Pricing Competition in a Duopoly with Self-adapting Strategies

Youri Kaminsky, Tobias Maltenberger, Mats Pörschke, Jan Westphal and Rainer Schlosser
Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

Keywords: Dynamic Pricing, Duopoly Competition, Price Anticipation, Self-adaptive Strategies, Cartel Formation.

Abstract: Online markets are characterized by highly dynamic price adjustments and steady competition. Many market participants adjust their prices in real-time to changing market situations caused by competitors' pricing strategies. In this paper, we examine price optimization within a duopoly under an infinite time horizon with mutually unknown strategies. The challenge is to derive knowledge about the opponent's pricing strategy automatically and to respond effectively. Strategy exploration procedures used to build a data foundation on a competitor's strategy are crucial in unknown environments and therefore need to be selected and configured with caution. We show how our models explore and exploit a competitor's price reaction probabilities. Moreover, we verify the quality of our learning approach against optimal strategies exploiting full information. In addition to that, we analyze the mutual interplay of two self-learning strategies. We observe a clear winning party over time as well as that they can also form a cartel when motivated accordingly.

1 INTRODUCTION

Online markets are becoming increasingly dynamic and competitive. Market participants can observe their competitors' prices and adjust their prices with high frequencies. Hence, to maximize their profits, merchants need to automatically adjust prices to respond to steadily changing market situations.

Given that online markets allow market participants to observe their competitors' prices in real-time, dynamic pricing strategies, which take into account the competitors' strategies by learning historical price reactions and gradually adjusting the own strategy accordingly, are getting implemented more frequently.

However, efficiently determining optimal price reactions to maximize long-term profits in competitive markets is anything but trivial, especially for wide price ranges and large numbers of goods. In online markets, both perishable goods (e.g., food products (Tong et al., 2020), event tickets (Sweeting, 2012), and seasonal pieces of clothing (Huang et al., 2014)) and durable goods (e.g., electronic devices (He and Chen, 2018) and licenses for software (Hajji et al., 2012)) are subject to automated price adjustment strategies. Oftentimes, these strategies follow a periodically recurring pattern over time (e.g., Edgeworth cycles) (Noel, 2007) (Noel, 2012). In the case of a duopoly, where two market participants are competing against each other, Edgeworth cycles entail that both market participants undercut each other until one market participant's lower bound is reached (e.g., the profit yields zero), and the market participant raises the price to secure future profits.

In this paper, we present a model for optimizing pricing strategies under duopoly competition in which sales probabilities are allowed to be an arbitrary function of competitor prices. We consider durable goods under an infinite time horizon. Our goal is to derive price response strategies that optimize the expected long-term future profits under uncertain environments by learning from the observed actions of the competitor and adapting to them effectively.

Our contributions are as follows. We derive mechanisms to find effective self-tuning responses against (i) fixed but unknown competitor strategies including deterministic as well as randomized (mixed) strategies. Based on these mechanisms, we analyze the interaction of (ii) two self-adapting strategies over time. Furthermore, we study (iii) how self-tuning strategies can be adapted to naturally form a cartel in which market participants settle on a fixed price and thereafter stop competing with price adjustments.

The remainder of this paper is structured as follows. In Section 2, we delve into related work regarding dynamic pricing models in general, and duopoly models in particular. Thereafter, Section 3 describes our infinite time horizon duopoly consisting of two competing market participants. In Section 4, we outline the theoretical framework on which our approach to determining optimized pricing strategies

Kaminsky, Y., Maltenberger, T., Pörschke, M., Westphal, J. and Schlosser, R.
Pricing Competition in a Duopoly with Self-adapting Strategies.
DOI: 10.5220/0010232900600071
In *Proceedings of the 10th International Conference on Operations Research and Enterprise Systems (ICORES 2021)*, pages 60-71
ISBN: 978-989-758-485-5

rests upon. Thereafter, in Section 4, we propose our concepts to tackle scenarios in which the competitor's strategies are unknown. For the case of an unknown competitor's strategy, it also provides an in-depth description of our self-adapting strategy for optimized price reactions. In Section 5, we evaluate our proposed pricing strategies for selected market setups. Section 6 summarizes our contributions.

2 RELATED WORK

Given that efficiently determining optimized prices for the sale of goods is one of the key challenges of revenue management, both comprehensive books (Phillips, 2005) (Talluri and Van Ryzin, 2006) (Gallego and Topaloglu, 2019) and conceptual papers (McGill and van Ryzin, 1999) (Bitran and Caldentey, 2003) cover the broad field of dynamic pricing. In addition, (den Boer, 2015) and (Chen and Chen, 2015) provide an extensive overview over dynamic pricing developments in recent years.

Most existing models consider so-called myopic customers who arrive and decide. Instead, (Levin et al., 2009), (Liu and Zhang, 2013), and (Schlosser, 2019b) analyze dynamic pricing models with customers who strategically time their purchase by anticipating future prices in advance.

(Adida and Perakis, 2010), (Tsai and Hung, 2009), and (Do Chung et al., 2011) study dynamic pricing models under competition with limited demand information by employing robust optimization techniques and learning approaches. Especially in the area of demand learning, however, the vast majority of techniques is not flexible enough to be widely adopted in practice. In the area of data-driven approaches to dynamic pricing, (Schlosser and Boissier, 2018) analyzes stochastic dynamic pricing models in competitive markets with multiple offer dimensions, such as price, quality, and rating.

(Gallego and Wang, 2014) considers a continuous time multi-product oligopoly for differentiated perishable goods by harnessing optimality conditions to solve the multi-dimensional dynamic pricing problem. In a more general oligopoly model for the sale of perishable goods, (Gallego and Hu, 2014) analyzes structural characteristics of equilibrium strategies.

(Martínez-de Albéniz and Talluri, 2011) studies duopoly pricing models for identical products. Since the sale of perishable goods is typically subject to incomplete market information, (Schlosser and Richly, 2018) looks at dynamic pricing strategies in a finite horizon duopoly with partial information.

(Schlosser and Boissier, 2017) analyze optimal repricing strategies in a stochastic infinite time horizon duopoly. (Schlosser, 2019a) extends this work by including endogenous reference price effects and price anticipations. The authors consider both known and unknown competitor strategies. However, they use an entirely different demand setup and price exploration mechanism to anticipate competitor prices. Moreover, they do not study cartel formation.

3 MODEL DESCRIPTION

We consider a scenario, where two competing market participants A and B want to sell goods on online marketplaces. Those marketplaces allow frequent price adjustments based on data that were collected on competitors' pricing strategies. Nowadays, computing power enables those competing market participants to perform market analyses for thousands of product frequently to support almost real-time price anticipation strategies. For this work, we have several assumptions that abstract away from a real price competition but allow space for exploration.

The product supply of each market participant is considered to be unlimited. Hence, we assume that both market participants have the ability to reorder an arbitrary amount of a product at any time.

All of our models focus on discrete prices only (i.e., both competing market participants are only allowed to price their product at one of the predefined prices $prices = \{p_1, \ldots, p_n\}$). None of our models makes distributional assumptions on the set of potential prices. Thus $p_1, \ldots, p_n$ may follow any potential distribution including non-uniform ones. While the assumption of a discrete price model allows for simplification of the model computation, real-world scenarios can still be mapped to our setting. In most markets, the products' smallest difference in price is a cent. Thus, most competitions can be easily simulated by our models. Moreover, the coherent costs c, $c \geq 0$ (e.g., for delivery) are predefined, as they do not change over time. In most cases, these coherent costs do not play a huge factor for computing optimal strategies, so we chose to set $c = 0$ for our experiments. As a consequence, in our experiments, a sale's net profit equals the offer price.

As most of the products on big online marketplaces are present for a long period of time and do not need to be sold until a specific date, we consider the time horizon under which we perform our pricing analyses to be infinite. However, we can use a discounting factor δ, $0 < \delta < 1$ to express the desire to gain profits early. Moreover, we decided to use a discrete time model with some adjusting screws, that

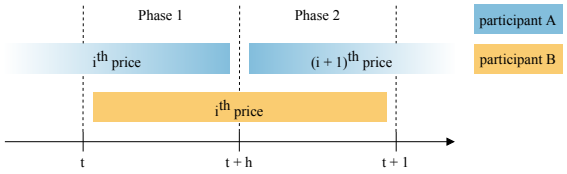


Figure 1: Discrete time model used across all models.

allow for very different scenarios.

We only allow price adjustments at specific predefined times, as most marketplaces do not allow for continuous price adaptions. While market participant A might react to the current market situation at t and $t+1$, market participant B might react at $t+h$ and $t+1+h, h \in (0,1)$. A visualization of this time model can be found in Figure 1. The hyper-parameter h allows for simulation of various different scenarios. While $h = 0.5$ results in a fair duopoly competition, $h \neq 0.5$ results in a biased scenario. Here, one can think of one competitor being able to have access to more computing power and thus, on average, reacting faster to price adjustments by the other party.

Furthermore, we divide the presented opponent strategies into deterministic and stochastic strategies. Deterministic strategies are characterized by only allowing for a single price reaction to a given price. Meanwhile, stochastic strategies have a larger pool of reactions for a single price from which they can choose one. In our case, stochastic strategies select the reaction randomly, although not necessarily uniformly distributed. This allows for interesting observations, as the optimized strategy against an unpredictable opponent can be counter intuitive. In general, a market participant's strategy can be characterized by a probability distribution of how to respond to a certain competitor price. In this context, the probability that B reacts to A's price $p_A \in prices$ (under a delay h) with the price $p_B \in prices$ is denoted by

$$P_{react}\left(p_A, p_B\right) : (prices, prices) \rightarrow [0,1]. \quad (1)$$

Further, as we do not represent different product conditions (e.g., used or new) or seller ratings in our models, customers can only base their buying decision on the two competitors' prices at time t. As demand learning is not in focus, we assume that the customer's behavior is known or has already been estimated. In our models, one customer arrives at each time interval $[t, t+1]$ and chooses to buy a product based on the current price level. After deciding to buy, the customer purchases from the competitor with the lower price or randomly chooses a competitor if the offer prices are equal. The probability that a customer buys a product of market participant A is described as

$$P_{buy_A}\left(p_A, p_B\right) : (prices, prices) \rightarrow [0,1]. \quad (2)$$

where p_A denotes the offer price of participant A and p_B denotes the offer price of participant B, respectively. Note that the sales probability of participant A can be summarized as a function which depends on (i) the current competitor price p_B and (ii) the price p_A chosen by participant A for one period. However, it may also include (iii) the competitor's price reaction p'_B and (iv) the reaction delay h of participant B. Hence, based on (1), (2) can be expressed via conditional probabilities $P_{buy_A}\left(p_A, p_B \mid h, p'_B\right)$.

Resulting from that, the expected total future profit G of market participant A given both player's strategies can be computed by evaluating

$$E\left(G\right) = \sum_{t=0}^{\infty} \delta^t \cdot P_{buy_A}\left(p_{A_t}, p_{B_t}\right) \cdot p_{A_t}.$$

The objective is to maximize this expected profit.

4 SOLUTION PROPOSITION

In Section 4.1, we describe our basic optimization model to solve the problem defined in Section 3 for known inputs. Based on this model, in Section 4.2, we address the case when the competitor's strategy is unknown. In Section 4.3, we study the case when both participants use adaptive learning strategies. Finally, in Section 4.4, we analyze how cartels form.

4.1 Basic Optimization Model

Taking participant A's perspective based on assumed buying probabilities (2) for one period (with reaction time h) as well as assumed price reaction probabilities (1), the value function $V(p_B)$ of the duopoly problem can be solved using dynamic programming methods (e.g., value iteration) with T steps using initial values for $V_T\left(p_B\right)$ via $t = 0, 1, \ldots, T-1$, $p_B \in prices$,

$$\begin{aligned} V_t\left(p_B\right) = \max_{p_A \in prices} \Big\{ & \sum_{p'_B \in prices} P_{react}\left(p_A, p'_B\right) \\ & \cdot \left(P_{buy_A}\left(p_A, p_B \mid h, p'_B\right) \cdot p_A + \delta \cdot V_{t+1}\left(p'_B\right)\right) \Big\}. \end{aligned} \quad (3)$$

The associated price reaction policy (i.e., how to respond to participant B's price p_B) is determined by the arg max of (3) derived at the last step of the recursion in $t = 0$. Note that the number of recursion steps T and the starting values $V_T\left(p_B\right)$ can be chosen such that the approximation satisfies a given accuracy (based on the discount factor and the maximum attainable reward). In addition to that, when solving (3) repeatedly with slight changes (e.g., with updated price reaction probabilities), suitable starting values of previous solutions can be used.

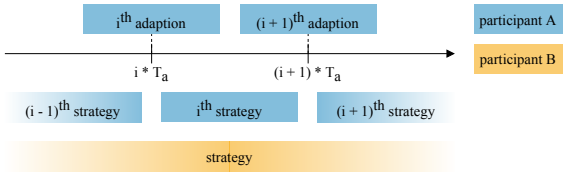


Figure 2: Participant A competing against an unknown pricing strategy by adjusting their own strategy over time.

4.2 Dealing with Unknown Strategies

We assume that market participant A does not know market participant B's strategy and that participant B's strategy is fixed and does not change over time.

The objective of A is to approximate B's strategy based on the reactions to proposed prices while still being competitive on the market. Therefore, A continuously adjusts their strategy after a fixed number of time steps T_a to have a competitive strategy while gathering more data about the reactions of the opponent B. This process is visualized in Figure 2. The reactions are recorded in a two dimensional data structure tr as follows. $tr[p_A, p_B]$ is the total number of times B reacted with price p_B to A's price p_A. Furthermore, we define $\#tr(p_A)$ as the total number of times A has seen a reaction to price p_A as

$$\#tr(p_A) = \sum_{p_B \in prices} tr[p_A, p_B].$$

After T_a time steps, participant A computes the best anticipation strategy via (3) using the estimated probability distribution over B's recorded price reactions so far with $P_{react}(p_A, p_B)$ from (1). If there are no reactions recorded for p_A, we assume a uniform distribution over all available prices. Therefore,

$$p_A, p_B \mapsto \begin{cases} \dfrac{1}{|prices|}, & \text{if } \#tr(p_A) = 0 \\ \dfrac{tr[p_A, p_B]}{\#tr(p_A)}, & \text{otherwise.} \end{cases}$$

Participant A acts according to the computed strategy for the next T_a time steps, until they do the next strategy computation also taking the newly collected data into consideration. The size of T_a should be as small as possible to update the participant's strategy often and is only limited by the available computation time and the available computational resources.

Over time, participant A gets to know B's strategy, as the observed distribution over the price reactions will become closer to the expected distribution. In the optimal case, this model will deliver the same price anticipation strategy as the competitor's price response probabilities would be exactly known. However, if A receives an unprofitable reaction for a specific price, it is likely that the model will not propose this price in the future again. A's strategy might get stuck and will not change in the future. In order to counteract this behavior, we need to motivate the model to explore. We call exploring the act of proposing prices that have not seen enough reactions, even though these prices would not be part of the optimal anticipation strategy that could be build based on the recorded price reactions. The participant is able to gather new reactions and extend their data foundation significantly. Below, we propose two procedures to explore participant B's pricing strategy. Note that both mechanisms differ from the one used in (Schlosser, 2019a), where artificially added observations of high price reactions of the competitor are used to organize the price exploration in an incentive-driven framework based on an optimistic initiation.

Assurance. For a specific number of time steps T_i the participant randomly proposes prices that have not received enough reactions. By doing so, the participant gains more confidence in the next strategy evaluation. We will only apply this procedure in the first T_i time steps to build a profound first strategy, but it is also reasonable to apply this procedure at a later point in time (e.g., when the strategy has not changed for a long time). In the former case, there is no recorded data and if $T_i \leq |prices|$ A proposes a different price each time step. If $T_i > |prices|$ then A will start over and proposes every price at least once before proposing it a second time. Which price is proposed exactly will be decided randomly to account for $|prices| \mod T_i \not\equiv 0$. During exploration, the model only cares about gaining new information about participant B's strategy and neither takes competitiveness nor profits into account. Afterwards, participant A continuously adjusts its procedure as described before.

Incentive. The price anticipation (1) is modified in order to motivate the model to include prices in its strategy that have not seen enough reactions by participant B yet. The way to motivate depends on the customer's buying behavior. We search for the combination of prices p_A*, p_B* that gives participant A the highest possible profit in the next iteration. Therefore, we utilize the part of the value function (3) for calculating the immediate profit as follows:

$$p_A*, p_B* = \underset{p_A, p_B \in prices}{\arg\max} P_{buy_A}(p_A, p_B) \cdot p_A.$$

It is desirable for the algorithm to propose a price that receives the reaction p_B* because participant A can react with p_A* and will then gain the highest possible profit. The algorithm needs to decide whether

high immediate profit is worth risking long-term profits. This way, participant A slowly adjusts its strategy by taking new prices into consideration until enough reactions for every available price were received. $P_{react}(p_A, p_B)$ for $\lambda \in \mathbb{R}^+$ is defined as

$$P_{react}(p_A, p_B) = \begin{cases} \frac{tr[p_A, p_B] + \lambda}{\#tr(p_A) + \lambda}, & \text{if } p_B = p_B* \\ \frac{tr[p_A, p_B]}{\#tr(p_A) + \lambda}, & \text{otherwise.} \end{cases}$$

With λ, it is to some extent possible to control how many reactions for a price participant A would like to receive before deeming this price as unprofitable. If $\lambda \approx 0$ and if A receives an unprofitable reaction for a specific price, this price will not be proposed again as we do not assume a desirable price reaction in the future. However, if we choose λ to be larger (e.g., $\lambda \geq 1$), there need to be multiple undesirable price reactions before they outweigh the possible chance of a high profit in the next iteration.

Both procedures have their own advantages and disadvantages. The advantage of the *Assurance* exploration procedure is that participant A gains a sparse but broad data foundation very quickly. After initial exploration, participant A is able to build their first competitive strategy. Furthermore, if participant B uses a deterministic strategy and $T_i \geq |prices|$, the evaluated strategy after exploration will not change in later strategy adaptions as A has already seen every possible reaction from B. In this case, A fully reveals the strategy after T_i time steps. A major downside of this procedure is that participant A does not care about profits for T_i time steps. As we consider an infinite event horizon, it is negligible if the competitor does not work efficiently for a finite number of time steps. If we instead consider a real world market situation, the competitor might not be able to survive the exploration phase. Therefore T_i needs be chosen wisely and in proportion to $|prices|$. It is not feasible to try out most of the available prices if $|prices|$ is large.

In this case, it might be better to use the *Incentive* approach. The participant considers profits and losses starting from the first proposed price and progressively explores prices that have not seen a reaction because exploring is part of strategy evaluation. On the other hand, it might take the *Incentive* procedure several strategy adaptions before every price has been proposed at least once and even more rounds of strategy adaptions before the incentive weight λ has been smoothed out completely.

It is worth noting that both procedures do not take interpolation into account. For example, in real-world scenarios where prices $p_A - 1$ and $p_A + 1$ are unprofitable, it is very likely that price p_A is also unprofitable. However, as both procedures have not seen a reaction for p_A, they are influenced to propose this price. Therefore, both procedures have the problem that they might propose prices unnecessarily. Additionally, both exploration procedures have one hyperparameter that each needs to be tuned. We will discuss choosing T_i and λ further in Section 5.2.

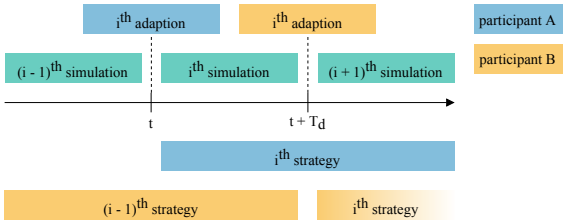


Figure 3: Two competing self-learning strategies over time.

4.3 Competing Self-adaptive Strategies

The last model that we present is an extension of the one presented in Section 4.2. Instead of competing with an unknown but fixed strategy, both parties can adapt their pricing strategies over time to react to the current market pricing situation and the other participant's pricing strategy. Similar to the model presented in Section 4.2, both participants need a data foundation to base their strategy decision on. We, therefore, collect the respective opponent's price reactions over time in the data structure tr. After a predefined number T_d of price reactions, one market participant is allowed to analyze their collected price reactions in order to adapt their own pricing strategy. Another T_d price reactions later, the other market participant reacts to the changed market situation.

A visualization of the procedure can be found in Figure 3. The collection of the mentioned T_d price reactions is grouped together in the referenced data collection lasting for T_d time steps. A data collection block represents the real market competition. All the tracked price reactions are passed into the strategy adaption of the respective participant at time step t. The computation of the newly adapted strategy is the same as the one from Section 4.2. The participant's $(i-1)^{th}$ strategy is replaced with the i^{th} strategy, which will be used for the next two data collection blocks while participant B still uses their $(i-1)^{th}$ strategy. The next data collection block starting at t runs another T_d time steps. At $t + T_d$, participant B updates their strategy which will be used within the subsequent two data collection blocks.

The model presented in this section mainly differs from the one presented in Section 4.2 by the two strategies changing the over time. Reaction data col-

lected at $t = 0$ will probably be outdated at some later point t_i which might result in inaccurate price reaction strategies. The model needs to anticipate that. In order to do so, we introduce a vanishing of values within our tr data structure. After a pricing strategy adaption was performed, the values are multiplied with a constant factor $\alpha \in [0,1]$. This allows to decide between different intensities of keeping all of the collected, but possibly outdated data. For example, with $\alpha = 1$ all recorded reactions will be kept and with $\alpha = 0$ the data structure will be reset. While $\alpha = 0$ leads to better anticipation strategies when just respecting the last simulation run, $\alpha > 0$ is expected to account for the long-term trend and to be less prone to over-fitting the own strategy on a single data collection run.

4.4 Incentivizing Cartel Formations

Additionally, we present a way to allow both market participants to form a cartel in which they constantly price their products equally. Note, instead of predefining a cartel price in advance, we study the case whether it is possible to modify our self-tuning price anticipation/optimization framework such that two of our independently applied learning strategies form a cartel *without* direct communication.

In order to determine the best cartel price, we reuse a modified version of the presented formula to find the optimal incentive price as follows:

$$p* = \underset{p \in prices}{\arg\max}\ P_{buy_A}(p,p) \cdot p.$$

Further, in our models, the adaption of the response policy derived by (3) is organized as follows. We manually overwrite the reaction of market participant A to $p*$ with $p*$. Thus, market participant A signals to market participant B its willingness to support a cartel price. The rest of our approach to define price reactions and to decide on prices, as described in Section 4.2 and Section 4.3, remains unchanged.

5 EXPERIMENTAL EVALUATION

In this section, we study the performance of our different approaches from Sections 4.2, 4.3, and 4.4. To do so, we consider numerical examples, where the buying behavior and the competitor's strategies to be learned are defined in Section 5.1.

5.1 Setup

In Section 5.1.1, we define example approaches for deterministic and stochastic strategies. Thereafter, in Section 5.1.2, we analyze the customer behavior.

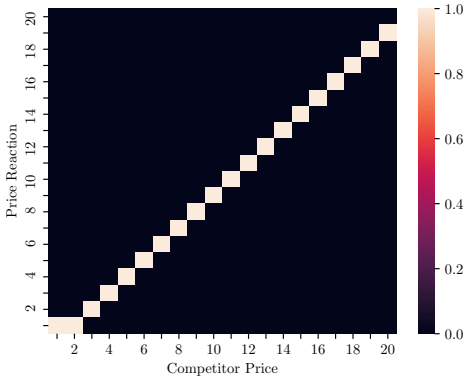


Figure 4: Visualization of price response probabilities for *Underbid* as an example for a deterministic strategy.

5.1.1 Test Strategies of the Competitor

We introduce two different groups of pricing strategies with a single representative each that will be referred to in the subsequent evaluation.

Deterministic. We test strategies that always react with the same price for a proposed price. Their behavior can be formally described as

$$\forall p_A \in prices : \exists!\, p_B : P_{react}(p_A, p_B) = 1.$$

Among those included strategies, the simplest and widely used is a strategy we call *Underbid*. The other participant's price is underbid by one unit (e.g., Δ) but respects the minimum available price. This strategy can be expressed by the response function

$$F : prices \to prices,\ p \mapsto max(min(prices),\ p - \Delta).$$

An exemplary visualization of strategy F can be found in Figure 4. We used twenty possible prices, $prices_{20} = \{\Delta, 2\Delta, \ldots, 20\}$, $\Delta = 1$. Each cell shows the probability that the competitor reacts with p_B to a current price p_A. In other words, each cell shows the result of $P_{react}(p_A, p_B)$. Resulting from that, a column contains a distribution over all price reactions p_B to a given price p_A. As *Underbid* is a deterministic strategy, in each column a single p_B makes up 100% of the occurrences. This can be clearly seen in Figure 4.

Stochastic. The second group of strategies we want to consider contains stochastic strategies only. Those are characterized by their non-deterministic behavior. A given price p_A might result in different price reactions p_B. One can compare this behavior with playing multiple pricing strategies at the same time. Figure 5 shows a stochastic strategy which, using the indicator

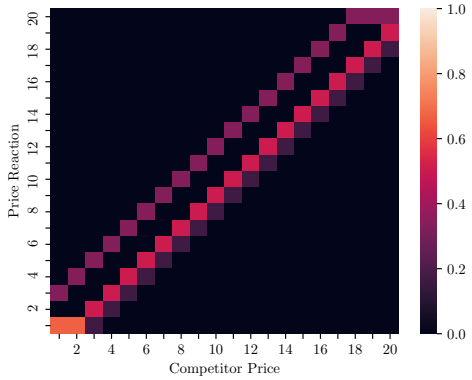


Figure 5: Visualization of an exemplary stochastic strategy.

function $1_{\{\cdot\}}$, can be described as

$$P_{react}(p_A, p_B) = \begin{cases} 1/2, & \cdot 1_{\{p_B = max(p_A - 1,\ min(prices))\}} \\ +1/6, & \cdot 1_{\{p_B = max(p_A - 2,\ min(prices))\}} \\ +1/3, & \cdot 1_{\{p_B = min(p_A + 2,\ max(prices))\}}. \end{cases}$$

5.1.2 Customer Buying Behavior

The subsequent evaluation will use a fixed customer buying strategy. Nonetheless, all models are capable of handling arbitrary customer buying behavior. We decided to pick a realistic buying behavior in order to make this evaluation as practical as possible. We define the probability that a customer buys a product given the current prices p_A and p_B as follows:

$$buy : (prices, prices) \to (0,1] :$$
$$(p_A, p_B) \mapsto 1 - \frac{\min(p_A, p_B)}{\max(prices) + 1}.$$

The customer is more likely to buy a product if the minimal price on the market is lower. If both prices are very high, it is less likely that the customer buys a product. As mentioned earlier, we assume that the customer always chooses the lower price. If both proposed prices are the same, the customer randomly chooses one market participant's product. Therefore, the probability that the customer decides participant A's product offer is defined via

$$dec_A : (prices, prices) \to [0,1] :$$
$$(p_A, p_B) \mapsto \begin{cases} 1, & \text{if } p_A < p_B \\ 1/2, & \text{if } p_A = p_B \\ 0, & \text{otherwise.} \end{cases}$$

Consequently, in the context of (1) and (2) in (3), the resulting buying probability is described by

$$P_{buy_A}(p_A, p_B \mid h, p'_B) = h \cdot buy(p_A, p_B) \cdot dec_A(p_A, p_B) + (1-h) \cdot buy(p_A, p'_B) \cdot dec_A(p_A, p'_B).$$

5.2 Results for Unknown Strategies

In the evaluation of the model for the unknown opponent's strategy, we compare how long the different exploration procedures *Assurance* and *Incentive* take to approximate the real opponent's strategy and what their profits are along the way. The quality of different learning strategies can be verified by comparing them to the optimal strategy, which can be obtained by solving (3) for the opponent's strategy.

We evaluate the two exploration procedures (*Assurance* and *Incentive*) in the setting described in Section 5.1, where the underlying opponent's strategy is either *Underbid* or *Stochastic*. We use the discount factor $\delta = 0.99$ and intervals with $h = 0.5$. We deduced $T = 100$ to be sufficient for the strategy anticipation. We choose the number of time steps after which A adjusts their strategy as $T_a = 1$. This, in return, implies that the strategy is reevaluated after every new price reaction from B.

In the following, we compare the *Assurance* procedure under different numbers of time steps for exploration T_i and the *Incentive* procedure under different incentive weights λ respectively. For this purpose, we choose T_i and λ as follows:

$$T_i \in \{0, 10, 20, 40, 100\}, \lambda \in \{0.001, 0.5, 1, 2, 5\}.$$

We use $T_i = 0$ and $\lambda = 0.001$ to get a good baseline for each approach. In order to get a profound impression of the calculated strategy at a specific time step t, we run the simulation $S = 1000$ times for $T_S = 100$ time steps. The average of A's expected profits in these simulations is divided by the simulation length T_S and will be denoted E_t. Therefore, E_t is A's expected profit with the strategy used at time step t. We denote O to be the expected profit that is achieved when the *optimal strategy* is used. O is constant as the optimal strategy does not change over time. If $E_t \approx O$, we know that A either found the optimal strategy or another strategy that produces very similar profits. If this keeps up for a greater number of time steps, A successfully identified B's strategy. An example is visualized in Figure 6. The figure depicts the development of expected profits E_t over time when utilizing the *Assurance* procedure. We used $prices = prices_{20}$ with *Underbid* as B's underlying strategy and $T_i = 20$ time steps for initial exploration.

As discussed in Section 4.2, A will be able to find the optimal strategy because B's strategy is deterministic and $T_i \geq |prices|$. This can be seen clearly in Figure 6. During exploration with *Assurance*, A's expected profits are mediocre but after exploration, the expected profits are equal to the optimal profits. In order to compare the two procedures over a longer

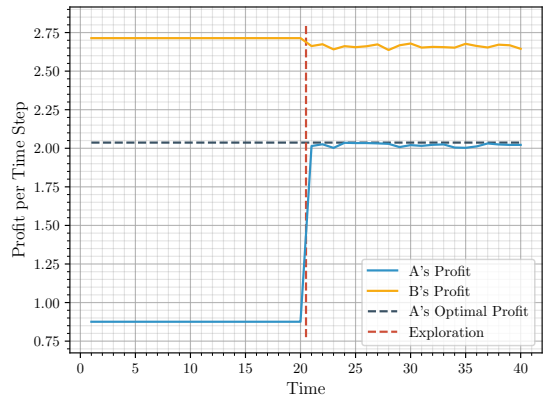


Figure 6: Profit per time step against *Underbid* on 20 prices with with *Assurance* exploration and $T_i = 20$.

time period, we set A's cumulative profit in proportion to the cumulative profit under the optimal strategy. A's cumulative profit will be denoted CE_t and the cumulative profit under the optimal strategy will be denoted CO_t. Consequently, we have $CE_t = \sum_{i=1}^{t} E_i$ and $CO_t = \sum_{i=1}^{t} O = t \cdot O$. If $CE_t = CO_t$, A found the optimal strategy or another one that achieves equal profits. Furthermore, in order for the cumulative profits to be approximately equal, the strategy needs to be used for several time steps to account for inferior strategies applied in the past. We will call $\frac{CE_t}{CO_t}$ the profit ratio. Figure 7 depicts the profit ratio of the *Assurance* and *Incentive* procedure with their respective configurations for 400 time steps.

Assurance and Underbid Strategy (Figure 7a). The figure shows that it takes a long time for larger T_i to account for losses during exploration as we know that A would find the optimal strategy after $T_i = 20$ time steps. $T_i = 10$ seems to have found the optimal strategy after initial losses as well. This can be seen because the plot is approached by $T_i = 20$. These two configurations as well as $T_i = 40$ and $T_i = 100$ will approach the optimal profit ratio of 1 on the infinite event horizon. With $T_i = 0$ the model was not able to find the optimal strategy which explains why its plot is being overtaken by that of $T_i = 20$.

Assurance and Stochastic Strategy (Figure 7b). The figure shows that for a stochastic strategy more exploration is needed. The configuration $T_i = 0$ and $T_i = 10$ converge to the same point, which means that the additional exploration did not contain any beneficial information. $T_i = 20$ results in a higher profit ratio but similar to the deterministic scenario, it takes very long for larger T_i to account for missed profits during the exploration phase.

Incentive and Underbid Strategy (Figure 7c). In the figure we can see that every configuration after some initial profits experiences a drop in profit ratio. The reason for that is because the model tries out less profitable prices after gaining enough information about profitable prices.

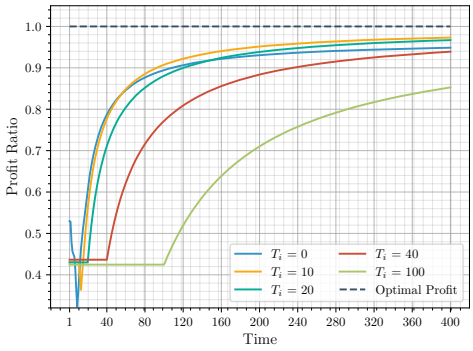
This is visualized in Figure 8 for $\lambda = 1$. The model continuously tries out higher prices. For a lower λ proposing these prices happens very fast. That is the reason why the drop for lower λ is greater compared to larger λ. Model configurations with larger λ take longer to gain confidence for the profitable prices before trying out less profitable prices. This also means that larger λ take longer to accept that the opponent strategy is deterministic. It is therefore not surprising that the order of profit ratio at $t = 400$ is the ascending order of λ. Every configuration is able to find the optimal strategy. However, we see that the larger λ the longer it takes for the model to be certain.

Incentive and Stochastic (Figure 7d). The plots in this figure have a similar shape compared to the same procedure with *Underbid* as the underlying opponent strategy. The drop is less significant which should be due to the *Stochastic* being a more forgiving strategy compared to *Underbid*. $\lambda = 0.001$ seems to not have received enough opponent reactions which can be seen as the plot is stagnating for larger t. Moreover, larger λ perform equally well.

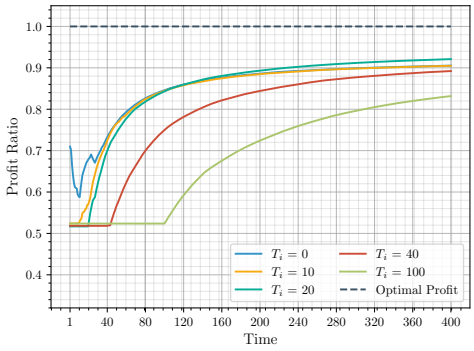
Comparing the results, we decide that the *Incentive* procedure should be preferred over the *Assurance* procedure for exploration. The *Incentive* procedure produces higher profit ratio compared to *Assurance* procedure. This is because the later needs exploration for $T_i \geq |prices|$ time steps in order to produce a good strategy which can be clearly seen in the scenario of the *Stochastic* strategy. However, if T_i is too large it takes very long to compensate the exploration phase. For the *Incentive* procedure $\lambda \approx 1$ seems to be ideal. Moreover, configurations with large λ take too long to be confident about the opponent's strategy while configurations with small λ are considerably less likely to propose a price multiple times.

5.3 Results for Self-adaptive Strategies

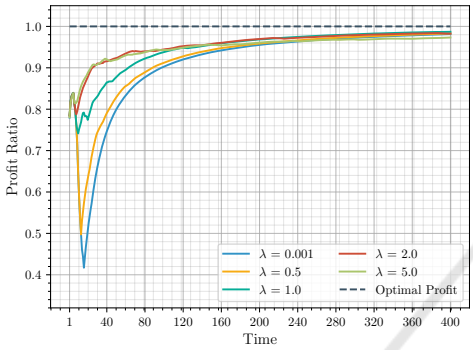
The evaluation of the interaction between two self adapting strategies will be divided into three major parts. The first of those runs the competition with two identically configured models and observes how the competition affects each of these. The second part focuses on the parameter α and its effect on model's performance. The final part examines whether both



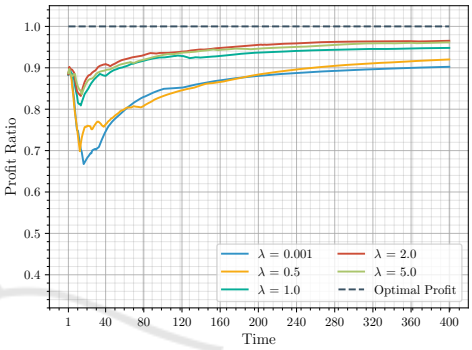
(a) *Assurance* exploration against *Underbid*



(b) *Assurance* exploration against *Stochastic*



(c) *Incentive* exploration against *Underbid*



(d) *Incentive* exploration against *Stochastic*

Figure 7: Profit ratios for *Assurance* and *Incentive* exploration against *Underbid* and *Stochastic* over 400 time steps.

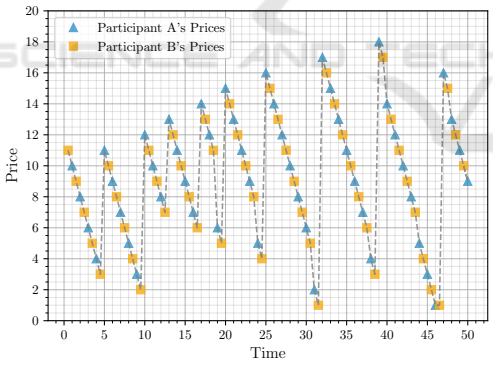


Figure 8: Market simulation of the *Incentive* procedure with $\lambda = 1$ competing against *Underbid* on 20 prices.

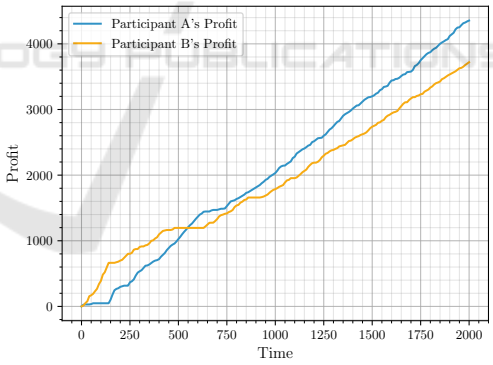
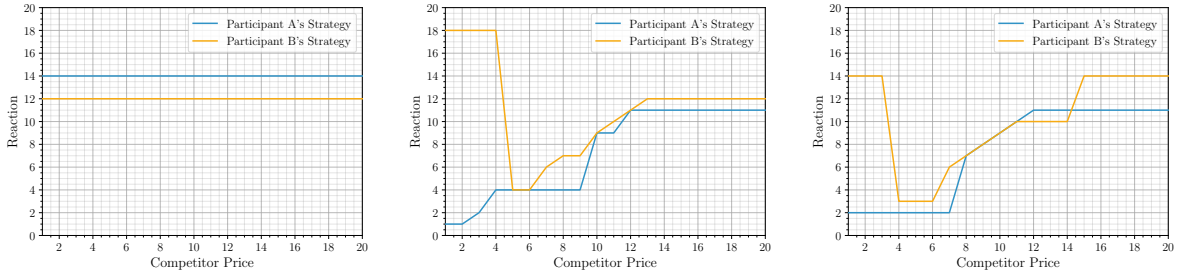


Figure 9: Profit progression of two competing, self-adapting price strategies.

strategies can form a cartel. For all of these tests, we use an interval split $h = 0.5$, a discount factor $\delta = 0.99$ and an evaluation time horizon $T = 50$. Strategy updates are performed frequently with $T_d = 10$ in order to shorten the initial exploration phase. We simulate each configuration for 2000 time points to account for long term effects. Additionally, the models use the *Incentive* technique, presented in Section 4.2, to explore prices the competitor has not reacted to.

Identical Start Conditions. In this scenario, we let two identically configured models compete. Both models differentiate between new and old reactions, $\alpha \neq 1$. We identified $\alpha = 0.8$ as suitable to account for focusing on newer reactions while keeping track of old ones, too. Figure 9 shows the progress of the profits of both strategies. In the beginning, market participant *B* is ahead due to the fact that the strategy of market participant *B* has one period of additional data during the strategy reevaluation. Therefore, it can finish its exploration phase earlier. However, the



(a) Both strategies at $t = 100$ (b) Both strategies at $t = 400$ (c) Both strategies at $t = 1000$

Figure 10: Progression of both strategies at different time points t.

earlier update has its disadvantages, too. Market participant B notices first that they have to restore a low price level at some point to gain higher profits in the future. However, in this scenario market participant A can exploit this behavior by choosing a low price, so participant B has to increase the price level. Therefore, market participant A never has to increase the price level itself, but can force market participant B to do so once the price level drops too low. Accordingly, market participant A wins out at some point and never looses the profit lead again. We see that market participant B is not able to stop the downward trend once it started. When competing for an extended period of time, both strategies enter a loop of chosen prices, which both models profit from. While market participant B looses the competition, its strategy is still optimal from its point of view. The alternative of matching a low price of market participant A is not lucrative, as there is no guarantee that market participant A will restore the price level and market participant B looses profit in the long run. When looking at the strategy evolution, we see that both models start with similar strategies to explore the respective competitor's strategy. Figure 10a shows that market participant B is ahead during the exploration phase. Both strategies evaluate the competitor reactions from most profitable to least profitable. As we can see, participant B is already using price 12, while participant A is still evaluating the more profitable price of 14.

Figure 10b shows the learning progress of both strategies at $t = 400$. Market participant B has learned that it has to restore the price level at some point, while market participant A exploits B's strategy by matching lower prices in order to force B to raise the price level afterwards. After 1000 time periods, both strategies do not change any more. The final strategies are presented in Figure 10c. We can see that both strategies underbid each other in the mid price ranges. However, they differ in their behavior once the price drops too low and also in their price reaction on too high market prices. Moreover, market participant A already deems prices below 7 as too low, while market participant B only deems below 3 as too low.

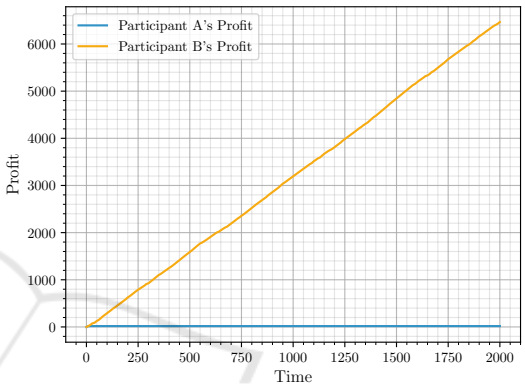


Figure 11: Profit progression of two competing, self-adapting price strategies with $\alpha_A = 0$ and $\alpha_B = 1$.

α Deviations. While the previous section focused on $\alpha = 0.8$, this section investigates the impact of selected α values on the models' performances. Figure 11 shows the competition of two extreme α values (i.e., $\alpha_A = 0$ and $\alpha_B = 1$). We see that participant B wins the competition very decidedly. $\alpha = 0$ is observed to be the worst possible setting as the incentive based learning has to start over and over again. This is due to the fact that the model loses every recorded price reaction after each strategy evaluation. Therefore, previously played prices appear to be new to the unsuspecting model. In the short term, this strategy can work out because it focuses on high profit prices first and the model with $\alpha = 1$ wants to learn about all prices instead. However, this effect is mitigated as participant B updates its strategy earlier.

Therefore, we present a more competitive setting where participant A uses $\alpha = 1$ and participant B uses $\alpha = 0.5$. An α value of 0.5 allows a model to focus on newer reactions, yet not loosing information on older ones. Figure 12 shows the profits of both competing strategies over time. We can see that participant B's

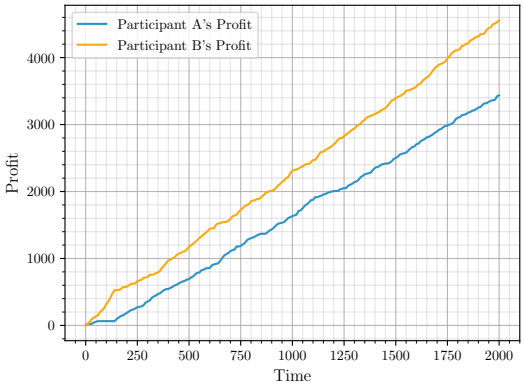


Figure 12: Profit progression of two competing, self-adapting price strategies with $\alpha_A = 1$ and $\alpha_B = 0.5$.

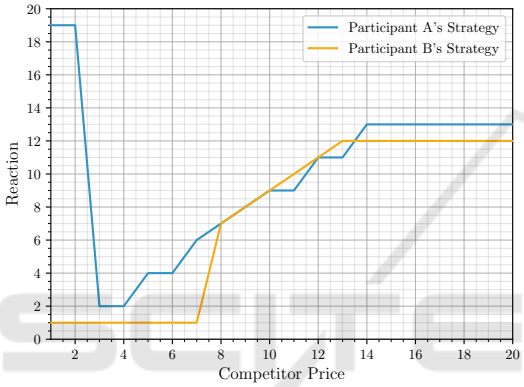


Figure 13: Strategy comparison at $t = 1000$ with $\alpha_A = 1$ and $\alpha_B = 0.5$.

strategy using $\alpha = 0.5$ is more effective.

As we saw earlier, participant B is at a structural disadvantage. Nonetheless, B is able to win over participant A due to the superior α value. In contrast to the first experiment, participant B is able to force A to restore the price level, as we can see in Figure 13. Participant A is not able to remove misleading reactions from the early exploration. Thus, market participant A is not able to compete with market participant B who adapts its strategy accordingly.

5.4 Results for Cartel Formations

The following focuses on a cartel formation using two self-learning strategies. While the previous experiments showed that the two models do not form a cartel on their own, the introduction of an artificial price reaction as discussed in Section 4.3 helps with that. Figure 14 shows the product prices p_A and p_B of both market participants over time.

We observe that both strategies explore the possible prices at first, as we see a lot of different prices

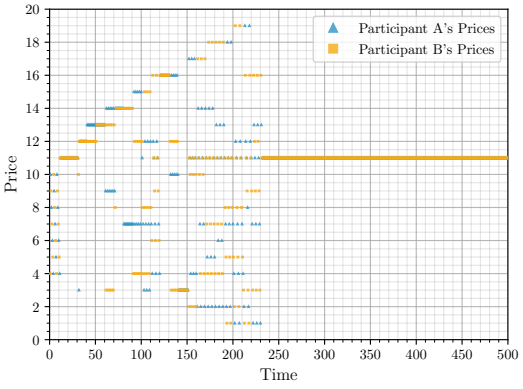


Figure 14: Price history with market participant A's artificial cartel price reaction.

played. Both participants choose the cartel price some time, but they do not form a cartel instantly. Given that participant A will always react to a price $p_B = p_B* = 11$ with $p_A = p_A* = 11$, we see that participant B at some point $t \approx 250$ decides to consistently react to p_A* with p_B* in order to form a cartel. After the formation phase, both parties continue to stick with the cartel price. As expected, the earned profits of both participants are high and both strategies outperform their competing counterparts.

6 CONCLUSION

In recent times, market participants try to adapt their prices more frequently to gain a competitive edge. Online markets offer optimal conditions to employ dynamic pricing strategies, as it is easy to observe competitor's prices and to change the own price. We analyze optimized pricing strategies for different scenarios. In all of these, we compete in a duopoly and operate under an infinite time horizon. Additionally, we allow for an arbitrary functional dependency between the sale probability and the current market situation consisting of two competitors' prices.

Firstly, we show how to explore the competitor's strategy efficiently while losing a minimum profit. We try out two different ways of estimating the competitor strategy. On the one hand, we simply cycle through prices to gain more information about specific prices. On the other hand, we use an incentive approach for motivating the model to try out prices that have not been proposed before. We find that the incentive approach should be preferred over the other as profits are considered during exploration.

Secondly, we let our self-learning strategies interact with each other. Both of the strategies estimate the respective competitor's strategy and adapt their price

responses in fixed intervals. We observe that equal strategies evolve over an extended period of time, but stop evolving at some point. Afterwards, neither strategy is changed again. When comparing different strategies, we observe that diminished knowledge of past price reactions outperforms settings without any as well as those with unlimited backward reaction tracking. Moreover, we slightly modify one strategy such that it prefers playing a cartel price. We show that both strategies stop competing once they discover the cartel price. Although customers suffer from the high price, it is the most beneficial scenario for both market participants due to high profits.

REFERENCES

Adida, E. and Perakis, G. (2010). Dynamic pricing and inventory control: Uncertainty and competition. *Operations Research*, 58:289–302.

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

Chen, M. and Chen, Z.-L. (2015). Recent developments in dynamic pricing research: Multiple products, competition, and limited demand information. *Production and Operations Management*, 24:704–731.

den Boer, A. V. (2015). Dynamic pricing and learning: Historical origins, current research, and new directions. *Surveys in Operations Research and Management Science*, 20:1–18.

Do Chung, B., Li, J., Yao, T., Kwon, C., and Friesz, T. L. (2011). Demand learning and dynamic pricing under competition in a state-space framework. *IEEE Transactions on Engineering Management*, 59:240–249.

Gallego, G. and Hu, M. (2014). Dynamic pricing of perishable assets under competition. *Management Science*, 60:1241–1259.

Gallego, G. and Topaloglu, H. (2019). *Revenue Management and Pricing Analytics*. Springer.

Gallego, G. and Wang, R. (2014). Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. *Operations Research*, 62:450–461.

Hajji, A., Pellerin, R., Léger, P.-M., Gharbi, A., and Babin, G. (2012). Dynamic pricing models for erp systems under network externality. *International Journal of Production Economics*, 135:708.

He, Q.-C. and Chen, Y.-J. (2018). Dynamic pricing of electronic products with consumer reviews. *Omega*, 80:123–134.

Huang, Y.-S., Hsu, C.-S., and Ho, J.-W. (2014). Dynamic pricing for fashion goods with partial backlogging. *International Journal of Production Research*, 52:4299–4314.

Levin, Y., McGill, J., and Nediak, M. (2009). Dynamic pricing in the presence of strategic consumers and oligopolistic competition. *Management Science*, 55:32–46.

Liu, Q. and Zhang, D. (2013). Dynamic pricing competition with strategic customers under vertical product differentiation. *Management Science*, 59:84–101.

Martínez-de Albéniz, V. and Talluri, K. (2011). Dynamic price competition with fixed capacities. *Management Science*, 57:1078–1093.

McGill, J. and van Ryzin, G. (1999). Revenue management: Research overview and prospects. *Transportation Science*, 33:233–256.

Noel, M. (2007). Edgeworth price cycles, cost-based pricing, and sticky pricing in retail gasoline markets. *The Review of Economics and Statistics*, 89:324–334.

Noel, M. (2012). Edgeworth price cycles and intertemporal price discrimination. *Energy Economics - ENERG ECON*, 34.

Phillips, R. L. (2005). *Pricing and Revenue Optimization*. Stanford University Press.

Schlosser, R. (2019a). Dynamic pricing under competition with data-driven price anticipations and endogenous reference price effects. *Journal of Revenue and Pricing Management*, 16:451–464.

Schlosser, R. (2019b). Stochastic dynamic pricing with strategic customers and reference price effects. *ICORES 2019*, pages 179–188.

Schlosser, R. and Boissier, M. (2017). Optimal price reaction strategies in the presence of active and passive competitors. *ICORES 2017*, pages 47–56.

Schlosser, R. and Boissier, M. (2018). Dynamic pricing under competition on online marketplaces: A data-driven approach. *International Conference on Knowledge Discovery and Data Mining*, pages 705–714.

Schlosser, R. and Richly, K. (2018). Dynamic pricing strategies in a finite horizon duopoly with partial information. *ICORES 2018*, pages 21–30.

Sweeting, A. (2012). Dynamic pricing behavior in perishable goods markets: Evidence from secondary markets for major league baseball tickets. *Journal of Political Economy*, 120:1133–1172.

Talluri, K. T. and Van Ryzin, G. J. (2006). *The Theory and Practice of Revenue Management*. Springer.

Tong, T., Dai, H., Xiao, Q., and Yan, N. (2020). Will dynamic pricing outperform? theoretical analysis and empirical evidence from o2o on-demand food service market. *International Journal of Production Economics*, 219:375–385.

Tsai, W.-H. and Hung, S.-J. (2009). Dynamic pricing and revenue management process in internet retailing under uncertainty: An integrated real options approach. *Omega*, 37:471–481.