DUOPOLY

Abstract.

1. Baseline Model

Let us start with a duopoly market with differentiated products where two firms are myopic optimizers. The demand for good $i \in \{1, 2\}$ produced by firm i is

$$q_i = A - Bp_i + Cp_j \qquad i \neq j \in \{1, 2\} \tag{1.1}$$

where $A, B, C > 0$. Since $C > 0$, two goods are strategic complements. Let us assume that $B - C > 0$ (that is, the demand of good i is more elastic to the price of good i than the price of good $j \neq i$) to make the Nash equilibrium and the cartel outcome well defined. The best response $b_i(p_j)$ of firm i against firm j's price p_j is

$$b_i(p_j) = \frac{A + Cp_j}{2B} \tag{1.2}$$

The unique Nash equilibrium price is

$$p^N = \frac{2A}{2B - C}.$$

Let Π^N be the Nash equilibrium payoff of a duopolist.

The second benchmark is the Stackelberg outcome. Firm i (follower) behaves according to (1.2) to choose p^F, while firm $j(\neq i)$ chooses p^L

$$p^L = \arg\max_{p_j} p_j(A - Bp_j + Cb_i(p_j)).$$

A simple calculation shows that the leader chooses

$$p^L = \frac{A(2B + C)}{2(2B^2 - C^2)}$$

and the follower chooses

$$p^F = \frac{A(4B^2 + 2BC - C^2)}{4B(2B^2 - C^2)}.$$

Let Π^L and Π^F be the payoffs of the Stackelberg leader and the Stackelberg follower, respective.

The third benchmark is the cartel outcome (p_1^c, p_2^c) that solves the joint maximization problem:

$$\max_{(p_1,p_2)} p_1 q_1 + p_2 q_2.$$

A simple calculation shows

$$p_1^C = p_2^C = \frac{A}{2(B - C)}.$$

Let Π^C be the payoff from the cartel outcome.

One can show

$$\Pi^L > \Pi^N \text{ and } \Pi^C > \Pi^F.$$

Date: August 23, 2023.

2. Behavioral Assumption

The behavioral assumption behind a Nash equilibrium is that firm i treats firm j's action as a constant and optimizes against the constant function:

$$p_j = a_{0,i}. \tag{2.3}$$

Let S^0 be the set of all functions that can be represented as (2.3), where a duopolist perceives the opponent's price as an unobserved constant. If firm i perceives that firm $j(\neq i)$ reacts to firm i's price as in (1.2). Let S^1 be the set of all linear reaction functions

$$p_j = \alpha_{0,i} + \alpha_{1,i} p_i \qquad \alpha_{1,i} \neq 0,\ i \neq j \in \{1,2\}. \tag{2.4}$$

Let us consider a game where each duopolist first chooses a specification from $S_i \in \{S^0, S^1\}$ $i \in \{1,2\}$. Given (S_1, S_2), each firm uses a forecasting rule $f_i \in S_i$, and chooses a best response by solving

$$\max_{p_i} p_i \left(A - Bp_i + Cf_i(p_i)\right). \tag{2.5}$$

Thanks to the linearity, the maximization problem has a unique solution. Let b_i be the solution of (2.5). Under the assumption of the linear market demand, the linear reaction function is a correctly specified model of the response of the other duopolist.

We define the equilibrium concept for the specification selection game. Let $f_i \in S_i \in \{S^0, S^1\}$ be the forecast by duopolist i about p_j $j \neq i$. Define $p_j = f_i(p_i)$ as the forecasted price of firm j conditioned on p_i of firm i.

Definition 2.1. *(S_1, S_2) is a self-confirming equilibrium specification, if $S_1, S_2 \in \{S^0, S^1\}$, and $\exists (f_1, f_2) \in (S_1, S_2)$ such that*

$$b_i = \arg\max_{p_i} p_i \left(A - Bp_i + Cf_i(p_i)\right),$$

$$b_j = f_i(b_i) \qquad \forall i \neq j \in \{1,2\}.$$

To be an equilibrium, the price of firm i must be an optimal choice given its perceived law of motion $p_j = f_i(p_i)$ about how firm j responds to firm i, and the belief must be confirmed at the optimal choice. The notion of equilibrium is weaker than Nash equilibrium, as we admit misspecified belief about the opponent's response outside of the equilibrium price.

Let us consider the specification choice game between two duopolists. Firm $i \in \{1,2\}$ chooses specification $S_i \in \{S^0, S^1\}$. For each pair (S_1, S_2) of specifications, each firm chooses a self-confirming equilibrium forecasting rule of the opponent's behavior to calculate its best response specification and receives the self-confirming equilibrium payoff accordingly.

The problem is that a pair of specifications can admit multiple self-confirming equilibria. As an example, let us consider (S^1, S^1) pair of specifications. A canonical specification in S^1 is

$$p_j = \alpha_{0i} + \alpha_{1i} p_i \qquad i \neq j \in \{1,2\}.$$

First, suppose that $\alpha_{11} = \alpha_{12} = 0$, and

$$p_2 = a_{0,1} \tag{2.6}$$
$$p_1 = a_{0,2} \tag{2.7}$$
$$p_1 = \frac{A + Ca_{0,1}}{2B} \tag{2.8}$$
$$p_2 = \frac{A + Ca_{0,2}}{2B} \tag{2.9}$$
$$p_1 = p_2. \tag{2.10}$$

(2.6) and (2.7) imply the perfect foresight condition of duopolists 1 and 2, respectively. (2.8) and (2.9) imply that the price of duopolist 1 and 2 must be a best response. The self-confirming equilibrium price is exactly the Nash equilibrium price

$$p_1 = p_2 = p^N = \frac{A}{2B - C}.$$

Second, consider an asymmetric case where $\alpha_{11} = 0$, $\alpha_{02} = \frac{A}{2B}$,

$$\alpha_{12} = \frac{C}{2B}$$

$$\begin{aligned} p_2 &= a_{01} && (2.11)\\ p_1 &= a_{02} + a_{12}p_2 && (2.12)\\ p_1 &= \frac{A + Ca_{01}}{2B} && (2.13)\\ p_2 &= \frac{A + Ca_{02}}{2(B - Ca_{12})} && (2.14)\\ & && (2.15) \end{aligned}$$

In (2.12), we substitute p_1 by (2.13) and p_2 by (2.11),

$$\frac{A + Ca_{0,1}}{2B} = a_{0,2} + a_{1,2}a_{0,1}$$

which must hold as an identify of $a_{0,1}$. Thus,

$$a_{0,2} = \frac{A}{2B} \text{ and } a_{1,2} = \frac{A}{2B}$$

which implies

$$p_1 = p^F \text{ and } p_2 = p^L$$

and

$$p_1 = \Pi^F \text{ and } p_2 = \Pi^L.$$

Finally, consider (S^1, S^1). The equilibrium conditions are

$$\begin{aligned} p_2 &= a_{0,1} + a_{1,1}p_1 && (2.16)\\ p_1 &= a_{0,2} + a_{1,2}p_2 && (2.17)\\ p_1 &= \frac{A + Ca_{0,1}}{2(B - Ca_{11})} && (2.18)\\ p_2 &= \frac{A + Ca_{0,2}}{2(B - Ca_{12})} && (2.19)\\ & && (2.20) \end{aligned}$$

Suppose that

$$a_{0,1} = a_{0,2} = 0 \text{ and } a_{1,1} = a_{1,2} = 1$$

from which

$$p_1 = p_2 = p^C$$

and

$$\Pi_1 = \Pi_2 = \Pi^C$$

follow.

We have constructed three different self-confirming equilibira for (S^1, S^1), which induce many other self-confirming equilibria. Following the same logic, we can construct multiple self-confirming equilibria for (S^0, S^1) and (S^1, S^0). We need a criterion to select a self-confirming equilibrium for each pair of specifications to make the specification selection game as a tool to explain why a duopolist chooses a particular specification.

3. Learning Dynamics

Each duopolist chooses a specification for the behavior of the opponent. Let S^0 be the collection of all 0-th order polynomial that assumes the price of the other player as a fixed but unobservable constant:

$$p_j = a_{0i}.$$

Let S^1 be the collection of all 1-st order polynomial

$$p_j = \alpha_{0i} + \alpha_{1i} p_i \qquad i, j \in \{1, 2\} \tag{3.21}$$

A specification from S^1 presumes that firm j's reaction function is an affine function of firm i's price. Since it is a common knowledge that the two goods are strategic complemnts, it is natural to impose a restriction that

$$a_{1i} \geq 0 \qquad \forall i.$$

To ensure the stability of the learning dynamics, we impose

$$a_{1i} \leq 1 \qquad \forall i.$$

After each duopolist chooses a specification, the duopoly game is played and the coefficient of each specification is updated according to the least square learning algorithm. Let $\alpha_{it} = (\alpha_{0i,t}, \alpha_{1i,t})$ be the estimator of the coefficients. Under S^0, $\alpha_{1i,t} \equiv 0$. Given $\alpha_{i,t-1}$, $b_i(\alpha_{i,t-1})$ is the best response to the estimated reaction curve. The actual price in period t is

$$p_{i,t} = b_i(\alpha_{i,t-1}) + \epsilon_{i,t}.$$

For example, suppose that each duopolist chooses S^1, assuming that the reaction function of the other firm is (3.21), and updates the coefficient according to

$$\begin{aligned} \begin{bmatrix} \alpha_{0i,t} \\ \alpha_{1i,t} \end{bmatrix} &= \begin{bmatrix} \alpha_{0i,t-1} \\ \alpha_{1i,t-1} \end{bmatrix} + \lambda R_{i,t-1}^{-1} \begin{bmatrix} 1 \\ p_{i,t} \end{bmatrix} \left(p_{j,t} - \alpha_{0i,t-1} - \alpha_{1i,t-1} p_{i,t} \right) \\ R_{i,t} &= R_{i,t-1} + \lambda \left(\begin{bmatrix} 1 & p_{i,t} \\ p_{i,t} & p_{i,t}^2 \end{bmatrix} - R_{i,t-1} \right). \end{aligned}$$

We can define the recursive least square estimation process for other combinations of specifications accordingly.

Given (S^i, S^j) $i, j \in \{0, 1\}$, let $\Pi^{\lambda}_{ij,k,t}$ be the profit of firm k in period t if duopolist 1 chooses S^i and duopolist 2 chooses S^j.

Proposition 3.1.

$$\lim_{\lambda \to 0} \lim_{t \to \infty} (a_{0,i,t}, a_{1,i,t}) = (p_i^N, 0)$$

in distribution.

$$\lim_{\lambda \to 0} \lim_{t \to \infty} \mathsf{E}\Pi^{\lambda}_{ij,t} = \Pi^N.$$

As long as $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are independent, $p_{1,t}$ and $p_{2,t}$ should not have any correlation. Thus, the slope of the estimated reaction function must be 0 "on average." If we select a self-confirming equilibrium which is a stable stationary solution of the least square learning dynamics, the payoff matrix of the specification game is written in terms of the payoff of duopolist 1:

$$\begin{array}{c} \\ S^0 \\ S^1 \end{array} \begin{array}{c} \begin{array}{cc} S^0 & \quad S^1 \end{array} \\ \begin{pmatrix} \Pi^N, \Pi^N & \Pi^N, \Pi^N \\ \Pi^N, \Pi^N & \Pi^N, \Pi^N \end{pmatrix} \end{array}. \tag{3.22}$$

The stability of the mean dynamics of the learning algorithm is a widely used selection criterion. The mean dynamics of ODE of the estimated specification reduces the set of the self-confirming equilibria to a unique equilibrium where the duopolists play the Nash equilibrium regardless of the specification. For any combination of specifications, the stable stationary payoff from the learning dynamics is the Nash equilibrium payoff. The mean dynamics cannot explain why a duopolist chooses a particular specification of the response of the other duopolist.

Instead, we examine the "perturbed" game with small $\lambda > 0$ for a larger t.

$$\begin{array}{cc} & \begin{array}{cc} S^0 & S^1 \end{array} \\ \begin{array}{c} S^0 \\ S^1 \end{array} & \begin{pmatrix} \Pi^\lambda_{00,1t}, \Pi^\lambda_{00,1t} & \Pi^\lambda_{01,1t}, \Pi^\lambda_{01,2t} \\ \Pi^\lambda_{10,1t}, \Pi^\lambda_{10,2t} & \Pi^\lambda_{11,1t}\Pi^\lambda_{11,2t} \end{pmatrix}. \end{array} \tag{3.23}$$

Definition 3.2. *We say that S^i is a risk dominant strategy of duopolist 1 if*

$$\frac{1}{2}\left[\Pi^\lambda_{i1,1t} + \Pi^\lambda_{i2,1t}\right] \geq \frac{1}{2}\left[\Pi^\lambda_{j1,1t} + \Pi^\lambda_{j2,1t}\right]$$

and is a strictly risk dominant strategy if the weak inequality is replaced by a strict inequality. We can define the risk dominance for duopolist 2 in the same manner.

Without any prior knowledge of the opponent's choice of specification, duopolist 1 assigns an uninformative prior (i.e., uniform) distribution over the strategy space of duopolist 2, $\{S^1, S^2\}$. A risk dominant strategy is a best specification against the uninformative prior.

By exploiting the large deviation properties of the learning dynamics for each pair of specifications, we show that (S^1, S^1) is a risk dominant outcome.

Theorem 3.3. $\forall \nu > 0$, $\exists \lambda(\nu) > 0$ *such that* $\forall \lambda \in (0, \lambda(\nu))$, $\exists T(\lambda, \nu)$ *such that* $\forall t \geq T(\lambda, \nu)$, S^1 *is the best response against a mixed strategy over* $\{S^0, S^1\}$ *assigning probability at least* ν *to* S^1.

If S^1 is a best response any any mixed strategy over $\{S^0, S^1\}$, then S^1 is a dominant strategy. It takes only a small amount of probability assessment that the opponent is playing S^1 to justify S^1 as an optimal choice.

4. Proof

4.1. **Learning Dynamics under** (S^1, S^1)**.** We examine the dynamics of duopolist 1 in the neighborhood of the stable stationary point that is the Nash equilibrium. The analysis of the learning dynamics of duopolist 2 follows from symmetry. Let us consider the learning dynamics under (S^1, S^1). The reaction function

$$p_2 = \alpha_{01} + \alpha_{11} p_1$$

is paramterized by $(\alpha_{01}, \alpha_{11})$. Instead, it is more convenient to write the reaction function centered around the average price.

$$p_2 - \overline{p}_2 = \alpha_{11}(p_1 - \overline{p}_1).$$

Since $p_i = b_i(\alpha) + \epsilon_i$,

$$\overline{p}_{i,t} = \frac{1}{t}\sum_{s=1}^{t} b_i(\alpha) + \epsilon_i$$

implying that $\overline{p}_{i,t}$ evolves according to the time scale of the sample average of $\alpha_{i,t}$. Since we are interested in the case where $t \to \infty$, $\overline{p}_{i,t}$ evolves infinitely more slowly than $\alpha_{i,t}$, allowing us treating $\overline{p}_{i,t} = \overline{p}_i$.

The best response of duopolist 1 is therefore

$$b_1 = \frac{A + C(\overline{p}_2 - \overline{p}_2\alpha_1)}{2(B - C\alpha_1)}$$

since

$$\alpha_{01} = \overline{p}_2 - \alpha_{11}\overline{p}_1.$$

The associated ODE is

$$\dot{\alpha}_{11} = \frac{\mathsf{E}(p_1 - \overline{p}_1)(p_2 - \overline{p}_2)}{\mathsf{E}(p_1 - \overline{p}_1)^2} - \alpha_{11}.$$

Since $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are mutually independent, $\mathsf{E}(p_1 - \overline{p}_1)(p_2 - \overline{p}_2) = 0$. Thus,

$$\dot{\alpha}_{11} = -\alpha_{11}$$

and similarly,

$$\dot{\alpha}_{12} = -\alpha_{12}$$

which proves that $(\alpha_{11}, \alpha_{12}) = (0, 0)$ is the stable stationary point of the mean dynamics.

We are interested in the dynamics of $(\alpha_{1i,t}, \overline{p}_{i,t})_{i\in\{1,2\}}$ and the corresponding variables of duopolist 1. In principle, we have to keep track of four variables. Because we have two stochastic shocks, we can analyze the stochastic process of two variables. We choose $(\alpha_{11,t}, \alpha_{12,t})$. We can infer the dynamic properties of $(\overline{p}_1, \overline{p}_2)$ according to

$$\lim_{\lambda\to 0}\lim_{t\to\infty} \overline{p}_i = \lim_{\lambda\to 0}\lim_{t\to\infty} \mathsf{E}b_i(\alpha_{1i}) \qquad i \in \{1, 2\}.$$

We are interested in the dynamics around a small neighborhood of $(\alpha_{11}, \alpha_{12}) = (0, 0)$ where $(\overline{p}_1, \overline{p}_2) = (p^N, p^N)$. Instead of the "usual" distribution of $(\epsilon_{1,t}, \epsilon_{2,t})$, let us consider the dynamics of $(\alpha_{11,t}, \alpha_{12,t})$ following a series of "unusual" events, namely T numbers of perfectly correlated $(\alpha_{11,t}, \alpha_{12,t})$. We shall calculate T later, whose size plays a vital role in understanding the large deviation properties of the learning dynamics under (S^1, S^1).

Suppose that $\epsilon_{i,t}$ is a binomial distribution, assining an equal probability over $\{\epsilon, -\epsilon\}$ for $\epsilon > 0$. According to the law of large numbers, the empirical frequency of (ϵ, ϵ), $(\epsilon, -\epsilon)$, $(-\epsilon, \epsilon)$ and $(-\epsilon, -\epsilon)$ must be close to $1/4$ over any small positive (clock) time interval.

Instead of a "usual" sample path, we examine an "unusual" sample path where a perfectly positively correlated shocks $(\epsilon_{1,t}, \epsilon_{2,t})$ occur for T rounds, followwed by a "usual" events where each of four realizations of $(\epsilon_{1,t}, \epsilon_{2,t})$ occur with probability close to $1/4$. We claim that $\exists T < \infty$ such that after T periods of perfectly correlated shocks, the gradient vector of the mean dynamics of $(\alpha_{11,t}, \alpha_{12,t})$ is pointing away from the locally stable point $(0, 0)$.

Suppose that T number of perfectly corrected $(\epsilon_{1,t}, \epsilon_{2,t})$ have occurred, which is a rare event. The probability of such an event would be 4^{-T} if $\epsilon_{i,t}$ is a binomial distribution. During the T periods, the updating process of $\alpha_{11,t}$ and $\alpha_{12,t}$ is perfectly synchronized. As a result $p_{1,t} = p_{2,t}$, which makes $\alpha_{11,t} = \alpha_{12,t} \to 1$ as more perfectly correlated shocks arrive. At the same time, $b_i(\alpha_{1i,t}, \overline{p}_t)$ increases as $\alpha_{1i,t}$ increases from 0. On average, $p_{i,t} \geq p^N$, and therefore, $\overline{p}_t \geq p^N$ on average. However, since $\overline{p}_t$ is a simple average of $b_i(\alpha_{i,t}) + \epsilon_{i,t}$, $\overline{p}_{i,t}$ evolves at an infinitely slower time scale than $\alpha_{i,t}$ and can be treated as a constant while investigating the dynamics of $\alpha_{i,t}$.

After T rounds of perfectly correlated shock, let us consider the evolution of $(\alpha_{11}, \alpha_{12})$, which in a small neighborhood of $(0, 0)$ where $a_{11} = a_{12} = a$ and $\overline{p}_1 = \overline{p}_2 = \underline{p}$. Since (a_{11}, a_{12}) is in a small neighborhodd of $(0, 1)$, so is $(\overline{p}_1, \overline{p}_2)$ in a small neighborhood of (p^N, p^N). We have yet to show that we can choose σ^2 sufficiently small so that (a_{11}, a_{12}) is in a small neighborhood of $(0, 0)$ and T does not explode as $\sigma^2 \to 0$.

Each duopolist is facing T price vectors, $\{(p_{1,t}, p_{2,t})\}_{t=1}^T$ where $p_{1,t} = p_{2,t}$. Consider $t = T + 1$. Since $\alpha_{1i,t}$ is selected to fit the data best, $\alpha_{1i,t} \to 1$. Moreover, $\mathsf{E}\alpha_{11,t} = \mathsf{E}\alpha_{12,t}$ and $\overline{p}_1 = \overline{p}_2$, which implies $\mathsf{E}b_1(\alpha_{11,t}) = \mathsf{E}b_2(\alpha_{12,t})$ for $t \geq T$ on average. Therefore,

$$\mathsf{E}(b_1(\cdot) - \overline{p}_1)(b_2(\cdot) - \overline{p}_2) = \mathsf{E}(b_1(\cdot) - \overline{p}_1)^2.$$

Since $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are orthogonal in $t \geq T$,

$$\mathsf{E}(\epsilon_{1,t} - \overline{p}_1)(\epsilon_{2,t} - \overline{p}_2) = 0.$$

Therefore,

$$\mathsf{E}(p_{1,t}-\overline{p}_{1,t})(p_{2,t}-\overline{p}_{2,t}) = \mathsf{E}(b_1(\cdot)-\overline{p}_{1,t})^2.$$

Similarly,

$$\mathsf{E}(p_{1,t}-\overline{p}_{1,t})^2 = \mathsf{E}(b_1(\cdot)-\overline{p}_{1,t})^2+\sigma^2.$$

After a tedious calculation, we have

$$\dot{\alpha}_{1i} = \frac{\left[\frac{A+C\overline{p}(1-\alpha_{1i})}{2(B-C\alpha_{1i})}-\overline{p}_i\right]^2}{\left[\frac{A+C\overline{p}(1-\alpha_1)}{2(B-C\alpha_1)}-\overline{p}\right]^2+\sigma^2}-\alpha_1 \qquad \forall i$$

that dictates the "normal" evolution of α_t, following T number of unusual shocks. We can write the right hand side of ODE as

$$f_{11}(\alpha_{1i},\overline{p}_i) \equiv \frac{\left[\frac{A+C\overline{p}(1-\alpha_{1i})}{2(B-C\alpha_{1i})}-\overline{p}_i\right]^2}{\left[\frac{A+C\overline{p}_i(1-\alpha_{1i})}{2(B-C\alpha_{1i})}-\overline{p}_i\right]^2+\sigma^2}-\alpha_{1i} \qquad \forall i. \tag{4.24}$$

Suppose that $\overline{p}_i = p^N$ $\forall i$. For a sufficiently small $\sigma^2>0$, there are three solutions of α_{1i} satisfying

$$f_{11}(\alpha_{1i},\overline{p}^N)=0.$$

One solution is $\alpha_{1i}=0$, and the other two solutions $\underline{\alpha}$ and $\overline{\alpha}$ satisfying

$$0<\underline{\alpha}<\overline{\alpha}<1.$$

Thus, if $\alpha_{1i}>\underline{\alpha}$ where $\overline{p}=p^N$, then $\dot{\alpha}_{1i}>0$, which make $(\alpha_{11,t},\alpha_{12,t})$ increase simultaneously to $(\overline{\alpha},\overline{\alpha})$ maintaining perfect coordination. The perfect coordination is a consequence of the identical right handside of the mean dynamics, rather than the perfect correlation of $\epsilon_{1,t}$ and $\epsilon_{2,t}$. As $(\alpha_{11,t},\alpha_{12,t})$ increases, so does $(p_{1,t},p_{2,t})$ on average, because $b_i(\cdot)$ increases $\forall i$.

Since $\overline{p}_i$ also increases as $\alpha_{1,i}$ increases, we need to check how $\underline{\alpha}$ changes as $\overline{p}_i$ increases from p^N, to ensure that the ODE trajectory where $\dot{\alpha}_{1i}>0$ is sustainable until α_{1i} approaches $\overline{\alpha}$ in a small neighborhood of 1. A simple calculation shows that $\underline{\alpha}$ decreases as $\overline{p}_i$ increases, implying that if $\alpha_{1i}>\underline{\alpha}$ so that $\dot{\alpha}_{1i}>0$ when $\overline{p}_i=p^N$, then $\dot{\alpha}_{1i}>0$ until α_{1i} reaches $\overline{\alpha}$. We report a few useful properties of $\underline{\alpha}$ and $\overline{\alpha}$.

Proposition 4.1.

$$\lim_{\sigma_2\to 0}\underline{\alpha}(\sigma^2)=0,\ \lim_{\sigma_2\to 0}\overline{\alpha}(\sigma^2)=1 \ \textit{and} \ \lim_{\sigma^2\to 0}\frac{\underline{\alpha}(\sigma^2)}{\sigma^2}=2B^2.$$

Proof. We are interested in the range of α_1 where $\dot{\alpha}_1>0$. If $\overline{p}=p^N$, the right hand side of ODE vanishes at three values of α_{1i} solving

$$\alpha_{1i}^2(1-\alpha_{1i}) = 2\alpha_{1i}(B-C\alpha_{1i})^2\sigma^2.$$

$\alpha_{1i}=0$ which is the stable stationary point of the mean dynamics. The other two values, $\underline{\alpha}$ and $\overline{\alpha}$, are the solutions of

$$\alpha_{1i}(1-\alpha_{1i}) = 2(B-C\alpha_{1i})^2\sigma^2$$

or equivalently,

$$(1+2\sigma^2C^2)\alpha_{1i}^2-(2\sigma^2BC+1)\alpha_{1i}+2\sigma^2B^2=0.$$

Let $\underline{\alpha}<\overline{\alpha}$ be the two real valued solutions which exist if σ^2 is sufficiently small. As $\sigma^2\to 0$, $\underline{\alpha}\to 0$ and $\overline{\alpha}\to 1$. Since

$$\overline{\alpha}\underline{\alpha} = \frac{2\sigma^2B^2}{1+2\sigma^2C^2},$$

$$\frac{\underline{\alpha}}{\sigma^2} = \frac{2B^2}{1+2\sigma^2C^2} \tag{4.25}$$

from which the last part of the proposition follows. $\square$

To highlight the fact that $\underline{\alpha}$ and $\overline{\alpha}$ depend on σ^2, we often write $\underline{\alpha}(\sigma^2)$ and $\overline{\alpha}(\sigma^2)$. Given (p^N, p^N), let us consider two of its neighborhood of $(\alpha_1, \alpha_1) = (0, 0)$.

$$\mathcal{N}_{\underline{\alpha}}(0,0) = \{(\alpha_1, \alpha_1) \mid \|(\alpha_1, \alpha_1) - (0,0)\| \leq \underline{\alpha}(\sigma^2)\}$$

and

$$\mathcal{N}_{\mu}(0,0) = \{(\alpha_1, \alpha_1) \mid \|(\alpha_1, \alpha_1) - (0,0)\| \leq \mu\}$$

for a small fixed $\mu > 0$.

As $\sigma^2 \to 0$, $\underline{\alpha}(\sigma^2) \to 0$. We can choose σ^2 sufficiently small so that $\underline{\alpha}(\sigma^2) < \mu$ so that $\mathcal{N}_{\underline{\alpha}} \subset \mathcal{N}_{\mu}$. We are interested in the probability that $(\alpha_{11,t}, \alpha_{12,t})$ moves out of $\mathcal{N}_{\mu}(0,0)$ for each combination of specifications.

4.2. **Rate function for (S^1, S^1).** Note that from $(\underline{\alpha}, \underline{\alpha})$, the trajectory of the mean dynamics lead to $(\overline{a}, \overline{a})$. Thus, if an escape path reaches a small neighborhood of $(\underline{\alpha}, \underline{\alpha})$, the path can hit the boundary of $N_{\mu}(0,0)$ with probability 1. Thus, the escape probability from $\mathcal{N}_{\mu}(0,0)$ is equal to the escape probability to a small neighborhood of $(\underline{\alpha}, \underline{\alpha}) \in \partial N_{\underline{\alpha}}(0,0)$ from $(0,0)$. Thus,

$$\lim_{\lambda \to 0} -\lambda \log \mathsf{P}\left((\alpha_{1,t}, \alpha_{1,t}) \notin \mathcal{N}_{\mu}(0,0)\right) = \lim_{\lambda \to 0} -\lambda \log \mathsf{P}\left((\alpha_{1,t}, \alpha_{1,t}) \notin \mathcal{N}_{\underline{\alpha}}(0,0)\right) \equiv \rho_{11}(\underline{\alpha}, \sigma^2).$$

By (4.25),

$$\lim_{\sigma^2 \to 0} \rho_{11}(\underline{\alpha}, \sigma^2) = \lim_{\sigma^2 \to 0} \rho_{11}(\mu, \sigma^2) = \rho_{11}^* < \infty. \tag{4.26}$$

Rough speaking, (4.26) implies that it does not take infinitely many number of unusual shocks to push $(\alpha_{11,t}, \alpha_{12,t})$ from a small neighborhood of $(0,0)$. The increment of $(\alpha_{11,t}, \alpha_{12,t})$ resulting from each "unusual" shock becomes smaller as $\sigma^2 \to 0$. Since $\underline{\alpha}(\sigma^2) \to 0$ sufficiently quickly, it does not take infinitely many unusual shocks to reach a small neighborhood of $(\underline{\alpha}, \underline{\alpha})$. From there, the mean dynamics pushes $(\alpha_{11,t}, \alpha_{12,t})$ to the boundary of $\mathcal{N}_{\mu}(0,0)$, not requiring a sequence of unusual shocks.

4.3. **Rate function for Other Combination of Specifications.** Let us consider the case of (S^0, S^0). The stable stationary point of the mean dyanmics of $(\alpha_{11,t}, \alpha_{12,t})$ is $(0,0)$. For a fixed $\mu > 0$, let $\rho_{00}(\mu, \sigma^2)$ be the large deviation rate function. Under specification S^0, $\alpha_{1,i} = 0$ and the feedback from the other player's action to $\alpha_{1,i}$ is shut off. The only way the escape path reaches $\partial N_{\mu}(0,0)$ is through a sequence of "unusual" shocks. Thus,

$$\lim_{\sigma^2 \to 0} \rho_{00}(\mu, \sigma^2) = \infty.$$

Following the same reasoning, we conclude that $\forall i, j \in \{0, 1\}$ with $i + j \leq 1$,

$$\lim_{\sigma^2 \to 0} \rho_{ij}(\mu, \sigma^2) = \infty. \tag{4.27}$$

As $\sigma^2 \to 0$, the increment of $(\alpha_{11,t}, \alpha_{12,t})$ becomes smaller. Consequently, it takes more unusual shocks to move $(\alpha_{11,t}, \alpha_{12,t})$ the same distance μ away from the stable stationary point $(0,0)$ to the boundary of $\mathcal{N}_{\mu}(0,0)$.

4.4. **Average Payoff.** If $(\alpha_{11}, \alpha_{12}) \in \mathcal{N}_{\mu}(0,0)$ for a small $\mu > 0$, the maximum difference between firm 1's payoff and Π^N is bounded by μM for some $M > 0$. Let M' be the largest profit firm 1 can generates if $(\alpha_{11}, \alpha_{12}) \notin N_{\mu}(0,0)$. Since (α_1, α_1) is contained in a compact set, $M' < \infty$.

Suppose that the initial condition of $(\alpha_{11,1}, \alpha_{12,1}) = (0,0)$. The weak convergence theory implies that $\forall \gamma > 0$, $\forall \tau > 0$, $\exists \lambda(\gamma) > 0$ such that $\forall \lambda \in (0, \lambda(\gamma))$

$$\mathsf{P}\left(\exists t \in \{1, \ldots, \lceil \tau/\lambda \rceil\},\ |\Pi_{ij} - \Pi_{ij,t}^{\lambda}| > \gamma\right) < \gamma.$$

The large deviation theory provides a better approximation. $\forall \gamma > 0$, $\forall \tau > 0$, $\exists \rho_\gamma > 0$, $\exists \lambda_\gamma > 0$ such that $\forall \lambda \in (0, \lambda_\gamma)$

$$\mathsf{P}\left(\exists t \in \{1, \ldots, \lceil \tau/\lambda \rceil\},\ |\Pi_{ij} - \Pi^\lambda_{ij,t}| > \gamma\right) < e^{-\frac{\tau \rho_\gamma}{\lambda}}.$$

For a fixed $\lambda > 0$, we choose small $\mu > 0$, and $\sigma^2 > 0$ so that

$$\rho_{11}(\underline{\alpha}, \sigma^2) \leq -\lambda \log \frac{4\mu M}{M' - \mu M}. \tag{4.28}$$

Since $\rho_{11} < \infty$, $\forall \lambda > 0$, we can choose sufficiently small $\mu > 0$ to satsify (4.28).

$\Pi^\lambda_{11,t} - \Pi^N$ is bounded from below by

$$\mu M (1 - e^{-\frac{\rho_{11}(\underline{\alpha},\sigma^2)}{\lambda}}) + M' e^{-\frac{\rho_{11}(\underline{\alpha},\sigma^2)}{\lambda}} \equiv g^\lambda_{11}(\mu, \sigma^2). \tag{4.29}$$

since the escape trajectory from $(\underline{\alpha}, \underline{\alpha})$ to $(\overline{\alpha}, \overline{\alpha})$ follows the trajectory of the mean dynamics. Thus, with probability $1 - e^{-\frac{\tau \rho_\mu}{\lambda}}$

$$\Pi^\lambda_{11,t} \geq \Pi^N + f^\lambda_{11}(\mu, \sigma^2).$$

As $\sigma^2 \to 0$,

$$g^\lambda_{11}(\mu, \sigma^2) \to \mu M (1 - e^{-\frac{\rho^*_{11}}{\lambda}}) + M' e^{-\frac{\rho^*_{11}}{\lambda}}.$$

On the other hand, for other combinations of specifications, $f_{ij}(\mu, \sigma^2)$ (where $i, j \in \{0, 1\}$ but $i + j \leq 1$) is bounded from above by

$$\mu M (1 - e^{-\frac{\rho_{ij}(\mu,\sigma^2)}{\lambda}}) + M' e^{-\frac{\rho_{ij}(\mu,\sigma^2)}{\lambda}} \equiv g_{ij}(\mu, \sigma^2) \tag{4.30}$$

Thus, with probability $1 - e^{-\frac{\tau \rho_\mu}{\lambda}}$

$$\Pi^\lambda_{ij,t} \leq \Pi^N + g_{ij}(\mu, \sigma^2).$$

For fixed $\lambda, \mu > 0$,

$$\lim_{\sigma^2 \to 0} \mu M (1 - e^{-\frac{\rho_{ij}(\underline{\alpha},\sigma^2)}{\lambda}}) + M' e^{-\frac{\rho_{ij}(\mu,\sigma^2)}{\lambda}} = \mu M$$

since $\rho_{ij}(\mu, \sigma^2) \to \infty$ as $\sigma^2 \to 0$.

4.5. **Risk Dominance.** To show that S^1 is the risk dominant strategy, we have to show that if the opponent plays S^0 and S^1 with an equal probability, then the average payoff from S^1 is larger than the average payoff S^0. A simple calculation shows that

$$\begin{aligned} \Pi^\lambda_{11,t} - \Pi^\lambda_{01,t} &\geq g^\lambda_{11}(\mu, \sigma^2) - g^\lambda_{01}(\mu, \sigma^2) \\ \Pi^\lambda_{10,t} - \Pi^\lambda_{00,t} &\geq -g^\lambda_{10}(\mu, \sigma^2) - g^\lambda_{00}(\mu, \sigma^2). \end{aligned}$$

with probability $1 - e^{-\frac{\rho_\Pi}{\lambda}}$. Thus,

$$\frac{1}{2}\left(\Pi^\lambda_{11,t} + \Pi^\lambda_{10,t}\right) - \frac{1}{2}\left(\Pi^\lambda_{01,t} + \Pi^\lambda_{00,t}\right) \geq \frac{1}{2}\left(g^\lambda_{11}(\mu, \sigma^1) - \sum_{i+j \leq 1} g^\lambda_{ij}(\mu, \sigma^2).\right).$$

Recall that $\rho^*_{11} < \infty$ and $\lim_{\sigma^2 \to \infty} \rho_{ij}(\mu, \sigma^2) = \infty$ if $i + j \leq 1$. If

$$\rho^*_{11} < -\lambda \log \frac{4\mu M}{M' - \mu M}, \tag{4.31}$$

$\exists \overline{\sigma}^2 > 0$ such that $\forall \sigma^2 < \overline{\sigma}^2$,

$$g^\lambda_{11}(\mu, \sigma^2) - \sum_{i+j<2} g^\lambda_{ij}(\mu, \sigma^2) > 0$$

implying that S^1 is the risk dominant strategy. Moreover,

$$\lim_{\sigma^2 \to 0} \frac{\sum_{i+j<2} g_{ij}^{\lambda}(\mu, \sigma^2)}{g_{11}^{\lambda}(\mu, \sigma^2)} = 0.$$

Define ν^* satisfying

$$\nu^* \Pi_{11,t}^{\lambda} + (1-\nu^*)\Pi_{10,t}^{\lambda} = \nu^* \Pi_{01,t}^{\lambda} + (1-\nu^*)\Pi_{00,t}^{\lambda}.$$

clearly, $\lim_{\sigma^2 \to 0} \nu^* = 0$, from which the last part of the theorem follows.

5. Formal Description

The behavior of duopolist $i \in \{1,2\}$ is described as an algorithm that updates $(\pi_{i,t}; \alpha_{0i,t}^0; \alpha_{0i,t}^1, \alpha_{1i,t}^1; \overline{\Pi}_{i,t}^0, \overline{\Pi}_{i,t}^1)$ recursively, where $\pi_{i,t}$ is the probability assigned to the hypothesis that the opponent's behavior is specified as S^1, $\alpha_{0i,t}^0$ is the coefficient for the specification of S^0

$$p_j = \alpha_{0i,t}^0$$

and $(\alpha_{0i,t}^1, \alpha_{1i,t}^1)$ is the coefficients of the linear reaction function in S^1

$$p_j = \alpha_{0i,t}^1 + \alpha_{1i,t}^1 p_i.$$

$\overline{\Pi}_{i,t}^k$ is the average payoff of duoplist i at the end of period t under the hypothesis that the opponent's action is specified according to S^k.

At the beginning of period t, duopolist i is given

$$\left(\pi_{i,t-1}; \alpha_{0i,t-1}^0; \alpha_{0i,t-1}^1, \alpha_{1i,t-1}^1; \overline{\Pi}_{i,t-1}^0, \overline{\Pi}_{i,t-1}^1\right).$$

Duopolist i solves

$$b_{i,t}^0 = \arg\max_p p(A - Bp + C\alpha_{0i,t-1}^0)$$

and

$$b_{i,t}^1 = \arg\max_p p\left(A - Bp + C\left(\alpha_{0i,t-1}^1 + \alpha_{1i,t-1}^1 p\right)\right)$$

so that

$$b_{i,t}^0 = \frac{A + C\alpha_{0i,t-1}^0}{2B}$$

and

$$b_{i,t}^1 = \frac{A + C\alpha_{0i,t-1}^1}{2(B - C\alpha_{1i,t-1}^1)}.$$

Duopolist i experiments by adding a white noise $\epsilon_{i,t}$ whose variance is $\sigma_i^2 > 0$ so that

$$p_{i,t}^0 = b_{i,t}^0 + \epsilon_{i,t}$$

and

$$p_{i,t}^1 = b_{i,t}^1 + \epsilon_{i,t}.$$

Duopolist i chooses the price in period t according to

$$p_{i,t} = (1 - \pi_{i,t-1})p_{i,t}^0 + \pi_{i,t-1}p_{i,t}^1.$$

After observing $(p_{1,t}, p_{2,t})$, duopolist i updates the parameters. We can calculate

$$\Pi_{i,t}^0 = p_{i,t}^0(A - Bp_{i,t}^0 + Cp_{j,t})$$

and

$$\Pi_{i,t}^1 = p_{i,t}^1\left(A - Bp_{i,t}^1 + Cp_{j,t}\right)$$

from which we update the average expected payoff

$$\begin{aligned} \overline{\Pi}^0_{i,t} &= \overline{\Pi}^0_{i,t-1} + \lambda \left(\Pi^0_{i,t} - \overline{\Pi}^0_{i,t-1} \right) \\ \overline{\Pi}^1_{i,t} &= \overline{\Pi}^1_{i,t-1} + \lambda \left(\Pi^1_{i,t} - \overline{\Pi}^1_{i,t-1} \right) \end{aligned}$$

and

$$\pi_{i,t} = \pi_{i,t-1} + \lambda \left(\mathbb{I} \left(\overline{\Pi}^1_{i,t} > \overline{\Pi}^0_{i,t} \right) - \pi_{i,t-1} \right).$$

The coefficients for each specification under S^0 and S^1 are updated according to the least square algorithm.

$$\alpha^0_{0i,t} = \alpha^0_{0i,t-1} + \lambda \left(p_{j,t} - \alpha^0_{0i,t-1} \right)$$

and

$$\begin{aligned} \begin{bmatrix} \alpha^1_{0i,t} \\ \alpha^1_{1i,t} \end{bmatrix} &= \begin{bmatrix} \alpha^1_{0i,t-1} \\ \alpha^1_{1i,t-1} \end{bmatrix} + \lambda R^{-1}_{i,t-1} \begin{bmatrix} 1 \\ p^1_{i,t} \end{bmatrix} \left(p_{j,t} - \alpha^1_{0i,t-1} - \alpha^1_{1i,t-1} p^1_{i,t} \right) \\ R_{i,t} &= R_{i,t-1} + \lambda \left(\begin{bmatrix} 1 & p^1_{i,t} \\ p^1_{i,t} & (p^1_{i,t})^2 \end{bmatrix} - R_{i,t-1} \right). \end{aligned}$$

Duopolist i starts period $t+1$ with $(\pi_{i,t}; \alpha^0_{0i,t}; \alpha^1_{0i,t}, \alpha^1_{1i,t}; \overline{\Pi}^0_{i,t}, \overline{\Pi}^1_{i,t})$.

References

[1] John Asker, Chaim Fershtman, and Ariel Pakes. Artificial Intelligence and Pricing: The Impact of Algorithm Design. UCLA, Tel Aviv and Harvard, 2021.

[2] Martino Banchio and Giacomo Mantegazza. Adaptive Algorithms and Collusion via Coupling. Stanford University, 2022.

[3] Zach Y. Brown and Alexander MacKay. Competition in Pricing Algorithms. Michigan and Harvard, 2021.

[4] Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–97, October 2020.

[5] In-Koo Cho and Kenneth Kasa. Gresham's Law of Model Averaging. *American Economic Review*, 107(11):1–29, 2017.

[6] In-Koo Cho and Akihiko Matsui. Learning Aspiration in Repeated Games. *Journal of Economic Theory*, 124(2), 2005.

[7] In-Koo Cho, Noah Williams, and Thomas J. Sargent. Escaping Nash Inflation. *Review of Economic Studies*, 69:1–40, 2002.

[8] Karsten T. Hansen, Kanishka isra, and Mellesh M. Pai. Algorithmic Collusion: Supra-Competitive Prices via Independent Algorithms. UCSD and Rice, 2021.

[9] Ludo Waltman and Uzay Kaymak. Q-Learning Agents in a Cournot Oligopoly Model. *Journal of Economic Dynamics and Control*, 32:3275–3298, 2008.