University of Amsterdam

ALGORITHMIC COLLUSION UNDER ASYMMETRIC PRICING FREQUENCIES

ANOUCK RIETVELD

Thesis BSc Econometrics

June, 2022

Firms increasingly use algorithms to price services and goods in online and offline markets. Recent studies show that using pricing technologies can result in collusion, which is detrimental to consumers. This thesis experimentally examines whether collusion occurs in a duopoly market where firms use asymmetric pricing frequencies. The findings suggest that supra-competitive outcomes do arise. However, the algorithms do not always converge to an equilibrium. Additionally, the average price levels set by both firms are higher when one firm has a faster pricing frequency compared to the setting where pricing frequencies are symmetric. These findings show that it is essential when investigating algorithmic collusion to account for heterogeneity in frequency technologies. Further research is necessary to develop adequate policies and laws concerning algorithmic collusion.

KEYWORDS: algorithmic pricing, collusion, Q-learning, multi-agent reinforcement learning, algorithmic decision-making, pricing algorithms, machine learning.

Anouck Rietveld: anouckrietveld@gmail.com
Supervisor: Dr. M. A. L. Koster

This document is written by Anouck Rietveld who declares to take full responsibility for the contents of this document. I declare that the text and the work presented in this document are original and that no sources other than those mentioned in the text and its references have been used in creating it. UvA Economics and Business is responsible solely for the supervision of completion of the work and submission, not for the contents.

Table of Contents

1. INTRODUCTION

Automated pricing has become widely used over the past few years with the rise of newly discovered machine learning techniques (Calvano et al. (2020a)). Firms increasingly adopt self-learning algorithms to set their prices since these algorithms can handle vast quantities of data, interact and converge to an optimal strategy (Assad et al. (2020)). Automating the price-setting process allows firms to make pricing decisions with unprecedented speed and sophistication (Brown and MacKay (2021)). Furthermore, the advent of online marketplaces booms the diffusion of artificial intelligence-powered algorithms for setting prices (Brown and MacKay (2021)). Online sales embody an increasing number of different markets, such as the insurance, accommodation, and automobile industry. Moreover, offline markets are also gradually applying pricing algorithms, for example, gas stations (Schechner (2017)).

These algorithms learn to set the most optimal price by developing pricing strategies from scratch. The concern with these machine-learning programs is that they learn to collude without communicating with each other while only being instructed to maximise profits (Calvano et al. (2020a)). Collusion is when firms collectively set the prices of their goods and services above the competitive level to earn higher profits, which would be undesirable from the consumers' perspective (Harrington (2018)). Unsurprisingly this topic raises extensive interest from antitrust authorities, economic organisations, and competition-law experts since this collusion would conflict with the competition policies and would be detrimental to consumers ((OECD (2017), Competition Bureau (2018), Autorité de la Concurrence and Bundeskartellamt (2019)).

Over the past three years, the literature on algorithmic collusion has expanded. There are contributions from the fields of economics, law and computer science. Some studies take an experimental approach, where firms set their prices simultaneously in a simulated market. These papers illustrate that the reinforcement learning algorithms learn to play a collusive strategy, resulting in supra-competitive prices. One paper by Klein (2019) deviates from the simultaneous price-setting and adopts a sequential pricing framework. One firm sets its price every even period, the other every uneven period. Also, in this context, the outcomes are collusive. Nonetheless, when looking at empirical data, most algorithms used in the market will not set their prices at the same time or frequency argued by Brown and MacKay (2021). Another theoretical finding by Brown and MacKay (2021) suggests that firms with different price-setting frequencies will generate higher prices. The simultaneous Bertrand equilibrium may be the exception rather than the rule in the market. Therefore, creating a simulation where firms set their prices at different frequencies will better reflect reality, which will be the main focus of this thesis.

Apart from the theoretical results by Brown and MacKay (2021) on asymmetric pricing frequencies, there is little known about whether machine learning algorithms that have asymmetric pricing frequencies will learn to collude and how essential it is to account for this asymmetry. If it is true that asymmetric frequencies lead to higher prices for both firms compared to the symmetric case, then it would mean that the problem may be even more substantial than the existing literature makes it seem. Furthermore, accounting for asymmetric pricing frequencies reflects reality better. Brown and MacKay (2021) report that pricing patterns in empirical frequency data are inconsistent with the standard simultaneous price-setting model.

It is relevant to understand the effects of algorithms on competition by demonstrating to what extent and the conditions under which collusive outcomes may emerge. Motivated by these facts, this paper investigates whether and how algorithms can learn to collude. More specifically, a model is developed in which two agents (the firms) set their prices with asymmetric frequency using Q-learning algorithms. In doing so, this thesis aims to answer the following research question: are Q-learning algorithms able to learn supra-competitive strategies in a duopoly model with asymmetric pricing frequencies? In addition, with subquestion: to what extent is accounting for asymmetric pricing frequencies relevant when investigating algorithmic collusion?

The remainder of this thesis is organised as follows. First, the relevant literature is discussed in chapter 2. Chapter 3 discusses the methodology of how the research is conducted. The results are presented in chapter 4, and chapter 5 contains the discussion. Lastly, the conclusion is found in chapter 6.

2. THEORETICAL FRAMEWORK

The following paragraphs discuss the literature about algorithmic collusion. First, it considers the papers that take an empirical approach. Then it will look at studies that examine collusion in an experimental setting. Lastly, there will be focused on heterogeneous price-setting by discussing two theoretical papers.

2.1. *Empirical Studies*

Chen et al. (2016) developed a method to detect and analyse algorithmic pricing on Amazon Marketplace. They found that over five hundred sellers used algorithms to set their prices for many different types of goods and that these prices tended to be higher. However, they do not explicitly show that these higher prices are caused by using machine learning. Assad et al. (2020) published a paper providing the first empirical analysis of the causal association between algorithmic pricing and competition in the German retail gasoline market. They demonstrate that the adoption of algorithms increases margins by 9% in non-monopoly markets. Specifically, there is a margin increase of 28% in a duopoly. The higher price margins rise steadily over time, which could be consistent with some learning behaviour curve. In addition, they have observed many duopoly markets that feature asymmetric adoption of algorithmic pricing technology, meaning that one firm uses price-setting algorithms while the other does not. In those markets, on the other hand, there is no statistically significant change in margins. These studies provide some empirical evidence of algorithmic collusion, although it remains minimal.

2.2. *Experimental Studies*

Several papers have reported the anti-competitive behaviour of algorithms by simulations. The first to report the collusive behaviour of algorithms were Waltman and Kaymak (2008). They show that, in general, Q-learning algorithms learn to collude with each other in infinitely repeated Cournot oligopoly games. However, these supra-competitive prices could result from a failure to optimise since, in their paper, they did not test for equilibrium behaviour. Calvano et al. (2020b) have extended the research by analysing collusive behaviour in a repeated Bertrand competition by adding a reward-punishment mechanism. The presence of such property ensures that the collusive prices may be obtained in equilibrium. Their results also show that Q-learning algorithms consistently learn to set supra-competitive prices. The findings from Calvano et al. (2020b) are in line with the paper by Klein (2019), which again demonstrates how collusive strategies can arise from Q-learning algorithms. However, Klein (2019) diverges from the simultaneous price-setting mechanism of Calvano et al. (2020b) by choosing a sequential, infinitely repeated framework in which the two firms take turns setting prices. A similar approach to that of Calvano et al. (2020b) and Klein (2019) is taken by Abada and Lambin (2020), who focus specifically on the energy industry. Correspondingly to Calvano et al. (2020b), they assume the simultaneous move framework. Again they report that Q-learners learn to collude. They add a special feature to their market, namely that they also let the agents encounter an arbitrage problem where they can sell or buy inventory. A year later, Hansen et al. (2021) identified that even in the setting in which agents do not observe the competitor's prices, supra-competitive outcomes arise.

The papers all show by simulations that machine-learning algorithms are capable of collusive outcomes in different economic environments. The existing literature assumes that firms have symmetric technologies to set their prices. However, there is economic evidence that firms alter their prices at differing frequencies across markets. Klenow and Malin (2010) review micro-pricing studies on the role of price setting in business cycles. They find that the frequency of price changes differs considerably across commodities, and the timing of price changes is poorly synced across suppliers. Of the experimental studies, only Klein (2019) has considered the case of a sequential move framework. Nonetheless, the case of asymmetric price-setting frequencies has not been explored yet. Two articles consider the case of asymmetric pricing frequencies from a theoretical point of view, which will be discussed in the next paragraph.

2.3. *Asymmetric Pricing Frequency Studies*

Salcedo (2015) considers a theoretical duopoly model in which firms commit to mixed-strategy pricing algorithms in the short run and argues that the optimised algorithms will inevitably reach a collusive outcome. He also considers the asymmetric case in which one firm revises its algorithm more frequently than the other and shows that in the limit, this will result in a unique equilibrium outcome in which the firm that is committed to one algorithm will be the Stackelberg leader with monopoly prices. However, his results assume that the firm's algorithms will be periodically revealed to their competitors that can 'decode' the other algorithm for a certain period. After, it can adjust its algorithm to account for this. Therefore, it is unclear whether these collusive outcomes will also hold in practice, where decoding is not always possible.

Brown and MacKay (2021) look more carefully at heterogeneous pricing technologies and their implications on prices. They perform a theoretical analysis in which they argue that a critical feature in pricing algorithms is the conditioning on the rival's prices. In their analysis, they focus on Markov perfect equilibria as in Maskin and Tirole (1988). First of all, using a dataset of high-frequency prices, their paper exemplifies pricing patterns consistent with algorithmic software, and secondly, the patterns contradict any simultaneous price-setting behaviour. Furthermore, they develop a model in which firms do not play the simultaneous Bertrand pricing game but a repeated game, expressed as sequence of single-period stage games. Compared to the Bertrand-Nash equilibrium, prices and margins increase without the algorithms learning to collude or failing to learn to optimise. Their theory predicts that when there is an asymmetry in price adoption, the firms with a faster price-setting algorithm have lower prices in equilibrium than firms with less sophisticated technology. The lower price for the firm with superior technology is because the best response of the pricing algorithm is often to undercut the competitor's prices. The more significant the asymmetry in pricing technology, the higher the prices and margins should be. Furthermore, they stress that the firms are incentivised to self-select into asymmetric pricing frequencies since symmetric frequencies will yield the Bertrand-Nash equilibrium and, therefore, a lower profit. The results of Brown and MacKay (2021) illustrate that it is essential to account for heterogeneity in pricing frequency.

This paper aims to simulate an economic environment where agents set their prices using Q-learning algorithms and a sequential pricing mechanism with asymmetric pricing frequencies. The frequency feature of the pricing algorithm has not yet been studied by simulations before. Moreover, it will reflect the real-world better since, in reality, firms do not update their prices simultaneously and with the same frequency. Furthermore, it will not be assumed that there is knowledge of the competitor's algorithms as in Brown and MacKay (2021), since not in all markets this knowledge is available. The problem of collusive outcomes notwithstanding will be viewed in this paper as quite general and will not be focused on a particular market. The problem of collusive pricing is still in its infancy, and there is an enormous scope for future research that incorporates the features of specific markets.

3. METHODOLOGY

This paper explores the collusive reinforcement learning behaviour in a duopoly where the two firms are asymmetric in their pricing frequency. The model is based on the environment by Maskin and Tirole (1988) and extends the results by Klein (2019). All agents are Q-learners. This section describes the economic environment in which the simulations will take place. Additionally, it will elaborate more on the Q-learning algorithm and which analyses will be performed on the results.

3.1. *Asymmetric Pricing Frequency Duopoly Model*

Competition takes place between two firms $(i = 1, 2)$ in an infinitely repeated discrete time t $(t = 0, 1, 2, ...)$, where two consecutive periods are equal sized. Prices are taken from a discrete set between 0 and 1, $P = \{0, \frac{1}{k}, \frac{2}{k}, ..., 1\}$. The agents cannot set their price in units smaller than a certain number, this restriction is enforced to ensure that optimal reactions exist. In order to compare the addition of asymmetric price setting with the results of Klein (2019), the analysis is restricted to the simple setting in which there are homogeneous goods with linear demand. Furthermore, it is assumed that the goods produced are perfect substitutes, and that there are no marginal or fixed costs. At time t, the profits for the two firms are $\pi_i(p_i, p_j)$ which is a function of the two firm's current prices p_i and p_j, where $j \in \{1,2\} \setminus i$, and is given by $\pi_i(p_i, p_j) = p_i D_i(p_i, p_j)$. Let $D_i(p_i, p_j)$ denote the market demand function as a function of own price p_i and opponent's price p_j. The firms share the market equally when they charge the same price. Considering these assumptions, the following profit function for firm i is retrieved

$$\pi_i(p_i, p_j) = \begin{cases} p_i(1-p_i), & \text{if } p_i < p_j \\ 0.5p_i(1-p_i), & \text{if } p_i = p_j \\ 0, & \text{if } p_i > p_j \end{cases} \tag{1}$$

The value p that maximises (1) is the monopoly price p^M, referred to as the collusive or joint-profit maximising price $p^C = 0.5$ [1]. Related to this, the profit π_i of the two firms, $\pi_i = 0.125$ [2]. Firms discount future profits with a discount factor $\gamma \in [0,1)$. It is often thought to be close to 1 since it is expected that firms can change prices quite quickly, and the periods between setting the prices are small (t is small). Firms have the incentive to maximise their stream of profits, which gives the following maximisation problem

$$\max \sum_{s=0}^{\infty} \gamma^s \pi_i(p_{i,t+s}, p_{j,t+s}) \tag{2}$$

The two firms have different frequencies in when they will change their price $p_i \in P$. Each firm i start of the simulation by updating its price at $t = 0$ to 0. The pricing frequencies for both firms are defined as two subsets $\mathbb{N}_i(t_i), \mathbb{N}_j(t_j) \subset \mathbb{N}$. These subsets contain the period numbers in which the firm can update its price. For example when firm i has a pricing frequency of 5, it will update its price at $t_i = 0, 5, 10, \ldots$. Therefore $\mathbb{N}_i(t_i) = \{0, 5, 10, \ldots\}$. And when firm j has a pricing frequency of 2, $\mathbb{N}_j(t_j) = \{0, 2, 4, \ldots\}$. When the two firms can both update their price in the same period, the firm that chooses their price second in that turn will not know what price the first firm has chosen before. This will be revealed at the end of the period when they get their profit and update their Q-table. When they do not change their price in a certain period,

[1] This can be seen from taking the derivative for the profit for the first two cases, when $p_i < p_j$ and $p_i = p_j$: (1) $\frac{\partial p_i(1-p_i)}{\partial p_i} = 1 - 2p_i = 0 \Rightarrow p_i = 0.5$; (2) $\frac{\partial 0.5 p_i(1-p_i)}{\partial p_i} = 0.5 - p_i = 0 \Rightarrow p_i = 0.5$.

[2] Obtained by simply plugging in p^C in the expression for the profit.

they will keep the last price they choose and will not update the Q-table. Furthermore, besides the deterministic frequency case, the firms will have a stochastic pricing frequency which will incorporate some degree of uncertainty in the model. Which means for firm i that on average it updates its price every 5 periods. This will be done by taking a number from a uniform distribution from 0 to 1 and when the number is (in this case) smaller than $\frac{1}{6}$ the firm is allowed to change its price. Same will hold for firm j but then with an average pricing frequency of 2.

The Markov property is imposed, which in this case means that the strategy of an agent only depends on the last price change $p_{j,t-1}$ by the other firm, and his own last price change $p_{i,t-1}$. The states are therefore the price pair of the two prices chosen by the firms. Since there are seven different prices to choose from, there will be $7^2 = 48$ different price combinations of prices therefore states.

3.2. *Q-learning Algorithm*

As mentioned before in this paper, the focus lies on the model-free Q-learning algorithm. This model was developed by Watkins (1989) initially to deal with stationary Markov decision processes. In this paper, it will be used for infinitely repeated games. For choosing Q-learning, there are several reasons. First of all, Q-learning is widely used among computer scientists (Johnson et al. (2020)), making it more likely that companies use similar kinds of algorithms. Second, the algorithms are simple and consist of just a few parameters, making the interpretation easier. Furthermore, this choice of algorithm is in line with some articles discussed (Waltman and Kaymak (2008), Calvano et al. (2020b), Klein (2019), Abada and Lambin (2020)), which makes it easier to see what the impact of asymmetric pricing is by comparing the results.

Agents attempt to find the optimal strategy to accomplish their goal while not knowing what the demand in the market is, and consequently the profits they will obtain. Their goal is maximising the expected present value of the stream of profits

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \pi_t\right] \tag{3}$$

where $\gamma \in [0, 1)$ is a discount factor. The algorithms have a Q-function $Q_i(p_{it}, s_t)$ that calculates the optimal long-run value of the action p_{it} given the current state $s_t = (p_{i,t-1}, p_{j,t-1}) \in S$ for firm $i, j \in \{1, 2\}$ and $i \neq j$. The firms both have their own Q-matrix, for which it holds that $Q : |P| \times |S| \to \mathbb{R}$, assuming a discrete state set. Each time t, the agent selects a price p_{it}, observes a profit $\pi_i(p_{it}, p_{jt})$ and enters a new state s_{t+1}, the entry $Q_i(p_{it}, s_t)$ is updated accordingly

$$Q_i(p_{it}, s_t) \leftarrow Q_i(p_{it}, s_t) + \alpha \left[\pi_i(p_{it}, p_{jt}) + \gamma \max_p Q(p, s_{t+1}) - Q_i(p_{it}, s_t) \right] \tag{4}$$

where $\alpha \in (0, 1)$ is a step size parameter that reflects the learning rate of the algorithm. In order to make sure that all actions are tried in all states to approximate the true Q-matrix, the algorithm has to keep experimenting with new actions. The Q-learning must balance exploration and exploitation. Maintaining exploration is accounted for by using ϵ-greedy exploration. With probability $\epsilon_t \in [0, 1]$ it selects a random action (price) among all actions, and with probability $1 - \epsilon_t$ it exploits, meaning that it chooses the action that is the best reply in that state based on the Q-table.

$$p_{it} = \begin{cases} \sim U\{P\} & \text{with probability } \epsilon_t \\ = \arg\max_p Q_i(p, s_t) & \text{with probability } 1 - \epsilon_t \end{cases} \tag{5}$$

here $U\{P\}$ is over action set P a discrete uniform distribution. $\epsilon_t = (1 - \theta)^t$, where $\theta > 0$ is the decay parameter. ϵ_t is set such that initially the algorithm chooses in a random fashion, but over time it will make the greedy choice more. To ensure this θ is set such that the probability of exploration decreases from 100% at the beginning to 0.1% in the middle of the run, and 0.00001% at the end. In all simulations the number of learning periods, $T = 300,000$, θ is set to 4.605065×10^{-5}. A pseudocode of the algorithm as used in the baseline simulation is provided in Table 1.

3.3. *Performance Assessment*

After the simulations have been conducted, the following two things will be tested or calculated. There will be tested for the reward-punishment effect and the final profitability will be calculated.

3.3.1. *Reward-Punishment Scheme*

The observation of supra-competitive prices is not a genuine proof of collusion. Harrington (2018) mentions that for economists, the existence of collusion involves the existence of a reward-punishment scheme, which provides the incentives for firms to price above the competitive level. In general extensive form games, the existence of such a scheme can be tested

Algorithm 1: Pseudocode Asymmetric Q-Learning (Baseline Simulation)

```
Set demand and learning parameters;
Set frequency firm 1 and 2;
Initialise {p_1, p_2} = 0 and {π_1(p_1, p_2), π_2(p_1, p_2)} = 0 for t = 0;
for Every period do
    for Every firm do
        if Firm is allowed to do action then
            Choose action p_it according to policy function (5)
        else
            Action p_it = p_{i,t-1}
        end
    end
    if Firm was allowed to do action then
        Calculate profit π_i(p_1t, p_2t) based on (1);
        Update Q_i(p_it, s_t) according to (4)
    end
end
```

by the one-shot-deviation principle described by Hendon et al. (1996). The following result holds for repeated games: "*For a given combination of strategies of the opponents, a player's strategy is optimal from any stage of the game if and only if there is no stage of the game from which the player can gain by changing his strategy there, keeping it fixed at all other stages*". The existence of this property helps to show whether a combination of strategies in a repeated game constitutes a (subgame) perfect equilibrium. For this can be tested by letting one firm near the end of the simulation deviate from its strategy by forcing it to choose the price $\frac{1}{6}$, and observing the following responses. In the presence of the reward-punishment scheme, the deviating firm will first face a short period of higher profit, which is followed by a downward price spiral. Ultimately, this will lead to a net profit loss for the deviating firm. The deviating firm is thus punished for deviating by losing profit. When this effect is present, it ensures that the supra-competitive results may be obtained in equilibrium. Furthermore, this way of testing for equilibrium results is similarly used by Calvano et al. (2020b) and Klein (2019). The outcome of a run is considered collusive when the reward-punishment effect is present, and profitability is above the competitive profit benchmark of 0.0611. This benchmark is not trivial and is taken from Klein (2019) and more details can be found in his paper.

3.3.2. *Final Profitability*

The final profitability will be evaluated by calculating the average profitability of a firm of the final 1,000 periods of the 300,000 in total. This is done because the profits of the firms can be dynamic and therefore fluctuate. Taking the average over the final 1,000 periods will yield a stable value that can be evaluated.

$$\Pi_i = \frac{1}{1000} \sum_{i=T-1000}^{T} \pi_i(p_{it}, p_{jt}) \tag{6}$$

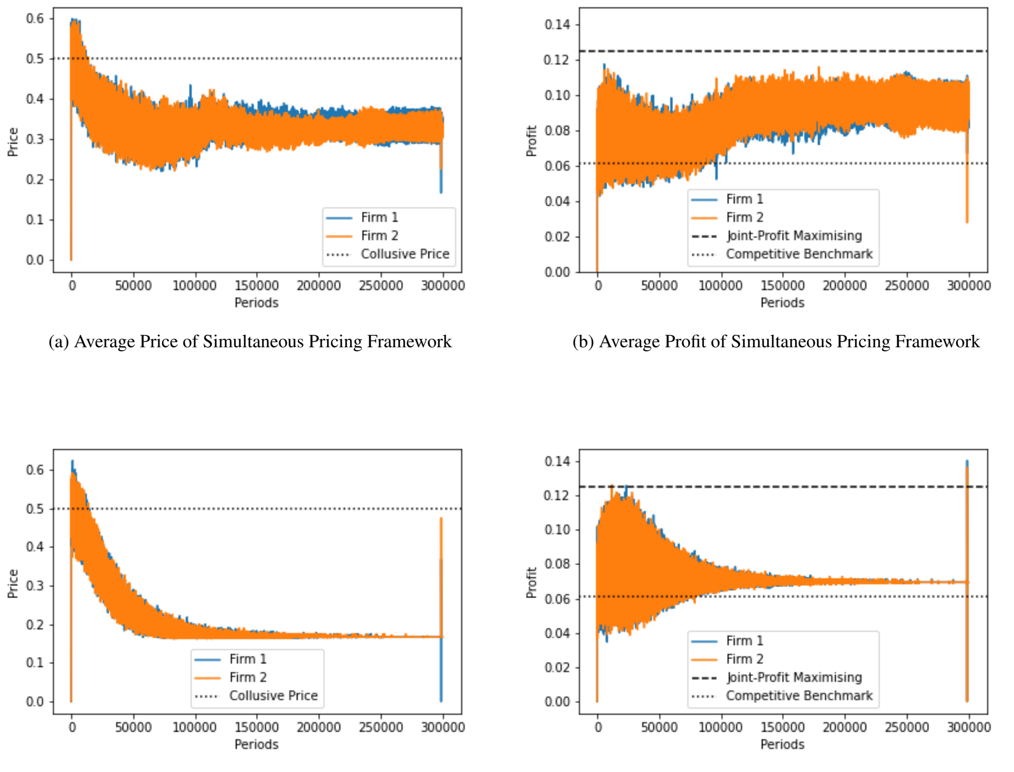
The profitability will be compared with the joint-profit maximising value of 0.125 (when both firms set their price equal to 0.5) and the non-trivial competitive benchmark of 0.0611.

4. RESULTS

This section is divided into three subsections. First, the basic results of the previous literature were replicated with the simultaneous and sequential price-setting framework and will be reviewed. Secondly, the results of the heterogeneous price-setting will be discussed. The third section contains more in-depth analyses of variations in the firms' frequencies. All simulations were carried out with the following settings. The amount of learning periods is $T = 300,000$. Furthermore, the discrete set of price choices consists of $k = 6$ intervals, resulting in seven different prices. The learning parameter α is set to $\alpha = 0.3$. This parameter determines to what extent the algorithm lets new information override old information. When α is set close to 0, the learning will be slow. When α is too close to 1, the agents do not attain any value to prior knowledge. The discount factor γ determines the importance of future rewards. When γ is close to 0, the agent will only consider immediate rewards, while a factor close to 1 makes the agent strive for a long-term high reward. There is chosen for $\gamma = 0.95$, because periods are generally quite small. Finally, the Q-values were initiated with all zeros, as well as the starting prices.

4.1. *Simultaneous and Sequential Simulation*

The basic results of the existing literature on the simultaneous and sequential pricing framework were replicated. This is to show the similar performance of this model and to compare the results with the asymmetric setting. For the simultaneous and sequential context, the average prices and profits are plotted the 300,000 learning periods. The averages were calculated out of 100 runs per game. For the simultaneous price-setting environment, both firms changed their price every period. The prices were chosen from the discrete set $P = \{0, \frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6}, \frac{5}{6}, 1\}$. For the sequential price-setting framework, firm 1 started by setting the first price and changed the price every even numbered period. Firms 2, on the other hand, changed its price every uneven period. Figure 1 shows that when the two Q-learning algorithms face each other, the average profits converge to a level above the competitive benchmark at 0.0611 but below the joint-profit maximising level at 0.125. This finding holds for both the simultaneous and sequential frameworks. Additionally, the algorithms, on average, converged to a price below the collusive price $p^C = 0.5$. In both scenarios, the pricing and profit patterns of firms 1 and 2 seem quite similar; there is a great overlap between the blue and the orange area.



(a) Average Price of Simultaneous Pricing Framework

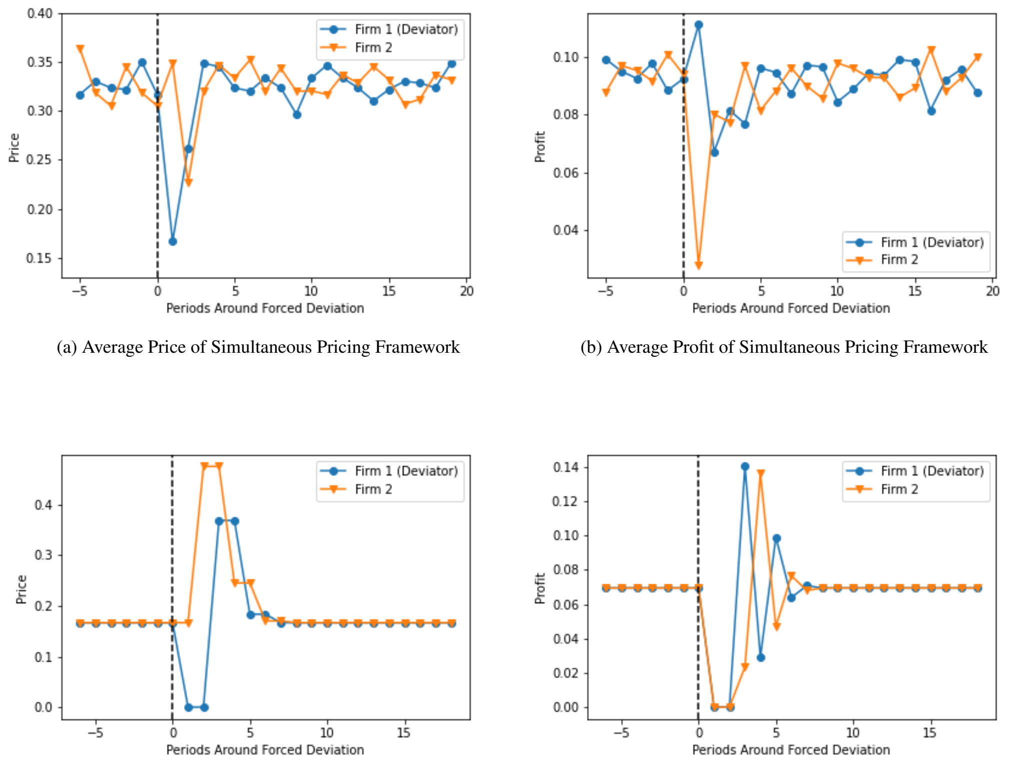
(b) Average Profit of Simultaneous Pricing Framework

(c) Average Price of Sequential Pricing Framework

(d) Average Profit of Sequential Pricing Framework

FIGURE 1.—Average Prices and Profits in 100 Runs

A deviation was forced at $T = 299,000$ for firm 1 to $p_1 = \frac{1}{6}$, to test for the one-deviation property. For the sequential context, the deviation was forced to $p_1 = 0$, because the average price level was already around $\frac{1}{6}$. These deviations are shown in Figure 2. The initialising time of deviation is modelled in the figures by period 0. Both strategies seemed to possess the reward-punishment effect: the deviation of firm 1 triggered a downward price spiral that led to a net profit loss for firm 1 after an initial one-period profit gain. This punishment is temporary since, in the subsequent periods, the firms returned to the original profit level.



(a) Average Price of Simultaneous Pricing Framework

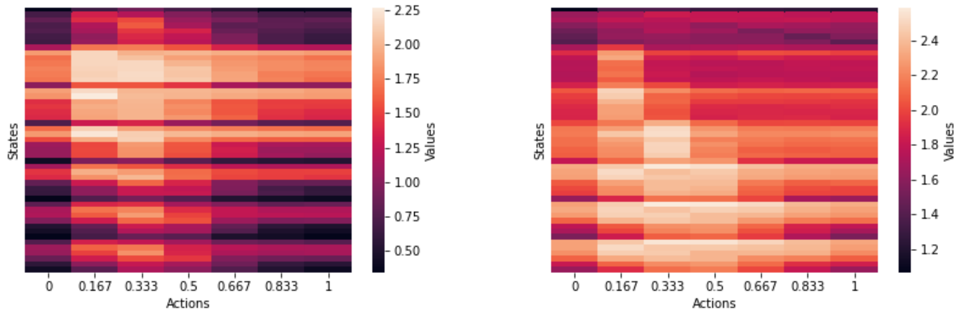
(b) Average Profit of Simultaneous Pricing Framework

(c) Average Price of Sequential Pricing Framework

(d) Average Profit of Sequential Pricing Framework

FIGURE 2.—Average Prices and Profits After Forced Deviation in 100 Runs

Lastly, the final profitability for both firms was determined for all of the 100 runs. For the simultaneous and sequential cases, this was for firm 1 respectively 41 and 59 times. Firm 2 had the highest final profitability respectively in 18 and 23 runs. In the simultaneous setting in almost half of the runs, 41 times, both firms had the same final profitability. All of these were above the competitive benchmark of 0.0611. In 18 runs, the firms had equal profitability in the sequential framework. Again, all of them were above the competitive benchmark.



(a) Average Values Q-Table Firm 1

(b) Average Values Q-Table Firm 2

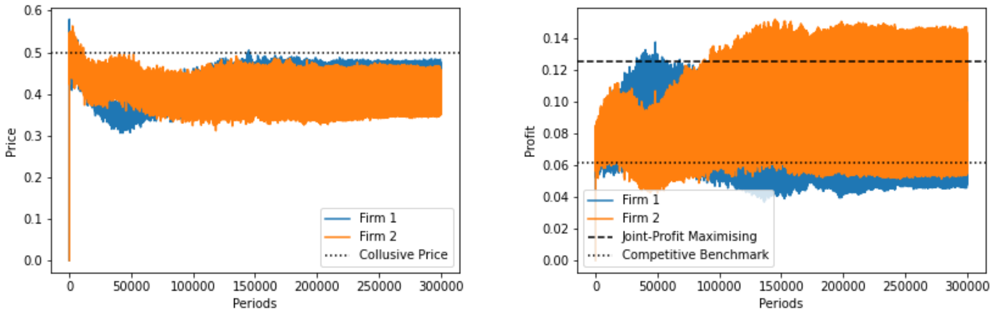
FIGURE 3.—Average Q-Tables Out of 300 Runs

4.2. *Baseline Simulation*

For the baseline simulation, the game was played 300 times. Firm 1 could adjust its price every 5 periods, while firm 2 could do so every 2. A heatmap for firm 1 and firm 2 of the average Q-tables in the 300 runs is displayed in Figure 3. Both heat maps are sufficiently filled, which shows thorough exploration. On the x-axis the actions are shown and on the y-axis the states are found, from top to bottom $(0,0), (0, \frac{1}{6}), ..., (1, \frac{5}{6}), (1,1)$. The first price is the price chosen by firm 1 and the second price by firm 2.

At the end of the 300 runs, the average prices and profits out of the 300,000 periods were calculated for all runs. This is displayed in Figure 4, which shows the average price and profit for every period t for both firms. In Figure 4a it is shown that on average firms set their price below the collusive price $p^C = 0.5$. When looking at the average profit in 4b, there is a wider variance in the resulting profits compared to the simultaneous and sequential framework. Nevertheless, the majority of the profits are above the competitive benchmark. When comparing Figure 4a with Figure 1a and 1c, the average level of prices was higher when asymmetric price frequencies were adopted compared to the context of equal pricing frequencies.

For both firms, the final profitability was calculated in every run and afterwards the final profitability was compared. It was determined which of the firms had the highest final profitability of that run. Figure 5 shows the number of times each firm had the highest final profitability out of all the 300 runs. Firm 2 had 209 times the highest profitability, and firm 1 did 91 times. The firms never ended up having the same average price.



(a) Average Price of Baseline Simulation

(b) Average Profit of Baseline Simulation

FIGURE 4.—Average Prices and Profits in 300 Runs

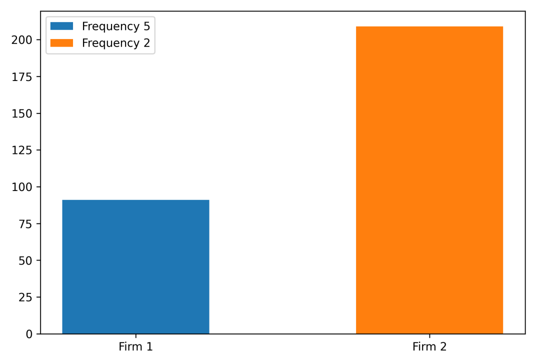
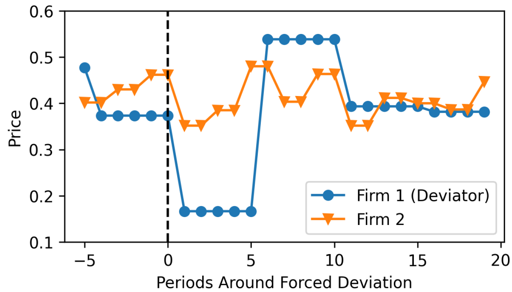


FIGURE 5.—Highest Final Profitability Out of 300 Runs

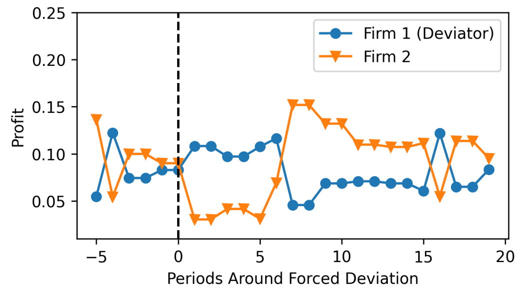
Finally, a deviation was forced in every run to test for the reward-punishment effect at $T = 299,000$. In Figure 6 the subsequent responses to a forced deviation of firm 1 are shown. Again the reward-punishment pattern can be deducted from the figure, however, this time less obvious. When firm 1 deviated, it got an initial boost in its profit, which resulted afterwards in a lower net profit (except in period 16).

4.3. *Varying Frequency Simulation*

The simulation was run again after two adjustments were made in the pricing frequency. First, the frequencies of both firms were changed to stochastic frequencies. Secondly, the frequency of firm 1 was modified multiple times to examine the effect of the degree of frequency heterogeneity on the final profitability.

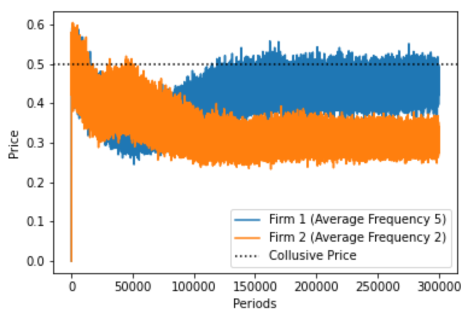


(a) Average Price of Baseline Simulation

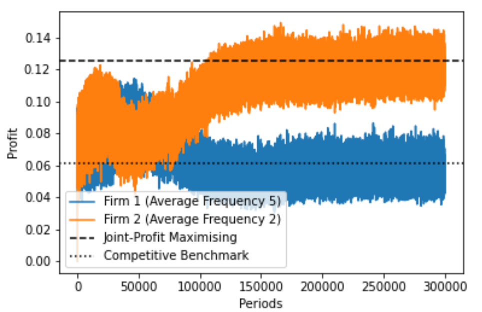


(b) Average Profit of Baseline Simulation

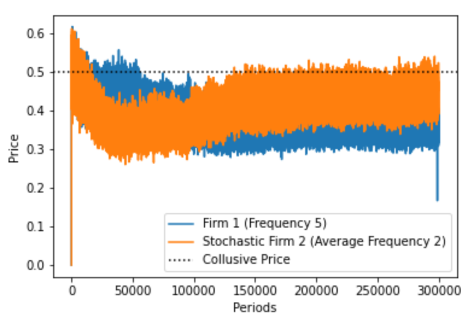
FIGURE 6.—Average Prices and Profits After Forced Deviation in 300 Runs



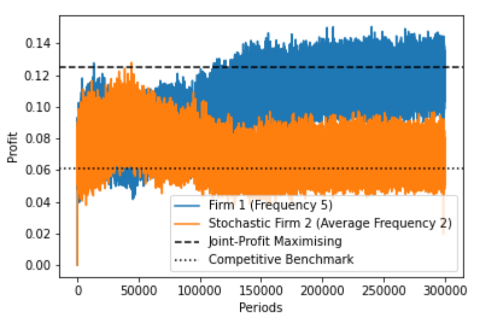
(a) Average Price of Two Stochastic Firms



(b) Average Profit of Two Stochastic Firms



(c) Average Price with One Stochastic Firm



(d) Average Profit with One Stochastic Firm

FIGURE 7.—Average Prices and Profits in 100 Runs

4.3.1. *Stochastic Frequencies*

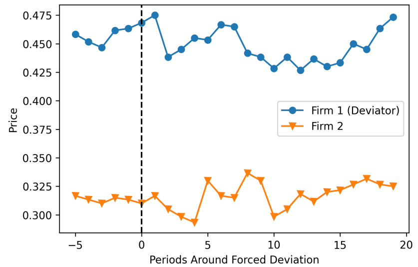
In the real world, firms may not change their price with a fixed frequency scheme. Incorporating some randomness in the price updating scheme could influence the learning of the algorithms, which could lead to different outcomes. Therefore, the firms played the same game, this time with a stochastic frequency of 5 (firm 1) and 2 (firm 2). Secondly, the game was played again when only one firm had a stochastic average frequency of again 2 (firm 2).

Two Stochastic Frequencies The baseline model was performed again after changing the frequencies to stochastic ones. Again 100 runs were performed, each consisting of 300,000 periods. Again the average prices and profits are plotted in Figure 7a and Figure 7b as well as the average prices and profits after the forced deviation in 8a and Figure 8b. Again, comparing Figure 7a with Figure 1a and Figure 1c, it seems that the prices of the stochastic asymmetric firms were higher compared to the prices set in the simultaneous and sequential setting. Furthermore, firm 2 (with higher frequency) set a lower average price than firm 1. At last, firm 1 had in 8 of the runs the highest final profitability, whereas firm 2 did in 75 cases. In 17 runs, they ended up having an equal final profitability.

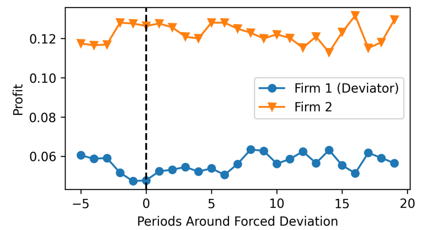
One Stochastic Frequency The same simulation was conducted again, this time, with only one firm having a stochastic frequency. Firm 1 had a fixed frequency of 5, and firm 2 had stochastic one of 2. Figure 7c and Figure 7d display the results of the average prices and profits. Again almost all prices were below the collusive price. However, the majority of average profits were above the competitive benchmark. When looking at the deviation in Figure 8c and Figure 8d, there is no such thing as a reward-punishment scheme detectable. After the deviation by firm 1, the net profits of the deviating firm do not go down, thus there is no punishment. The average prices displayed in Figure 7c indicate again that higher prices arise in the asymmetric stochastic firm setting compared to the setting of identical frequencies. Additionally, this time it seems that the prices of the firm with the higher frequency were higher than those of the firm with the lower frequency. Finally, in 89 of the runs, firm 1 had the highest final profitability and 11 times, firm 2 did.

4.3.2. *Changing Frequency Differences*

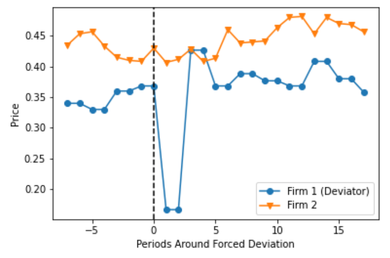
The simulation of 100 runs ran again, but for every simulation, the frequency of firm 1 was changed. The frequencies of firm 1 were chosen from the set: $\{2, 3, 5, 7, 9, 11, 13, 15, 17\}$, while the frequency of firm 2 was kept at 2. After every simulation, the final profitability for both firms was calculated. The final profitability is plotted in Figure 9. When the firms have the same frequency, the average final profitability is about equal. However, when increasing the frequency difference between both firms, the firms' average final profitability diverged. When the frequency difference between the two firms increased, as well did the profitability



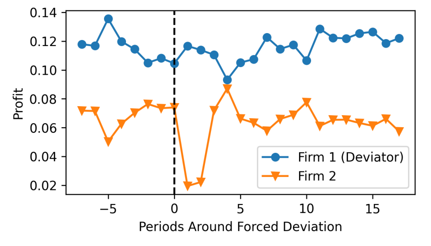
(a) Average Price of Two Stochastic Firms



(b) Average Profit of Two Stochastic Firms



(c) Average Price with One Stochastic Firm



(d) Average Profit with One Stochastic Firm

FIGURE 8.—Average Prices and Profits After Forced Deviation in 100 Runs

difference. There seemed to be an advantage in having a higher frequency when looking at profitability.

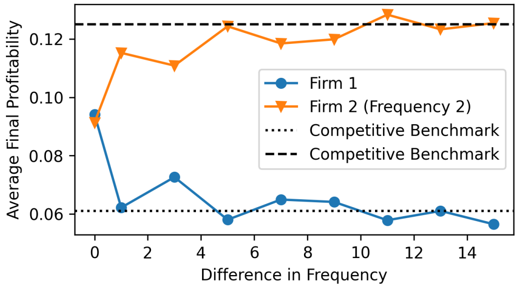


FIGURE 9.—Average Final Profitability for Changing Frequency Differences

5. DISCUSSION

The most important finding of this thesis is that, on average supra-competitive outcomes seem to arise in the asymmetric frequency setting. However, in the baseline simulation, none of the firm's final profitability is equal. There is always one firm 'winning' the simulation game, the firms do not end up with a joint-maximising outcome. Moreover, the outcomes can result in equilibrium due to the reward-punishment effect. Besides, this effect also shows that the algorithms do not fail to optimise. Additionally, it seems that the firm with the higher frequency is favoured when considering the final profitability.

Concerning the stochastic frequencies, the firms seem to diverge more in their strategy, one firm chooses on average higher prices than the other. This divergence in strategy is probably due to the randomness incorporated in the pricing moments, making it more complicated for the other firm to learn an optimal strategy, especially for the firm with the lower frequency. These outcomes can also appear in equilibrium due to the one-deviation property.

Overall, there is a clear contrast when comparing the asymmetric frequency results with symmetric ones. In the simultaneous and sequential scenarios, the algorithms seem to learn an optimal strategy more efficiently and converge to a fixed point in the grid. Also, the learned strategies of the firms show more similarities. Additionally, the prices of the asymmetric pricing setting seem to be higher compared to the symmetric case. This dissimilarity of the symmetric and asymmetric contexts shows the importance of accounting for frequency differences.

The results of the simultaneous and sequential pricing framework are in line with the existing literature. Waltman and Kaymak (2008), Calvano et al. (2020b), Abada and Lambin (2020), and Hansen et al. (2021) report the anti-competitive prices charged by the algorithms in the simultaneous setting, which in some cases are sustained by collusive strategies shown by the one-deviation property. Klein (2019) focuses on the sequential structure and mentions that the reinforcement learning algorithms learn to converge to collusive equilibria as the most important finding. These equilibria are on average supra-competitive, although below the joint-profit maximising level, similarly to the results of this thesis. Furthermore, Brown and MacKay (2021) show that asymmetry in pricing technology shifts the equilibrium behaviour. When one firm has a higher price-setting frequency, all firms can obtain higher prices. This universal price increase is reflected in the baseline simulation results and the stochastic frequency context. In all scenarios, the average prices are higher than when firms have symmetric pricing technologies. Another finding by Brown and MacKay (2021) is that they theoretically show that asymmetric pricing technologies are associated with asymmetric prices. Their study shows that firms with the faster pricing technology have lower prices than the inferior frequency firm in the duopoly case. Similar price patterns are reflected in the results of this thesis. The firm's prices with the superior frequency are lower for both the baseline simulation and the case when both firms adopt stochastic pricing frequencies. Interestingly, when only one firm adopts a stochastic pricing technology, the firm with the weaker pricing technology ends up in the lower price range. This deviant result could be because firm 2 being the stochastic one with an average frequency of 2. Therefore, there could be multiple price changes in a row in some periods, having a bigger chance of meeting the other firm in the same period and changing the price simultaneously. When this happens, it can make the model more similar to the simultaneous case, for which the theory on price dispersion by Brown and MacKay (2021) does not hold anymore.

There are some limitations to this study worth noting. First, there is chosen for a very stylised environment in which the simulations have been conducted. For example, there are only two firms, they both use the same Q-learning algorithm, and demand is deterministic. It is unclear to what extent these results are applicable in practice. However, since the research on this topic is relatively new, these simplifications and assumptions are also necessary to derive meaningful conclusions from the results and to be able to compare them to the existing literature. Likewise, there are some limitations to the Q-learning algorithm. The learning is relatively slow, and expanding the analysis to more challenging environments could take even more time. Besides, the speed influenced the number of runs conducted for every simulation. There are 300 runs for the baseline simulation and 100 runs for the others. Taking averages of more runs would give more precise estimates. Besides this, in reality, firms may not even use specifically the Q-learning algorithm used in these simulations but more expanded versions of this algorithm (Deep- or Double Q-learning). Or even a completely different algorithm. Although these are limitations, the reinforcement learning algorithm developed in this thesis makes it easier to make adjustments and interpret the results. In addition, Q-learning and variations there of have been reported as widely used machine learning algorithms. Finally, the reward-punishment scheme was not present in all simulations. In these cases, it is not assured that the results will appear in equilibrium and, therefore, would not be relevant. This could be solved by enhancing the algorithm or allowing for more extended learning periods.

More research is necessary to be able to develop fitting policy implications. Several issues stand out. The highly stylised economic environment is the first one. It is interesting to create a market with stochastic demand, with the presence of demand and cost shocks. Alternatively, firms can modify their pricing frequency during the game for a specified cost. These changes can intervene with the learning process of the algorithms, making it more demanding for the firms to work together. In addition, a market with more firms will reflect reality better than a duopoly model. Another critical issue is that the algorithms start competing without any prior knowledge. The question is if the algorithms are always this naive before they are used by corporations. Developing a simulation with algorithms that incorporate some economic structure in the model instead of a model-free environment could make learning collusive strategies for the algorithms easier. A possibility would be to train the model before letting it compete in a simulated market. Additionally, this thesis shows the importance of accounting for frequency differences, which requires more research. Incorporating frequency heterogeneity in more models could give further knowledge that can be used to develop appropriate competition policies. Finally, the market in this thesis is viewed as quite general. There is an enormous scope for future research that incorporates specific features of specific markets in the model—for example, distinguishing between online and offline markets.

Based on the current knowledge, some potential policy implications would be the following. There is primarily simulation-based theory on this topic, and a better empirical understanding of algorithmic collusion is indispensable. Extensive market research by authorities would make them better acquainted with this topic. For example, what algorithms are used precisely, in what markets, and on what information condition the algorithms their learning. In addition, the outcomes of this thesis show that when asymmetric frequency algorithms are present, both firms raise their price level compared to the symmetric situation. In order to prevent such price increases, policymakers would have to limit the conditioning on competitors' prices by the algorithms. This could be done by enforcing some laws that competitors' prices are not allowed to be incorporated into the algorithm. Moreover, imposing boundaries on web scraping or storing the competitors' recent prices could also be an option. Not only restricting the conditioning on competitors' prices is something that could be constrained, but in general, all information that goes into the algorithm could be controlled. Besides, the design of the market can be adjusted to ensure collusion does not occur, for example, by introducing another algorithm that maximises consumer or social welfare. Finally, collusion is considered by law when there is communication between competitors to raise prices. However, such contact is not present when algorithmic collusion emerges. Therefore, the competition laws should be changed to add the appearance of algorithmic collusion to the law.

6. CONCLUSION

In conclusion, to answer the research question, to what extent the asymmetric pricing frequency Q-learning algorithms can collude. The results show that the average profits are above the competitive benchmark when accounting for both stochastic and deterministic asymmetric frequencies. Furthermore, the average prices are higher than those in the simultaneous and sequential price-setting framework. However, it is not ensured that these outcomes occur in equilibrium for all events. The findings show that it is vital to account for asymmetric frequencies and that the simultaneous and sequential price-setting frameworks are just a glimpse of what the real world looks like, which answers this thesis' subquestion. The problem of algorithmic collusion is more complicated than it seems in most experimental papers, and further research is necessary to develop the appropriate policies against collusion in the market.

REFERENCES

ABADA, IBRAHIM AND XAVIER LAMBIN (2020): "Artificial intelligence: Can seemingly collusive outcomes be avoided?" *Working Paper, SSRN 3559308*. [4, 8, 19]

ASSAD, STEPHANIE, ROBERT CLARK, DANIEL ERSHOV, AND LEI XU (2020): "Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market," *CESifo Working Paper No. 8521*. [2, 4]

AUTORITÉ DE LA CONCURRENCE AND BUNDESKARTELLAMT (2019): "Algorithms and Competition," . [2]

BROWN, ZACH Y AND ALEXANDER MACKAY (2021): "Competition in Pricing Algorithms," *Harvard Business School Working Paper No. 20-067*. [2, 3, 5, 6, 19, 20]

CALVANO, EMILIO, GIACOMO CALZOLARI, VINCENZO DENICOLÒ, JOSEPH E. HARRINGTON, AND SERGIO PASTORELLO (2020a): "Protecting Consumers from Collusive Prices Due to AI," *Science*, 370, 1040–1042. [2]

CALVANO, EMILIO, GIACOMO CALZOLARI, VINCENZO DENICOLÒ, AND SERGIO PASTORELLO (2020b): "Artificial Intelligence, Algorithmic Pricing, and Collusion," *American Economic Review*, 110, 3267–3297. [4, 8, 10, 19]

CHEN, LE, ALAN MISLOVE, AND CHRISTO WILSON (2016): "An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace," in *Proceedings of the25th International World Wide Web Conference*, 1339–1349. [3]

COMPETITION BUREAU (2018): "Big data and Innovation: Implications for Competition Policy in Canada," *Government of Canada*. [2]

HANSEN, KARSTEN T, KANISHKA MISRA, AND MALLESH M PAI (2021): "Algorithmic Collusion: Supra-Competitive Prices via Independent Algorithms," . [4, 19]

HARRINGTON, JOSEPH E. (2018): "Developing Competition Law for Collusion by Autonomous Artificial Agents," *Journal of Competition Law and Economics*, 14, 331–363. [2, 9]

HENDON, EBBE, HANS JØRGEN JACOBSEN, AND BIRGITTE SLOTH (1996): "The One-Shot-Deviation Principle for Sequential Rationality," *Games and Economic Behavior*, 12, 274–282. [10]

JOHNSON, JUSTIN P, ANDREW RHODES, AND MATTHIJS WILDENBEEST (2020): "Platform Design When Sellers Use Pricing Algorithms," *Working Paper, SSRN 3691621*. [8]

KLEIN, TIMO (2019): "Autonomous Algorithmic Collusion: Q-Learning Under Sequential Pricing," *RAND Journal of Economics*. [2, 4, 5, 6, 8, 10, 19]

KLENOW, PETER J AND BENJAMIN A MALIN (2010): "Microeconomic Evidence on Price-Setting," *Handbook of Monetary Economics*, 3, 231–284. [5]

MASKIN, ERIC AND JEAN TIROLE (1988): "A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles," *Econometrica*, 56, 571–599. [5, 6]

OECD (2017): "Algorithms and Collusion: Competition Policy in the Digital Age," *Report*. [2]

SALCEDO, BRUNO (2015): "Pricing Algorithms and Tacit Collusion," *PhD Manuscript, Pennsylvania State University*. [5]

SCHECHNER, SAM (2017): "Why Do Gas Station Prices Constantly Change? Blame the Algorithm," *Wall Street Journal*. [2]

WALTMAN, LUDO AND UZAY KAYMAK (2008): "Q-learning Agents in a Cournot Oligopoly Model," *Journal of Economic Dynamics and Control*, 32, 3275–3293. [4, 8, 19]

WATKINS, CHRISTOPHER J. C. H. (1989): "Learning from Delayed Rewards," *PhD Manuscript, King's College London*. [8]