Strategic Experimentation with Private Arrival of Information and Competition *

Kaustav Das †

November 6, 2017

Abstract

This paper analyses private learning in a model of strategic experimentation under competition. I consider a two-armed bandit problem with one safe arm and one risky arm. A good risky arm can potentially experience two kinds of arrivals. One is publicly observable and the other is privately observed by the agent who experiences it. The safe arm experiences publicly observable arrivals according to a given intensity. Private arrivals yield no payoff. Only the first publicly observed arrival (in any of the arms) yields a payoff of 1 unit. Players start with a common prior about the quality of the risky arm. I construct a particular symmetric equilibrium. The full information optimal is considered as the benchmark. In the equilibrium constructed, conditional on no arrival along the risky arm, players tend to experiment too much if the prior is higher than a threshold and tend to experiment too little otherwise.

JEL Classification Numbers: C73, D83, O31.

Keywords: Two-armed Bandit, R&D competition, Duplication, Learning

---

*My greatest intellectual debt is to my mentor Kalyan Chatterjee. Thanks are due to Ufuk Akcigit, Pathikrit Basu, Martin Cripps, Robert Evans, Sanjeev Goyal, Vijay Krishna, In-Uck Park, Johannes Horner, Sven Rady, Larry Samuelson, Qingmin Liu, Hamid Sabourian, Caroline Thomas, Sonje Richie and Venky Venkateswaran for the helpful discussions with them. The remaining errors are my own. A substantial portion of the initial draft was prepared during my visit to Cambridge-INET, Faculty of Economics, University of Cambridge. I thank Hamid Sabourian and Sanjeev Goyal for their warm hospitality and also the University of Exeter Business School for allowing me to spend this time in Cambridge.

†Department of Economics, University of Exeter Business School; email: daskaus-tav84@gmail.com; K.Das@exeter.ac.uk
1 Introduction

This paper addresses the effect of private learning on the non-cooperative behaviour of players in a game of strategic experimentation under competition.

The trade-off between exploration and exploitation is faced by economic agents in many real-life situations. The two-armed bandit model has been extensively used in the economics literature to address this issue formally. This stylized model depicts the situation when an economic agent repeatedly chooses between alternative avenues (which are called arms in the formal analysis) to experiment along, with the ultimate objective to maximize the expected discounted payoff. In the course of experimenting along an arm, the agent upgrades the likelihood it attributes to the arm being capable of generating rewards. In the present work, I study a variant of the standard exponential two-armed bandit model (with one safe and one risky arm) which has both informational externalities and competition and we have private learning along the risky arm. In this two-armed bandit model, an agent not only learns from his experimentation but also learns from the experimentation experiences of others. This gives rise to informational externalities. On the other hand, in the model considered in this paper, only the first player to experience a reward can successfully convert it into a meaningful payoff. This feature captures the competition among the agents. In addition to these, I incorporate private learning by a player who experiments along the risky arm. This means that when a player experiments along the good risky arm, then apart from experiencing the reward (which is publicly observable), it also experiences private signals. A private signal does not yield any payoff, but it completely resolves the uncertainty to the player who experiences it. This is because private signals can be experienced only along a good risky arm. With these features in the current model, I construct a particular symmetric non-cooperative equilibrium. I show that compared to a full information benchmark (a social planner’s problem who observes everything and controls the actions of both the players), in the equilibrium constructed, conditional on no arrival there is too much experimentation along the risky arm if the players start with too high prior (probability that the risky arm is good) and too little experimentation if they start off with a low prior.

The situation described above can be observed in the pharmaceutical industry. Pharma companies operate in a competitive environment and carry out R&D activities to become the first inventor of a new drug. In the McKinsey study by Booth
and Zemmel, it is found that many companies, for clinical trials incline to opt for the novel targets discovered from the human genome project and computational analysis methods. Hence, companies are shifting their discovery portfolio more towards these riskier alternatives, and they decided to abandon the risk-adjusted systematic project development processes. One possible reason can be that the science behind these riskier candidates is very novel and quite attractive. Also, until the final development is done, these companies have many pre-clinical trials and clinical trials whose results are private to them. Results of these trials provide a signal to the companies about the probability of the successful innovation of the drug. In this paper, we analyse a similar situation in a stylised model where private signals are completely informative.

The incidence of private learning before the final success of innovation is also relevant in the academic world. In this context, the proof of Fermat’s last theorem by Sir Andrew Wiles is worth mentioning. This theorem was proved through a span of seven years in complete secrecy and in between Wiles did not reveal any of the interim findings. This may be because once an interim finding is revealed, nothing stops a competitor to use that interim result. This might result in competitor becoming the first one to make the final innovation. In this paper, we show that in a particular equilibrium, although agents explicitly do not reveal their interim private findings, it is implicitly conveyed through equilibrium behaviour.

The setting adopted in this paper has two homogeneous players, both of whom can access a common two-armed exponential bandit in continuous time. One of the arms is safe (S), and the other one is risky (R). A risky arm can either be good or bad. A player can either choose the risky arm which means he experiments or can choose the safe arm which means he exploits. When a player chooses the safe arm, he experiences the arrival of rewards according to a Poisson process with intensity $\pi_0 > 0$. The arrival of a reward is publicly observable. On the other hand, if a player experiments, then conditional on the risky arm is good, he can experience two kinds of arrivals. First, is the arrival of rewards which is according to a Poisson process with intensity $\pi_2 > \pi_0$. In addition to this, independent of the arrival of rewards, the player can experience informational arrival which follows a Poisson process with intensity $\pi_1 > 0$. An informational arrival is private to the player who experiences it. Only the first publicly observable arrival yields a positive payoff of 1 unit. Each player can observe the action of the other. They start off with a common prior $\rho$,
the probability with which the risky arm is good. As the game moves on, each player updates his beliefs as per his private arrival, the publicly observable arrivals and the actions of the players.

We first obtain the efficiency benchmark or the full information optimal of this model. This corresponds to the situation when both the players are controlled by a social planner, who can observe all arrivals experienced. Hence, both the players and the planner share a common belief about the state of the risky arm. The planner at each instant allocates each player to an arm. As soon as there is a publicly observable arrival, the experimentation ends. If one of the players experiences an informational arrival, then all uncertainties are resolved and both the players thereon are allocated to the risky arm (which, in fact, is now found to be good). The solution is of the threshold type. There exists a threshold belief $p^*$ such that conditional on no arrival, both players are made to experiment arm if $p > p^*$ and to choose the safe arm otherwise.

Next, we consider the non-cooperative game. We construct a particular symmetric equilibrium when the intensity of private arrival is lower than a threshold. In this particular equilibrium, on the equilibrium path, given the same information, actions will be identical across players. Hence, if the players start with a common prior, then on the equilibrium path both players would be experimenting if the prior exceeds a threshold $p^{*N}$. If initially, a player starts experimenting then conditional on observing nothing, it switches to the safe arm if the posterior is less than or equal to $p^{*N}$. Since the players are homogeneous and their actions are identical on the equilibrium path, players’ posterior, although private, will be identical across them. If a player, while experimenting experiences an informational arrival, then it keeps on experimenting as long the game continues. As stated, if initially a player starts experimenting and gets no arrival till the belief hits $p^{*N}$, then it switches to the safe arm. However, if it observes that its competitor has not switched, then it reverts to the risky arm again. This is because the action of the competitor gives the player a signal that an informational arrival has been experienced at the risky arm and thus it is good. If such an event occurs and the competitor switches to the safe arm after some time, then the player who had reverted to the risky arm would also follow suit, conditional on experiencing no informational arrival in between. This deters a player from not to switch to the safe arm when it is supposed to.

Having described the full information optimal and a non-cooperative equilibrium,
we try to analyse the nature of inefficiency in the equilibrium constructed. We observe that \( p^* N > p^* \). However, this will not help us to determine the nature of inefficiency in the non-cooperative interaction if there is any. This is because in the benchmark case, the beliefs are public and in the non-cooperative case, the beliefs are private. Moreover, the belief updating processes are different. In the non-cooperative game, movement of beliefs is sluggish. Hence, to determine the nature of inefficiency, we adopt a different method as follows.

First, for each initial prior, at which the planner makes both players to experiment, we calculate the duration for which the players are made to experiment, conditional on no observation. Then we compare this with the duration for which the firms would be in the risky arm in the equilibrium described above for the non-cooperative game, given the same prior.

If the prior is in the range \((p^*, p^* N)\), it is trivially true that in the constructed equilibrium, the duration for which the players experiment along the risky arm is less than that a planner would have wanted. This is because in this case, the duration of experimentation in the constructed equilibrium is 0. For priors exceeding \( p^* N \), we can identify a threshold \( p_0^* \in (p^* N, 1) \). If the initial prior is higher (lower) than this threshold, the duration for which the players experiment along the risky arm in the constructed equilibrium is higher (lower) than that a planner would have wanted. Hence, too much optimism results in excessive experimentation along the risky arm.

**Related Literature:** This paper contributes to the currently nascent literature on private learning in models of experimentation. Some recent related papers on this are Akcigit and Liu(2015) [1], Bimpikis and Drakopoulos(2014)[3], Guo and Roesler(2017)[9] and Heidhues, Rady and Strack (2015)[10]) and Dong (2017) [6].

Akcigit and Liu (2015) [1] analyse a two-armed bandit model with one risky and one safe arm. The risky arm could potentially lead to a dead end. Thus, private information is in the form of bad news. Inefficiency arises from the fact that there are wasteful dead-end replication and an early abandonment of the risky project. In the current paper, private information is in the form of good news about the risky arm. However, the present paper shows that there can still be an early abandonment of the risky project if to start with players are not too much optimistic about the quality of the risky line. Further, in the current paper, we have learning even when there is no information asymmetry.
Bimpikis and Drakopoulos (2014) consider a setting in which agents experiment with an opportunity of an unknown value. Information generated by the experimentation of an agent can be credibly communicated to others. They identify an optimal time $T > 0$ such that if agents commit not to share any information up to time $T$ and disclose all available information at time $T$, the extent of free riding is reduced.

Guo and Roesler (2017) study a collaboration model in continuous time. In both the good and the bad state, success can arrive at a positive rate. In the bad state, players may get a perfectly informative signal. Players have to exert costly effort to stay in the game and at any time they have the option to exit and secure a positive payoff from an outside option. Both the probability of success and private learning are directly related to the amount of effort exerted. An increase in the payoff from the outside option increases collaboration among agents.

Heidhues, Rady and Strack (2015) analyse a model of strategic experimentation where there are private payoffs. They take a two-armed bandit model with a risky arm and a safe arm. Players observe each other’s behaviour but not the realised payoffs. They communicate with each other via-cheap talk. Free riding problem can be reduced because of private payoffs, and there are conditions under which the cooperative solution can be supported as a perfect Bayesian equilibrium. The present paper differs from their work in the following ways. Firstly, we have private arrivals of information only. Secondly, players are rivals against each other.

Dong (2017) studies a model of strategic experimentation with two symmetric players where all actions and outcomes are public. However, one of the players is initially better informed about the state of nature. In the current paper, we have a competitive environment, and through a private outcome, both players can get privately informed.

This paper also contributes to the broad literature on strategic bandits. Some of the important papers in this area are Bolton and Harris (1999), Keller Rady and Thomas (2017) analyses a model where a decision maker chooses a stopping time for a project, and she gets private information gradually over time about whether the project will succeed or not. Rosenberg et al. (2013) also analyses a strategic experimentation game where they look at the effect of varying the observability of the experimentation outcomes. The current paper differs from this in three ways. First, there are both observable and unobservable kinds of outcomes. Secondly, the environment is competitive. Lastly, in the current paper, switching between arms is not irrevocable.

The survey by Horner and Skrzypacz (2016) gives a comprehensive picture of the current literature on this.
Cripps(2005) [12], Keller and Rady(2010) [13], Klein and Rady(2011) [16], Klein(2013) [15], Keller and Rady(2015) [14]. In most of these papers, there is under experimentation due to free riding and all learning is public. In the current paper, we show that in the presence of private learning in a competitive environment, depending on the initial prior, there can be both too much and too little experimentation in the same model. Another novel feature of the current paper is that it captures a setting which is both competitive and simultaneously has free-riding opportunities; the free-riding opportunities arise from the communication of private signals through equilibrium actions.

The rest of the paper is organised as follows. Section 2 discusses the Environment formally and the full information optimal solution. Section 3 discusses the non-cooperative game and the nature of inefficiency. Finally, section 4 concludes the paper.

2 Environment

Two players (1 and 2) face a common continuous time two-armed exponential bandit. Both players can access each of the arms. One of the arms is \textit{safe} (S) and the other one is \textit{risky} (R). A player experimenting along a safe arm experiences publicly observed arrivals according to a Poisson process with commonly known intensity $\pi_0 > 0$. A risky arm can either be \textit{good} or \textit{bad}. A player experimenting along a good risky arm can experience two kinds of arrivals. One of these is publicly observable, and it arrives according to a Poisson process with intensity $\pi_2 > \pi_0$. The other kind of arrival is only privately observable to the player who experiences it. It arrives according to a Poisson process with intensity $\pi_1 > 0$. Only the first public arrival (along any of the arms) yields a payoff of 1 unit to the player who experiences it.

Players start with a common prior $p^0$, which is the likelihood they attribute to the risky arm being good. Players can observe each other’s actions. Hence at each time point players update their beliefs using the public history (publicly observable arrivals and the actions of the players).

We start our analysis with the benchmark case, the social planner’s problem. The

\footnote{Das (2017) [5] shows that in a competitive environment with no private learning, we have too much experimentation.}

\footnote{Wong (2017) [21] also analyse competition and free-riding in the same environment. However, he does not have private learning.}
planner is benevolent and can observe all the arrivals experienced by the players. This can also be called the full information optimal.

2.1 The planner’s problem: The full information optimal

In this sub-section, we discuss the optimisation problem of a benevolent social planner who can completely control the actions of the players and can observe all the arrivals experienced by them. This is intended to be the efficient benchmark of the model described above. Before we move on to the formal analysis, we demonstrate the process of belief updating in this situation.

The action of the planner at time point \( t \) is defined by \( k_t \) \((k_t = 0, 1, 2)\). \( k_t \) is the number of players the planner makes to experiment along the risky arm. \( k_t(t \geq 0) \) is measurable with respect to the information available at the time point \( t \).

Let \( p_t \) be the prior at the time point \( t \). If there is no arrival over the time interval \( \Delta > 0 \), it must be the case that none of the players who were experimenting experienced any arrival. This is because the planner can observe all arrivals experienced by the players. Hence using Bayes’ rule we can posit that the posterior \( p_{t+\Delta} \) at the time point \( (t + \Delta) \) will be given by

\[
p_{t+\Delta} = \frac{p_t \exp^{-k_t(\pi_1 + \pi_2)}}{p_t \exp^{-k_t(\pi_1 + \pi_2)} + 1 - p_t}
\]

The above expression is decreasing in both \( \Delta \) and \( k \). Longer the planner has players experimenting without any arrival, more pessimistic they become about the likelihood of the risky arm being good. Also, as more players are experimenting without any arrival, higher is the extent to which the belief is updated downwards.

Let \( dt = \Delta \). As \( \Delta \to 0 \), the law of motion followed by the belief will be given as (we do away with the time subscript from now on):

\[
dp_t = -k(\pi_1 + \pi_2p_t(1 - p_t))
\]

As soon as the planner observes any arrival at the risky arm, the uncertainty is resolved. If it is an arrival which would have been publicly observable in the non-cooperative game, then the game ends. For the other kind of arrival, the planner gets to know for sure that it is a good risky arm and makes both the players to experiment until the first publicly observable arrival.
Let $v(p)$ be the value function of the planner. Then along with $k$, it should satisfy

$$v(p) = \max_{k \in \{0, 1, 2\}} \left\{ (2 - k) \pi_0 dt + kp \left[ \pi_2 dt + \pi_1 \frac{2\pi_2}{r + 2\pi_2} dt \right] + (1 - r dt) \left[ (1 - (2 - k)\pi_0 dt - kp(\pi_1 + \pi_2) dt) \right] (v(p) - v'(p)) kp(1 - p)(\pi_1 + \pi_2) dt \right\}$$

By ignoring the terms of the order $o(dt)$ and rearranging the above we obtain the following Bellman equation

$$rv = \max_{k \in \{0, 1, 2\}} \left\{ 2\pi_0 (1 - v) + k \left[ \left( (\pi_1 + \pi_2)p \left( \frac{\pi_2}{(\pi_1 + \pi_2)} + \frac{\pi_1}{(\pi_1 + \pi_2)} \frac{2\pi_2}{r + 2\pi_2} \right) \right] - (1 - p) v' \right\} - \pi_0 (1 - v) \right\}$$

The solution to the planner’s problem is summarised in the following proposition.

**Proposition 1** There exists a threshold belief $p^* = \frac{\pi_0}{\pi_2 + \frac{2\pi_0}{r + 2\pi_2}}$, such that if the belief $p$ at any point is strictly greater than $p^*$, the planner makes both the players to experiment, and if the belief is less than or equal to $p^*$, the planner makes both the players to exploit the safe arm. The planner’s value function is

$$v(p) = \begin{cases} \frac{2\pi_0}{r + 2\pi_2} p + C(1 - p)[\Lambda(p)] \frac{2\pi_2}{r + 2\pi_2} & : \text{if } p > p^*, \\ \frac{2\pi_0}{r + 2\pi_2} & : \text{if } p \leq p^*. \end{cases}$$

where $\Lambda = \frac{1-p}{p}$ and $C = \frac{2\pi_0}{r + 2\pi_2} \frac{2\pi_2}{r + 2\pi_2} p^* \frac{1}{(1-p)[\Lambda(p)] \frac{2\pi_2}{r + 2\pi_2}}$.

**Proof.**

We first conjecture that the optimal solution is of the type mentioned in the proposition and derive the value function. This implies, for $p > p^*$, $k = 2$ and from \( we obtain

$$v' + \frac{[r + 2(\pi_1 + \pi_2)p]}{p(1-p)2(\pi_1 + \pi_2)} v = \frac{2\pi_2 \left( r + 2(\pi_1 + \pi_2) \right)}{(r + 2\pi_2)2(\pi_1 + \pi_2)} \frac{1}{(1-p)}$$

The solution to this O.D.E is

$$v = \frac{2\pi_2}{(r + 2\pi_2)} p + C(1 - p)[\Lambda(p)] \frac{r}{2(\pi_1 + \pi_2)}$$

where $C$ is the integration constant.

For $p \leq p^*$, $v(p) = \frac{2\pi_0}{r + 2\pi_2}$. This is obtained by putting $k = 0$ in \( since left continuity of $v(p)$ can always be assumed, at $p = p^*$, the value matching condition
is satisfied and we have \( v(p^*) = \frac{2\pi_0}{r + 2\pi_0} \). Also, at \( p = p^* \), the planner is indifferent between making the players to experiment or to exploit. From 1 at \( p = p^* \) we have

\[
(\pi_1 + \pi_2)p^* \left( \frac{\pi_2}{\pi_1 + \pi_2} + \frac{\pi_1}{\pi_1 + \pi_2} \right) = \frac{2\pi_2}{r + 2\pi_2} - v(p^*) - (1 - p)\pi(p^*) = \pi_0(1 - v(p^*)) \tag{4}
\]

Since until any arrival belief can change only downwards, we have \( v'(p^*) = 0 \). Putting the value of \( v(p^*) \) in 4 we have

\[
p^* = \frac{-\pi_0}{\pi_2 + 2\pi_1(\pi_2 - \pi_0)} \tag{5}
\]

Finally, the value of the integration constant \( C \) is determined from the following value matching condition at \( p = p^* \).

\[
C = \frac{-\frac{2\pi_0}{r + 2\pi_0} - \frac{2\pi_2}{r + 2\pi_2}p^*}{(1 - p^*)[\Lambda(p)]^{2(\pi_1 + \pi_2)}}
\]

With standard verification arguments, we can check that the value function implied by the conjectured solution satisfies 1. This completes the proof of the proposition.

- From the expression of \( p^* \), we can see that in the absence of any informational arrival \( (\pi_1 = 0) \), the threshold is just the myopic belief. A higher rate of informational arrival increases the incentive to experiment and hence, the belief up to which the planner makes players experiment goes down. This effect is also dependant on the difference between the public arrival rates across the safe arm and a good risky arm. This is because ultimately any meaningful payoff is obtained only through a public arrival.

The following section describes the non-cooperative game and constructs a particular symmetric equilibrium.

### 3 The non-cooperative game

In this section, we consider the non-cooperative game between the players in the environment specified above. Players can observe each others’ actions. The informational arrival is only privately observable to the player who experiences it. Only the first publicly observable arrival yields a payoff of 1 unit to the player who experiences it.

\[ ^5 \text{For } \pi_1 = 0, \text{ we get back the result of the planner’s problem with homogeneous players in Das (2017)} \]
In the current model, there is inertia in the players’ decision to switch between arms or change the decision to switch arms. Formally this means if a player shifts from one arm to the another, then if he wants to shift back, he cannot do that before a time interval $\eta > 0$. Also, if at an instant a player decides not to switch to another arm, then to reconsider his decision he needs a time of $\eta > 0$. In the next subsection, we will construct a symmetric equilibrium when the inertia goes to zero ($\eta \to 0$).

3.1 Equilibrium

In this subsection, we construct a particular symmetric equilibrium of the game and study the equilibrium behaviour and outcomes when the inertia goes to zero. We suppose that the players start with a common prior about the type of the risky arm. We construct a symmetric equilibrium when the inertia goes to zero where each player conditional on experiencing no arrival, experiments if the probability he attaches to the risky arm being good is higher than a particular threshold. In the non-cooperative game, the belief is private as a player can also privately learn about the quality of the risky arm. However, in equilibrium, as players’ strategies are pure and actions are observable, on the equilibrium path conditional on no arrival players will have the same belief. This is similar to the approach followed by Bergemann and Hege (2005) [2]. If a player experiences a private arrival, then he never switches from the risky arm. If a player does not experience any arrival and shifts to the safe arm at the equilibrium threshold belief and observes the other player to stay back on the risky arm, then the former player infers from this that the other player has experienced an informational arrival.

Assuming the existence of the equilibrium conjectured in the preceding paragraph, we first determine the common threshold belief such that each player conditional on no arrival experiments as long as the belief is higher than this threshold. This is done in the following lemma.

Lemma 1 Suppose the symmetric equilibrium as conjectured above exists. Let $p^{*N}(p)$ be the threshold for the prior $p$, if the belief is higher than this then a player experiments and else he exploits. We have

$$p^{*N}(p) = \frac{\pi_0}{\pi_2 + \frac{\pi_1}{r+2\pi_2}(\pi_2 - \pi_0)\frac{p}{r+\pi_0}}$$
as the inertia $\eta \to 0$

**Proof.**

We assume that players start with a common prior. Suppose $p^*N(p)$ is the threshold belief. Given the prior, if a player does not get any arrival then he can calculate the probability with which the other player has experienced an informational arrival. Let this probability be $q_p$. At $p^*N$, given the other player’s strategy each player is indifferent between staying on the risky arm or to switch to the safe arm. When both players are on the safe arm then each gets a payoff of $v_s = \frac{\pi_0}{r + 2\pi_0}$. Similarly, if it is known with certainty that the risky arm is good and both players experiment, each gets $v_{rg} = \frac{\pi_0 + 2\pi_2}{r + 2\pi_2}$.

If a player decides to stay on the risky arm for an additional duration of $\eta$ then the payoff is

$$\theta_r = (1 - q_p)\{p_2\eta + p_1\eta p(1 - r\eta)(1 - \pi_0\eta) - \frac{\pi_2}{r + 2\pi_2}$$

$$+ (1 - r\eta)[1 - (p_1 + p_2)\eta - \pi_0\eta][\pi_0\eta + (1 - r\eta)(1 - \pi_0\eta - \pi_2\eta)v_s]\} + q_p\frac{\pi_2}{r + 2\pi_2}$$

If the player instead switches to the safe arm, then for the duration $\eta$ his payoff will be

$$\theta_s = (1 - q_p)\{\pi_0\eta + (1 - r\eta)(1 - 2\pi_0\eta)v_s\} + q_p\{\pi_0\eta + (1 - r\eta)(1 - (\pi_0 + \pi_2)\eta)\frac{\pi_2}{r + 2\pi_2}\}$$

At $p = p^*n$, we have $\theta_r = \theta_s$. We are constructing the equilibrium for $\eta \to 0$. This implies $\pi_0\eta + (1 - r\eta)(1 - \pi_0\eta - \pi_2\eta)v_s \approx v_s$ and $\pi_0\eta + (1 - r\eta)(1 - (\pi_0 + \pi_2)\eta)\frac{\pi_2}{r + 2\pi_2} \approx \frac{\pi_2}{r + 2\pi_2}$. After ignoring the terms of the order $\eta^2$, we get

$$(1 - q_p)\{p_2p^*N\eta + p_1p^*N\eta - \frac{\pi_2}{r + 2\pi_2} - (p_1 + p_2)p^*\eta v_s - \pi_0\eta v_s - v_s - r\eta v_s\} + q_p\frac{\pi_2}{r + 2\pi_2}$$

$$= (1 - q_p)\{\pi_0\eta - 2\pi_0\eta v_s + v_s - r\eta v_s\} + q_p\{\frac{\pi_2}{r + 2\pi_2}\}$$

$$\Rightarrow p^*N\{p_2 + p_1 - \frac{\pi_2}{r + 2\pi_2} - (p_1 + p_2)v_s\} = \pi_0 - \pi_0v_s$$

---

In this case, each player’s payoff is calculated as $v_s = \int_0^\infty e^{-rs}\frac{\pi_0}{2\pi_0}\frac{d(1-e^{-2\pi_0s})}{ds}ds = \frac{\pi_0}{r + 2\pi_0}$. 

12
Substituting the value of $v_s$, we get

$$P^* = \frac{\pi_0}{\pi_2 + \frac{\pi_1}{r+2\pi_2}(\pi_2 - \pi_0)\frac{r}{r+\pi_0}}$$

This concludes the proof of the lemma.

The above lemma shows that if there is an equilibrium as conjectured, the common threshold belief is given by $p^*$. We will now formally state the equilibrium strategies, the beliefs (both on the equilibrium path and off the equilibrium path) and derive the value function of the players. Later on, we will establish that the proposed equilibrium is indeed an equilibrium.

To begin with, let $k_i$ be the strategy of player $i$ ($i = 1, 2$). For each player $i$, $k_i$ is a mapping from the set of private beliefs to the set $\{0, 1\}$. $0$ ($1$) implies choosing the safe (risky) arm. Player $i$’s equilibrium strategy is given by

$$k_i = 1 \text{ for } p > p^* \text{ and } k_i = 0 \text{ for } p \leq p^*$$

We now state how beliefs evolve, both on and off the equilibrium path. To do this we first define following things. Let $A$ be the set of action profiles possible in the current game. We have

$$A = \{(R, R); (S, R); (R, S)\}$$

In any action profile, the first element denotes the action of player 1 and the second element is the action of player 2. For each player $i$, define $z_i$ such that $z_i \in \{0, 1\}$. $z_i = 1(0)$ denotes that player $i$ has (not) experienced an informational arrival. Similarly, for each player $i$ define $d_i$ such that $d_i \in \{0, 1\}$. $d_i = 1(0)$ denotes that player $i$ has (not) deviated at least once. Consider player 1. Suppose at an instant, his private belief is $p$. Based on the public observations and private observation, player 1 updates his beliefs as follows. Before we consider ranges of beliefs separately, it needs to be mentioned that as soon as $z_1 = 1$, then the belief of player 1 permanently becomes $p = 1$. This is represented as

$$p^u(a, z_1 = 1, d_1) = 1$$

Since we are constructing a symmetric equilibrium, the analysis will be analogous for the other player.
\[ \forall a \in A, \forall p \text{ and } \forall d_1. \]

First, consider the range \( p > p^*N \). If \( p^u \) is the updated belief then \( p^u((R, S), z_1 = 0, d_1) \) and \( p^u((S, R), z_1 = 0, d_1) \) follows the following law of motion:

\[
dp = -p(1 - p)(\pi_1 + \pi_2) dt
\]

\( p^u((R, R), z_1 = 0, d_1) \) satisfies

\[
dp = -p(1 - p)(2\pi_2 + \pi_1) dt
\]

Next, consider the range \( p \leq p^*N \). For \( a = (S, R) \) and \( a = (R, R) \), we have

\[
p^u(p, a, z_1 = 0, d_1) = 1
\]

and

\[
p^u(1, a, z_1 = 0, d_1) = 0
\]

for \( a = (R, S) \) and \( a = (S, S) \). This means if a player’s belief becomes 1 because of observing the other player not switching, then it falls to zero when he observes the other player at the safe arm. This ensures that a player has no incentive to deviate by not switching to the safe arm when he is supposed to switch. Finally, for \( a = (R, S) \), \( p^u(p < 1, a, z_1 = 0, d_1) \) follows \( dp_i = -(\pi_1 + \pi_2)p_i(1 - p_i) dt \) and for \( a = (R, R) \) \( p^u(p < 1, a, z_1 = 0, d_1) \) follows \( dp_i = -(2\pi_2 + \pi_1)p_i(1 - p_i) dt \).

This completes the description of the beliefs. We will now compute the payoffs of the players in the conjectured symmetric equilibrium. Both players start with a common prior. On the equilibrium path, the actions of the players are symmetric. For \( p > p^*N \), given \( k_j \) (\( j \neq i \)), player \( i \)'s value should satisfy

\[
v_i = \max_{k_i \in \{1, 0\}} \left\{ (1 - k_i)\pi_0 dt + k_ip\left[\pi_2 dt + \pi_1\frac{\pi_2}{r + 2\pi_2} dt\right] + (1-r dt)\left[1-(2-k_1-k_2)\pi_0 dt-[k_1(\pi_1+\pi_2)p dt+k_2\pi_2 p dt]\right][v_i-v_i'p(1-p)[k_i(\pi_1+\pi_2)+k_j\pi_2] dt] + k_j p dt \pi_1 \frac{\pi_2}{r + 2\pi_2} \right\}
\]

Since to player \( i \), \( k_j \) is given, by ignoring the term of the order \( 0(dt) \) and rearranging the above we can say that \( v_i \) along with \( k_i \) satisfies the following Bellman equation
\[ r v_i = \max_{k_i \in \{0, 1\}} \left\{ \left( [\pi_0 (1 - v_i)] + k_i [p (\pi_1 + \pi_2)] \right) - v_i - \pi_0 (1 - v_i) \right\} \]

\[- (1 - k_j) \pi_0 v_i - k_j [p \pi_2 v_i + \pi_2 p (1 - p) v_i'] + k_j p \pi_1 \pi_2 \frac{\pi_2}{r + 2 \pi_2} \] \hspace{1cm} (5)

For \( p > p^* N \), both players experiment. Putting \( k_i = k_j = 1 \) in (5), we get

\[ v_i' + \frac{v_i [r + (\pi_1 + 2 \pi_2) p]}{p (1 - p) (\pi_1 + 2 \pi_2)} = \frac{\pi_2 p [r + 2 \pi_1 + 2 \pi_2]}{(r + 2 \pi_2) p (1 - p) (\pi_1 + 2 \pi_2)} \] \hspace{1cm} (6)

Solving this O.D.E we obtain

\[ v_i = \frac{\pi_2}{r + 2 \pi_2} \frac{[r + 2 \pi_1 + 2 \pi_2] p}{r + \pi_1 + 2 \pi_2} + C_{rr}^m (1 - p) [\Lambda (p)] \frac{r}{\pi_1 + 2 \pi_2} \]

\[ \Rightarrow v_i = \frac{\pi_2}{r + 2 \pi_2} [1 + \frac{\pi_1}{r + \pi_1 + 2 \pi_2}] p + C_{rr}^m (1 - p) [\Lambda (p)] \frac{r}{\pi_1 + 2 \pi_2} \] \hspace{1cm} (7)

where \( C_{rr}^m \) is an integration constant and \( \Lambda (.) \) is as defined before. The integration constant is given by

\[ C_{rr}^m = \frac{[r + 2 \pi_1 + 2 \pi_2] p^{* N}}{(1 - p^{* N}) [\Lambda (p)] \pi_1 + 2 \pi_2} \frac{\pi_0}{r + 2 \pi_0} - \frac{\pi_2}{r + 2 \pi_2} \]

For \( p \leq p^{* N} \), both players get a payoff of \( v_s = \frac{\pi_0}{r + 2 \pi_0} \). We will now establish that the conjectured equilibrium is indeed an equilibrium. The following proposition does this.

**Proposition 2** There exists a threshold \( \pi_1^* > 0 \) such that if \( \pi_1 < \pi_1^* \), the conjectured equilibrium is indeed an equilibrium.

**Proof.**

We will first show that there exists a \( \pi_1^{* 1} > 0 \) such that for all \( \pi_1 < \pi_1^{* 1} \), the payoff for each player is increasing and convex for \( p > p^{* N} \). First we show convexity. From (7), we know that \( v_i \) is convex if \( C_{rr}^m > 0 \). \( C_{rr}^m \) is obtained from the value matching condition at \( p^{* N} \). This implies

\[ C_{rr}^m = \frac{\pi_0}{r + 2 \pi_0} - \frac{\pi_2}{r + 2 \pi_2} [1 + \frac{\pi_1}{r + \pi_1 + 2 \pi_2}] p^{* N} \]

\[ \frac{(1 - p^{* N}) [\Lambda (p^{* N}) \pi_1 + 2 \pi_2]}{1 - p^{* N}} \]

\[ \Rightarrow \]
As \( p \to 0 \), \( p^*N \to \frac{\pi_0}{\pi_2} \) and \( \frac{\pi_2}{r+2\pi_2} \left[ 1+\frac{\pi_1}{r+\pi_1+2\pi_2} \right] \to \frac{\pi_2}{r+2\pi_2} \). Hence, \( C_{rr}^n \to \frac{\pi_0}{r+2\pi_0} - \frac{\pi_2}{r+2\pi_2} p^*N \) > 0. From this, we can infer that there exists a \( \pi_1^{*1} > 0 \) such that for all \( \pi_1 < \pi_1^{*1} \), \( C_{rr}^n > 0 \).

\[ v'_i(p^*N) > 0 \] as long as \( p^*N > \frac{\pi_0}{\pi_2 + \frac{\pi_1}{r+2\pi_2} \left( 2\pi_2 - \pi_0 \right) + \frac{2\pi_0 \pi_1}{r+2\pi_2} \pi_2} \). Please refer to appendix [A] for a detailed proof. As this is true, we have \( v'_i \) to be strictly positive for all \( p > p^*N \).

Next, for each \( p > p^*N \), from (5) we can infer that no player has any incentive to deviate unilaterally if \( p(\pi_1 + \pi_2) \left\{ \left( \frac{\pi_2}{\pi_1 + \pi_2} + \frac{\pi_1}{\pi_1 + \pi_2} \right) - v_i - v'_i(1 - p) \right\} - \pi_0(1 - v_i) \). Since for \( p > p^*N \) both \( k_i = k_j = 1 \), this implies

\[
[r + \pi_0 + \frac{\pi_0 \pi_2}{(\pi_1 + \pi_2)}]v_i \geq \pi_0 + \frac{\pi_0 \pi_2}{(\pi_1 + \pi_2)} - \frac{\pi_2^2 (r + \pi_1 + 2\pi_2) p}{(\pi_1 + \pi_2)(r + 2\pi_2)} \tag{8}
\]

At \( p = \frac{\pi_0}{\pi_2} \), the above condition requires

\[
[r + \pi_0 + \frac{\pi_0 \pi_2}{(\pi_1 + \pi_2)}]v_i \geq \pi_0 - \frac{\pi_1 \pi_2 \pi_0}{(\pi_1 + \pi_2)(r + 2\pi_2)}\tag{9}
\]

\[ \Rightarrow v_i \geq \frac{\pi_0 - \frac{\pi_1 \pi_2 \pi_0}{(\pi_1 + \pi_2)(r + 2\pi_2)}}{[r + \pi_0 + \frac{\pi_0 \pi_2}{(\pi_1 + \pi_2)}]} \]

Since \( \pi_2 > \pi_0 \), we have

\[
\frac{\pi_0}{r + 2\pi_0} - \left[ \frac{\pi_0 - \frac{\pi_1 \pi_2 \pi_0}{(\pi_1 + \pi_2)(r + 2\pi_2)}}{[r + \pi_0 + \frac{\pi_0 \pi_2}{(\pi_1 + \pi_2)}]} \right] = \frac{\pi_0}{r + 2\pi_0} \frac{\pi_2 - \pi_0}{(r + 2\pi_0) r + \pi_0 + \frac{\pi_0 \pi_2}{(\pi_1 + \pi_2)}} > 0
\]

We know that at \( p = p^*N \), we have \( v_i(p^*N) = \frac{\pi_0}{r + 2\pi_0} \). For any positive \( \pi_1 \), we have

\[
\frac{\pi_0}{r + 2\pi_0} > \frac{\pi_0 + \frac{\pi_0 \pi_2}{(\pi_1 + \pi_2)} - \frac{\pi_2^2 (r + \pi_1 + 2\pi_2) p}{(\pi_1 + \pi_2)(r + 2\pi_2)}}{[r + \pi_0 + \frac{\pi_0 \pi_2}{(\pi_1 + \pi_2)}]}\]

at \( p = \frac{\pi_0}{\pi_2} \). Since \( p^*N \to \frac{\pi_0}{\pi_2} \) as \( \pi_1 \to 0 \), there exists a \( \pi_1^{*2} > 0 \) such that for all \( \pi_1 < \pi_1^{*2} \), we have

\[
v_i(p^*N) = \frac{\pi_0}{r + 2\pi_0} \geq \frac{\pi_0 + \frac{\pi_0 \pi_2}{(\pi_1 + \pi_2)} - \frac{\pi_2^2 (r + \pi_1 + 2\pi_2) p^*N}{(\pi_1 + \pi_2)(r + 2\pi_2)}}{[r + \pi_0 + \frac{\pi_0 \pi_2}{(\pi_1 + \pi_2)}]} \tag{10}
\]

Suppose \( \pi_1^* = \min\{\pi_1^{*1}, \pi_1^{*2}\} \). For any \( \pi_1 < \pi_1^* \), we have \( v_i \) to be strictly convex and increasing for all \( p > p^*N \) and condition (10) also holds. From this, we can infer that for all \( p > p^*N \), we have (8) holding good. From here, we can infer that no player
has any incentive to deviate for any $p > p^{*N}$. Finally, given how $p^{*N}$ is evaluated, for any $p < p^{*N}$, given the other player’s strategy, no player will have an incentive to experiment when he has not experienced any private arrival.

This concludes the proof that the conjectured equilibrium is indeed an equilibrium.

The above result can be interpreted intuitively. The equilibrium we construct involved both players playing a threshold type strategy. In our model, we have private learning along the risky arm which reveals the risky arm to be good. In the equilibrium constructed, on the equilibrium path any private signal received by a player is communicated to the other player through his equilibrium action. This means a player can reap the benefit out of an arrival experienced by his competitor. This brings in an aspect of free-riding in an implicit manner. Thus, although our setting is competitive, there is an aspect of implicit free-riding. From the existing literature, we know that in a model of strategic experimentation free-riding is a hindrance to the existence of equilibria where both players use threshold type strategies. In the current model, higher is the value of $\pi_1$, higher is the implicit free-riding effect. This explains why we need the intensity of arrival of information to be less than a threshold for the constructed equilibrium to exist.

### 3.2 Inefficiency in Equilibrium

In this subsection, we discuss the possible distortions that might arise in the equilibrium constructed compared to the benchmark case. We begin this sub-section by observing that $p^* < p^{*N}$. However, because of private learning, one needs to be careful. At this juncture, it must be stated that by just comparing the threshold probabilities of switching ($p^*$ in the planner’s case and $p^{*N}$ in the non-cooperative case) we cannot infer whether we have too much or too-little experimentation in the non-cooperative equilibrium. This is because in the non-cooperative equilibrium, the informational arrival along the good risky arm is only privately observable and hence if the prior is greater than $p^{*N}$ then same action profile would give rise to a different system of beliefs. In the non-cooperative equilibrium the beliefs are private (although same across individuals) and in the benchmark case, it is public. In the present work, we determine the nature of inefficiency in the following manner.

For each prior, we first determine the duration of experimentation, conditional on
no arrival for both the benchmark case and the non-cooperative equilibrium. Then, we say that there is excessive (too little) experimentation in the non-cooperative equilibrium if starting from a prior, conditional on no arrival, the duration of experimentation along the risky arm is higher (lower) in the non-cooperative equilibrium.

The following proposition describes the nature of inefficiency in the non-cooperative equilibrium.

**Proposition 3** The non-cooperative equilibrium involves inefficiency. There exists a $p^0 \in (p^* N, 1)$ such that if the prior $p_0 > p^0$, then conditional on no arrival we have excessive experimentation and for $p_0 < p^0$ we have too little experimentation. By excessive experimentation we mean that starting from a prior the duration for which players experiment along the risky arm is more than that a planner would have liked to.

**Proof.** Let $t^n_{p_0}$ be the duration of experimentation along the risky line by the firms in the non-cooperative equilibrium described above when they start from the prior $p_0$. From the non-cooperative equilibrium described above we know that of the firms start out from the prior $p_0$ then they would carry on experimentation along the risky line until the posterior reaches $p^* N$. From the dynamics of the posterior we know that

$$dp_t = -(\pi_1 + 2\pi_2)p_t(1 - p_t) dt \Rightarrow dt = -\frac{1}{(\pi_1 + 2\pi_2)} \frac{1}{p_t(1 - p_t)} dp_t$$

$$t^n_{p_0} = -\frac{1}{(\pi_1 + 2\pi_2)} \int_{p_0}^{p^* N} \left[ \frac{1}{p_t} + \frac{1}{(1 - p_t)} \right] dp_t$$

$$\Rightarrow t^n_{p_0} = \frac{1}{(\pi_1 + 2\pi_2)} [\log[\Lambda(p^* N)] - \log[\Lambda(p_0)]]$$

Let $t^p_{p_0}$ be the duration of experimentation along the risky line a planner would have wanted if the firms start out from the prior $p_0$. Then from the equation of motion of $p_t$ in the planner’s problem we have

$$dp_t = -2(\pi_1 + \pi_2)p_t(1 - p_t) dt \Rightarrow dt = -\frac{1}{2(\pi_1 + \pi_2)} \frac{1}{p_t(1 - p_t)} dt$$

$$\Rightarrow t^p_{p_0} = \frac{1}{(2\pi_1 + 2\pi_2)} [\log[\Lambda(p^*)] - \log[\Lambda(p_0)]]$$
We have excessive experimentation when \( t_{p_0}^n > t_{p_0}^p \). This is the case when
\[
\frac{1}{\pi_1 + 2\pi_2} [\log[\Lambda(p^*N)] - \log[\Lambda(p_0)]] > \frac{1}{(2\pi_1 + 2\pi_2)} [\log[\Lambda(p^*)] - \log[\Lambda(p_0)]]
\]
\[
\Rightarrow \pi_1 \log[\Lambda(p_0)] < (\pi_1 + \pi_2) \log[\Lambda(p^*N)] - (\pi_1 + 2\pi_2) \log[\Lambda(p^*)]
\]
Let \( \pi_1 \log[\Lambda(p_0)] \equiv \tau(p) \). Since logarithm is a monotonically increasing function and \( \Lambda(p) \) is monotonically decreasing in \( p \). Hence \( \tau(p) \) is monotonically decreasing in \( p \).

First, observe that \( \tau(1) = -\infty \).

The R.H.S can be written as
\[
\pi_1 \log[\Lambda(p^*N)] - (\pi_1 + 2\pi_2) [\log[\Lambda(p^*)] - \log[\Lambda(p^*N)]]
\]
Since \( [\log[\Lambda(p^*)] - \log[\Lambda(p^*N)]] > 0 \), we have
\[
\text{R.H.S} < \pi_1 \log[\Lambda(p^*N)] = \tau(p^*)
\]
Also since \( p^* \in (0, 1) \) and \( \log[\Lambda(p^*)] \) is finite we have the R.H.S satisfying
\[
2(\pi_1 + \pi_2) \log[\Lambda(p^*N)] - (\pi_1 + 2\pi_2) \log[\Lambda(p^*)] > 2(\pi_1 + \pi_2) \log[\Lambda(1)] - (\pi_1 + 2\pi_2) \log[\Lambda(p^*)] = -\infty
\]
These imply that \( \tau(1) < \text{R.H.S} \) and \( \tau(p^*) > \text{R.H.S} \)

Hence \( \exists \ p_0^* \in (p^*N, 1) \) such that for \( p_0 > p_0^* \), \( \tau(p_0) < \text{R.H.S} \) and for \( p_0 < p_0^* \), \( \tau(p_0) > \text{R.H.S} \). Hence if the prior exceeds \( p_0^* \), then there is excessive experimentation along the risky arm and if it is below the threshold there is too little experimentation along the risky arm in the non-cooperative equilibrium.

This concludes the proof of this proposition ■

In the non-cooperative equilibrium, distortion arises from two sources. First, is what we call the *implicit* free-riding effect. This comes from the fact that if a player experiences a private arrival of information, then the benefit from that is also reaped by the other competing player. This is possible here because we construct the equilibrium when the inertia goes to zero. In fact, if information arrival to firms would have been public, then the non-cooperative equilibrium would always involve free-riding. This follows directly from (12). Thus, this implicit free-riding effect
tends to reduce the duration of experimentation along the risky arm.

The other kind of distortion arises from the fact that information arrival is private and the probability that the opponent has experienced an arrival of information is directly proportional to the belief that the risky arm is good. Conditional on no observation, this makes the movement of the belief sluggish. This results in an increase in the duration of experimentation along the risky arm. The effect of distortion from the second (first) source dominates if the common prior to start with is higher (lower). This intuitively explains the result obtained in the above proposition.

4 Conclusion

This paper has analysed a tractable model to explore the situation when there can be private arrival of information. We show that there can be a non-cooperative equilibrium where depending on the prior we can have both too much and too little experimentation. The equilibrium is derived under the assumption that the inertia of players’ action goes to zero. It will be interesting to see how the results change if a player after switching to the safe arm is unable to revert back to the risky arm immediately (fixed positive inertia). Hence switching back to the risky arm is costly. In addition to it, once we introduce payoff from revealing informational arrival, then there might be situations where a player would have incentive to reveal a private observation. These issues will be addressed in my near future research.

References


[6] Dong, M., 2017 “Strategic Experimentation with Asymmetric Information ”, *mimeo, Bocconi University*


Appendix

A

Suppose both players are experimenting when \( p > \bar{p} \). Hence \( v_1 \) will be given by (7) and we have

\[
v'_1 = \frac{\pi_2 (r + 2 \pi_1 + 2 \pi_2)}{(r + 2 \pi_2)(r + \pi_1 + 2 \pi_2)} - C^m_{rr} \frac{\Lambda(p)}{r + \pi_1 + 2 \pi_2} \left[ 1 + \frac{r}{\pi_1 + 2 \pi_2} \right]
\]

Since \( v_1 \) satisfies the value matching condition at \( \bar{p} \), from (7) we obtain

\[
C^m_{rr} = \frac{\pi_0 \frac{r + 2 \pi_1}{r + 2 \pi_0} - \pi_2 \frac{r + 2 \pi_1 + 2 \pi_2}{r + \pi_1 + 2 \pi_2} \bar{p}}{(1 - \bar{p}) \Lambda(\bar{p})} \left[ \frac{1}{\pi_1 + 2 \pi_2} \right]
\]

This gives us

\[
v'_1 = \frac{\pi_2}{r + 2 \pi_2} \left[ \frac{r + 2 \pi_1 + 2 \pi_2}{r + \pi_1 + 2 \pi_2} \right] - \left[ \frac{\pi_0}{r + 2 \pi_0} - \frac{\pi_2}{r + 2 \pi_2} \frac{r + 2 \pi_1 + 2 \pi_2}{r + \pi_1 + 2 \pi_2} \right] \bar{p} \left[ 1 + \frac{r}{\pi_1 + 2 \pi_2} \right]
\]

The numerator of the above term is

\[
\frac{\pi_2 (r + 2 \pi_1 + 2 \pi_2)}{(r + 2 \pi_2)(r + \pi_1 + 2 \pi_2)} (1 - \bar{p}) - \left[ \frac{\pi_0}{r + 2 \pi_0} - \frac{\pi_2}{r + 2 \pi_2} \frac{r + 2 \pi_1 + 2 \pi_2}{r + \pi_1 + 2 \pi_2} \right] \bar{p} \left[ 1 + \frac{r}{\pi_1 + 2 \pi_2} \right]
\]

\[
= \frac{\pi_2 (r + 2 \pi_1 + 2 \pi_2)}{(r + 2 \pi_2)(r + \pi_1 + 2 \pi_2)} (1 - \bar{p}) - \left[ \frac{\pi_0}{r + 2 \pi_0} - \frac{\pi_2}{r + 2 \pi_2} \frac{r + 2 \pi_1 + 2 \pi_2}{r + \pi_1 + 2 \pi_2} \right] \bar{p} \left[ 1 + \frac{r}{\pi_1 + 2 \pi_2} \right]
\]

\[
= \frac{\pi_2 (r + 2 \pi_1 + 2 \pi_2)}{(r + 2 \pi_2)(r + \pi_1 + 2 \pi_2)} \left[ 1 - \frac{\pi_0 (r + \pi_1 + 2 \pi_2) \bar{p}}{(r + 2 \pi_0)(\pi_1 + 2 \pi_2)} \right]
\]
$\nu'(\bar{p})$ is positive if

$$\frac{\pi_2(r + 2\pi_1 + 2\pi_2)}{(r + 2\pi_2)(\pi_1 + 2\pi_2)} - \frac{\pi_0}{(r + 2\pi_0)} \frac{r + (\pi_1 + 2\pi_2)\bar{p}}{(\pi_1 + 2\pi_2)\bar{p}} > 0$$

$$\Rightarrow \bar{p} \left[ \frac{\pi_2(r + 2\pi_1 + 2\pi_2)}{(r + 2\pi_2)} - \frac{\pi_0}{r + 2\pi_0}(\pi_1 + 2\pi_2) \right] > \frac{r\pi_0}{(r + 2\pi_0)}$$

$$\Rightarrow \bar{p} \left[ \frac{\pi_2(r + 2\pi_1 + 2\pi_2)(r + 2\pi_0) - \pi_0(\pi_1 + 2\pi_2)(r + 2\pi_2)}{(r + 2\pi_2)(r + 2\pi_0)} \right] > \frac{r\pi_0}{(r + 2\pi_0)}$$

$$\Rightarrow \bar{p} \left[ \frac{r\pi_2(r + 2\pi_2) + r\pi_1(2\pi_2 - \pi_0) + 2\pi_0\pi_1\pi_2}{(r + 2\pi_2)} \right] > r\pi_0$$

$$\Rightarrow \bar{p} > \frac{\pi_0}{\pi_2 + \frac{\pi_1}{r + 2\pi_2} [2\pi_2 - \pi_0] + \frac{2\pi_0\pi_1\pi_2}{(r + 2\pi_2)r}} \equiv \bar{p}'$$