$. For Retailer 2, the 'mid variance retailer', $\mu = 2, \sigma = 1, a = 1, b = 3$, and for Retailer 3, the 'high variance retailer', $\mu = 2, \sigma = 5, a = 1, b = 3$.

(2012), which uses a Lagrangian relaxation technique to reduce the computationally intractable two echelon inventory allocation problem to a set of separate single retailer inventory ordering problems, can be applied to this setting. Under some general assumptions on the prices and holding costs, we show that the optimality bound proven in Marklund and Rosling (2012) also holds for the heuristic in our setting, implying that the heuristic is asymptotically optimal in the number of retailers N.

Using the heuristic as a proxy for the optimal solution, we study a simple, two-period demand forecasting model. We prove our main result, which is that under a further approximation that any inventory remaining at the retailers at the end of the first period cannot be carried over to the next period, the first period allocations are strictly decreasing in the variance of the second period forecast. Since the demand forecasting model captures the reduction in demand uncertainty over time that comes with learning, this result suggests that demand learning incentivizes the decision maker to reserve more inventory at the warehouse for later periods. Demand learning should thus have a similar effect on allocation decisions as risk pooling.

Additionally, we investigate the heuristic allocations when the retailer demands

have different levels of uncertainty. We show that the heuristic allocation policy depends on the available inventory at the warehouse: When the warehouse starting inventory is small, the decision maker should act more cautiously, and should allocate more inventory to retailers with low variances. When the warehouse starting inventory is large, on the other hand, the decision maker should instead favor a "riskier" policy that allocates more inventory to retailers with high variances.

Much remains to be said on the subject of demand learning in the two echelon setting. Our analysis is restricted to a stylized forecasting model with known parameters, and does not study the case where the parameters of the demand distributions are unknown and have to be learned over time. Although the heuristic can be combined with statistical learning methods in order to learn such unknown parameters, it is not clear whether it is possible to prove a non-trivial upper bound on the regret of this algorithm. Since we have only been able to show that the optimality gap between the heuristic and the exact solution to the allocation problem is sublinear in the number of retailers, and not in the length of the selling horizon, proving a non-trivial regret upper bound in terms of the length of the selling horizon would be difficult. We leave such questions to future work.

Chapter 5

Conclusions and Future Directions

This thesis studies three related topics in revenue and supply chain management, all motivated by constraints unique to fashion retail. Chapters 2 (Dynamic learning and pricing with model misspecification) and 3 (Feature-based dynamic pricing for fashion retail: A case study) look at feature-based dynamic learning and pricing problems. Chapter 4 (Inventory allocation with demand learning for seasonal consumer goods) studies an inventory allocation problem in a two echelon, i.e. single warehouse multiple retailer, setting with fixed warehouse inventory. All three of these chapters propose algorithms that are data-driven in that they allow the decision maker to update her demand beliefs and pricing or allocation decisions based on historical observations of sales.

However, while all three chapters are interested in demand learning, they take different perspectives on learning. Our work on dynamic pricing in Chapters 2 and 3 is specifically interested in a setting with parametric demand models, and where the unknown parameters have to be dynamically re-estimated as new sales data is observed. The feature-based dynamic pricing algorithms that we proposed are designed to learn these unknown parameters at an optimal rate. In Chapter 2, we also prove regret bounds that capture this rate of learning. Chapter 4, on the other hand, takes a more qualitative perspective on demand learning. We are interested in how demand learning affects the structure of the optimal allocation policy. To answer this question, we describe learning using a demand forecasting model whose parameters

are known to the decision maker. The decision maker is thus learning the demand realizations, rather than the form of demand, with time.

There are a number of broad directions for future research. Even within the domain of fashion retail, new business models bring technical challenges that require novel solution techniques. One such model is the omni-channel extension to our work on inventory allocation, which was discussed in the concluding section of Chapter 4. This would involve extending our brick-and-mortar store setting to one where the retailer has an online presence to complement its brick-and-mortar stores.

Another business model that could be interesting from a revenue management perspective is the online consignment store. As consumers of fashion products become increasingly eco conscious, there has been a growing interest in options that are more sustainable than the fast fashion model studied in this thesis. Online consignment stores, such as San Francisco based TheRealReal and ThredUp, present one such option by creating a venue where used fashion items can be both bought and sold at markdowns from the original sales price. As pricing managers at these retailers have to decide not only the prices at which to sell used items, but also the prices at which to buy these items, feature-based pricing takes on an added dimension.

Appendix A

Appendix to Chapter 2 (Dynamic Learning and Pricing with Model Misspecification)

A.1 A Different Regret Definition

In the literature on dynamic pricing with demand learning, it is standard to define regret relative to the clairvoyant who knows the *true* demand model. Let us refer to the clairvoyant defined in Section 2.3.1 as the "linear clairvoyant," and define a second clairvoyant, called the "true clairvoyant," who sets price $\tilde{p}_t = -\frac{f(\mathbf{x_t})}{2b}$ at each time period. Then we can define a second notion of regret, $\text{Regret}_2(T)$, in terms of the true clairvoyant:

$$\text{Regret}_2(T) = \sum_{t=1}^{T} \mathrm{E}[\tilde{p}_t D(\tilde{p}_t)] - \sum_{t=1}^{T} \mathrm{E}[p_t D(p_t)].$$

To see how $\text{Regret}(T)$ compares to $\text{Regret}_2(T)$, we can write

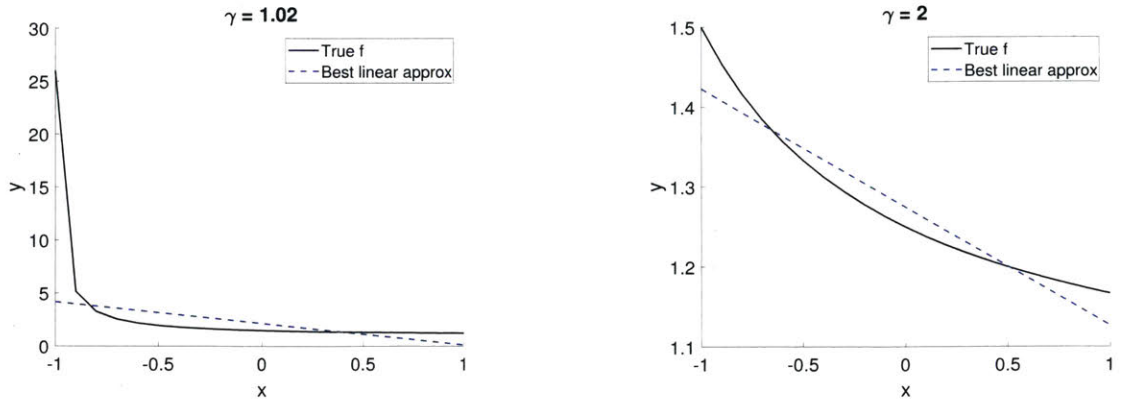
$$
\begin{aligned}
&\text{Regret}_2(T) \\
=&\text{Regret}(T) + \sum_{t=1}^{T} \text{E}[\tilde{p}_t D(\tilde{p}_t)] - \text{E}[p_t^* D(p_t^*)] \qquad \text{(A.1)} \\
=&\text{Regret}(T) \\
&+ \frac{T}{4|b|}\text{E}\left[\left(f(\mathbf{x_t}) - \text{E}\left[f(\mathbf{x_t})\begin{bmatrix}1 & \mathbf{x_t}^\top\end{bmatrix}\right]\right)\left(\text{E}\left[\begin{bmatrix}1 & \mathbf{x_t}^\top \\ \mathbf{x_t} & \mathbf{x_t}\mathbf{x_t}^\top\end{bmatrix}\right]\right)^{-1}\begin{bmatrix}1 \\ \mathbf{x_t}\end{bmatrix}\right)^2\right] \\
\geq&\frac{T}{4|b|}\text{E}\left[\left(f(\mathbf{x_t}) - \text{E}\left[f(\mathbf{x_t})\begin{bmatrix}1 & \mathbf{x_t}^\top\end{bmatrix}\right]\right)\left(\text{E}\left[\begin{bmatrix}1 & \mathbf{x_t}^\top \\ \mathbf{x_t} & \mathbf{x_t}\mathbf{x_t}^\top\end{bmatrix}\right]\right)^{-1}\begin{bmatrix}1 \\ \mathbf{x_t}\end{bmatrix}\right)^2\right]
\end{aligned}
$$

using closed form expressions for $\tilde{p}_t$ and p_t^*. This shows that the regret of any admissible pricing policy that assumes a misspecified demand model, relative to the true clairvoyant, grows linearly in T, and with the extent of model misspecification as captured by the expectation term in the second line. It reflects the fact that prices chosen by a seller who assumes a linear demand model may never converge to the optimal price $\tilde{p}_t$, because $\tilde{p}_t$ could depend nonlinearly on $\mathbf{x_t}$. We have also included additional numerical experiments using $\text{Regret}_2(T)$ as the benchmark, see Appendix A.2.3.

Throughout the rest of this paper, we mainly focus on $\text{Regret}(T)$ rather than $\text{Regret}_2(T)$. $\text{Regret}(T)$ is a more interesting performance metric as (A.1) shows that $\text{Regret}_2(T)$ of any admissible pricing policy affine in $\mathbf{x_t}$ is always $\theta(T)$, implying that it cannot be optimized in terms of T. The term "regret" thus refers to $\text{Regret}(T)$ in the rest of this paper unless stated otherwise.

A.2 Additional Numerical Results

In this section we expand on the numerical results in Section 2.4 by investigating how our results depend on the parameter settings. Section A.2.1 shows how the performance of the RPS algorithm depends on the choice of demand function. Section A.2.2 looks at its dependence on the dimension of the feature vector m, complementing



(a) $\gamma = 1.02$ (b) $\gamma = 2$

Figure A-1: $f(\mathbf{x})$ vs best linear approximation $a + \mathbf{c}'\mathbf{x}$ for $\gamma = 1.02, 2$

our theoretical results on the RPS algorithm's regret upper bound given in Section 2.3.

A.2.1 Dependence of regret on the demand function

We now investigate how the results of our simulations depend on the demand function. In the IID setting studied in Section 2.4, the quasilinear demand model is of the form

$$D_t(p) = \frac{1}{2(\mathbf{x_t} + \gamma)} + 1 - 0.9p + \epsilon_t,$$

where $\gamma = 1.03$, while the closest linear approximation is

$$\hat{D}_t(p) \approx 2.05 - 0.90p - 1.76\mathbf{x_t}$$

As γ increases, the fit of the closest linear approximation of D_t for $\mathbf{x_t}$ uniformly distributed between $[-1, 1]$ improves, i.e. $\mathrm{E}[(f(\mathbf{x_t}) - a - \mathbf{c}'\mathbf{x_t})^2]$ decreases. Fig. A-1 illustrates this by comparing the function f with its best linear approximation on the interval $[-1, 1]$ for two values of γ, $\gamma = 1.02$ and 2. Since model misspecification worsens as γ decreases, we would expect that the endogeneity effect is more significant for demand models with smaller values of γ. We ran the RPS and one-stage regression algorithms for $\gamma = 1.02, 1.03, 1.05,\ 1.1, 1.25, 1.5, 2.0$, keeping the price and parameter bounds the same as in the IID case numerical example with $\gamma = 1.03$. Table A.1,

which gives the estimates of the parameter b at the end of 5000 time periods averaged over 50 iterations, shows that for all γ, the RPS algorithm produces unbiased estimates of the parameter b. The one-stage regression algorithm estimates, on the other hand, are biased for smaller values of γ. As γ increases, the one-stage regression estimates of b improve. This is consistent with the observation that the endogeneity effect becomes more significant as γ decreases; the RPS algorithm, which corrects for endogeneity, produces unbiased parameter estimates for all γ, while the one-stage regression algorithm, which does not correct for endogeneity, only accurately estimates the parameters when the endogeneity effect becomes insignificant. Fig. A-2 plots the average cumulative regret (over 50 iterations) of the RPS and one-stage regression algorithms at the end of 5000 time periods for the different values of γ. The RPS algorithm outperforms the one-stage regression algorithm for $\gamma < 2.0$, and the improvement of RPS relative to one-stage generally increases as γ decreases and the endogeneity effect increases. However, for $\gamma = 2.0$, one-stage regression outperforms RPS algorithm; In the absence of endogeneity, parameters can be estimated more efficiently using a one-stage rather than a two-stage regression, and RPS loses its competitive edge.

Table A.1: Estimates of parameter b in Linear Demand Example

$\gamma =$	1.02	1.03	1.05	1.10	1.25	1.05	2.00s
RPS algo.	-0.94	-0.90	-0.91	-0.92	-0.90	-0.91	-0.90
One-stage reg.	-0.50	-0.50	-0.50	-0.53	-0.66	-0.77	-0.86

A.2.2 Dependence of regret on the feature vector dimension m

We conducted numerical experiments in an attempt to investigate the dependence of the results on m. For simplicity, we looked at a number of different settings without any model misspecification, with $T = 5000$ and m varying from 1 to 1001. Unfortunately, almost none of these settings yielded a clear regret trend, and showed the regret seesawing with increasing m. One possible explanation is that the asymptotic

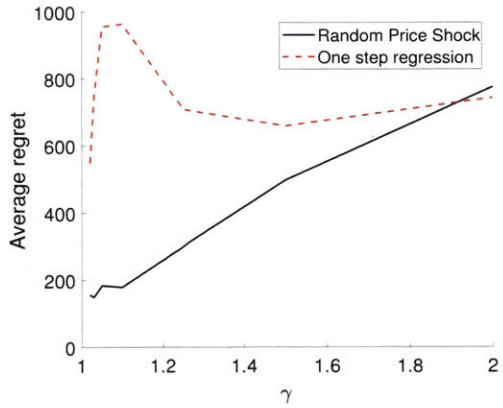


Figure A-2: Average regret over 50 iterations of RPS vs one-stage regression algorithms as γ is varied

dependence of the results on m only becomes detectable for larger values of m, which would be computationally infeasible to test.

However, for one of the settings tested, a clear regret trend was observed. Below, we report the results from this numerical experiment. The demand function is given by

$$D_t(p) = 2 - 0.7p + c^T \mathbf{x_t} + \epsilon_t.$$

For each m, the feature vectors $\mathbf{x_t}$ are drawn IID from the distribution $[-1, 1]^m$, and $\mathbf{c}$ is a vector of length m with the first entry set to 0.9 and all other entries set to 0. Note that $\|\mathbf{c}\|_1$ is constant for all m, and thus so is $\bar{c}$, on which our regret bound depends (see Eq (A.8) for the full statement of the IID regret bound in terms of all parameters). We set $c_{\max}$ to $\mathbf{c} + [0.5, 0.5, \ldots, 0.5]$ and $c_{\min}$ to $\mathbf{c} - [0.5, 0.5, \ldots, 0.5]$, and let the noise ϵ_t be normally distributed with mean 0 and variance 0.3. The price across all periods t is lower bounded by \$1.75 and upper bounded by \$8.25.

Fig. A-3 plots the regrets of the RPS algorithm for m = 1, 3, 5, 11, 51, 101, 201, 501, 1001, averaged over 10 iterations each. We can see that the regret of RPS is increasing with m, and that the growth of the regret with m appears to be $O((m+1)T)$, in accordance with our regret upper bound. This numerical example thus supports the idea that the regret of the RPS algorithm does indeed depend on m, and that there is a gap in terms of m between our lower and upper bounds.

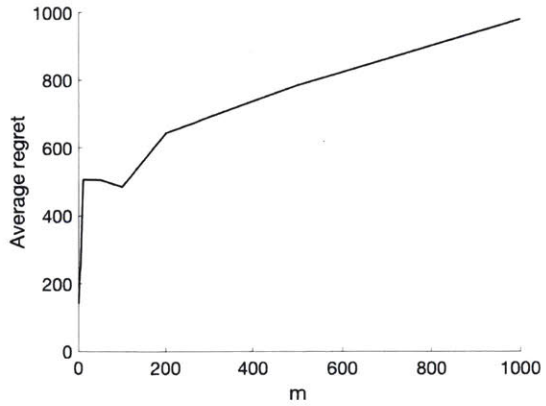
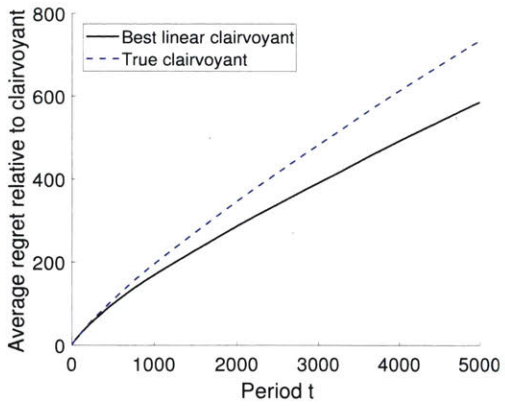


Figure A-3: Average regret over 10 iterations of the RPS algorithm as m is increased from 1 to 1001.

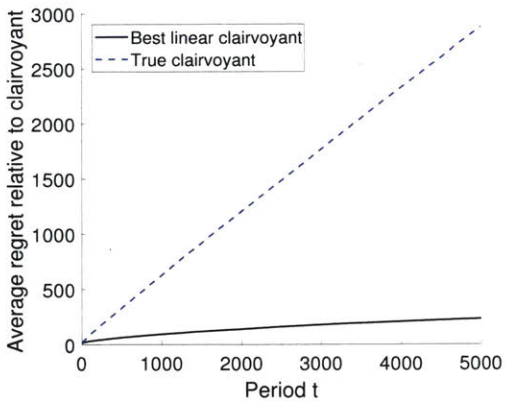
A.2.3 Regret relative to different clairvoyants

The above numerical experiments benchmark the performance of the RPS algorithm against the linear clairvoyant, who bases pricing decisions on the closest linear approximation of the true quasilinear demand model. Here we present additional numerical experiments benchmarking the performance of RPS against the true clairvoyant, who has full knowledge of the true quasilinear demand model, and sets price $\tilde{p}_t = -\frac{f(\mathbf{x})}{2b}$ at each time period. Fig. A-4a plots the results of repeating the IID setting experiments from Section 2.4; it plots the average regret of the RPS algorithm relative to both clairvoyants over 200 iterations and 5000 time periods. Similarly, Fig. A-4b plots the results of repeating the price ladder setting experiments from Section 2.4, and Fig. A-4c plots the results of repeating the non IID experiments from Section 2.4.

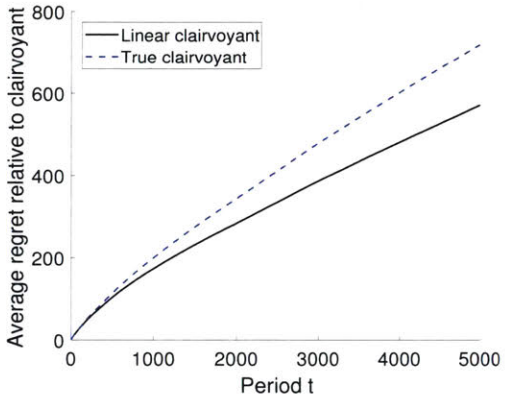
Fig. A-4a confirms the result that the regret of RPS relative to the true clairvoyant grows linearly with T in the IID setting. On the other hand, Fig. A-4b shows that, depending on the function f and the distribution of the feature vectors, the regret of RPS relative to the true clairvoyant need not grow linearly with T in the non IID setting. We can also observe from Figures A-4a - A-4c that the difference in revenue earned by the true clairvoyant and the revenue earned by the linear clairvoyant can vary considerably depending on the demand model and parameters; In the IID and price ladder settings, the extent of model misspecification is extremely large, while in



(a) IID setting – average regret



(b) Price ladder IID setting – average regret



(c) Non IID setting – average regret

Figure A-4: Average regret over 200 iterations of RPS algorithm relative to two different clairvoyants in IID and Price ladder IID settings

the non IID setting, the linear clairvoyant achieves nearly as much revenue as the true clairvoyant. One way the retailer could try to improve the fit of her demand model in the first two cases is by including higher order terms of $\mathbf{x_t}$ in the feature vector and performing polynomial regression; however we note that she faces a tradeoff in doing so: The regret bound stated in Theorem 1, shows that the regret of RPS is $O((m+1)\sqrt{T})$, i.e. including more terms of $\mathbf{x_t}$ in the feature vector could decrease the regret from model misspecification, but increase the regret due to parameter estimation errors.

A.3 Proofs for Theoretical Analysis

Notation. The following notations will be used in this section. We define $e := (a, \mathbf{c}^\mathsf{T})^\mathsf{T}$ and $\mathbf{e_t} := (\hat{a}_t, \hat{\mathbf{c}}_t^\mathsf{T})^\mathsf{T}$. Let $\tilde{\mathbf{x}} := (1, \mathbf{x}^\mathsf{T})^\mathsf{T}$, $M := \mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\mathsf{T}]$ and $M_t := \frac{1}{t-1}\sum_{j=1}^{t-1} \tilde{\mathbf{x}}_j \tilde{\mathbf{x}}_j^\mathsf{T}$.

A.3.1 Proof of Proposition 1

Proof. Proof of Proposition 1. Consider price $p_t' = -\frac{\alpha+\gamma^\mathsf{T}\mathbf{x_t}}{2\beta}$, where α, β, γ are measurable with respect to history $\mathcal{H}_{t-1}$. Since $p_t^* = -\frac{a+\mathbf{c}^\mathsf{T}\mathbf{x_t}}{2b}$, we have

$$\begin{aligned}
\mathrm{E}\left[p_t^* D(p_t^*) - p_t' D(p_t') \mid \mathcal{H}_{t-1}\right] =& \mathrm{E}\left[p_t^*(bp_t^* + f(x_t)) - p_t'(bp_t' + f(x_t)) \mid \mathcal{H}_{t-1}\right] \\
=& \mathrm{E}\left[p_t^*(bp_t^* + a + c^\mathsf{T}x_t) - p_t'(bp_t' + a + c^\mathsf{T}x_t) \mid \mathcal{H}_{t-1}\right] \\
& - \mathrm{E}\left[(p_t^* - p_t')(a + c^\mathsf{T}x_t - f(x_t)) \mid \mathcal{H}_{t-1}\right] \\
=& \mathrm{E}\left[p_t^*(bp_t^* - 2bp_t^*) - p_t'(bp_t' - 2bp_t^*) \mid \mathcal{H}_{t-1}\right] \\
& - \mathrm{E}\left[(p_t^* - p_t')(a + c^\mathsf{T}x_t - f(x_t)) \mid \mathcal{H}_{t-1}\right] \\
=& -b\mathrm{E}\left[(p_t^* - p_t')^2 \mid \mathcal{H}_{t-1}\right] \\
& - \mathrm{E}\left[(p_t^* - p_t')(a + c^\mathsf{T}x_t - f(x_t)) \mid \mathcal{H}_{t-1}\right].
\end{aligned}$$

To finish the proof, we shall prove that $\mathrm{E}\left[(p_t^* - p_t')(a + \mathbf{c}^\mathsf{T}\mathbf{x_t} - f(\mathbf{x_t})) \mid \mathcal{H}_{t-1}\right] = 0$. By definition, a, c is the optimal solution of the following least squares problem

$$\min_{a', \mathbf{c}'} \mathrm{E}\left[(f(\mathbf{x_t}) - (a' + \mathbf{c}'^\mathsf{T}\mathbf{x_t}))^2\right].$$

By first order conditions, we have

$$\mathrm{E}\left[a + \mathbf{c}^\mathsf{T}\mathbf{x_t} - f(\mathbf{x_t})\right] = 0, \quad \mathrm{E}\left[\mathbf{x_t}\left(a + \mathbf{c}^\mathsf{T}\mathbf{x_t} - f(\mathbf{x_t})\right)\right] = 0.$$

Since $\mathbf{x_t}$ is independent of the history $\mathcal{H}_{t-1}$, we have

$$\mathrm{E}\left[a + \mathbf{c}^\mathsf{T}\mathbf{x_t} - f(\mathbf{x_t}) \mid \mathcal{H}_{t-1}\right] = 0, \quad \mathrm{E}\left[\mathbf{x_t}\left(a + \mathbf{c}^\mathsf{T}\mathbf{x_t} - f(\mathbf{x_t})\right) \mid \mathcal{H}_{t-1}\right] = 0.$$

Therefore,

$$
\begin{aligned}
&\mathrm{E}\left[(p_t^* - p_t')(a + \mathbf{c}^\mathsf{T}\mathbf{x_t} - f(\mathbf{x_t})) \mid \mathcal{H}_{t-1}\right] \\
=&\mathrm{E}\left[(-\frac{a + \mathbf{c}^\mathsf{T}\mathbf{x_t}}{2b} + \frac{\alpha + \gamma^\mathsf{T}\mathbf{x_t}}{2\beta})(a + \mathbf{c}^\mathsf{T}\mathbf{x_t} - f(\mathbf{x_t})) \mid \mathcal{H}_{t-1}\right] \\
=&\mathrm{E}\left[(-\frac{a}{2b} + \frac{\alpha}{2\beta})\mathrm{E}\left[(a + \mathbf{c}^\mathsf{T}\mathbf{x_t} - f(\mathbf{x_t})) \mid \mathcal{H}_{t-1}\right]\right] \\
&+ \mathrm{E}\left[(-\frac{\mathbf{c}^\mathsf{T}}{2b} + \frac{\gamma^\mathsf{T}}{2\beta})\mathrm{E}\left[\mathbf{x_t}(a + \mathbf{c}^\mathsf{T}\mathbf{x_t} - f(\mathbf{x_t})) \mid \mathcal{H}_{t-1}\right]\right] \\
=&0.
\end{aligned}
$$

which implies that $\mathrm{E}\left[(p_t^* - p_t')(a + \mathbf{c}^\mathsf{T}\mathbf{x_t} - f(\mathbf{x_t})) \mid \mathcal{H}_{t-1}\right] = 0$. Then, applying the law of total expectation, we prove the theorem.

□

A.3.2 Proof of Theorem 1

Proof. Proof. Recall that the expected regret over the selling horizon is defined as

$$\text{Expected Regret}(T) = \sum_{t=1}^{T} \mathrm{E}[p_t^* D(p_t^*)] - \sum_{t=1}^{T} \mathrm{E}[p_t D(p_t)]. \tag{A.2}$$

First, let Q be a positive definite matrix such that $M = Q^2$ (Q must exist since M is positive definite). Then, let us define the event A_t as follows:

$$A_t = \left\{M_t \text{ is invertible and } \|QM_t^{-1}Q\|_2 \leq 2\right\}.$$

We can write the regret as

$$
\begin{aligned}
&\sum_{t=1}^{T} \mathrm{E}[p_t^* D_t(p_t^*) - p_t D_t(p_t)] \\
=&\sum_{t=1}^{T} \mathrm{E}[p_t^* D_t(p_t^*) - p_t D_t(p_t) | A_t] \cdot \mathbb{P}[A_t] + \mathrm{E}[p_t^* D_t(p_t^*) - p_t D_t(p_t) | A_t^C] \cdot \mathbb{P}[A_t^C] \\
\leq&\sum_{t=1}^{T} \mathrm{E}[p_t^* D_t(p_t^*) - p_t D_t(p_t) | A_t] \cdot \mathbb{P}[A_t] + \frac{(\bar{a}+\bar{c})^2}{2\underline{b}} \mathbb{P}[A_t^C] \\
\leq&\sum_{t=1}^{T} \mathrm{E}[p_t^* D_t(p_t^*) - p_t D_t(p_t) | A_t] \cdot \mathbb{P}[A_t] \\
&+ \frac{(\bar{a}+\bar{c})^2}{\underline{b}} 2(m+1) \exp\left(- \frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)} \right),
\end{aligned}
$$

where the second inequality follows from the definition of p_t^* and our assumptions on the boundedness of the true parameters $a, b, \mathbf{c}$, and the final inequality follows by bounding $\mathbb{P}(A_t^C)$ by Lemma 3, where $V := \mathrm{E}[(Q^{-1}\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\intercal Q^{-1} - I)^2]$. Since the second addend in the final line is $O(e^{-t})$, it is left to show that the first addend is $O(\sqrt{1/t})$.

We decompose it as follows:

$$
\begin{aligned}
\sum_{t=1}^{T} \mathrm{E}[p_t^* D_t(p_t^*) - p_t D_t(p_t) | A_t] \cdot \mathbb{P}[A_t] =& \sum_{t=1}^{T} \mathrm{E}[p_t^* D_t(p_t^*) - p_{g,t}^u D_t(p_{g,t}^u) | A_t] \cdot \mathbb{P}[A_t] \\
&+ \sum_{t=1}^{T} \mathrm{E}[p_{g,t}^u D_t(p_{g,t}^u) - p_{g,t} D_t(p_{g,t}) | A_t] \cdot \mathbb{P}[A_t] \\
&+ \sum_{t=1}^{T} \mathrm{E}[p_{g,t} D_t(p_{g,t}) - p_t D_t(p_t) | A_t] \cdot \mathbb{P}[A_t].
\end{aligned}
$$

Since $p_{g,t} = \mathrm{Proj}(p_{g,t}^u, [\underline{p}_t + \delta_t, \bar{p}_t - \delta_t])$ and the optimal price $\tilde{p}_t \in [\underline{p}_t, \bar{p}_t]$, we have

$$
\sum_{t=1}^{T} \mathrm{E}[p_{g,t}^u D_t(p_{g,t}^u) - p_{g,t} D_t(p_{g,t}) | A_t] \cdot \mathbb{P}[A_t] \leq \sum_{t=1}^{T} \bar{b}\delta_t^2 \cdot \mathbb{P}[A_t] \leq \sum_{t=1}^{T} \frac{\delta^2 \bar{b}}{4} \frac{1}{\sqrt{t}} \leq \frac{\bar{b}\delta^2\sqrt{T}}{2}.
$$

In addition, $p_t = p_{g,t} + \Delta p_t$, where Δp_t is generated independently from $p_{g,t}, \mathbf{x_t}$ and

the history $\mathcal{H}_{t-1}$ with variance $\delta_t^2 = \frac{\delta^2}{4\sqrt{t}}$. So

$$\begin{aligned}
&\sum_{t=1}^{T} \mathrm{E}[p_{g,t} D_t(p_{g,t}) - p_t D_t(p_t) | A_t] \cdot \mathbb{P}[A_t] \\
=&\sum_{t=1}^{T} \mathrm{E}[p_{g,t} D_t(p_{g,t}) - (p_{g,t} + \Delta p_t) D_t(p_{g,t} + \Delta p_t) | A_t] \cdot \mathbb{P}[A_t] \\
=&\sum_{t=1}^{T} \mathrm{E}[\Delta p_t(-2b p_{g,t} - f(\mathbf{x_t})) - b(\Delta p_t)^2 | A_t] \cdot \mathbb{P}[A_t] \\
=&\sum_{t=1}^{T} \mathrm{E}[-b(\Delta p_t)^2 | A_t] \cdot \mathbb{P}[A_t] \\
\leq&\sum_{t=1}^{T} -\frac{b\delta^2}{4}\frac{1}{\sqrt{t}} \leq \frac{\bar{b}\delta^2\sqrt{T}}{2}.
\end{aligned}$$

To finish the proof, we want to show that $\mathrm{E}[p_t^* D_t(p_t^*) - p_{g,t}^u D_t(p_{g,t}^u) | A_t] \cdot \mathbb{P}[A_t] = O(1/\sqrt{t})$. In the proof of Proposition 1, we show that

$$\mathrm{E}[p_t^* D_t(p_t^*) - p_t' D_t(p_t') | \mathcal{H}_{t-1}] = -b\mathrm{E}[(p_t^* - p_t')^2 | \mathcal{H}_{t-1}].$$

for any $p_t' = -\frac{\alpha + \gamma^\mathsf{T} \mathbf{x_t}}{2\beta}$ with α, β, γ measurable with respect to the history $\mathcal{H}_{t-1}$.

Since the event A_t depends on the history $\mathcal{H}_{t-1}$ and is independent of $\mathbf{x_t}$, this gives

$$\mathrm{E}[p_t^* D_t(p_t^*) - p_{g,t}^u D_t(p_{g,t}^u) | A_t] \cdot \mathbb{P}[A_t] = -b\mathrm{E}[(p_t^* - p_{g,t}^u)^2 | A_t] \cdot \mathbb{P}[A_t],$$

where $p_{g,t}^u = -\frac{\hat{a}_t + \hat{\mathbf{c}}_t^\mathsf{T} \mathbf{x_t}}{2\hat{b}_t}$ is the greedy price given the estimates $\hat{a}_t, \hat{b}_t, \hat{\mathbf{c}}_t$, and $p_t^* = -\frac{a + \mathbf{c}^\mathsf{T} \mathbf{x_t}}{2b}$ is the optimal price of the following linear model

$$D_t(p) = a + bp + \mathbf{c}^\mathsf{T} \mathbf{x_t} + \nu_t, \quad \forall p \in [\underline{p}_t, \bar{p}_t],$$

with $\nu_t = f(\mathbf{x_t}) - a - \mathbf{c}^\mathsf{T} \mathbf{x_t} + \epsilon_t$.

By the definition of $p^u_{t,g}$ and p^*_t, we have

$$\begin{aligned}
\mathrm{E}[(p^u_{t,g} - p^*_t)^2 | A_t] \cdot \mathbb{P}[A_t] =& \mathrm{E}\left[\left(\frac{a + \mathbf{c}^\mathsf{T}\mathbf{x_t}}{2b} - \frac{\hat{a}_t + \hat{\mathbf{c}}_t^\mathsf{T}\mathbf{x_t}}{2\hat{b}_t}\right)^2 | A_t\right] \cdot \mathbb{P}[A_t] \\
\leq& 2\mathrm{E}\left[\left(\frac{a + \mathbf{c}^\mathsf{T}\mathbf{x_t}}{2b} - \frac{a + \mathbf{c}^\mathsf{T}\mathbf{x_t}}{2\hat{b}_t}\right)^2 | A_t\right] \cdot \mathbb{P}[A_t] \\
&+ 2\mathrm{E}\left[\left(\frac{a + \mathbf{c}^\mathsf{T}\mathbf{x_t}}{2\hat{b}_t} - \frac{\hat{a}_t + \hat{\mathbf{c}}_t^\mathsf{T}\mathbf{x_t}}{2\hat{b}_t}\right)^2 | A_t\right] \cdot \mathbb{P}[A_t] \\
\leq& (\bar{a} + \bar{c})^2 \mathrm{E}\left[\left(\frac{1}{b} - \frac{1}{\hat{b}_t}\right)^2 | A_t\right] \cdot \mathbb{P}[A_t] \\
&+ \frac{1}{\underline{b}^2}\mathrm{E}\left[\left((a + \mathbf{c}^\mathsf{T}\mathbf{x_t}) - (\hat{a}_t + \hat{\mathbf{c}}_t^\mathsf{T}\mathbf{x_t})\right)^2 | A_t\right] \cdot \mathbb{P}[A_t],
\end{aligned}$$

where the second line follows from the inequality $(x+y)^2 \leq 2x^2 + 2y^2$, and the third line follows from the fact that the true parameters $a, \mathbf{c}$ satisfy $\|a\| \leq \bar{a}$ and $\|\mathbf{c}\|_1 \leq \bar{c}$, as well as from the fact that $\hat{b}_t \in [-\bar{b}, -\underline{b}]$.

Now, for demand parameter b', let h be the function $h(b') = \frac{1}{b'}$. The gradient of h, denoted by ∇h, is given by $\nabla h(b') = -\frac{1}{b'^2}$, and we have $|\nabla h(b')|^2 = \frac{1}{b'^4} \leq \frac{1}{\underline{b}^4}$. Then by the Mean Value Theorem, we have

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{1}{b} - \frac{1}{\hat{b}_t}\right)^2 | A_t\right] \cdot \mathbb{P}[A_t] &\leq \frac{1}{b^4}\mathrm{E}[(b - \hat{b}_t)^2 | A_t] \cdot \mathbb{P}[A_t] \\
&\leq \frac{1}{b^4}\mathrm{E}[(b - \hat{b}_t)^2]. \qquad \text{(A.3)}
\end{aligned}$$

By Lemma 2, we immediately have $\mathrm{E}[(\hat{b}_t - b)^2] = \mathrm{O}(1/\sqrt{t})$. Now we will bound the error in the estimates of a and $\mathbf{c}$, namely $\mathrm{E}[\left((\mathbf{e} - \mathbf{e_t})^\mathsf{T}\tilde{\mathbf{x}}_\mathbf{t}\right)^2 | A_t]$. Note that $\mathbf{e_t}$ is measurable with history $\mathcal{H}_{t-1}$ and $\tilde{\mathbf{x}}_\mathbf{t}$ is independent of $\mathcal{H}_{t-1}$, so

$$\begin{aligned}
\mathrm{E}[\left((\mathbf{e} - \mathbf{e_t})^\mathsf{T}\tilde{\mathbf{x}}_\mathbf{t}\right)^2 | A_t] =& \mathrm{E}\left[(\mathbf{e} - \mathbf{e_t})^\mathsf{T}\mathrm{E}[\tilde{\mathbf{x}}_\mathbf{t}\tilde{\mathbf{x}}_t^\mathsf{T} | A_t, \mathcal{H}_{t-1}](\mathbf{e} - \mathbf{e_t})\right] \\
=& \mathrm{E}\left[(\mathbf{e} - \mathbf{e_t})^\mathsf{T} M (\mathbf{e} - \mathbf{e_t}) | A_t\right] = \mathrm{E}\left[\|\mathbf{e} - \mathbf{e_t}\|_M^2 | A_t\right],
\end{aligned}$$

where $\|\mathbf{y}\|_A := \sqrt{\mathbf{y}^\mathsf{T} A \mathbf{y}}$ for any positive definite matrix A.

By the definition of Algorithm 1, assuming that M_t is invertible, $\mathbf{e_t} - \mathbf{e}$ can be

written as

$$\mathbf{e_t} - \mathbf{e} = \text{Proj}\left(M_t^{-1} \frac{\sum_{s=1}^{t-1} \tilde{\mathbf{x}}_s (p_s(b - \hat{b}_t) + \epsilon_s)}{t-1} \right). \tag{A.4}$$

Then we have

$$\begin{aligned}
\mathrm{E}[\left((\mathbf{e} - \mathbf{e_t})^\mathsf{T} \tilde{\mathbf{x}}_\mathbf{t}\right)^2 |] &= \mathrm{E}[\|\mathbf{e_t} - \mathbf{e}\|_M^2 \mid A_t] \cdot \mathbb{P}[A_t] \\
&\leq \mathrm{E}[\|\mathbf{e_t} - \mathbf{e}\|_M^2 \mid A_t] \cdot \mathbb{P}[A_t] + \mathrm{E}[4(\mathbf{e}^\mathsf{T} \tilde{\mathbf{x}}_\mathbf{t})^2 + 4(\mathbf{e_t}^\mathsf{T} \tilde{\mathbf{x}}_\mathbf{t})^2] \cdot \mathbb{P}[A_t^C] \\
&\leq \mathrm{E}[\|\mathbf{e_t} - \mathbf{e}\|_M^2 \mid A_t] \cdot \mathbb{P}[A_t] + 16\bar{b}^2 p_{\max}^2 \mathbb{P}[A_t^C] \\
&\leq \mathrm{E}[\|\mathbf{e_t} - \mathbf{e}\|_M^2 \mid A_t] \cdot \mathbb{P}[A_t] \qquad \text{(A.5)} \\
&\quad + 16\bar{b}^2 p_{\max}^2 \cdot 2(m+1) \exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)} \right). \qquad \text{(A.6)}
\end{aligned}$$

The third line follows from the assumption that the true parameter $\mathbf{e} \in E$. In the last step, we bound $\mathbb{P}(A_t^C)$ by Lemma 3, where $V := \mathrm{E}[(Q^{-1}\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\mathsf{T}Q^{-1} - I)^2]$. Since Eq (A.6) is $O(e^{-t})$, it is left to show that Eq (A.5) is $O(\sqrt{1/t})$.

We write Eq (A.5) as

$$\begin{aligned}
&\mathrm{E}[\|\mathbf{e_t} - \mathbf{e}\|_M^2 \mid A_t] \cdot \mathbb{P}[A_t] \\
&\leq \mathrm{E}\left[\|QM_t^{-1}QQ^{-1} \frac{\sum_{s=1}^{t-1} \tilde{\mathbf{x}}_s (p_s(b - \hat{b}_t) + \nu_s)}{t-1} \|^2 \mid A_t \right] \mathbb{P}[A_t] \\
&\leq \mathrm{E}\left[\|QM_t^{-1}Q\|_2^2 \cdot \|Q^{-1}\|_2^2 \cdot \| \frac{\sum_{s=1}^{t-1} \tilde{\mathbf{x}}_s (p_s(b - \hat{b}_t) + \nu_s)}{t-1} \|^2 \mid A_t \right] \mathbb{P}[A_t] \\
&\leq \mathrm{E}\left[4 \cdot \frac{1}{\lambda_{\min}(M)} \cdot 2 \left(\| \frac{\sum_{s=1}^{t-1} \tilde{\mathbf{x}}_s p_s (b - \hat{b}_t)}{t-1} \|^2 + \| \frac{\sum_{s=1}^{t-1} \tilde{\mathbf{x}}_s \nu_s}{t-1} \|^2 \right) \mid A_t \right] \mathbb{P}[A_t] \\
&\leq \mathrm{E}\left[\frac{8}{\lambda_{\min}(M)} \left(\| \frac{\sum_{s=1}^{t-1} \tilde{\mathbf{x}}_s p_s (b - \hat{b}_t)}{t-1} \|^2 + \| \frac{\sum_{s=1}^{t-1} \tilde{\mathbf{x}}_s \nu_s}{t-1} \|^2 \right) \right] \\
&\leq \mathrm{E}\left[\frac{8}{\lambda_{\min}(M)} \left((m+1) p_{\max}^2 (b - \hat{b}_t)^2 + \| \frac{\sum_{s=1}^{t-1} \tilde{\mathbf{x}}_s \nu_s}{t-1} \|^2 \right) \right] \\
&= \frac{8}{\lambda_{\min}(M)} \left((m+1) p_{\max}^2 \mathrm{E}[(b - \hat{b}_t)^2] + \frac{1}{(t-1)^2} \mathrm{E}[\| \sum_{s=1}^{t-1} \tilde{\mathbf{x}}_s \nu_s \|^2] \right). \qquad \text{(A.7)}
\end{aligned}$$

The first inequality holds by Eq (A.4) and the assumption that the true parameter $\mathbf{e} \in E$. The second inequality holds from the submultiplicative property of the spectral norm. By the definition of Q, we have $\|Q^{-1}\|_2 = 1/\sqrt{\lambda_{\min}(M)}$. The third inequality uses the definition of event A_t and the fact $\|\mathbf{x}+\mathbf{y}\|^2 \leq 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2$. The fourth inequality simply uses the definition of conditional expectation. The fifth inequality uses the assumptions that $\|\mathbf{x_t}\|_\infty \leq 1$ and $p_j \leq p_{\max}$.

It has already been established using Lemma 2 that $\mathrm{E}[(b-\hat{b}_t)^2]$ is $O(1/\sqrt{t})$, so the first term of Eq (A.7) is $O(1/\sqrt{t})$. For the second term, note that $(\tilde{\mathbf{x}}_s, \nu_s)$ is independent of $(\tilde{\mathbf{x}}_{s'}, \nu_{s'})$ for $s \neq s'$. Furthermore, by the first order condition of the least squares estimator, we have $\mathrm{E}[\nu_t] = 0$ and $\mathrm{E}[\mathbf{x_t}\nu_t] = 0$. So for each s, $\mathrm{E}[\tilde{\mathbf{x}}_s\nu_s \mid \mathcal{H}_{s-1}] = \mathrm{E}[\tilde{\mathbf{x}}_s\nu_s] = 0$. Thus,

$$\begin{aligned}
\frac{1}{(t-1)^2}\mathrm{E}[\|\sum_{s=1}^{t-1}\tilde{\mathbf{x}}_s\nu_s\|^2] =& \frac{1}{(t-1)^2}\sum_{s=1}^{t-1}\mathrm{E}[\|\tilde{\mathbf{x}}_s\nu_s\|^2] \\
=& \frac{1}{(t-1)^2}\sum_{s=1}^{t-1}\mathrm{E}[\|\tilde{\mathbf{x}}_s(f(\mathbf{x_t}) - a - \mathbf{c}^\intercal\mathbf{x_t} + \epsilon_t)\|^2] \\
\leq& \frac{(m+1)}{t-1}3(\bar{f}^2 + 4\bar{b}^2p_{\max}^2 + \sigma^2),
\end{aligned}$$

where the last step uses the fact that $(x+y+z)^2 \leq 3(x^2+y^2+z^2)$ and $\|\tilde{\mathbf{x}}_s\|^2 \leq m+1$. Therefore, by Eq (A.7), $\mathrm{E}[\|\mathbf{e}-\mathbf{e_t}\|_M^2] \leq O(1/\sqrt{t}) + O((m+1)/t) = O(1/\sqrt{t})$ as desired.

Dependence on $m, \underline{b}, \bar{b}$ and other parameters By combining constant factors, the expected regret of RPS algorithm over N periods can be bounded by

$$O\left(\frac{\bar{b}^2(p_{\max}^2+1)}{\underline{b}^4}\frac{(\bar{f}^2+\sigma^2+\bar{b}^2p_{\max}^2)}{\delta^2}(1+p_{\max}^2\frac{m+1}{\lambda_{\min}(M)})\sqrt{T}\right) + O((m+1)\log T), \tag{A.8}$$

where the pre-factor in the first big O notation only contains an absolute constant. □

A.3.3 Proof of Theorem 2

Proof. Proof. We will prove that the lower bound of regret is $\Omega(\sqrt{T})$ even if *the model is correctly specified.* Suppose there is no model misspecification, i.e. the

demand function is given by

$$D_t(p) = a + bp + \mathbf{c}^\mathsf{T}\mathbf{x_t} + \epsilon_t.$$

We assume feature vector $\mathbf{x_t}$ is i.i.d. and sampled uniformly from $[-1/2, 1/2]^m$, and demand noise ϵ_t is i.i.d. normal with variance 1. By the first order condition, the optimal price that any non-anticipating pricing policy can charge at period t is $p_t^* = (a + \mathbf{c}^\mathsf{T}\mathbf{x_t})/(-2b)$.

By Lemma 4, we can assume without loss of generality that the seller uses a linear pricing strategy π at period t given by $p_t = S_t + (U_t)^\mathsf{T}\mathbf{x_t}$, where S_t and U_t are measurable with respect to the history $\mathcal{H}_{t-1} = \sigma(x_1, \epsilon_1, \ldots, x_{t-1}, \epsilon_{t-1})$. Denote the regret incurred by the seller at the end of T periods as Regret(T). By Proposition 1, we have

$$\begin{aligned}
\text{Regret}(T) &= -b\mathrm{E}[(p_t - p_t^*)^2] \\
&= -b\mathrm{E}[(S_t + (U_t)^\mathsf{T}\mathbf{x_t} - S^* - (U^*)^\mathsf{T}\mathbf{x_t})^2] \\
&= -b\left\{\mathrm{E}[(S_t - S^*)^2] + \sum_{k=1}^{m}\mathrm{E}[(U_{t,k}x_{t,k} - U_k^*x_{t,k})^2]\right\} \\
&= -b\left\{\mathrm{E}[(S_t - S^*)^2] + \frac{1}{12}\sum_{k=1}^{m}\mathrm{E}[(U_{t,k} - U_k^*)^2]\right\}, \qquad \text{(A.9)}
\end{aligned}$$

where $S^* = -a/(2b)$, $U^* = -\mathbf{c}/(2b)$, the third line follows since $\mathrm{E}[\mathbf{x_t}] = 0$ and $\mathbf{x_t} = (x_{t,1}, \ldots, x_{t,m})$ has independent entries for our particular choice of $\mathbf{x_t}$, and the last line is because each entry of $\mathbf{x_t}$ has variance $\frac{1}{12}$.

Now we use the Van Trees inequality (Gill and Levit, 1995), a Bayesian version of the Crámer-Rao inequality, to lower bound the regret of any admissible policy. The proof below is a generalization of the proof of Theorem 1 in Keskin and Zeevi (2014). Suppose the parameters $\theta = (a, b, \mathbf{c})$ belong to compact sets $\theta = A \times B \times C$, where $A = [-\bar{a}, \bar{a}]$, $B = [-\bar{b}, -\underline{b}]$, $C = \{\mathbf{c}' \in \mathbb{R}^m : \sum_{k=1}^m |\mathbf{c}'_k| \leq \bar{c}\}$. We can construct a prior distribution on θ with density function λ which is positive on the interior and 0 on

the boundary of θ. We finish the proof by showing for any pricing policy that

$$\mathrm{E}_\lambda\left[\mathrm{Regret}_\theta(T)\right] = \Omega(\sqrt{T}),$$

where $\mathrm{Regret}_\theta(T)$ is the regret associated with a particular (unknown) parameter θ, and $\mathrm{E}_\lambda[\cdot]$ is the expectation operator on parameter θ under distribution λ. The above result immediately implies that there exists some parameter θ with regret $\Omega(\sqrt{T})$ for any pricing policy, namely

$$\max_{\theta\in\theta}\left\{\mathrm{Regret}_\theta(T)\right\} \geq \mathrm{E}_\lambda\left[\mathrm{Regret}_\theta(T)\right] = \Omega(\sqrt{T}).$$

Let $f_t(H_t \mid \theta)$ be the joint probability density function of history

$$H_t = (x_1, p_1, D_1, \cdots, x_t, p_t, D_t)$$

under parameter θ and a particular pricing policy $p_s = \pi(H_{s-1}, \mathbf{x_s})$. By our assumption that $\mathbf{x_t}$ is uniform and ϵ_t is normal, we have

$$f_t(H_t \mid \theta) = \Pi_{j=1}^{t}\phi(D_j - a - bp_j - \mathbf{c}^\mathsf{T}x_j),$$

where ϕ is the density function of the standard normal distribution. The Fisher information matrix of θ given history H_t is

$$\mathcal{I}_t(\theta) = \mathrm{E}_\theta\left[\nabla_\theta \log f_t(H_t \mid \theta)\cdot\left(\nabla_\theta \log f_t(H_t \mid \theta)\right)^\mathsf{T}\right] = \mathrm{E}_\theta\left[\sum_{j=1}^{t}\begin{bmatrix} 1 & x_j^\mathsf{T} & p_j \\ x_j & x_jx_j^\mathsf{T} & p_jx_j^\mathsf{T} \\ p_j & p_jx_j & p_j^2 \end{bmatrix}\right]. \tag{A.10}$$

Define function $g(\theta) = [a/(2b), 1, \mathbf{c}/(2b)]^\mathsf{T}$ and function $S(\theta) = -a/(2b) = S^*$. Applying the multivariate Van Trees inequality to S_t, which is an estimate of $S(\theta)$ based on history H_{t-1}, gives

$$\mathrm{E}_\lambda[\mathrm{E}_\theta[(S_t - S(\theta))^2]] \geq \frac{\mathrm{E}_\lambda[g(\theta)^\mathsf{T}\nabla S(\theta)]^2}{\mathrm{E}_\lambda[g(\theta)^\mathsf{T}\mathcal{I}_{t-1}(\theta)g(\theta)] + \tilde{I}(\lambda)}, \tag{A.11}$$

where $\tilde{\mathcal{I}}(\lambda)$ is the Fisher information of θ given prior λ. We have

$$g(\theta)^{\mathsf{T}} \cdot (\nabla S(\theta)) = \left[\frac{a}{2b}, 1, \frac{c}{2b}\right]^{\mathsf{T}} \cdot \left[-\frac{1}{2b}, \frac{a}{2b^2}, 0\right] = \frac{a}{4b^2}.$$

By Eq (A.10) and $p_j^* = -(a + \mathbf{c}^{\mathsf{T}} x_j)/(2b)$, one can show that

$$g(\theta)^{\mathsf{T}} \mathcal{I}_{t-1}(\theta) g(\theta) = \mathrm{E}_\theta \left[\sum_{j=1}^{t-1} (p_j - p_j^*)^2\right] \leq \mathrm{E}_\theta \left[\sum_{j=1}^{T} (p_j - p_j^*)^2 = \mathrm{Regret}_\theta(T)\right].$$

Substituting the equations above into Eq (A.11), we get

$$\mathrm{E}_\lambda[\mathrm{E}_\theta[(S_t - S(\theta))^2]] \geq \frac{\left(\mathrm{E}_\lambda[\frac{a}{4b^2}]\right)^2}{\mathrm{E}_\lambda[\mathrm{Regret}_\theta(T)] + \tilde{I}(\lambda)}. \tag{A.12}$$

Similarly, for each $k = 1, \ldots, m$, by letting $U_k(\theta) = U_k^* = -c_k/(2b)$ and applying Van Trees inequality, we get

$$\mathrm{E}_\lambda[\mathrm{E}_\theta[(U_{t,k} - U_k(\theta))^2]] \geq \frac{(\mathrm{E}_\lambda[\frac{c_k}{4b^2}])^2}{\mathrm{E}_\lambda[\mathrm{Regret}_\theta(T)] + \tilde{I}(\lambda)}. \tag{A.13}$$

Combining (A.9), (A.12), (A.13), and summing over $t = 1, \ldots, T$, we have

$$\mathrm{E}_\lambda\left[\mathrm{Regret}_\theta(T)\right] \geq \sum_{t=1}^{T} \underline{b} \left\{ \frac{\left(\mathrm{E}_\lambda[\frac{a}{4b^2}]\right)^2 + \frac{1}{12}\sum_{k=1}^{m} \mathrm{E}_\lambda[\frac{c_k}{4b^2}])^2}{\mathrm{E}_\lambda[\mathrm{Regret}_\theta(T)] + \tilde{I}(\lambda)} \right\} = \frac{\Omega(mT)}{\mathrm{E}_\lambda[\mathrm{Regret}_\theta(T)] + \tilde{I}(\lambda)}.$$

Note that $\tilde{I}(\lambda)$ is a constant independent of T. Consequently, we have

$$\mathrm{E}_\lambda[\mathrm{Regret}_\theta(T)] \geq \sqrt{\Omega(T)} - \frac{\tilde{I}(\lambda)}{2} = \Omega(\sqrt{T}).$$

□

A.3.4 Proof of Theorem 3.

Proof. Proof. In the following, let $p_{t,u}^* := -\frac{a+\mathbf{c}'\mathbf{x_t}}{2b}$. We can decompose the regret into the loss due to imperfect knowledge of the true demand model, and the loss due to

price experimentation, namely

$$\begin{aligned}\text{Regret}(T) &= \sum_{t=1}^{T} \mathrm{E}[p_t^* D_t(p_t^*)] - \mathrm{E}[p_t D_t(p_t)] \\ &= \sum_{t=1}^{T} \mathrm{E}[p_t^* D_t(p_t^*)] - \mathrm{E}[p_{g,t} D_t(p_{g,t})] + \mathrm{E}[p_{g,t} D_t(p_{g,t})] - \mathrm{E}[p_t D_t(p_t)].\end{aligned}$$

The loss from price experimentation is upper bounded by

$$\begin{aligned}\sum_{t=1}^{T} \mathrm{E}[p_{g,t} D_t(p_{g,t})] - \mathrm{E}[p_t D_t(p_t)] &= -b \sum_{t=1}^{T} \mathrm{E}[\Delta p_t^2] \\ &= -b \sum_{t=1}^{T} \mathrm{E}[(q_{i_t} - q_{i_t-1})(q_{i_t+1} - q_{i_t}) t^{-1/3}] \\ &\leq 3\bar{b}\bar{\delta}^2 T^{2/3}.\end{aligned}$$

where the last line uses the assumption that $q_i - q_{i-1} \leq \bar{\delta}$ for $i = 1, \ldots, N-1$.

The loss from parameter estimation is upper bounded by

$$\begin{aligned}&\sum_{t=1}^{T} \mathrm{E}[p_t^* D_t(p_t^*)] - \mathrm{E}[p_{g,t} D_t(p_{g,t})] \\ =&\mathrm{E}[(p_{t,u}^* - (p_t^* - p_{t,u}^*) D_t(p_{t,u}^* - (p_t^* - p_{t,u}^*)] - \mathrm{E}[p_{g,t} D_t(p_{g,t})] \qquad \text{(A.14)} \\ \leq& K \mathrm{E}[|p_{t,u}^* - p_t^* + p_{t,u}^* - p_{t,g}|] \qquad \text{(A.15)} \\ \leq& K(\mathrm{E}[|p_{t,u}^* - p_t^*|] + \mathrm{E}[|p_{t,u}^* - p_{t,g}|]) \\ \leq& 2K \mathrm{E}[|p_{t,u}^* - p_{t,g}|].\end{aligned}$$

The first line, (A.14), follows from the fact that

$$\begin{aligned}\mathrm{E}[p_t^* D_t(p_t^*)] &= \mathrm{E}[(p_{t,u}^* + (p_t^* - p_{t,u}^*) D_t(p_{t,u}^* + (p_t^* - p_{t,u}^*)] \\ &= \mathrm{E}[(p_{t,u}^* - (p_t^* - p_{t,u}^*) D_t(p_{t,u}^* - (p_t^* - p_{t,u}^*)],\end{aligned}$$

by the symmetry of the function $p \mapsto \mathrm{E}[p D_t(p)]$ around its maximizer $p = p_{t,u}^*$. The second line, (A.15), follows from the mean value theorem since $\mathrm{E}[p D_t(p)]$ is a

differentiable function of p. By the mean value theorem, we have, for any $p_1, p_2 \in \{q_1, \ldots, q_N\}$, that

$$|\mathrm{E}[p_1 D_t(p_1)] - \mathrm{E}[p_2 D_t(p_2)]| \leq \max_{p \in \{q_1, \ldots, q_M\}} |\frac{dpD(p)}{dp}| \leq 2|b|p_{\max} + \bar{f},$$

thus (A.15) follows by setting $K = 2|b|p_{\max} + \bar{f}$. Finally, the third line follows from the triangle inequality, and the last line follows from the fact that $|p^*_{t,u} - p^*_t| \leq |p^*_{t,u} - p_{t,g}|$ since $p^*_t = \arg\min_{q \in \{q_1, \ldots, q_N\}} |p^*_{t,u} - q|$.

It remains to bound $\mathrm{E}[|p^*_{t,u} - p_{t,g}|]$. Since $\mathrm{E}[|p^*_{t,u} - p_{t,g}|] \leq \sqrt{\mathrm{E}[|p^*_{t,u} - p_{t,g}|^2]}$, we can then bound $\mathrm{E}[|p^*_{t,u} - p_{t,g}|^2]$ using the same argument made in the proof of Theorem 1, giving an upper bound of

$$\frac{8}{\lambda_{\min}(M)}(m+1)p^2_{\max}\mathrm{E}[(b - \hat{b}_t)^2] + O(\frac{m+1}{t-1}).$$

Lemma 2 can be applied to bound the term $\mathrm{E}[(b - \hat{b}_t)^2]$. Then, using the identity $\sqrt{x+y+z} \leq \sqrt{x} + \sqrt{y} + \sqrt{z}$ for $x, y, z \geq 0$, we can bound $\mathrm{E}[|p^*_{t,u} - p_{t,g}|]$ with

$$4\sqrt{2} \cdot \frac{p_{\max}(\bar{f} + \sigma + \bar{b}p_{\max})}{\delta\sqrt{\lambda_{\min}(M)}} \cdot \frac{\sqrt{m+1}}{t^{1/3}} + O(\sqrt{\frac{m+1}{t-1}})$$

Dependence on $m, \underline{b}, \bar{b}$ and other parameters By combining constant factors, the expected regret of the RPS algorithm over T periods can be bounded by

$$O\left((|b|p_{\max} + \bar{f})\,\frac{p_{\max}(\bar{f} + \sigma + \bar{b}p_{\max})}{\underline{\delta}\sqrt{\lambda_{\min}(M)}}\sqrt{m+1}T^{2/3}\right) + O\left(\sqrt{(m+1)T}\right),$$

where the pre-factor in the first big O notation only contains an absolute constant. □

A.3.5 Proof of Proposition 2

Proof. Proof. Consider the optimization problem

$$\max_{\alpha,\beta,\gamma} \sum_{t=1}^{T} \mathrm{E}[p^*_t D_t(p^*_t) | \{x_1, \ldots, \mathbf{x_t}\}] = \max_{\alpha,\beta,\gamma} \sum_{t=1}^{T} (-\frac{\alpha + \gamma^\mathsf{T}\mathbf{x_t}}{2\beta})(b(-\frac{\alpha + \gamma^\mathsf{T}\mathbf{x_t}}{2\beta}) + f(\mathbf{x_t})).$$

It is easy to see that for any optimal solution $(\alpha^*, \beta^*, \gamma^*)$, $(\alpha^* \frac{b}{\beta^*}, b, \gamma^* \frac{b}{\beta^*})$ is another optimal solution. Thus, setting $\beta = b$, we have the equivalent optimization problem

$$\max_{\alpha,\gamma} \sum_{t=1}^{T} (\alpha + \gamma^\mathsf{T} \mathbf{x_t})(2f(\mathbf{x_t}) - (\alpha + \gamma^\mathsf{T} \mathbf{x_t})).$$

Finally, note that

$$\arg\max_{\alpha,\gamma} \sum_{t=1}^{T} (\alpha + \gamma^\mathsf{T} \mathbf{x_t})(2f(\mathbf{x_t}) - (\alpha + \gamma^\mathsf{T} \mathbf{x_t})) = -\arg\min_{\alpha,\gamma} \sum_{t=1}^{T} (f(\mathbf{x_t}) - (\alpha + \gamma^\mathsf{T} \mathbf{x_t}))^2,$$

which proves Proposition 2. □

A.3.6 Proof of Theorem 4.

Proof. Proof. We decompose the regret as

$$\begin{aligned}
\text{Regret}(T) &= \sum_{t=1}^{T} \mathrm{E}[p_t^* D(p_t^*)] - \mathrm{E}[p_t D(p_t)] \\
&= \sum_{t=1}^{T} \mathrm{E}[p_t^* D(p_t^*)] - \mathrm{E}[p_{g,t}^u D(p_{g,t}^u)] \\
&\quad + \mathrm{E}[p_{g,t}^u D(p_{g,t}^u)] - \mathrm{E}[p_{g,t} D(p_{g,t})] + \mathrm{E}[p_{g,t} D(p_{g,t})] - \mathrm{E}[p_t D(p_t)].
\end{aligned}$$

Following the proof of the regret bound in the IID setting, the quantity in the final line is upper bounded by

$$2 \sum_{t=1}^{T} \bar{b} \delta_t^2 = 2 \sum_{t=1}^{T} \frac{\bar{b}\delta^2}{4} \frac{1}{t^{1/3}} \leq \frac{3}{2} \bar{b} \delta^2 T^{2/3}.$$

To bound the difference between the oracle's revenue and the revenue earned by the greedy prices, we let $y_t = D_t - b p_t = f(\mathbf{x_t}) + \epsilon_t$. Let $y_t' = D_t - \hat{b}_t p_t$. Let $\mathbf{e_t} = (a_t, c_t)$ and let $\mathbf{e_x} = (a_x, c_x)$ denote the parameters of the clairvoyant's demand model conditional on the realization $\{x_1, \ldots, \mathbf{x_t}\}$. Let $\mathrm{E}_\mathbf{x}$ denote the expectation

conditional on a realization $\{\mathbf{x_1}, \ldots, \mathbf{x_t}\}$, namely

$$\mathrm{E}_{\mathbf{x}}[\cdot] = \mathrm{E}[\cdot | \mathbf{x_t} \text{ for } t = 1 \ldots T].$$

By rewriting the demands and prices in terms of y_t and y'_t we have

$$\sum_{t=1}^{T} \mathrm{E}[p_t^* D(p_t^*)] - \mathrm{E}[p_{g,t}^u D_i(p_{g,t}^u)] = \frac{1}{4|b|} \sum_{t=1}^{T} \mathrm{E}[\mathrm{E}_{\mathbf{x}}[(\mathbf{e_t}^{\mathsf{T}} \tilde{\mathbf{x}}_\mathbf{t} - y'_t)^2 - (\mathbf{e_x}^{\mathsf{T}} \tilde{\mathbf{x}}_\mathbf{t} - y'_t)^2]] \tag{A.16}$$

$$+ \frac{1}{|b|} \mathrm{E}[\mathrm{E}_{\mathbf{x}}[(p_{t,g}^u)^2 (b^2 - \hat{b}_t^2)]]. \tag{A.17}$$

$$+ \frac{1}{2|b|} \mathrm{E}[\mathrm{E}_{\mathbf{x}}[(y'_t - y_t)(\mathbf{e_t}^{\mathsf{T}} \tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e_x}^{\mathsf{T}} \tilde{\mathbf{x}}_\mathbf{t})]] \tag{A.18}$$

$$+ \frac{1}{|b|} \mathrm{E}[\mathrm{E}_{\mathbf{x}}[y_t p_{t,g}^u (\hat{b}_t - b)]]. \tag{A.19}$$

First, we will bound (A.16). Define $M_t = I_{m+1} + \sum_{s=1}^{t} \tilde{\mathbf{x}}_s \tilde{\mathbf{x}}_s^{\mathsf{T}}$. The closed form expression for the estimator $\mathbf{e_t}$ at period t is $(M_t)^{-1}(\sum_{s=1}^{t-1} y_s \tilde{\mathbf{x}}_s)$. Expanding the expressions for $\mathbf{e_t}$ in (A.16), we see that most of the terms in the expansion are telescoping, giving

$$\sum_{t=1}^{T} \mathrm{E}_{\mathbf{x}}[(\mathbf{e_t}^{\mathsf{T}} \tilde{\mathbf{x}}_\mathbf{t} - y'_t)^2 - (\mathbf{e_x}^{\mathsf{T}} \tilde{\mathbf{x}}_\mathbf{t} - y'_t)^2] = \sum_{t=1}^{T} \mathrm{E}_{\mathbf{x}}[(y'_t)^2 \tilde{\mathbf{x}}_\mathbf{t}^{\mathsf{T}} M_t^{-1} \tilde{\mathbf{x}}_\mathbf{t}] \tag{A.20}$$

$$+ \mathrm{E}_{\mathbf{x}}[\|\mathbf{e_x} - \mathbf{e_1}\|^2 - (\mathbf{e_x} - \mathbf{e_{T+1}})^{\mathsf{T}} M_T (\mathbf{e_x} - \mathbf{e_{T+1}})] \tag{A.21}$$

$$+ \mathrm{E}_{\mathbf{x}}[(\mathbf{e_1}^{\mathsf{T}} \tilde{\mathbf{x}}_1)^2 + (\mathbf{e_x}^{\mathsf{T}} \tilde{\mathbf{x}}_{T+1})^2] \tag{A.22}$$

$$- \sum_{t=1}^{T} \mathrm{E}_{\mathbf{x}}[(\mathbf{e_{T+1}}^{\mathsf{T}} \tilde{\mathbf{x}}_{t+1})^2 \tilde{\mathbf{x}}_{t+1}^{\mathsf{T}} M_t^{-1} \tilde{\mathbf{x}}_{t+1} \tag{A.23}$$

$$- (\mathbf{e_{T+1}}^{\mathsf{T}} \tilde{\mathbf{x}}_{T+1})^2 - (\mathbf{e_x}^{\mathsf{T}} \tilde{\mathbf{x}}_1)^2].$$

Since $\mathbf{e_1} = (I + \tilde{\mathbf{x}}_1 \tilde{\mathbf{x}}_1^{\mathsf{T}})^{-1} \cdot 0 = 0$, $\|\mathbf{e_x} - \mathbf{e_1}\|^2 = \|\mathbf{e_x}\|^2$. Then since M_t is positive semi-definite for all t, (A.21) is upper bounded by $\|\mathbf{e_x}\|^2 \leq \bar{a}^2 + \bar{c}^2$. Since $\mathbf{e_1} = 0$ and $\mathbf{x_{T+1}}$ can be set to 0, (A.22) is 0. The final line is upper bounded by 0.

Finally, to bound Eq (A.20), we can write

$$\sum_{t=1}^{T} \mathrm{E}_{\mathbf{x}}[(y_t')^2 \tilde{\mathbf{x}}_t^\mathsf{T} M_t^{-1} \tilde{\mathbf{x}}_t] = \sum_{t=1}^{T} \mathrm{E}_{\mathbf{x}}[(f(\mathbf{x_t}) + \epsilon_t + (b - \hat{b}_t)p_t)^2 \tilde{\mathbf{x}}_t^\mathsf{T} M_t^{-1} \tilde{\mathbf{x}}_t]$$
$$\leq \sum_{t=1}^{T} (2\sigma^2 + 2\bar{f}^2 + 4\bar{b}^2 p_{\max}^2) \tilde{\mathbf{x}}_t^\mathsf{T} M_t^{-1} \tilde{\mathbf{x}}_t.$$

The second line follows from the fact that ϵ_t is independent of the other terms and that it is mean 0 and variance σ^2. We also use the boundedness of f, $\hat{b}_t$ and p_t. Finally, using the identity

$$\mathbf{x}^\mathsf{T}(\Sigma + \mathbf{x}\mathbf{x}^\mathsf{T})^{-1}\mathbf{x} = \frac{\det(\Sigma)}{\det(\Sigma + \mathbf{x}\mathbf{x}^\mathsf{T})}$$

for any matrix Σ, we have

$$(2\sigma^2 + 2\bar{f}^2 + 4\bar{b}^2 p_{\max}^2) \sum_{t=1}^{T} \tilde{\mathbf{x}}_t^\mathsf{T} M_t^{-1} \tilde{\mathbf{x}}_t \leq (2\sigma^2 + 2\bar{f}^2 + 4\bar{b}^2 p_{\max}^2) \sum_{t=1}^{T} 1 - \frac{\det(M_{t-1})}{\det(M_t)}$$
$$\leq (2\sigma^2 + 2\bar{f}^2 + 4\bar{b}^2 p_{\max}^2) \sum_{k=1}^{m+1} \log(1 + \lambda_k),$$

where the λ_js are the eigenvalues of $\sum_{t=1}^{T} \tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t^\mathsf{T}$. The sum of the λ_js is at most $T \cdot \max_t \|\tilde{\mathbf{x}}_t\|^2$, which in turn is at most $\sqrt{m+1}T$. Thus the last line is $O((m+1) \cdot \log(T(m+1)))$. Then (A.16) does not dominate the regret bound.

Now we will bound Eq (A.17). Using the definition $p_{g,t}^u = -\frac{\mathbf{e_t}^T \tilde{\mathbf{x}}_t}{2b_t}$ and the fact that $|b_t| \geq \underline{b}$ gives

$$\mathrm{E}_{\mathbf{x}}[(p_{t,g}^u)^2 (b^2 - \hat{b}_t^2)] \leq \frac{1}{2\underline{b}} \mathrm{E}_{\mathbf{x}}[(\mathbf{e_t}^T \tilde{\mathbf{x}}_t)^2 (b^2 - \hat{b}_t^2)]$$
$$\leq \frac{1}{\underline{b}} \mathrm{E}_{\mathbf{x}}[((\mathbf{e_t}^T \tilde{\mathbf{x}}_t - \mathbf{e_x}^T \tilde{\mathbf{x}}_t)^2 + (\mathbf{e_x}^T \tilde{\mathbf{x}}_t)^2)(b^2 - \hat{b}_t^2)]$$
$$\leq \frac{1}{\underline{b}} ((2\bar{b}^2) \mathrm{E}_{\mathbf{x}}[(\mathbf{e_t}^T \tilde{\mathbf{x}}_t - \mathbf{e_x}^T \tilde{\mathbf{x}}_t)^2] + (\bar{a} + \bar{c})^2 \mathrm{E}_{\mathbf{x}}[b^2 - \hat{b}_t^2]). \quad \text{(A.24)}$$

The second line follows from the identity $(x+y)^2 \leq 2x^2 + 2y^2$. The third line follows from the fact that $b^2 + \hat{b}_t^2 \leq 2\bar{b}^2$ due to the assumptions on b and the projection step

in the algorithm, as well as from the assumptions on $\mathbf{e_x}$. Now, to bound $\mathrm{E_x}[(\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t})^2]$ in Eq (A.24), note that we have

$$\sum_{t=1}^{T} \mathrm{E_x}[(\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t})^2] = \sum_{t=1}^{T} \mathrm{E_x}[(\mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t} - y_t)^2 - (\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - y_t)^2]. \tag{A.25}$$

This is because

$$\begin{aligned}
&\sum_{t=1}^{T} \mathrm{E_x}[(\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - y_t)^2 - (\mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t} - y_t)^2] \\
&= \sum_{t=1}^{T} \mathrm{E_x}[(\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t})^2] + \mathrm{E_x}[(\mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t} - y_t)\tilde{\mathbf{x}}_\mathbf{t}^T(\mathbf{e_t} - \mathbf{e_x})] \\
&= \sum_{t=1}^{T} \mathrm{E_x}[(\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t})^2] + \mathrm{E_x}[(\mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t} - y_t)\tilde{\mathbf{x}}_\mathbf{t}^T(\mathbf{e_t} - \mathbf{e_x})] \\
&= \sum_{t=1}^{T} \mathrm{E_x}[(\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t})^2],
\end{aligned}$$

where the second line follows from the fact that $y_t = f(\mathbf{x_t}) + \epsilon_t$ and $\mathrm{E_x}[\epsilon_t] = 0$, ϵ_t independent of $\mathbf{x_t}$, $\mathbf{e_x}$, $\mathbf{e_t}$, and the final line follows from the first order conditions of the minimization problem Eq (2.10), as given by Eq (2.11). Eq (A.25) thus implies that $\sum_{t=1}^{T} \mathrm{E_x}[(\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t})^2]$ is $O((m+1)\log(T(m+1))$.

To bound $\mathrm{E_x}[b^2 - \hat{b}_t^2]$ in Eq (A.24), we can write

$$\begin{aligned}
\mathrm{E_x}[b^2 - \hat{b}_t^2] &= \mathrm{E_x}[(b - \hat{b}_t)(b + \hat{b}_t)] \\
&\leq 2\bar{b}\mathrm{E_x}[|b - \hat{b}_t|] \\
&\leq 2\bar{b}\sqrt{\mathrm{E}[(\hat{b}_t - b)^2]} \\
&\leq 8\bar{b}\frac{\sqrt{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}}{\delta}\frac{1}{t^{1/3}},
\end{aligned}$$

where the second line follows from our assumed bounds on b and the projection step in the algorithm, the third line follows from Jensen's inequality since the function $x \mapsto x^2$ is convex, and the final line follows from Lemma 2. Then, $\sum_{t=1}^{T} \mathrm{E_x}([b^2 - \hat{b}_t^2]) \leq \frac{32}{3}\bar{b}\frac{\sqrt{\bar{f}^2+\sigma^2+\bar{b}^2p_{\max}^2}}{\delta}T^{2/3}$, which dominates the $O((m + 1\log(T(m+1))$ term

$\sum_{t=1}^{T} \mathrm{E}_{\mathbf{x}}[(\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t})^2]$, and implies that Eq (A.17) is $O(T^{2/3})$.

Similar ideas can be used to bound Eq (A.18) and (A.19). For Eq (A.18), using the identity $y'_t - y_t = (b - \hat{b}_t)p_t$, we have

$$\begin{aligned}
\sum_{t=1}^{T} \mathrm{E}_{\mathbf{x}}[(y'_t - y_t)(\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t})] &\leq 2\bar{b}p_{\max} \sum_{t=1}^{T} \mathrm{E}_{\mathbf{x}}[|\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t}|] \\
&\leq 2\bar{b}p_{\max} \sum_{t=1}^{T} \sqrt{\mathrm{E}_{\mathbf{x}}[(\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t})^2]} \\
&\leq 2\bar{b}p_{\max}\sqrt{T}\sqrt{\sum_{t=1}^{T} \mathrm{E}_{\mathbf{x}}[(\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e_x}^T\tilde{\mathbf{x}}_\mathbf{t})^2]}.
\end{aligned}$$

The first line follows from our assumption on b, and that $\hat{b}$ and p_t are projections onto bounded sets. The second line follows from using Jensen's inequality again, and the final step follows from the Cauchy-Schwarz theorem. Then, applying Eq (A.25) again, we see that Eq (A.18) is $O(\sqrt{(m+1)T}\log(T(m+1)))$.

Finally, each term of Eq (A.19) can be written as

$$\begin{aligned}
\mathrm{E}_{\mathbf{x}}[y_t p^u_{t,g}(\hat{b}_t - b)] &= \mathrm{E}_{\mathbf{x}}[f(\mathbf{x_t})p^u_{t,g}(\hat{b}_t - b)] \\
&\leq \frac{\bar{f}}{2\underline{b}}\mathrm{E}_{\mathbf{x}}[(\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t})(\hat{b}_t - b)] \\
&= \frac{\bar{f}}{2\underline{b}}\mathrm{E}_{\mathbf{x}}[(\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e}^T\tilde{\mathbf{x}}_\mathbf{t})(\hat{b}_t - b) + (\mathbf{e}^T\tilde{\mathbf{x}}_\mathbf{t})(\hat{b}_t - b)] \\
&\leq \frac{\bar{f}}{2\underline{b}}(2\bar{b}\mathrm{E}[|\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{e}^T\tilde{\mathbf{x}}_\mathbf{t}|] + (\bar{a} + \bar{c})\mathrm{E}_{\mathbf{x}}[|\hat{b}_t - b|]).
\end{aligned}$$

The second line follows from the definition of y_t and the fact that $\mathrm{E}[\epsilon_t] = 0$ and ϵ_t is independent from $p^u_{t,g}$ and $\hat{b}_t$. The final line follows from our assumption on b, and that $\hat{b}$ and p_t are projections onto bounded sets. We have already shown that $\sum_{t=1}^{T} \mathrm{E}[|\mathbf{e_t}^T\tilde{\mathbf{x}}_\mathbf{t} - e^T\tilde{\mathbf{x}}_\mathbf{t}|]$ is $O((m+1\log(T(m+1))$, and that $\sum_{t=1}^{T} \mathrm{E}_{\mathbf{x}}[|\hat{b}_t - b|]$ is $O(T^{2/3})$. Then Eq (A.19) is $O(T^{2/3})$, which implies that the RPS algorithm is $O(T^{2/3})$ as well, thus concluding the proof.

Dependence on $m, \bar{a}, \bar{b}, \bar{c}$ and other parameters By combining constant fac-

tors, the expected regret of the RPS algorithm over T periods can be bounded by

$$O\left(\bar{b}\delta^2 + \frac{(\bar{a}+\bar{c})^2}{\underline{b}}\frac{\sqrt{\bar{f}^2+\sigma^2+\bar{b}^2p_{\max}^2}}{\delta}\left((\bar{a}+\bar{c})\bar{b}+\frac{1}{\underline{b}}\right)T^{2/3}\right)$$
$$+ O\left(\sqrt{(m+1)T}\log(T(m+1)) + (m+1)\log(T(m+1))\right)$$

where the pre-factor in the first big O notation only contains an absolute constant.

□

A.3.7 Lemmas

Lemma 2 (Bound on $\hat{b}_t$). $\mathrm{E}[(\hat{b}_t - b)^2]$ *can be bounded as follows:*

- *When Algorithm 1 is applied to the IID setting, for $t \geq 4$, we have*

$$\mathrm{E}[(\hat{b}_t - b)^2] \leq 12 \cdot \frac{\bar{f}^2+\sigma^2+\bar{b}^2p_{\max}^2}{\delta^2} \cdot \frac{1}{\sqrt{t}}.$$

- *When Algorithm 2 is applied to the price ladder setting, for $t \geq 2$, we have*

$$\mathrm{E}[(\hat{b}_t - b)^2] \leq 4 \cdot \frac{\bar{f}^2+\sigma^2+\bar{b}^2p_{\max}^2}{\underline{\delta}^2} \cdot \frac{1}{t^{2/3}}.$$

- *When Algorithm 3 is applied to the non IID setting, for $t \geq 4$ we have*

$$\mathrm{E}[(\hat{b}_t - b)^2] \leq 12 \cdot \frac{\bar{f}^2+\sigma^2+\bar{b}^2p_{\max}^2}{\delta^2} \cdot \frac{1}{t^{2/3}}.$$

Proof. Proof. Define the constant α_1 such that

$$\alpha_1 = \begin{cases} \frac{1}{4} & \text{in the IID setting,} \\ \frac{1}{6} & \text{in the price ladder setting,} \\ \frac{1}{6} & \text{in the non IID setting.} \end{cases}$$

We will first consider the **IID and non IID settings**, where prices are drawn from continuous price intervals at each time period. Using the definitions of b_t in

Algorithms 1 and 3, $\hat{b}_t = \text{Proj}(\hat{b}_t^u, B)$, where $\hat{b}_t^u = \frac{\sum_{s=1}^{t-1} \Delta p_s D_s}{\sum_{s=1}^{t-1} \Delta p_s^2}$. Since the true parameter $b \in B$, we have

$$\begin{aligned}
\mathrm{E}[(\hat{b}_t - b)^2] &\leq \mathrm{E}[(\hat{b}_t^u - b)^2] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s D_s}{\sum_{s=1}^{t-1} \Delta p_s^2} - b\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s (f(\mathbf{x_s}) + \epsilon_s + b p_{g,s} + b\Delta p_s)}{\sum_{s=1}^{t-1} \Delta p_s^2} - b\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s (f(\mathbf{x_s}) + \epsilon_s + b p_{g,s})}{\sum_{s=1}^{t-1} \Delta p_s^2}\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s (f(\mathbf{x_s}) + \epsilon_s + b p_{g,s})}{\sum_{s=1}^{t-1} \frac{\delta^2}{4} s^{-2\alpha_1}}\right)^2\right].
\end{aligned}$$

In the last equality, we used the fact that $\Delta p_s^2 = \frac{\delta^2}{4} s^{-2\alpha_1}$.

Note that Δp_s's for all s are mutually independent, independent of $\mathbf{x_s}$, and have mean 0, so

$$\begin{aligned}
&\mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s (f(\mathbf{x_t}) + \epsilon_s + b p_{g,s})}{\sum_{s=1}^{t-1} \frac{\delta^2}{4} s^{-2\alpha_1}}\right)^2\right] \\
=&\mathrm{E}\left[\frac{\sum_{s=1}^{t-1} \Delta p_s^2 (f(\mathbf{x_s}) + \epsilon_s + b p_{g,s})^2}{(\sum_{s=1}^{t-1} \frac{\delta^2}{4} s^{-2\alpha_1})^2}\right] \\
\leq&\mathrm{E}\left[\frac{\sum_{s=1}^{t-1} 3\Delta p_s^2 (f(\mathbf{x_s})^2 + \epsilon_t^2 + b^2 p_{g,j}^2)}{(\sum_{s=1}^{t-1} \frac{\delta^2}{4} s^{-2\alpha_1})^2}\right] \\
\leq&12 \cdot \frac{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}{\delta^2} \cdot \frac{1}{\sum_{s=1}^{t-1} s^{-2\alpha_1}}. \qquad \text{(A.26)}
\end{aligned}$$

We used the fact that $(x+y+z)^2 \leq 3(x^2+y^2+z^2)$. In the last step, we used the definition that $\Delta p_s^2 = \frac{\delta^2}{4} s^{-2\alpha_1}$ and the assumption that $f(\mathbf{x_s}), b, p_{g,s}$ are bounded.

Now consider the **price ladder setting**. Using the definitions of b_t in Algorithm 2, $\hat{b}_t = \text{Proj}(\hat{b}_t^u, B)$, where $\hat{b}_t^u = \frac{\sum_{s=1}^{t-1} \Delta p_s D_s}{\sum_{s=1}^{t-1} (q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1}}$. Since the true parameter

$b \in B$, we have

$$\begin{aligned}
\mathrm{E}[(\hat{b}_t - b)^2] &\leq \mathrm{E}[(\hat{b}_t^u - b)^2] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s D_s}{\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s})s^{-2\alpha_1}} - b\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s(f(\mathbf{x_s}) + \epsilon_s + bp_{g,s} + b\Delta p_s)}{\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s})s^{-2\alpha_1}} - b\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s(f(\mathbf{x_s}) + \epsilon_s + bp_{g,s})}{\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s})s^{-2\alpha_1}}\right)^2\right].
\end{aligned}$$

The last line follows from the fact that $\mathrm{E}[\Delta p_s^2 | p_{g,t} = q_{i_s}] = (q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s})s^{-2\alpha_1}$.

As before, Δp_s's for all s are mutually independent, independent of $\mathbf{x_s}$, and have mean 0, so

$$\mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s(f(\mathbf{x_t}) + \epsilon_s + bp_{g,s})}{\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s})s^{-2\alpha_1}}\right)^2\right] \tag{A.27}$$

$$\begin{aligned}
&= \mathrm{E}\left[\frac{\sum_{s=1}^{t-1} \Delta p_s^2(f(x_s) + \epsilon_s + bp_{g,s})^2}{(\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s})s^{-2\alpha_1})^2}\right] \\
&\leq \mathrm{E}\left[\frac{\sum_{s=1}^{t-1} 3\Delta p_s^2(f(x_s)^2 + \epsilon_t^2 + b^2 p_{g,j}^2)}{(\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s})s^{-2\alpha_1})^2}\right] \\
&\leq 3 \cdot \mathrm{E}\left[\frac{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}{\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s})s^{-2\alpha_1}}\right] \\
&\leq 3 \cdot \frac{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}{\underline{\delta}^2} \cdot \frac{1}{\sum_{s=1}^{t-1} s^{-2\alpha_1}}.
\end{aligned} \tag{A.28}$$

We used the fact that $(x+y+z)^2 \leq 3(x^2+y^2+z^2)$. The second to last step uses the definition that $\mathrm{E}[\Delta p_s^2 | p_{g,t} = q_{i_s}] = (q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s})s^{-2\alpha_1}$, and the assumption that $f(\mathbf{x_s}), b, p_{g,s}$ are bounded. The last step uses the assumption that $q_i - q_{i-1} \geq \underline{\delta}$ for $i = 1, \ldots, N+1$.

Now for the **IID, non IID and price ladder settings**,

$$\sum_{s=1}^{t-1} s^{-2\alpha_1} \geq \int_{y=1}^{t} y^{-2\alpha_1} dy = \frac{1}{1-2\alpha_1}(t^{1-2\alpha_1} - 1),$$

and we have for $t \geq 4$ that

$$\frac{1}{\sum_{s=1}^{t-1} s^{-2\alpha_1}} \leq 2(1-2\alpha_1)t^{2\alpha_1 - 1}. \tag{A.29}$$

Substituting (A.29) into (A.26) and (A.28) respectively, we prove the lemma in the IID, price ladder and non IID settings.

□

Lemma 3 (Bound on $\|QM_t^{-1}Q\|_2$). *Let* $M = \mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\mathsf{T}]$, $V = \mathrm{E}[(Q^{-1}\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\mathsf{T}Q^{-1} - I)^2]$ *and* $M_t = \frac{1}{t-1}\sum_{s=1}^{t-1} \tilde{\mathbf{x}}_s\tilde{\mathbf{x}}_s^\mathsf{T}$. *For any* $t \geq 2$, M_t *is invertible and* $\|QM_t^{-1}Q\|_2 \leq 2$ *with probability at least*

$$1 - 2(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right).$$

Proof. Proof. For any $s = 1, \ldots, t-1$, we have $\mathrm{E}[I - Q^{-1}\tilde{\mathbf{x}}_s\tilde{\mathbf{x}}_s^\mathsf{T}Q^{-1}] = 0$, where I is the identity matrix. In addition, for an arbitrary matrix A, it holds that $\|A\|_2 \leq \|A\|_F$, so by $\|\tilde{\mathbf{x}}_s\|_\infty \leq 1$, we have

$$\begin{aligned}
\lambda_{\max}(I - Q^{-1}\tilde{\mathbf{x}}_s\tilde{\mathbf{x}}_s^\mathsf{T}Q^{-1}) \leq & \|I - Q^{-1}\tilde{\mathbf{x}}_s\tilde{\mathbf{x}}_s^\mathsf{T}Q^{-1}\|_2 \\
\leq & \|Q^{-1}\|_2\|M - \tilde{\mathbf{x}}_s\tilde{\mathbf{x}}_s^\mathsf{T}\|_2\|Q^{-1}\|_2 \\
\leq & \|Q^{-1}\|_2\|M - \tilde{\mathbf{x}}_s\tilde{\mathbf{x}}_s^\mathsf{T}\|_F\|Q^{-1}\|_2 \\
\leq & \frac{1}{\sqrt{\lambda_{\min}(M)}} \cdot 2(m+1) \cdot \frac{1}{\sqrt{\lambda_{\min}(M)}} = \frac{2(m+1)}{\lambda_{\min}(M)}.
\end{aligned}$$

Note that we used the submultiplicative property of the spectral norm. Since $\{\tilde{\mathbf{x}}_s\}$ are independent and identically distributed, we apply the matrix Bernstein bound

(Lemma 5) with $\alpha = (t-1)/2$ to yield

$$\begin{aligned}
&\mathbb{P}\left[\lambda_{\max}\left(\sum_{s=1}^{t-1}\frac{I - Q^{-1}\tilde{\mathbf{x}}_s\tilde{\mathbf{x}}_s^\top Q^{-1}}{t-1}\right) > \frac{1}{2}\right] \\
\leq &(m+1)\exp\left(-\frac{t^2/2}{\|(t-1)V\|_2 + 2(m+1)t/(3\lambda_{\min}(M))}\right) \\
= &(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right).
\end{aligned}$$

By an identical argument, we also have

$$\begin{aligned}
&\mathbb{P}\left[\lambda_{\max}\left(-\sum_{s=1}^{t-1}\frac{I - Q^{-1}\tilde{\mathbf{x}}_s\tilde{\mathbf{x}}_s^\top Q^{-1}}{t-1}\right) > \frac{1}{2}\right] \\
\leq &(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right).
\end{aligned}$$

Thus we have

$$\begin{aligned}
&\mathbb{P}[\|I - Q^{-1}M_tQ^{-1}\|_2 > \frac{1}{2}] && \text{(A.30)} \\
= &\mathbb{P}[\max\{\lambda_{\max}(I - Q^{-1}M_tQ^{-1}), \lambda_{\max}(Q^{-1}M_tQ^{-1} - I)\} > \frac{1}{2}] \\
\leq &2(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right). && \text{(A.31)}
\end{aligned}$$

We can write $Q^{-1}M_tQ^{-1} = I + (Q^{-1}M_tQ^{-1} - I)$, then by Weyl's inequality,

$$\begin{aligned}
\lambda_{\min}(Q^{-1}M_tQ^{-1}) &\geq \lambda_{\min}(I) + \lambda_{\min}(Q^{-1}M_tQ^{-1} - I) \\
&\geq 1 - \|Q^{-1}M_tQ^{-1} - I\|_2
\end{aligned}$$

By Eq (A.31), with probability at least

$$1 - 2(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right),$$

we have $\lambda_{\min}(Q^{-1}M_tQ^{-1}) \geq 1/2$. Since $Q^{-1}M_tQ^{-1} = Q^{-1}\frac{\sum_{s=1}^{t-1}\tilde{\mathbf{x}}_s\tilde{\mathbf{x}}_s^\top}{t-1}Q^{-1}$ is positive

semidefinite, $\lambda_{\min}(Q^{-1}M_tQ^{-1}) > 0$ implies that it is invertible. Then

$$\|QM_t^{-1}Q\|_2 = \frac{1}{\lambda_{\min}(Q^{-1}M_tQ^{-1})} \leq 2.$$

This proves the lemma. □

Lemma 4 (Optimal Policy Structure for Linear Demand). *Suppose the true demand function is linear, given by*

$$D_t(p) = a + bp + \mathbf{c}^\mathsf{T}\mathbf{x_t} + \epsilon.$$

Then, it is optimal for the seller to use a linear pricing policy of the form $p_t = S_t + (U_t)^\mathsf{T}\mathbf{x_t}$, *where* S_t *and* U_t *are measurable with respect to* $\mathcal{H}_{t-1}$.

Proof. Proof. Suppose the seller uses a pricing policy $\pi(\mathcal{H}_{t-1}, \mathbf{x_t}) = \pi_t(\mathbf{x_t})$ at period t, where function $\pi_t(\cdot)$ is measurable with respect to t and could be nonlinear. We denote by $\tilde{\mathrm{E}}[\cdot]$ the conditional expectation operator $\mathrm{E}[\cdot \mid \mathcal{H}_{t-1}]$. Let S and U be the optimal solution of the following least squares problem:

$$\max_{s\in\mathbb{R}, u\in\mathbb{R}^m} \tilde{\mathbb{E}}\left[\left(\pi_t(\mathbf{x_t}) - s - u^\mathsf{T}\mathbf{x_t}\right)^2\right].$$

Clearly, S and U are measurable with respect to $\mathcal{H}_{t-1}$. By the first order condition, the optimal solution (S, U) satisfies

$$\tilde{\mathbb{E}}[\pi_t(\mathbf{x_t}) - S - U^\mathsf{T}\mathbf{x_t}] = 0, \quad \tilde{\mathbb{E}}\left[\mathbf{x_t}\left(\pi_t(\mathbf{x_t}) - S - U^\mathsf{T}\mathbf{x_t}\right)\right] = 0. \tag{A.32}$$

Now, let us compare the conditional expected revenue of price $\pi_t(\mathbf{x_t})$ and price

$S + U^{\mathsf{T}}\mathbf{x_t}$. We have

$$
\begin{aligned}
&\tilde{\mathbb{E}}\left[\pi_t(\mathbf{x_t})D_t(\pi_t(\mathbf{x_t})) - (S + U^{\mathsf{T}}\mathbf{x_t})D_t(S + U^{\mathsf{T}}\mathbf{x_t})\right] \\
=&\tilde{\mathbb{E}}\left[\pi_t(\mathbf{x_t}) \cdot \left(a + b\pi_t(\mathbf{x_t}) + \mathbf{c}^{\mathsf{T}}\mathbf{x_t}\right) - (S + U^{\mathsf{T}}\mathbf{x_t})(a + b \cdot (S + U^{\mathsf{T}}\mathbf{x_t}) + \mathbf{c}^{\mathsf{T}}\mathbf{x_t})\right] \\
=&b\tilde{\mathbb{E}}\left[(\pi_t(\mathbf{x_t}))^2 - (S + U^{\mathsf{T}}\mathbf{x_t})^2\right] + \tilde{\mathbb{E}}\left[(a + \mathbf{c}^{\mathsf{T}}\mathbf{x_t})(\pi_t(\mathbf{x_t}) - S - U^{\mathsf{T}}\mathbf{x_t})\right] \qquad \text{(A.33)} \\
=&b\tilde{\mathbb{E}}\left[(\pi_t(\mathbf{x_t}))^2 - (S + U^{\mathsf{T}}\mathbf{x_t})^2\right] \\
=&b\left\{\tilde{\mathbb{E}}\left[(\pi_t(\mathbf{x_t}) - S - U^{\mathsf{T}}\mathbf{x_t})^2\right] + 2\tilde{\mathbb{E}}\left[(S + U^{\mathsf{T}}\mathbf{x_t})(\pi_t(\mathbf{x_t}) - S - U^{\mathsf{T}}\mathbf{x_t})\right]\right\} \qquad \text{(A.34)} \\
=&b\tilde{\mathbb{E}}\left[(\pi_t(\mathbf{x_t}) - S - U^{\mathsf{T}}\mathbf{x_t})^2\right] \leq 0.
\end{aligned}
$$

The second term of Eq (A.33) and the second term of Eq (A.34) are both zero because of the first order condition Eq (A.32). In the last step, recall that the price sensitivity parameter $b < 0$.

By taking the expectation over history $\mathcal{H}_{t-1}$, we have

$$
\mathrm{E}\left[\pi_t(\mathbf{x_t})D_t(\pi_t(\mathbf{x_t})) - (S + U^{\mathsf{T}}\mathbf{x_t})D_t(S + U^{\mathsf{T}}\mathbf{x_t})\right] \leq 0,
$$

so if $p_t = \pi_t(\mathbf{x_t})$ is a nonlinear pricing policy, it is dominated by a linear pricing policy $p_t = S + U^{\mathsf{T}}\mathbf{x_t}$.

□

Lemma 5 (Matrix Bernstein bound, Tropp (2012)). *Consider a finite sequence X_k of independent, random, self-adjoint matrices with dimension d. Assume that each random matrix satisfies*

$$
\mathrm{E}[X_k] = 0 \text{ and } \lambda_{\max}(X_k) \leq R \text{ almost surely,}
$$

then for all $t \geq 0$,

$$
\mathbb{P}\left[\lambda_{\max}(\sum_k X_k) \geq t\right] \leq d\exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right) \text{ where } \sigma^2 = \|\sum_k \mathrm{E}[X_k^2]\|_2.
$$

Bibliography

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434.

Appendix B

Appendix to Chapter 3 (Feature-based Dynamic Pricing for Fashion Retail: A Case Study)

In this Appendix, we bridge the gap between the synthetic numerical simulations in Section 2.4 of Chapter 2 and the experiments on real world fashion retail data in Section 3.5 of Chapter 3. Using the fashion retail dataset and ground truth demand model introduced in Chapter 3, we compare the estimated revenues earned by RPS with the following dynamic pricing algorithms that we benchmarked RPS' performance against in Section 2.4:

- *Greedy algorithm*: The greedy algorithm (Algorithm 4) operates by estimating the demand parameters at each time period using linear regression, then setting the price to the optimal price assuming that the estimated parameters are the true parameters. This algorithm has been shown to be asymptotically optimal by Qiang and Bayati (2016) in a linear demand model setting with features, and with the availability of an incumbent price, but in general is known to suffer from *incomplete learning*, i.e., insufficient exploration in price Keskin and Zeevi (2014).

- *One-stage regression*: This algorithm introduces randomized price shocks to

force price exploration, but uses a one-stage regression instead of a two-step regression as in RPS to learn the parameters. A full description of the one-stage regression algorithm (Algorithm 5) is given below. The one-stage regression algorithm is analogous to the class of semi-myopic algorithms introduced by Keskin and Zeevi (2014), which use (deterministic) price perturbations to guarantee sufficient exploration. However, Algorithm 5 does not consider the price endogeneity effect caused by model misspecification in the estimation process.

The variant of RPS that we benchmark against competing dynamic pricing algorithms is as follows: It implements batch pricing, and simultaneously sets prices for all I_t items within each subclass at the start of every week t. It selects prices from a price ladder, denoted by $\{q_1, \ldots, q_{N_t}\}$. It ignores all the other fixed inventory and markdown pricing constraints listed in Section 3.3, but assumes that for each item grandparent-district-week tuple, the price charged is restricted to within 20% of the historical price charged by the retailer. This constraint functions as a proxy for the markdown pricing constraint while allowing us to adhere to the modeling framework of Chapter 2.

The algorithm statement of this batch-pricing variant of RPS is given in Algorithm 7. The pseudocodes for the one-stage regression and greedy analogues of Algorithm 7 are omitted, as the one-stage regression analogue simply replaces the two-step regression procedure with a one-step regression, and the greedy analogue simply sets price $p_t \leftarrow p_{g,t}$, and uses a one-step regression.

We can now use our ground truth demand model from Section 3.4 of Chapter 3 to benchmark the RPS algorithm against the greedy and one-stage regression algorithms. For each subclass, we run all three algorithms from the start to the end of the respective selling horizon. During the first two weeks, the algorithms operate by setting the price of each item as the sum of the historical price chosen by the retailer and a random component, until sufficiently many demand observations have been collected to uniquely determine the parameter c_S. To estimate the counterfactual demand that would have resulted from RPS choosing a particular price, we first calculate the corresponding expected demand using our linear-random forest hybrid

Algorithm 7 Random Price Shock (RPS) algorithm with batch updating.

input: parameter bounds $B = [-\bar{b}, -\underline{b}]$

initialize: choose $\hat{a}_1 = 0$, $\hat{b}_1 = -\bar{b}$, $\hat{c}_1 = 0$

for $t = 1, \ldots, T$ **do**

  for items $j = 1, \ldots, I_t$ **do**

    set $i \leftarrow I_1 + \ldots + I_{t-1} + j$

    set $S \leftarrow S \cup \{i\}$

    given $\mathbf{x_t}$, set unconstrained greedy price: $p^u_{g,t} \leftarrow -(\hat{a}_t + \hat{\mathbf{c}}_\mathbf{t}^T \mathbf{x_t})/(2\hat{b}_t)$

    find $l_t = \arg\min_{l \in \{1,\ldots,N_t\}} |q_l - p^u_{g,t}|$ and set constrained greedy price: $p_{g,t} \leftarrow q_{l_t}$

    generate an independent random variable

$$\Delta p_t \leftarrow \begin{cases} q_{l_t} - q_{l_t - 1} \text{ w.p. } \frac{q_{l_t+1} - q_{l_t}}{(q_{l_t+1} - q_{l_t-1}) t^{1/3}} \\ q_{l_t+1} - q_l \text{ w.p. } \frac{q_{l_t} - q_{l_t-1}}{(q_{l_t+1} - q_{l_t-1}) t^{1/3}} \\ 0 \text{ w.p. } 1 - t^{-1/3} \end{cases}$$

    set price $p_t \leftarrow p_{g,t} + \Delta p_t$

    observe demand $d_t = D_t(p_t)$

  end for

  set $\hat{b}_{t+1} \leftarrow \text{Proj}(\frac{\sum_{s=1}^t \Delta p_t d_t}{\sum_{s=1}^t \Delta p_t^2}, B)$

  set $(\hat{a}_{t+1}, \hat{\mathbf{c}}_{\mathbf{t+1}}) = \arg\min_{a', \mathbf{c'}} \sum_{s=1}^t (a' + \mathbf{c'}^\mathsf{T} \mathbf{x_s} - (d_s - \hat{b}_s p_s))^2 + \| \begin{bmatrix} a' \\ \mathbf{c'} \end{bmatrix} \|^2 + (a' + \mathbf{c'}^\mathsf{T} \mathbf{x_{t+1}})^2$

end for

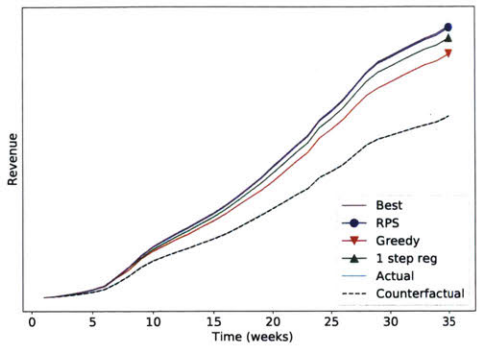
model, then add this expected demand to the prediction error of the random forest model, which we assume to be the demand noise.

Fig. B-1 gives the cumulative revenue, averaged over 100 iterations, of all three algorithms for each of the four subclasses. For reference, the actual revenues earned by the retailer as well as the projected revenue of the retailer in our estimated demand model, are also indicated, though of course, we cannot draw a fair comparison between the retailer's revenue and the revenue of RPS as the retailer's pricing scheme was subject to additional constraints that RPS did not take into account.

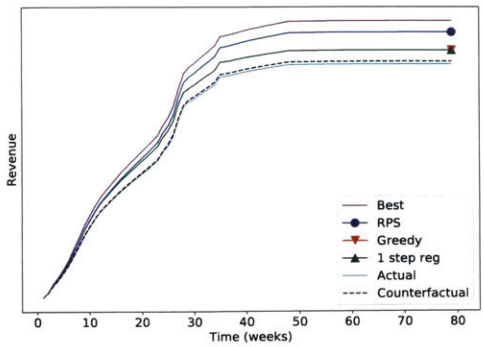
Comparing the revenue of RPS with those of the greedy and one-stage regression algorithms, however, we see that RPS clearly outperforms the other two algorithms. Table B.2 lists the summary statistics over 100 iterations of the cumulative revenue earned by RPS at the end of 35 weeks relative to those of the greedy and one-stage regression algorithms. The results show that the average revenue earned by RPS is between 7–20% higher than the average revenue earned by the greedy algorithm, and between 3–20% higher than the revenue earned by the one-stage regression algorithm. Further, the 95% confidence intervals in Table B.2 shows that RPS outperforms one-stage regression and the greedy algorithm with high probability.

The difference in revenues comes from the biased parameter estimates produced by the greedy and one-stage regression algorithms. Looking at Table B.1, we see that these two algorithms significantly underestimate the price sensitivity parameter b, while RPS alone estimates b accurately. This is consistent with our expectation that price endogeneity is present due to model misspecification: all the tested algorithms assume that demand is a linear function in features, while the true demand function is estimated from random forest, which can be highly nonlinear in features. Thus, our RPS algorithm successfully learns the demand elasticity even in the presence of endogeneity. In addition, we suspect that price endogeneity is also caused by the fact that the prices charged by our algorithms were restricted to within 20% of the historical prices charged by the retailer. These historical prices, as we have discussed in Section 3.4 of Chapter 3, are likely to be correlated with demand noise.

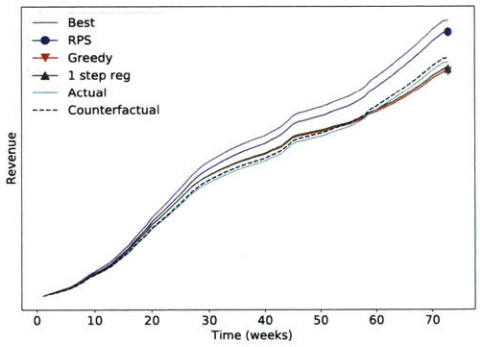
Finally, we compare the revenue of RPS to the best possible (clairvoyant) revenue



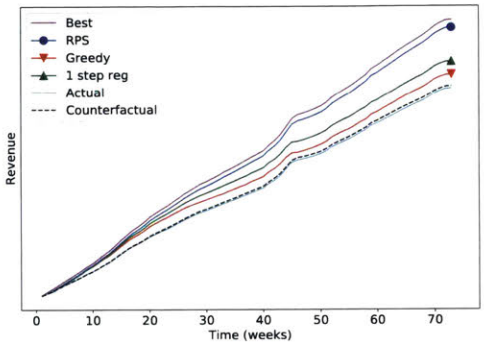
(a) Subclass 1



(b) Subclass 2



(c) Subclass 3



(d) Subclass 4

Figure B-1: Average revenue over 100 iterations of different algorithms

given full knowledge of the demand function, see Fig. B-1 and Table B.2. We find that the linear function estimated by the RPS algorithm in fact provides a good approximation for the true nonlinear demand function, as the revenue earned by RPS is very close to the clairvoyant revenue for all four subclasses. While this result might at first seem to be at odds with the demand prediction errors of 30-40% reported in Section 3.4 of Chapter 3, we note that demand prediction errors measure *absolute* differences between the estimated and actual demand. On the other hand, since demand is the sum of expected demand and a noise component, a price that deviates from the optimal linear price (which only optimizes revenue assuming a linear approximation of expected demand) may actually increase revenues relative to this optimal price.

Table B.1: Estimates of parameter b (with 95% confidence interval)

Subclass	True Value	RPS	Greedy	One-stage reg
1	-0.278	-0.279	-0.085	-0.119
		(-0.306,-0.254)	(-0.085,-0.085)	(-0.133,-0.107)
2	-0.280	-0.276	-0.100	-0.100
		(-0.369,-0.179)	(-0.100,-0.100)	(-0.101,-0.100)
3	-0.383	-0.377	-0.128	-0.122
		(-0.489,-0.233)	(-0.128,-0.128)	(-0.136,-0.100)
4	-0.383	-0.375	-0.149	-0.131
		(-0.461,-0.296)	(-0.153,-0.142)	(-0.131,-0.131)

Table B.2: Comparison of estimated revenues earned by various algorithms (with 95% confidence interval)

Subclass	RPS vs Greedy	RPS vs One-stage reg	RPS vs Clairvoyant
1	7.91%	3.32%	-0.85%
	(7.84%, 8.00%)	(2.50%, 4.62%)	(-0.92%, -0.78%)
2	6.98%	6.97%	-3.53%
	(3.03%, 8.33%)	(2.92%, 8.31%)	(-7.09%, -2.31%)
3	21.23%	20.30%	-3.18%
	(21.23%, 22.27%)	(17.62%, 21.32%)	(-4.74%, -2.35%)
4	21.04%	15.28%	-1.77%
	(19.19%, 21.59%)	(13.56%, 17.35%)	(-3.26%, -1.31%)

Appendix C

Appendix to Chapter 4 (Inventory Allocation with Demand Learning for Seasonal Consumer Goods)

C.1 Proofs for theoretical analysis

C.1.1 Proof of Lemma 1

Proof. Since we have assumed that the demands $D_{i,t}$ are discrete and bounded for all i, t, (P2) is a deterministic problem with finitely many constraints. These constraints are affine in $a_{i,t}$ and therefore convex. It remains to show that the objective function of (P2) is convex. We will show that this is true given the assumption $p_{i,t}+h_{i,t} \geq \alpha p_{i,t+1}$.

Given a demand realization $D_{i,t}$, define $\tilde{L}_{i,t}(a_{i,t}) := \alpha^{t-1} h_{i,t}[a_{i,t} + x_{i,t} - D_{i,t}]^+ + \alpha^{t-1} p_{i,t}[D_{i,t} - a_{i,t} - x_{i,t}]^+$ and define $\tilde{L}_{i,0}(a_{i,0}) := 0$ for each i. We will prove that $\sum_{t=1}^{T} \tilde{L}_{i,t}$ is convex, which is sufficient to prove the theorem. First, we will expand

the expression for $\tilde{L}_{i,t}(a_{i,t})$ as follows:

$$\begin{aligned}
\tilde{L}_{i,t}(a_{i,t}) &= \alpha^{t-1} h_{i,t}[a_{i,t} + x_{i,t} - D_{i,t}]^+ + \alpha^{t-1} p_{i,t}[D_{i,t} - a_{i,t} - x_{i,t}]^+ \\
&= \alpha^{t-1} h_{i,t} \max\{0, \sum_{s=u}^{t} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t\} \\
&+ \alpha^{t-1} p_{i,t} \max\{0, D_{i,t} - a_{i,t} - \max\{0, \sum_{s=u}^{t-1} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t-1\}\} \\
&= \alpha^{t-1} h_{i,t} \max\{0, \sum_{s=u}^{t} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t\} \\
&+ \alpha^{t-1} p_{i,t} \max\{D_{i,t} - a_{i,t}, \max\{0, \sum_{s=u}^{t-1} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t-1\}\} \quad \text{(C.1)} \\
&- \alpha^{t-1} p_{i,t} \max\{0, \sum_{s=u}^{t-1} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t-1\}. \quad \text{(C.2)}
\end{aligned}$$

Note that the sum of the two addends in (C.1) is convex, since the pointwise maximum of convex functions is convex, and $\alpha, h_{i,t}, p_t \ge 0$. We will now show by induction that

$$-\alpha^{t-1} p_{i,t} \max\{0, \sum_{s=u}^{t-1} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t-1\} + \sum_{s=1}^{t-1} \tilde{L}_{i,s}(a_{i,s}) \quad \text{(C.3)}$$

is convex. This will prove that $\sum_{t=1}^{T} \tilde{L}_{i,t}$, as the sum of convex functions, is convex in $\{a_{i,s}, s = 1, \ldots, t-1\}$. For $t = 0$, (C.3) is 0 and therefore clearly convex. Suppose now that (C.3) is convex for some $1 \le t < T$. We will show that it must be convex

for $t+1$ as well. By using the same expansion in (C.1) we have

$$
\begin{aligned}
&- \alpha^t p_{i,t+1} \max\{0, \sum_{s=u}^{t} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t\} + \sum_{s=1}^{t} \tilde{L}_{i,s}(a_{i,s}) \\
= &- \alpha^t p_{i,t+1} \max\{0, \sum_{s=u}^{t} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t\} \\
&+ \alpha^{t-1} h_{i,t} \max\{0, \sum_{s=u}^{t-1} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t\} \\
&+ \alpha^{t-1} p_{i,t} \max\{D_{i,t} - a_{i,t}, \max\{0, \sum_{s=u}^{t-1} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t-1\}\} \\
&- \alpha^{t-1} p_{i,t} \max\{0, \sum_{s=u}^{t-1} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t-1\}\} + \sum_{s=1}^{t-1} \tilde{L}_{i,s}(a_{i,s}).
\end{aligned}
$$

By the induction hypothesis, the term in the final line is convex. The sum of the remaining terms is equal to

$$
\begin{aligned}
&- \alpha^t p_{i,t+1} \max\{0, a_{i,t} - D_{i,t}, a_{i,t} - D_{i,t} + \sum_{s=u}^{t-1} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t\} \\
&+ \alpha^{t-1} h_{i,t} \max\{0, \sum_{s=u}^{t-1} a_{i,s} - D_{i,s}, 1 \le u \le t\} \\
&+ \alpha^{t-1} p_{i,t} \max\{0, D_{i,t} - a_{i,t}, \max\{0, \sum_{s=u}^{t-1} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t-1\} \\
= &\alpha^{t-1}(h_{i,t} - \alpha p_{i,t+1})(a_{i,t} - D_{i,t}) \\
&+ \alpha^{t-1}(h_{i,t} - \alpha p_{i,t+1}, +p_{i,t}) \max\{0, D_{i,t} - a_{i,t}, \max\{\sum_{s=u}^{t-1} a_{i,s} - D_{i,s},\ \forall 1 \le u \le t-1\}\}.
\end{aligned}
$$

The first addend in the final line is linear and therefore convex. The term $\max\{0, \dots, \}$ in the second addend is also convex as it is the pointwise maximum of convex functions. Finally, since we have assumed $h_{i,t} + p_{i,t} \ge \alpha p_{i,t+1}$, the second addend in the final line is convex, proving the induction hypothesis and completing the proof. $\square$

C.1.2 Proof of Theorem 5

Proof. Let LB denote the value of the dual (D1). Since our heuristic allocates inventory according to the solution of (D1), $\{a^*_{i,t}(\lambda^*)\}$, until the warehouse runs out of inventory, UB is only different from LB when $\sum_{i=1}^{N}\sum_{t=1}^{T} a_{i,t} > w_0$. If the warehouse runs out of inventory, the heuristic incurs additional lost sales costs not accounted for in the dual problem. Denoting $p_{\max} := \max\{p_{i,t}, i = 1,\ldots,N, t = 1,\ldots,T\}$, we have

$$\begin{aligned} UB &\leq LB + p_{\max}\mathrm{E}[[\sum_{i=1}^{N}\sum_{t=1}^{T} a^*_{i,t}(\lambda^*) - w_0]^+] \\ &\leq LB + p_{\max}\mathrm{E}[[\sum_{i=1}^{N}\sum_{t=1}^{T} a^*_{i,t}(\lambda^*) - \mathrm{E}[\sum_{i=1}^{N}\sum_{t=1}^{T} a^*_{i,t}(\lambda^*)]]^+] \\ &\leq LB + p_{\max}\mathrm{E}[|\sum_{i=1}^{N}\sum_{t=1}^{T} a^*_{i,t}(\lambda^*) - \mathrm{E}[\sum_{i=1}^{N}\sum_{t=1}^{T} a^*_{i,t}(\lambda^*)]|] \\ &\leq LB + p_{\max}\sqrt{\mathrm{Var}[\sum_{i=1}^{N}\sum_{t=1}^{T} a^*_{i,t}(\lambda^*)]}. \end{aligned} \tag{C.4}$$

The second line follows because we have strong duality between (P2) and its dual (D1), implying that the primal feasibility constraint

$$\mathrm{E}[\sum_{t=1}^{T}\sum_{i=1}^{N} a^*_{i,t}(\lambda^*)] \leq w_0$$

is satisfied for the optimal allocation policy $\{a^*_{i,t}(\lambda^*), i = 1,\ldots,N, t = 1,\ldots,T\}$. The final line follows from Jensen's inequality, which applies because of the convexity of the square function.

We will now bound $\mathrm{Var}[\sum_{i=1}^{N}\sum_{t=1}^{T} a^*_{i,t}(\lambda^*)]$. We will do so using a different argument from Marklund and Rosling (2012), since the latter assumes that demand is independent from time period to time period. We show that even without this assumption, a bound of the same order in terms of N can be achieved, and that our

bound is in fact an improvement in terms of T. We have

$$\begin{aligned}\mathrm{Var}[\sum_{i=1}^{N}\sum_{t=1}^{T} a^*_{i,t}(\lambda^*)] &= \sum_{i=1}^{N} \mathrm{Var}[\sum_{t=1}^{T} a^*_{i,t}(\lambda^*)] \\ &\leq \sum_{i=1}^{N} \frac{1}{4} D^2_{\max} T^2 \\ &= D^2_{\max} N T^2 \,.\end{aligned}$$

The equality in the first line follows from the assumption that the retailers experience independent demands. Since the allocations $a^*_{i,t}$ are solutions to a decoupled optimization problem, they are independent. The inequality in the second line follows from Popoviciu's inequality on variances, using the assumption that $D_{i,t}$ is known to be upper bounded by $D_{\max}$ for each i, t. Then the optimal $a^*_{i,t}$ must satisfy $0 \leq a^*_{i,t}(\lambda^*) \leq D_{\max}$.

The second addend in (C.4) is thus $O(\sqrt{N})$, which proves the theorem.

□

C.1.3 Proof of Theorem 6

Proof. Given dual variable λ, the optimal time 2 allocation to each retailer is the solution to the minimization problem

$$\min_{a_{i,2} \text{ s.t. } a_{i,2} \geq 0} \lambda a_{i,2} + p\mathrm{E}[[D_{i,1} + \rho\epsilon_i - a_{i,2}]^+] + h\mathrm{E}[[a_{i,2} - (D_{i,2} + \rho\epsilon_i)]^+] \,.$$

It is easy to see that this is a convex optimization problem, and that the optimal $a^*_{i,2}$ is given by $a^*_{i,2} = D_{i,1} + \rho F^{-1}_{\epsilon}(\frac{p-\lambda}{p+h})$, where $F^{-1}_{\epsilon_i}$ represents the inverse CDF of the demand noise ϵ_i. This expression is independent of the first period allocation decision $a_{i,1}$. The optimal time 1 allocation to each retailer is then the solution to the minimization problem

$$\min_{a_{i,1} \text{ s.t. } a_{i,1} \geq 0} \lambda a_{i,1} + p\mathrm{E}[[D_{i,1} - a_{i,1}]^+] + h\mathrm{E}[[a_{i,1} - D_{i,1}]^+] \,.$$

Again, this is clearly a convex optimization problem, and the optimal $a^*_{i,1}$ is given by the newsvendor level $a^*_{i,1} = F^{-1}_{D_{i,1}}(\frac{p-\lambda}{p+h})$, where $F^{-1}_{D_{i,1}}$ represents the inverse CDF of the first period demand at retailer i.

For a given λ, the total expected allocation across retailers and time periods is thus given by

$$\sum_{i=1}^{N} F^{-1}_{D_{i,1}}(\frac{p-\lambda}{p+h}) + \mathrm{E}[D_{i,1}] + \rho F^{-1}_{\epsilon}(\frac{p-\lambda}{p+h}) \ ,$$

and for each ρ, the corresponding dual variable $\lambda^*(\rho)$ that solves (D3) must satisfy

$$F^{-1}_{D_{i,1}}(\frac{p-\lambda}{p+h}) + \mathrm{E}[D_{i,1}] + \rho F^{-1}_{\epsilon}(\frac{p-\lambda}{p+h}) = \frac{w_0}{N} \ ,$$

by complementary slackness, and the fact that we have assumed that the retailers are identical. Now set $w_{\max} = N\mathrm{E}[D_{i,1}]$. For $w_0 \leq w_{\max}$, since $D_{i,1} > 0$, $\lambda^*(\rho)$ must be sufficiently large for all ρ such that

$$F^{-1}_{\epsilon}(\frac{p-\lambda^*(\rho)}{p+h}) < 0 \ .$$

(Since ϵ_i is symmetrical about its mean, 0, this implies that $\lambda^*(\rho)$ must be sufficiently large that $(p-\lambda^*(\rho))/(p+h) < 1/2$). Now for fixed ρ, suppose we increase ρ by some $\Delta\rho$. For any λ such that $\lambda \geq \lambda^*(\rho)$

$$0 \leq F^{-1}_{D_{i,1}}(\frac{p-\lambda}{p+h}) \leq F^{-1}_{D_{i,1}}(\frac{p-\lambda^*(\rho)}{p+h}) \ ,$$

and

$$0 > \rho F^{-1}_{\epsilon}(\frac{p-\lambda^*(\rho)}{p+h}) > (\rho+\Delta\rho) F^{-1}_{\epsilon}(\frac{p-\lambda}{p+h}) \ ,$$

where both inequalities follow from the fact that when $\lambda \geq \lambda^*(\rho)$, $(p-\lambda)/(p+h) < 1/2$. Thus

$$F^{-1}_{D_{i,1}}(\frac{p-\lambda}{p+h}) + \mathrm{E}[D_{i,1}] + (\rho+\Delta\rho) F^{-1}_{\epsilon}(\frac{p-\lambda}{p+h}) \neq \frac{w_0}{N} \ ,$$

and we must have $\lambda^*(\rho) > \lambda^*(\rho+\Delta\rho)$. Since the optimal first period allocation $a^*_{i,1}(\rho)$ is given by $a^*_{i,1}(\rho) = F^{-1}_{D_{i,1}}(\frac{p-\lambda^*(\rho)}{p+h})$, it is strictly decreasing in λ, and hence strictly

increasing in ρ. □

C.1.4 Proof of Theorem 7

Proof. For a given starting warehouse inventory level $w_{0,1}$, let the associated dual variable be λ. Each retailer i's subproblem D2 gives the following explicit and implicit forms for the optimal allocation to retailer i in each period:

In the second period, we must solve the optimization problem

$$\min_{a_{i,2}, a_{i,2} \geq 0} \lambda a_{i,2} + p\mathrm{E}[[D_{i,2} - a_{i,2} - [a_{i,1} - D_{i,1}]^+]^+] + h\mathrm{E}[[a_{i,2} + [a_{i,1} - D_{i,1}]^+ - D_{i,2}]^+] .$$

If we equivalently write it as

$$\begin{aligned} \min_{a_{i,2}, a_{i,2} \geq 0} \lambda a_{i,2} + p\mathrm{E}[[D_{i,1} - a_{i,1}]^+] + p\mathrm{E}[[D_{i,2} - a_{i,2} - [a_{i,1} - D_{i,1}]^+]^+] \\ + h\mathrm{E}[[a_{i,2} + [a_{i,1} - D_{i,1}]^+ - D_{i,2}]^+] , \end{aligned} \tag{C.5}$$

we see that this reformulated optimization problem is jointly convex in $a_{i,1}$ and $a_{i,2}$ as follows: the first term in the summand is linear in $a_{i,2}$, the final term in the summand (associated with holding costs) is convex since it is the composition of a convex increasing function and a convex function, and the sum of the second and third terms (associated with lost sales) can be expanded as

$$\begin{aligned} V_2(a_{i,1}, D_{i,1}) =& p\mathrm{E}[[D_{i,1} - a_{i,1}]^+] + p\mathrm{E}[[D_{i,2} - a_{i,2} - [a_{i,1} - D_{i,1}]^+]^+] \\ =& p\mathrm{E}[[D_{i,1} - a_{i,1}]^+] + p\mathrm{E}[\max\{D_{i,2} - a_{i,2}, [a_{i,1} - D_{i,1}]^+\}] \\ & - p\mathrm{E}[[a_{i,1} - D_{i,1}]^+] \\ =& p\mathrm{E}[\max\{D_{i,2} - a_{i,2}, a_{i,1} - D_{i,1}, 0\}] - (D_{i,1} - a_{i,1}) . \end{aligned}$$

This is the sum of the pointwise maximum of affine functions, which is convex, and an affine function, which is also convex. Thus the reformulated optimization problem (C.6) is jointly convex in $a_{i,1}$ and $a_{i,2}$.

We can then differentiate the objective function with respect to $a_{i,2}$, and get the

first order condition

$$\begin{aligned}(h+\lambda)\mathbb{P}[D_{i,2}-[a_{i,1}-D_{i,1}]^+\leq a_{i,2}] =&(p-\lambda)\mathbb{P}[D_{i,2}-[a_{i,1}-D_{i,1}]^+> a_{i,2}]\\ a^*_{i,2} =&\max\{F_i^{-1}(\frac{p-\lambda}{p+h}),[a_{i,1}-D_{i,1}]^+\}\\ &-[a_{i,1}-D_{i,1}]^+. \end{aligned} \tag{C.6}$$

Here F_i denotes the CDF of retailer i's demand, and F_i^{-1} denotes the inverse CDF of this demand. Then by the convexity of the period 2 allocation problem in $a_{i,2}$ (convexity in $a_{i,1}$ will be used later), the period 2 optimal order up to level is thus $F_i^{-1}((p-\lambda)/(p+h))$.

The first period allocation is then the solution to the optimization problem

$$\min_{a_{i,1}} \lambda a_{i,1}+h\mathrm{E}[[a_{i,1}-D_{i,1}]^+]+V_2(a_{i,1},D_{i,1})\,.$$

Using the well known property that $\pi(x)=\min_{g(y,x)\leq 0} f(y,x)$ is convex given that f,g are jointly convex in x,y, we know that V_2 is convex in $a_{i,1}$. Then the objective function of the first period optimization problem, as the sum of convex functions, is convex in $a_{i,1}$, and any solution of the first order conditions will be a global optimum. Using the expression for the period 2 optimal order up to level, $F_i^{-1}((p-\lambda)/(p+h))$, we can expand V_2 and write

$$\begin{aligned}\min_{a_{i,1}}\ &\lambda a_{i,1}+p\mathrm{E}[[D_{i,1}-a_{i,1}]^+]+h\mathrm{E}[[a_{i,1}-D_{i,1}]^+]+\lambda\mathrm{E}[[F_i^{-1}(\frac{p-\lambda}{p+h})-[a_{i,1}-D_{i,1}]^+]^+]\\ &+p\mathrm{E}[[D_{i,2}-\max\{F_i^{-1}(\frac{p-\lambda}{p+h}),[a_{i,1}-D_{i,1}]^+\}]^+]\\ &+h\mathrm{E}[[\max\{F_i^{-1}(\frac{p-\lambda}{p+h}),[a_{i,1}-D_{i,1}]^+\}-D_{i,2}]^+]\,,\end{aligned}$$

which is equivalently

$$\min_{a_{i,1}} \lambda a_{i,1} + p\mathrm{E}[[D_{i,1} - a_{i,1}]^+] + h\mathrm{E}[[a_{i,1} - D_{i,1}]^+] + \lambda\mathrm{E}[D_{i,2} - [a_{i,1} - D_{i,1}]^+]$$
$$+ (p-\lambda)\mathrm{E}[[D_{i,2} - \max\{F_i^{-1}(\frac{p-\lambda}{p+h}), [a_{i,1} - D_{i,1}]^+\}]^+]$$
$$+ (h+\lambda)\mathrm{E}[[\max\{F_i^{-1}(\frac{p-\lambda}{p+h}), [a_{i,1} - D_{i,1}]^+\} - D_{i,2}]^+]\,. \tag{C.7}$$

Now suppose $w_{\min}$ is the minimum inventory level such that for $w_{0,1} \geq w_{\min}$, $\lambda = 0$. Such an $w_{\min}$ exists since demand at all the retailers is almost surely upper bounded by the parameter b. Then if $\lambda = 0$, the second period order-up-to level is $F_i^{-1}(p/(p+h))$. Since we have assumed that $p > h$, we have $p/(p+h) < 1/2$. Then, since demand follows the truncated normal distribution, for $i > j$, we have $\mathbb{P}[D_{i,t} > d] > \mathbb{P}[D_{j,t} > d]$ when $d > \mu$, implying that the second period order-up-to level is increasing in i.

As for the optimal first period allocation when $\lambda = 0$, suppose we know that $a_{i,1}^* \leq F_i^{-1}(p/(p+h))$. Then, differentiating the objective function of (C.7) evaluated on any $a_{i,1} \leq F_i^{-1}(p/(p+h))$ gives

$$\lambda - p\mathbb{P}[D_{0,i} > a_{0,i}] + h\mathbb{P}[D_{0,i} \leq a_{0,i}]\,.$$

Then the left derivative of (C.7) is 0 when $a_{0,i}^* = F_i^{-1}(p/(p+h))$. By the differentiability of (C.7), the derivative of (C.7) is also 0 when $a_{0,i}^* = F_i^{-1}(p/(p+h))$, and by the convexity of (C.7), the optimal first period allocation is $F_i^{-1}(p/(p+h))$, i.e. the same as the second period order-up-to level. This implies that the optimal first period allocation is also increasing in i.

To analyze the case when the starting warehouse inventory $w_{0,1}$ is small, and λ is large, we will construct $w_{\max}$ by noting that demand is truncated normal, which implies that there exists some $t_{\max}$, $0 < t_{\max} < 1/2$ such that the demand probability density functions $f_i(t)$ is increasing in i for all $t \leq t_{\max}$. Then set $w_{\max}$ as the starting warehouse inventory that corresponds to dual variable $\lambda = p - (p+h)F_i(t_{\max}/2)$. For $w_{0,1}$, $w_{0,1} \leq w_{\max}$, the associated dual variable λ satisfies $\lambda \geq p - (p+h)F_i(t_{\max}/2)$.

For such $w_{0,1}$ that satisfies $w_{0,1} \leq w_{\max}$, since the associated λ satisfies $\lambda \geq p-(p+$

$h)F_i(t_{\max}/2) \geq (p-h)/2$, we have $(p-\lambda)/(p+h) \leq 1/2$ and thus, $F_i^{-1}((p-\lambda)/(p+h))$ is strictly decreasing in i by our assumption that demand is truncated normal, which implies that for $i > j$, $\mathbb{P}[D_{i,t} > d] < \mathbb{P}[D_{j,t}] > d]$ when $d < \mu$.

We now complete the proof by showing that the optimal first period allocation $a_{i,1}$ is also decreasing in i for $w_{0,1}$ that satisfies $w_{0,1} \leq w_{\max}$. We do so by writing the first order conditions of (C.7):

$$\begin{aligned} p - \lambda =& (p + h - \lambda)\mathbb{P}[D_{i,1} \leq a_{i,1}] \\ &+ (p - \lambda)\mathbb{P}[D_{i,1} \leq a_{i,1} - F_i^{-1}(\frac{p-\lambda}{p+h})] \\ &+ (h - p + 2\lambda)\mathbb{P}[D_{i,1} + D_{2,i} \leq a_{i,1} \mid D_{i,1} \leq a_{i,1}] \\ &- F_i^{-1}(\frac{p-\lambda}{p+h})\mathbb{P}[D_{i,1} \leq a_{i,1} - F_i^{-1}(\frac{p-\lambda}{p+h})]\,. \end{aligned} \tag{C.8}$$

We claim that $2F_i^{-1}((p-\lambda)/(p+h))$ is an upper bound on the optimal $a_{i,1}$ when $w_{0,i} \leq w_{\max}$. Suppose instead that $a_{i,1} > 2F_i^{-1}((p-\lambda)/(p+h))$. Then the right hand side of (C.8) is at least

$$\begin{aligned} &(p + h - \lambda)\mathbb{P}[D_{i,1} \leq 2F_i^{-1}(\frac{p-\lambda}{p+h})] \\ &+ (p - \lambda)\mathbb{P}[D_{i,1} \leq F_i^{-1}(\frac{p-\lambda}{p+h})] + (h - p + 2\lambda)\mathbb{P}[D_{i,1} \leq F_i^{-1}(\frac{p-\lambda}{p+h})] \\ \leq& 2(p + h - \lambda)\mathbb{P}[D_{i,1} \leq F_i^{-1}(\frac{p-\lambda}{p+h})] + (p - \lambda)\mathbb{P}[D_{i,1} \leq F_i^{-1}(\frac{p-\lambda}{p+h})] \\ &+ (h - p + 2\lambda)\mathbb{P}[D_{i,1} \leq F_i^{-1}(\frac{p-\lambda}{p+h})] \\ =& \mathbb{P}[D_{i,1} \leq F_i^{-1}(\frac{p-\lambda}{p+h})](2p + 3h - \lambda) \\ =& \frac{p-\lambda}{p+h}(p + h + p + 2h - \lambda) \\ >& p - \lambda. \end{aligned}$$

The last line follows from the fact that $w_{0,i} \leq w_{\max}$ implies that the associated dual variable λ satisfies $\lambda > (p-h)/2$, and the third line follows from our assumption

of truncated normal demand that is symmetrical about the mean, which implies that

$$\begin{aligned}\mathbb{P}[D_{i,1} \leq 2F_i^{-1}(\frac{p-\lambda}{p+h})] =& \mathbb{P}[D_{i,1} \leq F_i^{-1}(\frac{p-\lambda}{p+h})] \\ &+ \mathbb{P}[F_i^{-1}(\frac{p-\lambda}{p+h}) \leq D_{i,1} \leq 2F_i^{-1}(\frac{p-\lambda}{p+h})] \\ \leq& 2\mathbb{P}[D_{i,1} \leq F_i^{-1}(\frac{p-\lambda}{p+h})],\end{aligned}$$

when $(p-\lambda)/(p+h) < 1/2$. Thus the right hand side of (C.8) is strictly greater than the left hand side, leading to a contradiction.

For each $a_{i,1}$ such that $a_{i,1} < 2F_i^{-1}((p-\lambda)/(p+h))$, we will show that each of the summands on the right hand side of (C.8) is strictly increasing in both $a_{i,1}$ and i. This will imply that the heuristic first period allocations $a_{i,1}^*$ are strictly decreasing in i, proving the theorem.

Consider the first summand of (C.8). Since $\lambda \leq p$, this term is strictly increasing in $a_{i,1}$. Further, since $a_{i,1} \leq 2F_i^{-1}((p-\lambda)/(p+h)) \leq t_{\max}$, by our definition of $t_{\max}$, the summand is strictly increasing in $a_{i,1}$.

Now consider the second summand of (C.8). Again, since $\lambda \leq p$, this term is increasing in $a_{i,1}$. And since $a_{i,1} - F_i^{-1}((p-\lambda)/(p+h)) \leq 2F_i^{-1}((p-\lambda)/(p+h)) \leq t_{\max}$, this summand is also increasing in $a_{i,1}$.

Finally, consider the third summand of (C.8). We have set $w_{0,i}$ such that the associated dual variable λ satisfies $\lambda > (p-h)/2$. Thus the coefficient of the probability term, $h-p+2\lambda$, is positive, and the summand is increasing in $a_{i,1}$. We can write the probability term as

$$\begin{aligned}&\mathbb{P}[D_{i,1} + D_{2,i} \leq a_{i,1} \mid D_{i,1} \leq a_{i,1} - F_i^{-1}(\frac{p-\lambda}{p+h})]\mathbb{P}[D_{i,1} \leq a_{i,1} - F_i^{-1}(\frac{p-\lambda}{p+h})] \\ &= \int_a^{a_{i,1}} \int_a^{s-a} f_i(t) f_i(s-t) dt ds.\end{aligned}$$

For t such that $a \leq t \leq s-a$, and s such that $s \leq a_{i,1}$, we also have $a \leq t \leq a_{i,1} \leq 2F_i^{-1}((p-\lambda)/(p+h)) \leq t_{\max}$. Thus $f_i(t)$ is increasing in i for each t within the bounds of the integral. Similarly, for each t and s within the bounds of the integral,

we have $a \leq s - t \leq s - a$. By the same argument, $f_i(s - t)$ is increasing in i. Thus the probability term in the third summand of (C.8) is increasing in i, proving the claim and the theorem. $\square$

Bibliography

Adelman, D. and Mersereau, A. J. (2008). Relaxations of weakly coupled stochastic dynamic programs. *Operations Research*, 56(3):712–727.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Azoury, K. S. (1988). *Bayesian Policies for Dynamic Inventory Models*. PhD thesis, UCLA.

Azoury, K. S. and Miller, B. L. (1984). A comparison of the optimal ordering levels of bayesian and non-bayesian inventory models. *Operations Research*, 30(8):993–1003.

Azoury, K. S. and Warmuth, M. K. (2001). Relative loss bounds for on-line density estimation with the exponential family of distributions. *Machine Learning*, 43(3):211–246.

Badanidiyuru, A., Langford, J., and Slivkins, A. (2014). Resourceful contextual bandits. In *Proceedings of The 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pages 1109–1134. PMLR.

Ban, G.-Y. and Keskin, N. B. (2017). Personalized dynamic pricing with machine learning. available at https://ssrn.com/abstract=2972985.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

Bertsimas, D. and Kallus, N. (2016). Pricing from observational data. *arXiv preprint arXiv:1605.02347*.

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

Besbes, O. and Zeevi, A. (2012). Blind network revenue management. *Operations Research*, 60(6):1537–1550.

Besbes, O. and Zeevi, A. (2015). On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Science*, 61(4):723–739.

Bijmolt, T. H., Heerde, H. J. v., and Pieters, R. G. (2005). New empirical generalizations on the determinants of price elasticity. *Journal of marketing research*, 42(2):141–156.

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

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

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

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, Learning, and Games*. Cambridge University Press.

Chen, X., Owen, Z., Pixton, C., and Simchi-Levi, D. (2015). A statistical learning approach to personalization in revenue management. Available at SSRN: https://ssrn.com/abstract=2579462.

Clark, A. J. and Scarf, H. (1960). Optimal policies for a multi-echelon inventory problem. *Management Science*, 6(4):475–490.

Cohen, M. C., Lobel, I., and Paes Leme, R. (2016). Feature-based dynamic pricing. *Available at SSRN*.

Cooper, W. L., Homem-de Mello, T., and Kleywegt, A. J. (2006). Models of the spiral-down effect in revenue management. *Operations Research*, 54(5):968–987.

Cooper, W. L., Homem-de Mello, T., and Kleywegt, A. J. (2015). Learning and pricing with models that do not explicitly incorporate competition. *Operations research*, 63(1):86–103.

Dana Jr., J. D. and Petruzzi, N. C. (2001). Note: The newsvendor model with endogenous demand. *Management Science*, 47(11):1488–1497.

den Boer, A. V. (2015). Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science*, 20(1):1–18.

den Boer, A. V. and Zwart, B. (2013). Simultaneously learning and optimizing using controlled variance pricing. *Management Science*, 60(3):770–783.

Ding, X., Puterman, M., and Bisi, A. (2002). The censored newsvendor and optimal acquisition of information. *Operations Research*, 50(3):517–527.

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

Eppen, G. and Schrage, L. (1981). Centralized ordering policies in a multiwarehouse system with lead times and random demand. *SCHWARZ*, pages 51–69.

Federgruen, A. and Zipkin, P. (1984). Approximations of dynamic, multilocation production and inventory problems. *Management Science*, 30(1):69–84.

Ferreira, K., David, S.-L., and Wang, H. (2017). Online network revenue management using thompson sampling. *Operations Research (forthcoming).*

Ferreira, K. J., Lee, B. H. A., and Simchi-Levi, D. (2015). Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1):69–88.

Fisher, M., Gallino, S., and Li, J. (2017). Competition-based dynamic pricing in online retailing: A methodology validated with field experiments. *Management Science (forthcoming).*

Fisher, M., Gallino, S., and Li, J. (2018). Competition-based dynamic pricing in online retailing: A methodology validated with field experiments. *Management Science*, 64(6):2473–2972.

Fisher, M. and Rajaram, K. (2000). Accurate retail testing of fashion merchandise: Methodology and application. *Management Science*, 19(3):266–278.

Fisher, M. and Raman, A. (1996). Reducing the cost of demand uncertainty through accurate response to early sales. *Operations Research*, 44(1):87–99.

Gallego, G. and van Ryzin, G. (1994). Optimal dynamic pricing of inventories with stochastic demands over finite horizons. *Management Science*, 40(8):947–1068.

Gallien, J., Mersereau, A. J., Garro, A., Mora, A. D., and Vidal, M. N. (2017). Initial shipment decisions for new products at zara. *Operations Research*, 63(2):269–286.

Gill, R. D. and Levit, B. Y. (1995). Applications of the van trees inequality: a bayesian cramér-rao bound. *Bernoulli*, pages 59–79.

Greene, W. H. (2003). *Econometric analysis (5th edition).* Pearson.

Hsu, D., Kakade, S. M., and Zhang, T. (2014). Random design analysis of ridge regression. *Foundations of Computational Mathematics*, 14(3):569–600.

Jackson, P. L. (1988). Stock allocation in a two-echelon distribution system or "what to do until your ship comes in". *Management Science*, 34(7):1–17.

Jackson, P. L. and Muckstadt, J. A. (1989). Risk pooling in a two-period, two-echelon inventory stocking and allocation problem. *Naval Research Logistics*, 36:1–26.

Javanmard, A. and Nazerzadeh, H. (2016). Dynamic pricing in high-dimensions. *arXiv preprint arXiv:1609.07574.*

Keskin, N. B. and Zeevi, A. (2014). Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research*, 62(5):1142–1167.

Kuhlmann, R. (2004). Why is revenue management not working? *Journal of Revenue and Pricing Management*, 2(4):378.

Li, J., Granados, N., and Netessine, S. (2014). Are consumers strategic? structural estimation from the air-travel industry. *Management Science*, 60(9):2114–2137.

Li, J., Netessine, S., and Koulayev, S. (2016). Price to compete... with many: How to identify price competition in high dimensional space. available at https://ssrn.com/abstract=2651045.

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

Marklund, J. M. and Rosling, K. (2012). Lower bounds and heuristics for supply chain stock allocation. *Operations Research*, 60(1):iii–248.

McGavin, E. J., Schwarz, L. B., and Ward, J. E. (1993). Two-interval inventory-allocation policies in a one-warehouse n-identical-retailer distribution system. *Management Science*, 39(9):1092–1107.

McGavin, E. J., Schwarz, L. B., and Ward, J. E. (1997). Balancing retailer inventories. *Operations Research*, 45(6):780–989.

Petrin, A. and Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47(1):370–379.

Phillips, R., Şimşek, A. S., and Van Ryzin, G. (2015). The effectiveness of field price discretion: Empirical evidence from auto lending. *Management Science*, 61(8):1741–1759.

Qiang, S. and Bayati, M. (2016). Dynamic pricing with demand covariates. *Available at SSRN 2765257*.

Smith, S. A. and Achabal, D. D. (1998). Clearance pricing and inventory policies for retail chains. *Management Science*, 44(3):285–300.

Talluri, K. T. and Van Ryzin, G. J. (2005). *The theory and practice of revenue management*. Springer.

Van Ryzin, G. and McGill, J. (2000). Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels. *Management Science*, 46(6):760–775.

Veeraraghavan, S. and Vaidyanathan, R. (2011). Measuring seat value in stadiums and theaters. *Production and Operations Management*, 21(1):49–68.

Vulcano, G., van Ryzin, G., and Chaar, W. (2010). Choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management*, 12(3):371–392.

Wang, Z., Deng, S., and Ye, Y. (2011). Close the gaps: A learning-while-doing algorithm for a class of single-product revenue management problems. *CoRR*, abs/1101.4681.

Wang, Z., Deng, S., and Ye, Y. (2014). Close the gaps: A learning-while-doing algorithm for single-product revenue management problems. *Operations Research*, 62(2):318–331.