No Regret Learning in Oligopolies: Cournot vs Bertrand

Uri Nadav* Georgios Piliouras†

Abstract

Cournot and Bertrand oligopolies constitute the two most prevalent models of firm competition. The analysis of Nash equilibria in each model reveals a unique prediction about the stable state of the system. Quite alarmingly, despite the similarities of the two models, their projections expose a stark dichotomy. Under the Cournot model, where firms compete by strategically managing their output quantity, firms enjoy positive profits as the resulting market prices exceed that of the marginal costs. On the contrary, the Bertrand model, in which firms compete on price, predicts that a duopoly is enough to push prices down to the marginal cost level. This suggestion that duopoly will result in perfect competition, is commonly referred to in the economics literature as the "Bertrand paradox".

In this paper, we move away from the safe haven of Nash equilibria as we analyze these models in disequilibrium under minimal behavioral hypotheses. Specifically, we assume that firms adapt their strategies over time, so that in hindsight their average payoffs are not exceeded by any single deviating strategy. Given this no-regret guarantee, we show that in the case of Cournot oligopolies, the unique Nash equilibrium fully captures the emergent behavior. Notably, we prove that under natural assumptions the daily market characteristics converge to the unique Nash. In contrast, in the case of Bertrand oligopolies, a wide range of positive average payoff profiles can be sustained. Hence, under the assumption that firms have no-regret the Bertrand paradox is resolved and both models arrive to the same conclusion that increased competition is necessary in order to achieve perfect pricing.

[1]Department of Computer Science, Stanford University, Stanford CA 94305. E-mail: urinadav@cs.stanford.edu.

[2]Department of Computer Science, Cornell University, Ithaca NY 14853. Supported by NSF grant CCF-0325453. E-mail: gpil@cs.cornell.edu.

1 Introduction

Oligopoly theory deals with the fundamental economic problem of competition between two or more firms. In this work we study the conditions under which an oligopoly arrives at stability. We focus on the two most notable models in oligopoly theory: Cournot oligopoly[7], and Bertrand oligopoly[5]. In the Cournot model, firms control their production level, which influences the *market price*. In the Bertrand model, firms choose the price to charge for a unit of product, which affects the *market demand*.

Competition among firms in an oligopolistic market is a setting of strategic interaction, and is therefore analyzed within a game theoretic framework. Cournot and Bertrand oligopolies are modeled as strategic games, with continuous action sets (either production levels or prices). In both models the revenues of a firm are the product of the firm's part of the market times the price; In addition, a firm incurs a production cost, which depends on its production level.

In the most simple oligopoly model, the firms play a single game, where they all take actions simultaneously. All the firms produce the same good; the demand for this product is a linear in the total production; the cost of production is fixed per unit of production. In this oligopolistic market, a Nash equilibrium in pure strategies exists in both Cournot and Bertrand models. Interestingly, despite the strong similarity between these models, the Nash equilibrium points are very different: in Bertrand oligopoly, Nash equilibrium drives prices to their competitive levels, that is, the price equals the cost of production, while in Cournot oligopoly, the price in the unique Nash equilibrium is strictly above its competitive level. Liu [14] showed that the uniqueness of equilibrium in the linear demand, linear cost model, carries on to correlated equilibrium. Yi [23] have extended Liu's work to the case of Cournot oligopoly where firms produce different products, that are strategic substitutes, and to the case of weakly convex production cost functions.

Equilibrium analysis alone, however, cannot capture the dynamic nature of markets. In the real world, trading is performed over long periods of time, which gives firms the chance to adjust their actions e.g, their prices or production levels. If we assume that the essential market attributes remain unchanged, then this situation gives rise to a repeated game, obtained by repeated play of the original simultaneous, one shot game.

One approach for analyzing the repeated oligopolistic game, is through studying the Nash equilibrium of the repeated game. This models a situation where the firms "commit" to a strategy, and their joint commitment forms an equilibrium (see [15], Chapter 12.D). In practice however, an important feature of an oligopolistic market is that different firms are not perfectly informed about different aspects of the market, e.g., the attributes of the other participants, and cannot pre-compute, or agree on a Nash equilibrium of the repeated game before they begin interacting.

A more pragmatic approach for studying such repeated interactions is through the analysis of adaptive behavior dynamics (see, [12, 24]). The goal here is to investigate the evolution of the repeated game, when the agents (firms) play in accordance to some "natural" rule of behavior. In the setting of an oligopolistic market, we would want a natural behavior to comply with "rationality" and hopefully give rise to some sort of profit maximization on the side of a firm. Another natural aspiration is that our behavior rules should be "distributed", which means that firms should be able to make their choices in each period based only on their own payoffs, and independently of other firms (in most markets, a firm cannot tell with certainty what are the payoffs, and costs of other firms). The central question, in such a setting, is whether the behavior dynamics finally converge, as this would imply long term stability of the market.

Dynamic behavior in Cournot and Bertrand oligopolies have been studied before. Cournot [7] considered the simple best response dynamics, where at every step of the repeated game, firms react to what happened in the market on the previous step. Cournot showed that in the case of a duopoly, the simultaneous best response dynamics converges to the unique Nash equilibrium of the one shot game, i.e., after sufficiently many steps the two firms will play their Nash equilibrium strategies on every subsequent step. However, this result does not generalize to an arbitrary number of firms, as shown by Theocharis [20].

Milgrom and Roberts [16, 17] were the first to explore connections between Cournot competition and super-modular games as a way to show convergence results for learning dynamics. In their work (as well as in followup papers [2, 21]) Cournot duopolies as well as specific models of Cournot oligopolies are shown to exhibit strategic complementarities. This identifying property of supermodular games is shown to imply convergence to Nash equilibrium for a specific class of learning dynamics, known as adaptive choice behavior. This class of learning dynamics encompasses best-response dynamics and Bayesian learning but is generally orthogonal to the class of dynamics that we will be focusing on.

In this paper we are interested in dynamic behaviors where firms minimize their long term *regret*[1]. Regret compares the firm's average utility to that of the best fixed constant action (e.g., constant production level in Cournot, and constant price in Bertrand). Having no-regret means that no deviating action would significantly improve the firm's utility (see [6]). Several learning algorithms [25], [13],are known to offer such guarantees, as their average regret bounds are $o(T)$, where T is the number of time steps.

Regret minimization procedures prescribe to some rather desirable requirements in regards to modeling market behavior. Firstly, they are rational, in the sense that an agent is given guarantees on her own utility regardless of how the other agents act. Moreover, they are distributed, since an agent needs to be aware only of her own utility. Many of the no regret procedures[11] are rather intuitive, as they share the idea that agents increase the probability of choosing actions that have been performing well in the past. Several learning procedures are known to be of no-regret, but more importantly, the assumption is not tied to any specific algorithmic procedure, but merely captures successful long-term behavior. Lastly, no-regret guarantees can be achieved even in the "multi-armed bandit" setting[3, 10, 1], where the input for the algorithm consists only of the payoffs received. This feature is important in the case that firms are not fully aware of the market structure (i.e., demand function), and are maybe even uncertain regarding their own production costs.

In the most relevant result to our work, Even-Dar *et al.*[9] study no regret dynamics in a class of games that includes Cournot competition with linear inverse market demand function, and convex costs functions. They show that the *average* production level of every firm, as well as it's *average* profits, converge to the ones in the unique Nash equilibrium of the one shot game.

1.1 Our results

In this work we examine the behavior of no regret dynamics in Cournot and Bertrand oligopoly models.

In the classic model of Bertrand oligopoly [5], it is well known that oligopolies with more than two firms exhibit several trivial Nash equilibria but in all of them the prices are equal to the marginal costs and all players make zero profit (Bertrand paradox). This phenomenon has been

[1] Regret is sometimes also referred to as external regret.

	Bertrand	Cournot with perfect substitutes	Cournot with product differentiation
Nash equilibria	Infinite, Unique prices, Unique profits	Unique	Unique
Correlated equilibria	**Infinite**[2]	Unique	Unique
No Regret	**Infinite, Different prices, Different profits**	**Infinite, Unique prices, Different profits**	**Unique**

Table 1: Overview of results

verified for correlated equilibria only for the case of a duopoly[2], where correlated equilibria are unique[16]. In our work, we show that under no-regret behavior the zero-profit postulate does not hold even in the case of two players. In fact, we show that not only does the market not necessarily converge to zero profit outcomes, but that the players can actually enjoy significant profits. In summary, our main results for Bertrand oligopolies under no-regret have as follows:

1. The Bertrand paradox does not hold anymore; firms enjoy non-zero profits under no-regret behavior.
2. Moreover, the identified profits can be rather significant when the number of players is small (e.g. 17% of optimal profits in the case of a duopoly). Profits however, tend to go to zero quickly as the number of firms increases.

Interestingly, our observations about no-regret behavior in Bertrand oligopolies agree to a large extent both with experimental work [8], as well as with empirical observations about real world oligopolistic markets [19].

The study of correlated equilibria [14, 23] as well as of no-regret dynamics in [9] in Cournot oligopolies, has been an area of interest in both economics as well as computer science. In our work, we analyze a model of Cournot equilibria, which is a strict generalization of all the previously examined models, under no-regret dynamics. In fact, our results can be extended to all dynamics, in which each player's average payoff dominates the one they would receive if they always deviated to their respective Nash equilibrium strategy. This is a strict generalization of no-regret dynamics, since no-regret dynamics must fare well against all fixed strategies. In a novel approach in this line of work, we consider the evolution of the market not only from the perspective of the firms (individual production levels, profits), but also from the consumers' perspective (aggregate production level, prices) which leads to new insights. As a result, we can prove a single unifying message for all models examined: the daily prices converge to their level at Nash equilibrium.

In summary, our main results for Cournot oligopolies under no-regret have as follows:

[2] In [22] it is claimed that the correlated equilibria of Bertrand games are unique under some special cases.

1. In Cournot oligopoly with linear inverse demand function, and weakly convex costs, when every firm experiences no-regret, the empirical distribution of the daily overall production level, as well as of the daily prices, converges to a single point that corresponds to the Nash equilibrium of the one shot game.

2. When the firms produce products that are not perfect substitutes i.e., when even the tiniest of *product differentiation* is introduced, the empirical distributions of all market characteristics including the daily production levels of every firm converge to their levels in Nash equilibrium.

3. Some product differentiation is necessary in order to alleviate the nondeterminism of the day-to-day behavior on the side of the firms.

Table 1 summarizes what is known about equilibrium, and no-regret in Cournot and Bertrand oligopolies.

2 Preliminaries

2.1 Models of Oligopoly

We formally define Cournot oligopoly, and Bertrand oligopoly, as strategic games, with continuous action space.

Definition 2.1. *A Cournot oligopoly is a game between n firms, where the strategy space S_i of firm i is the span of its production level q_i. Typically, S_i is defined to be the interval $[0, \infty)$. The utility function for firm i is $u_i(q_1, \ldots, q_n) = P_i(q_1, \ldots, q_n)q_i - c_i(q_i)$, where P_i is the market inverse demand function for the good of firm i, which maps the vector of production levels to a market clearing price in $\mathbb{R}^+$.*

Our focus is on the case of linear inverse demand function. The utility of a firm i as a function of the firms' production levels is $u_i(q_1, \ldots, q_n) = (a - bQ)q_i - c_i(q_i)$, where a and b are positive constants, and $Q = \sum_i q_i$ denotes the total product supply. In Section 4 we consider an extension of Cournot oligopoly with perfect substitutes, to the case of product differentiation, where the price of firm i depends in an asymmetric manner on his own production level, and the production levels of the other firms. In this case the market inverse demand function $P^i(q)$ is given by $P^i(q) = a_i - b_i q_i - b_i \gamma_i Q_{-i} = a_i - b_i(1-\gamma_i)q_i - b_i \gamma_i Q$, where γ_i denotes the degree of product differentiation between products, $0 < \gamma_i \leq 1$, $b_i > 0$.

Definition 2.2. *A Bertrand oligopoly is a strategic game between n firms, where the strategy space P_i of firm i is its declared price p_i, which lies in the interval of all possible prices $[0, \infty)$, and its utility function is*
$u_i(p_1, \ldots, p_n) = D_i(p_1, \ldots, p_n)p_i - c_i(D_i(p_1, \ldots, p_n))$, where D_i is the market demand function of firm i, that maps from the vector of firms prices to a demand in $\mathbb{R}^+$.

We consider Bertrand oligopoly with a linear demand function, in which the market demand is equally shared among the firms with the least price:

$$D_i(p_1, \ldots, p_n) = \begin{cases} 0 & p_i > p_j, \text{ for some } j \\ \frac{a-p_i}{b(m+1)} & p_i \leq p_j \text{ for all } j, \text{ and } m = |\{j \neq i | p_j = p_i\}| \end{cases}$$

Intuitively, this means that the market demand goes down linearly as the minimal announced price increases. If the minimal price has been offered by more than one firms, these firms share the market demand equally.

2.2 Regret Minimization

We will give a formal definition of having no-regret in an online sequential problem.

Definition 2.3. *An online sequential problem consists of a feasible set $F \in \mathbb{R}^m$, and an infinite sequence of functions $\{f^1, f^2 \ldots, \}$, where $f^t : \mathbb{R}^m \to \mathbb{R}$.*

At each time step t, an online algorithm selects a vector $x^t \in \mathbb{R}^m$. After the vector is selected, the algorithm receives f^t, and collects a payoff of $f^t(x^t)$. All decisions must be made *online*, in the sense that an algorithm does not know f^t before selecting x^t, i.e., at each time t, a (possibly randomized) algorithm can be thought of as a mapping from a history of functions up to time t, $f^1, \ldots, f^{t-1}$, to the set F.

Given an algorithm $\mathcal{A}$ and an online sequential problem $(F, \{f^1, f^2, \ldots\})$, if $\{x^1, x^2, \ldots\}$ are the vectors selected by $\mathcal{A}$, then the payoff of $\mathcal{A}$ until time T is $\sum_{t=1}^{T} f^t(x^t)$. The payoff of a static feasible vector $x \in F$, is $\sum_{t=1}^{T} f^t(x)$. Regret compares the performance of an algorithm with the best static action in hindsight:

Definition 2.4. *The* external regret *of algorithm $\mathcal{A}$, at time T is defined as*

$$\mathcal{R}(T) = \max_{x \in F} \sum_{t=1}^{T} f^t(x) - \sum_{t=1}^{T} f^t(x^t).$$

An algorithm is said to have no-external regret, if for every online sequential problem, its regret at time T is $o(T)$.

The regret of a firm in a repeated oligopoly game: Consider the case of n firms that engage in a repeated Cournot (alternatively Bertrand) oligopoly game, and suppose that $\{x^t\}_{t=1}^{\infty}$ is a sequence of vectors, where x^t represents the production levels (alternatively, prices), set by the firms at time t. The regret of firm i at time T is defined as $\mathcal{R}_i(T) = \max_{y \in S_i} \sum_{t=1}^{T} u_i(y, x_{-i}^t) - \sum_{t=1}^{T} \left(u_i(x^t)\right)$, where u_i is the utility function of firm i, and S_i is the strategy set of i.

3 Bertrand Oligopolies

We will be focusing on the case where are the all firms share the same linear cost function (i.e. $C_i(x) = cx$ for all i). The set of Nash equilibria of this game consists of all price vectors such that the prices of at least two firms are equal to c, whereas all others are greater than c. Although there exist multiple Nash equilibria, all of them imply the same market prices where the firms sell at marginal cost and hence no profit is being made. On the contrary, we will show that firms can achieve positive payoffs while experiencing no-regret. Moreover, we will show that infinitely many positive profit vectors are sustainable under no-regret guarantees.

We will show that by producing a probability distribution on outcomes of Bertrand oligopolies such that when the market outcomes are chosen according to this distribution, then each player's expected payoff is at least as large as the expected payoff of her best deviating strategy, given that all other players follow the distribution. More formally, we will produce a probability measure F on (P, Σ),[4] such that for all i, p_i'

$$\int_P [u_i(p_i, p_{-i}) - u_i(p_i', p_{-i})] dF(p) \geq 0$$

Such probability distributions are referred to as *coarse correlated equilibria (CCE)*[24]. It is straightforward to check that, any market history whose empirical distribution of outcomes converges to a CCE imposes no regret on the involved players. Indeed, the average profits of the players, will converge to their expected values, which by definition of the CCE exhibit no-regret. Conversely, any CCE can give rise to such a history, merely by infinitely choosing outcomes according to it. Therefore it suffices to prove that we can achieve positive payoffs payoffs in a CCE. Our constructions are inspired by observations regarding the structure of Nash in Bertrand games made in [4].

Theorem 3.1. *All symmetric linear Bertrand games exhibit coarse correlated equilibria (CCE) in which all players exhibit positive profits.*

Proof. We denote $(p-c)(a-p)/b$ by $\pi(p)$, which is equal to the utility function when the winning player is unique. This function in strictly increasing in [c, (a+c)/2]. As a result, we can define the following distribution:

$$F_0(p) = \begin{cases} 0 & p \leq \beta \\ 1 - (\frac{\pi(\beta)}{\pi(p)})^{\frac{1}{n-1}} & \beta < p < \gamma \\ 1 & p \geq \gamma \end{cases} \tag{1}$$

where $\beta > c$ and $\gamma \leq (a+c)/2$. Before, we construct the CCE, we will examine some properties of the mixed strategy profile where each player chooses a strategy according to $F_0(p)$. We will show that each action in the support of the mixed strategy $F_0(p)$ is optimal [3] except from β.

The probability distribution $F_0(p)$ sets $p = \gamma$ with probability $(\frac{\pi(\beta)}{\pi(\gamma)})^{\frac{1}{n-1}}$. The rest of the probability distribution is atomless, that is $Pr(p = x | x < \gamma) = 0$. Suppose that the rest n-1 players play according to this distribution. The expected payoff for playing price $\beta \leq p < \gamma$ would be equal to:

$$E[u] = [1 - F_0(p)]^{n-1} \pi(p) = \pi(\beta)$$

Next, we will compute the expected payoff for playing β. The only way for someone to win when playing β is for everyone else to be playing β. However, in this case they share the pot. So,

$$E[u] = [(\frac{\pi(\beta)}{\pi(\gamma)})^{\frac{1}{n-1}}]^{n-1} \frac{\pi(\gamma)}{n} = \frac{\pi(\beta)}{n}$$

Also, just to complete the picture, the expected cost for playing $p > \gamma$ is 0 and the expected profit for playing $p < \beta$ is less than $\pi(\beta)$. Lastly, let us compute the expected utility of the players when all of them play according to this strategy distribution. In this case and if we denote $(\frac{\pi(\beta)}{\pi(\gamma)})^{\frac{1}{n-1}}$ as ρ, we have that:

[4] P is the set of all strategy (price) profiles and Σ is the Borel σ-algebra on it

[3] given that all players play according to $F_0(p)$

$$E[u] = (1-\rho)\pi(\beta) + \rho\frac{\pi(\beta)}{n} = (1 - \frac{n-1}{n}\rho)\pi(\beta)$$

Now, we will define a probability distribution over outcomes of the Bertrand games and we will prove that it is a CCE. we will be using three prices α, β, γ such that $c < \alpha < \beta < \gamma \leq (a+c)/2$. With probability 1/2 all players play α and with probability 1/2 all players play according to F_0. Regarding the expected payoff for each player, we have that with probability 1/2 they all share the profit at price α and with probability 1/2 they gain the precomputed payoff of the defined mixed strategy profile. Specifically,

$$E[u] = 1/2\frac{\pi(\alpha)}{n} + 1/2[(1 - \frac{n-1}{n}\rho)\pi(\beta)]$$

In order for this to be a CCE it must be the case any deviating player cannot increase his payoff by deviating to a single strategy given that the rest of the players keep playing according to this distribution. Let us examine what are the best deviating strategies for a player. First a player can deviate and play $\alpha - \epsilon$ for some small $\epsilon > 0$. Her expected payoff in that case is essentially $\pi(\alpha)$ since she will always be winning the competition. It is obvious that any strategy less than that is clearly worse for him since π is increasing in the range $[0, \alpha] \subset [0, (a+c)/2]$. Another good deviating strategy for the player is to play a strategy in $[\beta, \gamma)$ since this is a best response to the second probability distribution. Actually, given that a player deviates to a price which is greater than α her best choice is to deviate to any price in the $[\beta, \gamma)$ range. This is true because the only way to incur payoff at this point is to maximize her payoff when her opponents play according to $F_0(p)$. As we have seen, the player achieves a maximum expected payoff of $\pi(\beta)$ when playing within that range. So, the best deviating strategy is either $\alpha - \epsilon$ or something in the range $[\beta, \gamma)$. If our current expected payoff exceeds the payoffs at these points then our distribution is a CCE. So, we wish to have:

$$1/2\frac{\pi(\alpha)}{n} + 1/2[(1 - \frac{n-1}{n}\rho)\pi(\beta)] \geq \pi(\alpha), \text{ and } 1/2\frac{\pi(\alpha)}{n} + 1/2[(1 - \frac{n-1}{n}\rho)\pi(\beta)] \geq 1/2\ \pi(\beta)\ .$$

Let us try to analyze each relation separately:

$$\begin{aligned} 1/2\frac{\pi(\alpha)}{n} + 1/2[(1 - \frac{n-1}{n}\rho)\pi(\beta)] \geq \pi(\alpha) \quad &\Leftrightarrow \\ 1/2[(1 - \frac{n-1}{n}\rho)\pi(\beta)] \geq (1 - \frac{1}{2n})\pi(\alpha) \quad &\Leftrightarrow \\ (1 - \frac{n-1}{n}\rho)\frac{n}{2n-1} \geq \frac{\pi(\alpha)}{\pi(\beta)} \end{aligned}$$

Similarly, from the second inequality we have:

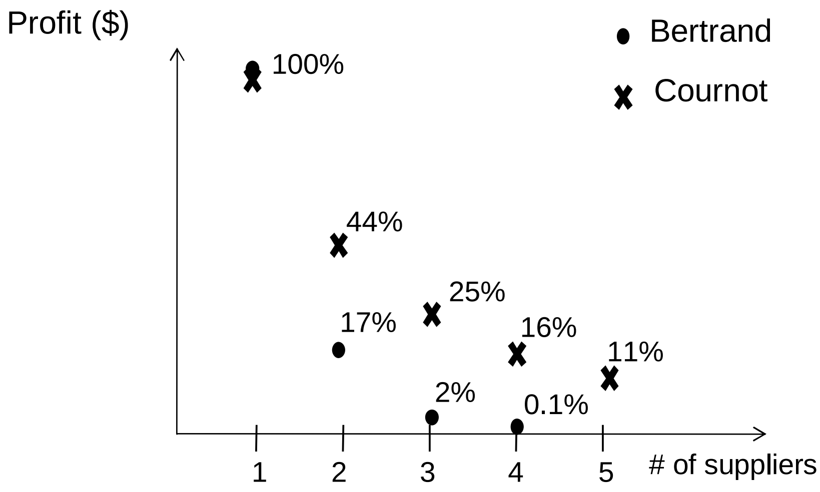

Figure 1: Profit decrease

$$\begin{aligned} 1/2\frac{\pi(\alpha)}{n} + 1/2[(1-\frac{n-1}{n}\rho)\pi(\beta)] \geq 1/2\ \pi(\beta) &\quad \Leftrightarrow \\ \frac{\pi(\alpha)}{n} + [(1-\frac{n-1}{n}\rho)\pi(\beta)] \geq \pi(\beta) &\quad \Leftrightarrow \\ \frac{\pi(\alpha)}{n} \geq \frac{n-1}{n}\rho\pi(\beta) &\quad \Leftrightarrow \\ \frac{\pi(\alpha)}{\pi(\beta)} \geq (n-1)\rho & \end{aligned}$$

So, in order for our probability distribution over outcomes to be a CCE, it suffices that we choose α, β so that:

$$(n-1)\rho \leq \frac{\pi(\alpha)}{\pi(\beta)} \leq (1-\frac{n-1}{n}\rho)\frac{n}{2n-1}$$

However $\pi(\alpha), \pi(\beta)$ are positive payoffs in the range $(0, \frac{(a-c)^2}{4b}]$ with $\pi(\alpha) < \pi(\beta)$. So, by choosing proper α, β we have reproduce any number in the range $(0,1)$. Hence, all we have to do is show that we can choose ρ appropriately such that:

$$(n-1)\rho \leq (1-\frac{n-1}{n}\rho)\frac{n}{2n-1}$$

as well as $(n-1)\rho < 1$ and $0 < (1-\frac{n-1}{n}\rho)\frac{n}{2n-1}$. Again, by manipulating the given inequality we get:

$$(n-1)\rho \leq (1-\frac{n-1}{n}\rho)\frac{n}{2n-1} \Leftrightarrow (n-1+\frac{n-1}{2n-1})\rho \leq \frac{n}{2n-1}$$

It suffices to choose $\rho = \frac{1}{2n-1}$ and $\frac{\pi(\alpha)}{\pi(\beta)} = \frac{n-1}{2n-1}$ to satisfy all inequalities. However, $\rho = (\frac{\pi(\beta)}{\pi(\gamma)})^{\frac{1}{n-1}}$. So, we have that we need to choose β and γ such that $\frac{\pi(\beta)}{\pi(\gamma)} = (\frac{1}{2n-1})^{n-1}$. So, given any

$\pi(\gamma) \in (0, \pi(\frac{a+c}{2})] = (0, \frac{(a-c)^2}{4b}]$, we can define β, α such that the distribution we have defined is a CCE. The expected payoffs of all players are positive in this CCE and can vary widely. Hence, no regret behavior can support infinitely many different positive average payoff profiles, in contrast to Bertrand's paradox. Finally, this construction establishes that increased competition is necessary for converging to marginal cost pricing. □

As we see in the figure 1, the profitability of the families of Bertrand no-regret histories we have identified, decreases much faster than the profitability of the no-regret histories in the Cournot oligopolies as the number of agents (firms) increases. In fact, for $n = 4$ players we see that essentially the prices reach the level of marginal costs as profitability drops to zero. This theoretical projection is in perfect agreement both with experimental work in the case of Bertrand games [8], as well as with empirical observations about real world oligopolistic markets [19]. Specifically, "the rule of three", as is presented in [19], states that in most markets three major players will emerge (e.g. ExxonMobil, Texaco and Chevron in petroleum). In order for the smaller companies to be successful they need to specialize and address niche markets. Our works suggests a possible quantitative explanation behind this phenomenon, as a result of the steep drop in profitability in the case of Bertrand markets.

4 Cournot Oligopolies

We will be analyzing a generalization of the Cournot model with product differentiation that was introduced by Yi[23]. By exploring ideas from that work, we will show how we can generalize its results and prove tight convergence guarantees in the case of no-regret algorithms. Our model will be the Cournot competition in the case of linear demand functions with symmetric product differentiation, where the inverse demand function $P^i(q)$ is given by $P^i(q) = a_i - b_i q_i - b_i\gamma_i Q_{-i} = a_i - b_i(1-\gamma_i)q_i - b_i\gamma_i Q$, where γ_i denotes the degree of product differentiation between products, $0 < \gamma_i \leq 1$, $b_i > 0$ and $Q = \sum_i q_i$ denotes the total product supply. We will assume that the cost functions are convex and twice continuously differentiable. We denote by $q^* = (q_1^*, \ldots, q_n^*)$ a pure Nash equilibrium of the one-shot game, which is known to exist by [18]. Finally, Q^* denotes the aggregate production level at the Nash.

Lemma 4.1. *Let q_i^τ, Q^τ denote respectively the production level of company i and the aggregate production level in period τ of a differentiated Cournot market with differentiation levels γ_i for each product. If each player's regret converges to zero, then*

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t}\left(\frac{\gamma_i - 1}{\gamma_i}\sum_i (q_i^\tau - q_i^*)^2 - (Q^\tau - Q^*)^2\right) \quad = \quad 0$$

Proof. By assumption, we have that each player i experiences vanishing regret against any deviating action $s_i \in S_i$. Specifically, we can apply this to their respective Nash equilibrium actions q_i^*.

$$\frac{1}{t}\sum_{\tau=1}^{t} u_i(q^\tau) = \max_{s_i\in S_i}\frac{1}{t}\sum_{\tau=1}^{t}\left(u_i(s_i, q_{-i}^\tau) - \mathcal{R}_i(t)\right) \geq \frac{1}{t}\sum_{\tau=1}^{t} u_i(q_i^*, q_{-i}^\tau) - \frac{\mathcal{R}_i(t)}{t} \tag{2}$$

Let us denote the difference $u_i(q^\tau) - u_i(q_i^*, q_{-i}^\tau)$ as $\Delta(u_i^\tau)$. Equation 2 allows us to bound $\sum_{\tau=1}^{t}\Delta(u_i^\tau)$ from below. Next, we will work on bounding this quantity from above.

$$\begin{aligned}\sum_{\tau=1}^{t}\Delta(u_i^\tau) &= \sum_{\tau=1}^{t}\left(P^i(q^\tau)q_i^\tau - C_i(q_i^\tau) - \left(P^i(q_i^*, q_{-i}^\tau)q_i^* - C_i(q_i^*)\right)\right)\\ &= \sum_{\tau=1}^{t}\left(P^i(q^\tau)q_i^\tau - P^i(q_i^*, q_{-i}^\tau)q_i^* - (C_i(q_i^\tau) - C_i(q_i^*))\right)\\ &= \sum_{\tau=1}^{t}(q_i^\tau - q_i^*)\left(P^i(q^\tau) - C_i'(\bar{q}_i^\tau) - b_i q_i^*\right) \qquad (3)\end{aligned}$$

The last line is derived from the mean value theorem [3] and the fact that $P^i(q_i^*, q_{-i}^\tau) = P^i(q_i^\tau, q_{-i}^\tau) + b_i(q_i^\tau - q_i^*)$.

$$\begin{aligned}\sum_{\tau=1}^{t}\Delta(u_i^\tau) &= \sum_{\tau=1}^{t}(q_i^\tau - q_i^*)\left(P^i(q^\tau) - C_i'(\bar{q}_i^\tau) - b_i q_i^*\right)\\ &\leq \sum_{\tau=1}^{t}(q_i^\tau - q_i^*)\left(P^i(q^\tau) - P^i(q^*) - \left(C_i'(\bar{q}_i^\tau) - C_i'(q_i^*)\right)\right)\\ &= \sum_{\tau=1}^{t}(q_i^\tau - q_i^*)\left(b_i(\gamma_i - 1)(q_i^\tau - q_i^*) - b_i\gamma_i(Q^\tau - Q^*) - C_i''(\tilde{q}_i^\tau)(\bar{q}_i^\tau - q_i^*)\right)\end{aligned}$$

where the inequality in the second line follows by the fact at a Nash equilibrium q^*, we have that

$$\frac{\partial u_i(q^*)}{\partial q_i} = P^i(q^*) - b_i q_i^* - C_i'(q_i^*) \leq 0.$$

Finally, the last line is derived by another application of the mean value theorem[4] and the definition of the demand functions P^i. Now, we will take the following weighted sum of the resulting inequalities over all players $i \in N$.

$$\sum_i \frac{1}{b_i\gamma_i}\sum_{\tau=1}^{t}\Delta(u_i^\tau) \leq \sum_{\tau=1}^{t}\left(\frac{\gamma_i - 1}{\gamma_i}\sum_i (q_i^\tau - q_i^*)^2 - (Q^\tau - Q^*)^2 - \sum_i \frac{C_i''(\tilde{q}_i^\tau)}{b_i\gamma_i}(q_i^\tau - q_i^*)(\bar{q}_i^\tau - q_i^*)\right)$$

By dividing the above inequality with t and combining with 2 we conclude that:

$$-\sum_i \frac{\mathcal{R}_i(t)/b_i\gamma_i}{t} \leq \frac{1}{t}\sum_{\tau=1}^{t}\left(\frac{\gamma_i - 1}{\gamma_i}\sum_i (q_i^\tau - q_i^*)^2 - (Q^\tau - Q^*)^2 - \sum_i \frac{C_i''(\tilde{q}_i^\tau)}{b_i\gamma_i}(q_i^\tau - q_i^*)(\bar{q}_i^\tau - q_i^*)\right) \quad (4)$$

Given the definition of regret minimizing behavior we have that for all i, $\limsup_{t\to\infty}\frac{\mathcal{R}_i(t)}{t} \leq 0$, therefore the $\limsup_{t\to\infty}$ of the second terms will greater or equal to 0. Ifhe cost functions are weakly convex all three terms in the summand are less or equal to 0. As a result, we have that:

[3] There exists $\bar{q}_i^\tau$ between q_i^τ, q_i^* such that $C_i(q_i^\tau) - C_i(q_i^*) = C_i'(\bar{q}_i^\tau)(q_i^\tau - q_i^*)$.

[4] There exists $\tilde{q}_i^\tau$ between $\bar{q}_i^\tau, q_i^*$ such that $C_i'(\bar{q}_i^\tau) - C_i'(q_i^*) = C_i''(\tilde{q}_i^\tau)(q_i^\tau - q_i^*)$.

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t}\left(\frac{\gamma_i - 1}{\gamma_i}\sum_i (q_i^\tau - q_i^*)^2 - (Q^\tau - Q^*)^2 - \sum_i \frac{C_i''(\tilde{q}_i^\tau)}{b_i\gamma_i}(q_i^\tau - q_i^*)(\bar{q}_i^\tau - q_i^*)\right) = 0 \quad (5)$$

The lemma follows immediately. □

Depending on the details of the Cournot model, we have the following cases:

A) Perfect substitutes

This is to the simplest case of Cournot competition and was the model analyzed by Even-Dar et. al. in [9]. We have that $\gamma_i = 1$ and $C_i''(q_i) \geq 0$ for all i, q_i. Equation 5 implies that:

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t}(Q^\tau - Q^*)^2 = 0$$

Intuitively, this equation suggests that if we exclude a statistically insignificant (sublinear) number of periods for the history of our no-regret play, then for the rest of the history the overall production levels converge to Q^τ (and therefore the prices ($P^i(q) = a_i - b_iQ,$)) converge to their levels at the Nash equilibrium Q^*.

Theorem 4.2. *Suppose that n firms participate in a homogeneous Cournot oligopoly game of perfect substitutes with linear demand ($P^i(q) = a_i - b_iQ,$) and convex cost functions. If all firms experience no-regret as t grows to infinity, then given any $\epsilon > 0$, for all but $o(t)$ periods τ in $[1, t]$ we have that $|Q^\tau - Q^*| < \epsilon$.*

We should stress here that this is a statement about the day-to-day behavior (i.e. aggregate production levels) instead of average behavior as in [9](Theorem 3.1.). In particular, this statement implies that the average action vector and the average utility of each player converge to their respective levels at the Nash equilibrium, a result that has been shown in [9]. Given the convergence of the day-to-day characteristics of the market prices and total supply, it is rather tempting to try to prove a similar statement about the convergence of the action vector and utilities of the firms and not merely of their averages. Here, we show that this cannot be the case by providing sufficient conditions for a market history to be of no-regret.

This is essentially a negative result, so it suffices to prove that this holds for as simple a model as possible. Therefore, we will focus on the special case of the fully symmetric Cournot oligopoly ($a_i = a$ and $b_i = b$) with linear cost functions. It is well known that these games exhibit a unique Nash $q^* = (q_1^*, q_2^*, \ldots, q_2^*)$ where $q_i^* = (a - (n+1)c_i - \sum_{j\in N} c_j)/((n+1)b)$.

Theorem 4.3. *Suppose that n firms participate in a homogeneous Cournot oligopoly game with linear demand ($P^i(q) = a - bQ,$) and linear cost functions and let $q*$ denote the unique Nash of this game. Any market history, where for all time periods τ, $Q^\tau = Q^*$ and where the time average $\hat{q}_i$ of each player's actions converges to her Nash strategy q_i^* does not induce regret to any player.*

An immediate corollary of the above theorem is that one cannot hope to prove convergence of the day-to-day action profiles in any model that generalizes the basic linear Cournot model. Surprisingly, if we introduce product differentiation in the market, then we can actually prove convergence of all attributes (i.e. action profiles, profits, prices e.t.c) of the market.

B) Symmetric product differentiation

In this case, we have that $0 < \gamma_i < 1$ and $C_i''(q_i) \geq 0$ for all i, q_i. Equation 5 implies that for all firms i,

$$\limsup_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} (q_i^\tau - q_i^*)^2 = 0$$

Theorem 4.4. *Suppose that n firms participate in a differentiated good Cournot oligopoly game with linear demand ($P^i(q) = a_i - q_i - \gamma Q, 0 < \gamma < 1$). If all firms experience no-regret, then given any $\epsilon > 0$, as t grows to infinity, for all but $o(t)$ periods τ in $[1, t]$ we have that $|q_i^\tau - q_i^*| < \epsilon$, where q^* is the unique Nash equilibrium.*

References

[1] J. Abernethy, E. Hazan, and A. Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. *COLT*, 2008.

[2] R. Amir. Cournot oligopoly and the theory of supermodular games. *Games and Economic Behavior*, 1996.

[3] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 2002.

[4] M. Baye and J. Morgan. A folk theorem for one-shot bertrand games. *Economic Letters*, 65(1):59–65, 1999.

[5] J. Bertrand. Theorie mathematique de la richesse sociale. *Journal des Savants.*, 67:499–508, 1883.

[6] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning and Games.* Cambridge University Press, 2006.

[7] A. A. Cournot. Recherches sur les principes mathmatiques de la thorie des richesses.. 1838.

[8] M. Dufwenberg and U. Gneezy. Price competition and market concentration: An experimental study. *International Journal of Industrial Organization 18*, 2000.

[9] E. Even-Dar, Y. Mansour, and U. Nadav. On the convergence of regret minimization dynamics in concave games. In *41st ACM Symposium on Theory of Computing. STOC*, 2009.

[10] A. Flaxman, A. T. Kalai, and B. McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. In *SODA*, 2005.

[11] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119139, 1997.

[12] D. Fudenberg and D. K. Levine. *The theory of learning in games.* MIT press, 1998.

[13] E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

[14] L. Liu. Correlated equilibrium of cournot oligopoly competition. *Journal of Economic Theory*, 1996.

[15] A. Mas-Colell, M. Winston, and J. Green. *Microeconomic Theory.* Oxford University Press, 1995.

[16] P. Milgrom and J. Roberts. Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica*, 58(6):1255–77, November 1990.

[17] P. Milgrom and J. Roberts. Adaptive and sophisticated learning in repeated normal form games. *Games and Economic Behavior*, pages 82–100, February 1991.

[18] J. Rosen. Existense and uniqueness of equilibrium points for concave n-person games.

[19] J. N. Sheth and R. S. Sisodia. *The Rule of Three: Surviving and Thriving in Competitive Markets.*

[20] R. D. Theocharis. On the stability of the cournot solution on the oligopoly problem. *The Review of Economic Studies.*, 27(2):133–134, 1960.

[21] X. Vives. *Oligopoly Pricing: Old Ideas and New Tools.* The MIT Press, 2001.

[22] J. Wu. Correlated equilibrium of bertrand competition. In *WINE*, 2008.

[23] S.-S. Yi. On the existence of a unique correlated equilibrium in cournot oligopoly. *Economic Letters*, 54:235–239, 1997.

[24] H. Young. *Strategic Learning and Its Limits.* Oxford University Press, Oxford, 2004.

[25] M. Zinkevich. Online convex programming and generalized infitesimal gradient ascent. *ICML*, 2003.

A Proofs from Section 4

proof of Theorem 4.2. We will prove this by contradiction. Indeed, suppose that this did not hold, then there would exist $\epsilon > 0$ such that it would not be the case that for all but $o(t)$ periods τ in $[1,t]$ we have that $|Q^\tau - Q^*| < \epsilon$. Namely, if we define $s_t = \{\tau | \tau \in [1,t] \text{ and } |Q^\tau - Q^*| \geq \epsilon\}$ then there exists $c > 0$ such for all k there exists $t \geq k$ such that $|s_t| \geq ct$. Hence, we can define an infinite subsequence $t_0, t_1, \ldots$ such that for all k, $|s_{t_k}| \geq ct_k$. Hence for this subsequence, we have that for all k, $\frac{1}{t_k}\sum_{\tau=1}^{t_k}(Q^\tau - Q^*)^2 \geq c\epsilon^2 > 0$. Therefore, we reach a contradiction, since $\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t}(Q^\tau - Q^*)^2 = 0$. □

proof of Theorem 4.3. As we have shown in equation 3 for all deviating strategies q_i' we have that:

$$\sum_{\tau=1}^{t}\left(u_i(q^\tau) - u_i(q_i', q_{-i}^\tau)\right) = \sum_{\tau=1}^{t}(q_i^\tau - q_i')\left(P^i(q^\tau) - C_i'(\bar{q}_i^\tau) - bq_i'\right)$$

where $\bar{q}$ between q_i^τ, q_i^* is such that $C_i(q_i^\tau) - C_i(q_i^*) = C_i'(\bar{q}_i^\tau)(q_i^\tau - q_i^*)$. However, in this case we have that since $C_i(x) = c_i x$, $C_i'(x) = c_i$ for all x. We can also substitute $P^i(q^\tau)$ with $a - bQ^\tau$. As a result, we derive that

$$\begin{aligned}
& \frac{1}{t}\sum_{\tau=1}^{t}\left(u_i(q^\tau) - u_i(q_i', q_{-i}^\tau)\right) \\
= \; & \frac{1}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i')\cdot\left(a - bQ^\tau - c_i - bq_i'\right) \\
= \; & \frac{1}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i')\left(a - bQ^* - c_i - bq_i'\right) && (6) \\
= \; & \frac{1}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i')\left(a - \frac{na - \sum_i c_i}{n+1} - c_i - bq_i'\right) && (7) \\
= \; & \frac{1}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i')\left(\frac{a + \sum_i c_i - (n+1)c_i}{n+1} - bq_i'\right) \\
= \; & \frac{1}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i')b(q_i^* - q_i') && (8) \\
= \; & \frac{b(q_i^* - q_i')}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i') \\
= \; & b(q_i^* - q_i')(\hat{q}_i - q_i')
\end{aligned}$$

where line 6 is derived by hypothesis. Also, lines 7, 8 follow from the fact that the unique Nash $q^* = (q_1^*, q_2^*, \ldots, q_2^*)$ of these games is of the form

$$q_i^* = \frac{a - (n+1)c_i - \sum_{j\in N} c_j}{(n+1)b}.$$

Lastly, since by hypothesis we have that $\lim_{t\to\infty} \hat{q}_i = q_i^*$, we derive that for each player i and all deviating actions q_i' we have that:

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} \left(u_i(q^\tau) - u_i(q_i', q_{-i}^\tau)\right) = b(q_i^* - q_i')^2 \geq 0$$

Hence in any such market history no players experience regret. □

proof of Theorem 4.4. We will prove this by contradiction. Indeed, suppose that this did not hold, then there would exist $\epsilon > 0$ such that it would not be the case that for all but $o(t)$ periods τ in $[0, t]$ we have that $|q_i^\tau - q_i^*| < \epsilon$. Namely, if we define $s_t = \{\tau | \tau \in [1, t] \text{ and } |q_i^\tau - q_i^*| \geq \epsilon\}$ then there exists $c > 0$ such for all k there exists $t \geq k$ such that $|s_t| \geq ct$. Hence, we can define an infinite subsequence $t_0, t_1, \ldots$ such that for all k, $|s_{t_k}| \geq ct_k$. Hence for this subsequence, we have that for all k, $\frac{1}{t_k} \sum_{\tau=1}^{t_k} (q_i^\tau - q_i^*)^2 \geq c\epsilon^2 > 0$. Therefore, we reach a contradiction, since $\limsup_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} (q_i^\tau - q_i^*)^2 = 0$. □