Soft Computing (2021) 25:11711–11733
https://doi.org/10.1007/s00500-021-06047-y

SOFT COMPUTING IN DECISION MAKING AND IN MODELING IN ECONOMICS

Novel pricing strategies for revenue maximization and demand learning using an exploration–exploitation framework

Dina Elreedy[1] · Amir F. Atiya[1] · Samir I. Shaheen[1]

Accepted: 14 July 2021 / Published online: 25 July 2021

Abstract
The price demand relation is a fundamental concept that models how price affects the sale of a product. It is critical to have an accurate estimate of its parameters, as it will impact the company's revenue. The learning has to be performed very efficiently using a small window of a few test points, because of the rapid changes in price demand parameters due to seasonality and fluctuations. However, there are conflicting goals when seeking the two objectives of revenue maximization and demand learning, known as the learn/earn trade-off. This is akin to the exploration/exploitation trade-off that we encounter in machine learning and optimization algorithms. In this paper, we consider the problem of price demand function estimation, taking into account its exploration–exploitation characteristic. We design a new objective function that combines both aspects. This objective function is essentially the revenue minus a term that measures the error in parameter estimates. Recursive algorithms that optimize this objective function are derived. The proposed method outperforms other existing approaches.

Keywords Revenue management · Dynamic pricing · Demand learning · Exploration–exploitation trade-off · Price experimentation · Sequential decision problems

1 Introduction

In the field of business, companies offer products and services, and they seek to maximize the revenue achieved by these sales. Determining the right price is crucial for obtaining the optimal revenue, and this is controlled by the well-known price-demand relation. Setting a high price will drive customers away and therefore reduce demand. On the other hand, choosing a low price will lead to increasing demand, but lower revenue due to the lower price. Companies attempt to set an optimal price that maximizes revenue based on their knowledge of price-demand relation. However, the shape or the parameters are not known beforehand, and have to be inferred from actual selling situations. This may have corporations test a number of different prices, in order to learn the demand curve parameters.

Some firms could perform the price experimentation as a part of the market research phase before the actual business operation. For example, the companies selling their products on the internet can utilize digital price tags to gather price-demand data for online customers (den Boer 2015). However, other firms could have business constraints on the frequency of price changes for their products (Cheung et al. 2017; Chen and Chao 2019; Rhuggenaath et al. 2019)). Moreover, excessive price experimentation may lead to a long initial period of non-optimal pricing, and will therefore compromise the revenue. On the other hand, too little experimentation may be insufficient to discover accurate parameter values.

Generally, price experimentation is used to learn the demand model by testing a number of prices in order to estimate the price demand relation. This is known in literature as the learning problem. On the other hand, companies should also seek to choose the optimal price that maximizes the gained revenue, which is known as the earning problem. Typically, there is an inherent trade-off between these two problems, named as the learning/earning trade-off (Rothschild 1974; Cheung et al. 2017). It is akin to the trade-

✉ Dina Elreedy
dinaelreedy@eng.cu.edu.eg

Amir F. Atiya
amir@alumni.caltech.edu

Samir I. Shaheen
sshaheen@ieee.org

[1] Computer Engineering Department, Cairo University, Giza 12613, Egypt

off of exploration versus exploitation that we encounter in machine learning and evolutionary optimization (Tokic 2010; Črepinšek et al. 2013; Rezaei and Safavi 2020; Jerebic et al. 2021; Mahesh and Sushnigdha 2021).

Fast and accurate estimation of the demand curve becomes particularly important for the novel field of dynamic pricing for revenue management (Bertsimas and Perakis 2006; Besbes et al. 2014; den Boer 2012; den Boer 2015). Dynamic pricing means pricing the product in a time varying way, according to the changes in demand, in order to maximize revenue (Ibrahim and Atiya 2016). Dynamic pricing has proved its powerful impact in various applications such as hotel revenue management (Bayoumi et al. 2013), airline industry (McAfee and Te Velde 2006), mobile data services (Elreedy et al. 2017), electricity (Triki and Violi 2009), and e-services (Xia and Dube 2007).

The problem with dynamic pricing is that firms usually do not know the underlying demand price relation that characterizes customers' response upon any price change. Moreover, the price demand curve shifts frequently with time and with seasonal fluctuations (which is the reason why we would apply dynamic pricing). The learning window is therefore too short, and one has to make the most out of few data.

Another factor that could result in sudden shift in demand is catastrophic events such as wars, economic downturns, or pandemics like COVID-19. Also some lesser effects, such fluctuations of demand by season or due to shift in fashion tastes, lead to smaller and more gradual shifts of the demand curve. This necessitates speedy learning of the new demand relation. A timely algorithm that can quickly track the new demand variations, like the methods we propose here, would be very useful.

In this paper, we make use of the knowledge in the machine learning field of the exploration versus exploitation concept, in order to solve the problem of price demand function determination. Initially, the algorithm is more focused on exploration. It is a discovery phase with the goal being to accurately estimate the parameters of the price demand relation. Gradually, the algorithm shifts to exploitation, where it puts more attention toward revenue maximization (rather than exploration of the parameter space). Moreover, in the proposed approach we make use of machine learning approaches and signal processing algorithms, to explore efficient algorithms for learning the demand function. Specifically, we propose an objective function that is a combination of revenue and accuracy of the parameter estimates. Revenue is the ultimate goal that needs to be optimized. However, parameter estimate accuracy will positively impact future revenue. This is a novel formulation that can combine the effects of exploration and exploitation. By having a decaying weighting coefficient for the accuracy term of the objective function, exploration will gradually make way for exploitation as time goes by.

Essentially, the proposed approach formulates a sequential optimization problem, where the objective function is the revenue minus a term that measures the error or uncertainty in the price demand parameters. We propose three different formulations for handling the problem, where each corresponds to a different way of defining the parameter uncertainty.

We use a simple parametric model, assuming a linear demand curve. We consider a simple parametric model for several reasons: first, at early time steps, not much information is available, which hinders the performance of nonparametric models. In addition, generally, parametric models are less computationally intensive than nonparametric ones. Another argument raised by Keskin and Zeevi (2014) is that the linear demand function could approximate any demand function especially that firms usually do not use a very broad range of prices, they rather experiment with prices around a certain predefined price or within a certain range, where such predefined prices are set according to business considerations and marketing conditions. Operating at a narrow range means that a linear model is approximately valid. Finally, linear demand models are the dominant models used in the operations research and the economics literature (Lobo and Boyd 2003; Bertsimas and Perakis 2006; Cheng 2008; Keskin and Zeevi 2014; Besbes and Zeevi 2015).

In our work, we apply the recursive linear regression model proposed by Atiya et al. (2005) for estimating the demand curve due to its efficiency since it fits the sequential nature of the problem. We provide Sect. 4 for briefly describing the recursive linear regression model and presenting its formulation. The purpose of this work is to propose several simple, closed-form, efficient, and effective pricing strategies that can be conveniently applied by firms for revenue maximization and demand learning. We conduct a set of experiments to our proposed pricing strategies, to some standard baseline pricing strategies, and to some pricing methods in the literature. The experiments show that our proposed formulations outperform the competing methods and benchmarks in terms of the achieved revenue.

The main contributions of this work are summarized as follows:

- To the best of our knowledge, the explicit incorporation of model uncertainty is essentially novel in the context of managing the exploration versus exploitation trade-off.
- In this work, we propose several novel formulations incorporating the target objective function (revenue) and model uncertainty.
- This work presents different pricing methods that are simple and easy to implement taking into account business considerations of pricing constraints and little price experimentation.
- We apply our proposed pricing methods to real and synthetic datasets, and they achieve superior performance in

terms of the gained revenue compared to the other pricing methods in the literature including: myopic pricing, myopic pricing with dithering (Lobo and Boyd 2003), and controlled variance pricing (CVP) (den Boer and Zwart 2013).

The paper is organized as follows: Sect. 2 presents a literature review. Section 3 presents the problem formulation. Section 4 briefly describes the recursive formulation of linear regression model that is applied in our experiments. Then, our proposed pricing formulations are represented in Sect. 5. After that, Sect. 6 presents experimental results. The results are further analyzed in Sect. 7. Finally, Sect. 8 concludes the paper and mentions potential future work.

2 Related work

2.1 Dynamic pricing with demand learning

In this section, we review the work in the literature considering dynamic pricing in case of unknown demand price curve. Our work relates to the literature in both operations research and sequential optimization. Regarding the operations research literature, there are several contributions handling dynamic pricing with demand learning (comprehensive reviews are provided in Araman and Caldentey (2010); Aviv and Vulcano (2012); den Boer (2012); den Boer (2015)).

We discuss dynamic pricing in two main settings: with no inventory restrictions (i.e., infinite inventory) and finite inventory where there is a limitation on the supply of products/services to sell.

2.1.1 Infinite inventory

One intuitive dynamic pricing strategy is the greedy or myopic pricing where at each time step, the price is chosen so as to maximize the immediate revenue. Definitely, this policy is myopic and sub-optimal since this pricing strategy does not learn the demand curve parameters.

Lobo and Boyd (2003) propose a basic simple pricing policy for linear demand learning of a single product based on the simple myopic pricing policy. The authors modify the myopic pricing and introduce some exploration to it by adding a random perturbation to the myopic price.

Another work extending the simple myopic pricing is the work by den Boer and Zwart (2013). The proposed pricing policy, named controlled variance pricing (CVP), chooses the optimal price given the current estimate of the model (like myopic greedy pricing). However, the CVP policy imposes a constraint that the chosen price is not very close to the average of the prices previously selected. This constraint ensures diversity of chosen prices and incorporates some exploration to enhance the accuracy of estimating demand model parameters.

Since price experimentation is costly as pointed out in the introduction (see Sect. 1), Cheung et al. (2017) propose a dynamic pricing model with unknown demand function, and under the constraint of having a limited number of price adjustments for demand learning. The authors propose a pricing policy minimizing the worst-case regret, $O(log^m T)$, where T is the length of the sales horizon and m is the maximum limit of number of price changes. However, their model assumes that the demand function belongs to a finite set of functions.

Besbes and Zeevi (2015) investigate how model misspecification could affect revenue loss. They consider a multi-period single product pricing problem and prove that some pricing strategies based on two parameter linear demand models could converge to near-optimal pricing decisions even in case of model misspecification.

Keskin and Zeevi (2014) handle pricing not only for a single product, but also for multiple products along finite, T-time step horizon. They propose some variants of the greedy iterative least squares strategy which utilizes sequential model learning, and myopic price optimization given the learned model.

Carvalho and Puterman (2005) consider the dynamic pricing problem in the context of online pricing over the internet. They model the individual customer's response to price change as a binary random variable following binomial distribution. Their proposed pricing method maximizes the one-step look-ahead revenue using Taylor series expansion to approximate the next step revenue. Their proposed method outperforms myopic pricing. Further, Elreedy et al. (2021) develop a multi-step look-ahead pricing policy for uncertain linear demand models. Their approach incorporates future revenues into the objective function by maximizing the expected multi-step look-ahead revenue in addition to the immediate revenue. They implement two methods considering a single and two look-ahead revenues. Their approach outperforms the myopic pricing.

2.1.2 Finite inventory

There are various contributions in literature that handle finite inventory setting in the dynamic pricing with learning problem, where the seller has a fixed finite number of products to sell over a sales horizon. An example of the work considering the finite inventory setting is the work by Aviv and Pazgal (2002). The authors develop a Bayesian dynamic pricing control model where customers arrive according to a Poisson process with unknown arrival rate. However, the customer's potential buying probability is assumed to be known. Prices are derived by solving a differential equa-

tion, and in case of no solution of the equation, one of the these simple heuristics is applied: fixed pricing policy, certainty equivalent pricing (CEP), and a basic pricing policy that ignores demand uncertainty and uses initial expected values for demand parameters.

Araman and Caldentey (2009) consider a similar problem setting of finite inventory. They model the dynamic pricing problem as an intensity control problem, and propose a heuristic pricing policy based on approximating the value function of the underlying problem.

Farias and Van Roy (2010) handle dynamic pricing with finite inventory, in case of unknown demand. They consider maximizing the expected discounted revenue over an infinite time horizon. In their model, they assume that a customer buys the product/service only if his reservation price equals or exceeds the seller's price. The authors propose a heuristic pricing strategy named as decay balancing. They show that their proposed decay balancing strategy outperforms certainty equivalent pricing (CEP) (Aviv and Pazgal 2002) and the greedy strategy proposed by Araman and Caldentey (2009). In addition, the authors extend their model to handle sellers with multiple branches.

Another piece of work that considers dynamic pricing with finite inventory is proposed by Bertsimas and Perakis (2006). Since the dynamic pricing problem is a sequential optimization problem, the authors develop dynamic programming based models considering both competitive and non-competitive marketing environments, assuming perishable products. However, since dynamic programming considers the whole state space, it is intractable. Consequently, the authors propose several lower-dimensional approximations. The proposed pricing policies outperform the myopic pricing; however, these methods are still computationally intensive.

Another piece of work done by Wang et al. (2014) applies a nonparametric demand model for pricing with finite inventory constraint. The proposed model applies a sequence of shrinking pricing intervals before choosing a price within each iteration. This model achieves low regret bounds $O(n^{-1/2})$; however, it is computationally intensive.

Cao et al. (2019) develop a Bayesian pricing method for a single product in a finite time horizon with unknown customers' arrival rate. The authors assume that the customers' buying behavior is affected by the reference price. They formulate the dynamic pricing problem with the imposed assumptions using Bayesian dynamic programming. Moreover, they study how demand learning is influenced by having sufficient inventory. In addition, they analyze the impact of the reference price on the gained revenue. Price et al. (2019) use a Gaussian Process methodology to track and estimate the dynamic changes in demand, taking into consideration the necessity to unconstrain the demand (estimating the true demand in case inventory is assumed unlimited from finite inventory data). The Gaussian Process is a machine learning/statistical approach that models data as a joint multivariate Gaussian (Atiya et al. 2020).

Some dynamic pricing approaches do not use a fixed price for all customers, they rather tailor a different price per customer based on each customer's buying behavior, commonly known in the literature as personalized pricing (Aydin and Ziya 2009; Diao et al. 2011). A piece of work that develops an adjusted price per customer in case of unknown demand is presented by Morales-Enciso and Branke (2012). In this paper, the authors assume a different potential buying probability per customer. They develop two different pricing policies. One of them chooses the price maximizing the expected improvement of revenue. On the other hand, the other pricing policy selects the price maximizing the summation of expected immediate revenue and expected revenue of the next time step. However, the myopic greedy pricing policy outperforms both of their proposed pricing methods.

Another work adopting personalized pricing is developed by Ban and Keskin (2020). In this work, the authors develop a personalized pricing policy that learns the customer behavior over time horizon T. In their work, the authors model the customer behavior as a d-dimensional feature vector where only s out of the d features are the personalized ones. The authors analyze their proposed policy and prove that the expected regret of their policy is $O(s\sqrt{T}(\log d + \log T))$.

Not only product pricing, but also option pricing exhibits uncertainty in the financial market environment as indicated by (Ji and Zhou 2015; Sun et al. 2018; Chen et al. 2019; Gao et al. 2021). Several works study option pricing under the uncertain stock market. Chen et al. (2019) examine pricing the European call options under a fuzzy environment. Furthermore, Gao et al. (2021) investigate pricing the Asian rainbow option under the uncertain stock model. The authors model assets' prices as uncertain processes, and they derive pricing formulas for the Asian rainbow option.

Crises such as COVID-19 usually result in a tremendous change of customers' purchase behavior. Liu et al. (2020) analyze the impact of COVID-19 on the demand price relation. They develop a Bayesian approach for learning the demand function. In their work, the authors handle a single-product periodic-review inventory system. They adopt a multiplicative demand model where the demand is defined as the product of a price function and a random perturbation term representing the fluctuations in the market environment. The authors formulate the dynamic pricing problem as a Bayesian dynamic program to learn the demand distribution.

2.2 Studies of the exploration–exploitation trade-off

Exploration versus exploitation trade-off is studied in many contexts including: reinforcement learning (Ishii et al. 2002;

Tokic 2010; Asiain et al. 2019), dynamic pricing (Araman and Caldentey 2009; Harrison et al. 2012; den Boer and Zwart 2013; Besbes and Zeevi 2015), evolutionary optimization (Črepinšek et al. 2013; Singh and Deep 2019), sequential optimization (Martinez-Cantin et al. 2009), sequential design (Crombecq et al. 2011), and online advertising (Li et al. 2010). Furthermore, the exploration–exploitation trade-off is investigated in the context of multi-armed bandit problem setting (Auer et al. 2002; Vermorel and Mohri 2005; Valizadegan et al. 2011; Besbes et al. 2014).

Multi-armed bandit (MAB) is a class of sequential decision-making problems originally developed by Thompson (1933); Robbins (1985). Multi-armed bandit problems aim to maximize rewards, but under uncertainty and incomplete feedback about rewards, so there is a trade-off between performing an action that gathers information regarding reward (exploration), and making a decision that maximizes the immediate reward given the information gathered so far (exploitation) (Audibert et al. 2009). Many problems can be formulated using the multi-armed bandit setting such as our target problem: dynamic pricing with unknown demand (den Boer 2012), online advertising (Pandey et al. 2007), and clinical trials (Villar et al. 2015).

Trovo et al. (2015) utilize the multi-armed bandit formulation for solving the revenue maximization problem in case of unknown demand model. They propose two pricing policies that are, essentially, refined versions of the upper confidence bound (UCB) algorithm proposed by Auer (2002) to adapt the pricing problem. In addition, Rhuggenaath et al. (2020) develop an auction pricing algorithm based on one of the main multi-armed bandit algorithms: Thompson Sampling (Thompson 1933, 1935).

Reinforcement learning is extensively applied in dynamic pricing frameworks (Kutschinski et al. 2003; Cheng 2008; Han et al. 2008; Rana and Oliveira 2015). As an example of using reinforcement learning for dynamic pricing with unknown demand is the work developed by Cheng (2008) where Q-learning is applied for learning the value function, with the objective of revenue maximization. However, the reinforcement learning approach is computationally expensive, and under the constraint of having limited price experimentation. Accordingly, reinforcement learning could be challenging for the underlying problem of dynamic pricing with unknown demand curve.

Deep learning (Shrestha and Mahmood 2019) and deep reinforcement learning (Arulkumaran et al. 2017; Caviglione et al. 2020) have gained much interest in recent years. Kastius and Schlosser (2021) employ deep reinforcement learning for dynamic pricing. The authors mainly apply Deep Q-Networks (DQN) to model market competitors in e-commerce. Moreover, they develop another pricing model using a policy gradient algorithm named soft actor-critic (SAC). Furthermore, the work developed by Zhong et al. (2021) applies deep reinforcement learning to dynamic pricing in regenerative electric heating.

Recently, active learning has proved its powerfulness, especially in applications where the cost of data collection is significant (Settles 2009; Fazakis et al. 2019). Elreedy et al. (2019) propose an active learning framework for handling the exploration–exploitation trade-off in optimization problems. They apply the proposed framework to the dynamic pricing with demand learning problem.

Another approach for optimizing multiple contradictory objectives is the multi-objective evolutionary algorithms which seek to find Pareto-optimal solutions (Schaffer 1985; Curiel et al. 2012). An example of the multi-objective evolutionary algorithms is the multi-objective differential evolution (DE) algorithm developed by Awad et al. (2017). Another work by Srinivasan and Kamalakannan (2018) introduces a multi-objective genetic algorithm (MOGA) for analyzing financial data for risk management. However, generally, the performance of evolutionary algorithms is highly dependent on the applied crossover, mutation, and selection strategies. Recently, Farahani and Hajiagha (2021) employ meta-heuristic algorithms: social spider optimization (SSO) and bat algorithm (BA) along with artificial neural networks for stock price forecasting. However, generally, the performance of evolutionary algorithms is highly dependent on the applied crossover, mutation, and selection strategies.

Recently, fuzzy optimization has been applied to uncertain environments, especially in financial markets as indicated by Bisht and Srivastava (2019). For example, Li et al. (2020) design a multi-objective fuzzy optimization algorithm for portfolio selection of time-inconsistent investors.

Several game theoretical approaches have been developed for dynamic pricing in different contexts such as smart grids by Tang et al. (2019) and resource pricing by Zhu et al. (2020). For example, Zhu et al. (2020) design a dynamic pricing model for cloud computing services using game theory. Specifically, the authors model pricing and resource allocation as a Stackelberg game in order to resolve the conflict of maximizing revenues for both the software as a service (SaaS) providers that deliver software services and the infrastructure as a service (IaaS) providers that offer the infrastructure.

3 Problem formulation

In this work, we use a linear price demand model (or price elasticity model), as typically used in the economics/finance literature. The price is the main controlling variable for demand. We assume a monopolist seller who has a sufficient inventory to satisfy all potential demand, which is known in literature as infinite inventory setting. Our work considers pricing a single product over a finite selling horizon T.

We formulate a dynamic pricing problem for the case of unknown demand as a sequential optimization problem. Our work is algorithmic in general and attempts to derive efficient algorithms for tackling this problem. At each time step n, a price p_n is chosen so as to maximize a certain utility function incorporating the two objectives of demand estimation and revenue maximization. For any new price p_n, we observe the corresponding demand D_n, and this pair (p_n, D_n) is considered an extra data point that can fine tune more accurate parameter estimates (for the price demand relation). We apply a weighted least squares recursive formulation for updating these parameter estimates given the new acquired data point (p_n, D_n). This process iterates until the number of iterations defining the horizon T is reached.

The linear demand model equation is defined as follows:

$$D = a + bp + \epsilon \tag{1}$$

such that $b < 0$ and $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Let $x = [1 \quad p]^T$, so we can express the linear regression problem as:

$$y = \beta^T x + \epsilon \tag{2}$$

where $\beta = [a \quad b]^T$.

4 Preliminaries: recursive formulation of weighted linear regression

In this section, we briefly describe the weighted linear regression model developed by Atiya et al. (2005) that we employ in our proposed optimization strategies. We apply such a recursive regression model because it conforms with the sequential nature of the dynamic pricing problem in case of unknown demand where at each time step a new price is tested, and the model is updated accordingly. Moreover, it becomes more computationally efficient, due to the sequential update nature.

4.1 Estimating model's parameter vector β and its covariance matrix Σ_β

In this subsection, we present the recursive formulations of the weighted linear regression for the regression model parameter's vector β, and its covariance matrix Σ_β using the work presented in (Atiya et al. 2005).

Let x_n be the d-dimensional vector example, picked at time n, and let y_n be the predicted response variable, which defines the demand in our problem. In addition, let $\hat{\beta}$ be the $d \times 1$ estimated coefficient vector $[\hat{a} \quad \hat{b}]^T$ used for the linear prediction, for the linear demand estimation problem ($d = 2$) according to Eq. (1). A discounted error function is defined as follows:

$$E(T) = \sum_{n=1}^{T} \gamma^{T-n} \left[{x_n}^T \hat{\beta} - y_n \right]^2 \tag{3}$$

where γ is the discount factor, such that $0 < \gamma \leq 1$, and usually γ is set close to 1. Define the matrix X, where the rows of X are the input vectors ${x_n}^T$. Similarly, let y represent the vector of target outputs y_n, and let W denote the discount matrix, which is a $n \times n$ diagonal matrix with $W_{nn} = \gamma^{T-n}$. Then, the estimated model parameter $\hat{\beta}$ is given by the least square solution formula according to (Atiya et al. 2005) as follows:

$$\hat{\beta} = (X^T W X)^{-1} X^T W y \tag{4}$$

However, evaluating Eq. (4) in a continuous manner is computationally extensive, so recursive formulas are used.

Similarly, as indicated in (Atiya et al. 2005), the covariance matrix of β can be calculated as follows:

$$\Sigma_\beta = \sigma^2 (X^T W X)^{-1} \tag{5}$$

When a new data point comes at instant n, the parameter vector is updated recursively. According to (Atiya et al. 2005), the recursive update for the model parameter $\beta(n)$ in terms of previous estimates is:

$$\hat{\beta}(n) = \hat{\beta}(n-1) + \frac{\Sigma_\beta(n-1) x_n \left[y_n - {x_n}^T \hat{\beta}(n-1) \right]}{\sigma^2 \gamma + {x_n}^T \Sigma_\beta(n-1) x_n} \tag{6}$$

Similarly, the recursive formula for the covariance matrix $\Sigma_\beta(n)$ can be written as follows:

$$\Sigma_\beta(n) = \frac{1}{\gamma} \Sigma_\beta(n-1) - \frac{\Sigma_\beta(n-1) x_n {x_n}^T \Sigma_\beta(n-1)}{\sigma^2 \gamma^2 + \gamma {x_n}^T \Sigma_\beta(n-1) x_n} \tag{7}$$

4.2 Estimating variance of random error term (σ^2)

In the last subsection, we showed the recursive formulas for the regression model's parameter vector β, and covariance matrix Σ_β using the work of Atiya et al. (2005). However, there is still an unknown parameter not explicitly considered in (Atiya et al. 2005), namely the variance σ^2 of the ϵ error term. Accordingly, in this subsection, we estimate the variance parameter σ^2 recursively using the maximum likelihood estimator.

The likelihood function can be expressed as:

$$\mathcal{L}(\sigma^2, \beta) = \prod_{n=1}^{T} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{\sum_{n=1}^{T} -\gamma^{T-n} (y_n - \beta^T x_n)^2}{2\sigma^2}} \tag{8}$$

where T denotes the number of data points used in the estimate and γ is the discount factor of the weighted linear regression. Accordingly, the log likelihood can be calculated as follows:

$$l(\sigma^2, \beta) = -T \log \sigma - T \log \sqrt{2\pi} - \frac{\sum_{n=1}^{T} \gamma^{T-n}(y_n - \beta^T x_n)^2}{2\sigma^2} \quad (9)$$

Maximizing the log likelihood in Eq. (9) results in the following estimate $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{\sum_{n=1}^{T} \gamma^{T-n}(y_n - \beta^T x_n)^2}{T} \quad (10)$$

which represents an estimate of the variance of data.

A recursive version of the above formula can be written as

$$\hat{\sigma}^2(n) = \frac{\gamma(n-1)}{n}\sigma^2(n-1) + \frac{e^2(n)}{n} \quad (11)$$

where $e(n) = y_n - \beta^T x_n$.

5 Formulations of pricing policies

In the proposed dynamic pricing formulations, we seek to optimize both objectives of maximizing the immediate revenue (exploitation), and minimizing the uncertainty of demand model parameters (exploration). This is achieved by combining the two objectives into one hybrid utility function in three different ways. At each time step n, the price value maximizing the expected utility is used as the pricing for the next period. This price choice would simultaneously achieve good revenue and provide some exploration to test different portions of the price space in order to obtain better parameter estimates. Every successive step would provide gains in parameter accuracy, until ultimate exploration would almost no longer be necessary, and exploitation (i.e., focusing on just maximization of the revenue) would dominate.

The general form of the considered constrained optimization problem at any time step n can be expressed as follows:

$$max_{p^*} \;\; E[U(p^*)_n | \beta_{n-1}] \;\; s.t. \;\; p_l \leq p^* \leq p_u \quad (12)$$

where β_{n-1} is the estimated regression model parameters at time step $n-1$, U_n is the utility to be maximized, and p_l and p_h are imposed price bounds which are set by business owners to keep the prices in a controlled range. The utility function $U(p^*)_n$ consists of the revenue $R(p^*)$ for the selected price (exploitation term), minus a term that measures the uncertainty or error in parameter estimates (exploration term). The coefficient multiplier of the exploration term η, presented in the three formulations (Eqs. 13, 21 and 26, decays with iteration, as the initial emphasis on exploration will gradually give way to more exploitation as we proceed with the iterations. After solving the constrained optimization problem defined in Eq. (12), then the price at time step n, p_n is set to p^*. We propose three different formulations, with each suggesting a different parameter uncertainty term.

Exploration means inspecting the parameter space, and in the process narrowing down onto the true parameter values, thereby reducing the uncertainty. At the beginning uncertainty is high, but the more we explore, the more information about the parameters will be uncovered and uncertainty will decrease.

In the three proposed formulations, exploration is performed by minimizing different forms of model parameters' variances. The reason for adopting the variances of model parameters to express exploration is that the ultimate objective of exploration is minimizing the model estimation error. Furthermore, the model estimation error can be expressed in terms of the variances of the model parameters due to the bias-variance decomposition of the learning model error (Geman et al. 1992; Taieb and Atiya 2015; Elreedy and Atiya 2019). The model bias results from model misspecification. On the other hand, the model variance is caused by the disparity of the model performance when learning using different sets of training samples. Increasing training data points reduces the model variance (Elreedy and Atiya 2019).

5.1 Formulation 1

The first proposed utility function aggregates the immediate revenue $R(p^*)$, and the model uncertainty expressed in terms of the total summation of variances of the estimated model parameters $tr[\Sigma_\beta]$, i.e., equal to the trace of covariance matrix Σ_β. However, for keeping units consistent, the square root of the trace of $[\Sigma_\beta]$ is taken. Consequently, the utility function of a certain price p^* is defined as:

$$U(p^*) = R(p^*) - \eta\sqrt{tr[\Sigma_\beta]} \quad (13)$$

where η represents the trade-off parameter between exploitation (choosing a price maximizing the gained revenue) and exploration (choosing a price minimizing the model uncertainty). We consider η to be exponentially decreasing in time according to Eq. (14). At early iterations, more emphasis is imposed on exploration in order to have better estimate for the demand model parameters. However, at later iterations since the model estimates improve over time, more attention should be devoted to the ultimate goal of revenue maximization. This setting of η is applied for all of the three formulations, and it is given by:

$$\eta = \eta_0 e^{-\alpha n} \quad (14)$$

where n is the time step and $\alpha > 0$. Taking the expectation of the utility function defined in Eq. (13):

$$E[U(p^*)_n] = E[R(p^*)_n] - \eta\sqrt{tr[\Sigma_\beta(n)]} \tag{15}$$

The expected revenue $E[R(p^*)]$ for linear demand model is calculated as follows:

$$E[R(p^*)] = p^*(a + bp^*) = b{p^*}^2 + ap^* \tag{16}$$

Substituting from Eq. (7) and Eq. (16) into Eq. (15) results in:

$$E[U(p^*)_n] = b{p^*}^2 + ap^* - \eta\frac{1}{2\sqrt{tr[\Sigma_{\beta(n)}]}}tr \left[\frac{1}{\gamma}\Sigma_\beta(n-1) - \frac{\Sigma_\beta(n-1)x_n{x_n}^T\Sigma_\beta(n-1)}{\sigma^2\gamma^2 + \gamma{x_n}^T\Sigma_\beta(n-1)x_n}\right] \tag{17}$$

where $x_n = [1 \;\; p^*]^T$.
Since our target is to find the price p^* that maximizes the expected utility function defined in Eq. (17), we evaluate the derivative of $E[U(p^*)_n]$ w.r.t. p^*:

$$\frac{\partial E[U(p^*)_n]}{\partial p^*} = a + 2bp^* + \eta\frac{tr\left[\frac{1}{g^2(p^*)}\Sigma_\beta^2(n-1)Z(p*)\right]}{2\sqrt{tr[\Sigma_{\beta(n)}]}} \tag{18}$$

where $g(p^*) = (\sigma^2\gamma^2 + \gamma[\sigma_a^2 + 2\sigma_{ab}p^* + {\sigma_b}^2{p^*}^2])$, and $Z(p*)$ is a 2×2 matrix with elements: Z_{11}, Z_{12}, and Z_{22} given as

$$\begin{aligned} Z_{11} &= -2\gamma(\sigma_{ab} + {\sigma_b}^2p^*) \\ Z_{12} &= \gamma(\sigma^2\gamma + \sigma_a^2 - \sigma_b^2{p^*}^2) \\ Z_{22} &= \gamma(2\sigma^2\gamma p^* + 2\sigma_a^2p^* + 2\sigma_{ab}{p^*}^2) \end{aligned} \tag{19}$$

Then, by equating Eq. (18) to zero and solving the resulting equation, we can get the price p^* maximizing the expected utility function at time step n using a simple one-dimensional search.

$$\frac{\partial E[U(p^*)_n]}{\partial p^*} = a + 2bp^* + \eta\frac{tr\left[\frac{1}{g^2(p^*)}\Sigma_\beta^2(n-1)Z(p*)\right]}{2\sqrt{tr[\Sigma_{\beta(n)}]}} = 0 \tag{20}$$

The details of the derivative computation of the expected utility of this formulation, defined in Eq. (17), can be found in Appendix A.

5.2 Formulation 2

Similar to the first formulation, we define a utility function in terms of the immediate revenue $R(p^*)$ and model uncertainty. However, the model uncertainty in this formulation is expressed as a summation of normalized standard deviations of model parameters σ_a and σ_b We normalize the standard deviations σ_a and σ_b in order to have the uncertainty relative to the value of the parameters. For example, consider a problem where $a = 1000$, and another one where $a = 10$, and if the standard deviation $\sigma_a = 5$, this value for the uncertainty in parameter a would be more significant for the case of $a = 10$ than for $a = 1000$.
The proposed utility function can then be written as:

$$U(p^*) = R(p^*) - \eta\left(\frac{\sigma_a}{a} + \frac{\sigma_b}{|b|}\right) \tag{21}$$

Calculating the expectation of the utility function defined in Eq. (21):

$$E[U(p^*)_n] = E[R(p^*)_n] - \eta\left(\frac{\sigma_a(n)}{a} + \frac{\sigma_b(n)}{|b|}\right) \tag{22}$$

Using Eq. (7) and the definition of $g(p^*)$ in formulation 1, Sect. 5.1, accordingly, the expected utility can be calculated as:

$$E[U(p^*)_n] = E[R(p^*)_n] - \eta\left(\frac{\Sigma_\beta(n)_{11}}{a} + \frac{\Sigma_\beta(n)_{22}}{|b|}\right) \tag{23}$$

The first derivative of the expected utility $\frac{\partial E[U(p^*)_n]}{\partial p^*}$ with respect to p^* can be evaluated as follows; the details are presented in Appendix A.

$$\begin{aligned} \frac{\partial E[U(p^*)_n]}{\partial p^*} &= a + 2bp^* \\ &+\eta\frac{\gamma}{2ag^2(p^*)\sqrt{\Sigma_{\beta_{11}}(n)}} \\ &\times\Big({p^*}^2({\sigma_{ab}}^3 - \sigma_{ab}\sigma_a^2\sigma_b^2) + p^*(\sigma^2\gamma\sigma_{ab}^2 \\ &-\sigma_a^4\sigma_b^2) + \sigma^2\gamma\sigma_a^2\sigma_{ab}\Big) \\ &+\eta\frac{\gamma}{2|b|\sqrt{\Sigma_{\beta_{22}}(n)}g^2(p^*)} \times \Big(p^*(\sigma^2\gamma\sigma_b^4 - \sigma_{ab}^2\sigma_b^2 \\ &+\sigma_b^4\sigma_a^2) + (\sigma^2\gamma\sigma_{ab}\sigma_b^2 + \sigma_b^2\sigma_a^2\sigma_{ab} - \sigma_{ab}^3)\Big) \end{aligned} \tag{24}$$

Similar to formulation 1, by equating Eq. (24) to zero, and solving the resulting equation, we can get the price p^* maximizing the expected utility function at time step n using a simple one-dimensional search.

$$\frac{\partial E[U(p^*)_n]}{\partial p^*} = a + 2bp^* + \eta \frac{\gamma}{2ag^2(p^*)\sqrt{\Sigma_{\beta 11}(n)}} \times \Big(p^{*2}(\sigma_{ab}{}^3 - \sigma_{ab}\sigma_a^2\sigma_b^2) + p^*(\sigma^2\gamma\sigma_{ab}^2 - \sigma_a^4\sigma_b^2) + \sigma^2\gamma\sigma_a^2\sigma_{ab} \Big) + \eta \frac{\gamma}{2|b|\sqrt{\Sigma_{\beta 22}(n)}g^2(p^*)} \Big(p^*(\sigma^2\gamma\sigma_b^4 - \sigma_{ab}^2\sigma_b^2 + \sigma_b^4\sigma_a^2) + (\sigma^2\gamma\sigma_{ab}\sigma_b^2 + \sigma_b^2\sigma_a^2\sigma_{ab} - \sigma_{ab}^3) \Big) = 0 \quad (25)$$

The details of deriving the derivative of the expected utility defined in Eq. (22) are presented in the appendix.

5.3 Formulation 3

For the third proposed formulation, we define the utility function in terms of the immediate revenue $R(p^*)$, but the focus here is on the uncertainty of the immediate revenue $\sigma_{R(p^*)}$, instead of uncertainty of demand model parameters. The intuition for including uncertainty of revenue in the model is to promote the potential of selecting prices that maximize the expected revenue with high confidence. Thus, the utility function is defined as:

$$U(p^*) = R(p^*) - \eta\sigma_{R(p^*)} \quad (26)$$

where $\sigma_{R(p^*)}$ is the standard deviation of revenue. Taking the expectation of the utility function:

$$E[U(p^*)_n] = E[R(p^*)_n] - \eta\sigma_{R(p^*)_n} \quad (27)$$

Given the linear elasticity demand model defined in Eq. (1), the standard deviation of revenue $\sigma_{R(p^*)}$ can be calculated as follows:

$$\sigma_{R(p^*)} = p^*\sigma_y = p^*\sqrt{x^{*T}\Sigma_{\beta(n-1)}x^* + \sigma^2} \quad (28)$$

Accordingly, the utility function can be expressed as:

$$E[U(p^*)_n] = E[R(p^*)_n] - \eta p^*\sqrt{x^{*T}\Sigma_{\beta(n-1)}x^* + \sigma^2} \quad (29)$$

The derivative of expected utility with respect to p^*, $\frac{\partial E[U(p^*)_n]}{\partial p^*}$ is evaluated as follows:

$$\frac{\partial E[U(p^*)_n]}{\partial p^*} = a + 2bp^* - \eta \frac{2\sigma_b{}^2 p^{*2} + 3\sigma_{ab}p^* + \sigma_a{}^2 + \sigma^2}{\sqrt{(\sigma^2 + \sigma_a^2 + 2\sigma_{ab}p^* + \sigma_b^2 p^{*2})}} \quad (30)$$

As the two formulations above, by equating Eq. (30) to zero, and solving the resulting equation, we can get the price p^* maximizing the expected utility function at time step n using a simple one-dimensional search.

$$\frac{\partial E[U(p^*)_n]}{\partial p^*} = a + 2bp^* - \eta \frac{2\sigma_b{}^2 p^{*2} + 3\sigma_{ab}p^* + \sigma_a{}^2 + \sigma^2}{\sqrt{(\sigma^2 + \sigma_a^2 + 2\sigma_{ab}p^* + \sigma_b^2 p^{*2})}} = 0 \quad (31)$$

We provide the details of calculating the derivative of the expected utility defined in Eq. (27) in Appendix A.

6 Experiments

To test the performance of the proposed approaches, we have applied them to different pricing problems. In order to explore the standing of the proposed methods compared to other existing approaches, we have also applied some benchmark or baseline price demand estimation methods, and some other algorithms proposed in the literature.

6.1 Benchmarks

One benchmark pricing strategy that we apply is the basic myopic pricing policy, which selects the price maximizing the immediate revenue at each time step. Such price is estimated as $\frac{-\hat{a}}{2\hat{b}}$ for the standard linear demand model. Clearly, this pricing strategy greedily focuses on exploitation only. In addition, we compare our proposed methods to two other strategies from the literature, the myopic pricing with dithering proposed by Lobo and Boyd (2003), and the controlled variance pricing (CVP) policy proposed by den Boer and Zwart (2013). We have briefly described these methods in Sect. 2.

Furthermore, we investigate a strategy consisting of two phases: exploration then exploitation. In this strategy the first phase of exploration (for example in the first half of the period) is essentially performed in order to obtain an accurate estimate of model parameters. In the next phase (the remaining portion of the considered period), we use the estimated model, and apply pure exploitation by applying the greedy myopic pricing policy. We consider two variants of this two-phase approach: the random-myopic policy where the exploratory phase is performed by selecting random prices, and then exploitation is performed by means of myopic pricing. Similarly, the second approach is the uncertain-myopic pricing whereby the exploratory phase is performed by minimizing the model uncertainty, expressed as the summation of variances of the two model parameters a and b. Follow-

ing this, the exploitation phase is performed using myopic pricing.

6.2 Performance metrics

We evaluate the performance of the different pricing policies with respect to two main objectives. The primary objective is revenue maximization, while the secondary objective is the accuracy of the estimated demand. The revenue management objective is basically the revenue gain, or a normalized version of the total discounted revenue $Rev(T)$ achieved in the considered time period, as follows:

$$Rev\ Gain = \frac{Rev(T)}{Rev_{opt}} = \frac{\sum_{n=1}^{T} \gamma^{n-1} R(n)}{\sum_{n=1}^{T} \gamma^{n-1} R_{opt}} \tag{32}$$

where $R(n)$ is the revenue in step n and R_{opt} is the optimal revenue given the true model parameters a and b, which is calculated as:

$$R_{opt} = p_{opt}(a + bp_{opt}) = bp^2{}_{opt} + a\,p_{opt} \tag{33}$$

where p_{opt} is the optimal price, which equals to $\frac{-a}{2b}$ for our case of linear demand model where a and b are the ground truth values for the linear demand model parameters.

Simplifying $\sum_{n=1}^{T} \gamma^{n-1}$ by using the summation of geometric series formula, this becomes:

$$Rev\ Gain = \frac{Rev(T)}{Rev_{opt}} = \frac{\sum_{n=1}^{T} \gamma^{n-1} R(n)}{(1-\gamma^T)/(1-\gamma) R_{opt}} \tag{34}$$

In addition to evaluating the gained revenue, we test whether the final price converges to the true optimal price by measuring the deviation of the price p_T, at last iteration T, from the true optimal price p_{opt}.

$$\delta_p = \frac{|p_T - p_{opt}|}{p_{opt}} \tag{35}$$

Concerning the demand model estimation accuracy, we evaluate it in terms of the deviation of the final estimated demand model parameters $\hat{\beta}_T$, at iteration T, from the true parameter's vector β as shown in Eq.(36):

$$\delta_\beta = \frac{||\beta - \hat{\beta}_T||_2}{||\beta||_2} \tag{36}$$

6.3 Experimental setup

The simulation proceeds as follows: after generating a pool of price-demand data, we start with a very limited number of points, $N_0 = 3$ points (less than three points cannot give any sensible initial parameter estimate). Then, we train a regression model to obtain an initial estimate for the model parameters β_0, and the corresponding covariance matrix Σ_{β_0}. After that, we apply the proposed sequential optimization methods (which maximize the utility function) in order to obtain the optimal price at iteration n, denoted as p_n. The optimization is under the constraint that p_n is within the pricing interval defined by the seller where the minimum allowable price is p_l, and the maximum possible price p_u, i.e., $p_l \leq p_n \leq p_u$. Once the price is determined, the demand D_n is observed. It follows the linear demand model (Eq. (1)), with of course the error term ϵ giving random fluctuations around the true demand line. We use this point (p_n, D_n) to update the model estimates β and Σ_β using recursive weighted linear regression update equations (Eqs. 6 and 7). The simulation loop continues till reaching a certain predefined number of iterations T. For each dataset, we run the experiment 100 runs and we present the average results over the runs.

One can observe from the equations of three proposed utility functions (Eqs. 17, 23 and 29) that the true values of demand model a and b are present in parts of the formulas that determine the price. However, since the demand model parameters are unknown, we use current estimates of model parameters $\hat{a}_{n-1}$ and $\hat{b}_{n-1}$, respectively, at each time step n.

In our experiments, we set the number of iterations T to 100, and the discount factor of the weighted linear regression, γ is set to 0.99. Since the optimization problem is over one variable, the price p, any simple grid search over the pricing values could be used. In our implementation, we use the interior point optimization algorithm (Byrd et al. 1999). Regarding the exploration–exploitation hyper-parameter α presented in Eq. (14), we set α such that at the last iteration T, where the exploration is nearly diminished, η equals to a small value: $\eta = 0.25$. For η_0, we use values that make the weights (impacts) of the two underlying objectives of revenue and model uncertainty comparable at the first iteration.

In our implementation, for the considered two-phase benchmark strategies we use the same number of iterations for the exploration phase as for the exploitation phase, i.e., 50 for each. Regarding the myopic pricing with dithering method (Lobo and Boyd 2003), we set the amount of dithering to 0.1.

We use a unified method for estimating the demand model parameters for all pricing methods, which is the weighted recursive linear regression described in Sect. 4 in order to have a fair comparison among the different pricing policies.

6.4 On price-demand elasticity

In our experiments, we test several values for the demand slope parameter b in order to explore the performance for three main cases of demand elasticity ranges (to be described shortly). Elasticity is defined as the ratio of the percentage change in demand change to the percentage change in price

change (see Eq. (37) and refer to (Gillespie 2014; Gwartney et al. 2014)).

$$Elasticity = \frac{\Delta_D\%}{\Delta_p\%} \tag{37}$$

where Δ_p denotes the price change, and Δ_D is the corresponding demand change. The elasticity parameter is related to the slope of linear demand model b in Eq. (2). Naturally, demand elasticity is negative because of the inverse relation between price and demand.

The demand-price elasticity varies for different types of products or services. Demand can be inelastic (elasticity < 1), e.g., for necessities or indispensable products, neutrally elastic (elasticity ≈ 1), and elastic (elasticity > 1), e.g., for luxury goods. We test the performance of our proposed methods for each of these three cases by setting appropriate values for the elasticity parameter b.

6.5 Experiments using synthetic datasets

First, we apply our proposed methods as well as the other pricing methods and benchmarks to artificial datasets. The advantage of using artificial data is that the true model parameters $\beta = [a \quad b]^T$ are known. Therefore, the revenue gain can be accurately estimated with the knowledge of the true optimal revenue. Moreover, the estimation error of demand model parameter's vector β can be accurately evaluated. We create synthetic datasets by generating several price points and then assuming linear demand model, we calculate the corresponding demands using Eq. (1). We adopt different values for the standard deviation σ of the error term ϵ, so that we can analyze the impact of the error term on the different pricing policies, and evaluate their immunity toward errors. Moreover, we use different values for the variance of the error term because it can be conceived as aggregating all other influencing factors that may be hard to model, such as competition, seasonality, or perishability of the products.

We generate twenty different synthetic datasets using diverse values for parameters a, b, and σ. Specifically, we investigate different values for the parameter b including the three demand elasticity cases of inelastic, neutral, and elastic demands. The detailed results for revenue gain, parameter accuracy, and price convergence are represented in Tables 1, 2, and 3, respectively.

Tables 1, 2, and 3 represent the gain in revenue, the estimation error of model parameter's vector β, and the percentage error of the estimated price with respect to the optimal price, respectively. These tables show the results averaged over the twenty synthetic datasets in case of low error setting and high error setting.

In order to investigate the behavior of different pricing methods over time horizon T, we provide, as an example, the figures for one artificial dataset with $a = 1000$, $b = -1$, and $\sigma = 200$. Figure 1 shows the cumulative discounted revenue for different methods over time steps of the horizon. Figure 2 shows the model percentage error for regression coefficients β using different pricing methods at different time steps. Figure 3 represents the chosen price at different iterations by different methods.

Table 1 Revenue gain of different methods, averaged over twenty different synthetic datasets over two different settings of the standard deviation of the error term

Method	Low error setting (%)	High error setting (%)	Average (%)
Form2	**98.88**	**96.09**	**97.49**
Form1	98.11	92.47	95.29
CVP	95.77	94.25	95.01
Form3	94.95	88.52	91.73
Myopic	93.67	77.26	85.46
Myopic-dith	94.40	76.30	85.35
Rand-Myopic	79.22	78.81	79.02
Uncertain-Myopic	51.93	47.78	49.86

he methods are sorted descendingly according to their average revenue gain over the two settings of the standard deviation of the error term. The bold entries represent the maximum revenue gain per column (over all strategies)

Table 2 Percentage error in estimating model parameter's vector β of different methods, averaged over twenty different synthetic datasets over two different settings of the standard deviation of the error term

Method	Low Error Setting (%)	High Error Setting (%)	Average (%)
Uncertain-Myopic	**0.64**	**2.68**	**1.66**
Rand-Myopic	0.77	3.27	2.02
Form3	1.04	3.53	2.29
CVP	1.11	4.08	2.60
Form2	1.44	5.64	3.54
Myopic	1.58	5.76	3.67
Myopic-dith	1.53	5.84	3.69
Form1	1.70	6.01	3.85

The methods are sorted ascendingly according to their average percentage model error over the two settings of the standard deviation of the error term. The bold entries represent the minimum model error per column (over all strategies)

6.6 Experiments using real parameter sets

To have more realistic parameter values, we have adopted seven real datasets of nineteen different products described in Table 4. First, we have gathered some data online though surveys. The dataset is a transportation ticket pricing data,

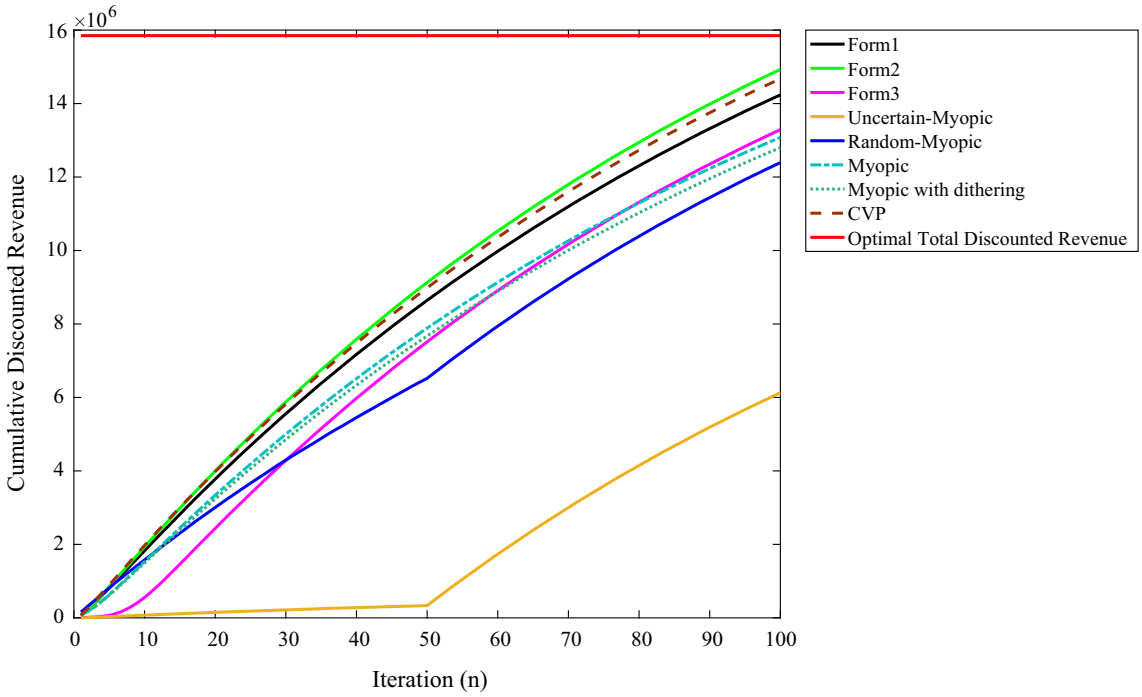

Fig. 1 Cumulative discounted revenue using different formulations for the synthetic dataset $a = 1000$, $b = -1$, and $\sigma = 200$

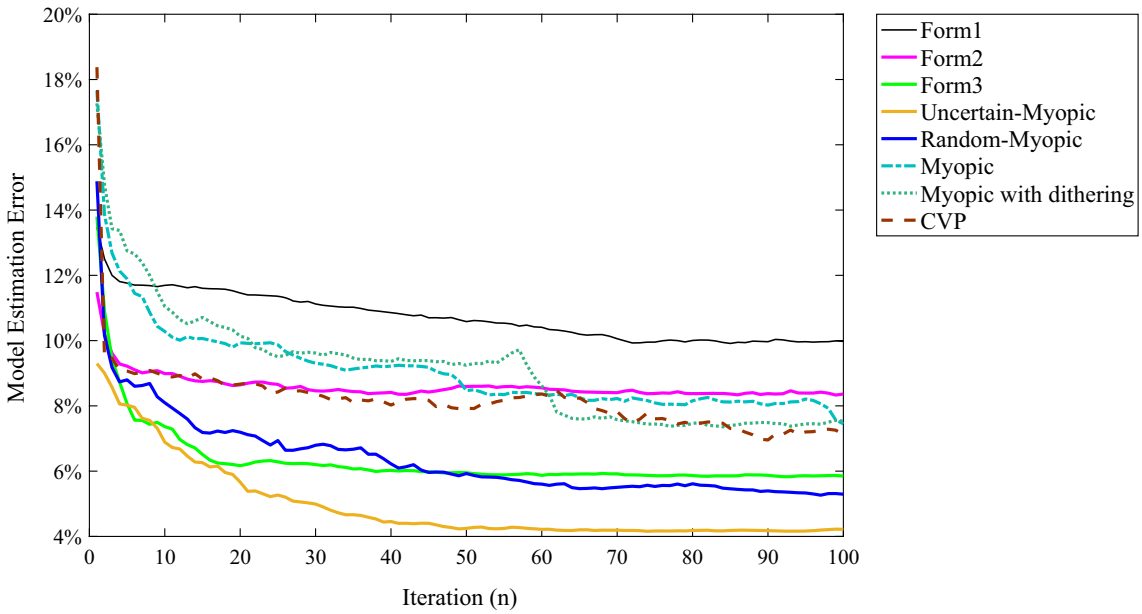

Fig. 2 Estimated regression model percentage error using different formulations for the synthetic dataset $a = 1000$, $b = -1$, and $\sigma = 200$

where we ask users about the minimum and the maximum fares they would pay for an economy class bus ticket between any generic certain two cities. We collected 41 responses from different users. In order to have data in the form of price and demand pairs, we perform the following. For each price, we calculate the corresponding demand as the number of users who can afford this price according to their stated minimum and maximum prices.

Another dataset is the so-called beef dataset. It is obtained from the USDA Red Meats Yearbook Library (2001). Similarly, the sugar dataset is adopted from Schultz (1933). The

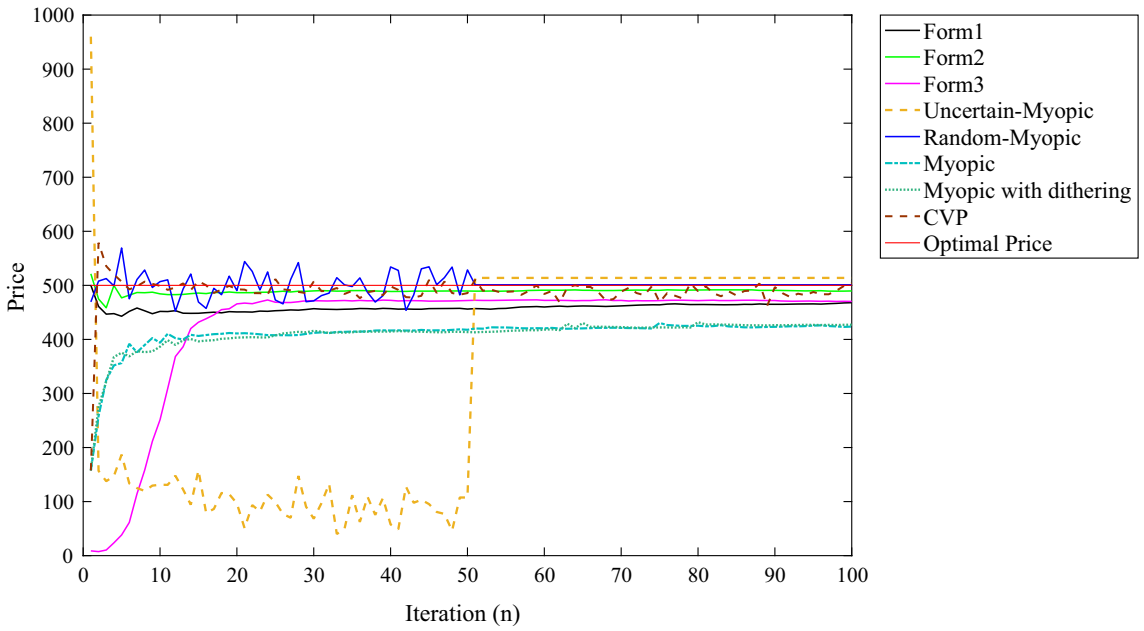

Fig. 3 Estimated prices using different formulations for the synthetic dataset $a = 1000$, $b = -1$, and $\sigma = 200$

Table 3 Percentage error of the final estimated price p_T for all methods, averaged over twenty different synthetic datasets over two different settings of the standard deviation of the error term

Method	Low Error Setting (%)	High Error Setting (%)	Average (%)
Rand-Myopic	0.19	**0.72**	**0.45**
Form3	0.26	1.49	0.88
Uncertain-Myopic	**0.13**	1.68	0.91
Form2	0.33	1.68	1.01
Form1	0.33	2.08	1.20
CVP	0.84	1.89	1.36
Myopic	4.87	20.44	12.66
Myopic-dith	4.15	21.28	12.71

The methods are sorted ascendingly according to their average price deviation over the two settings of the standard deviation of the error term. The bold entries represent the minimum price deviation per column (over all strategies)

Table 4 A description for the real-world datasets

Dataset	Size	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$
Transport	41	41.3778	−0.1378	3.3902
Beef	91	30.0515	−0.0465	0.5670
Sugar	18	1.3576	−0.3184	0.0292
Spirits	69	4.4651	−1.2723	0.0573
Coke	20	50.5700	−0.3406	1.9319
Café-Burger	1351	189.6795	−7.1411	15.6471
Café-1	1351	54.2005	−2.0234	6.1317
Café-2	1351	47.7671	−2.2588	4.1790
Café-3	1351	108.9627	−5.2635	8.6968
HOBBIES-1-001	146	206.9059	− 21.5411	7.3863
HOBBIES-1-028	274	132.2558	− 13.3455	6.9413
HOBBIES-1-046	72	341.6164	− 16.4479	7.1373
HOBBIES-1-207	274	231.1926	− 77.4905	7.8952
HOBBIES-2-045	199	79.4778	− 25.4614	6.5352
HOUSEHOLD-1-164	115	232.3479	−52.0884	20.9894
HOUSEHOLD-2-089	274	104.2676	− 21.9714	19.1267
HOUSEHOLD-2-505	274	132.2804	−21.3876	9.7262
FOODS-3-754	165	332.5571	−48.8230	9.4418
FOODS-3-799	274	71.9769	− 28.8564	5.2721

spirits dataset is obtained from Durbin and Watson (1950), and the coke dataset is acquired from Sun (2011).

In addition, we have considered a large sales dataset of a café offering four products at a single store Zavarella (2018). The four offered products are a burger and three other meals. Each product has 1351 sales transactions. For the café dataset, we estimate the demand model for each product separately. In the experimental results, the four products are denoted as: Café-1, Café-2, Café-3, and Café-4. Furthermore, we have used the Walmart retail goods dataset offered by the University of Nicosia, named as M5 Forecasting dataset of Nicosia (2020). This dataset is quite large; it has 42,840 sales records for 3049 products of three main categories (Hobbies, Household, and Food) placed in seven departments. The M5 Forecasting dataset includes the sales data of ten stores of Walmart in three states. For space limitations, we present the results of ten different products of the M5 Fore-

casting dataset. We follow the naming convention described in (Makridakis et al. 2020) for the dataset products shown in Table 4. For example, $HOUSEHOLD-2-505$ defines a household product at department 2 with id 505.

In our sequential optimization framework, the selected price p_n at each time step n could potentially be outside the available prices provided in the dataset. Thus, we use any of the real datasets mainly for estimating linear demand model parameter's vector β only. Then, we generate data using the estimated parameters, with the same methodology described in Sect. 6.5. The regression model coefficients a and b are estimated using ordinary least squares linear regression. For the error variance parameter σ^2, we estimate it using the maximum likelihood estimator (Eq. (10) with $\gamma = 1$). The estimated parameter values for the adopted real datasets are presented in Table 4.

The following tables summarize the results of our conducted experiments on the real datasets described in Table 4, for the different pricing policies. Tables 5, 6, 7 represent the gain in revenue, the estimation error of model parameter's vector β, and the percentage error of the estimated price with respect to the optimal price, respectively.

7 Discussion

From the experiments, we observe the following findings. We categorize our findings with respect to the three performance evaluation aspects: revenue gain, model estimation error, and price convergence.

7.1 Revenue gain

- From the presented results, we can observe that our proposed models generally outperform the competing methods in terms of the achieved revenue, for most of the synthetic and real datasets as indicated in Tables 1 and 5, respectively. They obtain on average better results compared to the standard benchmarks of random-myopic, uncertain-myopic, and myopic pricing, as well as the two state-of-the art methods in revenue management literature: myopic with dithering (Lobo and Boyd 2003), and controlled variance pricing (den Boer and Zwart 2013). The reason for this outperformance of our proposed methods over other approaches is the way we incorporate both aspects of the target objective which is the gained revenue, and model uncertainty into one hybrid utility function with the aim of maximizing the immediate revenue in addition to having better estimates of model parameters that help maximizing future revenues.

- Regarding the myopic pricing policy, typically, it yields sub-optimal performance due to its greedy nature. Even for myopic pricing with dithering, the dithering level, which is a major hyper-parameter for balancing exploration and exploitation, is a user input parameter. It turns out that this method's performance is not significantly better than the myopic pricing policy. One can observe from Figure 1 that the curves of myopic pricing and myopic pricing with dithering are very close. In addition, Tables 1 and 5 show that the two methods obtain very close average revenue gains over the synthetic and real datasets, respectively.

- It could be inferred from Figure 1 that our proposed methods, especially formulation 1 and formulation 2, convincingly have superior performance in terms of the gained revenue over other methods over the whole time horizon T. In addition, Tables 1 and 5 indicate that formulation 2 is the best performing method on average, over the synthetic and real datasets, respectively.

- Table 5 demonstrates that our proposed methods, especially the second formulation, outperform other benchmarks for the real datasets. Furthermore, one can observe that the myopic and myopic with dithering strategies perform comparably well and this occurs since the inherent random error in the data is low as shown in Table 4. Similarly, for synthetic datasets with low error settings, our first and second formulations surpass the performance of other methods in terms of the gained revenue over, as indicated in Table 1.

- For the more challenging synthetic datasets with high error settings, our proposed methods have robust performance and outperform other methods in terms of the gained revenue as shown in Table 1. This is due to the explicit incorporation of model uncertainty in the underlying objective functions we optimize.

7.2 Model estimation error

- Regarding the model estimation error, Tables 2 and 6 show that the uncertain-myopic benchmark achieves the minimum estimation error. This result is essentially reasonable since the first phase of uncertain-myopic is totally devoted for minimizing model uncertainty, by explicitly minimizing the trace of covariance matrix Σ_β. Accordingly, the uncertain-myopic benchmark obtains accurate model parameters' estimates. However, the first, relatively long, exploration phase compromises revenue, and accordingly the uncertain-myopic benchmark obtain poor revenues as indicated in Tables 1 and 5 for the syn-

Table 5 Revenue gain of different methods for the nineteen products of the seven real datasets

Dataset	Form1 (%)	Form2 (%)	Form3 (%)	Uncertain-Myopic (%)	Rand-Myopic (%)	Myopic (%)	Myopic-dith (%)	CVP (%)
Transport	97.08	**98.01**	90.91	44.09	79.05	97.58	96.93	95.55
Beef	99.55	**99.61**	97.18	54.67	79.28	99.56	99.55	96.17
Sugar	98.97	99.71	99.23	72.78	87.59	**99.93**	99.86	97.32
Spirits	99.50	99.94	99.65	55.40	73.41	**99.97**	99.85	95.55
Coke	99.39	99.46	97.02	42.33	79.52	99.39	**99.48**	96.15
Café-1	98.20	98.81	95.04	44.09	79.85	**98.85**	98.77	95.98
Café-2	97.35	**98.39**	93.87	44.40	80.12	98.33	98.19	95.66
Café-3	98.27	**98.88**	95.18	47.00	81.61	98.66	98.40	96.08
Café-4	97.68	**98.82**	95.19	45.64	80.46	98.64	98.73	95.96
HOBBIES-1-001	98.84	**99.82**	98.03	51.79	80.59	**99.82**	**99.81**	96.15
HOBBIES-1-028	98.38	**99.57**	97.37	53.62	82.11	99.52	99.49	96.14
HOBBIES-1-046	99.65	**99.93**	98.69	46.31	80.46	**99.93**	**99.93**	96.18
HOBBIES-1-207	98.04	**99.79**	99.28	70.37	83.67	99.74	99.70	96.60
HOBBIES-2-045	96.23	**99.11**	98.27	73.46	86.48	**99.12**	98.99	96.79
HOUSEHOLD-1-164	96.43	**98.91**	97.16	73.58	89.87	98.69	98.68	96.71
HOUSEHOLD-2-089	90.81	**96.06**	94.58	58.77	80.26	95.48	95.24	94.47
HOUSEHOLD-2-505	97.70	**99.31**	97.03	63.33	85.21	99.29	99.30	96.53
FOODS-3-754	98.74	99.75	98.54	56.96	81.80	**99.80**	99.78	96.23
FOODS-3-799	97.44	99.26	98.63	88.62	94.08	**99.29**	99.07	98.02
Average	97.80	**99.11**	96.89	57.22	82.39	99.03	98.93	96.22

The bold entries represent the maximum revenue gain per column (over all strategies)

Table 6 Percentage error in estimating model parameter's vector β of different methods for the nineteen products of the seven real datasets

Dataset	Form1 (%)	Form2 (%)	Form3 (%)	Uncertain-Myopic (%)	Rand-Myopic (%)	Myopic (%)	Myopic-dith (%)	CVP (%)
Transport	6.10	5.64	3.19	**1.60**	2.24	5.66	5.74	3.56
Beef	0.95	0.96	0.76	**0.42**	0.47	0.97	0.88	0.77
Sugar	1.78	1.83	1.83	**0.76**	0.80	2.10	1.95	1.25
Spirits	1.20	1.22	1.03	**0.31**	0.42	1.24	1.25	0.69
Coke	3.31	3.17	2.29	**0.63**	0.97	3.60	3.34	1.90
Café-1	5.31	7.76	3.84	**1.32**	1.89	7.27	7.79	3.70
Café-2	7.36	8.61	5.10	**1.74**	2.97	8.95	9.40	4.84
Café-3	5.46	6.80	3.68	**1.48**	2.22	7.76	7.01	4.16
Café-4	6.25	7.60	4.46	**1.36**	2.08	8.24	8.16	3.73
HOBBIES-1-001	2.51	2.93	2.20	**0.64**	1.03	3.01	3.00	1.62
HOBBIES-1-028	3.82	4.55	2.94	**0.97**	1.28	4.71	4.81	2.38
HOBBIES-1-046	1.77	1.77	1.42	**0.39**	0.48	1.63	1.80	0.98
HOBBIES-1-207	2.24	3.53	2.52	**0.92**	1.25	3.75	3.46	1.93
HOBBIES-2-045	5.60	7.84	6.51	**2.37**	3.18	8.03	8.10	4.62
HOUSEHOLD-1-164	5.48	7.71	5.07	**2.12**	3.65	8.68	8.53	5.17
HOUSEHOLD-2-089	11.94	15.61	9.34	**4.29**	6.95	15.09	15.11	8.27
HOUSEHOLD-2-505	5.38	6.40	4.03	**1.68**	2.12	6.22	6.33	3.40
FOODS-2-754	2.24	3.34	2.23	**0.51**	0.88	3.23	2.93	1.44
FOODS-2-799	5.60	7.47	9.05	**2.97**	4.21	8.11	8.19	5.06
Average	4.44	5.51	3.76	**1.39**	2.06	5.70	5.67	3.13

The bold entries represent the minimum model error per column (over all strategies)

Table 7 Percentage error of the final estimated price p_T for all methods for the nineteen products of the seven real datasets

Dataset	Form1 (%)	Form2 (%)	Form3 (%)	Uncertain-Myopic (%)	Rand-Myopic (%)	Myopic (%)	Myopic-dith (%)	CVP (%)
Transport	1.45	2.11	1.36	0.60	**0.31**	1.93	3.52	2.26
Beef	**0.04**	0.33	0.06	0.31	0.18	0.18	0.20	0.38
Sugar	0.24	0.17	0.23	0.08	**0.04**	0.09	2.31	0.12
Spirits	0.23	0.25	0.17	**0.04**	**0.04**	0.31	3.27	2.27
Coke	**0.01**	0.89	0.03	0.07	0.11	1.28	0.94	0.08
Café-1	1.54	2.77	1.38	0.41	**0.32**	2.09	2.29	0.91
Café-2	0.77	**0.08**	0.55	0.64	0.79	0.48	0.72	2.45
Café-3	0.45	0.17	0.40	1.00	0.63	1.15	**0.10**	2.15
Café-4	0.43	**0.26**	0.41	0.74	0.51	0.34	0.53	0.95
HOBBIES-1-001	0.06	0.09	**0.01**	0.05	0.24	0.06	0.92	1.84
HOBBIES-1-028	0.04	0.27	**0.03**	**0.02**	0.05	0.25	1.31	2.43
HOBBIES-1-046	0.20	**0.01**	0.08	0.12	0.08	0.12	0.39	0.90
HOBBIES-1-207	0.11	0.31	0.17	0.30	**0.01**	0.05	2.73	0.73
HOBBIES-2-045	0.97	0.38	**0.02**	0.06	0.24	0.05	2.98	3.49
HOUSEHOLD-1-164	0.56	0.98	0.64	**0.18**	1.45	0.78	3.24	2.88
HOUSEHOLD-2-089	4.20	1.89	0.75	1.12	1.35	**0.36**	0.70	4.52
HOUSEHOLD-2-505	0.82	0.41	**0.10**	**0.09**	0.22	0.69	0.57	0.96
FOODS-2-754	0.09	0.07	0.38	0.13	**0.02**	0.04	1.54	1.01
FOODS-2-799	1.64	1.08	0.86	**0.24**	0.64	0.84	4.31	3.78
Average	0.73	0.66	0.40	**0.33**	0.38	0.58	1.71	1.80

The bold entries represent the minimum deviation error per column (over all strategies)

thetic and real datasets, respectively.

- Similarly, the random-myopic benchmark obtains relatively accurate model estimates as indicated in Tables 2 and 6 since the first phase is pure exploration via random sampling. In addition, the CVP method achieves low estimation error rates for synthetic and real datasets as shown in Tables 2 and 6, respectively. The reason for that is that the CVP method inherently imposes emphasis on exploration by ensuring the diversity of the chosen prices in order to improve the regression model accuracy. Consequently, the CVP method results in near-optimal values for model parameters as well. However, similar to the uncertain-myopic baseline, both of the random-myopic and the CVP method compromise the gained revenue. However, the CVP method achieves more robust performance in terms of the gained revenue, especially in high error settings since it inherently emphasizes exploration through choosing diverse prices, as indicated in Table 1.

- Tables 2 and 6 demonstrate that our proposed methods achieve comparable performance in terms of model estimation error. Moreover, our proposed methods mainly emphasize the ultimate objective: the utility (revenue) maximization, while treating the convergence to the true model parameters as an important, but a secondary objective. Furthermore, there is a trade-off between parameter estimation accuracy (exploration) and revenue maximization (exploitation). Too much focus on parameter estimation may be at the expense of some foregone revenue, and vice versa. This is valid in the short term ahead. However, in the long run, better parameter accuracy should positively impact revenue. Therefore, it is imperative to attempt to improve the accuracy, if possible without too much sacrifice in revenue.

- The myopic and myopic with dithering policies obtain poor estimates for model parameters, especially for high noisy datasets as represented in Table 2.

- Figure 2 shows the model estimation error for one noisy synthetic dataset, as an example. Beside the final estimates represented in Table 2, here we seek to investigate the performance of different methods over time. One can observe from Fig. 2 that over iterations, the model estimation is enhanced, and this is intuitive because more training points are added as iterations go on. It can be noticed that our proposed third formulation achieves comparable performance to the best performing benchmarks uncertain-myopic and random-myopic, and these results agree with the results of the Monte Carlo simulation presented in Table 2.

7.3 Price convergence

- Figure 3 shows that in the initial period the price changes rapidly, often going up and down. The algorithm is literally exploring the space in order to learn the price demand model. Later in the iterations the price stabilizes. It now enters the exploitation phase, whereby it narrows down on the price that maximizes revenue.
- Regarding the price convergence to the optimal price p_{opt}, Tables 3 and 7 indicate that the best performing methods are our defined random-myopic baseline for the synthetic datasets and the uncertain-myopic method for the real datasets, respectively. However, our three proposed formulations produce comparable results. The two-phase pricing policies: random-myopic and uncertain-myopic perform well with regard to price convergence because sufficient exploration during the first phase leads to fairly accurate parameter estimates (see Tables 2 and 6 for synthetic and real datasets results, respectively).

- For the other methods including: myopic and myopic with dithering methods, they have a considerable deviation error from the optimal price, especially for the high error setting according to Table 3. These methods do not converge to optimal prices because they do not obtain accurate model estimates. Accordingly, they do not reveal the true demand model, and cannot converge to the optimal pricing.

- Figure 3 shows the price convergence to the optimal price which is the obtained price if the true model parameters are known, for one noisy synthetic dataset. For all methods, along iterations, the convergence improves due to the corresponding enhancement in model estimation presented in Fig. 2. It can be inferred that all of our proposed methods produce promising results (near-optimal prices). Figure 3 indicates that the random-myopic benchmark performs the best in terms of price convergence, but at the expenses of sacrificing revenues as indicated in Table 1 and Fig. 1.

8 Conclusions and future work

In this work, we have proposed several dynamic pricing strategies for revenue maximization with demand learning. The proposed methods seek to balance the trade-off between exploitation (revenue maximization) and explo-

ration (demand model estimation). We compare our proposed methods to different benchmarks and popular methods in literature with respect to different aspects including: the total discounted gained revenue, the accuracy of the estimated demand model, and the price convergence to optimal price. We test the pricing methods using different twenty synthetic datasets with different parameter settings and error settings, and seven different real datasets including nineteen different products.. The experiments show a significant performance improvement of our pricing strategies, especially in terms of the gained revenue, while achieving comparable performance in demand learning. Moreover, our pricing policies are easy to analyze and implement since we use simple formulations. Furthermore, our proposed methods are computationally efficient as we apply regression model with incremental updates. For future research directions, we can extend our proposed methods to different demand models such as exponential, and logit demand functions. Furthermore, other factors could be taken into consideration in demand estimation to maximize the obtained revenue such as market environment and customers', and competitor's related features. Finally, we may thoroughly investigate the impact of the counterfeit products on demand, and we could develop pricing strategies with the aim of combating the counterfeiting adverse effects.

Author Contributions Dina Elreedy, Amir F. Atiya and Samir I. Shaheen were involved in the conceptualization; Dina Elreedy and Amir F. Atiya were involved in the formal analysis; Dina Elreedy and Amir F. Atiya were involved in the methodology; Amir F. Atiya and Samir I. Shaheen contributed to the project administration; Amir F. Atiya and Samir I. Shaheen contributed to resources; Dina Elreedy contributed to software; Amir F. Atiya and Samir I. Shaheen were involved in the supervision; Dina Elreedy and Amir F. Atiya were involved in the validation; Dina Elreedy, Amir F. Atiya and Samir I. Shaheen contributed to the writing.

Funding This research received no external funding.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflicts of interest The authors of this work declare no conflict of interest.

Code availability The source code of the current work is available from the corresponding author on reasonable request.

Informed consent Informed consent was obtained from all individual participants included in the study.

A Derivation of utility derivatives for the three proposed formulations

Formulation 1

According to Section 5, the expected utility of our first formulation is defined as:

$$E[U(p^*)_n] = bp^{*2} + ap^* - \eta \frac{1}{2\sqrt{tr[\Sigma_{\beta(n)}]}} tr\Big[\frac{1}{\gamma}\Sigma_\beta(n-1) - \frac{\Sigma_\beta(n-1)x_n x_n{}^T \Sigma_\beta(n-1)}{\sigma^2\gamma^2 + \gamma x_n{}^T \Sigma_\beta(n-1)x_n}\Big]$$

The first derivative of the expected utility, $\frac{\partial E[U(p^*)_n]}{\partial p^*}$, can be calculated as:

$$\frac{\partial E[U(p^*)_n]}{\partial p^*} = a + 2bp^* - \eta \frac{1}{2\sqrt{tr[\Sigma_{\beta(n)}]}} \frac{\partial tr[\frac{1}{\gamma}\Sigma_\beta(n-1) - \frac{\Sigma_\beta(n-1)x_n x_n{}^T \Sigma_\beta(n-1)}{\sigma^2\gamma^2+\gamma x_n{}^T \Sigma_\beta(n-1)x_n}]}{\partial p^*} \tag{38}$$

Since $tr[A+B] = tr[A] + tr[B]$, then $\frac{\partial tr[\Sigma_\beta(n)]}{\partial p^*}$ would be:

$$\frac{\partial tr[\Sigma_\beta(n)]}{\partial p^*} = \frac{\partial tr[\frac{1}{\gamma}\Sigma_\beta(n-1)]}{\partial p^*} - \frac{\partial tr[\frac{\Sigma_\beta(n-1)x_n x_n{}^T \Sigma_\beta(n-1)}{\sigma^2\gamma^2+\gamma x_n{}^T \Sigma_\beta(n-1)x_n}]}{\partial p^*} \tag{39}$$

It can be observed that the first derivative term in Eq. (39) evaluates to zero. Evaluating the second term of Eq. (39), and letting x_n be denoted as x^*:

$$\frac{\partial tr[\Sigma_\beta(n)]}{\partial p^*} = \frac{\partial tr[\frac{\Sigma_\beta(n-1)x^* x^{*T} \Sigma_\beta(n-1)}{\sigma^2\gamma^2+\gamma x^{*T} \Sigma_\beta(n-1)x^*}]}{\partial p^*} \tag{40}$$

Let $A = \frac{x^* x^{*T} \Sigma_\beta(n-1)}{\sigma^2\gamma^2+\gamma x^{*T} \Sigma_\beta(n-1)x^*}$, accordingly Eq. (40) can be evaluated as follows:

$$\frac{\partial tr[\Sigma_\beta(n)]}{\partial p^*} = \frac{\partial tr[\Sigma_\beta(n-1)A\Sigma_\beta(n-1)]}{\partial p^*} \tag{41}$$

However, from trace properties $tr[BAC] = tr[ACB]$, then:

$$\frac{\partial tr[\Sigma_\beta(n)]}{\partial p^*} = \frac{\partial tr[A\Sigma_\beta^2(n-1)]}{\partial p^*} \tag{42}$$

Then, from trace derivative properties:

$$\frac{\partial tr[AB]}{\partial x} = \frac{\partial < A^T, B >_F}{\partial x} = < B^T, \frac{\partial A}{\partial x} >_F + < A^T, \frac{\partial B}{\partial x} >_F = tr[B\frac{\partial A}{\partial x} + A\frac{\partial B}{\partial x}] \quad (43)$$

Accordingly, substitute from Eq.(43) into Eq. (42) where $B = \Sigma_\beta^2(n-1)$ and $x = p^*$, accordingly $\frac{\partial B}{\partial x}$ evaluates to zero, and Eq. (42) is simplified to:

$$\frac{\partial tr[\Sigma_\beta(n)]}{\partial p^*} = tr\left[\Sigma_\beta^2(n-1)\frac{\partial A}{\partial p^*}\right] \quad (44)$$

Simplifying matrix A:

$$A = \frac{\begin{pmatrix} 1 & p^* \\ p^* & p^{*2} \end{pmatrix}}{\sigma^2\gamma^2 + \gamma[\sigma_a^2 + 2\sigma_{ab}p^* + {\sigma_b}^2 p^{*2}]} \quad (45)$$

where $\Sigma_\beta(n-1) = \begin{pmatrix} {\sigma_a}^2 & \sigma_{ab} \\ \sigma_{ab} & {\sigma_b}^2 \end{pmatrix}$.

Then, evaluating $\frac{\partial A}{\partial p^*}$:

$$\frac{\partial A}{\partial p^*} = \frac{1}{(\sigma^2\gamma^2 + \gamma(\sigma_a^2 + 2\sigma_{ab}p^* + {\sigma_b}^2 p^{*2}))^2} \left[(\sigma^2\gamma^2 + \gamma(\sigma_a^2 + 2\sigma_{ab}p^* + {\sigma_b}^2 p^{*2}))\begin{pmatrix} 0 & 1 \\ 1 & 2p^* \end{pmatrix} - \gamma\begin{pmatrix} 1 & p^* \\ p^* & p^{*2} \end{pmatrix}(2\sigma_{ab} + 2\sigma_b^2 p^*)\right] \quad (46)$$

Let $g(p^*) = (\sigma^2\gamma^2 + \gamma[\sigma_a^2 + 2\sigma_{ab}p^* + {\sigma_b}^2 p^{*2}])$, then:

$$\frac{\partial A}{\partial p^*} = \frac{1}{g^2(p^*)}Z(p*) = \frac{1}{g^2(p^*)}\begin{pmatrix} Z_{11} & Z_{12} \\ Z_{12} & Z_{22} \end{pmatrix} \quad (47)$$

where $Z(p*)$ matrix elements: Z_{11}, Z_{12}, and Z_{22} are evaluated as follows:

$$\begin{aligned} Z_{11} &= -2\gamma(\sigma_{ab} + {\sigma_b}^2 p^*) \\ Z_{12} &= \gamma(\sigma^2\gamma + \sigma_a^2 - \sigma_b^2 p^{*2}) \\ Z_{22} &= \gamma(2\sigma^2\gamma p^* + 2\sigma_a^2 p^* + 2\sigma_{ab}p^{*2}) \end{aligned} \quad (48)$$

Substituting from Eq. (44) and Eq. (47) into Eq. (38):

$$\frac{\partial E[U(p^*)_n]}{\partial p^*} = a + 2bp^* + \eta\frac{1}{2\sqrt{tr[\Sigma_{\beta(n)}]}}tr\Big[\frac{1}{g^2(p^*)}\Sigma_\beta^2(n-1)Z(p*)\Big] \quad (49)$$

Formulation 2

As presented in Section 5, as defined in Section 5, can be evaluated as follows:

$$E[U(p^*)_n] = E[R(p^*)_n] - \eta\Big(\frac{\sqrt{\Sigma_\beta(n)_{11}}}{a} + \frac{\sqrt{\Sigma_\beta(n)_{22}}}{|b|}\Big)$$

Using Eq. (7) to substitute for $\Sigma_\beta(n)$, and let $A = \Sigma_\beta(n-1)x^*{x^*}^T\Sigma_\beta(n-1)$, and calculate the derivative of utility $\frac{\partial E[U(p^*)_n]}{\partial p^*}$ with respect to p^*, this results in the following equation:

$$\frac{\partial E[U(p^*)_n]}{\partial p^*} = a + 2bp^* + \eta\Big(\frac{1}{2a\sqrt{\Sigma_\beta(n)_{11}}}\frac{\partial}{\partial p^*}\Big[\frac{A_{11}}{g(p^*)}\Big] + \frac{1}{2|b|\sqrt{\Sigma_\beta(n)_{22}}}\frac{\partial}{\partial p^*}\Big[\frac{A_{22}}{g(p^*)}\Big]\Big) \quad (50)$$

where A_{11} is the first row and column entry in matrix A. A_{11} and A_{22} are evaluated as follows:

$$\begin{aligned} A_{11} &= {\sigma_{ab}}^2 p^{*2} + 2{\sigma_a}^2\sigma_{ab}p^* + {\sigma_a}^4 \\ A_{22} &= {\sigma_b}^4 p^{*2} + 2\sigma_{ab}{\sigma_b}^2 p^* + {\sigma_{ab}}^2 \end{aligned} \quad (51)$$

Substituting Eq. (51) into Eq. (50) and evaluating $\frac{\partial}{\partial p^*}[\frac{A_{11}}{g(p^*)}]$ and $\frac{\partial}{\partial p^*}[\frac{A_{22}}{g(p^*)}]$ terms results in:

$$\frac{\partial E[U(p^*)_n]}{\partial p^*} = a + 2bp^* + \eta\Big(\Big[\frac{2g(p^*)(\sigma_{ab}{\sigma_a}^2 + \sigma_{ab}^2 p^*) - 2\gamma A_{11}(\sigma_{ab} + \sigma_b^2 p^*)}{2ag^2(p^*)\sqrt{\Sigma_{\beta_{11}}(n)}}\Big] + \Big[\frac{2g(p^*)(\sigma_{ab}{\sigma_b}^2 + \sigma_b^4 p^*) - 2\gamma A_{22}(\sigma_{ab} + \sigma_b^2 p^*))}{2bg^2(p^*)\sqrt{\Sigma_{\beta_{22}}(n)}}\Big]\Big) \quad (52)$$

Simplifying Eq. (52) results in the following equation:

$$\frac{\partial E[U(p^*)_n]}{\partial p^*} = a + 2bp^* + \eta\frac{\gamma}{2ag^2(p^*)\sqrt{\Sigma_{\beta_{11}}(n)}} \times \Big(p^{*2}({\sigma_{ab}}^3 - \sigma_{ab}\sigma_a^2\sigma_b^2) + p^*(\sigma^2\gamma\sigma_{ab}^2 - \sigma_a^4\sigma_b^2) + \sigma^2\gamma\sigma_a^2\sigma_{ab}\Big) + \eta\frac{\gamma}{2|b|\sqrt{\Sigma_{\beta_{22}}(n)}g^2(p^*)}\Big(p^*(\sigma^2\gamma\sigma_b^4 - \sigma_{ab}^2\sigma_b^2 + \sigma_b^4\sigma_a^2) + (\sigma^2\gamma\sigma_{ab}\sigma_b^2 + \sigma_b^2\sigma_a^2\sigma_{ab} - \sigma_{ab}^3)\Big) \quad (53)$$

Formulation 3

The expected utility of the third proposed formulation, the expected utility of our second formulation is defined as:

$$E[U(p^*)_n] = E[R(p^*)_n] - \eta p^* \sqrt{x^{*T} \Sigma_{\beta(n-1)} x^* + \sigma^2}$$

The derivative of the expected utility $U(p^*)_n$ w.r.t. p^* is calculated as follows:

$$\frac{\partial E[U(p^*)_n]}{\partial p^*} = a + 2bp^* - \frac{\eta}{2\sqrt{p^{*2}(x^{*T} \Sigma_{\beta(n-1)} x^* + \sigma^2)}} \frac{\partial \left(p^{*2}(x^{*T} \Sigma_{\beta(n-1)} x^* + \sigma^2) \right)}{\partial p^*} \tag{54}$$

where $\Sigma_\beta(n-1) = \begin{pmatrix} {\sigma_a}^2 & \sigma_{ab} \\ \sigma_{ab} & {\sigma_b}^2 \end{pmatrix}$. Thus, the derivative of expected utility with respect to p^*, $\frac{\partial E[U(p^*)_n]}{\partial p^*}$ can be simplified into:

$$\frac{\partial E[U(p^*)_n]}{\partial p^*} = a + 2bp^* - \frac{\eta \left(2p^*(\sigma_a{}^2 + 2\sigma_{ab} p^* + p^{*2}\sigma_b{}^2 + \sigma^2) + 2p^{*2}(\sigma_{ab} + \sigma_b{}^2 p^*) \right)}{2\sqrt{p^{*2}(x^{*T} \Sigma_{\beta(n-1)} x^* + \sigma^2)}} \tag{55}$$

Simplifying Eq. (55) results in the following equation:

$$\frac{\partial E[U(p^*)_n]}{\partial p^*} = a + 2bp^* - \eta \frac{2\sigma_b{}^2 p^{*2} + 3\sigma_{ab} p^* + \sigma_a{}^2 + \sigma^2}{\sqrt{(\sigma^2 + \sigma_a^2 + 2\sigma_{ab} p^* + \sigma_b^2 p^{*2})}} \tag{56}$$

References

Araman VF, Caldentey R (2009) Dynamic pricing for nonperishable products with demand learning. Op Res 57(5):1169–1188

Araman VF, Caldentey R (2010) Revenue management with incomplete demand information. Wiley Encyclopedia of Operations Research and Management Science

Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA (2017) Deep reinforcement learning: a brief survey. IEEE Sig Proces Mag 34(6):26–38

Asiain E, Clempner JB, Poznyak AS (2019) Controller exploitation-exploration reinforcement learning architecture for computing near-optimal policies. Soft Comput 23(11):3591–3604

Atiya AF, Aly MA, Parlos AG (2005) Sparse basis selection: new results and application to adaptive prediction of video source traffic. IEEE Trans Neural Netw 16(5):1136–1146

Atiya AF, Abdel-Gawad AH, Fayed HA (2020) A new monte carlo based exact algorithm for the gaussian process classification problem. Adv Mathe Mod Appl 5(3):261–288

Audibert JY, Munos R, Szepesvári C (2009) Exploration-exploitation tradeoff using variance estimates in multi-armed bandits. Theor Comput Sci 410(19):1876–1902

Auer P (2002) Using confidence bounds for exploitation-exploration trade-offs. J Mach Learn Res 3:397–422

Auer P, Cesa-Bianchi N, Fischer P (2002) Finite-time analysis of the multiarmed bandit problem. Mach Learn 47(2–3):235–256

Aviv Y, Pazgal A (2002) Pricing of short life-cycle products through active learning. Olin School of Business, Washington University, St, Louis, Tech. rep

Aviv Y, Vulcano G (2012) Dynamic list pricing. In: The Oxford handbook of pricing management

Awad NH, Ali MZ, Duwairi RM (2017) Multi-objective differential evolution based on normalization and improved mutation strategy. Nat Comput 16(4):661–675

Aydin G, Ziya S (2009) Personalized dynamic pricing of limited inventories. Op Res 57(6):1523–1531

Ban GY, Keskin NB (2020) Personalized dynamic pricing with machine learning: High dimensional features and heterogeneous elasticity. Forthcoming, Management Science

Bayoumi AEM, Saleh M, Atiya AF, Aziz HA (2013) Dynamic pricing for hotel revenue management using price multipliers. J Rev Pric Manag 12(3):271–285

Bertsimas D, Perakis G (2006) Dynamic pricing: A learning approach. In: Mathematical and computational models for congestion charging, Springer, pp 45–79

Besbes O, Zeevi A (2015) On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. Manag Sci 61(4):723–739

Besbes O, Gur Y, Zeevi A (2014) Optimal exploration-exploitation in a multi-armed-bandit problem with non-stationary rewards. arXiv preprint arXiv:14053316

Bisht DC, Srivastava PK (2019) Fuzzy optimization and decision making. In: Advanced fuzzy logic approaches in engineering science, IGI Global, pp 310–326

den Boer AV (2015) Dynamic pricing and learning: historical origins, current research, and new directions. Surv Op Res Manage Sci 20(1):1–18

den Boer AV, Zwart B (2013) Simultaneously learning and optimizing using controlled variance pricing. Manag Sci 60(3):770–783

Byrd RH, Hribar ME, Nocedal J (1999) An interior point algorithm for large-scale nonlinear programming. SIAM J Optim 9(4):877–900

Cao P, Zhao N, Wu J (2019) Dynamic pricing with bayesian demand learning and reference price effect. Eur J Op Res 279(2):540–556

Carvalho AX, Puterman ML (2005) Learning and pricing in an internet environment with binomial demands. J Rev Pric Manag 3(4):320–336

Caviglione L, Gaggero M, Paolucci M, Ronco R (2020) Deep reinforcement learning for multi-objective placement of virtual machines in cloud datacenters. Soft Comput pp 1–20

Chen B, Chao X (2019) Parametric demand learning with limited price explorations in a backlog stochastic inventory system. IISE Trans 51(6):605–613

Chen HM, Hu CF, Yeh WC (2019) Option pricing and the greeks under gaussian fuzzy environments. Soft Comput 23(24):13351–13374

Cheng Y (2008) Dynamic pricing decision for perishable goods: a q-learning approach. In: Wireless communications, networking and mobile computing. WiCOM'08. 4th International Conference on, IEEE, pp 1–5

Cheung WC, Simchi-Levi D, Wang H (2017) Dynamic pricing and demand learning with limited price experimentation. Ope Res 65(6):1722–1731

Črepinšek M, Liu SH, Mernik M (2013) Exploration and exploitation in evolutionary algorithms: a survey. ACM Comput Surv (CSUR) 45(3):35

Crombecq K, Gorissen D, Deschrijver D, Dhaene T (2011) A novel hybrid sequential design strategy for global surrogate modeling of computer experiments. SIAM J Sci Comput 33(4):1948–1974

Curiel IT, Di Giannatale SB, Herrera JA, Rodríguez K (2012) Pareto frontier of a dynamic principal-agent model with discrete actions: an evolutionary multi-objective approach. Comput Econ 40(4):415–443

den Boer A (2012) Dynamic pricing and learning. PhD thesis, Vrije Universiteit Amsterdam, naam instelling promotie: VU Vrije Universiteit Naam instelling onderzoek: VU Vrije Universiteit

Diao J, Zhu K, Gao Y (2011) Agent-based simulation of durables dynamic pricing. Syst Eng Proc 2:205–212

Durbin J, Watson GS (1950) Testing for serial correlation in least squares regression: I. Biometrika 37(3/4):409–428

Elreedy D, Atiya AF (2019) A comprehensive analysis of synthetic minority oversampling technique (smote) for handling class imbalance. Inform Sci 505:32–64

Elreedy D, Atiya AF, Fayed H, Saleh M (2017) A framework for an agent-based dynamic pricing for broadband wireless price rate plans. J Simul, pp 1–15

Elreedy D, Atiya F, A, I Shaheen S, (2019) A novel active learning regression framework for balancing the exploration-exploitation trade-off. Entropy 21(7):651

Elreedy D, Atiya AF, Shaheen SI (2021) Multi-step look-ahead optimization methods for dynamic pricing with demand learning. IEEE Access

Farahani MS, Hajiagha SHR (2021) Forecasting stock price using integrated artificial neural network and metaheuristic algorithms compared to time series models. Soft Comput, pp 1–31

Farias VF, Van Roy B (2010) Dynamic pricing with a prior on market response. Op Res 58(1):16–29

Fazakis N, Kanas VG, Aridas CK, Karlos S, Kotsiantis S (2019) Combination of active learning and semi-supervised learning under a self-training scheme. Entropy 21(10):988

Gao R, Wu W, Liu J (2021) Asian rainbow option pricing formulas of uncertain stock model. Soft Comput, pp 1–25

Geman S, Bienenstock E, Doursat R (1992) Neural networks and the bias/variance dilemma. Neural Comput 4(1):1–58

Gillespie A (2014) Foundations of economics. Oxford University Press

Gwartney JD, Stroup RL, Sobel RS, Macpherson DA (2014) Economics: Private and public choice. Nelson Education

Han W, Liu L, Zheng H (2008) (2008) Dynamic pricing by multiagent reinforcement learning. Electronic Commerce and Security. International Symposium on, IEEE, pp 226–229

Harrison JM, Keskin NB, Zeevi A (2012) Bayesian dynamic pricing policies: learning and earning under a binary prior distribution. Manag Sci 58(3):570–586

Ibrahim MN, Atiya AF (2016) Analytical solutions to the dynamic pricing problem for time-normalized revenue. Eur J Op Res 254(2):632–643

Ishii S, Yoshida W, Yoshimoto J (2002) Control of exploitation-exploration meta-parameter in reinforcement learning. Neural Netw 15(4–6):665–687

Jerebic J, Mernik M, Liu SH, Ravber M, Baketarić M, Mernik L, Črepinšek M (2021) A novel direct measure of exploration and exploitation based on attraction basins. Exp Syst Appl 167:114353

Ji X, Zhou J (2015) Option pricing for an uncertain stock model with jumps. Soft Comput 19(11):3323–3329

Kastius A, Schlosser R (2021) Dynamic pricing under competition using reinforcement learning. J Rev Pric Manag, pp 1–14

Keskin NB, Zeevi A (2014) Dynamic pricing with an unknown demand model: asymptotically optimal semi-myopic policies. Op Res 62(5):1142–1167

Kutschinski E, Uthmann T, Polani D (2003) Learning competitive pricing strategies by multi-agent reinforcement learning. J Econ Dyn Control 27(11–12):2207–2218

Li W, Wang X, Zhang R, Cui Y, Mao J, Jin R (2010) Exploitation and exploration in a performance based contextual advertising system. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 27–36

Li Y, Wang B, Fu A, Watada J (2020) Fuzzy portfolio optimization for time-inconsistent investors: a multi-objective dynamic approach. Soft Comput 24(13):9927–9941

Library CUM (2001) Musdaers electronic data archive, red meats yearbook. "http://usda.mannlib.cornell.edu/"

Liu J, Pang Z, Qi L (2020) Dynamic pricing and inventory management with demand learning: a bayesian approach. Comput Op Res 124:105078

Lobo MS, Boyd S (2003) Pricing and learning with uncertain demand. In: INFORMS revenue management conference

Mahesh A, Sushnigdha G (2021) A novel search space reduction optimization algorithm. Soft Comput pp 1–28

Makridakis S, Spiliotis E, Assimakopoulos V (2020) The m5 accuracy competition: results, findings and conclusions. Int J Forecast

Martinez-Cantin R, de Freitas N, Brochu E, Castellanos J, Doucet A (2009) A bayesian exploration-exploitation approach for optimal online sensing and planning with a visually guided mobile robot. Autonom Rob 27(2):93–103

McAfee RP, Te Velde V (2006) Dynamic pricing in the airline industry. Forthcoming in handbook on economics and information systems, Ed: TJ Hendershott, Elsevier

Morales-Enciso S, Branke J (2012) Revenue maximization through dynamic pricing under unknown market behaviour. In: OASIcs-OpenAccess Series in Informatics, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, vol 22

of Nicosia TU (2020) M5 forecasting - accuracy. https://www.kaggle.com/c/m5-forecasting-accuracy

Pandey S, Agarwal D, Chakrabarti D, Josifovski V (2007) Bandits for taxonomies: A model-based approach. In: Proceedings of the 2007 SIAM international conference on data mining, SIAM, pp 216–227

Price I, Fowkes J, Hopman D (2019) Gaussian processes for unconstraining demand. Eur J Op Res 275(2):621–634

Rana R, Oliveira FS (2015) Dynamic pricing policies for interdependent perishable products or services using reinforcement learning. Exp Syst Appl 42(1):426–436

Rezaei F, Safavi HR (2020) Guaspso: a new approach to hold a better exploration-exploitation balance in pso algorithm. Soft Comput 24(7):4855–4875

Rhuggenaath J, da Costa PRdO, Akcay A, Zhang Y, Kaymak U (2019) A heuristic policy for dynamic pricing and demand learning with limited price changes and censored demand. 2019 IEEE international conference on systems. Man and Cybernetics (SMC), IEEE, pp 3693–3698

Rhuggenaath J, da Costa PRdO, Zhang Y, Akcay A, Kaymak U (2020) Dynamic pricing using thompson sampling with fuzzy events. In: International conference on information processing and management of uncertainty in knowledge-based systems, Springer, pp 653–666

Robbins H (1985) Some aspects of the sequential design of experiments. In: Herbert Robbins Selected Papers, Springer, pp 169–177

Rothschild M (1974) A two-armed bandit theory of market pricing. J Econ Theory 9(2):185–202

Schaffer JD (1985) Multiple objective optimization with vector evaluated genetic algorithms. In: Proceedings of the first international conference on genetic algorithms and their applications (1985) Lawrence Erlbaum Associates. Publishers, Inc

Schultz H (1933) A comparison of elasticities of demand obtained by different methods. Econometrica J Econ Soc pp 274–308

Settles B (2009) Active learning literature survey. University of Wisconsin-Madison Department of Computer Sciences, Tech. rep
Shrestha A, Mahmood A (2019) Review of deep learning algorithms and architectures. IEEE Access 7:53040–53065
Singh A, Deep K (2019) Exploration-exploitation balance in artificial bee colony algorithm: a critical analysis. Soft Comput 23(19):9525–9536
Srinivasan S, Kamalakannan T (2018) Multi criteria decision making in financial risk management with a multi-objective genetic algorithm. Comput Econ 52(2):443–457
Sun Y (2011) Coke demand estimation dataset. http://leeds-faculty.colorado.edu/ysun/doc/Demand_estimation_worksheet.doc
Sun Y, Yao K, Dong J (2018) Asian option pricing problems of uncertain mean-reverting stock model. Soft Comput 22(17):5583–5592
Taieb SB, Atiya AF (2015) A bias and variance analysis for multistep-ahead time series forecasting. IEEE Trans Neural Netw Learn Syst 27(1):62–76
Tang R, Wang S, Li H (2019) Game theory based interactive demand side management responding to dynamic pricing in price-based demand response of smart grids. Appl Energy 250:118–130
Thompson WR (1933) On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika 25(3/4):285–294
Thompson WR (1935) On the theory of apportionment. Am J Math 57(2):450–456
Tokic M (2010) Adaptive ε-greedy exploration in reinforcement learning based on value differences. In: Annual conference on artificial intelligence, Springer, pp 203–210
Triki C, Violi A (2009) Dynamic pricing of electricity in retail markets. 4OR 7(1):21–36
Trovo F, Paladino S, Restelli M, Gatti N (2015) Multi–armed bandit for pricing. In: Proceedings of the european workshop on reinforcement learning (EWRL)
Valizadegan H, Jin R, Wang S (2011) Learning to trade off between exploration and exploitation in multiclass bandit prediction. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 204–212
Vermorel J, Mohri M (2005) Multi-armed bandit algorithms and empirical evaluation. In: European conference on machine learning, Springer, pp 437–448
Villar SS, Bowden J, Wason J (2015) Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. Stat Sci A Rev J Inst Math Stat 30(2):199
Wang Z, Deng S, Ye Y (2014) Close the gaps: a learning-while-doing algorithm for single-product revenue management problems. Op Res 62(2):318–331
Xia CH, Dube P (2007) Dynamic pricing in e-services under demand uncertainty. Prod Op Manag 16(6):701–712
Zavarella L (2018) Price elasticity dataset. https://towardsdatascience.com/price-elasticity-data-understanding-and-data-exploration-first-of-all-ae4661da2ecb
Zhong S, Wang X, Zhao J, Li W, Li H, Wang Y, Deng S, Zhu J (2021) Deep reinforcement learning framework for dynamic pricing demand response of regenerative electric heating. Appl Energy 288:116623
Zhu Z, Peng J, Liu K, Zhang X (2020) A game-based resource pricing and allocation mechanism for profit maximization in cloud computing. Soft Comput 24(6):4191–4203