Dynamic Learning and Pricing with Model Misspecification

Dynamic Learning and Pricing with Model Misspecification

Mila Nambiar
Operations Research Center, Massachusetts Institute of Technology mnambiar@mit.edu

David Simchi-Levi
Institute for Data, Systems, and Society, Department of Civil and Environmental Engineering, and Operations Research Center, Massachusetts Institute of Technology, dslevi@mit.edu

He Wang
School of Industrial and Systems Engineering, Georgia Institute of Technology, he.wang@isye.gatech.edu

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 offline simulations gauging the performance of RPS on a large fashion retail dataset, and find that is expected to earn 8–20% more revenue on average than competing algorithms that do not account for price endogeneity.

Key words: revenue management; pricing; parameter estimation; endogeneity; model misspecification; fashion retail

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 paper, 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 often 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. 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 revenue performance within the assumed model family (see Proposition 1). Somewhat surprisingly, we find that this is not always true, as the model parameters may converge to another estimate (denoted by a circle dot) with worse revenue performance. Sometimes, the updated model parameter (circle dot) may even have worse revenue performance than the initial model estimation (white "x")!

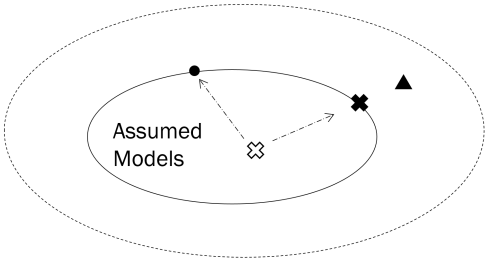


Figure 1 The dynamics of parameter estimates under model misspecification.

The reason why the estimated parameters are inconsistent is because 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 paper, 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 the scope of this paper.

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 to get an unbiased and consistent estimate of the model parameter 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. IV method is a widely used econometric method to obtain unbiased estimates of coefficients of endogenous variables. The IV method 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 the IV method 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 endogenous heavily relies on the fact that those price perturbations are randomized.

In Section 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.

We test the numerical performance of the RPS algorithm using synthetic data in Section 4. The experiments demonstrated how the RPS algorithm obtains unbiased estimation in the presence of price endogeneity. We also compared the RPS algorithm with other pricing algorithms proposed in the literature.

In collaboration with Oracle Retail, we performed simulation experiments to gauge the performance of the RPS algorithm in a real-world setting. These experiments were performed on a dataset provided by Oracle Retail, consisting of three years' worth of data on customer transactions and product feature information at an anonymous chain of brick-and-mortar department stores. Based on this dataset, we built a "ground truth" demand model. Then, using the ground truth model as

a stand-in for the true demand, we simulated the performance of the RPS algorithm, allowing it to price the items in the historical dataset based on their features. The procedures we used to build the ground truth model and the results of our experiments are reported in Section 4.2.

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. 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, our paper considers 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.

Our paper considers 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).

Our paper is particularly focused on a dynamic learning and pricing problem that contains contextual information (i.e. features). Here we compare our paper with 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, our paper studies 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 closer to ours 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) made 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 approaches to pricing with features in a batch learning setting where historical data are given as input (Chen et al. 2015, Bertsimas and Kallus 2016). Our paper differs from these works 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 paper, 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, in this paper 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 the paper are understood to be column vectors. For any vector $x \in \mathbb{R}^k$, we denote its transpose by $x^\intercal$ and denote its Euclidean norm by $\|x\| := \sqrt{x^\intercal x}$. We let $\|x\|_1$ be the ℓ_1 norm of x, defined as $\|x\|_1 = \sum_i |x_t|$. We let $\|x\|_\infty$ be the ℓ_∞ norm, defined as $\|x\|_\infty = \max_i |x_t|$. For any square matrix $M \in \mathbb{R}^{k\times k}$, we denote its transpose by $M^\intercal$, its inverse by M^{-1} and its trace by $\mathrm{tr}(M)$; if M is also symmetric $(M = M^\intercal)$, 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 $\|M\|_2 = \sqrt{\lambda_{\max}(M^\intercal M)}$. We denote the Frobenius norm of M by $\|M\|_F$, namely $\|M\|_F = \sqrt{\mathrm{tr}(M^\intercal M)}$.

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, $x_t \in \mathbb{R}^m$, which represents exogenous information that may affect demand in the current period. We assume that feature vectors x_t are sampled independently for $t = 1, 2, \ldots, T$ from a fixed but unknown distribution with bounded support. (In Section 3.5, we will relax the IID assumption of x_t and assume an arbitrary sequence of random feature vectors.) Without loss of generality, we assume $x_t \in [-1, 1]^m$ after appropriate scaling. Moreover, we assume that the matrix

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

is positive definite.[1]

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

$$D_t(p) = bp + f(x_t) + \epsilon_t, \quad \forall p \in [\underline{p}_t, \bar{p}_t]. \tag{1}$$

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 $f(x_t)$ is bounded for all x_t such that $|f(x_t)| \le \bar{f}$. The latter assumption would follow immediately from the fact that the set of all features x_t is compact if f were continuous. The last term ϵ_t in Eq (1) represents a demand noise. Without loss of generality, we assume ϵ_t has zero mean conditional of x_t: $\mathrm{E}[\epsilon_t \mid x_t] = 0$; otherwise, the conditional mean $\mathrm{E}[\epsilon_t \mid x_t]$ can be shifted into function $f(x_t)$. We assume that ϵ_t has bounded second moment ($\mathrm{E}[\epsilon_t^2] \le \sigma^2, \forall t$), and is independent of historical data (x_j, ϵ_j) for all $1 \le j \le t-1$. However, the distribution of ϵ_t is allowed to vary over time. We refer to Eq (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 x_t and demand noise ϵ_t. We also assume there exist constants $\delta > 0$ and $p_{\max}$ such that $\bar{p}_t \le p_{\max}$ and $\bar{p}_t - \underline{p}_t \ge \delta$ for all t. Given features x_t, we denote the optimal price for the true demand model (as a function of x_t) by $\tilde{p}_t(x_t) = -\frac{f(x_t)}{2b}$. We assume that the optimal price $\tilde{p}_t(x_t) \in [\underline{p}_t, \bar{p}_t]$ 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.

2.1. 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 Section 4.2. 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.

As another example, feature-based pricing is also used for personalized financial services. Phillips et al. (2015) described 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 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. 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 (1). The seller assumes a linear demand model as

$$D_t(p) = a + bp + c^\intercal x_t + \nu_t, \quad \forall p \in [\underline{p}_t, \bar{p}_t], \tag{2}$$

where $a \in \mathbb{R}$ and $c \in \mathbb{R}^m$ 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 c are bounded, and that there exist $\bar{a}$, $\bar{c}$ such that $|a| \leq \bar{a}$ and $\|c\|_1 \leq \bar{c}$, but not necessarily 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 Section 4.2, 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, 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 (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.

2.3. Price Endogeneity Caused by Model Misspecification and Other Factors

By comparing the true quasi-linear demand model (cf. Eq (1)) and the misspecified linear model (cf. Eq (2)), it is easily verified that the error term in the misspecified linear model is equal to $\nu_t = f(x_t) - (a + c^\top x_t) + \epsilon_t$. This error term ν_t is composed of an approximation error, $f(x_t) - (a + c^\top x_t)$, which is correlated with features x_t, and a random noise ϵ_t, which is uncorrelated with features.

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

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

$$p_t = \pi(\mathcal{H}_{t-1}, x_t, \underline{p}_t, \bar{p}_t). \tag{3}$$

Because the error term ν_t is correlated with features x_t while price p_t is a function of 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 $\mathrm{E}[\nu_t p_t \mid x_t] = \mathrm{E}[\nu_t \cdot \pi(\mathcal{H}_{t-1}, x_t, \underline{p}_t, \bar{p}_t) \mid x_t] \neq 0$. 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 the paper, 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

$$d_t = a + bp_t + c^\top x_t + \nu_t, \tag{4}$$

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, 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, 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 Section 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 paper 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 paper 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.

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 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 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 x to a set S by $\text{Proj}(x, S) = \arg\min_{x' \in S} \|x - x'\|$.) 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 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

Algorithm 1 Random Price Shock (RPS) algorithm.

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

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

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

set $\delta_t \leftarrow \frac{\delta}{2} t^{-\frac{1}{4}}$

given x_t, set unconstrained greedy price: $p^u_{g,t} \leftarrow -\frac{\hat{a}_t + \hat{c}_t^\top x_t}{2\hat{b}_t}$

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 $(\hat{a}_{t+1}, \hat{c}_{t+1}) \leftarrow \arg\min \sum_{s=1}^t (d_s - \hat{b}_{t+1} p_s - \alpha - \gamma^\top x_s)^2$

end for

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 x_t and error term ν_t.

Indeed, the error term ν_t contains an approximation part $f(x_t)-(a+c^\mathsf{T}x_t)$ due to model misspecification, so the least squares parameters a and c will be biased if x_t and ν_t are correlated. The reason why we can ignore the correlation between x_t and ν_t in pricing decisions is that computing an optimal price for the linear model, namely $-(a+c^\mathsf{T}x_t)/(2b)$, only requires an unbiased estimate of the *aggregated* effect of the feature vector on demand, which is measured by the numerator $a+c^\mathsf{T}x_t$ rather than the treatment effect of each individual component of x_t. Unbiased estimates of $a+c^\mathsf{T}x_t$ and b can be provided by the RPS algorithm. However, when x_t is endogenous, the RPS cannot guarantee that the estimates of a, c are unbiased component-wise.

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

$$D_t(p) = bp + f(x_t) + \epsilon_t, \quad \forall t = 1, \dots, T \tag{5}$$

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 $\tilde{p}_t(x_t) = -\frac{f(x_t)}{2b}$ cannot be expressed as an affine function of x_t. In Appendix A of E-companion, we show that if the sequence of p_t for $t = 1, \dots, T$ is chosen based on linear demand models, the model misspecification error is quantified by

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

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 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:

$$D_t(p) = a + bp + c^\mathsf{T}x_t + \nu_t, \quad \forall p \in [\underline{p}_t, \bar{p}_t], \tag{6}$$

where a and c are population least squares estimates of $f(x_t)$: $a, c = \arg\min_{\alpha,\gamma} \mathrm{E}[\|f(x_t) - (\alpha + \gamma^\mathsf{T} x_t)\|^2]$. It can be shown by solving first order conditions that a, c are given by the closed form expression:

$$\begin{bmatrix} a \\ c \end{bmatrix} = \left(\mathrm{E}\left[\begin{bmatrix} 1 & x_t^\mathsf{T} \\ x_t & x_t x_t^\mathsf{T} \end{bmatrix} \right] \right)^{-1} \mathrm{E}\left[\begin{bmatrix} f(x_t) \\ f(x_t) x_t \end{bmatrix} \right]. \tag{7}$$

One can view linear model (6) as the projection of the true quasi-linear model (1) to the linear model family (see Fig. 1). Let $p_t^*(x_t) = -\frac{a + c^\mathsf{T} x_t}{2b}$ be the optimal price under the best linear model given by Eq (6). The proposition below shows that the linear demand function (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 $p_t' = -\frac{\alpha + \gamma^\mathsf{T} x_t}{2\beta}$ that is affine in features 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 $p_t^*(x_t)$, namely*

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

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

$$\mathrm{Regret}(T) = \sum_{t=1}^{T} \mathrm{E}[p_t^*(x_t) D(p_t^*(x_t))] - \sum_{t=1}^{T} \mathrm{E}[p_t D(p_t)], \tag{8}$$

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 (8), we use the optimal price of the best linear model (6), $p_t^*(x_t)$, instead of the absolute optimal price for the true quasi-linear model (5), $\tilde{p}_t(x_t)$. The reason is that the optimal price $p_t^*(x_t)$ of model (6) gives the highest *achievable* revenue if the seller is restricted to making pricing decisions using linear demand models. Should we replace $p_t^*(x_t)$ by $\tilde{p}_t(x_t)$ in the definition of regret in (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.

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* (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 $M = \mathrm{E}[(1, x_t)(1, x_t)^\intercal]$, 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 x_t except through the parameter $\lambda_{\min}(M)$. The proof of Theorem 1 shows explicitly how the regret depends on these parameters, see Appendix C of E-companion.

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 (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.

Remark 1. (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 (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.

3.3. A Lower Bound of 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 C. 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 B.2 of E-companion). 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.

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 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.

We benchmark the performance of admissible pricing algorithms against a clairvoyant who knows the "best" linear demand model given by (6), and selects price

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

upon observing feature x_t. Then the expected regret, as before, is given by

$$\text{Regret}(T) = \sum_{t=1}^{T} \mathrm{E}[p_t^* D(p_t^*)] - \sum_{t=1}^{T} \mathrm{E}[p_t D(p_t)], \tag{9}$$

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

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

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

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

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

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 $(\hat{a}_{t+1}, \hat{c}_{t+1}) \leftarrow \arg\min \sum_{s=1}^t (d_s - \hat{b}_{t+1} p_s - \alpha - \gamma^\top x_s)^2$

end for

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 $1 \le i \le N$. 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 $\mathrm{E}[\Delta p_t] = 0$ and $\mathrm{Var}[\Delta p_t]$ is a decreasing function of t. While Algorithm 1 sets Δp_t such that $\mathrm{Var}[\Delta p_t] \propto \frac{1}{\sqrt{t}}$, our modified RPS algorithm sets Δp_t such that $\mathrm{Var}[\Delta p_t] \propto \frac{1}{t^{1/3}}$. 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.

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 (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 C of E-companion. 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 $p_t^* = \text{Proj}(-\frac{a+c^\intercal x_t}{2b}, \{q_1, q_2, \ldots, q_N\})$ 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.

3.5. Non-IID features

Previously, we assumed that the features $\{x_t\}_{t=1}^T$ 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 $\{x_t\}_{t=1}^T$ 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.

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

$$\hat{D}_t(p) = a_x + bp + c_x^\intercal x_t, \quad \forall p \in [\underline{p}_t, \bar{p}_t], \tag{10}$$

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

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

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

$$\begin{bmatrix} a_x \\ c_x \end{bmatrix} = \left(\sum_{t=1}^T \begin{bmatrix} 1 & x_t^\intercal \\ x_t & x_t x_t^\intercal \end{bmatrix} \right)^{-1} \sum_{t=1}^T \begin{bmatrix} f(x_t) \\ f(x_t) x_t \end{bmatrix}. \tag{11}$$

Notice that Eq (10) describes the linear model that best approximates $f(x_t)$ under the empirical distribution given $\{x_t\}_{t=1}^T$.

We assume that the parameters (a_x, b, c_x) are bounded as follows: $|a_x| \le \bar{a}$, $\underline{b} \le |b| \le \bar{b}$, $\|c_x\|_1 \le \bar{c}$. The seller is assumed to know the bounds on b, $\underline{b}$ and $\bar{b}$, but not the bounds on a_x and 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, c_x, and the expected revenue achieved by that pricing policy.[2] We note here that although the clairvoyant has full knowledge of the realization $\{x_t\}_{t=1}^T$ at the start of the selling horizon,

[2] Again, we could define regret relative to the "true clairvoyant," who knows the *true* demand model and sets price $\tilde{p}_t = -\frac{f(x_t)}{2b}$ at each time period. But this definition can result in a linear regret (see details in Appendix A of E-companion). Namely if $\{x_t\}$ happens to be IID, $\text{E}\left[\sum_{t=1}^T \tilde{p}_t(x_t) D(\tilde{p}_t(x_t)) - \sum_{t=1}^T p_t D(p_t)\right] = \Omega(T)$. 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.

any admissible pricing policy does not know this realization, and can only observe the history $\mathcal{H}_{t-1} = \{p_1, x_1, d_1, \ldots, p_{t-1}, x_{t-1}, d_{t-1}\}$. The expected regret is given by

$$\text{Regret}(T) = \sum_{t=1}^{T} \mathrm{E}[p_t^* D(p_t^*)] - \sum_{t=1}^{T} \mathrm{E}[p_t D(p_t)], \tag{12}$$

where $p_t^* = -\frac{a_x + c_x^\top x_t}{2b}$ are the price chosen by the clairvoyant upon observing feature x_t. The expectation in Eq (12) is taken over all random quantities, including the features 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 $p_t^*(x) = -\frac{a_x + c_x^\top x}{2b}$ for all t and for any features x. Analogous to Proposition 1, Proposition 2 below shows that the linear demand function (10) gives the highest revenue among all linear demand functions, justifying our choice of regret benchmark.

PROPOSITION 2. *Given a particular realization $\{x_t\}_{t=1}^T$ of the features, consider price $p_t' = -\frac{\alpha + \gamma^\top x_t}{2\beta}$ 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*

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

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 x_s, we use b_s as an estimate 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 x_s. This modification allows us to extend the analysis to an arbitrary sequence of features $\{x_t\}_{t=1}^T$. 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 C of E-companion, 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

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 x_t, set unconstrained greedy price: $p^u_{g,t} \leftarrow -\frac{\hat{a}_t + \hat{c}_t^\top x_t}{2\hat{b}_t}$

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 $(\hat{a}_{t+1}, \hat{c}_{t+1}) \leftarrow \arg\min \sum_{s=1}^t (d_s - \hat{b}_{t+1} p_s - \alpha - \gamma^\top x_s)^2$

end for

the prediction of sequences in the presence of feature information. In their setup, a sequence $\{(y_1, g(y_1)), (y_2, g(y_2)), \ldots\}$ is observed, where the y_ns are d-dimensional feature vectors and the function g determining the outcome variable $g(y_t)$ is potentially nonlinear. The goal is to predict the outcomes $g(y_n)$ for each n based on the observations $\{(y_i, g(y_i)), i = 1 \ldots n-1\}$. 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 $\{x_t\}$ 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 $f(x)$ 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 C.

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. The first set of simulations, presented in Section 4.1, makes use of synthetic data. These experiments 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, and show that the regret growth matches our theoretical guarantees from the previous section, thus validating our theoretical analysis. The second set of simulations, presented in Section 4.2, is based on higher-dimensional fashion retail data provided by Oracle Retail. These simulation experiments serve to gauge the performance of the RPS algorithm in a real-world setting. Both sets of simulations benchmark 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.

4.1. Numerical Experiments with Synthetic Data

Each of the simulations in this section is run over a selling horizon of length 5000 periods and repeated 200 times, and compares the performance of the RPS algorithm with the performance of the following three algorithms:

- *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

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

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.

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 (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 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. 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. 2b confirms that the regret of the RPS algorithm is $O(\sqrt{T})$. Finally, Table 1, which provides summary

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: $p^u_{g,t} \leftarrow -\frac{\hat{a}_t + \hat{c}_t^\top x_t}{2\hat{b}_t}$

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

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 $(\hat{a}_{t+1}, \hat{b}_{t+1}, \hat{c}_{t+1}) \leftarrow \arg\min_{\alpha, \beta \in B, \gamma} \sum_{s=1}^{\mathsf{T}} (d_s - \alpha - \beta p_s - \gamma^\top x_s)^2$

end for

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 B.1 of E-companion, we include additional numerical experiment 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 B.2 of E-companion. 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, \$0.90, ..., \$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. 2c) while the regret of the RPS algorithm (Algorithm 2) is $O(T^{2/3})$ (Fig. 2d). The summary statistics of the parameter estimates of the competing algorithm (Table 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 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 $[\min_t\{a(t)\} - 0.5, \max_t\{a(t)\}] = [1.9,\ 2.6]$, b lies in the interval $[-1.2,\ -0.1]$ and c lies in the interval $[\min_t\{c(t)\} - 0.5, \max_t\{c(t)\}] = [-7.3,\ 0.3]$.

Results Fig. 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. 2f. Table 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 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

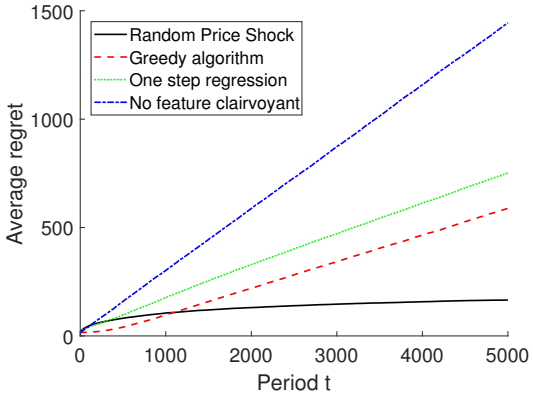
Table 2 End of selling horizon parameter estimates in the price ladder setting

	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

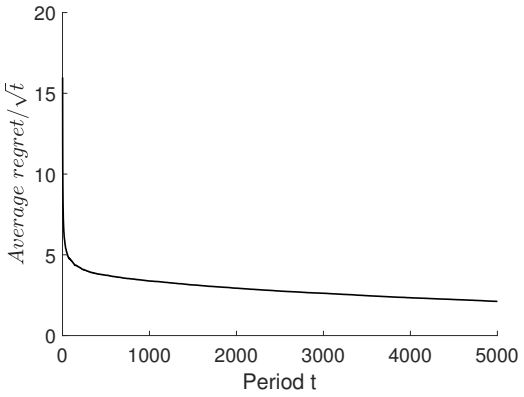
f

Table 3 End of selling horizon parameter estimates in the non IID setting

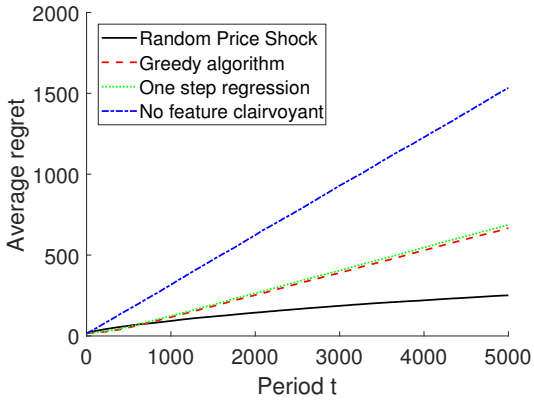
	True value	RPS algo.	Greedy algo.	One step reg.
Mean ($\hat{a}_T$)	-1.38	-1.35	-1.49	-0.87
Median ($\hat{a}_T$)	-1.38	-1.37	-1.50	-0.88
Mean ($\hat{b}_T$)	-0.90	-0.91	-0.16	-0.40
Median ($\hat{b}_T$)	-0.90	-0.91	-0.16	-0.40
Mean ($\hat{c}_T$)	-6.63	-6.60	-3.95	-4.40
Median ($\hat{c}_T$)	-6.63	-6.66	-3.97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



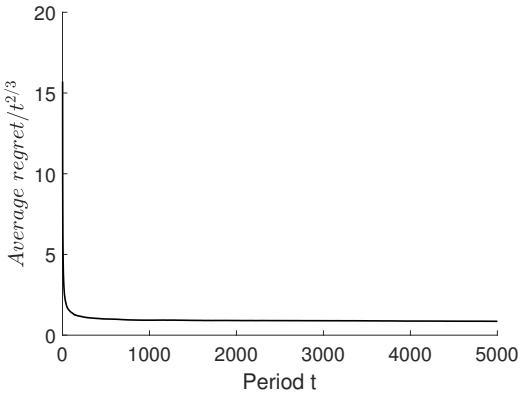
(a) IID setting – average regret



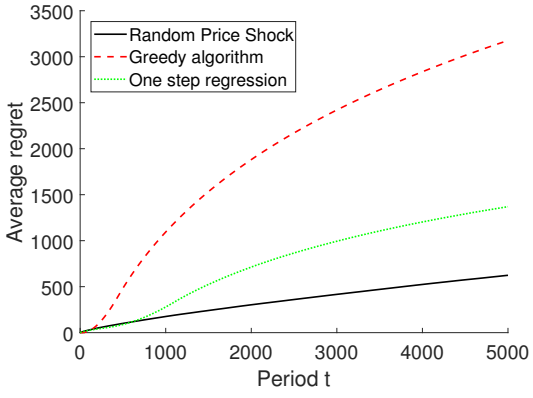
(b) IID setting – avg regret scaled by $1/\sqrt{t}$



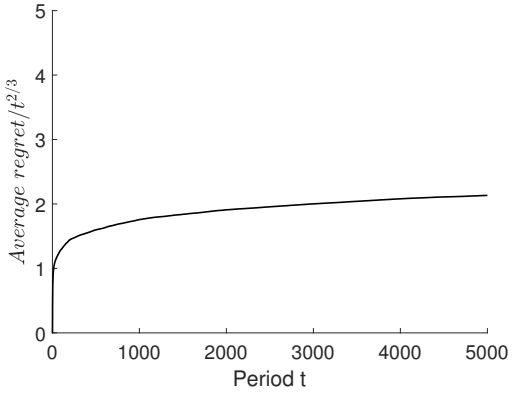
(c) Price ladder setting – average regret



(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 Average regret and scaled regret in IID, price ladder and non IID settings

4.2. Case Study

In collaboration with Oracle Retail, we performed simulation experiments to gauge the performance of the RPS algorithm in a real-world setting. These experiments were performed on a dataset

provided by Oracle Retail, consisting of three years' worth of data on customer transactions and product feature information at an anonymous chain of brick-and-mortar department stores.

The goal of our experiments was to 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. The process of building our demand model is described first. 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 algorithm over the selling horizon, allowing it to price the items in the historical dataset based on their feature information. As in the computational experiments in Section 4 with synthetic data, we evaluated the performance of the RPS algorithm by comparing its estimated revenue with the estimated revenues of the greedy and the one-stage regression algorithms (cf. Algorithms 4 and 5). The details of our experiments, from building our demand model to running simulations, reported below.

Data Processing. We had access to two main types of datasets:

1. Customer transaction data – This dataset consists of customer transactions from August 2012 to July 2015. Purchased items are in the categories of fashion, furniture and housewares. 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.
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.

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 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.

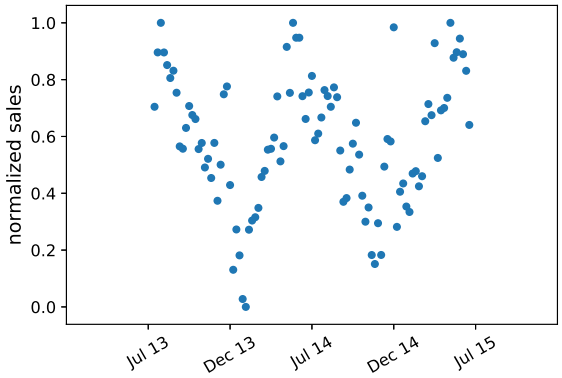


Figure 3 Seasonality of Demand

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:

$$D^S_{i,d,w} = b_S p_{i,d,w} + f_S(x_{i,d,w}) + \epsilon^S_{i,d,w}. \tag{13}$$

Here, $p_{i,d,w}$ represents the price of item i offered in district d and week w, and feature vector $x_{i,d,w}$ represents the item-specific features and seasonal information. This function is linear in price and possibly nonlinear in the features $x_{i,d,w}$, and is analogous to the single product demand function we defined in Eq (1).

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 metrics 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 began the demand estimation process by estimating the parameter b_S in (13) for each subclass S. Our initial approach was to simply apply ordinary linear regression (OLS) on the historical data. We used standard variable selection techniques and measured the accuracy of the estimated model by randomly splitting the dataset into a training set and a testing set in the ratio 70:30. However, the first column of Table 4 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.

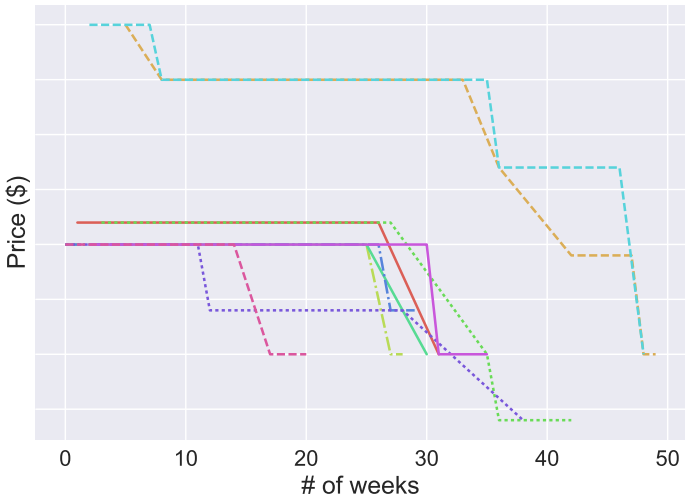


Figure 4 **Markdown pricing: each trajectory represents the price of one item from the category.**

We thus attempted to correct for endogeneity by using the two-stage least squares (2SLS) method. In the absence of the availability of cost-side variables, our choice of instrumental variable was, for each item grandparent-district-week tuple, the average price of all item grandparents sold in other districts during the same week. 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. 4). 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.

Table 4 **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)

The corrected price coefficient estimates with 2SLS for all four subclasses are given in the second column of Table 4 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 (13). Substituting our 2SLS estimates of the demand elasticity b_S from Table 4 into Eq (13), 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 5 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 5 **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%

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: $D^S_{i,d,w} = a_S + (b_S x_{i,d,w}) p_{i,d,w} + c_S^\intercal x_{i,d,w} + \epsilon_{i,d,w}$. 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: $D^S_{i,d,w} = a^S_{i,d,w} + b^S_1 p_{i,d,w} + b^S_2 \bar{p}_w + c_S^\intercal x_{i,d,w} + \epsilon_{i,d,w}$, 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 (13) where the function f_S is our random forest estimator. In Table 6, 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 (13) as the counterfactual demand model in our simulations.

Table 6 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%

Simulation Results After estimating the ground truth demand model, we ran numerical simulations to determine how well the RPS algorithm could learn the model parameters and set prices.

The RPS algorithm and its variants presented in Section 3 cannot be directly applied to this retail setting, since they assume a single product setting where demand is observed sequentially. In our fashion retail setting, however, we price all items in a subclass simultaneously at the start of every week, and observe demand for these items in batches at the end of each week. We therefore modify the RPS for the batch updating setting. The algorithm statement is given in Algorithm 6. It assumes that I_t items are sold in a particular week t, and denotes the price ladder at each week t by $\{q_1, \ldots, q_{N_t}\}$.

In addition to modifying the pricing algorithm, we imposed the following price constraints on the output of the algorithm:

Algorithm 6 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 x_t, set unconstrained greedy price: $p^u_{g,t} \leftarrow -\frac{\hat{a}_t + \hat{c}_t^T x_t}{2\hat{b}_t}$

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

end for

1. *Price ladder*: All prices chosen by a pricing algorithm had to be rounded to end with 99 cents, e.g. \$3.99, \$5.99, \$7.99. This constraint was also imposed by the retailer on historical prices from the dataset.
2. *Price bounds*: For each item grandparent-district-week tuple, the price charged by a pricing algorithm was restricted to within 20% of the historical price charged by the retailer. This had the effect of ensuring that the prices charged were appropriate (e.g. a \$100 item could not be priced at \$1), and did not deviate wildly from time period to time period.

We ran the RPS algorithm with batch updating on a week-by-week basis over a 35 week horizon. During the first two weeks, the algorithm set 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 had 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 calculated the corresponding expected demand using our random forest model, then added this expected demand to the prediction error of the random forest model, which we assumed to be the demand noise.

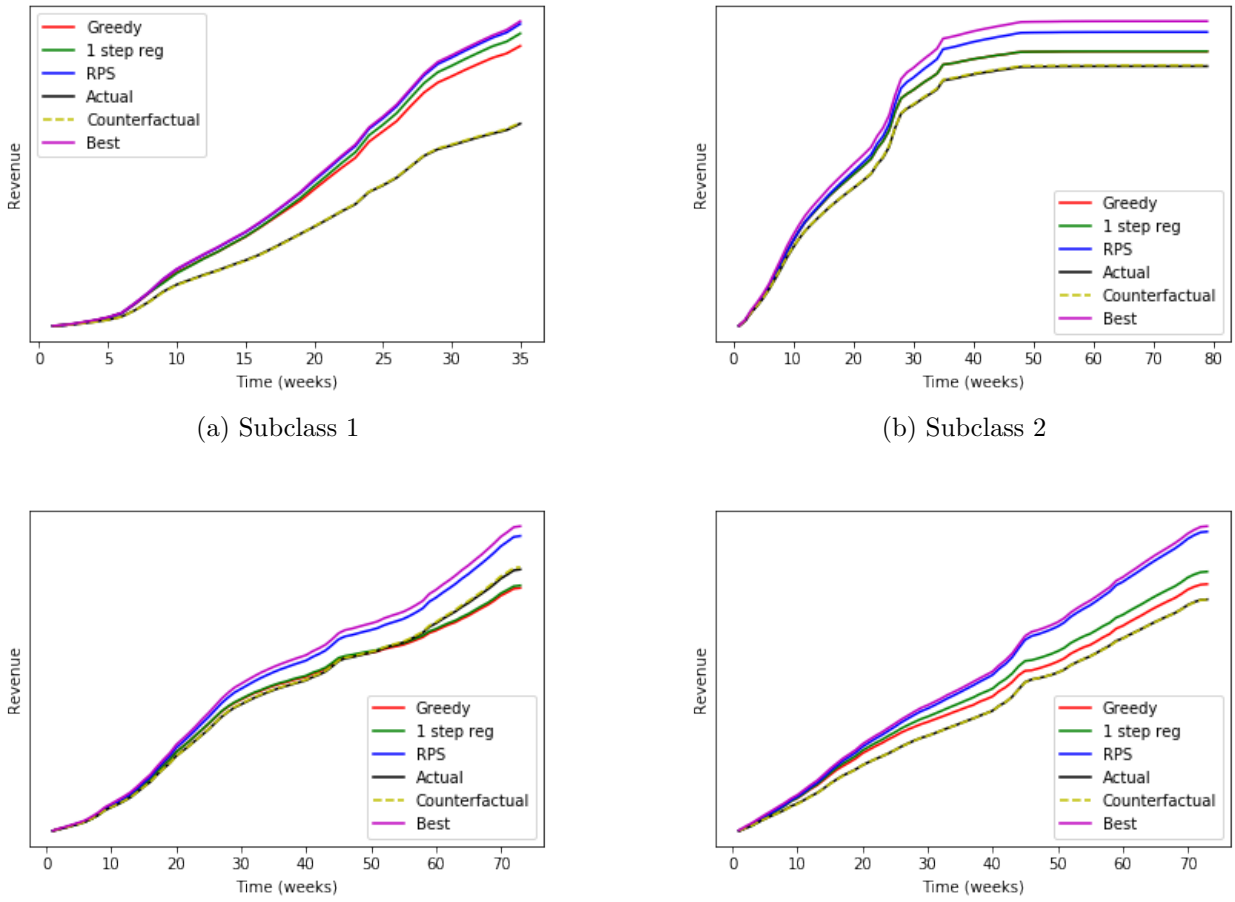
We also ran batch updating variants of the greedy and one-stage regression algorithms, which we subjected to the same pricing constraints as RPS. The pseudocode of these algorithms is omitted as they are modified from Algorithms 4 and 5 in a similar manner to RPS.

Fig. 5 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. Note that 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.[3] 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 8 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 8 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 7, 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 above, are likely to be correlated with demand noise.

Finally, we compare the revenue of RPS to the best possible (clairvoyant) revenue given full knowledge of the demand function, see Fig. 5 and Table 8. 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 close to the clairvoyant revenue for all four subclasses.

[3] In addition to the price ladder constraint, the actual prices set by the retailer satisfy a markdown constraint, which stipulates that prices must decrease monotonically with time towards the end of the selling horizon. We excluded markdown constraints in all the algorithms tested, because we were not given information by the retailer on how to formulate markdown constraints.



(a) Subclass 1

(b) Subclass 2

(c) Subclass 3

(d) Subclass 4

Figure 5 Average revenue over 100 iterations of different algorithms

Table 7 Estimates of parameter b (with 95% confidence interval)

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

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

Table 8 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% (7.84%, 8.00%)	3.32% (2.50%, 4.62%)	-0.85% (-0.92%, -0.78%)
2	6.98% (3.03%, 8.33%)	6.97% (2.92%, 8.31%)	-3.53% (-7.09%, -2.31%)
3	21.23% (21.23%, 22.27%)	20.30% (17.62%, 21.32%)	-3.18% (-4.74%, -2.35%)
4	21.04% (19.19%, 21.59%)	15.28% (13.56%, 17.35%)	-1.77% (-3.26%, -1.31%)

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.

Finally, we have demonstrated the real-world applicability of our model and algorithm through a case study in collaboration with Oracle Retail, involving a large fashion retail dataset from a chain of brick and mortar department stores. This case study shows how our model can be extended beyond its single product setting, to a setting where multiple products are sold simultaneously from week to week. Using our historical data, we have performed offline simulations gauging the performance of RPS in this setting. The results of our simulations are very promising and show that the RPS algorithm is expected to earn 8-20% more revenue on average than competing algorithms that do not account for price endogeneity.

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.

Acknowledgments

The authors gratefully acknowledge the generous support from Oracle Corporation through an External Research Office grant. They thank Su-Ming Wu, Setareh Borjian, Sajith Vijayan from Oracle Retail Global Business Unit for their suggestions and valuable feedback in this work and for providing the data used in the case study of Section 4. The authors also thank the department editor, associate editor, and reviewers for insightful comments that helped improve the manuscript.

References

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

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.

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.

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.

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.

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.

Fisher, M., Gallino, S., and Li, J. (2017). Competition-based dynamic pricing in online retailing: A methodology validated with field experiments. *Management Science (forthcoming)*.

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.

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.

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*.

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.

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.

Appendix. E-Companion to "Dynamic Learning and Pricing with Model Misspecification"

A. 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 3.1 as the "linear clairvoyant," and define a second clairvoyant, called the "true clairvoyant," who sets price $\tilde{p}_t = -\frac{f(x_t)}{2b}$ 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

$$\begin{aligned}
\text{Regret}_2(T) &= \text{Regret}(T) + \sum_{t=1}^{T} \mathrm{E}[\tilde{p}_t D(\tilde{p}_t)] - \mathrm{E}[p_t^* D(p_t^*)] \qquad (14)\\
&= \text{Regret}(T) + \frac{T}{4|b|}\mathrm{E}\left[\left(f(x_t) - \mathrm{E}\left[f(x_t)\begin{bmatrix}1 & x_t^\mathsf{T}\end{bmatrix}\right]\left(\mathrm{E}\left[\begin{bmatrix}1 & x_t^\mathsf{T}\\ x_t & x_t x_t^\mathsf{T}\end{bmatrix}\right]\right)^{-1}\begin{bmatrix}1\\ x_t\end{bmatrix}\right)^2\right]\\
&\geq \frac{T}{4|b|}\mathrm{E}\left[\left(f(x_t) - \mathrm{E}\left[f(x_t)\begin{bmatrix}1 & x_t^\mathsf{T}\end{bmatrix}\right]\left(\mathrm{E}\left[\begin{bmatrix}1 & x_t^\mathsf{T}\\ x_t & x_t x_t^\mathsf{T}\end{bmatrix}\right]\right)^{-1}\begin{bmatrix}1\\ x_t\end{bmatrix}\right)^2\right]
\end{aligned}$$

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 x_t. We have also included additional numerical experiment using $\text{Regret}_2(T)$ as the benchmark, see Appendix B.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 (14) shows that $\text{Regret}_2(T)$ of any admissible pricing policy affine in 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.

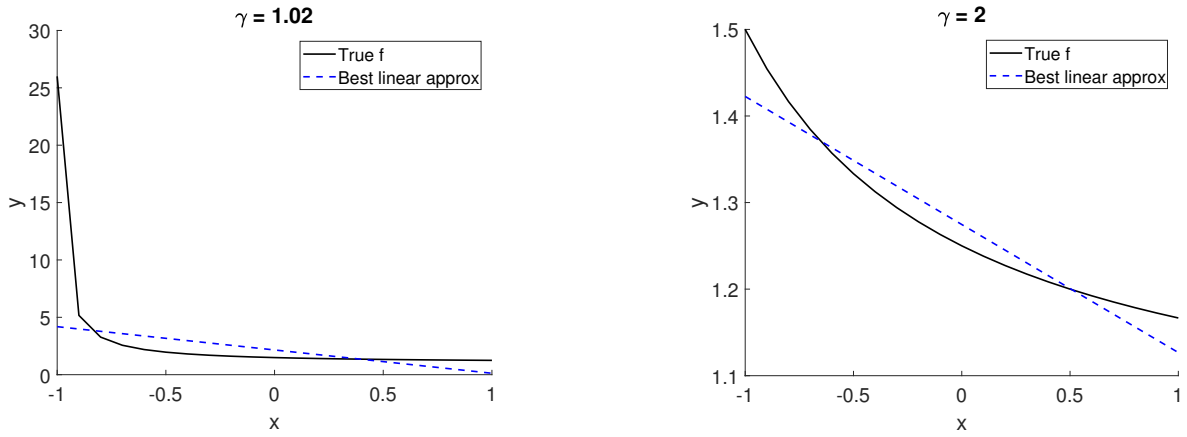
B. Additional Numerical Results

In this section we expand on the numerical results in Section 4 by investigating how our results depend on the parameter settings. Section B.1 shows how the performance of the RPS algorithm depends on the choice of demand function. Section B.2 looks at its dependence on the dimension of the feature vector m, complementing our theoretical results on the RPS algorithm's regret upper bound given in Section 3.

B.1. Dependence of regret on demand function

We now investigate how the results of our simulations depend on the demand function. In the IID setting studied in Section 4, the quasilinear demand model is of the form

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



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

Figure 6 $f(x)$ **vs best linear approximation** $a + c'x$ **for** $\gamma = 1.02, 2$

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

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

As γ increases, the fit of the closest linear approximation of D_t for x_t uniformly distributed between $[-1, 1]$ improves, i.e. $\mathrm{E}[(f(x_t) - a - c'x_t)^2]$ decreases. Fig. 6 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 9, 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. 7 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.

B.2. Dependence of regret on 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 $T = 5000$

Table 9 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

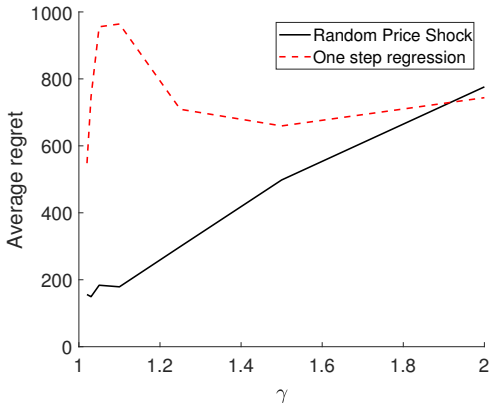


Figure 7 Average regret over 50 iterations of RPS vs one-stage regression algorithms as γ is varied

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 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

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

For each m, the feature vectors x_t are drawn IID from the distribution $[-1,1]^m$, and c is a vector of length m with the first entry set to 0.9 and all other entries set to 0. Note that $\|c\|_1$ is constant for all m, and thus so is $\bar{c}$, on which our regret bound depends (see Eq (21) for the full statement of the IID regret bound in terms of all parameters). We set $c_{\max}$ to $c + [0.5, 0.5, \ldots, 0.5]$ and $c_{\min}$ to $c - [0.5, 0.5, \ldots, 0.5]$, 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. 8 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.

B.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

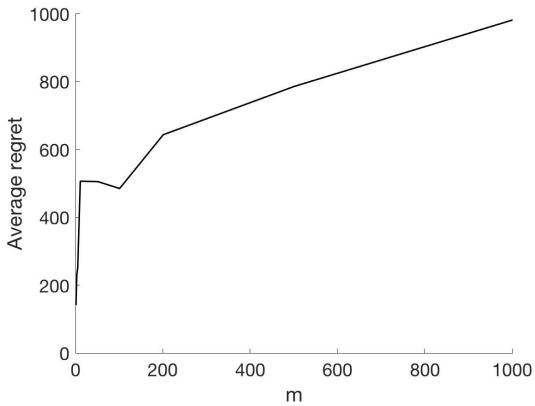


Figure 8 Average regret over 10 iterations of the RPS algorithm as m is increased from 1 to 1001.

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 $\tilde{p}_t = -\frac{f(x)}{2b}$ at each time period. Fig. 9a plots the results of repeating the IID setting experiments from Section 4.1; it plots the average regret of the RPS algorithm relative to both clairvoyants over 200 iterations and 5000 time periods. Similarly, Fig. 9b plots the results of repeating the price ladder setting experiments from Section 4.1, and Fig. 9c plots the results of repeating the non IID experiments from Section 4.1.

Fig. 9a 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. 9b 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 9a - 9c 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 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 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 x_t in the feature vector could decrease the regret from model misspecification, but increase the regret due to parameter estimation errors.

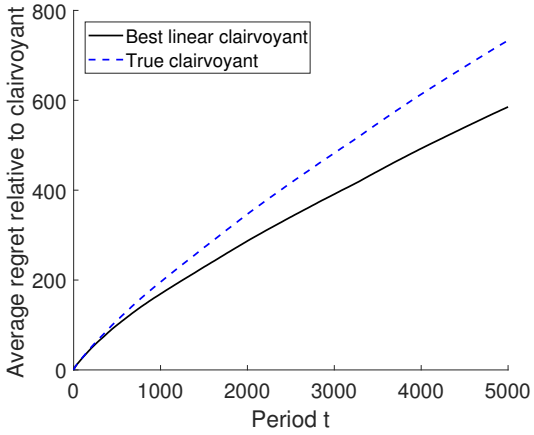
C. Appendix: Proofs for Theoretical Analysis

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

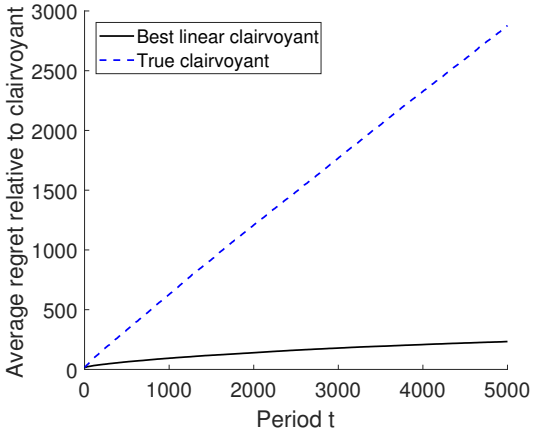
C.1. Proof of Proposition 1

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

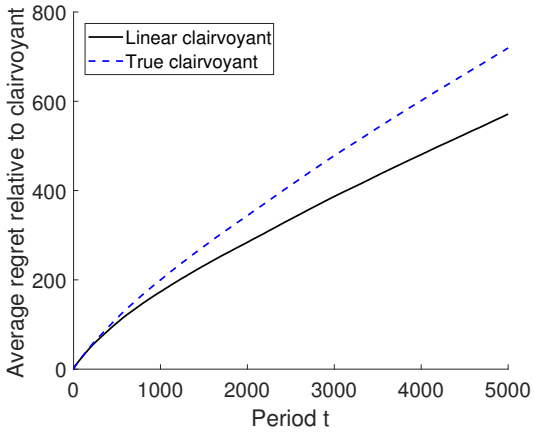
$$\begin{aligned}\mathrm{E}\left[p^*_t D(p^*_t) - p'_t D(p'_t) \mid \mathcal{H}_{t-1}\right] =& \mathrm{E}\left[p^*_t(bp^*_t + f(x_t)) - p'_t(bp'_t + f(x_t)) \mid \mathcal{H}_{t-1}\right] \\ =& \mathrm{E}\left[p^*_t(bp^*_t + a + c^\mathsf{T} x_t) - p'_t(bp'_t + a + c^\mathsf{T} x_t) - (p^*_t - p'_t)(a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



(a) IID setting – average regret



(b) Price ladder IID setting – average regret



(c) Non IID setting – average regret

Figure 9 Average regret over 200 iterations of RPS algorithm relative to two different clairvoyants in IID and Price ladder IID settings

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

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

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

By first order conditions, we have

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

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

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

Therefore,

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

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

C.2. Proof of Theorem 1

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

$$\text{Expected Regret}(T)=\sum_{t=1}^{T}\mathrm{E}[p_t^*D(p_t^*)]-\sum_{t=1}^{T}\mathrm{E}[p_tD(p_t)]. \tag{15}$$

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:

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

We can write the regret as

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

where the second inequality follows from the definition of p_t^* and our assumptions on the boundedness of the true parameters a,b,c, and the final inequality follows by bounding $\mathbb{P}(A_t^C)$ by Lemma 2, where $V:=\mathrm{E}[(Q^{-1}\tilde x\tilde x^\mathsf{T}Q^{-1}-I)^2]$. 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:

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

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

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

In addition, $p_t = p_{g,t} + \Delta p_t$, where Δp_t is generated independently from $p_{g,t}, x_t$ and the history $\mathcal{H}_{t-1}$ with variance $\delta_t^2 = \frac{\delta^2}{4\sqrt{t}}$. So

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

To finish the proof, we want to show that $\mathrm{E}[p_t^* D_t(p_t^*) - p^u_{g,t} D_t(p^u_{g,t}) | A_t] \cdot \mathbb{P}[A_t] = O(1/\sqrt{t})$. 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 $p_t' = -\frac{\alpha + \gamma^\mathsf{T} x_t}{2\beta}$ 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 x_t, this gives

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

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

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

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

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

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

where the second line follows from the inequality $(x+y)^2 \le 2x^2 + 2y^2$, and the third line follows from the fact that the true parameters a, c satisfy $\|a\| \le \bar{a}$ and $\|c\|_1 \le \bar{c}$, 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 $|\nabla h(b')|^2 = \frac{1}{b'^4} \le \frac{1}{\underline{b}^4}$. Then by the Mean Value Theorem, we have

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{1}{b} - \frac{1}{\hat{b}_t} \right)^2 | A_t \right] \cdot \mathbb{P}[A_t] &\le \frac{1}{b^4} \mathrm{E}[(b - \hat{b}_t)^2 | A_t] \cdot \mathbb{P}[A_t] \\
&\le \frac{1}{b^4} \mathrm{E}[(b - \hat{b}_t)^2]. \qquad (16)
\end{aligned}$$

By Lemma 1, 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 c, namely $\mathrm{E}[\left((e-e_t)^\mathsf{T}\tilde{x}_t\right)^2 | A_t]$. Note that e_t is measurable with history $\mathcal{H}_{t-1}$ and $\tilde{x_t}$ is independent of $\mathcal{H}_{t-1}$, so

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

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

By the definition of Algorithm 1, assuming that M_t is invertible, $e_t - e$ can be written as

$$e_t - e = \mathrm{Proj}\left(M_t^{-1}\frac{\sum_{s=1}^{t-1}\tilde{x}_s(p_s(b-\hat{b}_t)+\epsilon_s)}{t-1}\right). \tag{17}$$

Then we have

$$\begin{aligned}\mathrm{E}[\left((e-e_t)^\mathsf{T}\tilde{x}_t\right)^2 |] &= \mathrm{E}[\|e_t-e\|_M^2 \mid A_t]\cdot\mathbb{P}[A_t] \\ &\le \mathrm{E}[\|e_t-e\|_M^2 \mid A_t]\cdot\mathbb{P}[A_t] + \mathrm{E}[4(e^\mathsf{T}\tilde{x}_t)^2 + 4(e_t^\mathsf{T}\tilde{x}_t)^2]\cdot\mathbb{P}[A_t^C] \\ &\le \mathrm{E}[\|e_t-e\|_M^2 \mid A_t]\cdot\mathbb{P}[A_t] + 16\bar{b}^2p_{\max}^2\mathbb{P}[A_t^C] \\ &\le \mathrm{E}[\|e_t-e\|_M^2 \mid A_t]\cdot\mathbb{P}[A_t] \qquad (18) \\ &\quad + 16\bar{b}^2p_{\max}^2\cdot 2(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right). \qquad (19)\end{aligned}$$

The third line follows from the assumption that the true parameter $e \in E$. In the last step, we bound $\mathbb{P}(A_t^C)$ by Lemma 2, where $V := \mathrm{E}[(Q^{-1}\tilde{x}\tilde{x}^\mathsf{T}Q^{-1} - I)^2]$. Since Eq (19) is $O(e^{-t})$, it is left to show that Eq (18) is $O(\sqrt{1/t})$.

We write Eq (18) as

$$\begin{aligned}\mathrm{E}[\|e_t-e\|_M^2 \mid A_t]\cdot\mathbb{P}[A_t] &\le \mathrm{E}\left[\|QM_t^{-1}QQ^{-1}\frac{\sum_{s=1}^{t-1}\tilde{x}_s(p_s(b-\hat{b}_t)+\nu_s)}{t-1}\|^2 \mid A_t\right]\mathbb{P}[A_t] \\ &\le \mathrm{E}\left[\|QM_t^{-1}Q\|_2^2\cdot\|Q^{-1}\|_2^2\cdot\|\frac{\sum_{s=1}^{t-1}\tilde{x}_s(p_s(b-\hat{b}_t)+\nu_s)}{t-1}\|^2 \mid A_t\right]\mathbb{P}[A_t] \\ &\le \mathrm{E}\left[4\cdot\frac{1}{\lambda_{\min}(M)}\cdot 2\left(\|\frac{\sum_{s=1}^{t-1}\tilde{x}_sp_s(b-\hat{b}_t)}{t-1}\|^2 + \|\frac{\sum_{s=1}^{t-1}\tilde{x}_s\nu_s}{t-1}\|^2\right) \mid A_t\right]\mathbb{P}[A_t] \\ &\le \mathrm{E}\left[\frac{8}{\lambda_{\min}(M)}\left(\|\frac{\sum_{s=1}^{t-1}\tilde{x}_sp_s(b-\hat{b}_t)}{t-1}\|^2 + \|\frac{\sum_{s=1}^{t-1}\tilde{x}_s\nu_s}{t-1}\|^2\right)\right] \\ &\le \mathrm{E}\left[\frac{8}{\lambda_{\min}(M)}\left((m+1)p_{\max}^2(b-\hat{b}_t)^2 + \|\frac{\sum_{s=1}^{t-1}\tilde{x}_s\nu_s}{t-1}\|^2\right)\right] \\ &= \frac{8}{\lambda_{\min}(M)}\left((m+1)p_{\max}^2\mathrm{E}[(b-\hat{b}_t)^2] + \frac{1}{(t-1)^2}\mathrm{E}[\|\sum_{s=1}^{t-1}\tilde{x}_s\nu_s\|^2]\right). \qquad (20)\end{aligned}$$

The first inequality holds by Eq (17) and the assumption that the true parameter $e \in E$. 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 $\|x+y\|^2 \le 2\|x\|^2 + 2\|y\|^2$. The fourth inequality simply uses the definition of conditional expectation. The fifth inequality uses the assumptions that $\|x_t\|_\infty \le 1$ and $p_j \le p_{\max}$.

It has already been established using Lemma 1 that $\mathrm{E}[(b-\hat{b}_t)^2]$ is $O(1/\sqrt{t})$, so the first term of Eq (20) is $O(1/\sqrt{t})$. For the second term, note that $(\tilde{x}_s, \nu_s)$ is independent of $(\tilde{x}_{s'}, \nu_{s'})$ for $s \neq s'$. Furthermore, by the first order condition of the least squares estimator, we have $\mathrm{E}[\nu_t] = 0$ and $\mathrm{E}[x_t \nu_t] = 0$. So for each s, $\mathrm{E}[\tilde{x}_s \nu_s \mid \mathcal{H}_{s-1}] = \mathrm{E}[\tilde{x}_s \nu_s] = 0$. Thus,

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

where the last step uses the fact that $(x+y+z)^2 \leq 3(x^2+y^2+z^2)$ and $\|\tilde{x}_s\|^2 \leq m+1$. Therefore, by Eq (20), $\mathrm{E}[\|e-e_t\|_M^2] \leq O(1/\sqrt{t}) + O((m+1)/t) = O(1/\sqrt{t})$ 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

$$O\left(\frac{\bar{b}^2(p_{\max}^2+1)}{\underline{b}^4}\frac{(\bar{f}^2+\sigma^2+\bar{b}^2p_{\max}^2)}{\delta^2}(1+p_{\max}^2\frac{m+1}{\lambda_{\min}(M)})\sqrt{T}\right) + O((m+1)\log T), \tag{21}$$

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

C.3. Proof of Theorem 2

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

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

We assume feature vector 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 $p_t^* = (a + c^\mathsf{T}x_t)/(-2b)$.

By Lemma 3, we can assume without loss of generality that the seller uses a linear pricing strategy π at period t given by $p_t = S_t + (U_t)^\mathsf{T}x_t$, 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 $\mathrm{Regret}(T)$. By Proposition 1, we have

$$\begin{aligned}\mathrm{Regret}(T) &= -b\mathrm{E}[(p_t - p_t^*)^2] \\ &= -b\mathrm{E}[(S_t + (U_t)^\mathsf{T}x_t - S^* - (U^*)^\mathsf{T}x_t)^2] \\ &= -b\left\{\mathrm{E}[(S_t-S^*)^2] + \sum_{k=1}^m \mathrm{E}[(U_{t,k}x_{t,k} - U_k^*x_{t,k})^2]\right\} \\ &= -b\left\{\mathrm{E}[(S_t-S^*)^2] + \frac{1}{12}\sum_{k=1}^m \mathrm{E}[(U_{t,k} - U_k^*)^2]\right\},\end{aligned} \tag{22}$$

where $S^* = -a/(2b)$, $U^* = -c/(2b)$, the third line follows since $\mathrm{E}[x_t] = 0$ and $x_t = (x_{t,1}, \ldots, x_{t,m})$ has independent entries for our particular choice of x_t, and the last line is because each entry of 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 $\theta = (a, b, c)$ belong to compact sets $\Theta = A \times B \times C$, where $A = [-\bar{a}, \bar{a}]$, $B = [-\bar{b}, -\underline{b}]$, $C = \{c' \in \mathbb{R}^m : \sum_{k=1}^m |c'_k| \le \bar{c}\}$. 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

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

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 $p_s = \pi(H_{s-1}, x_s)$. By our assumption that x_t is uniform and ϵ_t is normal, we have

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

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

$$\mathcal{I}_t(\theta) = \mathrm{E}_\theta\left[\nabla_\theta \log f_t(H_t \mid \theta) \cdot (\nabla_\theta \log f_t(H_t \mid \theta))^\mathsf{T}\right] = \mathrm{E}_\theta\left[\sum_{j=1}^t \begin{bmatrix} 1 & x_j^\mathsf{T} & p_j \\ x_j & x_j x_j^\mathsf{T} & p_j x_j^\mathsf{T} \\ p_j & p_j x_j & p_j^2 \end{bmatrix}\right]. \tag{23}$$

Define function $g(\theta) = [a/(2b), 1, c/(2b)]^\mathsf{T}$ 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

$$\mathrm{E}_\lambda[\mathrm{E}_\theta[(S_t - S(\theta))^2]] \ge \frac{\mathrm{E}_\lambda[g(\theta)^\mathsf{T}\nabla S(\theta)]^2}{\mathrm{E}_\lambda[g(\theta)^\mathsf{T}\mathcal{I}_{t-1}(\theta)g(\theta)] + \tilde{I}(\lambda)}, \tag{24}$$

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 (23) and $p_j^* = -(a + c^\mathsf{T} x_j)/(2b)$, one can show that

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

Substituting the equations above into Eq (24), we get

$$\mathrm{E}_\lambda[\mathrm{E}_\theta[(S_t - S(\theta))^2]] \ge \frac{\left(\mathrm{E}_\lambda[\frac{a}{4b^2}]\right)^2}{\mathrm{E}_\lambda[\mathrm{Regret}_\theta(T)] + \tilde{I}(\lambda)}. \tag{25}$$

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

$$\mathrm{E}_\lambda[\mathrm{E}_\theta[(U_{t,k} - U_k(\theta))^2]] \ge \frac{(\mathrm{E}_\lambda[\frac{c_k}{4b^2}])^2}{\mathrm{E}_\lambda[\mathrm{Regret}_\theta(T)] + \tilde{I}(\lambda)}. \tag{26}$$

Combining (22), (25), (26), and summing over $t = 1, \ldots, T$, we have

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

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

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

C.4. Proof of Theorem 3.

Proof. In the following, let $p^*_{t,u} := -\frac{a+c'x_t}{2b}$. 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

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

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

The loss from parameter estimation is upper bounded by

$$\begin{aligned}\sum_{t=1}^{T} \mathrm{E}[p_t^* D_t(p_t^*)] - \mathrm{E}[p_{g,t} D_t(p_{g,t})] &= \mathrm{E}[(p^*_{t,u} - (p_t^* - p^*_{t,u}) D_t(p^*_{t,u} - (p_t^* - p^*_{t,u})] - \mathrm{E}[p_{g,t} D_t(p_{g,t})] \qquad (27) \\ &\le K \mathrm{E}[|p^*_{t,u} - p_t^* + p^*_{t,u} - p_{t,g}|] \qquad (28) \\ &\le K(\mathrm{E}[|p^*_{t,u} - p_t^*|] + \mathrm{E}[|p^*_{t,u} - p_{t,g}|]) \\ &\le 2K \mathrm{E}[|p^*_{t,u} - p_{t,g}|].\end{aligned}$$

The first line, (27), follows from the fact that

$$\mathrm{E}[p_t^* D_t(p_t^*)] = \mathrm{E}[(p^*_{t,u} + (p_t^* - p^*_{t,u}) D_t(p^*_{t,u} + (p_t^* - p^*_{t,u})] = \mathrm{E}[(p^*_{t,u} - (p_t^* - p^*_{t,u}) D_t(p^*_{t,u} - (p_t^* - p^*_{t,u})],$$

by the symmetry of the function $p \mapsto \mathrm{E}[pD_t(p)]$ around its maximizer $p = p^*_{t,u}$. The second line, (28), follows from the mean value theorem since $\mathrm{E}[pD_t(p)]$ 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

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

thus (28) 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 $|p^*_{t,u} - p_t^*| \le |p^*_{t,u} - p_{t,g}|$ 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 $\mathrm{E}[|p^*_{t,u} - p_{t,g}|] \le \sqrt{\mathrm{E}[|p^*_{t,u} - p_{t,g}|^2]}$, 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

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

Lemma 1 can be applied to bound the term $\mathrm{E}[(b - \hat{b}_t)^2]$. Then, using the identity $\sqrt{x+y+z} \le \sqrt{x} + \sqrt{y} + \sqrt{z}$ for $x, y, z \ge 0$, 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

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

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

C.5. Proof of Proposition 2

Proof. Consider the optimization problem

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

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

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

Finally, note that

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

which proves Proposition 2. □

C.6. Proof of Theorem 4.

Proof. We decompose the regret as

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

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

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

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

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

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

$$\sum_{t=1}^{T}\mathrm{E}[p_t^*D(p_t^*)]-\mathrm{E}[p_{g,t}^uD_i(p_{g,t}^u)]=\frac{1}{4|b|}\sum_{t=1}^{T}\mathrm{E}[\mathrm{E}_x[(e_t^\mathsf{T}\tilde{x}_t-y_t')^2-(e_x^\mathsf{T}\tilde{x}_t-y_t')^2]] \tag{29}$$

$$+\frac{1}{|b|}\mathrm{E}[\mathrm{E}_x[(p_{t,g}^u)^2(b^2-\hat{b}_t^2)]]. \tag{30}$$

$$+\frac{1}{2|b|}\mathrm{E}[\mathrm{E}_x[(y_t'-y_t)(e_t^\mathsf{T}\tilde{x}_t-e_x^\mathsf{T}\tilde{x}_t)]] \tag{31}$$

$$+\frac{1}{|b|}\mathrm{E}[\mathrm{E}_x[y_tp_{t,g}^u(\hat{b}_t-b)]]. \tag{32}$$

First, we will bound (29). Define $M_t = I_{m+1} + \sum_{s=1}^{t} \tilde{x}_s \tilde{x}_s^\intercal$. The closed form expression for the estimator e_t at period t is $(M_t)^{-1}(\sum_{s=1}^{t-1} y_s \tilde{x}_s)$. Expanding the expressions for e_t in (29), we see that most of the terms in the expansion are telescoping, giving

$$\sum_{t=1}^{T} \mathrm{E}_x[(e_t^\intercal \tilde{x}_t - y_t')^2 - (e_x^\intercal \tilde{x}_t - y_t')^2] = \sum_{t=1}^{T} \mathrm{E}_x[(y_t')^2 \tilde{x}_t^\intercal M_t^{-1} \tilde{x}_t] \tag{33}$$

$$+ \mathrm{E}_x[\|e_x - e_1\|^2 - (e_x - e_{T+1})^\intercal M_T (e_x - e_{T+1})] \tag{34}$$

$$+ \mathrm{E}_x[(e_1^\intercal \tilde{x}_1)^2 + (e_x^\intercal \tilde{x}_{T+1})^2] \tag{35}$$

$$- \sum_{t=1}^{T} \mathrm{E}_x[(e_{t+1}^\intercal \tilde{x}_{t+1})^2 \tilde{x}_{t+1}^\intercal M_t^{-1} \tilde{x}_{t+1} - (e_{T+1}^\intercal \tilde{x}_{T+1})^2 - (e_x^\intercal \tilde{x}_1)^2].$$

Since $e_1 = (I + \tilde{x}_1 \tilde{x}_1^\intercal)^{-1} \cdot 0 = 0$, $\|e_x - e_1\|^2 = \|e_x\|^2$. Then since M_t is positive semi-definite for all t, (34) is upper bounded by $\|e_x\|^2 \le \bar{a}^2 + \bar{c}^2$. Since $e_1 = 0$ and x_{T+1} can be set to 0, (35) is 0. The final line is upper bounded by 0.

Finally, to bound Eq (33), we can write

$$\begin{aligned}\sum_{t=1}^{T} \mathrm{E}_x[(y_t')^2 \tilde{x}_t^\intercal M_t^{-1} \tilde{x}_t] &= \sum_{t=1}^{T} \mathrm{E}_x[(f(x_t) + \epsilon_t + (b - \hat{b}_t)p_t)^2 \tilde{x}_t^\intercal M_t^{-1} \tilde{x}_t] \\ &\le \sum_{t=1}^{T} (2\sigma^2 + 2\bar{f}^2 + 4\bar{b}^2 p_{\max}^2) \tilde{x}_t^\intercal M_t^{-1} \tilde{x}_t.\end{aligned}$$

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

$$x^\intercal (\Sigma + xx^\intercal)^{-1} x = \frac{\det(\Sigma)}{\det(\Sigma + xx^\intercal)}$$

for any matrix Σ, we have

$$\begin{aligned}(2\sigma^2 + 2\bar{f}^2 + 4\bar{b}^2 p_{\max}^2) \sum_{t=1}^{T} \tilde{x}_t^\intercal M_t^{-1} \tilde{x}_t &\le (2\sigma^2 + 2\bar{f}^2 + 4\bar{b}^2 p_{\max}^2) \sum_{t=1}^{T} 1 - \frac{\det(M_{t-1})}{\det(M_t)} \\ &\le (2\sigma^2 + 2\bar{f}^2 + 4\bar{b}^2 p_{\max}^2) \sum_{k=1}^{m+1} \log(1 + \lambda_k),\end{aligned}$$

where the λ_js are the eigenvalues of $\sum_{t=1}^{T} \tilde{x}_t \tilde{x}_t^\intercal$. The sum of the λ_js is at most $T \cdot \max_t \|\tilde{x}_t\|^2$, which in turn is at most $\sqrt{m+1}T$. Thus the last line is $O((m+1) \cdot \log(T(m+1)))$. Then (29) does not dominate the regret bound.

Now we will bound Eq (30). Using the definition $p_{g,t}^u = -\frac{e_t^T \tilde{x}_t}{2b_t}$ and the fact that $|b_t| \ge \underline{b}$ gives

$$\begin{aligned}\mathrm{E}_x[(p_{t,g}^u)^2 (b^2 - \hat{b}_t^2)] &\le \frac{1}{2\underline{b}} \mathrm{E}_x[(e_t^T \tilde{x}_t)^2 (b^2 - \hat{b}_t^2)] \\ &\le \frac{1}{\underline{b}} \mathrm{E}_x[((e_t^T \tilde{x}_t - e_x^T \tilde{x}_t)^2 + (e_x^T \tilde{x}_t)^2)(b^2 - \hat{b}_t^2)] \\ &\le \frac{1}{\underline{b}} ((2\bar{b}^2) \mathrm{E}_x[(e_t^T \tilde{x}_t - e_x^T \tilde{x}_t)^2] + (\bar{a} + \bar{c})^2 \mathrm{E}_x[b^2 - \hat{b}_t^2]).\end{aligned} \tag{36}$$

The second line follows from the identity $(x+y)^2 \le 2x^2 + 2y^2$. The third line follows from the fact that $b^2 + \hat{b}_t^2 \le 2\bar{b}^2$ due to the assumptions on b and the projection step in the algorithm, as well as from the assumptions on e_x. Now, to bound $\mathrm{E}_x[(e_t^T \tilde{x}_t - e_x^T \tilde{x}_t)^2]$ in Eq (36), note that we have

$$\sum_{t=1}^{T} \mathrm{E}_x[(e_t^T \tilde{x}_t - e_x^T \tilde{x}_t)^2] = \sum_{t=1}^{T} \mathrm{E}_x[(e_x^T \tilde{x}_t - y_t)^2 - (e_t^T \tilde{x}_t - y_t)^2]. \tag{37}$$

This is because

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

where the second line follows from the fact that $y_t = f(x_t) + \epsilon_t$ and $\mathrm{E}_x[\epsilon_t] = 0$, ϵ_t independent of x_t, e_x, e_t, and the final line follows from the first order conditions of the minimization problem Eq (10), as given by Eq (11). Eq (37) thus implies that $\sum_{t=1}^{T} \mathrm{E}_x[(e_t^T\tilde{x}_t - e_x^T\tilde{x}_t)^2]$ is $O((m+1)\log(T(m+1))$.

To bound $\mathrm{E}_x[b^2 - \hat{b}_t^2])$ in Eq (36), we can write

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

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 1. Then, $\sum_{t=1}^{T} \mathrm{E}_x([b^2 - \hat{b}_t^2]) \le \frac{32}{3}\bar{b}\frac{\sqrt{\bar{f}^2+\sigma^2+\bar{b}^2p_{\max}^2}}{\delta}T^{2/3}$, which dominates the $O((m+1\log(T(m+1))$ term $\sum_{t=1}^{T} \mathrm{E}_x[(e_t^T\tilde{x}_t - e_x^T\tilde{x}_t)^2]$, and implies that Eq (30) is $O(T^{2/3})$.

Similar ideas can be used to bound Eq (31) and (32). For Eq (31), using the identity $y_t' - y_t = (b - \hat{b}_t)p_t$, we have

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

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 (37) again, we see that Eq (31) is $O(\sqrt{(m+1)T}\log(T(m+1)))$.

Finally, each term of Eq (32) can be written as

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

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 $\sum_{t=1}^{T} \mathrm{E}[|e_t^T \tilde{x}_t - e^T \tilde{x}_t|]$ is $O((m+1\log(T(m+1))$, and that $\sum_{t=1}^{T} \mathrm{E}_x[|\hat{b}_t - b|]$ is $O(T^{2/3})$. Then Eq (32) 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 factors, the expected regret of the RPS algorithm over T periods can be bounded by

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

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

C.7. Lemmas

LEMMA 1 **(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*
$$\mathrm{E}[(\hat{b}_t - b)^2] \leq 12 \cdot \frac{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}{\delta^2} \cdot \frac{1}{\sqrt{t}}.$$
- *When Algorithm 2 is applied to the price ladder setting, for $t \geq 2$, we have*
$$\mathrm{E}[(\hat{b}_t - b)^2] \leq 4 \cdot \frac{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}{\underline{\delta}^2} \cdot \frac{1}{t^{2/3}}.$$
- *When Algorithm 3 is applied to the non IID setting, for $t \geq 4$ we have*
$$\mathrm{E}[(\hat{b}_t - b)^2] \leq 12 \cdot \frac{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}{\delta^2} \cdot \frac{1}{t^{2/3}}.$$

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

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

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 x_s, and have mean 0, so

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s (f(x_t)+\epsilon_s+bp_{g,s})}{\sum_{s=1}^{t-1} \frac{\delta^2}{4} s^{-2\alpha_1}}\right)^2\right] &= \mathrm{E}\left[\frac{\sum_{s=1}^{t-1} \Delta p_s^2 (f(x_s)+\epsilon_s+bp_{g,s})^2}{(\sum_{s=1}^{t-1} \frac{\delta^2}{4} s^{-2\alpha_1})^2}\right] \\
&\le \mathrm{E}\left[\frac{\sum_{s=1}^{t-1} 3\Delta p_s^2 (f(x_s)^2+\epsilon_t^2+b^2 p_{g,j}^2)}{(\sum_{s=1}^{t-1} \frac{\delta^2}{4} s^{-2\alpha_1})^2}\right] \\
&\le 12 \cdot \frac{\bar{f}^2+\sigma^2+\bar{b}^2 p_{\max}^2}{\delta^2} \cdot \frac{1}{\sum_{s=1}^{t-1} s^{-2\alpha_1}} \qquad (38)
\end{aligned}$$

We used the fact that $(x+y+z)^2 \le 3(x^2+y^2+z^2)$. 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 $f(x_s), b, p_{g,s}$ are bounded.

Now consider the **price ladder setting**. Using the definitions of b_t in Algorithm 2, $\hat{b}_t = \mathrm{Proj}(\hat{b}_t^u, B)$, 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

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

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 x_s, and have mean 0, so

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s (f(x_t)+\epsilon_s+bp_{g,s})}{\sum_{s=1}^{t-1} (q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1}}\right)^2\right] &= \mathrm{E}\left[\frac{\sum_{s=1}^{t-1} \Delta p_s^2 (f(x_s)+\epsilon_s+bp_{g,s})^2}{(\sum_{s=1}^{t-1} (q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1})^2}\right] \\
&\le \mathrm{E}\left[\frac{\sum_{s=1}^{t-1} 3\Delta p_s^2 (f(x_s)^2+\epsilon_t^2+b^2 p_{g,j}^2)}{(\sum_{s=1}^{t-1} (q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1})^2}\right] \\
&\le 3 \cdot \mathrm{E}\left[\frac{\bar{f}^2+\sigma^2+\bar{b}^2 p_{\max}^2}{\sum_{s=1}^{t-1} (q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1}}\right] \\
&\le 3 \cdot \frac{\bar{f}^2+\sigma^2+\bar{b}^2 p_{\max}^2}{\underline{\delta}^2} \cdot \frac{1}{\sum_{s=1}^{t-1} s^{-2\alpha_1}}. \qquad (39)
\end{aligned}$$

We used the fact that $(x+y+z)^2 \le 3(x^2+y^2+z^2)$. 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 $f(x_s), b, p_{g,s}$ are bounded. The last step uses the assumption that $q_i - q_{i-1} \ge \underline{\delta}$ for $i = 1, \ldots, N+1$.

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

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

and we have for $t \geq 4$ that

$$\frac{1}{\sum_{s=1}^{t-1} s^{-2\alpha_1}} \leq 2(1-2\alpha_1)t^{2\alpha_1-1}. \tag{40}$$

Substituting (40) into (38) and (39) respectively, we prove the lemma in the IID, price ladder and non IID settings. □

LEMMA 2 **(Bound on $\|QM_t^{-1}Q\|_2$).** *Let* $M = \mathrm{E}[\tilde{x}\tilde{x}^\mathsf{T}]$, $V = \mathrm{E}[(Q^{-1}\tilde{x}\tilde{x}^\mathsf{T}Q^{-1} - I)^2]$ *and* $M_t = \frac{1}{t-1}\sum_{s=1}^{t-1}\tilde{x}_s\tilde{x}_s^\mathsf{T}$. *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. For any $s = 1, \ldots, t-1$, we have $\mathrm{E}[I - Q^{-1}\tilde{x}_s\tilde{x}_s^\mathsf{T}Q^{-1}] = 0$, where I is the identity matrix. In addition, for an arbitrary matrix A, it holds that $\|A\|_2 \leq \|A\|_F$, so by $\|\tilde{x}_s\|_\infty \leq 1$, we have

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

Note that we used the submultiplicative property of the spectral norm. Since $\{\tilde{x}_s\}$ are independent and identically distributed, we apply the matrix Bernstein bound (Lemma 4) with $\alpha = (t-1)/2$ to yield

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

By an identical argument, we also have

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

Thus we have

$$\begin{aligned}
\mathbb{P}[\|I - Q^{-1}M_tQ^{-1}\|_2 > \frac{1}{2}] &= \mathbb{P}[\max\{\lambda_{\max}(I - Q^{-1}M_tQ^{-1}), \lambda_{\max}(Q^{-1}M_tQ^{-1} - I)\} > \frac{1}{2}] \\
&\leq 2(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right).
\end{aligned} \tag{41}$$

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 (41), 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 $Q^{-1}M_tQ^{-1} = Q^{-1}\frac{\sum_{s=1}^{t-1}\tilde{x}_s\tilde{x}_s^\mathsf{T}}{t-1}Q^{-1}$ 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 3 **(Optimal Policy Structure for Linear Demand)**. *Suppose the true demand function is linear, given by*

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

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

Proof. Suppose the seller uses a pricing policy $\pi(\mathcal{H}_{t-1}, x_t) = \pi_t(x_t)$ at period t, where function $\pi_t(\cdot)$ is measurable with respect to t and could be nonlinear. We denote by $\tilde{\mathbb{E}}[\cdot]$ 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:

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

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

$$\tilde{\mathbb{E}}[\pi_t(x_t) - S - U^\mathsf{T} x_t] = 0, \quad \tilde{\mathbb{E}}\left[x_t\left(\pi_t(x_t) - S - U^\mathsf{T} x_t\right)\right] = 0. \tag{42}$$

Now, let us compare the conditional expected revenue of price $\pi_t(x_t)$ and price $S + U^\mathsf{T} x_t$. We have

$$\begin{aligned}
&\tilde{\mathbb{E}}\left[\pi_t(x_t) D_t(\pi_t(x_t)) - (S + U^\mathsf{T} x_t) D_t(S + U^\mathsf{T} x_t)\right] \\
=&\tilde{\mathbb{E}}\left[\pi_t(x_t)\cdot\left(a + b\pi_t(x_t) + c^\mathsf{T} x_t\right) - (S + U^\mathsf{T} x_t)(a + b\cdot(S + U^\mathsf{T} x_t) + c^\mathsf{T} x_t)\right] \\
=&b\tilde{\mathbb{E}}\left[(\pi_t(x_t))^2 - (S + U^\mathsf{T} x_t)^2\right] + \tilde{\mathbb{E}}\left[(a + c^\mathsf{T} x_t)(\pi_t(x_t) - S - U^\mathsf{T} x_t)\right] \qquad (43) \\
=&b\tilde{\mathbb{E}}\left[(\pi_t(x_t))^2 - (S + U^\mathsf{T} x_t)^2\right] \\
=&b\left\{\tilde{\mathbb{E}}\left[(\pi_t(x_t) - S - U^\mathsf{T} x_t)^2\right] + 2\tilde{\mathbb{E}}\left[(S + U^\mathsf{T} x_t)(\pi_t(x_t) - S - U^\mathsf{T} x_t)\right]\right\} \qquad (44) \\
=&b\tilde{\mathbb{E}}\left[(\pi_t(x_t) - S - U^\mathsf{T} x_t)^2\right] \le 0.
\end{aligned}$$

The second term of Eq (43) and the second term of Eq (44) are both zero because of the first order condition Eq (42). In the last step, recall that the price sensitivity parameter $b < 0$.

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

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

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

LEMMA 4 **(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*

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

then for all $t \ge 0$,

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

References

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