

Buchali, Katrin; Grüb, Jens; Muijs, Matthias; Schwalbe, Ulrich

**Conference Paper**

## Strategic Choice of Price-Setting Algorithms

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2023: Growth and the "sociale Frage"

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Buchali, Katrin; Grüb, Jens; Muijs, Matthias; Schwalbe, Ulrich (2023) : Strategic Choice of Price-Setting Algorithms, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2023: Growth and the "sociale Frage", ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/277695>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Strategic Choice of Price-Setting Algorithms

Katrin Buchali<sup>1</sup>

Jens Grüb<sup>1,\*</sup>

Matthias Muijs<sup>1</sup>

Ulrich Schwalbe<sup>1</sup>

February 2023

## Abstract

Recent experimental simulations have shown that autonomous pricing algorithms are able to learn collusive behavior and thus charge supra-competitive prices without being explicitly programmed to do so. These simulations assume, however, that both firms employ the identical price-setting algorithm based on Q-learning. Thus, the question arises whether the underlying assumption that both firms employ a Q-learning algorithm can be supported as an equilibrium in a game where firms can choose between different pricing rules. Our simulations show that when both firms use a learning algorithm, the outcome is not an equilibrium when alternative price setting rules are available. In fact, simpler price setting rules as for example meeting competition clauses yield higher payoffs compared to Q-learning algorithms.

**JEL-Codes:** D43, D83, L13, L49

**Keywords:** pricing algorithms, algorithmic collusion, reinforcement learning

## 1 Introduction

Recent experimental simulations have shown that autonomous pricing algorithms are able to learn collusive behavior without being explicitly programmed to do so (Calvano et al. 2020; Klein 2021). The pricing algorithm used in both studies is Q-learning, which is a method of reinforcement learning that learns autonomously through trial and error in a similar way humans do.

Both studies consider a duopolistic market with Bertrand competition where firms either produce differentiated goods and set prices simultaneously (Calvano et al. 2020) or offer

---

<sup>1</sup> University of Hohenheim, Institute of Economics

\* Corresponding author: j.grueb@uni-hohenheim.de; Postal Address: Universität Hohenheim, Institut für Volkswirtschaftslehre, Lehrstuhl für Mikroökonomik (520C), 70593 Stuttgart, Germany.

a homogeneous good and price sequentially (Klein 2021). The firms employ the identical price-setting algorithm based on Q-learning and the simulations show that these self-learning pricing algorithms are indeed capable of learning collusive behavior and thus charge supra-competitive prices. This confirms what has already been suspected in several earlier contributions such as Ezrachi and Stucke (2016), Mehra (2016), and Ezrachi and Stucke (2017).

The two models assume that firms employ the same pricing algorithm. In practice, however, many different pricing rules are used – ranging from simple rules of thumb to complex learning algorithms (Calzolari and Hanspach 2022). Hence, the question arises whether the underlying assumption that both firms employ the same algorithm based on Q-learning is justified. Stated otherwise: Does the choice of Q-learning algorithms by the two firms represent an equilibrium or would they use alternative algorithms that promise a higher payoff if they have the choice between different pricing rules? If none of the firms have an incentive to use Q-learning algorithms as pricing mechanisms, then the supposedly concerning results are merely theoretical.

In what follows, within the framework of a simple game-theoretic model, we want to determine whether the assumption implicit in the two simulation models of Calvano et al. (2020) and Klein (2021) that both firms would use price-setting algorithms based on Q-learning can be justified, i.e. whether this is the equilibrium outcome of a game. For this purpose, the same market environment is considered as in the previously mentioned models, i.e. a market with two identical firms that offer a heterogeneous product and compete with prices. The strategies of the firms, however, are not the prices, but the rules according to which the prices are set. It is assumed here that firms have a choice between four different pricing rules. First, we are considering two types of learning algorithms, namely the Q-learning algorithm in Calvano et al. and Klein’s models, and a multi-armed-bandit algorithm (MAB). We then examine two types of pricing rules that are not based on learning algorithms and are frequently observed in practice: a meeting competition clause and a simple rule to undercut the rival’s price. Using the simulation method of Calvano et al. (2020), the results indicate that in this game of algorithm choice, strategy combinations that involve only self-learning algorithms do not constitute a Nash equilibrium.

## 2 Pricing Algorithms and Experiments

The two reinforcement learning algorithms we consider are Q-learning and a MAB algorithm (Sutton and Barto 2018) where the agents interact in an environment that has a finite number of states. They select actions from a given set in order to maximize a reward that an agent receives in each period. The rewards and the transitions between

the states of the environment are determined by the chosen actions. Both algorithms use Q-values which give the expected reward for a given action. In contrast to MABs, for Q-learning, these values also depend on the states. Once an action is selected and the agents have received their reward, they observe the next state and update the respective Q-value.

Agents follow a policy that determines which action is chosen in each period. It is assumed that the agents employ an  $\varepsilon$ -greedy strategy, which is a method of balancing exploration and exploitation. Exploration allows an agent to improve its current information about an action while exploitation is aimed at getting the most reward by exploiting the agent's current action-value estimates by selecting the "greedy" action. By choosing randomly between exploration and exploitation the discounted reward obtained over all periods is maximized.<sup>2</sup>

There are two important parameters that influence how well these algorithms learn, the learning rate and the discount factor. The learning rate  $\alpha$  determines how much weight is put on newly learned Q-values. The discount factor  $\delta$  determines how much emphasis is placed on future rewards as actions also influence the transitions to the next state.

The other two price setting rules considered are a meeting competition clause and a simple price heuristic. An example of a meeting competition clause is a price guarantee, i.e. consumers can claim a price discount up to the difference to the lowest price in the market.<sup>3</sup> An example of a price heuristic is the automatic pricing option at *Amazon*. The *Amazon Seller Central Europe* website states that "[f]or example, you can create a rule that stays 0,10 EUR below the Buy Box price".<sup>4</sup> The meeting competition clause is designed to always set a price that is closest to the perfectly collusive price and automatically match any lower price set by a competitor. The heuristic is implemented such that it always undercuts the competitor's price by the smallest monetary unit provided that the price is not below marginal cost. The heuristic matches the competitor's price if it is the lowest price in the action set.

It is assumed in the simulations that the action sets have a lower bound of 10 percent below the price in a non-cooperative Nash equilibrium and an upper bound of 10 percent above the price under perfect collusion. This interval is then equally divided into  $k = 6$  prices. Exploration follows an  $\varepsilon$ -greedy strategy where the exploration probability in the first round is one and decreases asymptotically towards zero as the number of rounds

---

<sup>2</sup> An alternative MAB to the one considered here is the EXP3 algorithm that uses weights instead of Q-values to determine whether to explore or exploit. The weights are updated exponentially in each period. This ensures that more emphasis is put on profitable arms of the bandit.

<sup>3</sup> For a survey on different types of price relationship agreements see Office of Fair Trading (2012).

<sup>4</sup> [https://sellercentral.amazon.de/gp/help/external/201995750?language=en\\_DE&ref=efph\\_201995750\\_cont\\_202166010accessedon25.03.2022](https://sellercentral.amazon.de/gp/help/external/201995750?language=en_DE&ref=efph_201995750_cont_202166010accessedon25.03.2022).

increases.<sup>5</sup> The learning rate is  $\alpha = 0.1$  and the discount factor is  $\delta = 0.95$ . The per-period reward of a firm is equal to the profit it obtained in that period. Each experiment consists of ten runs with each run lasting for a maximum of 2,000,000 periods. A run is terminated earlier if both agents chose the same action or repeat the same cycle of actions for 10,000 consecutive periods. The results of the simulations are reported in the next section.

### 3 Game and Equilibria

Instead of setting prices, firms simultaneously choose one of four algorithms which in turn sets the price autonomously. These algorithms include Q-learning (Q), multi-armed bandits (B), meeting competition clauses (M) and price heuristics (H) such that the strategy set of player  $i$  is characterized by  $S_i = \{Q, B, M, H\}$ .

In our experiment, we assume that two firms  $i, j = 1, 2$  with  $i \neq j$  produce heterogeneous goods with marginal costs of  $c = 1$  and compete in prices. The demand each firm faces is described by a logit demand function which is also used by Calvano et al. (2020):

$$D_i(p_i, p_j) = \frac{\exp(\frac{2-p_i}{0.25})}{\exp(\frac{2-p_i}{0.25}) + \exp(\frac{2-p_j}{0.25}) + 1}.$$

The prices and profits under oligopolistic competition are  $p_i^N \approx 1.47$  and  $\pi_i^N \approx 0.22$ , respectively. Perfect collusion is characterized by price  $p_i^C \approx 1.92$  and profits  $\pi_i^C \approx 0.34$ . As described in the previous section, the action set is given by  $A_i = \{1.323, 1.485, 1.647, 1.810, 1.972, 2.134\}$ .

The payoffs for this game are given by the results from the pricing algorithm experiments as averages over the last 1,000 periods over all runs and are shown in Table 1<sup>6</sup>

	Q	B	M	H
Q	0.314, 0.307	0.291, 0.247	0.336, 0.336	0.225, 0.407
B	0.247, 0.291	0.228, 0.228	0.336, 0.336	0.225, 0.407
M	0.336, 0.336	0.336, 0.336	<i>0.336, 0.336</i>	<i>0.156, 0.156</i>
H	0.407, 0.225	0.407, 0.225	<i>0.156, 0.156</i>	<i>0.156, 0.156</i>

Table 1: Normal-form game with *logit* demand

The Nash equilibria in pure strategies are  $\{(M, M), (Q, H), (H, Q), (B, H), (H, B)\}$ . Obviously, the strategy combination  $(Q, Q)$  is not a Nash equilibrium.<sup>7</sup>

<sup>5</sup> After approximately 700,000 periods, the exploration probability drops below 0.1%.

<sup>6</sup> Italicized payoffs were not determined using simulations, but rather based on economic reasoning.

<sup>7</sup> We have employed the MAB algorithm instead of the EXP3 as used in den Boer et al. (2022) because it yields even higher total discounted payoffs.

To assess the robustness of the previous results, an alternative demand function given by

$$D_i(p_i, p_j) = 2 - p_i + \frac{1}{2}p_j$$

is considered. For this demand function, the prices and profits under oligopolistic competition are given by  $p_i^N = 2.0$  and  $\pi_i^N = 1.0$ . Perfect collusion leads to a price of  $p_i^C = 2.5$  and yields profits of  $\pi_i^C = 1.125$ . The action sets are therefore given by  $A_i = \{1.8, 1.99, 2.18, 2.37, 2.56, 2.75\}$ . The experimental results are reported in Table 2.

	<i>Q</i>	<i>B</i>	<i>M</i>	<i>H</i>
<i>Q</i>	1.074, 1.074	1.073, 1.015	1.123, 1.123	0.986, 1.186
<i>B</i>	1.015, 1.073	0.995, 0.995	1.123, 1.123	0.986, 1.186
<i>M</i>	1.123, 1.123	1.123, 1.123	<i>1.123, 1.123</i>	<i>0.880, 0.880</i>
<i>H</i>	1.186, 0.986	1.186, 0.986	<i>0.880, 0.880</i>	<i>0.880, 0.880</i>

Table 2: Normal-form game with *linear* demand

The Nash equilibria are exactly the same as before:  $\{(M, M), (Q, H), (H, Q), (B, H), (H, B)\}$  and thus independent of the underlying demand function.

## 4 Conclusion

Our simulations demonstrate that when both companies use a learning algorithm, the outcome is not an equilibrium when alternative price setting rules are available. This holds true for the Q-learning strategy as well as for the even easier to implement MAB. Thus, the situation described by Calvano et al. (2020), wherein firms employ such an algorithm which then learns to set supra-competitive prices, represents a situation wherein both firms would have an incentive to deviate from their strategy.

This result suggests that firms are more likely to use a simple pricing rule like price guarantees which are significantly cheaper to implement and also promise a higher payoff. Therefore, fears that learning algorithms may result in more opportunities for collusion appear unfounded, as even comparatively simple pricing rules seem to be more effective in producing cartel-like behavior.<sup>8</sup> Consequently, which antitrust regulations should be adopted in these markets depends on which pricing rules are available to the firms. At present, therefore, there seems to be no need for specific competition law regulations regarding algorithmic pricing. A closer examination of the anticompetitive effects of simple pricing rules would be more appropriate instead.

---

<sup>8</sup> That price matching guarantees lead to higher prices in online markets has been shown by Zhuo (2017).

## References

- Calvano, E., G. Calzolari, V. Denicolò, and S. Pastorello (2020): Artificial intelligence, algorithmic pricing, and collusion, *American Economic Review*, Vol. 110, No. 10, pp. 3267–3329.
- Calzolari, G. and P. Hanspach (2022): Pricing algorithms out of the box: a study of the repricing industry, forthcoming.
- den Boer, A. V., J. Meylahn, and M. P. Schinkel (2022): Artificial collusion: Examining supracompetitive pricing by Q-learning algorithms, Amsterdam Law School Research Paper No. 2022-25, Amsterdam Center for Law & Economics Working Paper No. 2022-06, DOI: 10.2139/ssrn.4213600.
- Ezrachi, A. and M. E. Stucke (2016): How pricing bots could form cartels and make things more expensive, *Harvard Business Review*, Vol. 27.
- Ezrachi, A. and M. E. Stucke (2017): Two Artificial Neural Networks Meet in an Online Hub and Change the Future (of Competition, Market Dynamics and Society), Oxford Legal Studies Research Paper No. 24/2017, University of Tennessee Legal Studies Research Paper No. 323, URL: <https://ssrn.com/abstract=2949434>.
- Klein, T. (2021): Autonomous algorithmic collusion: Q-learning under sequential pricing, *RAND Journal of Economics*, Vol. 52, No. 3, pp. 538–599.
- Mehra, S. K. (2016): Antitrust and the robo-seller: Competition in the time of algorithms, *Minnesota Law Review*, Vol. 100, pp. 1323–1375.
- Office of Fair Trading (2012): Can 'Fair' Prices be Unfair? A Review of Price Relationship Agreements, URL: [https://www.learlab.com/wp-content/uploads/2016/04/Can-%E2%80%99Fair%E2%80%99-Prices-Be-Unfair\\_-A-Review-of-Price-Relationship-Agreements.pdf](https://www.learlab.com/wp-content/uploads/2016/04/Can-%E2%80%99Fair%E2%80%99-Prices-Be-Unfair_-A-Review-of-Price-Relationship-Agreements.pdf).
- Sutton, R. S. and A. G. Barto (2018): *Reinforcement Learning: An Introduction*, Cambridge, Massachusetts: MIT Press, DOI: 10.1109/TNN.1998.712192.
- Zhuo, R. (2017): Do Low-Price Guarantees Guarantee Low Prices? Evidence from Competition between Amazon and Big-Box Stores, *Journal of Industrial Economics*, Vol. 65, No. 4, pp. 719–738.