Underinvestment and Market Structure*

by
Volker Nocke
STICERD, London School of Economics

The Toyota Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel 0207 955 7719

Discussion Paper
No.EI/22
February 1998

* This paper is based on chapter 2 of my Ph.D. thesis submitted at the London School of Economics. I am grateful to John Sutton for his advice and encouragement. I would also like to thank Markus Brunnermeier, Adam Ostaszewski, Martin Peitz, Sönje Reiche, Patrick Rey, Rani Spiegler and Jean Tirole for helpful comments. Moreover, I am grateful to seminar participants at various universities and conferences. Financial support by the European Commission (Marie Curie Fellowship) is gratefully acknowledged.

Abstract

This paper analyses a dynamic game of investment in R&D or advertising, where current investments change future market conditions. It investigates whether under-investment can be supported in equilibrium by the threat of escalation in investment outlays. When there are no spillovers, or there is full patent protection, underinvestment equilibria are shown to exist even though, by deviating, a firm can get a persistent strategic advantage. When there are strong spillovers and weak patent protection, underinvestment equilibria fail to exist. This implies that weaker patent protection can actually lead to *more* investment in equilibrium. Furthermore, potential entry is introduced into the model so as to address issues of market structure. It is shown that underinvestment equilibria can be stable with respect to further entry, independently of market size and entry cost. Finally, the "nonfragmentation" result of statis stage games (Shaked and Sutton, 1987) is proved to hold in this dynamic game. That is, fragmented outcomes cannot be suppoorted in any equilibrium, no matter how large the market, and despite the existence of under-investment equilibria.

Keywords: Dynamic game, investment, collusion, industry structure.
JEL classifications: L13, D43, O31.

1 Introduction

A central theme in the literature on investment is whether firms have sufficient incentives to invest. For instance, the rationale for patent protection is to give innovating firms the "right" incentives to engage in R&D. In oligopolistic markets, firms usually invest in order to gain a competitive advantage over their rivals. Because of this "business stealing effect", noncooperative investment levels tend to be higher than the level that maximises firms' joint profits. In a dynamic model of investment, the following question arises therefore naturally: 'Is it possible that firms "underinvest" in equilibrium only because investing more would lead to an escalation of investment outlays by rival firms?' In other words, 'Do tacitly collusive "underinvestment equilibria" exist?'

When analysing this question, there is a subtle but important distinction to be made between infinitely repeated games (supergames) and truly dynamic investment games. The theory of supergames, which has been applied mostly to price or quantity setting games, is well developed. Since oligopolistic interaction has the underlying structure of a prisoner's dilemma game, we know from the Folk Theorem that in supergames tacitly collusive equilibria always exist for a discount factor sufficiently large. By contrast, truly dynamic investment games, in which current actions change future payoffs, are not yet very well understood. In particular, the existence of tacitly collusive (underinvestment) equilibria is not obviously ensured. The reason is that, by deviating, a firm might change future market conditions and, thereby, gain a persistent strategic advantage over its rivals. In their paper on investment in capacity, Fudenberg and Tirole (1983) have given an example of the existence of underinvestment equilibria in dynamic investment games. However, in their continuous-time framework, they have muted, by construction, the above distinction in that a firm cannot leapfrog its rivals by deviating.

In R&D- and advertising-intensive industries, *endogenous* industry dynamics play a particularly important role in that current investments in product or process innovation, or in "goodwill", change not only current but also future market conditions. Since investments in R&D or advertising are sunk, these investments have a commitment value. This obviously suggests to model dynamic competition in R&D or advertising as a truly dynamic investment game rather than as an infinitely repeated game since, in the latter, tangible market conditions are assumed to be stationary.

In this paper, we explore the incentives of firms to invest, and to collude, in a dynamic (infinite horizon) game of investment in R&D or advertising. The focus is not on the dynamics of investments as such but rather on the commitment value of investments. We investigate, in particular, the issue of existence of tacitly collusive underinvestment equilibria when firms, by deviating, can get ahead of their rivals and, thereby, gain a considerable strategic advantage.

In our model, the existence of underinvestment equilibria depends crucially on the presence of spillover effects in the appropriation of the benefits from investment. When there are no spillovers or, alternatively, there is full patent protection, underinvestment equilibria exist as long as the investment cost function is sufficiently elastic, and the discount factor sufficiently large. However, when there are strong spillovers and no patent protection, underinvestment equilibria fail to exist, even for discount factors arbitrarily close to unity. This implies that a weakening in the degree of patent protection can actually lead to *more* investment in equilibrium. The reason is that firms have less incentives to invest when they cannot fully appropriate the benefits, and this reduction in the incentives to invest destroys the mechanism through which underinvestment can be supported in equilibrium. Our model thus casts doubt on the effectiveness of complete patent protection in fostering investment. This should be of particular concern since, as we show, underinvestment unambiguously reduces welfare.

The issue of existence of underinvestment equilibria in our model raises an important question for the analysis of market structure. The static version of our dynamic investment model satisfies the "nonconvergence property" (see Shaked and Sutton (1987) and Sutton (1991)): as market size becomes large, free entry does *not* lead to a fragmentation of the market.[1] In the limit when market size tends to infinity, the market share of the largest firm is bounded away from zero. This result is based on the "escalation mechanism". The larger the size of the market, the greater are the returns accruing to a firm from raising its investment outlays. This implies, under some conditions, that increases in market size are associated with a rising level of firms' investment outlays. In the limit, at least one and at most a finite number of firms will find it worthwhile to engage in an escalation of R&D or advertising spendings to capture a positive market share. Hence, concentration remains bounded away from zero, no matter how large the market. The escalation mechanism has been successfully tested by Sutton (1991) in his seminal work on advertising-intensive industries.[2]

However, the nonconvergence property has been obtained almost solely in static stage-game models. The open question is whether this result still holds in dynamic models. In a static model, the way to prove the nonconvergence property is to show that there always exists a profitable deviation for some firm in a large and fragmented market. This deviation consists in a sufficient rise in investment outlays so as to capture a positive market share. In a dynamic model, however, such a single deviation might be followed by a severe (and possibly complex) "punishment"

[1]In Sutton's (1991) terminology, the static version of our model is an "endogenous sunk cost model".

[2]For a recent study on R&D-intensive industries, see Sutton (1998).

strategy by rival firms. What is at issue here is that the existence of underinvestment equilibria (when there are no spillovers) implies that firms do not necessarily engage in an escalation of R&D or advertising spendings precisely because firms will otherwise be punished. But without an escalation mechanism at work, the nonconvergence property breaks down. This is the central question on market structure we address in the present paper. Our result is very reassuring: the nonconvergence property is robust to the existence of underinvestment equilibria in our dynamic investment game.

The plan of the paper is as follows. In section 2, we present the basic two-firm version of our model when there are no spillovers. The equilibrium analysis is given in section 3. This is, in section 4, followed by a comparison of welfare in the collusive underinvestment equilibria and the noncollusive investment equilibrium. In section 5, we introduce spillovers into the model. Then, in section 6, we turn to the analysis of market structure, and investigate whether the two-firm underinvestment equilibria are stable with respect to further entry, independently of market size and entry costs. In the following section, we turn to the central question on market structure: 'Does the nonconvergence property hold in our dynamic game, despite the existence of underinvestment equilibria?' Finally, in section 8, we conclude briefly.

2 The Basic Model

In this section, we present our basic dynamic model without spillovers. There are two firms, each offering a product variety. In each period, firms first decide how much to invest in R&D or advertising. Then, they compete in quantities. Investment is sunk, and persistently raises the consumers' willingness-to-pay for the product variety. By investing more than its rival, a firm can, therefore, get a competitive advantage. Here, we do not allow for entry of a third firm. The topic of potential entry, which is essential for the analysis of market structure, will be taken up in the second part of the paper.

We consider an infinite-horizon game of investment in R&D or advertising. The framework is essentially a dynamic version of the model in Sutton (1991). Time is discrete and indexed by t. There are two firms, $i = 1, 2$, and N consumers, indexed by l. Consumer preferences are defined over a 'quality good', produced in the industry under consideration, and an 'outside good' (or Hicksian composite commodity) whose price and attributes are assumed to be constant. There are two varieties of the quality good on offer, one by each firm. Consumers are assumed to value quality. More specifically, consumer l's utility in period t is given by

$$U^l(x_t^{l,1}, x_t^{l,2}, y_t^l) = \alpha^l \ln\left(\sum_{i=1}^{2} u_t^i x_t^{l,i}\right) + y_t^l \quad (1)$$

if $\sum_{i=1}^{2} u_t^i x_t^{l,i} > 0$, and $U^l(x_t^{l,1}, x_t^{l,2}, y_t^l) = -\infty$ otherwise. We denote by $x_t^{l,i} \geq 0$ and $y_t^l \geq 0$ the quantities consumed of firm i's variety of the quality good and the outside good, respectively; u_t^i is the quality of firm i's offering in period t, and α^l is a parameter, assumed to be strictly positive, that measures the intensity of consumer l's preferences for the quality good. Consumer income in each period is denoted by m^l. We assume $m^l \geq \alpha^l$, for all consumers l; otherwise we allow for arbitrary heterogeneity of consumers in the α's and in income. The quality index is normalised so that the 'basic version' of the quality good is of quality 1, i.e. $u_t^i \geq 1$. Note that, for $x_t^{l,i} > 0$, utility is strictly increasing in quality u_t^i.

In the quality good industry, firm i's period-t-cost of investment in R&D or advertising is given by

$$F(u_t^i; u_{t-1}^i) = F_0 \left(u_t^i\right)^{\beta} - F_0 \left(u_{t-1}^i\right)^{\beta} \tag{2}$$

if $u_t^i \geq u_{t-1}^i$, and $F(u_t^i; u_{t-1}^i) = 0$ otherwise, where $F_0 > 0$ and $\beta > 1$ are parameters that measure the effectiveness of R&D or advertising outlays in raising the consumers' willingness-to-pay. That is, we assume that the effectiveness of R&D or advertising outlays are subject to diminishing returns; for simplicity, we do not consider "adjustment costs". In the case of investment in advertising, the quality of firm i's period-t offering can be interpreted as the stock of firm i's "goodwill" accumulated up to period t.[3] Note that $F(u; u) = 0$; that is, investment costs are zero if a firm does not want to raise the quality of its product. We assume that quality does not depreciate; moreover, $u_t^i \geq u_{t-1}^i$. Both firms have constant and strictly positive marginal costs of production, c, that are independent of quality.

The time structure of the game is as follows. In each period, there are two stages. In the first stage, firms 1 and 2 simultaneously decide whether and how much to invest in quality improvement, and incur the fixed investment outlays. In the second stage, the two firms simultaneously decide how much to produce (quantity competition); consumers, taking price as given, decide how much to consume of each product, and prices are such that markets clear. Firm i's second-stage profit in period t is therefore given by $(p_t^i - c)x_t^i$, where p_t^i and x_t^i are price and quantity, respectively; firm i's total profit in period t is then $(p_t^i - c)x_t^i - F(u_t^i; u_{t-1}^i)$.

Consumers are assumed to maximise the discounted value of per-period-utility, taking the sequence of prices and qualities as given. Since, for simplicity, saving and storing are not allowed, this amounts to consumers maximising per-period-utility myopically. Firms maximise the discounted value of profits. The common discount factor is δ, $0 < \delta < 1$. All parameters of the model, and all moves in past periods and stages, are assumed to be common knowledge.

[3] The goodwill approach to advertising goes back to Nerlove and Arrow (1962).

3 Equilibrium Analysis: Escalation and Underinvestment

In this section, we turn to the equilibrium analysis of the basic model without spillovers. We first show the existence of a noncollusive "investment equilibrium" in which firms engage in an escalation of investment outlays. The central result of this section is developed in subsection 3.3, where we show the existence of a collusive "underinvestment equilibrium". Underinvestment can be supported in equilibrium by the credible threat of escalation in case of deviation, even though, by deviating, a firm can get a persistent strategic advantage over its rival.

In the equilibrium analysis, we confine attention to Markov strategies that depend on the tangible state only; hence, the relevant solution concept is that of Markov perfect equilibrium (MPE). Recall that every MPE is a subgame perfect equilibrium (SPE), even when strategies are *not* restricted to be Markov. The idea of this approach is that history should influence current actions only if it has a direct effect on the current environment, but not because players *believe* that history matters. Furthermore, the state-space approach greatly simplifies the equilibrium analysis; as Shapiro (1989) notes, it allows us to focus on strategic aspects of commitment. (For further justification of the approach, see Maskin and Tirole (1988).)

At each decision node, the state of the industry can be summarised by the current pair of qualities $(u^1, u^2) \in [1, \infty)^2$. Firm i's (pure) Markov action rule at stage 1 in period t is a mapping $s^i : (u^1_{t-1}, u^2_{t-1}) \longmapsto u^i_t$; at the second stage of the same period its action rule is a mapping $t^i : (u^1_t, u^2_t) \longmapsto x^i_t$. The state-space approach has real bite here in that it eliminates all bootstrap-type action rules in the output stage. Since quantity choice at a given stage 2 does not affect future payoff-relevant variables (qualities), the second stage in any given period can be analysed as a *one-shot game*.

3.1 Cournot Competition with Perceived Quality

The important result of this subsection is that, for all pairs of qualities, there exists a unique stage-2 Nash equilibrium in quantities. The associated equilibrium profit is given by equation (6), which will serve, in the remainder of the paper, as a reduced-form stage-2 profit function for the dynamic investment game. Below we present some routine calculations; they can easily be skipped by the reader.

Given qualities and prices, a consumer's optimisation problem in period t (stage 2) can be written as

$$\begin{array}{cl} \max\limits_{\{x^{l,i}\},y^l} & \alpha^l \ln(\sum_i u^i x^{l,i}) + y^l \\ \text{s.t.} & \sum_i p^i x^{l,i} + y^l \leq m^l \end{array}$$

where we have normalised the price of the outside good to one, and dropped time indices for convenience. This programme is equivalent to

$$\max_{y^l} \ \alpha^l \ln\left(\max_i \left\{\frac{u^i}{p^i}\right\}\right) + \alpha^l \ln(m^l - y^l) + y^l.$$

Hence, in equilibrium the quality-price ratio u^i/p^i must be the same for all firms i with positive market share. Solving the first-order condition yields $y^l = m^l - \alpha^l$, which is nonnegative by assumption. Total sales in the quality good industry, S, are therefore equal to

$$S \equiv \sum_{i=1}^{2} p^i x^i = \sum_{i=1}^{2} p^i \sum_{l=1}^{N} x^{l,i} = \sum_{l=1}^{N} (m^l - y^l) = \sum_{l=1}^{N} \alpha^l. \tag{3}$$

Given its rival's price, firm j's price in equilibrium is given by

$$p^j = \frac{u^j}{u^i} p^i \tag{4}$$

for $j \neq i$. Using equation (4) and the definition of S yields firm i's price as a function of firms' quantities:

$$p^i = \frac{S}{x^i + (u^j/u^i)x^j}.$$

Thus, given its rival's quantity, firm i sets x^i so as to maximise

$$x^i \left(\frac{S}{x^i + (u^j/u^i)x^j} - c\right). \tag{5}$$

Remark that this expression is strictly concave in x^i; it is zero at $x^i = 0$, and tends to $-\infty$ as $x^i \to \infty$. Its first derivative at $x^i = 0$ is strictly positive if $S/c > (u^j/u^i)x^j$. Hence, the first-order condition, which can be written as

$$\frac{S}{x^i + (u^j/u^i)x^j}\left(1 - \frac{x^i}{x^i + (u^j/u^i)x^j}\right) - c = 0,$$

gives a unique interior maximum if $S/c > (u^j/u^i)x^j$. Subtracting the two first-order conditions yields $x^i = x^j = x$. Simple calculations then give quantities, prices and profits in the unique stage-2 Nash equilibrium:

$$x = \frac{u^i/u^j}{(u^i/u^j + 1)^2} \frac{S}{c},$$

$$p^i = c\left(\frac{u^i}{u^j}+1\right),$$

and

$$\pi^i(u^1,u^2) = S\left(\frac{u^i/u^j}{u^i/u^j+1}\right)^2 \tag{6}$$

for $i,j = 1,2$, $i \neq j$. Observe that profits in stage 2 depend on the quality *ratio*, and market size, only. Furthermore, equation (6) has the nice and intuitive property that a firm's stage-2 profit is increasing in its own quality, and decreasing in its rival's quality; this differs from models of pure vertical product differentiation.

Above we have *assumed* that both firms will have positive market shares in equilibrium. To show uniqueness, we still have to prove that there does not exist an equilibrium with only one firm making (strictly) positive sales. From expression (5) it can be seen that firm i's unique best reply to any quantity x^j such that $x^j \geq (u^i/u^j)S/c$ is to set its own quantity equal to zero. If only one firm has a positive market share in equilibrium (firm j, say), then its price is given by $p^j = S/x^j$, and its profit by $S - cx^j$, which is strictly decreasing in x^j. Given that firm i sets its quantity x^i equal to zero, firm j therefore wants to set x^j strictly *below* $(u^i/u^j)S/c$. Hence, there is no (pure strategy) Nash equilibrium in quantities such that only one firm has a positive market share.

Finally, remark that, in equilibrium, each consumer is indifferent between the offerings by the two firms. Consumer l's period-t utility, in equilibrium, is therefore equal to

$$\begin{aligned} U^l(u_t^1,u_t^2) &= \alpha^l \ln\left(\alpha^l \frac{u_t^i}{p_t^i}\right) + m^l - \alpha^l \\ &= \alpha^l \ln\left(\alpha^l \frac{u_t^1 u_t^2}{u_t^1+u_t^2}\right) - \alpha^l \ln c + m^l - \alpha^l. \end{aligned} \tag{7}$$

3.2 Dynamic Investment: Escalation

Having solved each period's quantity competition stage, the dynamic game can now be viewed as a simple infinite-horizon investment game in which, in each period, the two firms simultaneously invest in quality, and firm i's payoff in period t is given by

$$\Pi^i(u_t^1,u_t^2) = \pi^i(u_t^1,u_t^2) - F(u_t^i;u_{t-1}^i). \tag{8}$$

Infinite-horizon dynamic games, like the present one, are notoriously difficult to analyse since neither do they have a stationary structure (like infinitely repeated games), nor can they be solved by backward induction (like finite-horizon games).

However, a class of subgame perfect equilibria can be found in our game by first viewing each firm's sequence of investment decisions as a single-player dynamic optimisation problem, holding the quality of the other player fixed. In this way, we can determine a region in the space of qualities (state variables) such that neither firm wants to invest further, *given that its rival will never invest again.* Since this region is associated with "high" quality levels, we can then, in a backward induction fashion, proceed to determine equilibria for subgames starting at "lower" quality levels.

Suppose that the current quality of firm i's offering is given by u_{-1}^i, with $u_{-1}^i \geq 1$. Holding firm j's quality, u^j, fixed forever, firm i's optimisation problem is then given by

$$\max_{\{u_\tau^i\}} \sum_{\tau=0}^{\infty} \delta^\tau \Pi^i(u_\tau^1, u_\tau^2), \tag{9}$$

with $u_\tau^j = u^j$ for all $\tau \geq 0$. Due to the additive separability of the investment cost function, the dynamics are conveniently simple: given that its rival will never invest again, it is optimal for firm i to do all its investment at once, and then cease investing forever.[4] That is, firm i's optimisation problem can be rewritten as

$$\max_{u^i} \frac{S}{1-\delta} \left(\frac{u^i/u^j}{u^i/u^j + 1} \right)^2 - F(u^i; u_{-1}^i),$$

and the optimal sequence of qualities is given by $u_\tau^i = u^*(u^j)$ for all $\tau \geq 0$, where $u^*(u^j)$ denotes the solution to the above problem.[5] Note that $u^*(u^j) \geq u_{-1}^i$ since a firm's stage-2 profit is strictly increasing in its own quality, so that it never pays to reduce quality; that is, sunk investments have commitment value. If $u^*(u^j) > u_{-1}^i$, then $u^*(u^j)$ is the solution to[6]

$$\max_{u^i} \frac{S}{1-\delta} \left(\frac{u^i/u^j}{u^i/u^j + 1} \right)^2 - F_0 \left(u^i \right)^\beta . \tag{10}$$

The concept of a best-reply (or reaction) function is a familiar one in the context of static games. Now, in a dynamic game, a best reply is defined relative to the

[4]We could get more "interesting" dynamics by allowing for "adjustment costs", for instance. However, this would complicate the analysis unnecessarily and not change the qualitative insights.

[5]Remark that firm i's stationary best-reply, $u^*(\cdot)$, depends on firm i's current quality, u_{-1}^i. For notational convenience, we drop this argument.

[6]To see that in this case it is indeed optimal to invest all at once, observe that the dynamic optimization problem (9) can be rewritten in the following way:

$$\max_{\{u_\tau^i\}} \sum_{\tau=0}^{\infty} \delta^\tau \left\{ S \left(\frac{u_\tau^i/u^j}{u_\tau^i/u^j + 1} \right)^2 - (1-\delta) F_0 \left(u_\tau^i \right)^\beta \right\} + F_0 \left(u_{-1}^i \right)^\beta ,$$

and that $u^*(u^j)$ maximizes the expression in curly brackets.

whole sequence of action rules only; the best reply to any given quality level is not well defined since it depends on future actions. What is well defined, however, is a firm's best reply to its rival's quality, given that the rival firm will never invest again. This is exactly how we have constructed $u^*(\cdot)$; we will, therefore, refer to $u^*(\cdot)$ as to the "stationary" best-reply function. In the following, we will characterise both firms' stationary reaction functions; due to symmetry, we can restrict ourselves to firm i's best-reply function.

Lemma 1 *If $\beta \geq 2$, firm i's stationary best reply, $u^*(u^j)$, is given by*

$$u^*(u^j) = \max\left\{\widehat{u}(u^j), u^i_{-1}\right\},$$

where u^i_{-1} is firm i's current quality, and $\widehat{u}(u^j)$ is the unique strictly positive solution to the first-order condition of (10), i.e.

$$\frac{2S}{1-\delta}\frac{\widehat{u}(u^j)u^j}{(\widehat{u}(u^j)+u^j)^3} - \beta F_0\left(\widehat{u}(u^j)\right)^{\beta-1} = 0. \tag{11}$$

(If $\beta = 2$ and $u^j \geq \sqrt{2S/(1-\delta)\beta F_0}$, there is no strictly positive solution to (11). In this case, $\widehat{u}(u^j) = 0$.)[7]

Remark that firm i's "interior" stationary best-reply function, $\widehat{u}(\cdot)$, does not depend on the initial quality u^i_{-1}; this is in contrast to $u^*(\cdot)$. The following lemma is straightforward to show. For all proofs that are not given in the text, the interested reader is referred to the appendix.

Lemma 2 *There is a unique intersection of the two interior (stationary) best-reply curves in $(0,\infty)^2$. This intersection corresponds to a symmetric state, $(\overline{u},\overline{u})$, where $\overline{u}$ is given by*

$$\overline{u} = \left(\frac{S}{4(1-\delta)\beta F_0}\right)^{\frac{1}{\beta}}. \tag{12}$$

Lemma 2 implies that if $(u^1_{-1}, u^2_{-1}) \leq (\overline{u},\overline{u})$, then $(\overline{u},\overline{u})$ is the unique intersection of the two stationary reaction curves. The stationary reaction curves are shown in Figure 1.

We can now define four regions in the space of qualities. In Region 1, $U^{(1)}$, the qualities of both firms are above their respective interior best-replies:

$$U^{(1)} \equiv \left\{(u^1,u^2) \in [1,\infty)^2 \mid u^i \geq \widehat{u}(u^j),\ i,j=1,2,\ i\neq j\right\}.$$

[7]Strictly speaking, the stage-2 reduced-form profit function is not defined for qualities below the minimum quality of 1. For expositional clarity, we extend function (6) to all nonnegative qualities u^i.

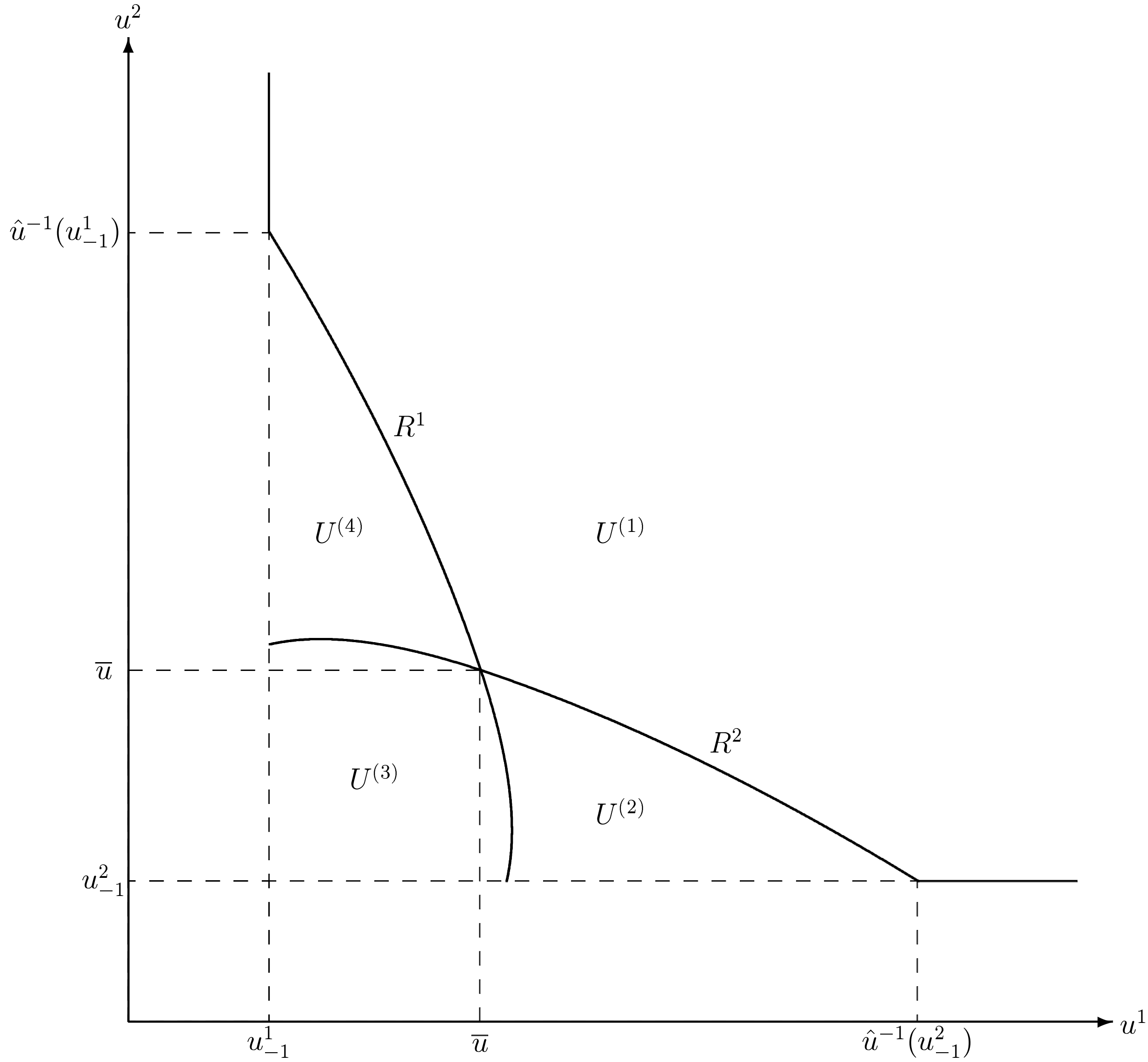

Figure 1: Stationary Reaction Curves

Graphically speaking, this is the region above the outer envelope of the two interior best-reply curves. Region 2 consists of the pairs of qualities such that firm 1's quality is above $\overline{u}$ and firm 2's quality is below its interior best-reply; that is,

$$U^{(2)} \equiv \left\{(u^1, u^2) \in [1, \infty)^2 \mid u^1 \geq \overline{u},\ u^2 < \widehat{u}(u^1)\right\}.$$

Region 4 is defined as Region 2, but firm indices are reversed. Finally, Region 3 encompasses all states that are below the symmetric intersection:

$$U^{(3)} \equiv \left\{(u^1, u^2) \in [1, \infty)^2 \mid u^i < \overline{u},\ i = 1, 2\right\}.$$

We are now in the position to determine a MPE of the dynamic investment game, starting from any state of the industry.

Proposition 1 *The following set of mappings from the current state, (u^1_{t-1}, u^2_{t-1}), to the space of feasible actions, $[1, \infty)$, induces a pure strategy for each firm. The induced strategy profile, Σ^{esc}, forms a MPE starting from any state.*

(i) If $(u^1_{t-1}, u^2_{t-1}) \in U^{(1)}$, then $s^i(u^1_{t-1}, u^2_{t-1}) = u^i_{t-1}$, $i = 1, 2$. ("No investment.")

(ii) If $(u^1_{t-1}, u^2_{t-1}) \in U^{(2)}$, then $s^1(u^1_{t-1}, u^2_{t-1}) = u^1_{t-1}$, and $s^2(u^1_{t-1}, u^2_{t-1}) = \widehat{u}(u^1_{t-1})$. ("Only firm 2 invests.")

(iii) If $(u^1_{t-1}, u^2_{t-1}) \in U^{(3)}$, then $s^i(u^1_{t-1}, u^2_{t-1}) = \overline{u}$, $i = 1, 2$. ("Both firms invest up to $\overline{u}$.")

(iv) If $(u^1_{t-1}, u^2_{t-1}) \in U^{(4)}$, then $s^1(u^1_{t-1}, u^2_{t-1}) = \widehat{u}(u^2_{t-1})$, and $s^2(u^1_{t-1}, u^2_{t-1}) = u^2_{t-1}$. ("Only firm 1 invests.")

Proof. Note that per-period net profits, $\Pi^i(u^1_t, u^2_t)$, are bounded above (by S), and each firm maximises the *discounted* sum of its per-period net profits. This implies that the one-stage deviation principle for infinite-horizon games applies (see Fudenberg and Tirole (1991)): it is impossible to gain by an infinite sequence of deviations when one cannot gain by a single deviation in any subgame.

Observe now that, in each state of the industry, there exists a unique intersection of the two stationary reaction curves. Remark further that, according to Σ^{esc}, the state of the industry will move at once to this unique intersection. Now, recall that the stationary best-reply function, $u^*(\cdot)$, gives the unique best reply, holding the rival firm's quality fixed forever. Furthermore, notice that the unique intersection lies in $U^{(1)}$, so that no firm will invest again *along the equilibrium path*. Thus, by definition of $u^*(\cdot)$, any single deviation that does not induce the nondeviant rival to (dis)invest again cannot be profitable. However, according to Σ^{esc}, a deviation can never induce the nondeviant firm to disinvest since investment costs are sunk. Finally, consider a single deviation that induces the nondeviant rival to invest in the following period. Since stage-2 profits are decreasing in the rival's quality, the deviant's payoff along this path would be smaller than the payoff if the nondeviant did not invest again. But if the nondeviant firm did not invest again, then the deviant's payoff would be even higher by not deviating at all. Hence, such a deviation cannot be profitable. ■

Comparative statics results are easily obtained. Investment along the equilibrium path is weakly increasing in the discount factor, δ, and the size of the market, S, and weakly decreasing in the cost parameters β and F_0.

Remarks. *(1)* Given the current pair of qualities, the stationary reaction curves of the dynamic game converge to the usual reaction curves of the corresponding static stage game as the discount factor δ goes to zero. Hence, as $\delta \to 0$, strategy profile Σ^{esc} converges (in the space of action rules) to the unique Nash equilibrium of the associated static game. That is, the "investment equilibrium" is simply the dynamic version of the static (noncollusive) equilibrium.

(2) The investment equilibrium has another nice property. Let us denote our dynamic investment game with payoff function (6) by Γ. Define the dynamic game Γ' as being equivalent to Γ, but with the following assumption on each firm's information set: starting from an initial state (u^1, u^2), which is common knowledge, each firm observes, in any period, calendar time and its own past moves only; the rival's quality level is unobservable. (This is equivalent to the assumption that, in the initial period, each firm has to "precommit" to the sequence of its future investments. Hence, Γ' is essentially a static game.) It is possible to show that, for any initial state, Γ' possesses a unique Nash equilibrium. Given (u^1, u^2), the unique equilibrium path of Γ' coincides with the equilibrium path induced by Σ^{esc} in the original game Γ. The unique Nash equilibrium of the modified game Γ' is often referred to as the *open-loop* (or *precommitment*) *equilibrium* of the original game Γ; see Fudenberg and Tirole (1991). Notice, however, that the result is due to the absence of adjustment costs in our investment cost function.

(3) Consider the following T-period truncation of the dynamic investment game Γ with payoff function (6): after T periods, $T \geq 1$, firms are restricted to choose the null action "no investment", i.e. $u^i_{t+1} = u^i_t$ for all $t \geq T$. The truncated game Γ^T possesses a unique SPE, which coincides with $\sum^{esc}$ (except for the fact that the SPE of the truncated game depends in a degenerate way on the entire history of the game). For more general investment cost functions with adjustment costs, the unique SPE of Γ^T would converge to $\sum^{esc}$ as $T \to \infty$.

These remarks, especially remark *(3)*, show that $\sum^{esc}$ is the natural noncollusive benchmark equilibrium in the dynamic investment game Γ. In the remainder of the paper, the investment equilibrium will, therefore, serve as the benchmark noncollusive equilibrium. We will call "underinvestment equilibrium" any equilibrium that exhibits less investment along the equilibrium path than this benchmark equilibrium.

3.3 Dynamic Investment: Underinvestment

Along the equilibrium path, induced by strategy profile Σ^{esc} from proposition 1, both firms engage in an "escalation" of R&D or advertising spendings up to the symmetric quality level $\overline{u}$ if the current state is "below" $(\overline{u}, \overline{u})$. In particular, if the current state is (u, u), where $u < \overline{u}$, then the state of the industry will move to $(\overline{u}, \overline{u})$, and stay there forever, even though *both* firms would prefer to stay at (u, u). (Since the stage-2 profit function, (6), depends on the ratio of qualities only, the stage-2 profit is the same in both states but, of course, moving to the higher state involves spending on R&D or advertising.) That is, both firms have an incentive to coordinate not to invest at all in order to avoid an escalation of R&D or advertising outlays, which is wasteful from their point of view.

Since we have established the existence of an MPE exhibiting escalation of investment, it might be possible to support "tacitly collusive" MPE, exhibiting little or no investment, by the threat of escalation in case of deviation. Formally, we consider a strategy profile, denoted by Σ^{coll}, that is induced by the following action rules. If $(u^1_{t-1}, u^2_{t-1}) = (u, u)$, then $s^i(u^1_{t-1}, u^2_{t-1}) = u$; if, however, $(u^1_{t-1}, u^2_{t-1}) \neq (u, u)$, then firms revert to strategy profile Σ^{esc}.

However, it is by no means obvious whether such an underinvestment equilibrium exists. Firstly, suppose the discount factor is (approximately) zero. Then, clearly, a deviation to $\widehat{u}(u)$, where $\widehat{u}(u) > u$ by definition of underinvestment, is profitable since the deviant firm does not care about future costs and stage-2 profits. Hence, by continuity of discounted payoffs in δ, there exists a profitable deviation for discount factors sufficiently small.

Secondly, suppose firm i deviates in period t by investing up to quality level u', $u' > \overline{u}$. According to strategy profile Σ^{coll}, the nondeviant will then, in period $t+1$, react and invest up to $\widehat{u}(u')$, where $\widehat{u}(u') < \overline{u}$; no further investment will occur. (By deviating to $u' \geq \widehat{u}^{-1}(u)$, firm i can even preempt any reaction by its rival.) Along this path, the deviant will make, in *each* period, higher stage-2 profits than in the symmetric underinvestment situation. That is, by deviating, firm i can get ahead of its rival, and ensure that it will always have the higher quality. These additional stage-2 profits have to be compared with the associated investment costs, which occur in period t only. Intuitively then, such a deviation should be profitable for a sufficiently large discount factor. However, the higher is the discount factor, the larger are the returns accruing from investment in R&D or advertising, and hence the higher is the level of investment associated with Σ^{esc}. That is, the larger is δ, the more expensive it is for the deviant firm to ensure itself a persistent strategic advantage over its rival.

The following proposition gives one of the main results of this paper.

Proposition 2 *If $\beta \geq 2$, and the discount factor is sufficiently large, underinvestment equilibria exist. In particular, suppose the current state is given by $(u_{t-1}^1, u_{t-1}^2) = (u, u)$, where the quality level u is arbitrary. Then there is a threshold discount factor $\widehat{\delta} \in (0,1)$, such that for all $\delta \in (\widehat{\delta}, 1)$, the path $(u_\tau^1, u_\tau^2) = (u, u)$, for all $\tau \geq t$, can be supported as a MPE, namely by strategy profile Σ^{coll}.*

It is straightforward to see that, under the conditions of of the proposition, asymmetric underinvestment equilibria exist as well; this is a consequence of stage-2 profits being continuous in qualities.

As we have already argued above, the higher is the discount factor δ, the larger is the increase in the discounted sum of stage-2 profits from deviating to a given quality level u', $u' > \overline{u}$. But with increasing δ, the stationary reaction curves move outwards so that deviating to u' for a given quality *ratio* $u'/\widehat{u}(u') > 1$ becomes more and more expensive. Now, when the investment cost function is sufficiently elastic, i.e. β is sufficiently large, then the cost effect dominates the stage-2 profit effect.[8].

Our results are reminiscent of the existence of "early stopping equilibria" in the dynamic model of investment in capacity by Fudenberg and Tirole (1983).[9] In their continuous-time model, firms face linear investment costs and an exogenous upper bound on the feasible flow of investment at each point in time.[10] These extreme assumptions directly imply that a firm cannot leapfrog its rival by deviating. Moreover, notice that, in a model of capacity investment, gross profits for both firms can be higher at low capacity levels than at high levels; this is due to the fact that competition in quantities is tougher when both firms have higher capacities. (This is an important difference to models of investment in R&D or advertising, where gross profits will, in general, not be higher in lower states; see the remark below.) Hence, in this model, underinvestment equilibria trivially exist for all discount factors. To

[8]Actually, the effect of an increase in β on the profits from deviation is rather subtle; there are two opposing effects. On the one hand, an increase in β makes the deviation to a given u' more expensive; on the other hand, it makes the response of the nondeviant rival less aggressive in that it decreases $\widehat{u}(u')$ for a given u'. For any *given* quality ratio $u'/\widehat{u}(u') > 1$, one can show that profits from deviation are first increasing, and then decreasing, in β.

[9]Reynolds (1987,1991) analyses Fudenberg and Tirole's model in a linear-quadratic differential game framework, where capacity depreciates over time.

[10]That is, Fudenberg and Tirole assume the information lag to be extremely short (zero) relative to the speed of investment, in contrast to our model. Their assumption seems to be more reasonable in the context of capacity investment than in the case of investment in R&D (or advertising).

see this, consider two points in the state space, “A” and “B”, where “A” exhibits lower capacity levels than “B”, but higher gross profits (stage-2 profits). Suppose, moreover, that the noncollusive benchmark equilibrium requires firms to invest from “A” to “B”. Clearly, “no investment” at “A” can be sustained in equilibrium, independently of the discount factor.[11]

Remark. Consider a model of capacity investment such as Fudenberg and Tirole (1983), but allow for more general cost functions and positive detection lags. We then claim the following. *In a model of investment in capacity, if a firm can leapfrog its rival by deviating, and thereby get forever higher gross profits (stage-2 profits), then, for discount factors sufficiently close to unity, such a deviation must be profitable.*[12] Hence, underinvestment cannot be supported. The idea behind this claim is simple. As the discount factor δ goes to one, the stationary reaction curves converge to some “limit curves”. (The reason is that firms have no incentives to build huge excess capacities which are worthless.) Thus, investment levels are bounded from above, and get “dwarfed” by the discounted sum of gross profits as $\delta \to 1$. In fact, Fudenberg and Tirole mostly confine attention to the case $\delta = 1$, where investment costs do not enter firms’ objective functions.

This is in sharp contrast to our model of investment in R&D or advertising, where the toughness of competition in investment levels is essentially independent of the level of investment. This implies that the stationary reaction curves do not converge to some limit curves as $\delta \to 1$, and investment costs do not get dwarfed. Hence, underinvestment can be supported in our model, even though, by deviating, a firm can ensure itself forever higher gross profits. We believe that this difference between investment in capacity and investment in R&D or advertising is an important one. In fact, this difference is closely related to Sutton’s (1991) distinction between exogenous and endogenous sunk cost industries, which is crucial for the analysis of industrial market structure; see section 7.[13]

In the analysis conducted so far, we have left out the important issue of potential

[11] Notice also that Fudenberg and Tirole’s motivation is quite different from ours: they focus on “mobility barriers”. In particular, they investigate the validity of the proposed Stackelberg solution in Spence (1979), where an early entrant in a new market can exploit its head start by strategic investment in capacity.

[12] Implicitly, we assume here that the increase in stage-2 profits is bounded away from zero, no matter how large the discount factor.

[13] The point is the following. If a firm can already serve the whole market with its capacity, any further increase in its capacity has no impact on the firm’s market share; this is the exogenous sunk cost case. In contrast, by outspending its rivals in fixed R&D or advertising outlays, a firm can steal business from its rivals and thus increase its market share, although the investment may not increase industry sales. This is the endogenous sunk cost case.

entry. We will generalise the model so as to allow for potential entry and an arbitrary number of active firms in the second part of the paper, namely in sections 6 and 7. But before turning to the analysis of market structure, we will first analyse welfare in the basic model and, then, introduce spillovers.

4 Welfare Analysis

The aim of this section is to compare "welfare" in the symmetric investment equilibrium and in an arbitrary (symmetric) underinvestment equilibrium. This is of great interest since any action by antitrust authorities is justified only if "tacit collusion" indeed reduces welfare.

As a welfare measure, we choose the sum of discounted profits and discounted utility; we will call this measure "net surplus". In our setting, this choice is natural and theoretically well justified since we use quasilinear preferences for this same reason. In particular, consumer utility is linear in "money" (outside numéraire good), and there are no income effects so that all profits can be redistributed to consumers without changing our analysis.

A priori it is not quite obvious whether or not net surplus is lower in an underinvestment equilibrium. Clearly, consumers' utility, prior to any redistribution of profits, is lower in an underinvestment equilibrium, simply because per-period utility is increasing in quality, and prices depend on the ratio of qualities only. However, firms' profits are unambiguously larger under "tacit collusion" since firms engage in less R&D or advertising. Indeed, from the viewpoint of a social planner, any R&D (or advertising) outlays by a second firm are wasted, holding prices fixed. (A social planner would set price equal to marginal cost, and one (subsidised) firm only would engage in investment and production.)

Nevertheless, the following proposition shows that the welfare comparison is unambiguous. But before stating and proving the proposition, we want to set the problem formally. Given an initial state (u_{-1}^1, u_{-1}^2) and any sequence of states, $\{(u_t^1, u_t^2)\}_{t=0}^{\infty}$, net surplus along this path is equal to

$$\sum_{t=0}^{\infty} \delta^t \left\{ \sum_{i=1}^{2} \Pi^i(u_t^1, u_t^2) + \sum_{l=1}^{N} U^l(u_t^1, u_t^2) \right\}, \tag{13}$$

with $U^l(\cdot)$ and $\Pi^i(\cdot)$ defined in (7) and (8), respectively. This assumes implicitly that at any stage 2 the industry is in equilibrium, depending on the current state. The equilibrium path associated with the symmetric investment equilibrium is $(u_t^1, u_t^2) = (\overline{u}, \overline{u})$ for all $t \geq 0$; in case of a symmetric underinvestment equilibrium it is $(u_t^1, u_t^2) = (u, u)$ for all $t \geq 0$, where, by definition of underinvestment, $\max\{u_{-1}^1, u_{-1}^2\} \leq u < \overline{u}$.

Proposition 3 *In the symmetric investment equilibrium welfare, as measured by (13), is higher than in any symmetric underinvestment equilibrium.*

The intuition behind the result is the following. Even in a second best world where two firms compete à la Cournot and the quality level is constrained to be identical for both firms, the noncollusive symmetric "investment" equilibrium exhibits too low a level of investment; this is true despite the presence of a business stealing effect. The reason is that the duopolists capture only a relatively small part of the surplus from R&D or advertising, and thus invest too little. Consequently, the problem of underinvestment is even more severe in any collusive "underinvestment" equilibrium.

5 Spillovers and Patents

So far we have assumed that a firm's investment cost function, $F(u_t^i; u_{t-1}^i)$, depends on its own quality level only. However, spillover effects are a pervasive phenomenon in many markets. For instance, a firm might be able to copy cheaply its rival's technology. The rationale for patents is, of course, to prevent such free-riding; but, in practice, firms can often "invent around" existing patents. (For a survey on spillovers and R&D, see De Bondt (1996).) Similarly, it might be profitable for a firm to imitate the design and packaging of a rival brand so as to free-ride on the rival's advertising outlays.

To highlight the effects of spillovers on the incentives for firms to invest and to collude, we make the following clearcut assumption. There are no immediate spillovers, but full spillovers after one period. More precisely, a firm can costlessly "copy" its rival's quality of last period. Consequently, $u_t^i \geq u_{t-1}^{\max} \equiv \max\left\{u_{t-1}^1, u_{t-1}^2\right\}$, and firm i's investment cost function (2) is replaced by

$$F(u_t^i; u_{t-1}^{\max}) = F_0 \left(u_t^i\right)^\beta - F_0 \left[u_{t-1}^{\max}\right]^\beta .$$

This can be thought of, for instance, as firms having one-period patent protection when imitation is virtually costless. Alternatively, there might be no patent protection but a time-lag of imitation.[14] Consequently, the payoff-relevant state at the start of period t is now given by $\left(u_{t-1}^{\max}, u_{t-1}^{\max}\right)$, i.e. it lies on the 45°-line in the (u^1, u^2)-space. To shorten notation, we denote the state by the scalar $u_{t-1}^{\max}$.

Intuition might suggest that the presence of spillover effects makes tacit collusion "easier" (supportable for lower discount factors) since, in the long run, firms will

[14] This is consistent with there being many paths that lead to a given quality level: by investing in R&D, a firm discovers its own path; but a firm has also the option to copy its rival's path which is protected for only one period.

always end up in a symmetric state. Hence, by deviating to a quality level above the symmetric investment quality, $\overline{u}$, a firm can no longer ensure itself higher stage-2 profits *ad infinitum* than in any (symmetric) underinvestment equilibrium. The proposition below shows, however, that the opposite result holds.

For expositional clarity, let us proceed along the lines of our earlier analysis. In the presence of spillovers, firm i's stationary reaction function has to be defined differently since $u_0^i \geq u_{-1}^{\max}$. Firm i's stationary best reply to u_0^j, $u^*(u_0^j)$, is now the solution to the following dynamic optimisation problem:

$$\max_{\{u_\tau^i\}} \sum_{\tau=0}^{\infty} \delta^\tau \Pi^i \left(u_\tau^1, u_\tau^2\right),$$

where $u_\tau^j = u_{\tau-1}^{\max}$ for $\tau \geq 1$, and $\Pi^i\left(u_\tau^1, u_\tau^2\right) = \pi^i\left(u_\tau^1, u_\tau^2\right) - F(u_\tau^i; u_{\tau-1}^{\max})$. The interior stationary best reply, $\widehat{u}(u_0^j)$, is defined as the unique positive solution[15] to the following programme:

$$\max_{u_0^i} \; S\left(\frac{u_0^i/u_0^j}{u_0^i/u_0^j+1}\right)^2 - F_0(u_0^i)^\beta + F_0[u_{-1}^{\max}]^\beta + \frac{\delta}{1-\delta}\frac{S}{4}, \tag{14}$$

where the last term is the discounted sum of stage-2 profits from $\tau = 1$ onwards that arise when both firms offer the same qualities. Clearly, the stationary best reply is given by $u^*(u_0^j) = \max\{\widehat{u}(u_0^j), u_{-1}^{\max}\}$.[16] As in the case without spillovers, one can show that there is a unique intersection of the interior (stationary) best-reply curves, namely at $(\overline{u}, \overline{u})$. The corresponding symmetric quality level, $\overline{u}$, is now

$$\overline{u} = \left(\frac{S}{4\beta F_0}\right)^{\frac{1}{\beta}}.$$

Note that this symmetric quality level equals the one without spillovers, as given by (12), when the discount factor is zero.

Since the payoff relevant state at the start of each period lies on the 45°-line in the state space, we have to modify the definitions of strategy profiles Σ^{esc} and Σ^{coll}. The noncollusive benchmark strategy profile $\Sigma^{esc'}$ is now induced by the following set of action rules:

(i) If $u_{t-1}^{\max} < \overline{u}$, then $s^i(u_{t-1}^{\max}) = \overline{u}$, $i = 1, 2$.

(ii) If $u_{t-1}^{\max} \geq \overline{u}$, then $s^i(u_{t-1}^{\max}) = u_{t-1}^{\max}$, $i = 1, 2$.

[15] We assume here, as before, that $\beta \geq 2$. If $\beta = 2$ and $u_0^j \geq \sqrt{2S/\beta F_0}$, then $\widehat{u}(u_0^j) = 0$.

[16] Notice that firm i's stationary best reply, $u^*(u_0^j)$, now depends not on u_{-1}^i, but on $u_{-1}^{\max}$. As before, we drop the additional argument for notational convenience.

Analogously to Σ^{coll}, the collusive strategy profile $\Sigma^{coll'}$ is defined as follows. If $u_{t-1}^{\max} = u$, then $s^i(u_{t-1}^{\max}) = u$, $i = 1, 2$; otherwise firms revert to $\Sigma^{esc'}$.

We can now state and prove another main result of our paper.

Proposition 4 *In the presence of spillovers, there exists a unique MPE, which is given by the noncollusive benchmark equilibrium $\Sigma^{esc'}$. That is, underinvestment cannot be sustained in equilibrium. In particular, the collusive strategy profile $\Sigma^{coll'}$ does not form an MPE.*

The intuition for proposition 4 is as follows. The existence of spillover effects reduces each firm's (noncooperative) incentive to invest, given its rival's quality. This implies that any "noncollusive" investment equilibrium, as supported by strategy profile $\Sigma^{esc'}$, exhibits low quality levels in the long run, relative to the case without spillovers. But any underinvestment equilibrium can only be enforced by the credible threat of escalation. In the presence of spillover effects, however, this threat is rather blunt.

Abstracting from strategic issues, in a world where costless imitation is possible, the individual incentives to get ahead of one's rival are exactly the same as in a "myopic" world with or without spillovers, where the discount factor is equal to zero. Now, clearly, if the discount factor were zero, underinvestment could not be an equilibrium outcome since, by definition of underinvestment, there are always short-run gains from some suitable deviation. However, introducing spillovers in our model is not equivalent to reducing the discount factor (to zero). To see this, notice that the optimal "myopic" deviation from state $u < \overline{u}$ is to invest up to quality $\widehat{u}(u)$. Suppose that, indeed, one firm deviates to $\widehat{u}(u)$ in, say, period t. If $\widehat{u}(u) < \overline{u}$, then, in period $t+1$, *both* firms will invest further, namely up to quality level $\overline{u}$. That is, the optimal myopic deviation requires to invest in both periods t and $t+1$, whereas the gain in stage-2 profits is confined to period t. Hence, when $\delta > 0$, the optimal deviation in the presence of spillovers is, by continuity, not identical to the optimal myopic deviation. Furthermore, it is a priori not obvious whether, for *large* δ, a profitable deviation exists at all. (Notice, however, that the case $\widehat{u}(u) < \overline{u}$ for $u < \overline{u}$ would not arise if qualities were *global* strategic substitutes.)

Let us compare the equilibrium investment level when there are spillovers to the investment level when there are no spillovers. Clearly, if firms do not collude in the latter case, then the investment level is higher than in the case with spillovers, holding fixed all parameters. But if firms do underinvest in the absence of spillovers, then the equilibrium investment level can be higher in the presence of spillovers. To see this, suppose the current state is given by u and choose parameters such that $\overline{u} > u$ in the presence of spillovers. Then, if the discount factor is sufficiently close to unity, there exists an equilibrium in the absence of spillovers such that no firms raises its quality level above u. Another interesting comparison is the following. Suppose the discount factor δ is such that, in the absence of spillovers, underinvestment can be supported in equilibrium. Now, if the size of the market in the presence of spillovers is $1/(1-\delta)$ times the market size in the absence of spillovers, then the quality level $\overline{u}$ is the same in both cases. However, proposition 4 implies that underinvestment cannot be supported in the presence of spillovers. In particular, a deviation from a quality level u, $u < \overline{u}$, to $\overline{u}$ is profitable if spillovers are present and firms use strategy profile $\Sigma^{coll'}$, but unprofitable if there are no spillovers and firms use Σ^{coll} – *although the path induced by the deviation is the same in both cases.* The reason is that market size was assumed to be larger in the presence of spillovers so that the deviant's gain in the period of deviation is larger.

Our result has important implications for the literature on patents. A recurrent theme throughout the whole literature is that patents give firms higher incentives to invest in R&D, and will hence result in higher equilibrium levels of investment. Now, in a world where technological spillovers are present, one can interpret our model without spillovers as representing the case of infinite patent length and breadth, while the extension with spillovers corresponds to the case of short patent length. As we have shown, a shorter patent length can lead to higher R&D in equilibrium simply because it *reduces* the incentives to invest, and hence destroys the mechanism through which underinvestment can be supported. In light of the welfare analysis conducted in the last section, this suggests that, for any given discount factor, there exists an "optimal" patent length that gives maximal incentives to invest but is just short enough so as to prevent firms from colluding in investment.

Remark. In this section, we have focussed on spillovers in the appropriation of the benefits from investment in R&D or advertising. In particular, we have assumed that spillovers are asymmetric in that technological laggards (or weak brands) profit from the investments of technological (or brand) leaders but *not* vice versa. This is a natural way of modelling spillovers in the present setup, and captures exactly what patent protection is about.

This differs from the way how spillovers are modelled in the literature on R&D cartels and joint ventures in the tradition of d'Aspremont and Jacquemin (1988) and Kamien, Muller, and Zang (1992). In this literature, process innovation is modelled as a static two-stage game. Spillovers directly affect the innovation process and are assumed to be immediate and symmetric: the innovation process of a technological leader benefits as much from the current investments of a technological laggard as the laggard can free-ride on the leader's current effort. When products are substitutes and there are strong positive spillovers, the "cooperative equilibrium" exhibits *higher* investment in R&D than the noncooperative Nash equilibrium, while the opposite result holds when spillovers are negative, or positive but weak. Notice that in these static models, the joint profit maximising "cooperative equilibrium" is *not* a Nash equilibrium.

In a recent paper, Kesteloot and Veugelers (1995) have analysed the standard two-stage model of this literature when it is infinitely repeated. In this setting, tacitly collusive SPE always exist for large discount factors. However, as we have already argued in the Introduction, the supergame framework is not very appropriate for modelling strategic interaction in investment since it fails to capture the commitment value of investment. Kesteloot and Veugelers focus on the question of how the threshold discount factor, $\widehat{\delta}$, above which tacit collusion can be sustained, varies with the magnitude of spillovers. They show that when the strength of the spillover effect is sufficiently high, then an increase in the magnitude of the spillover leads to a rise in $\widehat{\delta}$, i.e. collusion becomes "more difficult" to sustain. This is somewhat in line with our results. However, in the case of strong positive spillovers, welfare in the collusive equilibrium is higher than in the noncollusive one since it exhibits higher levels of investment; this is in contrast to our model. The intuition for their result is the following: the larger are the spillovers, the stronger are the incentives for a firm to invest less and to free-ride on the nondeviant's R&D expenditures. In our model, in contrast, it is always the nondeviant laggard that can free-ride in the following period on the deviant leader.

6 Potential Entry

We now turn to the analysis of market structure. In particular, we take up the issue of potential entry that has so far been kept aside in the analysis. The question is whether the high profits the incumbents make while underinvesting will trigger new entry. For this purpose, we extend the basic model by introducing an additional stage in each period at which further entry can occur.

By postulating a sufficiently high entry cost, the modeller could always ensure that it is not profitable for a new firm to enter the market. But for a given entry cost, entry would then still occur in sufficiently large markets. There is, however, an endogenous mechanism which might deter entry: the incumbents' threat of escalation. The aim of this section, therefore, is to investigate whether or not this threat of escalation is credible, and whether it successfully deters entry, no matter how large the market.This question is of interest for two reasons. First, it relates to the robustness of our two-firm underinvestment equilibrium. Second, it addresses a fundamental issue in the theory of market structure, namely whether or not concentration can be high in large markets.

The basic model is modified as follows. There are three players: the incumbents, firms 1 and 2, and a potential entrant, firm 3. In each period, there are now three stages. At stage 1, the potential entrant decides whether to enter or not if it has not yet decided to do so. If firm 3 decides to enter, it has to pay an entry fee ("setup cost") $\epsilon > 0$. At stage 2, the firms that are present in the market (the two incumbents, and firm 3 if it has decided to enter in this period or before) decide simultaneously whether and how much to invest in R&D or advertising. The potential entrant starts up with "zero" quality; its investment cost function in the period of entry is given by $F^e(u) = F_0 u^\beta$, and in all subsequent periods by (2). There are no spillovers. Finally, at stage 3, firms compete simultaneously in quantities. Consumers' utility is given by the natural extension of (1) to three varieties of the quality good. As before, all past actions are assumed to be common knowledge.

The equilibrium analysis proceeds along the lines of section 3. In period t, the state of the industry is given by the quality triple $(u_t^1, u_t^2, u_t^3) \in \Re^3$, where we adopt the convention that $u_t^3 = -1$ if firm 3 has not yet entered the market, and $u_t^3 = 0$ if firm 3 has entered the market but not yet invested in quality. A pure investment action rule is a mapping $s^i : (u_{t-1}^1, u_{t-1}^2, u_{t-1}^3) \mapsto u_t^i$; a pure output action rule is a mapping $t^i : (u_t^1, u_t^2, u_t^3) \mapsto x_t^i$. As before, the minimum quality (in order to make positive sales) is equal to one; therefore, the initial investment outlays necessary to produce the basic version of the quality good are equal to F_0.

As to the equilibrium analysis of the output stage, it is straightforward to show that there exists a unique pure strategy Nash equilibrium in quantities, given any state (u_t^1, u_t^2, u_t^3). If firm 3 has not yet entered the market, or not invested, then its stage-3 profit is zero, and the incumbents' equilibrium profits are given by (6). Otherwise, firm i's stage-3 equilibrium profits are given by

$$\pi^i(u_t^1, u_t^2, u_t^3) = \begin{cases} S\left(\frac{\sum_{k=1}^3 u_t^i/u_t^k - 2}{\sum_{k=1}^3 u_t^i/u_t^k}\right)^2 & \text{if } \sum_{k=1}^3 \frac{u_t^{\min}}{u_t^k} \geq 2 \\ S\left(\frac{u_t^i/u_t^j}{u_t^i/u_t^j + 1}\right)^2 & \text{if } \sum_{k=1}^3 \frac{u_t^{\min}}{u_t^k} < 2 \text{ and } u_t^i, u_t^j > u_t^{\min} \ (\ i \neq j) \\ 0 & \text{if } \sum_{k=1}^3 \frac{u_t^{\min}}{u_t^k} < 2 \text{ and } u_t^i = u_t^{\min}, \end{cases} \tag{15}$$

where $u_t^{\min} = \min\{u_t^1, u_t^2, u_t^3\}$.[17] Hence, in the three-firm equilibrium there exists a "quality window" such that a firm makes zero sales if its quality is too low relative to its rivals' qualities. But there will always be at least two firms making positive sales in equilibrium; this explains why we did not find any quality window in the two-firm case. Observe that $\pi^i(u_t^1, u_t^2, u_t^3)$ is continuous in all its arguments, despite the quality window.

The resulting subgame *after* entry of firm 3 can, in principle, be analysed analogously to the two-firm investment game, given the stage-3 profit function (15). How-

[17] For a general proof of the n-firm case, see the proof of Lemma 3 in the Appendix.

ever, entry is endogenous and might be deterred by the incumbents. We do not attempt here to investigate the three-firm case comprehensively. Rather, we focus on the question whether or not the two incumbents can be in a two-firm underinvestment equilibrium, and successfully deter entry by credibly threatening to engage in an escalation of R&D or advertising outlays in case of entry. The following proposition summarises our results.

Proposition 5 *There exists a $\widehat{\beta} > 2$ such that if $\beta \in [2, \widehat{\beta}]$, any two-firm underinvestment equilibrium is stable with respect to entry by a third firm. In particular, for δ sufficiently large, there exists a MPE such that $(u_\tau^1, u_\tau^2, u_\tau^3) = (u, u, -1)$ for all τ, with $u \leq \overline{u}$. This is true independently of market size and entry costs.*

We have thus shown that the same mechanism that supports underinvestment in equilibrium can be sufficient to deter further entry. The proposition illustrates that concentration can be high even in very large markets. Remark that, if $\beta \in [2, \widehat{\beta}]$, the (symmetric) investment equilibrium is also stable with respect to entry by a third firm but, of course, even without the threat of further escalation.

7 Market Structure and Nonconvergence

In the early literature on industrial market structure, the alleged negative relationship between market size and concentration, even though noted by some authors, has not received much attention. From a theoretical viewpoint, such a negative relationship was considered to be quite obvious: for a given level of "barriers to entry", an increase in market size should raise the profitability of incumbent firms, and thus trigger new entry, which would lead to a fall in concentration. However, the empirical evidence from cross-sectional studies was found to be rather weak.

It is only quite recently that the size-structure relationship has become a major focus of research. In his landmark book, "Sunk Costs and Market Structure", Sutton (1991) shows that the alleged size-structure relationship breaks down in certain groups of industries. In particular, Sutton makes the important distinction between "exogenous" and "endogenous" sunk cost industries. In exogenous sunk cost industries, the only sunk costs involved are the exogenously given setup costs; R&D and advertising outlays are insignificant. In endogenous sunk cost industries, on the other hand, the equilibrium level of sunk costs is endogenously determined by firms' investment decisions. Roughly, these are industries in which advertising or R&D "works" in that investments in some fixed outlays raise the consumers' willingness-to-pay (or reduces marginal costs of production). Sutton's predictions are that in exogenous sunk cost industries, the lower bound to concentration tends to zero as the market becomes large, whereas in industries for which the endogenous sunk cost model applies, the lower bound to concentration is bounded away from zero.[18]

The "fragmentation" result for exogenous sunk cost industries can be illustrated by reference to a simple two-stage game. At stage 1 ("entry stage"), firms decide simultaneously or sequentially whether or not to enter the industry. If they decide to do so, they have to pay an entry fee $\epsilon > 0$. At stage 2 ("output stage"), the firms that have entered the market compete in prices or quantities, according to some static oligopoly model. Firm i's stage-2 equilibrium payoff can be summarised by some reduced-form profit function $S\pi^i(n(S,\epsilon))$, where S denotes market size, and $n(S,\epsilon)$ the number of entrants. It is assumed that the number of potential entrants, $n_0(S,\epsilon)$, is sufficiently large; that is, $n_0(S,\epsilon) > n(S,\epsilon)$ ("free entry"). For a wide class of standard oligopoly models describing competition in the output stage, the equilibrium number of firms in the market tends to infinity as the market becomes large, i.e. $n(S,\epsilon) \to \infty$ as $S \to \infty$, and the market share of each firm converges to zero.[19]

In the endogenous sunk cost model, there is a further stage in which active firms make sunk investments in, say, R&D or advertising. The resulting game consists of three stages: the entry stage, the investment stage, and the output stage. The "nonfragmentation" or "nonconvergence" result states that, under some general conditions, the market share of the largest firm is bounded away from zero in *any* equilibrium. In some models a stronger result obtains: the number of active firms remains finite in the limit when $S \to \infty$. The reason is that, as the market becomes large, firms engage in an escalation of investment outlays which makes it increasingly expensive for rivals to capture a positive market share. Sutton calls this the "escalation mechanism".[20]

However, the nonconvergence result has been obtained almost solely in *static* stage games.[21] According to Sutton (1998), the open question is whether it still

[18] This result follows from an exercise in comparative statics with respect to market size. Notice that it is *not* assumed that market size increases *over time*.

[19] For another class of models (which allow, for instance, for multiproduct firms), multiple equilibria are endemic. The same model may permit fragmented equilibria, in which, for example, each firm offers one product, and concentrated equilibria, in which a single firm is crowding out the product space. The more general fragmentation result refers, therefore, to the *lower bound* to concentration. This has been dubbed the "bounds approach" to concentration. For a discussion, see Sutton (1991).

[20] For a precise statement of the conditions under which the nonconvergence property holds, see Shaked and Sutton (1987) and Sutton (1991). In the case of pure vertical product differentiation and price competition, the finiteness result has been first obtained by Shaked and Sutton (1983).

[21] Two exceptions are in Sutton (1998) and Hole (1997). Hole uses the Pakes-McGuire algorithm to simulate a stochastic dynamic model with incremental sunk costs. However, market size (and hence the average number of entrants) is kept small. Sutton analyses a rather special setting with

holds in dynamic investment games like ours. What is at issue is that the existence of underinvestment equilibria in our dynamic game implies that firms do not necessarily engage in an escalation of R&D or advertising outlays; but without an escalation mechanism at work, the nonconvergence property cannot hold. To make this point clear, let us consider the following example. Suppose that, for a given market size S, there exists a symmetric underinvestment equilibrium in which all $n(S, \epsilon)$ active firms offer quality u in each period, where $n(S, \epsilon)$ is such that any additional entrant would make an overall loss. Now, if this underinvestment equilibrium still holds under free entry when the market becomes large, then we are back in an "exogenous sunk cost world", in which each firm has to pay an exogenous setup cost of $\epsilon + F(u)$. Consequently, the nonconvergence result breaks down in this case. (Actually, one could allow quality u to increase with S, and still get that $n(S, \epsilon) \to \infty$ as $S \to \infty$, unless u increases too fast with S.)

Another way of seeing this point is the following. In a static stage game, the nonconvergence property is proved by showing that in a sufficiently large and fragmented market, there always exists a profitable deviation for some firm. This deviation consists in an escalation of fixed R&D or advertising outlays so as to capture a larger share of the market. Now, in a dynamic game such a deviation might not be profitable since it can trigger an escalation of investment spendings by rival firms, which is detrimental for the deviant firm's profit.

As to the result of section 6, this can be seen as an example of nonfragmentation in that two firms are able to deter further entry, no matter how large the market, as long as $\beta \in [2, \widehat{\beta}]$. However, this equilibrium is not unique; there is another equilibrium in which the two firms acquiesce, and further entry takes place.

To address the issue of nonconvergence, we have to modify the basic version of our dynamic investment game. The time structure is as in section 6; that is, there are three stages in each period: entry, investment, and quantity competition. There is an initial period (say, 0) before which there are no active firms, i.e. *all* firms are potential entrants in period 0. Entry costs as well as the investment cost functions for a new entrant and for an incumbent are as in section 6. The consumers' utility function can be generalised in an obvious way to an arbitrary number of firms offering each a variant of the quality good.

As before, the output stage in each period can be analysed as a one-shot game.

Lemma 3 *In any given stage 3, there exists a unique (Markov-)Nash equilibrium in quantities. Suppose there are $n(S)$ active firms. Re-label the firms such that firm 1 offers the highest quality, u^1, and firm $n(S)$ the lowest quality, $u^{n(S)}$. Then, in equilibrium, there is a "quality window" such that firms 1 to $\underline{n}(S)$ only make positive*

spillovers, in which our problem of interest, underinvestment, does not arise.

sales, where $\underline{n}(S)$ *is the maximum integer* z, $z \leq n(S)$, *such that* $\sum_{i=1}^{z}(u^z/u^i) > z-1$. *Firm i's stage-3 equilibrium profit is given by*[22]

$$\pi^i(u^1,...,u^i,...,u^{n(S)}) = \begin{cases} S\left(1 - \frac{\underline{n}(S)-1}{\sum_{j=1}^{\underline{n}(S)}(u^i/u^j)}\right)^2 & \textit{if } i \leq \underline{n}(S) \\ 0 & \textit{otherwise.} \end{cases} \quad (16)$$

Notice that the above labelling is not unique when two or more firms offer the same quality; in this case, however, each of these firms produces the same amount in equilibrium.

Using equation (16) as the reduced-form stage-3 profit function, we can now focus on the analysis of investment strategies. For technical convenience, we restrict attention to equilibria such that all investment, *along the equilibrium path*, occurs in the initial period, when the market opens.[23] This implies, in particular, that the number of active firms remains constant over time. However, we allow for all the "investment", "underinvestment" and "entry deterring" strategies we have considered earlier as well as for much more complex strategies. Of course, we allow strategies and any equilibrium path to depend on market size.

In a dynamic game, the lower bound to concentration for a given market size might be quite different from that in a static game. The open theoretical question is whether or not the asymptotic properties are the same, namely that the lower bound does not converge to zero as market size becomes large. In the section on potential entry, we have already seen that nonconvergence is a *possible* outcome in our model; what is at issue is whether or not it is a *necessary* outcome in that it occurs in all equilibria. The following proposition gives the central result on market structure.

Proposition 6 *The nonconvergence property holds in our dynamic game. In particular, in any MPE, the number of active firms,* $n(S)$, *remains finite as market size tends to infinity.*

The proof of this proposition is rather lengthy, and can be found in the appendix. Here, we just give a sketch of it. We first assume that there exists an equilibrium such that $n(S) \to \infty$ as $S \to \infty$, and then show that this leads to a contradiction. We pick the firm with the lowest quality in equilibrium and calculate an upper bound on its equilibrium profit. Then, we consider a carefully selected deviation for this

[22]For notational convenience, we describe the current state by the quality tuple of *active* firms only.

[23]This assumption is not essential; its only purpose is to pin down an (arbitrary) equilibrium path. Notice also that there is no "technological" reason why firms should not do all investment in the first period.

firm, which is a function of market size and its rivals' qualities. One can easily calculate the deviant's associated profits in the period of deviation. Since we do not make any restrictions on the "punishment strategies", we cannot say much, at this level of generality, about the deviant's profit in the periods *after* deviation. What we do know, however, is that these profits are nonnegative. This is sufficient to show that the deviation is profitable for large markets. Hence, the threat of future escalation does not have enough bite to prevent the deviating firm from escalating its investment outlays so as to capture a larger share of the market. This contradicts the initial assumption.

Our result is reassuring in that it shows the robustness of the nonfragmentation result to the existence of underinvestment equilibria in dynamic games. Remark that we have actually shown a "strong version" of the nonconvergence property to hold: the market share of *all* firms is bounded away from zero, no matter how large the market. Note also that proposition 6 does *not* imply that underinvestment equilibria necessarily break down when market size becomes large, as we have already seen in section 6.

8 Conclusion

In this paper, we have explored a dynamic game of investment in R&D or advertising. It is quite distinct from the applied literature on supergames since, in our model, current investments change future market conditions. From a game-theoretic viewpoint, the model is related to the dynamic game of capacity investment by Fudenberg and Tirole (1983). It differs from their continuous-time framework in that firms can leapfrog their rivals. Therefore, the existence of tacitly collusive equilibria is no longer ensured. In the first part of the paper, we have focussed on the issue of existence of underinvestment equilibria when firms have strong incentives to deviate and, thereby, to persistently improve their strategic position. In the second part, we have introduced potential entry into the model so as to address issues of market structure.

Using a state-space approach, we have shown that when strong spillovers in the appropriation of the benefits from investment are present, underinvestment equilibria fail to exist, while the opposite result holds without spillovers. This implies that a weakening in the degree of patent protection can actually lead to more R&D in equilibrium even though (or, rather, because) it reduces the individual incentives to invest. Furthermore, we have shown that underinvestment should be an issue of concern for antitrust authorities in that it unambiguously reduces welfare. This is especially true since detecting tacit collusion in R&D or advertising is likely to be much more difficult than detecting tacit collusion in price setting.

The existence of underinvestment equilibria has raised the question whether one of the central results on market structure in the I.O. literature, the "nonconvergence property" namely, breaks down in dynamic investment games. What has been at issue is that, in an underinvestment equilibrium, firms do not engage in an escalation of fixed investment outlays; but without an escalation mechanism at work, the nonconvergence property cannot hold. Our main result on market structure is very reassuring: the nonconvergence property is robust to the existence of underinvestment equilibria.

9 Appendix

Proof of Lemma 1. The first-order condition of (10) is given by

$$\varphi\left(u^i \mid u^j\right) \equiv \frac{2S}{1-\delta}\frac{u^i u^j}{(u^i+u^j)^3} - \beta F_0 \left(u^i\right)^{\beta-1} = 0. \tag{17}$$

Now, observe that $\varphi\left(u^i \mid u^j\right) \to -\infty$ as $u^i \to \infty$, $\varphi\left(0 \mid u^j\right) = 0$. Furthermore, for $u^i > 0$:

$$\varphi\left(u^i \mid u^j\right) > 0 \Leftrightarrow \frac{2S}{(1-\delta)\beta F_0} u^j > \left(u^i\right)^{\beta-2}\left(u^i+u^j\right)^3.$$

The l.h.s. of the last inequality is independent of u^i, and strictly positive for $u^j > 0$. If $\beta > 2$, then the r.h.s. tends to zero as $u^i \to 0$. If $\beta = 2$, then in the limit as $u^i \to 0$, the l.h.s. is larger than the r.h.s. if and only if $u^j < \sqrt{2S/(1-\delta)\beta F_0}$. Note also that for $\beta \geq 2$, the r.h.s. is strictly increasing in u^i. Therefore, if $\beta > 2$ (or if $\beta = 2$ and $u^j < \sqrt{2S/(1-\delta)\beta F_0}$), there exists a strictly positive $\widehat{u}(u^j)$ such that $\varphi\left(u^i \mid u^j\right) > 0$ if and only if $u^i < \widehat{u}(u^j)$, and $\varphi\left(u^i \mid u^j\right) < 0$ if and only if $u^i > \widehat{u}(u^j)$. Thus, $\widehat{u}(u^j)$ is the unique strictly positive solution of (17), and hence of (10). (If $\beta = 2$ and $u^j \geq \sqrt{2S/(1-\delta)\beta F_0}$, however, then $\widehat{u}(u^j) = 0$.)

This obviously implies

$$u^*(u^j) = \max\left\{\widehat{u}(u^j), u^i_{-1}\right\}$$

as long as $\beta \geq 2$. ■

Proof of Lemma 2. Suppose there exists an intersection of the two interior stationary best-reply curves at $(u^1, u^2) \in (0,\infty)^2$. By definition, $\widehat{u}(u^1) = u^2$ and $\widehat{u}(u^1) = u^2$. From lemma 1, the associated first-order conditions are given by (11), i.e.

$$\frac{2S}{1-\delta}\frac{u^1 u^2}{(u^1+u^2)^3} = \beta F_0 \left(u^i\right)^{\beta-1},\ i = 1, 2.$$

Since the left-hand side is the same for both firms, it follows immediately that $u^1 = u^2$. Simple calculations show that $u^1 = u^2 = \overline{u}$, as given by (12). ■

Proof of Proposition 2. Since the strategy profile Σ^{esc} forms an MPE, it is sufficient to show that there is no single profitable deviation when the current state is given by $(u^1_{t-1}, u^2_{t-1}) = (u, u)$. The proof is organised as follows. We first seek the optimal deviation for any player (due to symmetry, we can confine attention to an arbitrary firm), and then show that the associated net present value of future profits, Π^{dev}, is not larger than the corresponding value in case of nondeviation, Π^{coll}. We distinguish three cases.

Case *(i)*: Firm 1, say, deviates in period t by raising its quality to u', where $u < u' < \overline{u}$; that is, the state moves to (u', u) in period t. According to strategy profile Σ^{coll}, *both* firms will then invest further in period $t+1$, and the state of the industry will be given by $(u^1_\tau, u^2_\tau) = (\overline{u}, \overline{u})$ for all $\tau \geq t+1$. The associated discounted sum of profits for the deviant is equal to

$$\Pi^{dev} = S\left(\frac{u'/u}{u'/u+1}\right)^2 - (1-\delta)F_0(u')^\beta + F_0 u^\beta + \frac{\delta}{1-\delta}\frac{S}{4} - \delta F_0 \overline{u}^\beta, \tag{18}$$

while in case of nondeviation it is given by $\Pi^{coll} = S/[(1-\delta)4]$. Maximising Π^{dev} with respect to u' gives a first-order condition identical to (11); hence the condition is sufficient for a maximum. Note, however, that the unique positive solution to (11) might be larger than $\overline{u}$. (It is straightforward to show that this is indeed the case when $\overline{u} < (2+\sqrt{5})u$; we are dealing with this case in part (ii) of the proof. Hence, in the following we analyse the case when the reverse inequality holds. By choosing δ sufficiently close to 1 this can always be ensured.) Denote the optimal value of u' by $u'(u)$. Then, from (11), $(1-\delta)F_0[u'(u)]^\beta = (2/\beta)S[u'(u)]^2 u/[u'(u)+u]^3$. Substituting $(1-\delta)F_0[u'(u)]^\beta$ and $\overline{u}$ in (18) gives

$$\Pi^{dev} = S\left(\frac{u'(u)}{u'(u)+u}\right)^2 - \frac{2S[u'(u)]^2 u}{\beta[u'(u)+u]^3} + F_0 u^\beta + \frac{\delta}{1-\delta}\frac{S}{4}\left(\frac{\beta-1}{\beta}\right),$$

which is continuous in δ. Now, multiplying both sides by $(1-\delta)$, and taking the limit as δ goes to one, one gets

$$\lim_{\delta\to 1}(1-\delta)\Pi^{dev} = \frac{S}{4}\left(\frac{\beta-1}{\beta}\right) < \frac{S}{4} = \lim_{\delta\to 1}(1-\delta)\Pi^{coll}.$$

Hence, there exists a $\widehat{\delta}^{(i)} < 1$ such that for all $\delta \geq \widehat{\delta}^{(i)}$ deviation is not profitable.

Case *(ii)*: Suppose now that, in period t, firm 1 deviates to a quality u' such that $\widehat{u}^{-1}(u) > u' \geq \overline{u}$. In period $t+1$, firm 2 will then react and raise its quality to

$u^*(u') = \widehat{u}(u')$, where $u < \widehat{u}(u') \leq \overline{u}$. Hence, the sequence of states induced by the deviation will be given by $(u_\tau^1, u_\tau^2) = (u', u)$ for $\tau = t$, and $(u_\tau^1, u_\tau^2) = (u', \widehat{u}(u'))$ for $\tau \geq t+1$. The deviant's discounted sum of profits is thus equal to

$$\Pi^{dev} = S\left(\frac{u'/u}{u'/u+1}\right)^2 - F_0(u')^\beta + F_0 u^\beta + \frac{\delta}{1-\delta} S\left(\frac{u'/\widehat{u}(u')}{u'/\widehat{u}(u')+1}\right)^2. \tag{19}$$

Maximising this expression with respect to u' yields the first-order condition for optimal deviation:

$$2S\frac{u'u}{[u'+u]^3} - \beta F_0(u')^{\beta-1} + 2S\frac{\delta}{1-\delta}\frac{u'\left[\widehat{u}(u') - u'\frac{d\widehat{u}(u')}{du'}\right]}{[u'+\widehat{u}(u')]^3} = 0, \tag{20}$$

where $\widehat{u}(u')$ is implicitly defined by (11), and $d\widehat{u}(u')/du'$ can be obtained by implicit differentiation of (11):

$$\frac{d\widehat{u}(u')}{du'} = -\frac{\frac{2S}{1-\delta}\frac{\widehat{u}(u')[\widehat{u}(u')-2u']}{[u'+\widehat{u}(u')]^4}}{\frac{2S}{1-\delta}\frac{u'[u'-2\widehat{u}(u')]}{[u'+\widehat{u}(u')]^4} - \beta(\beta-1)F_0[\widehat{u}(u')]^{\beta-2}}.$$

In order to reduce the dimensionality of the problem, let us define u'_λ such that $\widehat{u}(u'_\lambda) = \lambda u'_\lambda$, where $\lambda \in (0,1]$. For a fixed λ, the first-order condition for the nondeviant's best reply to u'_λ, (11), can then be rewritten as

$$\frac{2S}{1-\delta}\frac{u'_\lambda(\lambda u'_\lambda)}{[u'_\lambda + \lambda u'_\lambda]^3} - \beta F_0[\lambda u'_\lambda]^{\beta-1} = 0.$$

Solving for u'_λ gives

$$u'_\lambda = \left(\frac{2S}{(1-\delta)\beta F_0}\frac{1}{\lambda^{\beta-2}(1+\lambda)^3}\right)^{\frac{1}{\beta}}. \tag{21}$$

This enables us to calculate $d\widehat{u}(u')/du'$ locally at $u' = u'_\lambda$, as a function of λ:

$$\left.\frac{d\widehat{u}(u')}{du'}\right|_{u'=u'_\lambda} = -\frac{\lambda(2-\lambda)}{2\lambda - 1 + (\beta-1)(1+\lambda)}, \tag{22}$$

which is strictly negative for $\lambda \in (0,1]$ and $\beta \geq 2$: the higher is the deviant's quality, the less will be invested by its rival.

We can interpret firm 1's "optimal deviation problem" as a choice of λ. The deviant's first-order condition, (20), for the optimal λ, denoted by λ_δ, can now be written as

$$2S\frac{u'_{\lambda_\delta}u}{[u'_{\lambda_\delta}+u]^3} - \beta F_0(u'_{\lambda_\delta})^{\beta-1} + 2S\frac{\delta}{1-\delta}\frac{u'_{\lambda_\delta}\left[\lambda_\delta u'_{\lambda_\delta} - u'_{\lambda_\delta}\left.\frac{d\widehat{u}(u')}{du'}\right|_{u'=u'_{\lambda_\delta}}\right]}{[\lambda_\delta u'_{\lambda_\delta} + u'_{\lambda_\delta}]^3} = 0,$$

where u'_{λ_δ} and $d\widehat{u}(u')/du'|_{u'=u'_{\lambda_\delta}}$ are given by (21) and (22), respectively. Multiplying both sides by $(1-\delta)$, taking the limit as δ goes to one, and simplifying, one gets

$$\xi(\lambda_1) \equiv \lambda_1^\beta + \lambda_1^{\beta-1} - \frac{1+\beta}{\beta}\lambda_1 - \frac{\beta-2}{\beta} = 0, \tag{23}$$

where $\lambda_1 = \lim_{\delta\to 1}\lambda_\delta$. This is the first-order condition for the optimal λ as a function of β in the limit when $\delta \to 1$. Since the sign of the coefficients in (23) changes once if $\beta \geq 2$, Descartes' sign rule tells us that (23) has exactly one (strictly) positive root.[24] Now, $\xi(1) = 1/\beta$, $\xi(0) = (2-\beta)/\beta \leq 0$ if $\beta \geq 2$, and $\xi'(0) < 0$. Hence, if $\beta \geq 2$, there exists exactly one $\lambda_1 \in (0,1]$ such that $\xi(\lambda_1) = 0$. Since an increase in u' corresponds to a *decrease* in λ, and $\xi(0) \leq 0$ and $\xi(1) > 0$, (23) defines indeed a maximum! The optimal choice of λ, in the limit when $\delta \to 1$, is therefore the unique $\lambda_1 \in (0,1]$ satisfying (23).[25] It is straightforward to show that λ_1 is strictly increasing in β,[26] and that $\lambda_1 \to 1$ as $\beta \to \infty$.[27].

Substituting u' in (19) by u'_{λ_1}, as given by (21), and substituting $\widehat{u}(u')$ by $\lambda_1 u'_{\lambda_1}$, multiplying both sides of (19) by $(1-\delta)$, and taking the limit as $\delta \to 1$, yields

$$\begin{aligned}
\lim_{\delta\to 1}(1-\delta)\Pi^{dev} &= \frac{S}{(1+\lambda_1)^2}\left(1 - \frac{2}{\beta\lambda_1^{\beta-2}(1+\lambda_1)}\right) \\
&= \frac{S}{(1+\lambda_1)^2}\left(1 - \frac{2}{\beta+1+(\beta-2)/\lambda_1}\right) \\
&= \frac{S}{(1+\lambda_1)^2}\left(\frac{(\beta-1)\lambda_1+\beta-2}{(\beta+1)\lambda_1+\beta-2}\right) \equiv \widehat{\Pi}^{dev}(\lambda_1,\beta),
\end{aligned}$$

[24] It is straightforward to generalize Descartes' sign rule, which has been developed for polynomials, to the case when the powers are not necessarily integers, but (more generally) rational numbers. To see this, define $\xi(x) \equiv a_0 + a_1 x^{b_1} + ... + a_n x^{b_n}$, where $b_i = p_i/q_i$ and $p_i, q_i \in N$. Suppose q is the smallest common denominator of the b_i's. Then, $\xi(x)$ can be rewritten as a polynomial: $\xi(x) = a_0 + a_1 y^{\widetilde{b}_1} + ... + a_n y^{\widetilde{b}_n}$, where $y \equiv x^{1/q}$ and $\widetilde{b}_i \equiv q p_i/q_i \in N$. As to irrational β's, one can show that, in our case, $\xi(x)$ has exactly one sign change at some positive x for any real (rational or irrational) $\beta \geq 2$.

[25] Here, we abstract from the lower bound on λ_1, which is given by $u/\widehat{u}^{-1}(u)$.

[26] Implicit differentiation of (23) gives

$$\frac{d\lambda_1}{d\beta} = -\frac{(\ln\lambda_1)(\lambda_1^\beta + \lambda_1^{\beta-1}) + (\lambda_1 - 2)/\beta^2}{\beta\lambda_1^{\beta-1} + (\beta-1)\lambda_1^{\beta-2} - (\beta+1)/\beta}.$$

Clearly, the numerator of the r.h.s. expression is negative for $\lambda_1 \leq 1$. As to the denominator, (23) implies that $\lambda_1^\beta + \lambda_1^{\beta-1} - (1+\beta)\lambda_1/\beta = (\beta-2)/\beta \geq 0$ if $\beta \geq 2$, and hence $\beta\lambda_1^{\beta-1} + (\beta-1)\lambda_1^{\beta-2} - (1+\beta)/\beta > \lambda_1^{\beta-1} + \lambda_1^{\beta-2} - (1+\beta)/\beta \geq 0$ if $\beta \geq 2$. That is, the denominator is positive, and hence $d\lambda_1/d\beta > 0$, for $\beta \geq 2$.

[27] To see this, suppose otherwise that $\lambda_1 \to k < 1$ as $\beta \to \infty$. Then, from (23), it follows that $\xi(\lambda_1) \to -k-1 < 0$ as $\beta \to \infty$. But this contradicts the definition of λ_1.

where the second equality follows from the definition of λ_1 in equation (23). Observe that $\partial\widehat{\Pi}^{dev}(\lambda_1,\beta)/\partial\lambda_1 < 0$. To find a suitable lower bound on λ_1, let us define

$$\eta(\lambda) \equiv \lambda^2 - \frac{\lambda}{\beta(\beta-1)} - \frac{\beta-2}{\beta}.$$

If $\beta \geq 2$, there is a unique strictly positive $\widehat{\lambda}_1$ such that $\eta(\widehat{\lambda}_1) = 0$; it is given by $\widehat{\lambda}_1 = (\beta-1)/\beta$. Furthermore, we have $\eta(\lambda) \geq \xi(\lambda)$ for all $\lambda \in (0,1]$, where $\xi(\lambda)$ is defined as in (23). Hence, $\widehat{\lambda}_1 \leq \lambda_1$ and $\widehat{\Pi}^{dev}(\widehat{\lambda}_1,\beta) \geq \widehat{\Pi}^{dev}(\lambda_1,\beta) = \lim_{\delta\to1}(1-\delta)\Pi^{dev}$.

Remark that if $\beta = 2$, then $\widehat{\lambda}_1 = 1/2$, and $\widehat{\Pi}^{dev}(1/2,2) = 4S/27 < S/4 = \lim_{\delta\to1}(1-\delta)\Pi^{coll}$. One can show that the total derivative of $\widehat{\Pi}^{dev}(\widehat{\lambda}_1(\beta),\beta)$ with respect to β is positive: the higher is the elasticity of the investment cost function, the higher is the upper bound on the profits from deviation. Finally note that $\widehat{\lambda}_1(\beta) \to 1$ as $\beta \to \infty$, and thus $\widehat{\Pi}^{dev}(\widehat{\lambda}_1(\beta),\beta) \to_{\beta\to\infty} S/4 = \lim_{\delta\to1}(1-\delta)\Pi^{coll}$. Hence, for all $\beta \geq 2$, $\lim_{\delta\to1}(1-\delta)\Pi^{dev} < \lim_{\delta\to1}(1-\delta)\Pi^{coll}$.

Because of continuity in δ, there exists therefore, for any $\beta \geq 2$, a threshold value $\widehat{\delta}^{(ii)} < 1$ such that for all $\delta \geq \widehat{\delta}^{(ii)}$, deviation is not profitable.

Case *(iii)*: Finally, suppose the deviant firm (firm 1, say) preempts any reaction by its rival. That is, in period t, firm 1 chooses a quality level u' such that $\widehat{u}(u') \leq u$; in the induced subgame, the state of the industry will then be given by $(u_\tau^1, u_\tau^2) = (u', u)$ for all $\tau \geq t$.

Since $\widehat{u}^{-1}(u) > u^*(u)$ (where the inverse of $\widehat{u}(\cdot)$ is defined over the decreasing part of $\widehat{u}(\cdot)$ only), the deviant firm chooses u' such that $\widehat{u}(u') = u$ so that its rival is *just* preempted. That is, the optimal preemptive deviation, u', is implicitly defined by

$$\phi(u') \equiv \frac{2S}{1-\delta}\frac{uu'}{(u+u')^3} - \beta F_0 u^{\beta-1} = 0. \tag{24}$$

Now, $\phi(u) > 0$ if and only if $u < \overline{u}$ (which is, of course, the relevant case of underinvestment, and can always be ensured by choosing δ sufficiently large), $\lim_{u'\to\infty}\phi(u') = -\beta F_0 u^{\beta-1} < 0$, and $\phi'(u') < 0$ for all $u' > u/2$. Thus, if $u < \overline{u}$, there exists a unique u', $u' > u$, such that $\phi(u') = 0$. Define $\psi(u') \equiv u'/(u+u')$, and note that $\psi(u') \in (1/2,1)$, and $\lim_{u'\to\infty}\psi(u') = 1$. Equation (24) can now be rewritten, and solved for u':

$$u' = \left(\frac{2S}{(1-\delta)\beta F_0}\frac{\psi(u')}{u^{\beta-2}}\right)^{1/2} - u.$$

The discounted sum of profits from deviation is then equal to

$$\Pi^{dev} = \frac{S}{(1-\delta)}\left[\frac{\left(\frac{2S}{(1-\delta)\beta F_0}\frac{\psi(u')}{u^{\beta-2}}\right)^{1/2} - u}{\left(\frac{2S}{(1-\delta)\beta F_0}\frac{\psi(u')}{u^{\beta-2}}\right)^{1/2}}\right]^2 - F_0\left[\left(\frac{2S}{(1-\delta)\beta F_0}\frac{\psi(u')}{u^{\beta-2}}\right)^{1/2} - u\right]^\beta + F_0 u^\beta,$$

and, hence,

$$\lim_{\delta\to 1}(1-\delta)\Pi^{dev} = \begin{cases} -\infty & \text{if } \beta > 2 \\ 0 & \text{if } \beta = 2 \\ S & \text{otherwise.} \end{cases}$$

If $\beta \geq 2$, there exists therefore a $\widehat{\delta}^{(iii)} < 1$ such that for all $\delta \geq \widehat{\delta}^{(iii)}$, $\Pi^{dev} < \Pi^{coll}$. ■

Proof of Proposition 3. Instead of directly comparing welfare in the two "classes" of equilibria under consideration, we opt for a more instructive proof. Let us define the symmetric "second best quality" as the common quality level, u^*, that maximises our welfare measure, (13), under the assumption that both firms compete à la Cournot at the output stage. Given any quality level $\widetilde{u}$, with $\max\{u^1_{-1}, u^2_{-1}\} \leq \widetilde{u}$, consider the sequence of quality pairs $(u^1_t, u^2_t) = (\widetilde{u}, \widetilde{u})$ for all $t \geq 0$. From (7), (8) and (12) it follows directly that the sum of discounted consumer utility along this path is given by

$$\sum_{t=0}^{\infty}\delta^t \sum_{l=1}^{N} U^l(u^1_t, u^2_t) = \frac{1}{1-\delta}\sum_{l=1}^{N}\left\{\alpha^l \ln\left(\frac{\alpha^l}{2}\widetilde{u}\right) - \alpha^l \ln c + m^l - \alpha^l\right\},$$

and the sum of discounted net profits by

$$\sum_{t=0}^{\infty}\delta^t \left\{\sum_{i=1}^{2}\Pi^i(u^1_t, u^2_t)\right\} = 2\left[\frac{1}{1-\delta}\frac{S}{4} - F_0\widetilde{u}^{\beta}\right] + F_0\left(u^1_{-1}\right)^{\beta} + F_0\left(u^2_{-1}\right)^{\beta}.$$

Maximising net surplus (i.e. the sum of the two expressions above) with respect to $\widetilde{u}$ yields the following first-order condition:

$$\gamma(\widetilde{u}) \equiv \frac{S}{1-\delta}\frac{1}{\widetilde{u}} - 2\beta F_0 \widetilde{u}^{\beta-1} = 0,$$

where we have used the fact that $\sum_l \alpha^l = S$. It is easy to see that $\gamma'(\widetilde{u}) < 0$ for all $\widetilde{u}$, i.e. net surplus is strictly concave in $\widetilde{u}$. Furthermore, $\gamma(1) > 0$ if $\overline{u} \geq 1$ as assumed, and $\gamma(\widetilde{u}) \to -\infty$ as $\widetilde{u} \to \infty$. Therefore, the second best symmetric quality level is uniquely defined by $\gamma(u^*) = 0$, and is equal to

$$u^* = \left(\frac{S}{2(1-\delta)\beta F_0}\right)^{\frac{1}{\beta}}.$$

Since $u^* > \overline{u} > u$, it follows immediately that welfare is higher in the symmetric investment equilibrium than in any symmetric underinvestment equilibrium. ■

Proof of Proposition 4. The proof proceeds in several steps.

(i) Suppose the current state, $u^{\max}_{-1}$, is "above" $\overline{u}$, i.e. $u^{\max}_{-1} \geq \overline{u}$. Then, at most one firm will invest along the equilibrium path in any given period. To prove this claim, we consider period t, and assume that $u_t^j \leq u_t^i = u_t^{\max}$ along the equilibrium path. It is easy to see that it is optimal for firm j to set $u_t^j = u_{t-1}^{\max}$: the choice of u_t^j does not affect j's payoff in the continuation game as long as $u_t^j \leq u_t^i$, and firm j's best reply from below to $u_t^{\max}$ is given by the solution to

$$\max_{u_t^j \in \left[u_{t-1}^{\max}, u_t^i\right]} S\left(\frac{u_t^j}{u_t^j + u_t^i}\right)^2 - F_o\left(u_t^j\right)^\beta + F_0\left(u_{t-1}^{\max}\right)^\beta,$$

which is equal to $u_{t-1}^{\max}$ since $u_{t-1}^{\max} \geq \overline{u}$.

(ii) We now claim that no firm will invest along the equilibrium path starting from any state $u_{-1}^{\max}$ above $\overline{u}$. To see this, suppose otherwise. From (i), we know that at most one firm invests in any given period. Assume firm i invests in period t (and hence firm j does not), and let $V^i(\cdot)$ denote firm i's value function. Then, firm i's discounted sum of profits (or value) in period t, $V^i\left(u_{t-1}^{\max}\right)$, must satisfy

$$\begin{aligned} V^i\left(u_{t-1}^{\max}\right) &= \max_{u \geq u_{t-1}^{\max}} S\left(\frac{u}{u + u_{t-1}^{\max}}\right)^2 - F_0 u^\beta + F_0\left(u_{t-1}^{\max}\right)^\beta + \delta V^i(u) \\ &\geq S/4 + \delta V^i\left(u_{t-1}^{\max}\right), \end{aligned}$$

where the inequality follows from the fact that i may decide not to invest, in which case $u_t^{\max} = u_{t-1}^{\max}$. This yields

$$V^i\left(u_{t-1}^{\max}\right) \geq \frac{S}{4(1-\delta)}.$$

Since each firm's gross profit is decreasing in its rival's quality, we obtain an upper bound on firm i's profit by assuming that firm j will never invest again in the future. Firm i's stationary best reply to any quality level above $\overline{u}$ is never to invest. Hence, for any $u \geq \overline{u}$,

$$V^i(u) \leq \frac{S}{4(1-\delta)}.$$

Combining the equations gives $V^i\left(u_{t-1}^{\max}\right) = S/\left[4(1-\delta)\right]$, $i = 1, 2$, and $u_t^{\max} = u_{t-1}^{\max}$.

(iii) Suppose now that $u_{t-1}^{\max} \leq \overline{u}$. We claim that $u_t^{\max} \leq \overline{u}$. To see this, suppose $u_t^{\max} \geq \overline{u}$ so that the continuation payoff for both firms is given by $V^i\left(u_t^{\max}\right) = S/\left[4(1-\delta)\right]$, $i = 1, 2$. Assume $\overline{u} \leq u_t^1 = u_t^{\max}$. Then the two value functions must satisfy

$$V^1\left(u_{t-1}^{\max}\right) = \max_{u^1 \geq \overline{u}} S\left(\frac{u_t^1}{u_t^2 + u_t^1}\right)^2 - F_0\left(u_t^1\right)^\beta + F_0\left(u_{t-1}^{\max}\right)^\beta + \delta\frac{S}{4(1-\delta)},$$

and

$$V^2\left(u_{t-1}^{\max}\right)=\max_{u^2\geq u_{t-1}^{\max}} S\left(\frac{u_t^i}{u_t^j+u_t^i}\right)^2-F_0\left(u_t^j\right)^{\beta}+F_0\left(u_{t-1}^{\max}\right)^{\beta}+\delta\frac{S}{4(1-\delta)}.$$

The only tuple (u_t^1,u_t^2) which satisfies both equations is given by $(\overline{u},\overline{u})$. Hence, $u_{t-1}^{\max}\leq\overline{u}$ implies $u_t^{\max}\in\left[u_{t-1}^{\max},\overline{u}\right]$.

(iv) Any equilibrium is symmetric. That is, along the equilibrium path, $u_t^1=u_t^2=u_t^{\max}$ for all t. Above, we have shown that this holds for the unique equilibrium starting from state $u_{-1}^{\max}\geq\overline{u}$. We now extend this result to the case where $u_{-1}^{\max}\leq\overline{u}$. From our earlier analysis, we know that $u_t^{\max}\in\left[u_{t-1}^{\max},\overline{u}\right]$ for all t. We claim that firm i has a profitable deviation if $u_t^i<u_t^j=u_t^{\max}$. Indeed, firm i's quality choice does not affect its continuation payoff provided that it does not invest more than its rival. But firm i's best reply from below is to set its quality equal to the rival's quality since $\widehat{u}(u)\geq u$ for any $u\leq\overline{u}$, , where $\widehat{u}(\cdot)$ is the interior stationary best reply function.

(v) Suppose that $u_{t-1}^{\max}\leq\overline{u}$ and $\widehat{u}\left(u_{t-1}^{\max}\right)\geq\overline{u}$, where $\widehat{u}(\cdot)$ denotes again the interior stationary best reply function. We then claim that $u_s^1=u_s^2=\overline{u}$ for all $s\geq t$. To see this, recall again that we obtain an upper bound on firm i's profit by assuming that firm j will never invest again in the future. Now, if $u_s^j<\overline{u}$, then firm i may deviate to $\widehat{u}\left(u_s^j\right)\geq\overline{u}$. In the equilibrium of the induced subgame, both firms will never invest again. By definition of $\widehat{u}(\cdot)$, the deviation must therefore be profitable.

(vi) Suppose that $u_{t-1}^{\max}\leq\overline{u}$ and $\widehat{u}\left(u_{t-1}^{\max}\right)\leq\overline{u}$. Note that the latter inequality holds if and only if $\overline{u}\geq(2+\sqrt{5})u_{t-1}^{\max}$. We now prove by contradiction that $u_s^1=u_s^2=\overline{u}$ for all $s\geq t$. Assume to the contrary that $u_t^{\max}<\overline{u}$ and consider a period-t deviation by firm 1 to quality level $\overline{u}$. The deviation induces the following sequence of quality levels: $(u_t^1,u_t^2)=(\overline{u},u)$, and $(u_s^1,u_s^2)=(\overline{u},\overline{u})$ for all $s\geq t+1$. The deviation is profitable if and only if

$$\begin{aligned}
& S\left(\frac{\overline{u}/u_t^{\max}}{1+\overline{u}/u_t^{\max}}\right)^2-F_0\overline{u}^{\beta}+F_0\left(u_t^{\max}\right)^{\beta}+\frac{\delta}{1-\delta}\frac{S}{4}>\frac{1}{1-\delta}\frac{S}{4}\\
\Leftrightarrow\quad & S\left(\frac{\overline{u}/u_t^{\max}}{1+\overline{u}/u_t^{\max}}\right)^2>\frac{S}{4}\left(\frac{\beta+1}{\beta}\right)-F_0\left(u_t^{\max}\right)^{\beta}\\
\Leftarrow\quad & \beta>\frac{1}{\left[4\left(\frac{\overline{u}/u_t^{\max}}{1+\overline{u}/u_t^{\max}}\right)^2-1\right]}.
\end{aligned}$$

For a given β, the l.h.s. of the last inequality is independent of $u_t^{\max}$ and $\overline{u}$, while the r.h.s. is strictly decreasing in the ratio $\overline{u}/u_t^{\max}$, for $\overline{u}\geq u_t^{\max}$. Since $\overline{u}\geq(2+\sqrt{5})u_t^{\max}$, the proof is complete if we can show that

$$\beta > \frac{1}{\left[4\left(\frac{2+\sqrt{5}}{1+2+\sqrt{5}}\right)^2 - 1\right]} = 0.61803.$$

But this inequality holds by assumption. ■

Proof of Lemma 3. The analysis of the $n(S)$-firm case proceeds analogously to that of the 2-firm case (see subsection 3.1). Denote by I the set of firms with positive equilibrium market share, i.e. $I \equiv \{i = 1, ..., n(S) \mid x^i > 0\}$. Since each consumer chooses the variant of the quality good with the highest quality-price ratio, all firms with positive sales must exhibit the same quality-price ratio in equilibrium. That is, firm j's equilibrium price is given by

$$p^j = \frac{u^j}{u^i} p^i$$

for $i, j \in I$, $i \neq j$. Using the definition of total sales, $S = \sum_{i \in I} p^i x^i$, one obtains

$$p^j = \frac{u^j S}{\sum_{i \in I} u^i x^i}.$$

Firm j's stage-3 profit can then be written as

$$x^j \left(\frac{u^j S}{\sum_{i \in I} u^i x^i} - c \right).$$

This expression is strictly concave in x^j, equal to zero at $x^j = 0$, and tends to $-\infty$ as $x^j \to \infty$. Its first derivative is strictly positive at $x^j = 0$ if and only if $u^j S > c \sum_{i \in I} u^i x^i$. Thus, the following first-order condition yields a unique interior maximum if $u^j S > c \sum_{i \in I} u^i x^i$:

$$\frac{S}{\sum_{i \in I} u^i x^i} - \frac{x^j u^j S}{\left(\sum_{i \in I} u^i x^i\right)^2} = \frac{c}{u^j}. \tag{25}$$

This gives x^j as a function of "weighted" aggregate output:

$$x^j = \frac{\sum_{i \in I} u^i x^i}{u^j} \left(1 - \frac{c \sum_{i \in I} u^i x^i}{u^j S} \right). \tag{26}$$

Summing (25) over all firms with positive market share, one gets

$$\sum_{i \in I} u^i x^i = \frac{S}{c} \frac{(\#I - 1)}{\sum_{i \in I} \frac{1}{u^i}},$$

where $\#I$ denotes the number of elements in I. Inserting the r.h.s. expression into (26), we obtain firm j's equilibrium output, price and profit:

$$x^j = \frac{S}{c} \frac{(\#I - 1)}{\sum_{i \in I} \frac{u^j}{u^i}} \left(1 - \frac{\#I - 1}{\sum_{i \in I} \frac{u^j}{u^i}}\right),$$

$$p^j = c \frac{\sum_{i \in I} \frac{u^j}{u^i}}{\#I - 1},$$

and

$$\pi^j(u^1, ..., u^{n(S)}) = S \left(1 - \frac{\#I - 1}{\sum_{i \in I} \frac{u^j}{u^i}}\right)^2$$

provided that $\sum_{i \in I} \frac{u^j}{u^i} > \#I - 1$, and $x^j = \pi^j = 0$ otherwise.[28]

It remains to show that, in *any* equilibrium, $I = \{1, ..., \underline{n}(S)\}$ and, hence, $\#I = \underline{n}(S)$. First, notice if $\sum_{i \in I \cup \{j\}} \frac{u^j}{u^i} > \#(I \cup \{j\}) - 1$, then $x^j > 0$, i.e. $j \in I$, in equilibrium; otherwise firm j could profitably deviate by producing $x^j = (S/c) \times \left(\#I / \sum_{i \in I \cup \{j\}} (u^j/u^i)\right) \left(1 - \left(\#I / \sum_{i \in I \cup \{j\}} (u^j/u^i)\right)\right)$. We now prove that there cannot be an equilibrium in which a product of some quality has zero market share while another offering of lower quality makes positive sales. That is, there are no firms k and l, $k < l$ such that $k \notin I$ and $l \in I$. To see this, suppose otherwise. From $k \notin I$, it follows that $\sum_{i \in I \cup \{k\}} (u^k/u^i) \leq \#I$, and from $l \in I$ that $\sum_{i \in I} (u^l/u^i) > \#I - 1$. It is easy to show that these two inequalities lead to a contradiction. This completes the proof. ■

Proof of Proposition 6. Suppose that there exists an equilibrium such that $n(S) \to \infty$ as $S \to \infty$. Below, we will prove that, for S sufficiently large, there will then exist a profitable deviation for some firm, contradicting the existence of such an equilibrium.

Consider firm $n(S)$. Remark first that $n(S) = \underline{n}(S)$; otherwise, if $n(S) > \underline{n}(S)$, firm $n(S)$'s stage-3 profit would be nil in each period, and firm $n(S)$'s profitable deviation would be not to enter the market. Now, oberserve that firm $n(S)$'s discounted sum of profits, prior to entry, is bounded above by

$$B(S) \equiv \frac{S}{(1 - \delta)[n(S)]^2} - F_0 \left(u_S^{n(S)}\right)^\beta - \epsilon,$$

where we allow quality $u_S^{n(S)}$ to depend directly on S. For notational convenience, we

[28]Remark that, in equilibrium, the condition for positive output is equivalent to the condition for an interior solution, $u^j S > c \sum_{i \in I} u^i x^i$.

will henceforth drop the subscript S. In equilibrium, clearly, $B(S) \geq 0$, and hence

$$\frac{(1-\delta)\epsilon}{S} + (1-\delta)F_0\frac{\left(u^{n(S)}\right)^{\beta}}{S} \leq \frac{1}{[n(S)]^2}.$$

Since, by assumption, $n(S) \to \infty$ as $S \to \infty$, it follows that

$$S\left(u^{n(S)}\right)^{-\beta} \to_{S\to\infty} \infty. \tag{27}$$

Now, consider the investment stage in an arbitrary period. Suppose that firm $n(S)$ deviates and invests up to quality level $u' > u^{n(S)}$, where u' is allowed to depend on S. Then, a sufficient condition for this deviation to be profitable is given by

$$\frac{S}{(1-\delta)[n(S)]^2} < S\left(1-\frac{n(S)-1}{\sum_{i=1}^{n(S)-1}\frac{u'}{u^i}+1}\right)^2 - F_0\left(u'\right)^{\beta} + F_0\left(u^{n(S)}\right)^{\beta}.$$

The expression on the l.h.s. is an upper bound on the discounted sum of stage-3 profits from nondeviation. The first term on the r.h.s. is a lower bound on stage-3 payoffs from deviation, and the remaining terms correspond to investment costs. (Actual payoffs from deviation might be higher for two reasons: firstly, the deviant firm might get positive stage-3 profits in future periods as well, and not only in the period of deviation, and secondly, the deviation might induce low-quality firms (such as firm $n(S)-1$) to fall out of the quality window.)

Let us now consider the following deviation:

$$u' = \left(Su^{h(n(S))}\right)^{\frac{1}{\beta+1}},$$

where $u^{h(n(S))}$ is the harmonic mean of firm $n(S)$'s rival qualities, i.e.

$$u^{h(n(S))} \equiv \frac{n(S)-1}{\sum_{i=1}^{n(S)-1}\frac{1}{u^i}}.$$

Notice that $u^{h(n(S))} \geq u^{n(S)} \geq 1$, and that $\left(Su^{h(n(S))}\right)^{\frac{1}{\beta+1}} > u^{n(S)}$ for S sufficiently large. Furthermore, $u^{n(S)}/\ u^{h(n(S))} \to 1$ as $S \to \infty$ since firm $n(S)$'s stage-3 profit is positive in equilibrium. This, in conjunction with (27), implies that $\lim_{S\to\infty} S\left[u^{h(n(S))}\right]^{-\beta} = \infty$. The sufficient condition for the deviation to be profitable can be written as

$$\begin{aligned}\frac{1}{(1-\delta)[n(S)]^2} &< \left(1-\frac{1}{\left(S\left[u^{h(n(S))}\right]^{-\beta}\right)^{\frac{1}{\beta+1}}+\frac{1}{n(S)-1}}\right)^2 \\ &\quad -S^{-1}F_0\left(Su^{h(n(S))}\right)^{\frac{\beta}{\beta+1}} + S^{-1}F_0\left(u^{n(S)}\right)^{\beta}.\end{aligned}$$

Remark that the l.h.s. of this inequality converges to zero as market size tends to infinity. Furthermore, it is straightforward to see that

$$\lim_{S\to\infty}\left\{\left(1-\frac{1}{\left(S\left[u^{h(n(S))}\right]^{-\beta}\right)^{\frac{1}{\beta+1}}+\frac{1}{n(S)-1}}\right)^2-S^{-1}F_0\left(Su^{h(n(S))}\right)^{\frac{\beta}{\beta+1}}\right\}=1$$

since $\lim_{S\to\infty} S\left[u^{h(n(S))}\right]^{-\beta}=\infty$. Hence, for S sufficiently large, the r.h.s. of the above inequality is larger than the l.h.s.; that is, there exists a profitable deviation in large markets. ■

References

d'Aspremont, C. and A. Jacquemin (1988). Cooperative and Noncooperative R&D in Duopoly with Spillovers. *American Economic Review*, 1133–1137.

De Bondt, R. (1996). Spillovers and Innovative Activities. *International Journal of Industrial Organization*, 1–28.

Fudenberg, D. and J. Tirole (1983). Capital as a Commitment: Strategic Investment to Deter Mobility. *Journal of Economic Theory*, 227–250.

Fudenberg, D. and J. Tirole (1991). *Game Theory.* MIT Press.

Hole, A. (1997). *Dynamic Non-Price Strategy and Competition: Models of R&D, Advertising and Location.* Ph. D. thesis, University of London.

Kamien, M. I., E. Muller, and I. Zang (1992). Research Joint Ventures and R&D Cartels. *American Economic Review*, 1293–1306.

Kesteloot, K. and R. Veugelers (1995). Stable R&D Cooperation with Spillovers. *Journal of Economics & Management Strategy*, 651–672.

Maskin, E. and J. Tirole (1988). A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition with Large Fixed Costs. *Econometrica*, 549–570.

Nerlove, M. and K. J. Arrow (1962). Optimal Advertising Policy Under Dynamic Conditions. *Economica*, 129–142.

Reynolds, S. S. (1987). Capacity Investment, Preemption and Commitment in an Infinite Horizon Model. *International Economic Review 28*(1), 69–88.

Reynolds, S. S. (1991). Dynamic Oligopoly with Capacity Adjustment Costs. *Journal of Economic Dynamics and Control 15*(3), 491–514.

Shaked, A. and J. Sutton (1983). Natural Oligopolies. *Econometrica 51*(5), 1469–1484.

Shaked, A. and J. Sutton (1987). Product Differentiation and Industrial Structure. *Journal of Industrial Economics 36*(2), 131–146.

Shapiro, C. (1989). Theories of Oligopoly Behavior. In R. Schmalensee and R. D. Willig (Eds.), *Handbook of Industrial Organization*, Volume 1, pp. 329–414. North Holland.

Spence, A. M. (1979). Investment Strategy and Growth in a New Market. *Bell Journal of Economics*, 1–19.

Sutton, J. (1991). *Sunk Costs and Market Structure*. Cambridge, Mass.: MIT Press.

Sutton, J. (1998). *Technology and Market Structure*. Cambridge, Mass.: MIT Press.