Dynamic Learning and Price Optimization with Endogeneity Effect

Mila Nambiar* David Simchi-Levi† He Wang‡

November 9, 2016

Abstract

In the demand estimation process, it is widely known that price endogeneity, or the correlation between price and demand noise, can cause biased estimation of price elasticity. We study a dynamic pricing model with contextual information and show that price endogeneity can arise in this setting, leading to incorrect estimation of model parameters and potentially suboptimal pricing decisions. To address the endogeneity effect we propose a "Random Price Shock" (RPS) algorithm, which dynamically generates independent price shocks to estimate price elasticity and maximizes revenue while controlling for the endogeneity effect. We show that the RPS algorithm has strong theoretical and numerical performance, and is robust to model misspecification.

Keywords: revenue management; pricing; parameter estimation; endogeneity; misspecification.

1 Introduction

With the growing availability of data in many revenue management applications, there has been significant interest recently in developing dynamic pricing algorithms for a data-rich environment. In such an environment, a seller observes some contextual information represented by a feature vector at the beginning of each time period; demand in this period is a function of both price and the observed feature vector. This is an extension of the traditional dynamic pricing setting where demand is only a function of price. By including additional information into the demand model, firms may obtain more accurate demand forecast and achieve higher revenue.

However, a pitfall when introducing feature vectors into the demand model is that it may lead to *price endogeneity* — an effect that occurs when price is correlated with demand noise. It is well known that price endogeneity can cause a problem when estimating the relationship between price and demand from historical data: if a demand estimation method ignores the endogeneity effect, it may produce a biased and inconsistent estimate of price elasticity, leading to ineffective pricing decisions.

The phenomenon of price endogeneity is widely studied in economics, marketing, and operations management literature. Empirical studies have found that price endogeneity exists and

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

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

‡Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, wanghe@mit.edu.

is quite significant in many real-world business settings. For example, in a survey of more than 80 marketing publications, [BHP05] found that the average price elasticity after controlling for endogeneity is (-3.74), while the average price elasticity if endogeneity is not accounted for is only (-2.47). In an empirical study of financial loan data, [PŞVR15] reported that controlling for endogeneity would increase the magnitude of price elasticity by 126% to 258%.

Despite its importance in practice, we have found only a few studies that consider price endogeneity in the dynamic pricing literature, where price is treated as a decision variable instead of some observational data. As we will see in Section 2, price endogeneity can arise in a dynamic pricing model with features, if either of the following scenarios happens: 1) the demand model is misspecified; 2) the admissible price set is correlated with demand noise. Both scenarios are common in practice. An example of the first scenario is when a seller assumes a linear demand model for feature vectors, while the true model is nonlinear. An example of the second scenario is when the admissible price sets are correlated with some intangible characteristics such product quality, but such intangible characteristics are not included in the feature vector.

The goal of this paper is to develop a simple and effective dynamic pricing algorithm that correctly estimates price elasticity and achieves good revenue performance when there is a price endogeneity effect. We outline the relevant academic literature in Section 1.1 and provide an overview of the main contributions of this paper in Section 1.2.

1.1 Literature Review

Our paper belongs to the recent line of research in revenue management which assumes demand models contain unknown parameters. In such a model, the sellers faces an exploration-exploitation tradeoff: towards the beginning of the selling season, the seller may offer several different prices to try to learn the unknown parameters; over time, the seller can exploit the parameter estimations to set a price that maximizes revenue. We refer readers to the survey by [dB15] for an overview of different problem settings and solution techniques used in this area.

In particular, our paper focuses on a demand model that contains unknown parameters as well as features. Among related works, [QB16] considers a linear demand model with features, and applies a greedy least squares method to estimate model parameters. [CLPL16] proposes a feature-based pricing algorithm to estimate model parameters when demand is binary. [JN16] studies a problem where feature vector is high dimensional and the demand parameter has some sparsity structure. We note that all three papers cited above assume that the demand model is correctly specified and the admissible price set is fixed over time, which essentially rules out the possibility of price endogeneity. In contrast, we propose a method in this paper for when these assumptions may not hold. [BK16] does consider estimating pricing models with endogeneity effect. However, prices are treated as observational data whereas in this paper, we consider dynamic pricing models with endogeneity effect in which prices are decision variables.

Another related work is [BZ15], which considers model misspecification in dynamic pricing caused by the seller approximating the true demand function by a linear demand function. Our paper complements their work by considering model misspecification on *features* rather than misspecification on *price*.

The econometrics literature has proposed various methods to identify model parameters with endogeneity effect; see e.g. [AP08]. A popular method used in the pricing literature to estimate discrete choice models is proposed by [BLP95]. [TVR06] also provide an overview of these methods with applications in revenue management. Roughly speaking, the existing solutions to price endogeneity can be classified into two main categories:

- *Instrumental variables*: This is the standard method used in empirical studies to estimate models with endogeneity effect. The method aims at finding *instrumental variables* that are correlated with price but are uncorrelated with demand noise. For example, [PŞVR15] uses the average price in neighboring regions as an instrumental variable to estimate demand models for auto loan. However, this method has several potential pitfalls: Firstly, the selection of instrumental variables is ad hoc and depends on specific applications and the availability of data. Secondly, an instrument that seems uncorrelated with demand can in fact be correlated with demand in unexpected ways. As a result, the instrumental variables method can be difficult to use for sellers who lack expertise in it.
- *Randomized trials*: This method offers prices that are randomly generated, and is often used in field experiments. For example, [FGL16] applies randomized prices in a field experiment to estimate an online retailer's demand model. Downsides of this method include the time-consuming nature of field experiments, and the fact that many retailers cannot afford to lose revenue by offering random prices.

To summarize, most existing solutions to price endogeneity are not well-suited for dynamic pricing. These methods are either time-consuming and costly like randomized trials, or complex to implement like instrumental variables method.

1.2 Overview of Contributions

To overcome the drawbacks in the existing solutions to price endogeneity, we propose in this paper a "random price shock" (RPS) algorithm. The RPS algorithm can provide an unbiased and consistent estimate of the relationship between demand and price while controlling for the endogeneity effect. The estimation process is simple and scalable, and prices are optimized and updated on the fly as demand parameters are being estimated.

The key idea of the RPS algorithm is adding random price perturbations to the original prices recommended by some price optimization model. The variances of these price perturbations are specified by the algorithm to balance the exploration-exploitation tradeoff. Intuitively, using a larger variance can help explore the demand function under different prices, while using a smaller variance can generate a price that is closer to the one that exploits the current parameter estimation.

In fact, the RPS algorithm can be viewed as a hybrid of randomized trials and instrumental variable method: it uses external randomization as in randomized trials, while the randomly generated price perturbation serves as a valid instrumental variable. Despite these similarities, the RPS algorithm enjoys two major advantages over the existing methods: Firstly, unlike randomized trials, the RPS algorithm does not require a separate period devoted to field experiments, as it performs price experimentation on the fly during revenue maximization. Secondly, unlike the instrumental variables method, RPS algorithm does not require a manual process to choose appropriate instruments or to collect data for these instruments.

In our model, defined in Section 2, we assume that the seller uses a *linear* demand function in both price and feature vectors. Under the assumption that the true demand model is also linear, Section 3 shows that the RPS algorithm identifies the correct model parameters in the presence of price endogeneity, and has an expected regret of $O(\sqrt{N})$ compared to a clairvoyant who knows the true model a priori, where N is the number of time periods. In the case where the true model is nonlinear, Section 4 shows the RPS algorithm is robust to model misspecification, and has the same $O(\sqrt{N})$ expected regret compared to a clairvoyant who knows the "best" linear model, i.e. the linear model associated with the highest revenue. Our regret bound matches the best possible lower bound of $\Omega(\sqrt{N})$ by [KZ14], who considers a special case of our model without features or endogeneity. Moreover, RPS algorithm improves the $O(\sqrt{N}\log N)$ upper bound obtained by [KZ14] for this special case. In Section 5, we use numerical experiments to test the performance of the RPS algorithm. We verify that the regret of the algorithm is indeed $O(\sqrt{N})$, and demonstrate that the RPS algorithm significantly outperforms several competing algorithms.

Notations

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 $a_n \le Cb_n$ for all n; we write $a_n = \Omega(b_n)$ is there exists a constant c such that $a_n \ge cb_n$ 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^T and denote its Euclidean norm by $\|x\| := \sqrt{x^T x}$. We let $\|x\|_1$ be the ℓ_1 norm of x, defined as $\|x\|_1 = \sum_i |x_i|$. We let $\|x\|_\infty$ be the ℓ_∞ norm, defined as $\|x\|_\infty = \max_i |x_i|$.

For any square matrix $M \in \mathbb{R}^{k\times k}$, we denote its transpose by M^T, its inverse by M^{-1} and its trace by $\mathrm{tr}(M)$; if M is also symmetric ($M = M^T$), we denote its largest eigenvalue by $\lambda_{\max}(M)$ and its smallest eigenvalue by $\lambda_{\min}(M)$. We let $\|M\|_2$ be the spectral norm of matrix M, defined by $\|M\|_2 = \sqrt{\lambda_{\max}(M^T M)}$. We denote the Frobenius norm of M by $\|M\|_F$. Recall that $\|M\|_F = \sqrt{\mathrm{tr}(M^T M)}$.

2 Model

We consider a firm selling heterogeneous products over a finite selling horizon. In each time period ($i = 1, 2, \ldots, N$), a single product indexed by i is offered. At the beginning of period i, the seller observes a feature vector, $x_i \in \mathbb{R}^m$, which represents the characteristics of product i and other available information that may affect demand, such as a competitor's price or a customer's preference. We assume that feature vectors x_i are sampled independently from a fixed but unknown distribution on $[-1, 1]^m$. Moreover, we assume that the matrix

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

is positive definite.[1]

Given feature x_i, we assume that the demand for period i as a function of price p is given

[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.

by:

$$D_i(p) = bp + f(x_i) + \epsilon_i, \quad \forall p \in [\underline{p}_i, \bar{p}_i]. \tag{1}$$

Here, parameter b is a fixed constant, and $f : \mathbb{R}^m \to \mathbb{R}$ is a fixed function. Both b and f are *unknown* to the seller. We assume that the demand function is strictly decreasing in p (i.e. $b < 0$), and $f(x_i)$ is bounded for all x_i by $|f(x_i)| \leq \bar{f}$. The latter assumption would follow immediately from the fact that the set of all x_i is compact if f were continuous. The last term ϵ_i represents a demand noise with $\mathrm{E}[\epsilon_i \mid x_i] = 0$. We assume that ϵ_i has bounded second moment ($\mathrm{E}[\epsilon_i^2] \leq \sigma^2, \forall i$), and is independent of (x_j, ϵ_j) for all $j \leq i-1$. 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 the feature vector.

We denote the admissible price range for product i, i.e. the range of prices from which the price of product i must be chosen, by $[\underline{p}_i, \bar{p}_i]$. The price range is determined by the seller's business constraints. For example, if the seller wants to keep its price lower than the price of a competitor, we can let $\bar{p}_i$ be the competitor's price as observed at the beginning of period i. We assume that $\underline{p}_i$ and $\bar{p}_i$ are exogenous random variables, and allow them to have arbitrary correlation with x_i and ϵ_i. We also assume there exists a constant upper bound $p_{\max}$ such that $\bar{p}_i \leq p_{\max}$ for all i.

In the paper, we consider a seller who uses a linear demand function to approximate the true demand function in Eq (1), although the true demand function can be nonlinear. In other words, the linear model may be *misspecified.* We consider two settings. In Section 3, we further assume that the model is correctly specified, that is, f is an affine function of x_i:

$$f(x_i) = a + c^T x_i. \tag{2}$$

We can then rewrite Eq (1) as

$$D_i(p) = a + bp + c^T x_i + \epsilon_i, \quad \forall p \in [\underline{p}_i, \bar{p}_i]. \tag{3}$$

Given x_i, we denote the optimal price associated with model (3) by $p_i^* = -\frac{a+c^T x_i}{2b}$. We assume that $p_i^* \in [\underline{p}_i, \bar{p}_i]$ for all i.

In Section 4, we relax the assumption of Eq (2) and assume f is an arbitrary function. We define the following linear demand model as a benchmark:

$$D_i(p) = a + bp + c^T x_i + \nu_i, \quad \forall p \in [\underline{p}_i, \bar{p}_i], \tag{4}$$

where a, c are least squares estimates of $f(x_i)$: $a, c = \arg\min_{a', c'} \mathrm{E}[\|f(x_i) - (a' + c'^T x_i)\|^2]$, and $\nu_i = f(x_i) - (a + c^T x_i) + \epsilon_i$. We will give further justification in Section 4 of why the least squares linear model is an appropriate benchmark to consider: we show that among all linear approximations of the true demand model in Eq (1), the optimal price of the linear model in Eq (4) has the highest expected revenue.

Given x_i, we denote the optimal price for the best linear model by $p_i^* = -\frac{a+c^T x_i}{2b}$, and the optimal price for the true demand model by $\tilde{p}_i = -\frac{f(x_i)}{2b}$. We assume that $\tilde{p}_i \in [\underline{p}_i, \bar{p}_i]$ for all i.

Example: Flash Sales in E-commerce

An application of the model defined above is price optimization for flash sales. In a flash sale event, an e-commerce retailer offers a product at a discounted price for a short period of time, typically ranging from a few hours to a few days. The discount and the time limit entice consumers to buy the product on the spot. Once the event is over, the product becomes unavailable, and the seller launches new flash sale events for other products. E-commerce stores using flash sales include Gilt Groupe and Rue La La, which offer high-end fashion products at discounted prices through limited time events. [McK15] reported that the flash sales industry was worth 3.8 billion USD in 2015, with an annual growth rate of approximately 17%.

In our model framework, each flash sale event can be represented as a time period. The feature vector x_i represents characteristics of the product offered in the i^{th} event. These characteristics could include brand, category, style, as well as other information like seasonality factors. The admissible price range $[\underline{p}_i, \bar{p}_i]$ is exogenously given according to the retailer's business constraints. For example, the retailer may set $\underline{p}_i$ as the product's purchasing cost, and set $\bar{p}_i$ as the product's regular price. Based on the feature vector, the retailer needs to determine a discounted price for the product within the price range.

2.1 Demand Learning and Pricing Policies

We assume that in both the linear demand and quasi-linear demand cases the true values of the parameters (a, b, c) are unknown to the seller when the selling season starts. The seller only knows that the parameters belong to some compact sets: $|a| \le \bar{a}$, $0 < \underline{b} \le |b| \le \bar{b}$, $\|c\|_1 \le \bar{c}$, where $\bar{a}, \bar{b}, \bar{c}$ are known to the seller.[2] The seller's goal is to estimate the values of (a, b, c), and select a price $p_i \in [\underline{p}_i, \bar{p}_i]$ for each period $i = 1, 2, \ldots, N$ sequentially in order to maximize her total expected revenue.

We denote the realized demand for product i by

$$d_i := D_i(p_i) = bp_i + f(x_i) + \epsilon_i,$$

and the history up to period $i-1$ is denoted by

$$\mathcal{H}_{i-1} = \{(x_1, p_1, d_1), \ldots, (x_{i-1}, p_{i-1}, d_{i-1})\}.$$

We say that $\pi(\cdot)$ is an admissible pricing policy if for any i, price p_i is a function of the realized history, the feature vector observed at the beginning of the current period, and the feasible price range, i.e., $p_i = \pi(\mathcal{H}_{i-1}, x_i, \underline{p}_i, \bar{p}_i)$.

The regret of any admissible pricing policy over a selling horizon of length N is defined as the difference in the expected revenue of a clairvoyant who knows the true demand parameters a, b, c and the expected revenue of certain pricing policy over the selling horizon. Mathematically, the

[2]Here, we use the ℓ_1 norm to bound the parameter c, because this guarantees that the mean demand is always bounded by a constant independent of the dimension of feature vector, m, namely $a + c^T x_i \le |a| + \|c\|_1 \|x_i\|_\infty \le \bar{a} + \bar{c}$. Intuitively, there should be an upper limit on the mean demand regardless of how much information is used for demand prediction.

expected regret is given by

$$\text{Regret}(N) = \sum_{i=1}^{N} \mathrm{E}[p_i^* D(p_i^*)] - \sum_{i=1}^{N} \mathrm{E}[p_i D(p_i)], \tag{5}$$

where the expectation is over all random quantities including features x_i, price ranges $[\underline{p}_i, \bar{p}_i]$, and the demand noise ϵ_i.

2.2 Causes of Price Endogeneity

Generally speaking, price endogeneity means that price and demand noise are correlated. For the linear model in Eq (3) specifically, endogeneity means that $\mathrm{E}[p_i \epsilon_i] \neq 0$ for some i. For the quasi-linear case in Eq (4), endogeneity means that $\mathrm{E}[p_i \nu_i] \neq 0$ for some i.

The reason why price endogeneity can arise in our model setting is that for a given pricing policy π, we have

$$p_i = \pi(\mathcal{H}_{i-1}, x_i, \underline{p}_i, \bar{p}_i).$$

Here, $\mathcal{H}_{i-1}$ is the realized history, x_i is the feature vector, and $\underline{p}_i, \bar{p}_i$ are boundaries for the feasible price range. All of these variables are random quantities, so they can be correlated with the demand noise ϵ_i (or ν_i); since p_i is a function of these random quantities, p_i is a random variable and may also be correlated with ϵ_i (or ν_i).

It is well known that in the presence of price endogeneity, common estimation methods such as ordinary least squares (OLS) would give biased and inconsistent parameter estimates. In the numerical experiments in Section 5, we show that when $i \to \infty$, the estimate of b generally does not converge to its true value if the endogeneity effect is not correctly addressed.

In our model setting, price endogeneity can arise from the following sources:

- *Hidden information in price set*: The admissible sets $[\underline{p}_i, \bar{p}_i]$ specified by the seller may contain some hidden information about demand that is not included in the feature vector x_i. For instance, in the earlier example of flash sales in e-commerce, the feature vector x_i represents the observable characteristics of the product offered in period i. However, the product may have some intangible properties such as design and quality that are difficult to characterize and thus not included in the feature vector. This is often the case for high-end fashion products sold by flash sale retailers like Gilt and Rue La La, where customers are sensitive to product design and quality, so these intangible properties can have a big impact on customer demand. Meanwhile, the price range $[\underline{p}_i, \bar{p}_i]$ is often influenced by these intangible properties such as product quality. This leads to a correlation between the demand noise ϵ_i and $\underline{p}_i, \bar{p}_i$ given feature x_i. Because price p_i must be chosen in the range $[\underline{p}_i, \bar{p}_i]$, it can be correlated with ϵ_i as a result.

  One tentative approach to avoid hidden information in the price set is to include the price bounds $\underline{p}_i$ and $\bar{p}_i$ in the feature vector x_i. This means that given x_i, the price boundaries $\underline{p}_i$ and $\bar{p}_i$ are deterministic, so they are uncorrelated with the noise ϵ_i. However, as we will see, this approach does not resolve the endogeneity issue, as it may cause a new problem: model misspecification.

- *Model misspecification*: The second source of endogeneity arises when the model is misspecified. We assume that the seller uses a linear demand model to estimate demand and optimize price. Although linear models are often used in practice [KZ14, BZ15], the true model may not be linear, as information in the feature vector can affect demand in a complex way. For example, if the price bounds $\underline{p}_i$ and $\bar{p}_i$ are included in the feature vector as proxies of the intangible features, we cannot guarantee that these intangible features (and thus $\underline{p}_i$ and $\bar{p}_i$) would affect demand in a linear fashion.

In the demand model given by Eq (4), we have $\nu_i = f(x_i) - (a + c^T x_i) + \epsilon_i$, i.e., the error term ν_i contains both the approximation error $f(x_i) - (a + c^T x_i)$ and a random noise ϵ_i. If the model is misspecified, then $f(x_i) - (a + c^T x_i) \neq 0$, so ν_i is not mean independent of x_i, namely $\mathrm{E}[\nu_i \mid x_i] \neq 0$. When the seller chooses a price p_i based on feature x_i, we may get $\mathrm{E}[\nu_i p_i \mid x_i] \neq 0$, which in turn leads to price endogeneity with $\mathrm{E}[\nu_i p_i] \neq 0$.[3] In Section 5, we use numerical experiments to verify that if price endogeneity is not addressed in a misspecified model, the estimated parameters would be biased and would not converge to the true parameters.

3 Random Price Shock Algorithm

In this section, we propose a dynamic pricing algorithm that solves the endogeneity problem by using randomized price shocks. We refer to it as the Random Price Shock (RPS) algorithm. As the number of periods (N) grows, the parameters estimated by the RPS algorithm are guaranteed to converge to the true parameters. Therefore, the prices chosen by the algorithm will also converge to the optimal prices.

The idea behind the RPS algorithm is to add a random price shock to the greedy price obtained from the current estimates of the parameters. The price shock is generated independently of the feature vector and the demand noise, allowing us to avoid the bias in the estimation process introduced by price endogeneity.

The RPS algorithm is presented in Algorithm 1. We define the projection operator as

$$\mathrm{Proj}(x, S) = \arg\min_{x' \in S} \|x - x'\|.$$

The RPS algorithm starts each iteration by checking if the feasible price range in this period is large enough for price perturbation. If not, an arbitrary price is selected. Otherwise, the algorithm computes the greedy price, $p_{g,i}$, and adds it to a random price shock, Δp_i. Note that the greedy price is projected to the interval $[\underline{p}_i + \delta_i, \bar{p}_i - \delta_i]$, so that the sum of greedy price and price shock is always in the feasible price range $[\underline{p}_i, \bar{p}_i]$.

After the demand in this period is observed, the algorithm updates parameter estimations by using a two-step regression method. First, the price parameter b is estimated by applying linear regression for d_i against Δp_i. It is important to note that we cannot estimate b by regressing d_i against the actual price p_i, since p_i may be endogenous and correlated with demand noise. The second step estimates the remaining parameters, a and c. Since we know that the true parameters belong to some compact sets (A, B and C), the least squares estimators are projected to these sets to improve estimation accuracy.

[3] Although ν_i is not mean independent of x_i, it is in fact mean zero and *uncorrelated* with x_i, namely $\mathrm{E}[\nu_i] = \mathrm{E}[\nu_i x_i] = 0$. This follows from the first order conditions of the least squares method. Should p_i be an affine function of x_i, i.e. $p_i = \alpha + \beta^T x_i$, we would have $\mathrm{E}[\nu_i p_i] = 0$ as well. However, since p_i is generally not an affine function of x_i in the presence of price range bounds $p_i \in [\underline{p}_i, \bar{p}_i]$, we do not have $\mathrm{E}[p_i \nu_i]$ in general.

Algorithm 1 Random Price Shock (RPS) algorithm.

input: parameter bounds $A = [-\bar{a}, \bar{a}]$, $B = [-\bar{b}, -\underline{b}]$, $C = \{c' \in \mathbb{R}^m : \sum_{k=1}^m |c'_k| \leq \bar{c}\}$
initialize: choose $\hat{a}_1, \hat{b}_1, \hat{c}_1$ to be arbitrary parameters in A, B and C
for $i = 1, \ldots, N$ **do**
  set $\delta_i \leftarrow i^{-\frac{1}{4}}$
  if $\bar{p}_i - \underline{p}_i \geq 2\delta_i$ **then**
    given x_i, set unconstrained greedy price: $p^u_{g,i} \leftarrow -\frac{\hat{a}_i + \hat{c}_i^T x_i}{2\hat{b}_i}$
    project greedy price: $p_{g,i} \leftarrow \text{Proj}(p^u_{g,i}, [\underline{p}_i + \delta_i, \bar{p}_i - \delta_i])$
    generate an independent random variable $\Delta p_i \leftarrow i^{-\frac{1}{4}}$ w.p. $\frac{1}{2}$ and $\Delta p_i \leftarrow -i^{-\frac{1}{4}}$ w.p. $\frac{1}{2}$
    set price $p_i \leftarrow p_{g,i} + \Delta p_i$
  else $(\bar{p}_i - \underline{p}_i < 2\delta_i)$
    choose an arbitrary price $p_i \in [\underline{p}_i, \bar{p}_i]$
  end if
  observe demand $d_i = D_i(p_i)$
  set $\hat{b}_{i+1} \leftarrow \text{Proj}(\frac{\sum_{j=1}^i \Delta p_j d_j}{\sum_{j=1}^i \Delta p_j^2}, B)$
  set $(\hat{a}_{i+1}, \hat{c}_{i+1}) \leftarrow \arg\min_{\alpha \in A, \gamma \in C} \sum_{j=1}^i (d_j - \hat{b}_{i+1} p_j - \alpha - \gamma^T x_j)^2$
end for

In the RPS algorithm, the variance of the price shock introduced at each time period (Δp_i) is an important tuning parameter. Intuitively, choosing a large variance of Δp_i means generating larger price perturbations, which can help us learn demand more quickly; choosing a small variance means that the price offered would be closer to the greedy price, which could earn more revenue if the greedy price is close to the optimal price. This is the classical "exploration-exploitation" tradeoff faced by many dynamic learning problems. In Algorithm 1, the variances of the price shocks are specifically chosen to balance the exploration-exploitation tradeoff and control the performance of the algorithm.

Lastly, one may ask why our algorithm considers the endogeneity of price p_i but ignores possible endogeneity in feature vector x_i. The reason is that p_i is a decision variable in our model, so it is critical to establish a causal relationship between demand and price p_i. On the other hand, the feature vector x_i is an exogenous random vector in our setting, thus it is not necessary to determine the causal relationship between demand and feature x_i for the sake of pricing, although such relationship may be of interest for other purposes.

3.1 Theoretical Analysis for Linear Demand

We first consider the case where the linear model is correctly specified (cf Eq (3)). Under the assumptions outlined in Section 2, we have the following upper bound on the regret of the RPS algorithm (Algorithm 1):

Theorem 1. *Under the linear model specified in* (3), *the regret of Algorithm 1 over N periods is bounded by* $O\left(\frac{1+m}{\lambda_{\min}(M)}\sqrt{N}\right)$.

Theorem 1 expresses the upper bound on regret in terms of the horizon length N, the dimension of features m, and the minimum eigenvalue of matrix $M = \mathrm{E}[(1, x_i)(1, x_i)^T]$, while the constant factor within the big O notation only depends on model parameters $\bar{a}, \underline{b}, \bar{b}, \bar{c}, \sigma^2$

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_i except through the parameter $\lambda_{\min}(M)$. The reader can refer to the proof for a detailed description of how the regret depends on these parameters. The full proof can be found in the appendix.

The main idea behind the proof 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 are chosen to have variance $\frac{1}{\sqrt{i}}$ at period i, the former part is bounded by $O(\sqrt{N})$. The latter 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 to prove that the estimated parameters converge sufficiently quickly to their true values.

Comparison with the upper bound in [KZ14] In a related paper, [KZ14] considers learning for a linear model without features, corresponding to a special case of our model where $m = 0$ and the price bounds $\underline{p}_i, \bar{p}_i$ are constants. They describe a family of "semi-myopic" pricing policies that ensure the price selected at time period i is both sufficiently deviated from the historical average of prices and sufficiently close to the greedy price. [KZ14] shows that such policies attain a worst case regret of at most $O(\sqrt{N}\log N)$. The result for the RPS algorithm in Theorem 1 thus improves the upper bound in [KZ14] by a factor of $\log N$. In addition, the RPS algorithm can be applied to a broader setting with features and price endogeneity.

A corresponding lower bound The regret $O(\sqrt{N})$ achieved by the RPS algorithm is optimal up to a constant factor, and cannot be improved by any other algorithm in the following sense: For a setting with no features and no endogeneity, [KZ14] proves a lower bound of $\Omega(\sqrt{N})$ on the *worst case* regret; this is equivalent to stating that no admissible pricing policy can achieve a worst case regret lower than $\Omega(\sqrt{N})$, where the worst case regret of a pricing policy is defined as the maximum regret over all possible values of the true parameters in a given compact set. The setting in [KZ14] is a special case of our model, corresponding to the case where $m = 0$ and $\lambda_{\min}(M) = 1$. For this special case, Theorem 1 establishes a worst case regret bound of $O(\sqrt{N})$, whose constant factor depends exclusively on $\bar{a}, \underline{b}, \bar{b}, \bar{c}, \sigma^2$ and $p_{\max}$. Since this bound matches the $\Omega(\sqrt{N})$ lower bound exactly, the theorem by [KZ14] states that no algorithm can achieve a smaller regret bound in N than the RPS algorithm.

4 Extension: Quasi-Linear Demand Model

4.1 Model and Objectives

We now consider the case where the seller assumes linear demand even though the demand model may be misspecified, and applies the RPS algorithm (Algorithm 1). Recall that for product i ($i = 1, \ldots, N$) the true demand function is given by

$$D_i(p) = bp + f(x_i) + \epsilon_i, \tag{6}$$

where $f(\cdot)$ is an arbitrary function.

Similar to the linear case in Section 3, we could compare the performance of our algorithm to that of a clairvoyant who knows the true model parameters, including the function f, a priori.

However, since the seller assumes that demand is an affine function of price and the feature vector, any feasible algorithm would not achieve the optimal revenue of model (6) if $f(x_i)$ is nonlinear.

In order to establish an achievable upper bound on the expected revenue, we consider a clairvoyant who uses the following linear demand function

$$\hat{D}_i(p) = a + bp + c^T x_i, \quad \forall p \in [\underline{p}_i, \bar{p}_i], \tag{7}$$

where a, c are least squares estimates of $f(x_i)$: $a, c = \arg\min_{a',c'} \mathrm{E}[\|f(x_i) - (a' + c'^T x_i)\|^2]$. It can be shown by solving first order conditions that a, c are given by the closed form expression:

$$\begin{bmatrix} a \\ c \end{bmatrix} = \left(\mathrm{E}\left[\begin{bmatrix} 1 & x_i^T \\ x_i & x_i x_i^T \end{bmatrix} \right] \right)^{-1} \mathrm{E}\left[\begin{bmatrix} f(x_i) \\ f(x_i) x_i \end{bmatrix} \right]. \tag{8}$$

Let $p_i^* = -\frac{a+c^T x_i}{2b}$ be the optimal price under the linear model given by Eq (7). The proposition below shows that the linear demand function with least squares parameters gives the highest revenue among all linear demand functions.

Proposition 1. *For any period $i = 1, \ldots, N$, consider price $p_i' = -\frac{\alpha+\gamma^T x_i}{2\beta}$ where α, β, γ are measurable with respect to history $\mathcal{H}_{i-1}$. Then*

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

By Proposition 1, if the seller uses a linear demand model $D_i'(p) = \alpha + \beta p + \gamma^T x_i$ for period i, the expected revenue of its optimal price $p_i' = -\frac{\alpha+\gamma^T x_i}{2\beta}$ is maximized when $p_i' = p_i^*$. Therefore, we can use the revenue under p_i^* as an upper bound benchmark. The expected regret incurred by pricing policy $\{p_i\}$ at the end of the selling horizon is defined as:

$$\text{Expected Regret}(N) = \sum_{i=1}^{N} \mathrm{E}\left[p_i^* D_i(p_i^*)\right] - \sum_{i=1}^{N} \mathrm{E}\left[p_i D_i(p_i)\right].$$

4.2 Theoretical Analysis for Quasi-Linear Demand

Since the RPS algorithm (Algorithm 1) fits the observed demand to a linear model, it can be applied with no modifications to the case where the true demand model is nonlinear but is approximated by a linear model. In fact, it can also be shown that the RPS algorithm incurs the same regret in the case of quasi-linear demand as in the case of linear demand. This result is stated in Theorem 2.

Theorem 2. *Under the quasilinear demand model in Eq* (1)*, the regret of Algorithm 1 over a selling horizon of length N is $O(\frac{1+m}{\lambda_{\min}(M)}\sqrt{N})$.*

The key step in the proof of Theorem 2 is to show that in the quasi-linear demand case, the loss in revenue can, as in the linear demand case, be bounded in terms of the expected difference between the true parameters a, b, c and the estimated parameters. The rest of the proof is then similar to the proof of Theorem 1, and we are able to establish the same regret bound for Theorem 2 as in Theorem 1. The constant factor within the big O notation of Theorem 2 only depends on model parameters $\bar{a}, \underline{b}, \bar{b}, \bar{c}$ and $p_{\max}, \sigma^2$. Although Theorem 1 is a special case of

Theorem 2, the proofs to these two theorems are given separately in Appendix A to enhance readability.

Theorem 2 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 which gives the highest expected revenue among all linear models, i.e, the demand model in Eq (7). The RPS algorithm achieves such robustness because it correctly addresses the price endogeneity effect introduced by misspecified models. In comparison, several existing methods for dynamic pricing with features discussed in the literature review (Section 1.1) do not consider price endogeneity and assume that demand models are exactly specified. In practice, we expect model misspecification to be common, because feature vectors contain various information – such as product characteristics or customer personalization – that can affect demand in a complex way. Therefore, robustness to model misspecification is a useful property for pricing algorithms with feature information.

5 Numerical Results

In this section we present numerical simulations confirming that the RPS algorithm learns the correct parameters of the demand function over the selling horizon, and that the regret incurred by the algorithm matches the theoretical guarantee of $O(\sqrt{N})$. Results in two particular scenarios are presented; In the first example, the demand function is correctly specified, and the source of endogeneity is the correlation between the noise ϵ_i and the price bounds $\underline{p}_i, \bar{p}_i$. In the second scenario, the demand model is misspecified while the price bounds $\underline{p}_i, \bar{p}_i$ are fixed. These examples illustrate the two sources of price endogeneity that we identified in Section 2. Each simulation is run over a selling horizon of length 5000 periods and repeated 200 times.

The simulations compare the performance of the RPS algorithm with the performance of the following three algorithms:

- *Greedy algorithm*: The greedy algorithm (Algorithm 2) 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 studied by [QB16] for a special case of linear model with features, but in general is known to suffer from *imcomplete learning*, i.e., insufficient exploration in price [KZ14].
- *One step regression*: This algorithm introduces randomized price shocks to force price exploration, but uses a one-step regression instead of a two-step regression as in RPS to learn the parameters. A full description of the one-step regression algorithm (Algorithm 3) is given below. The one-step regression algorithm is analogous to the class of semi-myopic algorithms introduced by [KZ14], which use (deterministic) price perturbations to guarantee sufficient exploration. However, Algorithm 3 does not consider the price endogeneity effect 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 0-mean distribution, to be part of the demand noise. Hence this clairvoyant will set prices to be $-\frac{a}{2b}$ at each time period. Such a pricing policy would be optimal in the absence of features but would evi-

dently incur regret linear in N when $m > 0$. This highlights the importance of considering demand covariates in dynamic pricing.

Algorithm 2 Greedy algorithm.

input: parameter bounds $A = [-\bar{a}, \bar{a}]$, $B = [-\bar{b}, -\underline{b}]$, $C = \{c' \in \mathbb{R}^m : \sum_{k=1}^m |c'_k| \le \bar{c}\}$
initialize: choose $\hat{a}_1, \hat{b}_1, \hat{c}_1$ to be arbitrary parameters in A, B and C
for $i = 1, \dots, N$ **do**
  given x_i, set unconstrained greedy price: $p^u_{g,i} \leftarrow -\frac{\hat{a}_i + \hat{c}_i^T x_i}{2\hat{b}_i}$
  project greedy price: $p_{g,i} \leftarrow \mathrm{Proj}(p^u_{g,i}, [\underline{p}_i, \bar{p}_i])$
  set price $p_i \leftarrow p_{g,i}$
  observe demand $d_i := D_i(p_i)$
  set $(\hat{a}_{i+1}, \hat{b}_{i+1}, \hat{c}_{i+1}) \leftarrow \arg\min_{\alpha \in A, \beta \in B, \gamma \in C} \sum_{j=1}^i (d_j - \alpha - \beta p_j - \gamma^T x_j)^2$
end for

Algorithm 3 One step regression

input: parameter bounds $A = [-\bar{a}, \bar{a}]$, $B = [-\bar{b}, -\underline{b}]$, $C = \{c' \in \mathbb{R}^m : \sum_{k=1}^m |c'_k| \le \bar{c}\}$
initialize: choose $\hat{a}_1, \hat{b}_1, \hat{c}_1$ to be arbitrary parameters in A, B and C
for $i = 1, \dots, N$ **do**
  set $\delta_i \leftarrow i^{-\frac{1}{4}}$
  if $\bar{p}_i - \underline{p}_i \ge 2\delta_i$ **then**
    given x_i, set unconstrained greedy price: $p^u_{g,i} \leftarrow -\frac{\hat{a}_i + \hat{c}_i^T x_i}{2\hat{b}_i}$
    project greedy price: $p_{g,i} \leftarrow \mathrm{Proj}(p^u_{g,i}, [\underline{p}_i + \delta_i, \bar{p}_i - \delta_i])$
    generate an independent random variable $\Delta p_i \leftarrow i^{-\frac{1}{4}}$ w.p. $\frac{1}{2}$ and $\Delta p_i \leftarrow -i^{-\frac{1}{4}}$ w.p. $\frac{1}{2}$
    set price $p_i \leftarrow p_{g,i} + \Delta p_i$
  else $(\bar{p}_i - \underline{p}_i < 2\delta_i)$
    choose an arbitrary price $p_i \in [\underline{p}_i, \bar{p}_i]$
  end if
  observe demand $d_i := D_i(p_i)$
  set $(\hat{a}_{i+1}, \hat{b}_{i+1}, \hat{c}_{i+1}) \leftarrow \arg\min_{\alpha \in A, \beta \in B, \gamma \in C} \sum_{j=1}^i (d_j - \alpha - \beta p_j - \gamma^T x_j)^2$
end for

5.1 Numerical Example for Linear Demand

Set up The first simulation considers the case where the source of endogeneity is the correlation between the noise ϵ_i and the price limits $\underline{p}_i, \bar{p}_i$. The demand function is given by

$$D_i(p) = 2.7 - 0.7p + 0.9x_i + \epsilon_i,$$

where x_i is one dimensional (i.e. $m = 1$) and uniformly distributed in $[-1,\ 1]$, and the noise ϵ_i is given by $\epsilon_i = v_i + w_i$. v_i is normally distributed with mean 0 and variance 0.3, while w_i is uniformly distributed in $[-0.6,\ -0.5] \cup [0.5,\ 0.6]$. The price at period i is upper bounded by

$$\bar{p}_i = \max\{-\frac{a}{2b} + 0.01, p^*_i + 0.01 + 2\max\{0, w_i\}\}$$

and lower bounded by

$$\underline{p}_i = \min\{-\frac{a}{2b} + 0.01, p_i^* - 0.01 + 2\min\{0, w_i\}\}$$

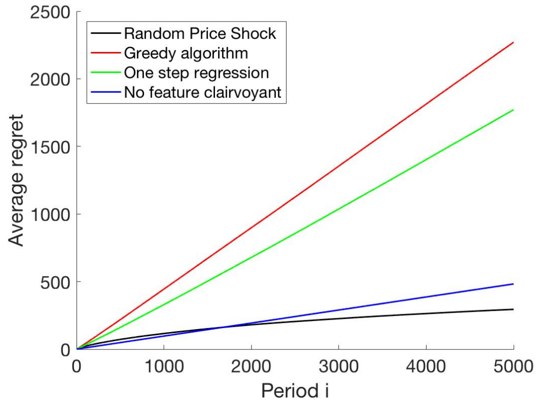
Since the bounds on the prices are functions of w_i, the noise ϵ_i is correlated with the price p_i. In this simulation, the one-step regression and RPS algorithms are run with price shocks set to $\frac{1}{(i+20)^{0.25}}$ instead of $1/i$ for convenience; this ensures that the sum of the greedy price and the price shock is always within the price bounds. It is equivalent to starting the clock at $i = 20$ and does not affect the growth rate of the regret. Finally, the retailer knows that a is in the interval $[2.2\ 3.2]$, b is in the interval $[-1\ -0.3]$ and c is in the interval $[0.4\ 1.4]$.

An example application for the above setting could be as follows: At each time period, a retailer sells a heterogeneous product with an observable characteristic x_i. Each product also has a quality factor, captured by the variable w_i, which affects demand for the product. The retailer does not include the value of w_i in the feature vector, but sets a higher price range for higher quality products and a lower price range for the prices of lower quality products. Thus the lower and upper bounds on price are correlated with the product quality, and hence the noise.

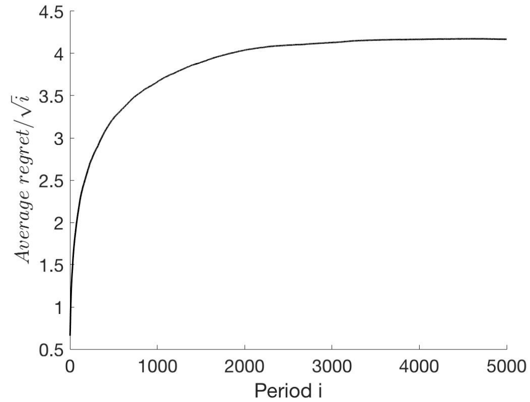
Results Figures 1a and 1b plot the increase in average regret (where the average is taken over 200 iterations) as the number of periods increases from 1 to 5000. Figure 1a shows that the regrets of the greedy algorithm, the one step regression algorithm and the clairvoyant who ignores features, grow linearly with i, while the regret of the RPS algorithm grows sub-linearly. Figure 1b, which plots the growth of the average regret at period i divided by $\sqrt{i}$ against the period i, confirms that the regret of RPS is $O(\sqrt{N})$. In this numerical example, the regrets incurred by the RPS algorithm and the no-feature clairvoyant are close for the first 2000 periods, before the RPS algorithm overtakes the clairvoyant. The no-feature clairvoyant still significantly outperforms the greedy algorithm and one step regression for the rest of the selling horizon, showing the extent of the effect of price endogeneity on the performance of the latter two algorithms.

Finally, Table 1 presents summary statistics of the estimates of the parameters a, b, c obtained by all pricing policies except the no-feature clairvoyant after 5000 periods (aggregated over 200 iterations). As expected, the greedy algorithm fails to estimate the parameters correctly. This is due to both incomplete learning and ignoring price endogeneity. The one step regression algorithm, while avoiding incomplete learning, also suffers from the effect of price endogeneity. Furthermore, both the greedy algorithm and the one step regression algorithm significantly over-estimate the parameter $b = -0.7$. The over-estimation of b is due to the fact that the price bounds, and hence the prices, are positively correlated with the noise. The RPS algorithm, on the other hand, produces close estimates of all the parameters.

Endogeneity effect We also measured price endogeneity by the correlation between the prices generated by all algorithms except the no feature clairvoyant and the noise. For each of the three algorithms, we calculated the correlation coefficient using data from all iterations and time periods. Table 2 shows the values obtained. As expected, the prices generated by each algorithm are positively correlated with the noise. Furthermore, the endogeneity effect is significant in this example, with the correlation coefficients being significantly deviated from 0. Notice that the endogeneity effect exists for the RPS algorithm as well, since the algorithm does not attempt to eliminate endogeneity; instead, the RPS algorithm is designed to correctly estimate parameters in the presence of endogeneity.



(a) Average regret over 200 iterations



(b) Average regret of RPS scaled by a factor of $\frac{1}{\sqrt{i}}$ at period i

Figure 1: Average regret in Linear Demand Example

Table 1: Estimates of parameters in Linear Demand Example

	True value	RPS algo.	Greedy algo.	One step reg.
Mean ($\hat{a}_i$)	2.70	2.70	2.20	2.56
Median ($\hat{a}_i$)	2.70	2.71	2.20	2.56
Mean ($\hat{b}_i$)	-0.70	-0.70	-0.30	-0.30
Median ($\hat{b}_i$)	-0.70	-0.71	-0.30	-0.30
Mean ($\hat{c}_i$)	0.90	0.90	0.40	0.43
Median ($\hat{c}_i$)	0.90	0.90	0.40	0.43

Table 2: Correlation coefficient (R) between price and noise in Linear Demand Example

	RPS algo.	Greedy algo.	One step reg.
R	0.59	0.22	0.57

5.2 Numerical Example for Quasi-Linear Demand

Set up The second simulation example considers the case where the source of endogeneity is a misspecified demand function. In this set up, demand is given by the quasilinear function

$$D_i(p) = \frac{1}{2(x_i + 1.03)} + 1 - 0.9p + \epsilon_i,$$

where x_i is uniformly distributed between $[-1, 1]$ and the noise ϵ_i is normally distributed with mean 0 and standard deviation 0.1. Using the closed-form expression in Eq (8), it can be seen that the linear demand model approximated by least squares is given by

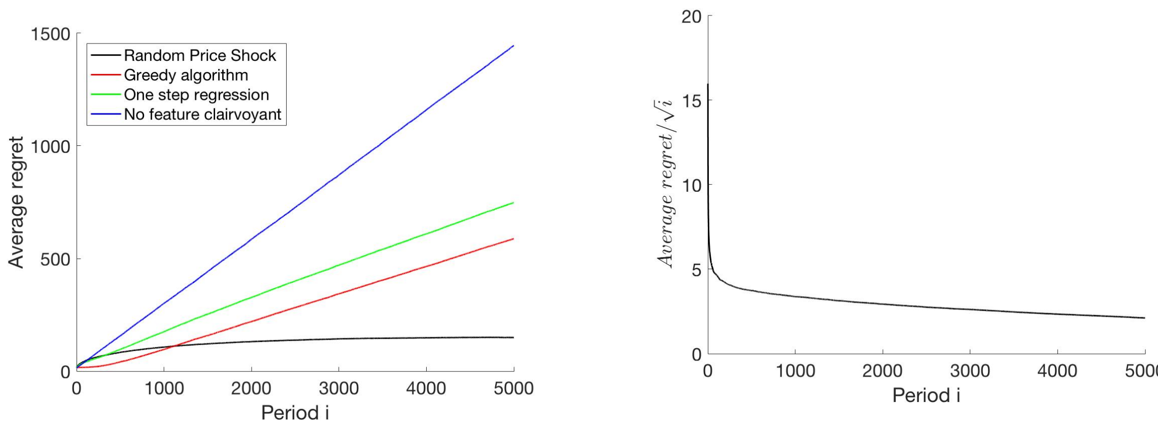
$$\hat{D}_i(p) \approx 2.05 - 0.90p - 1.76x_i,$$

where all coefficients are expressed to 2 decimal places.

The price range at period i is lower bounded by $\underline{p}_i = 0.69$ and upper bounded by $\bar{p}_i = 9.81$.

Note that unlike the previous numerical example, the price bounds in this example are constants and uncorrelated with the noise. Thus the sole source of endogeneity in this example is the misspecification of the demand function. 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 Figure 2a shows that in this numerical example, the regret of the greedy algorithm, the one step regression algorithm, and the clairvoyant who ignores features, grow linearly with i, and in all cases the regrets are higher than that of RPS after around 1000 iterations. Figure 2b confirms that the regret of the RPS algorithm is $O(\sqrt{N})$. Finally, Table 3, which provides summary statistics of the parameter estimates of all the pricing algorithms except the clairvoyant at the end of the selling horizon, shows that the RPS algorithm produces close estimates of all the parameters whereas the greedy and one step regression algorithms, even after 5000 iterations, produce parameter estimates that are significantly deviated from their true values.



(a) Average regret over 200 iterations

(b) Average regret of RPS scaled by a factor of $\frac{1}{\sqrt{i}}$ at period i

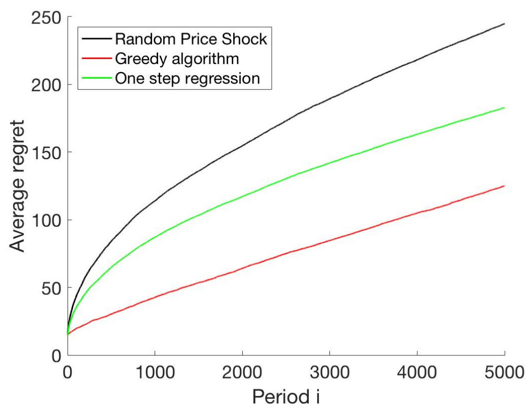
Figure 2: Average regret in Quasi-Linear Demand Example

Table 3: Estimates of parameters in Quasi-Linear Demand Example

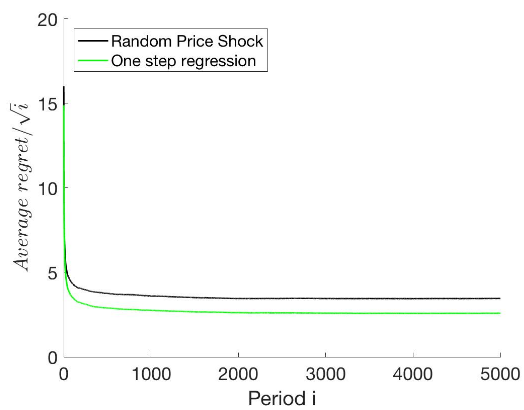
	True value	RPS algo.	Greedy algo.	One step reg.
Mean ($\hat{a}_i$)	2.05	2.04	1.50	1.50
Median ($\hat{a}_i$)	2.05	2.04	1.50	1.50
Mean ($\hat{b}_i$)	-0.90	-0.91	-0.50	-0.50
Median ($\hat{b}_i$)	-0.90	-0.89	-0.50	-0.50
Mean ($\hat{c}_i$)	-1.76	-1.74	-1.20	-1.20
Median ($\hat{c}_i$)	-1.76	-1.75	-1.20	-1.20

Endogeneity effect As noted earlier, the source of the endogeneity effect in the case of quasilinear demand is the fact that the price p_i is not an affine function of x_i in the presence of price range bounds and may thus be correlated with the noise $\nu_i = \epsilon_i + f(x_i) - a - c^T x_i$. To illustrate this, we repeat the above numerical example with the same parameter settings but with the price bounds set to $\underline{p}_i = -\infty$, $\bar{p}_i = +\infty$. The prices chosen by the RPS, greedy and one step regression algorithms are thus affine functions of x_i and are uncorrelated with the noise. Figure

3a shows the average regret (over 200 iterations) incurred by the three algorithms. Although the average regret of the greedy algorithm is lower than the regret of the other algorithms in this example, the growth of its regret is clearly linear. Figure 3b shows that the regret of the RPS and one step regression algorithms is $O(\sqrt{N})$. Furthermore, Table 4 shows that these two algorithms tend to produce close estimates of the parameters at the end of the selling horizon while the greedy algorithm yields poorer estimates of the parameters. This example illustrates the fact that the one step regression algorithm will incur $O(\sqrt{N})$ regret in the absence of price endogeneity, i.e. it is optimal without endogeneity, and confirms that the presence of price bounds is the source of the endogeneity effect when the demand function is misspecifed. It also shows that even without price endogeneity, the greedy algorithm will suffer from incomplete learning and the order of its regret will not be optimal in N. Finally, we note that in this setting, the one step regression outperforms the RPS algorithm. This is unsurprising, since in the absence of endogeneity parameters can be learned more efficiently through a one step rather than a two step regression.



(a) Average regret over 200 iterations



(b) Average regret scaled by a factor of $\frac{1}{\sqrt{i}}$ at period i

Figure 3: Average regret in Quasi-Linear Demand Example without price bounds

Table 4: Estimates of parameters in Quasi-Linear Demand Example without price bounds

	True value	RPS algo.	Greedy algo.	One step reg.
Mean ($\hat{a}_i$)	2.05	2.05	2.19	2.07
Median ($\hat{a}_i$)	2.05	2.06	2.21	2.07
Mean ($\hat{b}_i$)	-0.90	-0.91	-1.05	-0.92
Median ($\hat{b}_i$)	-0.90	-0.91	-1.04	-0.92
Mean ($\hat{c}_i$)	-1.76	-1.75	-1.87	-1.76
Median ($\hat{c}_i$)	-1.76	-1.75	-1.87	-1.77

6 Conclusion

To address the price endogeneity effect in dynamic pricing with contextual information, we have proposed in this paper a "random price shock" (RPS) algorithm. The RPS algorithm dynamically generates random price shocks to estimate price elasticity and maximize price while controlling for the endogeneity effect. It has many advantages over existing solutions to the endogeneity problem, being simpler to implement than the instrumental variables method while having strong numerical and theoretical performance. The expected regret of the RPS algorithm is $O(\sqrt{N})$, which matches the best possible lower bound of $\Omega(\sqrt{N})$ for linear pricing models in the special case where there are no features. We have also demonstrated that the RPS algorithm is robust to model misspecification and always converges to the linear demand model with least squares parameters. Model misspecification does not affect the regret of the RPS algorithm, which is also $O(\sqrt{N})$ in this case.

There are several directions for future work. One possible extension is to allow the demand function to take on more general forms. For example, demand could be a nonlinear function of price, or it could be represented by a generalized linear model. Another direction is to extend the single product setting studied in this paper to a a multi product setting, which could be represented using a discrete choice model. Finally, our model could be extended to incorporate business constraints such as limited inventory constraints.

A Appendix: Proofs for Theoretical Analysis

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

A.1 Proof of Theorem 1

Proof. Using the closed form expression for the optimal price at period i, $p_i^* = \frac{-(a+c^T x_i)}{2b}$, the expected regret over the selling horizon can be written as

$$\text{Expected Regret}(N) = \sum_{i=1}^{N} \mathrm{E}[p_i^* D(p_i^*)] - \sum_{i=1}^{N} \mathrm{E}[p_i D(p_i)] = -b \sum_{i=1}^{N} \mathrm{E}[(p_i - p_i^*)^2]. \tag{9}$$

Now suppose that at time period i we have $\bar{p}_i - \underline{p}_i < 2\delta_i$. In this case, RPS does not introduce any price shocks and simply selects an arbitrary price in the interval $[\underline{p}_i\ \bar{p}_i]$. Since p_i^* is contained in this interval, the expected revenue loss incurred in this period can be at most $4\bar{b}\delta_i^2 = 4\bar{b}/\sqrt{i}$, and the total expected revenue loss incurred during all time periods where $\bar{p}_i - \underline{p}_i < 2\delta_i$ is at most $4\bar{b}\sum_{i=1}^{N} 1/\sqrt{i} \le 8\bar{b}\sqrt{N}$. Therefore, without loss of generality, we can remove all time periods with $\bar{p}_i - \underline{p}_i < 2\delta_i$ from the selling horizon. The remaining proof will assume that in every time period i, the price bounds satisfy $\bar{p}_i - \underline{p}_i \ge 2\delta_i$.

We decompose the regret as

$$\begin{aligned}
\text{Expected Regret}(N) &= \sum_{i=1}^{N} -b\mathrm{E}[(p_i - p_i^*)^2] \\
&= \sum_{i=1}^{N} -b\left(\mathrm{E}[(p_{g,i} - p_i^*)^2] + 2\mathrm{E}[\Delta p_i (p_{g,i} - p_i^*)] + \mathrm{E}[(\Delta p_i)^2]\right) \\
&= \sum_{i=1}^{N} -b\left(\mathrm{E}[(p_{g,i} - p_i^*)^2] + \mathrm{E}[(\Delta p_i)^2]\right)
\end{aligned}$$

$$\leq \sum_{i=1}^{N} -b\left(\mathrm{E}[(p_{g,i}^u - p_i^*)^2] + \delta_i^2 + \mathrm{E}[(\Delta p_i)^2]\right). \tag{10}$$

The first equality uses the definition of p_i^*. The second and third equalities hold because $p_i = p_{g,i} + \Delta p_i$, and Δp_i is sampled independently from $p_{g,i}$ and p_i^*. To show the last step, we consider the two cases: (1) $p_i^* \in [\underline{p}_i + \delta_i, \bar{p}_i - \delta_i]$; (2) $p_i^* \in [\underline{p}_i, \underline{p}_i + \delta_i) \cup (\bar{p}_i - \delta_i, \bar{p}_i]$. In the first case, we have $(p_{g,i} - p_i^*)^2 \leq (p_{g,i}^u - p_i^*)^2$; in the second case, we have $(p_{g,i} - p_i^*)^2 \leq (p_{g,i}^u - p_i^*)^2 + \delta_i^2$.

By our choice of δ_i and Δp_i in Algorithm 1, we have

$$\sum_{i=1}^{N} \delta_i^2 = \sum_{i=1}^{N} \frac{1}{\sqrt{i}} \leq 2\sqrt{N}; \quad \sum_{i=1}^{N} \mathrm{E}[(\Delta p_i)^2] = \sum_{i=1}^{N} \frac{1}{\sqrt{i}} \leq 2\sqrt{N}.$$

To finish the proof, we want to show that $\mathrm{E}[(p_{g,i}^u - p_i^*)^2] = O(1/\sqrt{i})$.

By the definition of $p_{i,g}^u, p_i^*$, we have

$$\begin{aligned} \mathrm{E}[(p_{i,g}^u - p_i^*)^2] &\leq K\mathrm{E}\left[(b - \hat{b}_i)^2 + (a + c^T x_i - \hat{a}_i - \hat{c}_i^T x_i)^2\right] \\ &= K\mathrm{E}\left[(b - \hat{b}_i)^2 + (e^T \tilde{x}_i - e_i^T \tilde{x}_i)^2\right] \end{aligned} \tag{11}$$

for $K := \frac{(\bar{a}+\bar{c})^2}{4\underline{b}^4} + \frac{1}{4\underline{b}^2}$. To see this, for any demand parameters a', b', c', let h be the following function: $h(b', a' + c'^T x_i) = -\frac{a'+c'^T x_i}{2b'}$. We then have

$$p_i^* = h(b, a + c^T x_i), \quad p_{i,g}^u = h(\hat{b}, \hat{a}_i + \hat{c}_i^T x_i).$$

The gradient of h, denoted by ∇h, is given by $\nabla h(b', a' + c'^T x_i) = (\frac{a'+c'^T x_i}{2b'^2}, -\frac{1}{2b'})^T$. Since $|c'^T x_i| \leq \|c'\|_1 \|x_i\|_\infty \leq \|c'\|_1 \leq \bar{c}$, we have

$$\|\nabla h(b', a' + c'^T x_i)\|^2 = \frac{(a' + c'^T x_i)^2}{4(b')^4} + \frac{1}{4(b')^2} \leq \frac{(|a'| + \|c'\|_1)^2}{4(b')^4} + \frac{1}{4(b')^2} \leq K,$$

and Eq (11) follows from the Mean Value Theorem.

By Lemma 3, we immediately have $\mathrm{E}[(\hat{b}_i - b)^2] = \mathrm{O}(1/\sqrt{i})$. Now we will bound the error in the estimates of a and c, namely $\mathrm{E}[\left((e - e_i)^T \tilde{x}_i\right)^2]$. Note that e_i is measurable with history $\mathcal{H}_{i-1}$ and $\tilde{x}_i$ is independent of $\mathcal{H}_{i-1}$, so

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

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

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

$$e_i - e = \mathrm{Proj}\left(M_i^{-1} \frac{\sum_{j=1}^{i-1} \tilde{x}_j (p_j(b - \hat{b}_i) + \epsilon_j)}{i-1}, A \times C\right). \tag{12}$$

Since M is positive definitive, there exists a positive definitive matrix Q such that $M = Q^2$. Let

A_i be the following event:

$$A_i = \big\{M_i \text{ is invertible and } \|QM_i^{-1}Q\|_2 \leq 2\big\}.$$

Since $\|e_i - e\|^2 = \|\hat{a}_i - a\|^2 + \|\hat{c}_i - c\|^2 \leq 4\bar{a}^2 + 4\bar{c}^2$, we have

$$\begin{aligned}
\mathrm{E}[\|e_i - e\|_M^2] &= \mathrm{E}[\|e_i - e\|_M^2 \mid A_i] \cdot \mathbb{P}[A_i] + \mathrm{E}[\|e_i - e\|_M^2 \mid A_i^C] \cdot \mathbb{P}[A_i^C] \\
&\leq \mathrm{E}[\|e_i - e\|_M^2 \mid A_i] \cdot \mathbb{P}[A_i] + \lambda_{\max}(M)\mathrm{E}[\|e_i - e\|^2 \mid A_i^C] \cdot \mathbb{P}[A_i^C] \\
&\leq \mathrm{E}[\|e_i - e\|_M^2 \mid A_i] \cdot \mathbb{P}[A_i] + 4\lambda_{\max}(M)(\bar{a}^2 + \bar{c}^2)\mathbb{P}[A_i^C] \\
&\leq \mathrm{E}[\|e_i - e\|_M^2 \mid A_i] \cdot \mathbb{P}[A_i] && (13) \\
&\quad + 4\lambda_{\max}(M)(\bar{a}^2 + \bar{c}^2) \cdot 2(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(i-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right). && (14)
\end{aligned}$$

In the last step, we bound $\mathbb{P}(A_i^C)$ by Lemma 4, where $V := \mathrm{E}[(Q^{-1}\tilde{x}\tilde{x}^TQ^{-1} - I)^2]$. Since Eq (14) is $O(e^{-i})$, it is left to show that Eq (13) is $O(1/\sqrt{i})$.

We write Eq (13) as

$$\begin{aligned}
\mathrm{E}[\|e_i - e\|_M^2 \mid A_i] \cdot \mathbb{P}[A_i] &\leq \mathrm{E}\left[\|QM_i^{-1}QQ^{-1}\frac{\sum_{j=1}^{i-1}\tilde{x}_j(p_j(b - \hat{b}_i) + \epsilon_j)}{i-1}\|^2 \mid A_i\right]\mathbb{P}[A_i] \\
&\leq \mathrm{E}\left[\|QM_i^{-1}Q\|_2^2 \cdot \|Q^{-1}\|_2^2 \cdot \|\frac{\sum_{j=1}^{i-1}\tilde{x}_j(p_j(b - \hat{b}_i) + \epsilon_j)}{i-1}\|^2 \mid A_i\right]\mathbb{P}[A_i] \\
&\leq \mathrm{E}\left[4 \cdot \frac{1}{\lambda_{\min}(M)} \cdot 2\left(\|\frac{\sum_{j=1}^{i-1}\tilde{x}_jp_j(b - \hat{b}_i)}{i-1}\|^2 + \|\frac{\sum_{j=1}^{i-1}\tilde{x}_j\epsilon_j}{i-1}\|^2\right) \mid A_i\right]\mathbb{P}[A_i] \\
&\leq \mathrm{E}\left[\frac{8}{\lambda_{\min}(M)}\left(\|\frac{\sum_{j=1}^{i-1}\tilde{x}_jp_j(b - \hat{b}_i)}{i-1}\|^2 + \|\frac{\sum_{j=1}^{i-1}\tilde{x}_j\epsilon_j}{i-1}\|^2\right)\right] \\
&\leq \mathrm{E}\left[\frac{8}{\lambda_{\min}(M)}\left((m+1)p_{\max}^2(b - \hat{b}_i)^2 + \|\frac{\sum_{j=1}^{i-1}\tilde{x}_j\epsilon_j}{i-1}\|^2\right)\right] \\
&= \frac{8}{\lambda_{\min}(M)}\left((m+1)p_{\max}^2\mathrm{E}[(b - \hat{b}_i)^2] + \frac{1}{(i-1)^2}\mathrm{E}[\|\sum_{j=1}^{i-1}\tilde{x}_j\epsilon_j\|^2]\right).
\end{aligned} \tag{15}$$

The first inequality holds by Eq (12) and the assumption that the true parameter $e \in A \times C$. 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_i and the fact $\|x + y\|^2 \leq 2\|x\|^2 + 2\|y\|^2$. The fourth inequality simply uses the definition of conditional expectation. The fifth inequality uses the assumptions that $\|x_i\|_\infty \leq 1$ and $p_j \leq p_{\max}$.

It has already been established using Lemma 3 that $\mathrm{E}[(b - \hat{b}_i)^2]$ is $O(1/\sqrt{i})$, so the first term of Eq (15) is $O(1/\sqrt{i})$. For the second term, note that $(\tilde{x}_j, \epsilon_j)$ is independent of $(\tilde{x}_{j'}, \epsilon_{j'})$ for $j \neq j'$. Furthermore, for each j, $\mathrm{E}[\tilde{x}_j\epsilon_j \mid \mathcal{H}_{j-1}] = \mathrm{E}[\tilde{x}_j\epsilon_j] = 0$, and $\mathrm{E}[\|\tilde{x}_j\epsilon_j\|^2] \leq \mathrm{E}[(m+1)\epsilon_j^2] \leq$

$(m+1)\sigma^2$. Thus,

$$\frac{1}{(i-1)^2}\mathrm{E}[\|\sum_{j=1}^{i-1}\tilde{x}_j\epsilon_j\|^2] = \frac{1}{(i-1)^2}\sum_{j=1}^{i-1}\mathrm{E}[\|\tilde{x}_j\epsilon_j\|^2] \leq \frac{(m+1)\sigma^2}{i-1}. \tag{16}$$

Therefore, by Eq (15), $\mathrm{E}[\|e-e_i\|_M^2] \leq O(1/\sqrt{i}) + O(1/i) = O(1/\sqrt{i})$ as desired.

Dependence on constant factors By combining all the constant factors, the expected regret of RPS algorithm over N periods can be bounded by

$$O\left(\frac{\bar{b}(\bar{a}^2+\bar{c}^2+\underline{b}^2)}{\underline{b}^4}(\bar{a}^2+\bar{c}^2+\sigma^2+\bar{b}^2p_{\max}^2)(1+p_{\max}^2\frac{m+1}{\lambda_{\min}(M)})\sqrt{N}\right) + O(\log N),$$

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

□

A.2 Proof of Proposition 1

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

$$\begin{aligned}
\mathrm{E}[p_i^* D(p_i^*) - p_i' D(p_i')] =& \mathrm{E}\left[p_i^*(bp_i^* + f(x_i)) - p_i'(bp_i' + f(x_i))\right] \\
=& \mathrm{E}\left[p_i^*(bp_i^* + a + c^T x_i) - p_i'(bp_i' + a + c^T x_i) - (p_i^* - p_i')(a + c^T x_i - f(x_i))\right] \\
=& \mathrm{E}\left[p_i^*(bp_i^* - 2bp_i^*) - p_i'(bp_i' - 2bp_i^*)\right] - \mathrm{E}\left[(p_i^* - p_i')(a + c^T x_i - f(x_i))\right] \\
=& -b\mathrm{E}\left[(p_i^* - p_i')^2\right] - \mathrm{E}\left[(p_i^* - p_i')(a + c^T x_i - f(x_i))\right].
\end{aligned}$$

To finish the proof, we shall prove that $\mathrm{E}\left[(p_i^* - p_i')(a + c^T x_i - f(x_i))\right] = 0$. By definition, a, c is the optimal solution of the following least squares problem

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

By the first order condition, we have

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

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

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

Therefore,

$$\begin{aligned}
\mathrm{E}\left[(p_i^* - p_i')(a + c^T x_i - f(x_i))\right] =& \mathrm{E}\left[(-\frac{a+c^T x_i}{2b} + \frac{\alpha+\gamma^T x_i}{2\beta})(a + c^T x_i - f(x_i))\right] \\
=& \mathrm{E}\left[\mathrm{E}\left[(-\frac{a+c^T x_i}{2b} + \frac{\alpha+\gamma^T x_i}{2\beta})(a + c^T x_i - f(x_i)) \mid \mathcal{H}_{i-1}\right]\right] \\
=& \mathrm{E}\left[(-\frac{a}{2b} + \frac{\alpha}{2\beta})\mathrm{E}\left[(a + c^T x_i - f(x_i)) \mid \mathcal{H}_{i-1}\right]\right]
\end{aligned}$$

$$+\mathrm{E}\left[(-\frac{c^T}{2b}+\frac{\gamma^T}{2\beta})\mathrm{E}\left[x_i(a+c^Tx_i-f(x_i))\mid \mathcal{H}_{i-1}\right]\right]=0.$$

□

A.3 Proof of Theorem 2

Proof. We first consider all period i such that $\bar{p}_i - \underline{p}_i < 2\delta_i$. In this case, RPS does not introduce any price shocks and simply selects an arbitrary price in the interval $[\underline{p}_i\ \bar{p}_i]$. Recall that $\tilde{p}_i = -\frac{f(x_i)}{2b}$ is the optimal price for period i, and we have

$$\mathrm{E}[p_i^*D_i(p_i^*)] - \mathrm{E}[p_iD_i(p_i)] \le \mathrm{E}[\tilde{p}_iD_i(\tilde{p}_i)] - \mathrm{E}[p_iD_i(p_i)] = -b\mathrm{E}[(\tilde{p}_i - p_i)^2].$$

Since we assumed that p_i and $\tilde{p}_i \in [\underline{p}_i, \bar{p}_i]$, if $\bar{p}_i - \underline{p}_i < 2\delta_i$, we have $\mathrm{E}[(\tilde{p}_i - p_i)^2] \le 4\delta_i^2 = 4/\sqrt{i}$. The total expected revenue loss incurred during all time periods where $\bar{p}_i - \underline{p}_i < 2\delta_i$ is at most $4\bar{b}\sum_{i=1}^N 1/\sqrt{i} \le 8\bar{b}\sqrt{N}$. Therefore, without loss of generality, we can remove all time periods with $\bar{p}_i - \underline{p}_i < 2\delta_i$ from the selling horizon. The remaining proof will assume that in every time period i, the price bounds satisfy $\bar{p}_i - \underline{p}_i \ge 2\delta_i$.

We decompose the regret as

$$\begin{aligned}\sum_{i=1}^N \mathrm{E}[p_i^*D_i(p_i^*) - p_iD_i(p_i)] =& \sum_{i=1}^N \mathrm{E}[p_i^*D_i(p_i^*) - p_{g,i}^uD_i(p_{g,i}^u)] \\ &+ \sum_{i=1}^N \mathrm{E}[p_{g,i}^uD_i(p_{g,i}^u) - p_{g,i}D_i(p_{g,i})] + \sum_{i=1}^N \mathrm{E}[p_{g,i}D_i(p_{g,i}) - p_iD_i(p_i)].\end{aligned}$$

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

$$\sum_{i=1}^N \mathrm{E}[p_{g,i}^uD_i(p_{g,i}^u) - p_{g,i}D_i(p_{g,i})] \le \sum_{i=1}^N \bar{b}\delta_i^2 = \sum_{i=1}^N \frac{\bar{b}}{\sqrt{i}} \le 2\bar{b}\sqrt{N}.$$

In addition, $p_i = p_{g,i} + \Delta p_i$, where Δp_i is generated independently from $p_{g,i}$ and x_i. So

$$\begin{aligned}\sum_{i=1}^N \mathrm{E}[p_{g,i}D_i(p_{g,i}) - p_iD_i(p_i)] &= \sum_{i=1}^N \mathrm{E}[p_{g,i}D_i(p_{g,i}) - (p_{g,i} + \Delta p_i)D_i(p_{g,i} + \Delta p_i)] \\ &= \sum_{i=1}^N \mathrm{E}[\Delta p_i(-2bp_{g,i} - f(x_i)) - b(\Delta p_i)^2] \\ &= \sum_{i=1}^N \mathrm{E}[-b(\Delta p_i)^2] \\ &= \sum_{i=1}^N -b\frac{1}{\sqrt{i}} \le 2\bar{b}\sqrt{N}.\end{aligned}$$

In the remaining proof, we want to show that $\mathrm{E}[p_i^*D_i(p_i^*) - p_{g,i}^uD_i(p_{g,i}^u)] = O(1/\sqrt{i})$. By Proposition 1, we have

$$\mathrm{E}[p_i^*D_i(p_i^*) - p_{g,i}^uD_i(p_{g,i}^u)] = -b\mathrm{E}[(p_i^* - p_{g,i}^u)^2],$$

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

$$D_i(p) = a + bp + c^T x_i + \nu_i, \quad \forall p \in [\underline{p}_i, \bar{p}_i],$$

with $\nu_i = f(x_i) - a - c^T x_i + \epsilon_i$. By the first order condition of the least squares estimator, we have $\mathrm{E}[x_i \nu_i] = 0$.

The rest of the proof for showing $\mathrm{E}[(p^*_i - p^u_{g,i})^2] = O(1/\sqrt{i})$ is identical to the proof for Theorem 1 by replacing ϵ_i with ν_i. Note that the proof of Theorem 1 never used the condition that ϵ_i is mean independent of x_i, namely $\mathrm{E}[\epsilon_i|x_i] = 0$, after Eq (9); it only requires the weaker condition that ϵ_i and x_i are uncorrelated, namely $\mathrm{E}[\epsilon_i x_i]$, in Eq (16). Since $\mathrm{E}[x_i \nu_i] = 0$, the proof for Theorem 1 goes through for the quasilinear case as well. The last step in the proof for Theorem 1 (Eq (16)) is replaced by

$$\begin{aligned}
\frac{1}{(i-1)^2}\mathrm{E}[\|\sum_{j=1}^{i-1} \tilde{x}_j \nu_j\|^2] =& \frac{1}{(i-1)^2}\sum_{j=1}^{i-1} \mathrm{E}[\|\tilde{x}_j \nu_j\|^2] \\
=& \frac{1}{(i-1)^2}\sum_{j=1}^{i-1} \mathrm{E}[\|\tilde{x}_j (f(x_i) - a - c^T x_i + \epsilon_i)\|^2] \\
\leq& \frac{(m+1)}{i-1} 3(\bar{f}^2 + (\bar{a}+\bar{c})^2 + \sigma^2),
\end{aligned}$$

where the last step uses the fact that $(x+y+z)^2 \leq 3(x^2+y^2+z^2)$ and $\|\tilde{x}_j\|^2 \leq m+1$.

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

$$O\left(\frac{\bar{b}(\bar{a}^2 + \bar{c}^2 + \underline{b}^2)}{\underline{b}^4} (\bar{f}^2 + \sigma^2 + \bar{b}^2 p^2_{\max})(1 + p^2_{\max} \frac{m+1}{\lambda_{\min}(M)}) \sqrt{N} \right) + O(\log N),$$

where the pre-factor in the first big O notation only contains an absolute constant. Note that this bound matches the regret bound for the linear case in Theorem 1.

□

A.4 Lemmas

Lemma 3 (Bound on $\hat{b}_i$)**.** *With quasilinear demand in Eq (1), for $i \geq 2$, we have* $\mathrm{E}[(\hat{b}_i - b)^2] \leq 3(\bar{f}^2 + \sigma^2 + \bar{b}^2 p^2_{\max})/\sqrt{i}$.

Proof. Using the definition in Algorithm 1, $\hat{b}_i = \mathrm{Proj}(\hat{b}^u_i, B)$, where $\hat{b}^u_i = \frac{\sum_{j=1}^{i-1} \Delta p_j D_j}{\sum_{j=1}^{i-1} \Delta p_j^2}$. Since the true parameter $b \in B$, we have

$$\begin{aligned}
\mathrm{E}[(\hat{b}_i - b)^2] &\leq \mathrm{E}[(\hat{b}^u_i - b)^2] \\
&= \mathrm{E}\left[\left(\frac{\sum_{j=1}^{i-1} \Delta p_j D_j}{\sum_{j=1}^{i-1} \Delta p_j^2} - b \right)^2 \right] \\
&= \mathrm{E}\left[\left(\frac{\sum_{j=1}^{i-1} \Delta p_j (f(x_j) + \epsilon_j + b p_{g,j} + b\Delta p_j)}{\sum_{j=1}^{i-1} \Delta p_j^2} - b \right)^2 \right]
\end{aligned}$$

$$= \mathrm{E}\left[\left(\frac{\sum_{j=1}^{i-1}\Delta p_j(f(x_i)+\epsilon_j+bp_{g,j})}{\sum_{j=1}^{i-1}\Delta p_j^2}\right)^2\right]$$

$$= \mathrm{E}\left[\left(\frac{\sum_{j=1}^{i-1}\Delta p_j(f(x_i)+\epsilon_j+bp_{g,j})}{\sum_{j=1}^{i-1} j^{-\frac{1}{2}}}\right)^2\right].$$

In the last equality, we used the fact that $\Delta p_j^2 = j^{-1/2}$.

Note that Δp_j's for all j are mutually independent and have mean 0, so

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{\sum_{j=1}^{i-1}\Delta p_j(f(x_i)+\epsilon_j+bp_{g,j})}{\sum_{j=1}^{i-1} j^{-\frac{1}{2}}}\right)^2\right] &= \mathrm{E}\left[\frac{\sum_{j=1}^{i-1}\Delta p_j^2(f(x_j)+\epsilon_j+bp_{g,j})^2}{(\sum_{j=1}^{i-1} j^{-\frac{1}{2}})^2}\right] \\
&\leq \mathrm{E}\left[\frac{\sum_{j=1}^{i-1}3\Delta p_j^2(f(x_j)^2+\epsilon_i^2+b^2p_{g,j}^2)}{(\sum_{j=1}^{i-1} j^{-\frac{1}{2}})^2}\right] \\
&\leq 3\cdot\frac{\bar{f}^2+\sigma^2+\bar{b}^2p_{\max}^2}{\sum_{j=1}^{i-1} j^{-\frac{1}{2}}}
\end{aligned}$$

We used the fact that $(x+y+z)^2 \leq 3(x^2+y^2+z^2)$. In the last step, we used the definition that $\Delta p_j^2 = j^{-1/2}$ and the assumption that $f(x_i), b, p_{g,j}$ are bounded.

Finally, since

$$\sum_{j=1}^{i-1} j^{-\frac{1}{2}} \geq \int_{y=1}^{i} y^{-\frac{1}{2}}dy = 2\sqrt{i}-2,$$

we have for $i \geq 2$ that

$$\mathrm{E}[(\hat{b}_i - b)^2] \leq \frac{3}{2}\cdot\frac{\bar{f}^2+\sigma^2+\bar{b}^2p_{\max}^2}{\sqrt{i}-1} \leq 3\cdot\frac{\bar{f}^2+\sigma^2+\bar{b}^2p_{\max}^2}{\sqrt{i}}.$$

□

Lemma 4 (Bound on $\|QM_i^{-1}Q\|_2$)**.** *Let $M = \mathrm{E}[\tilde{x}\tilde{x}^T]$, $V = \mathrm{E}[(Q^{-1}\tilde{x}\tilde{x}^TQ^{-1}-I)^2]$ and $M_i = \frac{1}{i-1}\sum_{j=1}^{i-1}\tilde{x}_j\tilde{x}_j^T$. For any $i \geq 2$, M_i is invertible and $\|QM_i^{-1}Q\|_2 \leq 2$ with probability at least*

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

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

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

Note that we used the submultiplicative property of the spectral norm. Since $\{\tilde{x}_j\}$ are independent and identically distributed, we apply the matrix Bernstein bound (Lemma 5) with

$t = (i-1)/2$ to yield

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

By an identical argument, we also have

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

Thus we have

$$\begin{aligned}\mathbb{P}[\|I - Q^{-1}M_iQ^{-1}\|_2 > \frac{1}{2}] &= \mathbb{P}[\max\{\lambda_{\max}(I - Q^{-1}M_iQ^{-1}), \lambda_{\max}(Q^{-1}M_iQ^{-1} - I)\} > \frac{1}{2}] \\ &\le 2(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(i-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right). \end{aligned} \tag{17}$$

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

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

By Eq (17), with probability at least

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

we have $\lambda_{\min}(Q^{-1}M_iQ^{-1}) \ge 1/2$. Since $Q^{-1}M_iQ^{-1} = Q^{-1}\frac{\sum_{j=1}^{i-1}\tilde{x}_j\tilde{x}_j^T}{i-1}Q^{-1}$ is positive semidefinite, $\lambda_{\min}(Q^{-1}M_iQ^{-1}) > 0$ implies that it is invertible. Then

$$\|QM_i^{-1}Q\|_2 = \frac{1}{\lambda_{\min}(Q^{-1}M_iQ^{-1})} \le 2.$$

This proves the lemma.

□

Lemma 5 (Matrix Bernstein bound, [Tro12])**.** *Consider a finite sequence X_k of independent, random, self-adjoint matrices with dimension d. Assume that each random matrix satisfies*

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

then for all $t \ge 0$,

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

References

[AP08] Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2008.

[BHP05] Tammo HA Bijmolt, Harald J van Heerde, and Rik GM Pieters. New empirical generalizations on the determinants of price elasticity. *Journal of marketing research*, 42(2):141–156, 2005.

[BK16] Dimitris Bertsimas and Nathan Kallus. Pricing from observational data. *arXiv preprint arXiv:1605.02347*, 2016.

[BLP95] Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

[BZ15] Omar Besbes and Assaf Zeevi. On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Science*, 61(4):723–739, 2015.

[CLPL16] Maxime C Cohen, Ilan Lobel, and Renato Paes Leme. Feature-based dynamic pricing. *Available at SSRN*, 2016.

[dB15] Arnoud V den Boer. Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science*, 20(1):1–18, 2015.

[FGL16] Marshall Fisher, Santiago Gallino, and Jun Li. Competition-based dynamic pricing in online retailing: A methodology validated with field experiments. *Ross School of Business Paper No. 1265*, 2016.

[FLSL15] Kris Johnson Ferreira, Bin Hong Alex Lee, and David Simchi-Levi. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1):69–88, 2015.

[JN16] Adel Javanmard and Hamid Nazerzadeh. Dynamic pricing in high-dimensions. *arXiv preprint arXiv:1609.07574*, 2016.

[KZ14] N Bora Keskin and Assaf Zeevi. Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research*, 62(5):1142–1167, 2014.

[McK15] W McKitterick. Online fashion sample sales in the us. Technical report, Tech. rep., IBISWorld Industry Report OD5438, 2015.

[PŞVR15] Robert Phillips, A Serdar Şimşek, and Garrett Van Ryzin. The effectiveness of field price discretion: Empirical evidence from auto lending. *Management Science*, 61(8):1741–1759, 2015.

[QB16] Sheng Qiang and Mohsen Bayati. Dynamic pricing with demand covariates. *Available at SSRN 2765257*, 2016.

[Tro12] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.

[TVR06] Kalyan T Talluri and Garrett J Van Ryzin. *The theory and practice of revenue management*, volume 68. Springer Science & Business Media, 2006.