Entry, Exit, and Imperfect Competition in the Long Run

Rabah Amir* and Val E. Lambson†

Abstract: An infinite-horizon, stochastic model of entry and exit with sunk costs and imperfect competition is constructed. Simple examples provide insights into: (1) the relationship between sunk costs and industry concentration, (2) entry when current profits are negative, and (3) the relationship between entry and the length of the product cycle. A subgame perfect Nash equilibrium for the general dynamic stochastic game is shown to exist as a limit of finite-horizon equilibria. This equilibrium has a relatively simple structure characterized by two numbers per finite history. Under very general conditions, it tends to exhibit excessive entry and insufficient exit relative to a social optimum.

Key Words and Phrases: Entry, Exit, Dynamic games, Integer constraints.

JEL Classification Codes: C73, D43, L13

*CORE, 34 Voie du Roman Pays, 1348, Louvain-la-Neuve, Belgium. (Email: amir@core.ucl.ac.be).

†Brigham Young University, Economics Dept., Provo UT 84602, USA; email:vel@email.byu.edu

1 Introduction

Accepting the proposition that firms are forward-looking, economists have constructed myriad dynamic models of industry evolution, that is, models that incorporate time and (sometimes) uncertainty. For reasons of tractability, most of these have been two-period models. While two-period models have yielded many important insights, they suffer from the well-known drawback that the final period behaves like a static (one-period) model. This is a potential problem because firms expecting static behavior in the near future may behave differently than firms expecting continued dynamic behavior. Circumventing this problem are the stochastic infinite-horizon models of industry evolution. Most of these models—e.g. Jovanovic (1982), Dixit (1989), Lambson (1991), Rob (1991), Hopenhayn (1992), and Klepper (1996)—posit infinitesimally-sized price-taking firms. Thus they yield insights into competitive processes but are ill-suited for exploring traditional industrial organization questions regarding market structure and monopoly power.[1]

The present paper formulates a stochastic infinite-horizon model where the number of firms is required to be an integer. The framework generates endogenous entry and exit from shocks—such as changes in demand or factor prices—that are external to the firms. The infinite-horizon nature of the model avoids the "final period problem" while the requirement that there be an integer number of firms—rather than a continuum of infinitesimally-sized firms—makes the model suitable for addressing traditional industrial organization questions concerning imperfect competition. Furthermore, the integer constraint—and the resulting inability to satisfy equilibrium zero-profit conditions with equality—has economically relevant implications for the behavior of firms over time.

The first part of the paper—comprising Sections 2 through 5—contains a rather informal description of the model followed by several examples that offer insights into traditional industrial organization questions. The second part of the paper—comprising Sections 6 through 8—contains a formal description of the model, addresses technical equilibrium existence issues, and contains a rather general investigation of the welfare properties of equilibrium entry and exit when competition is imperfect. Although Sec-

[1]An interesting exception is due to Ericson and Pakes (1995). The relationship between their model and ours is discussed in Section 6.

tions 2 through 5 are rather informal, and the equilibria are presented intuitively without a careful definition of the equilibrium concept, the interested reader will have little difficulty verifying that the described outcomes are indeed associated with equilibria as formally defined in Section 6.

Section 2 contains an informal description of the model and defines the notation that is essential to the examples presented in Sections 3 through 5.

Section 3 questions the implicit assumption underlying much of the industrial organization literature that high sunk costs result in high concentration, that is, that high sunk costs reduce the number of firms that are active in an industry. It is shown by example that this belief depends on static thinking. In a dynamic environment sunk costs reduce both entry *and* exit. Thus, although high sunk costs mean that fewer firms enter in good times they also mean that more firms remain active in bad times; hence the effect on the average number of firms over time is ambiguous.

Section 4 considers the question of why new firms often lose money before eventually becoming profitable. This can be explained in a variety of ways that depend on imperfect information. (For example, imperfect capital markets may cause delays in achieving minimum efficient scale or learning-by-doing may be required before a firm's technology becomes profitable.) Our model suggests another possibility that depends critically on the requirement that the number of firms be an integer. The implicit lumpiness of the technology means that the zero-profit entry condition is not exactly satisfied except by chance; thus the expected present value of an entrant typically exceeds its entry cost. In equilibrium, the lure of these economic profits can induce firms to enter sooner than they would like—for example, before demand allows for positive operating profits—in order to "hold the place." (The example in Section 4 also suggests that unregulated entry tends to result in excessive entry, a result previously established by Mankiw and Whinston (1986) in the context of a two-stage model. This topic is revisited more generally in Section 8.)

Section 5 contains a result that is interpretable economically as well as relevant to the existence issues discussed in a later section. When a firm considers whether to enter, the number of periods over which it can recoup its sunk cost matters. In particular, it seems plausible that more firms would enter (or fewer firms exit) given a longer time horizon. Although this is true when firms are infinitesimally-sized price-takers—see Lambson (1992)—it is not true here. Thus, for example, lengthening the product cycle may not result in an increase in the number of initial entrants. Indeed, the level of entry (or exit) in a given period can change non-monotonically as the horizon lengthens.

Section 6 contains a formal description of the general, infinite-horizon, stochastic framework. The model is a dynamic stochastic game in discrete time with countably infinitely many time periods and potential firms. An equilibrium is a subgame perfect Nash equilibrium vector of the firms' entry and exit strategies.

Section 7 contains a general proof that a subgame perfect Nash equilibrium exists as the limit of finite-horizon equilibria. The strategic nature of entry and exit decisions, the integer constraint, and the non-monotonicities documented in Section 5 imply that straightforward extensions of proofs in models with continua of firms are inadequate to prove existence of equilibrium in the current framework.

Section 8 considers the welfare implications of unregulated entry and exit when there is imperfect competition. As noted above, Section 4 suggests that unregulated entry may result in too much entry relative to the socially optimal outcome. That there is a tendency for excessive entry in a two-stage model—where firms make entry decisions in the first stage and production decisions in the second stage—was established by Mankiw and Whinston (1986). The general, stochastic, infinite-horizon model of Sections 6 and 7 allows a deeper investigation. For the class of equilibria from Section 7, not only is there a tendency for excessive entry when times are good but there is also a tendency for insufficient exit when times are bad. Furthermore, these tendencies are not reversed in the present even when firms take into account that there will tend to be too many firms in the future. Thus, these equilibria tend to exhibit too many active firms in *every* period of *every* realization of the stochastic process. ("Tendency" means that the number of active firms never falls more than one short of the optimal number, and may greatly exceed it.) This result only applies to the "limit equilibria" constructed in Section 7, however. Section 8 also contains an example of an equilibrium in which entry falls more than one short of the optimal number of firms. Thus the tendency for an excessive number of firms in imperfectly competitive markets is not universal.

Section 9 contains concluding remarks.

2 The Model: An Informal Description

The reader who is interested in the technical details of the model is referred to Section 6. Here just enough detail is provided to allow the reader to follow the examples in Sections 3 through 5.

There are countably infinitely many firms with a countably infinite time horizon. A *market condition* is a description of all relevant exogenous variables in a given time period; for example, the description of a market condition may include factor prices, demand, entry costs, scrap values, and the regulatory environment among other things. Market conditions follow an exogenous stochastic process known to the firms.

At most finitely many firms are initially *active*, that is, capable of producing without paying an entry cost. The other firms are initially *inactive*. At the beginning of each period, firms observe the current market condition, say, m. Furthermore, they recall all past market conditions as well as all the firms' previous entry and exit decisions. Then inactive firms decide whether to pay an *entry cost* $\xi_m > 0$ to become active and active firms decide whether to accept a *scrap value* $\chi_m < \xi_m$ to become inactive. After entry and exit decisions are implemented, each active firm receives a current profit of $\pi_m(y)$, where y is the number of active firms. Thus π_m is a reduced-form profit function; it is assumed to be non-increasing in y. In equilibrium, each firm chooses an entry and exit strategy to maximize the expected present value of its profit stream given the entry and exit strategies of the other firms. Firms discount their profit streams with the same discount factor, $\delta \in (0, 1)$.

This concludes the informal description of the model. A formal description is contained in Section 6. The next three sections contain some examples.

3 Concentration and Entry Barriers

It is natural—but, as this section demonstrates, incorrect—to believe that higher sunk costs necessarily decrease the number of firms in an industry. Put another way, it is natural to believe that higher sunk costs increase an industry's concentration, typically defined as the combined market share of the largest firms. Partly because of this belief, much effort has gone into debating the empirical importance of entry barriers such as sunk costs.[2] Little of this literature has challenged the assertion that entry barriers and concentration are positively correlated.[3] This section challenges that assertion, showing by example that it depends critically on static reasoning. In a dynamic context, higher entry barriers indeed reduce the number of entrants when times are good but they also reduce the number of firms that exit when times are bad. The net effect on the average number of firms over time is ambiguous.

Suppose there are two market conditions: market condition H exhibits high demand and market condition L exhibits low demand. Suppose the initial market condition is H; thereafter market conditions follow the i.i.d. stochastic process characterized by the probabilities ρ_H and ρ_L of the respective market conditions. In market condition $m \in \{H, L\}$ inverse demand is $P = a_m - Q$ and production is costless, so Cournot operating profit is $\pi_m(y) = \left(\frac{a_m}{y+1}\right)^2$. Assume that both market conditions have the same entry cost ξ and the same scrap value χ. The discount factor is δ. Initially there are no active firms.

Using standard recursive methods, it follows that if y_m firms are active in market condition m, then a firm that never exits has an expected present value (before deducting

[2]A lot of this debate has centered around a perceived positive correlation between concentration and profit rates across industries. See, for example, Weiss (1974), Demsetz (1974), Dewey (1976), La Manna (1986) and Lambson (1987).

[3]An exception is Bernheim (1984) who, in a finite model of sequential entry, made the point that when future entry deterrence is more costly, current potential entrants may be less likely to enter. Bentolila and Bertola (1990) and Lambson (1992) point out that in dynamic competitive models higher sunk costs reduce both entry in good times and exit in bad times and that either effect may dominate on average.

the entry cost) of

$$\nu_H(y_H, y_L) = \pi_H(y_H) + \delta[\rho_H \nu_H(y_H, y_L) + \rho_L \nu_L(y_H, y_L)]$$

in the high demand market condition and an expected present value of

$$\nu_L(y_H, y_L) = \pi_L(y_L) + \delta[\rho_H \nu_H(y_H, y_L) + \rho_L \nu_L(y_H, y_L)]$$

in the low demand market condition. Solving these two equations for ν_H and ν_L yields

$$\nu_H(y_H, y_L) = \frac{(1 - \delta\rho_L)\pi_H(y_H) + (\delta\rho_L)\pi_L(y_L)}{(1 - \delta)} \tag{3.1}$$

and

$$\nu_L(y_H, y_L) = \frac{(1 - \delta\rho_H)\pi_L(y_L) + (\delta\rho_H)\pi_H(y_H)}{(1 - \delta)}. \tag{3.2}$$

Finally, a firm the enters in market condition H and then exits (for good) in the next realization of market condition L has an expected present value of $\nu_H^x(y_H, y_L) = \pi_H(y_H) + \delta[\rho_H \nu_H^x(y_H, y_L) + \rho_L \chi]$, which implies

$$\nu_H^x(y_H, y_L) = \frac{\pi_H(y_H) + \delta\rho_L\chi}{1 - \delta\rho_H}. \tag{3.3}$$

To construct the example, impose the following values: $a_H = 2$, $a_L = 1$, $\rho_H = \rho_L = .5$, $\delta = .9$, and $\chi = 1.35$. If the entry cost is the relatively high $\xi = 1.65$ then it will be argued that $y_H = y_L = 3$ in equilibrium, so the average number of active firms is 3. By contrast, if the entry cost is the relatively low $\xi = 1.39$ then it will be argued that $y_H = 4$ and $y_L = 1$ in equilibrium, so the average number of active firms is only 2.5. Thus, contrary to the conventional wisdom, average concentration over time is negatively related to entry barriers. Formal and complete descriptions of the equilibrium strategies are omitted in this section—such formalities are postponed until section 6—but a heuristic argument that the equilibria have these properties will be given.

First consider equilibrium when the entry cost is $\xi = 1.65$. To see that $y_H = y_L = 3$, suppose three firms adopt the strategy of entering and never exiting. Given the imposed values, (3.1) and (3.2) imply

$$\nu_H(3, 3) \simeq 1.656 > \xi > 1.469 \simeq \nu_L(3, 3) > \chi.$$

Thus, each of the three firms has expected present value exceeding entry cost when demand is high and exceeding scrap value when demand is low. A fourth firm would achieve an expected present value of $\nu_H(4,4) \simeq 1.1 < \xi$ if it entered immediately and never exited or an expected present value of $\nu_H^x(4,3) \simeq 1.4 < \xi$ if it entered and exited on the first occurrence of market condition L. (See (3.1) and (3.3).) Other attempts to enter and exit advantageously similarly fail to cover the entry cost.

Now consider equilibrium when the entry cost is only $\xi = 1.39$. To see that $y_H = 4$ and $y_L = 1$, suppose one firm adopts the strategy of entering immediately and never exiting while a group of other firms (three at a time) adopt the strategy of entering when (it is "their turn" and) the market condition H is realized, and exiting at the first subsequent occurrence of market condition L.[4] Given the imposed values, (3.3) implies that a firm that enters in market condition H and exits at the first occurrence of market condition L has an expected present value (discounted back to the period of entry) of $\nu_H^x(4,1) \simeq 1.395 > \xi$ in market condition H. The single firm that never exits has an expected present value—from (3.2)—of $\nu_L(4,1) \simeq 2.095 > \chi$ in market condition L; a second firm will not find it optimal to remain active in market condition L because $\nu_L(4,2) \simeq 1.331 < \chi$. Furthermore, a fifth firm considering entry in market condition H faces an expected present value of $\nu_H^x(5,1) \simeq 1.31 < \xi$ if it exits at the first subsequent occurrence of market condition L or an expected present value of $\nu_H(5,2) \simeq 1.12 < \xi$ if it never exits. Other attempts to enter (or remain active) "out of turn" similarly fail to cover the entry cost (or scrap value).

This example establishes that the long run average number of firms may be lower with lower entry barriers, contrary to the conventional wisdom. The intuition is that, although lower entry costs induce more entry when times are good, they also induce more exit when times are bad; hence the effect on the long-run average number of active firms is ambiguous.

[4]The meaning of "their turn" will be clear in section 7.

4 Preemptive Entry

Firms entering an industry often lose money initially before becoming profitable. This can be explained in a variety of ways that depend on imperfect information. (For example, imperfect capital markets may cause delays in achieving minimum efficient scale or learning-by-doing may be required before a firm's technology becomes profitable.) This section contains another possibility that depends critically on lumpy sunk costs, that is, on the requirement that the number of firms be an integer. The lumpiness of the technology means that, even allowing for entry, economic profits can be made; if the constraint that the number of firms must be an integer is binding—as will be the case except by chance—then the expected present value of entrants exceeds their entry costs. In equilibrium, the lure of these economic profits can induce firms to enter sooner than they would like—for example, before demand allows for positive operating profits—in order to "hold the place." Thus this explanation is related to the "rent dissipation" literature. (See, for example, Fudenberg and Tirole (1985, 1987).)

Suppose there are three market conditions: H, M, and L (meant to suggest high, medium, and low demand, respectively). Beginning with market condition H, market conditions follow an i.i.d. stochastic process characterized by the probabilities ρ_H, ρ_M, and ρ_L, respectively. In market condition $m \in \{H, M, L\}$ inverse demand is $P = a_m - Q$ and production requires a fixed cost of ϕ (but no marginal cost), so Cournot operating profit is $\pi_m(y) = \left(\frac{a_m}{y+1}\right)^2 - \phi$. Assume that all market conditions have the same entry cost ξ and the same scrap value χ. The discount factor is δ. Initially there are no active firms.

To construct the example, impose the following values: $\rho_H = .6$, $\rho_M = .3$, $\rho_L = .1$, $a_H = 15$, $a_M = 8$, $a_L = 3$, $\xi = 1$, $\chi = 0$, and $\phi = 2$. It will be argued that, given the imposed values, equilibrium exhibits $y_H = 8$ firms in market condition H, $y_M = 6$ firms in market condition M if the previous distinct market condition was H, $y_M = 5$ firms in market condition M if the previous distinct market condition was L, and $y_L = 2$ in market condition L. Thus, each time the market condition changes to H there is entry until eight firms are active. Each time the market condition changes to L there is exit until two firms remain active. The result of a change to market condition M depends on prior history: if there were previously eight active firms (because the change to M is from market condition H) then there is exit until six firms remain but if there were previously only two firms (because the change to M is from market condition L) then there is entry until five firms are active. The history-dependence of the number of active firms in market condition M reflects *hysteresis* caused by the sunk costs. (See Dixit (1989) or Lambson (1992).)

To see that this example exhibits entry with negative current profits infinitely often, note that $\pi_M(5) \simeq -.22$; thus whenever the market condition changes from L to M, three firms enter even though current profits are negative. They do so hoping to earn $\pi_H(8) \simeq .78$ often enough in the future to cover the entry cost before they are forced to exit when the market condition returns to L.

As in the previous section, a full description of the equilibrium strategies is omitted here, but a heuristic argument that equilibrium is as described will be given. Suppose initially (when the market condition is H) that eight firms enter, two with the intention of never exiting, two with the intention of exiting when either M or L is realized, and four with the intention of exiting only when L is realized. Thereafter, there are four kinds of market condition changes to consider: changes to H (from either M or L), changes to M from H, changes to M from L, and changes to L (from either H or M). The arguments for the first two cases will be sketched. The other cases are similar—albeit more complicated—and are omitted. In what follows, ν_{mn} is the expected present value of a firm in market condition m that intends to exit the next time the market condition exhibits lower demand than market condition n.

When the market condition changes to H, equilibrium exhibits entry until there are eight firms, where the marginal firm intends to exit the next time the market condition changes (and demand falls). Thus the present expected value of the marginal firm is ν_{HH} satisfying $\nu_{HH} = \pi_H(8) + \delta\rho_H\nu_{HH}$; so $\nu_{HH} = \pi_H(8)/(1-\delta\rho_H) \simeq 1.69 > \xi$. If, however, a ninth firm entered with the intention of exiting when demand next fell, its expected present value would be only $\pi_H(9)/(1-\delta\rho_H) \simeq .54 < \xi$. It is also possible to verify that other exit plans similarly fail to cover the entry cost.

When the market condition changes from H to M, all but six firms exit. The marginal remaining firm, intending to exit only when L is next realized (and understanding that there will be entry up to eight firms if H is realized), has an expected present value of ν_{MM} satisfying $\nu_{MM} = \pi_M(6) + \delta[\rho_M \nu_{MM} + \rho_H \nu_{HM}]$ where ν_{HM} satisfies $\nu_{HM} = \pi_H(8)+\delta[\rho_M \nu_{MM}+\rho_H \nu_{HM}]$. Solve these two equations to write $\nu_{MM} = [(1-\delta\rho_H)\pi_M(6)+\delta\rho_H\pi_H(8)]/(1-\delta\rho_M-\delta\rho_H) \simeq .53 > \chi$. By contrast, if a seventh firm remained active with the intention of exiting the first time market condition L is realized, its expected present value would only be $\nu_{MM} = [(1-\delta\rho_H)\pi_M(7)+\delta\rho_H\pi_H(8)]/(1-\delta\rho_M-\delta\rho_H) \simeq -.21 < \chi$. It is also possible to verify that other exit plans similarly fail to compensate the seventh firm for foregoing its scrap value. Thus only six firms remain active when the market condition changes from H to M.

In this example, equilibrium is clearly inefficient, even subject to the constraint that active firms are Cournot (not perfect) competitors. To see this, consider what happens when the market condition changes from L to M. With five active firms (as equilibrium requires), standard calculations establish that current consumer surplus is approximately 22.22 while current industry profit is approximately -1.11; thus current social surplus is approximately 21.11. If, by contrast, only four firms were active, current consumer surplus would be 20.48 while current industry profit would be 2.24, resulting in a current social surplus of 22.72. It follows that a social planner could increase the expected present value of social surplus relative to the equilibrium level by making the single change of allowing only four active firms (rather than five) when the market condition M follows the market condition L. Not only would it increase current social surplus in those cases, it would also save on entry costs. Entry is excessive when demand increases from low demand to medium demand.

It is interesting to note that the fifth firm would have an incentive to delay entry until H is realized, if it could do so without fear of another firm taking its place; by doing so it could avoid some losses while delaying payment of the entry cost. Equilibrium does not allow such delay, however, because of the incentive for other firms to enter instead: firms must enter to "hold the place" before it is socially optimal to do so.

The insight that unregulated entry can result in excessive entry when competition is imperfect is not new. It was formally established in a two-stage model by Mankiw and Whinston (1986). The stochastic infinite-horizon model constructed in Section 6 below allows a much more general investigation. First, however, Section 5 offers one more example that has economic meaning as well as implications for the technical issues that follow.

5 Entry, Exit, and Horizon Length

A longer time horizon gives entering firms (or, similarly, firms that refrain from exiting) more time to recoup entry costs (or, similarly, foregone scrap values); this effect tends to increase the current number of firms. However, a longer time horizon may also increase the future number of active firms, potentially making it more difficult to recoup entry costs or foregone scrap values; this effect tends to reduce the current number of active firms.

In dynamic models with infinitesimal price-taking firms there is no ambiguity: the first effect dominates. The intuition is not hard to grasp. In continuum models zero-profit conditions can be exactly satisfied, so a firm's equilibrium present value is maximized (and equal to the entry cost) when there is entry. Hence future increased entry due to a longer horizon implies that future profit streams are maximally attractive, and thus cannot reduce the attractiveness of current activity. It follows that the number of firms in any given period is (weakly) increasing in the time horizon. Lambson (1992) used this monotonicity to show that an infinite-horizon equilibrium can be constructed as the limit of finite-horizon equilibria when there is a continuum of infinitesimal price-taking firms.

When the number of firms must be an integer, by contrast, either effect may dominate, because zero-profit conditions need not be satisfied with equality. Since entry doesn't imply maximal present values, future increased entry due to a longer horizon may (but need not) reduce the attractiveness of future profit streams, make current activity less attractive, and thus reduce the current number of firms. This ambiguity suggests that the number of active firms in a given time period need not be monotonic in the horizon. The remainder of this section contains an example of the non-monotonicity phenomenon.

Thus monotonicity arguments that are useful for proving existence in the analogous model with a continuum of firms fail in the present circumstance.

Suppose inverse demand in period t is $P_t = a_t - Q_t$ and that there are no production costs, so Cournot operating profit in period t is $\left(\frac{a_t}{y+1}\right)^2$. Assume there is no scrap value—that is, $\chi_t = 0$ for all t—and let the entry cost be $\xi_t = 1$ in every period. The discount factor is $\delta = \frac{1}{2}$. Suppose the a_t values are given by $a_1 = 1$, $a_2 = 4-\varepsilon$, and $a_t = 6$ for $t \geq 3$. Since there are no scrap values, and since operating profits are always positive, there will never be any exit. Suppose no firms are initially active, and let y_t^T be the number of firms that are active in period t if the horizon length is $T \geq t$.

First consider the truncated model where firms expect the market to last for one period. Then y_1^1 firms will enter, where y_1^1 is the largest integer satisfying $\left(\frac{1}{y_1^1+1}\right)^2 \geq 1$. Thus

$$y_1^1 = 0.$$

In the one-period truncation there is no entry if there are initially no firms.

Next consider the two-period truncation, that is, let $T = 2$. Apply backward induction. Assuming (as turns out to be the case) that there is entry in period 2, y_2^2 is the largest integer satisfying $\left(\frac{4-\varepsilon}{y_2^2+1}\right)^2 \geq 1$. Thus for small $\varepsilon > 0$, $y_2^2 = 2$. Then y_1^2 is the largest integer satisfying $\left(\frac{1}{y_1^2+1}\right)^2 + \frac{1}{2}\left(\frac{4-\varepsilon}{\max(y_1^2,2)+1}\right)^2 \geq 1$. So for small $\varepsilon > 0$, $y_1^2 = 1$. In summary,

$$y_1^2 = 1;\ y_2^2 = 2.$$

In the two-period truncation one firm enters in period 1 and then an additional firm enters in period 2.

Finally, consider the three-period truncation, that is, let $T = 3$. Apply backward induction. Assuming (as turns out to be the case) that there is entry in each period, y_3^3 is the largest integer satisfying $\left(\frac{6}{y_3^3+1}\right)^2 \geq 1$. Thus $y_3^3 = 5$. Next, y_2^3 is the largest integer satisfying $\left(\frac{4-\varepsilon}{y_2^3+1}\right)^2 + \frac{1}{2}\left(\frac{6}{\max(y_2^3,5)+1}\right)^2 \geq 1$. Thus, for small $\varepsilon > 0$, $y_2^3 = 4$. Finally, y_1^3 is the largest integer satisfying $\left(\frac{1}{y_1^3+1}\right)^2 + \frac{1}{2}\left(\frac{4-\varepsilon}{\max(y_1^3,4)+1}\right)^2 + \frac{1}{4}\left(\frac{6}{\max(y_1^3,5)+1}\right)^2 \geq 1$. Thus $y_1^3 = 0$. In summary,

$$y_1^3 = 0;\ y_2^3 = 4;\ y_3^3 = 5.$$

In the three-period truncation, no firms enter in period 1, four firms enter in period 2, and an additional firm enters in period 3.

Notice the non-monotonicity of y_1^T: $y_1^1 = 0$, $y_1^2 = 1$, and $y_1^3 = 0$. The intuition for why this arises is as follows. In the one-period truncation no firm enters; demand doesn't allow even one firm to recoup the entry cost so quickly. In the two-period truncation, demand is higher in the second period, thus allowing two firms each to recoup much more than the entry cost by producing only in period 2. These super-normal profits occur because of the integer constraint: if a third firm entered it couldn't recoup its entry cost in period 2 alone. The lure of super-normal profits in period 2 (even discounted) suffices to lure one firm to enter in period 1. In the three-period truncation there is higher demand still in period 3, sufficient for five firms to more than recoup their entry costs just in period 3. Once again, the lure of super-normal profits in period 3 induces additional entry in period 2, but the example is constructed so that the super-normal profits are dissipated in period 2: the integer constraint does not bind very tightly. Thus there are no second-period super-normal profits to lure entrants in period 1. This non-monotonicity of the number of firms as the horizon-length increases implies that monotonicity arguments cannot be used to establish the existence of a limiting equilibrium in the infinite horizon. As will be seen, a more sophisticated diagonalization argument must be used to prove existence.

This concludes the simple, illustrative examples. With these guides to intuition in place, the general stochastic model is described (in Section 6), the existence of equilibrium is proved (in Section 7), and (in Section 8) the tendency for unregulated entry and exit to result in excessive numbers of firms when there is imperfect competition is investigated.

6 The Model: A Formal Description

Let $i \in \{1, 2, ...\} \equiv I$ index the countably infinite set of firms. Let $t \in \{1, 2, ...\} \equiv I$ index the countably infinite set of time periods. A *market condition* is a description of all relevant exogenous variables in a given time period; for example, the description of a market condition may include factor prices, demand, entry costs, scrap values, and the regulatory environment as well as other things. Let $m \in M$ index the countable (finite or infinite) set of market conditions. Let m_t be the market condition realized in period t. Let $h = (m_1, ..., m_t)$ be the market conditions realized in the first t periods (to be called a t-period *market history*), let H_t be the set of possible t-period market histories, and let $H = \cup_{t=1}^{\infty} H_t$ be the set of all possible finite market histories. Market conditions begin with a first market condition, m_1. Their evolution is governed by the exogenous stochastic process described by $\rho(\cdot \mid \cdot)$, where $\rho(g \mid h)$ is the probability that the market history $g \in H_\tau$ is observed given that the market history $h \in H_t$ is observed where $\tau \geq t$. Market histories are partially ordered by $\rho(\cdot \mid \cdot)$: $g \geq h$ iff $\rho(g \mid h) > 0$ and $g > h$ iff $g \geq h$ and $g \neq h$. The notation $f \in [h, g]$ will denote f such that $h \leq f \leq g$. For $h \in H_t$ and $t \geq 2$, the notation h^{-1} denotes the immediate predecessor of the market history h; that is, if $h \in H_t$ then $h^{-1} \in H_{t-1}$ and lists the first $t-1$ market conditions of h.

At most finitely many firms are initially *active* and the other firms are initially *inactive*. Beginning in period 1, firms choose whether to be active or inactive in each period. Call $a_0^i \in \{0, 1\}$ the *initial activity index* for firm i and call $a_h^i \in \{0, 1\}$ the *activity index* of firm i in the last period of the market history h (or the activity index of firm i at h); an index of 1 denotes activity and an index of 0 denotes inactivity. Call $a_0 = \{a_0^i\}_{i \in I}$ and $a_h = \{a_h^i\}_{i \in I}$ the *activity vectors* initially and at h, respectively. A pair (m, a) is called a *state*; it is interpreted as the current market condition and the previous period's activity vector. A t-period *history*, s, is a finite realization of states: $s = \{h, (a_{g^{-1}})_{g \leq h}\}$ for some $h \in H_t$. (If $g = (m_1)$—that is, if $t = 1$—then it is understood that $a_{g^{-1}} = a_0$.) Let S_t be the set of possible t-period histories and let $S = \cup_{t=1}^{\infty} S_t$ be the set of all possible finite histories.

At the beginning of each period $t \geq 1$, firms observe the history through time t and then make their entry and exit decisions. Specifically, inactive firms may enter (become active) by paying a cost ξ_h while active firms may exit (become inactive) and recoup a scrap value $\chi_h < \xi_h$ (which may be negative). After these decisions are made, active firms play a symmetric game (e.g. Cournot) yielding a current payoff of $\pi_h(y_h)$ to each active firm, where $y_h = \Sigma_{i \in I} a_h^i$ is the number of active firms at h. The definitions of ξ_h, χ_h, and π_h reflect a slight abuse of notation relative to the definitions of ξ_m, χ_m, and π_m in earlier sections. Although these variables depend on the current market condition rather than the entire market history, it is notationally convenient in what follows to use market history subscripts rather than market condition subscripts. Thus, for example, ξ_h is understood to be the entry cost in the market condition prevailing in the last period of the market history h.

A strategy for firm i is a decision rule $\sigma^i : S \to \{0,1\}$ that specifies firm i's activity index as a function of the history. Given a history $s \in S_t$, an induced strategy σ^i_s is the restriction of σ^i to histories that follow s (in the sense that their first t states are the list s). Let Σ^i_s be the set of firm i's induced strategies given s. The induced strategy profile $\sigma_s = \{\sigma^i_s\}_{i\in I}$ induces a *path*, that is, a stochastic sequence of activity vectors $\{a_g\}_{g\geq h}$. Associated with that path is a *value* for firm i given s:

$$\overline{V}^i_s(\sigma_s) = \sum_{\tau=t}^{\infty} \delta^{\tau-t} \sum_{g\geq h} \rho(g \mid h)\{-\max(0, a^i_g - a^i_{g-1})\xi_g + \max(0, a^i_{g-1} - a^i_g)\chi_g + a^i_g \pi_g(y_g)\},$$

where $\delta \in [0,1)$ is a discount factor. Equilibrium can now be formally defined.

Definition: A (subgame perfect Nash) equilibrium is a strategy profile $\widetilde{\sigma} = \{\widetilde{\sigma}\}_{i\in I}$ such that $\overline{V}^i_s(\widetilde{\sigma}_s) \geq \overline{V}^i_s(\sigma^i_s, \widetilde{\sigma}^{-i}_s)$ for all $i \in I$, for all $s \in S$, and for all $\sigma^i_s \in \Sigma^i_s$, where $\widetilde{\sigma}_s$ is the strategy profile induced by $\widetilde{\sigma}$ for each s and where $\widetilde{\sigma}^{-i}_s$ is $\widetilde{\sigma}_s$ with the i^{th} element removed.

This concludes the formal description of the model. Of all the industry dynamics models, the one due to Ericson and Pakes (1995) is closest to ours. Both models allow for strategic behavior, uncertainty, and endogenous entry and exit by an integer number of firms. Yet important differences also exist. On one hand, the Ericson-Pakes model is more general in that it allows for process-R&D investments by the firms, thereby making the evolution of technology heterogeneous and endogenous. By contrast, the market conditions in our model are driven by a stochastic process that is independent of firms' actions. On the other hand, our analysis exhibits greater robustness to the specification of the primitives. In particular, the nature of the stochastic process for market conditions is very general. By contrast, the existence of a pure-strategy Markov-perfect equilibrium in the Ericson-Pakes model crucially relies on the following rather restrictive condition on each firm's transition probability, mapping the current state and investment into (probability distributions over) the next state: its associated distribution function is concave in the investment level. (Ericson and Pakes use this condition to ensure that, given rivals' strategies, a firm's best response is unique.) The restrictive nature of this assumption is discussed in some detail by Amir (1996) who argues, in particular, that this excludes, among others, situations where the next state is a deterministic function of the current state and actions. Another important difference is that while the Ericson-Pakes proof is based on an abstract fixed-point argument, ours is constructive and yields as a by-product some useful insights into the properties of equilibrium.

7 Existence of Equilibrium

Under weak regularity conditions, this section proves that a subgame perfect Nash equilibrium always exists. The equilibrium constructed in the proof will be called a *limit-lifo* equilibrium because it is the limit of finite-horizon equilibria and the last firms to enter are the first firms to exit (last in first out). A heuristic sketch of the proof will first be presented, followed by the formal statement and proof of existence. The general method is to construct an integer pair $\{N_h, X_h\}$ for each market history h where $N_h \leq X_h$; N_h will turn out to be the number of firms in the last period of the market history h if there is entry—that is, if there are fewer than N_h active firms going into the period—while X_h will turn out to be the number of firms in the last period of the market history h if there is exit—that is, if there are more than X_h active firms going into the period. First, integer pairs $\{N_h^T, X_h^T\}$ are constructed for each T-period truncation of the model by backward induction; then a limiting argument is invoked as T increases without bound. For each T-period truncation, each $t \leq T$, and each $h \in H_t$, N_h^T is the largest integer such that the expected present value (through period T) of the firm that will be the first to exit is not less than the entry cost, given that the number of firms in the last period of each future market history g, say y_g, is determined by

$$y_g = \min\{X_g^T, \max\{N_g^T, y_{g^{-1}}\}\}. \tag{7.1}$$

Equation (7.1) says that there is entry in the last period of market condition g if fewer than N_g^T firms are active going into the period and that there is exit if there are more than X_g^T active firms going into the period. Similarly, X_h^T is the largest integer such that the expected present value (through period T) of the firm that will be the first to exit is not less than the scrap value, given that the number of firms will follow the path defined by (7.1).

Unfortunately, as seen in section 5, N_h^T and X_h^T are not necessarily monotonic in T, making it impossible to verify that limits exist. Instead, a diagonalization argument is constructed in such a way that, for each h, N_h and X_h can be derived as limits of an appropriately specified subsequence. Firms' strategies are then constructed so that the number of firms follows the limiting analog of (7.1),

$$y_g = \min\{X_g, \max\{N_g, y_{g^{-1}}\}\}, \tag{7.2}$$

and such that no firm ever has an incentive to deviate from its assigned strategy. This is done by partitioning the set of initially inactive firms into countably many countable subsets. Firms in each subset are ordered and then each subset is assigned a market history. Each inactive firm considers entry only in the last period of its assigned market history; if its index is sufficiently low it enters. Active firms consider exit each period according to a last-in-first-out rule. Any firm that deviates from its assigned behavior by becoming or remaining active becomes the "marginal firm"—that is, the firm expected to exit first. Any firm that deviates from its assigned behavior by becoming or remaining inactive loses its chance at further equilibrium activity in the induced subgame.

Of course, some assumptions are required. First, the natural assumption that operating profits do not increase with the number of firms is imposed:

A1. For all $h \in H$, $\pi_h(y_h)$ is non-increasing in y_h.

Next, it is proper to allow some or all of the X_h values to be infinite; this would correspond to situations where there is no exit regardless of the number of active firms (e.g. due to the absence of scrap values and fixed operating costs). However, it is desirable to bound entry in each period so that the number of active firms is always finite (although perhaps unbounded over time). This is accomplished as follows. For each $h \in H$, let B_h be the collection of subsets of H defined as follows. If $\beta_h \in B_h$, then (1) $g > h$ for all $g \in \beta_h$, and (2) if $g \in \beta_h$ and $g' \in \beta_h$ then it is not the case that $g > g'$. In words, all market histories in β_h follow h but not each other. Given $\beta_h \in B_h$, define

$$\alpha_h = \{f \geq h \mid f < g \text{ for some } g \in \beta_h\}.$$

In words, all market histories in α_h (weakly) follow h but do not follow any market history in β_h. The next assumption can now be stated.

A2. For each $h \in H$ there exists a positive integer $\overline{y}_h$ such that, for all $y > \overline{y}_h$,

$$\sup_{\beta_h \in B_h} \sum_{\tau=t}^{\infty} \delta^{\tau-t} \left[\sum_{g \in H_\tau \cap \alpha_h} \rho(g \mid h)\pi_g(y) + \sum_{g \in H_\tau \cap \beta_h} \rho(g \mid h)\chi_g \right] < \xi_h.$$

Assumption A2 is used to bound the number of active firms at each market history h. Intuitively, if there are $y_h > \overline{y}_h$ active firms in the last period of market condition h then the marginal firm—that is, the first of the y_h firms to exit—cannot have an expected present value exceeding the left hand side of the inequality in A2. Since this is less than the entry cost, no more than $\overline{y}_h$ can enter in the last period of h. Thus, since $\overline{y}_h$ is finite for all $h \in H$, the number of active firms must always be finite (although it may increase over time without bound).

Finally, a technical assumption is required to ensure that expected present values converge as the horizon lengthens. This is accomplished by assumption A3:

A3. There exists $\kappa \in (\delta, 1]$ and $M > 0$ such that, for all t and for all $h \in H_t$, $|\pi_h(1)| < M/\kappa^t$ and $|\chi_h| < M/\kappa^t$.

The theorem can now be stated and rigorously proved. Note that the proof is constructive, and thus provides a characterization of an equilibrium. In particular, the evolution of the number of active firms over time is governed by (7.2).

Theorem: If A1-A3 are satisfied then a (subgame perfect Nash) equilibrium exists.

Proof: For each positive integer T and each market history $h \in H$, define two integers, N_h^T and X_h^T as follows. For $t > T$ and $h \in H_t$, let $N_h^T = X_h^T = 0$. For $h \in H_T$, define N_h^T and X_h^T as the largest integer values of y_h satisfying $\pi_h(y_h) - \xi_h \geq 0$ and $\pi_h(y_h) - \chi_h \geq 0$, respectively, if well-defined. Otherwise, if $\pi_h(1) - \xi_h < 0$ let $N_h^T = 0$; if $\pi_h(1) - \chi_h < 0$ let $X_h^T = 0$; and if $\pi_h(y_h) - \chi_h \geq 0$ for all y_h let $X_h^T = \infty$. (A2 implies N_h^T is finite.) For

$t < T$, and having defined N_g^T and X_g^T for all $g \in H_\tau$ and all $\tau > t$, define N_h^T and X_h^T for each $h \in H_t$ as the largest integer values of y_h that satisfy, respectively,

$$V_h^T(y_h) \equiv \sum_{\tau=t}^{T} \delta^{\tau-t} \left[\sum_{g \in H_\tau \cap \Theta} \rho(g \mid h)\pi_g(y_g) + \sum_{g \in H_\tau \cap \Phi} \rho(g \mid h)\chi_g \right] \geq \xi_h$$

and

$$V_h^T(y_h) \equiv \sum_{\tau=t}^{T} \delta^{\tau-t} \left[\sum_{g \in H_\tau \cap \Theta} \rho(g \mid h)\pi_g(y_g) + \sum_{g \in H_\tau \cap \Phi} \rho(g \mid h)\chi_g \right] \geq \chi_h$$

where

$$\Theta \equiv \{g \geq h \mid y_f \geq y_h \forall f \in [h, g]\}$$

$$\Phi \equiv \left\{g > h \mid g^{-1} \in \Theta \text{ and } y_g < y_h\right\}$$

and

$$y_g = \min\left\{X_g^T, \max\left\{N_g^T, y_{g^{-1}}\right\}\right\}.$$

If N_h^T or X_h^T is not well-defined in this way then proceed as follows. If $V_h^T(1) - \xi_h < 0$ let $N_h^T = 0$; if $V_h^T(1) - \chi_h < 0$ let $X_h^T = 0$; and if $V_h^T(y_h) - \chi_h \geq 0$ for all y_h let $X_h^T = \infty$. Note that Θ, Φ, and the y_g values all depend on y_h, which dependency is suppressed for notational clarity. The set Θ contains those future market histories before any of the active firms at h exit. The set Φ contains those future histories at which, for the first time, some firm that was active at h exits. Thus $V_h^T(y_h)$ is interpretable as the value of the marginal firm, that is, the active firm at h that will exit first.

Now index H by the positive integers and let $g(j)$ be the market history assigned to the integer j. Consider the sequences $\left\{N_{g(1)}^T\right\}_{T \in I}$ and $\left\{X_{g(1)}^T\right\}_{T \in I}$. Since $\left\{N_{g(1)}^T\right\}_{T \in I}$ is bounded (by A2), there exists $I_1 = \{T_{11}, T_{12}, ...\} \subset I$ such that the subsequences $\left\{N_{g(1)}^T\right\}_{T \in I_1}$ and $\left\{X_{g(1)}^T\right\}_{T \in I_1}$ converge to, say, $N_{g(1)}$ and $X_{g(1)}$, respectively (where $X_{g(1)}$ may be infinite). Having defined I_{j-1}, define $I_j = \{T_{j1}, T_{j2}, ...\} \subset I_{j-1}$ so the subsequences $\left\{N_{g(j)}^T\right\}_{T \in I_j}$ and $\left\{X_{g(j)}^T\right\}_{T \in I_j}$ converge to, say, $N_{g(j)}$ and $X_{g(j)}$, respectively (where $X_{g(j)}$ may be infinite). Finally, define the diagonal sequence $\overline{I} = \{T_{11}, T_{22}, ...\}$. Then, by construction, $\left\{N_g^T\right\}_{T \in \overline{I}}$ and $\left\{X_g^T\right\}_{T \in \overline{I}}$ converge to N_g and X_g, respectively, for all $g \in H$ (where X_g may be infinite).

Note that, by construction, $V_h^T(N_h^T) \geq \xi_h > V_h^T(N_h^T + 1)$ and $V_h^T(X_h^T) \geq \chi_h > V_h^T(X_h^T + 1)$ for all $h \in H$ and for all T if N_h^T and X_h^T are positive and finite. Thus, by A3, if N_h and X_h are both positive and finite then $V_h(N_h) \geq \xi_h \geq V_h(N_h + 1)$ and $V_h(X_h) \geq \chi_h \geq V_h(X_h + 1)$ for all $h \in H$, where for all $h \in H$,

$$V_h(y_h) \equiv \sum_{\tau=t}^{\infty} \delta^{\tau-t} \left[\sum_{g \in H_\tau \cap \Theta} \rho(g \mid h) \pi_g(y_g) + \sum_{g \in H_\tau \cap \Phi} \rho(g \mid h) \chi_g \right]$$

where Θ and Φ are as before and $y_g = \min\{X_g, \max\{N_g, y_{g^{-1}}\}\}$. If X_h is infinite then $V_h(y_h) \geq \chi_h$ for all y_h.

Construct firms' strategies as follows. Let $\{I_h\}_{h \in H}$ partition the set of initially inactive firms into countably infinitely many countably infinite subsets. Intuitively, I_h is the set of firms assigned to consider entry at the market history h. Order the y_0 initially active firms by their indexes i and assign them the integers from 1 to y_0. Let $\iota_0(i)$ be the integer initially assigned to the initially active firm i. Now, given a path of activity vectors, $\{a_g\}_{g \leq h}$, and having defined $\iota_{h^{-1}}(i)$ for each active firm at h^{-1}, define $\iota_h(i)$ for each active firm at h as follows. First order the active firms at h that were also active at h^{-1} by $\iota_{h^{-1}}(i)$, then order new entrants that belong to the set I_h by their indexes i, and finally order new entrants that do not belong to the set I_h by their index numbers i. (A new entrant is a firm for which $a_h^i - a_{h^{-1}}^i = 1$.) Having ordered the active firms at h, assign the positive integers between 1 and y_h to them and let $\iota_h(i)$ be the integer assigned to the active firm i at h. For each initially inactive firm, define

$$\eta^i \equiv \sum_{j \in I_{h(i)} \cap \{j \mid j < i\}} (1 - a_{h(i)^{-1}}^j)$$

where $h(i)$ is the history such that $i \in I_{h(i)}$. Thus η^i is the number of firms in $I_{h(i)}$ with lower indexes than firm i that are not active coming into the last period of the market history $h(i)$. Firm i's equilibrium strategy $\widetilde{\sigma}^i$ is given by

$$\begin{aligned} \text{If } a_{h^{-1}}^i &= 0 \text{ then } \widetilde{a}_h^i = 1 \text{ iff } h = h(i) \text{ and } N_h - y_{h^{-1}} > \eta^i. \\ \text{If } a_{h^{-1}}^i &= 1 \text{ then } \widetilde{a}_h^i = 1 \text{ iff } X_h \geq \iota_h(i). \end{aligned}$$

Note that the construction of the strategies guarantees that

$$y_g = \min\{X_g, \max\{N_g, y_{g^{-1}}\}\}$$

for all $g \in H$ and that the most recent entrant is the first to exit. Thus the marginal firm's value—that is, the value of the active firm at h that will be the first to exit—is given by $V_h(y_h)$ along the equilibrium path.

Now consider all possible one-shot deviations at an arbitrary history s. First suppose an inactive firm deviates by entering when its equilibrium strategy dictates otherwise. By construction, it will be designated as the marginal firm and will earn $V_h(\widetilde{y}_h + 1) - \xi_h \leq 0$ where $\widetilde{y}_h$ is the equilibrium number of firms at h. Next suppose an inactive firm deviates by not entering when it is supposed to. It then receives a payoff of zero, which is no greater than the $V_h(\widetilde{y}_h) - \xi_h \geq 0$ it would gain by following its equilibrium strategy. Next suppose an active firm exits when its equilibrium strategy dictates otherwise. It then receives a payoff of χ_h which cannot exceed its equilibrium value of $V_h(\widetilde{y}_h) \geq \chi_h$. Finally, suppose an active firm does not exit when its strategy dictates to exit. Then it will be designated as the marginal firm and will have a value of $V_h(\widetilde{y}_h + 1) \leq \chi_h$, which cannot exceed its payoff of χ_h for exiting. It follows that no one-shot deviation from equilibrium—and hence no deviation from equilibrium—is profitable. Thus the constructed strategy profile constitutes a subgame perfect Nash equilibrium. **Q.E.D.**

8 Excessive Entry and Insufficient Exit

The traditional industrial organization literature often implicitly assumed that there is insufficient entry in imperfectly competitive markets. By contrast, Mankiw and Whinston (1986) established a tendency toward excessive entry in a simple two-stage framework where in the first stage firms make entry decisions and in the second stage active firms make production decisions. ("Tendency" means that the number of entrants never falls more than one short of the socially optimal number and can far exceed it, where the socially optimal number is what would be chosen by a social planner who could control entry but not production decisions.) Intuitively, firms do not take into account the effects of their activities on the profits of their competitors, and the negative effects of the marginally active firm on aggregate profits (including the additional entry costs) tend to overcome the positive effects on consumer surplus. The reason only a tendency can be established—that is, the number of active firms can be one less than optimal—reflects the integer constraint.

The general, stochastic, infinite-horizon model of Sections 6 and 7 allows a much stronger result. In the class of equilibria constructed in Section 7—that is, in limit-lifo equilibria—not only is there a tendency for excessive entry when times are good but there is also a tendency for insufficient exit when times are bad. Furthermore, these tendencies are not reversed in the present even when firms take into account that there will tend to be too many firms in the future. Thus, equilibrium tends to exhibit too many active firms in *every* period of *every* realization of the stochastic process.[5] This theorem requires additional notation and assumptions—analogous to those used by Mankiw and Whinston—which will now be introduced.

Assume each market condition is completely described by a differentiable downward-sloping inverse demand function and a twice differentiable, non-negative, convex cost function. Given an initial number of active firms, let $\mathbf{y}^* = \{y_h^*\}_{h \in H}$ be the socially optimal stochastic sequence of numbers of firms given that the social planner can control entry and exit but cannot control the production and pricing decisions of active firms. Formally, $\mathbf{y}^*$ maximizes

$$\begin{aligned} W(\mathbf{y}) &\equiv \sum_{\tau=1}^{\infty} \sum_{h \in H_\tau} \rho_h \left[\begin{array}{c} \int_0^{Q_{y_h}} P_h(s)ds - y_h c_h(q_{y_h}) \\ - \max\{0, y_h - y_{h^{-1}}\}\xi_h + \max\{0, y_{h^{-1}} - y_h\}\chi_h \end{array} \right] \\ &\equiv \sum_{\tau=1}^{\infty} \sum_{h \in H_\tau} \rho_h \left[S_h(y_h) - \max\{0, y_h - y_{h^{-1}}\}\xi_h + \max\{0, y_{h^{-1}} - y_h\}\chi_h \right] \end{aligned}$$

where $Q_{y_h} = y_h q_h$ is total output when there are y_h firms at h, P_h is the inverse demand function at h, c_h is the individual firm's cost function at h, q_{y_h} is the equilibrium per-firm output given y_h firms at the history h and ρ_h is the unconditional probability of the market history h occurring. $S_h(y_h)$ will be called the *current surplus* at h. For all h, assume (following Mankiw and Whinston):

A4. $Q_{y_h} > Q_{y'_h}$ for all $y_h > y'_h$ and $\lim_{y_h \to \infty} Q_{y_h} < \infty$.

A5. $q_{y_h} < q_{y'_h}$ for all $y_h > y'_h$.

[5] This results contrasts with Rob's (1991) continuum model where, due to learning externalities, the model exhibits insufficient entry.

A6. $P_h(Q_{y_h}) - c'_h(q_{y_h}) \geq 0$ for all y_h.

In words, in each period of each realization, industry output increases and is bounded in the number of firms, per-firm output is decreasing in the number of firms, and price is no less than the marginal cost given the number of firms. These assumptions are not very restrictive: Amir and Lambson (2000) provide (minimal) conditions sufficient for A4 and A5 to hold for Cournot equilibria. Specifically, defining $\Delta \equiv -P'(Q) + c''(q)$, they show that the first part of A4—namely $Q_{y_h} > Q_{y'_h}$ for all $y_h > y'_h$—holds if $\Delta > 0$ globally.[6] This is a very general condition, implied in particular by our assumption that P is downward-sloping and c is convex. They also show that A5 holds if $\Delta > 0$ globally and $\log P$ is a concave function, which still covers most examples of interest. On the other hand, they show that if $\Delta < 0$ globally (which requires a strongly concave cost function c), the opposite of the first part of A4 holds. Also, with $\Delta > 0$, A5 may fail if $\log P$ is convex. Thus, while A4 and A5 are not universal, they are satisfied under very broad conditions covering most cases of interest in Cournot models. A6 follows directly from the best reply property of Cournot equilibria.

Given an initial number of active firms, let $\mathbf{y}^e = \{y_h^e\}_{h \in H}$ be a stochastic limit-lifo equilibrium sequence of numbers of firms. The theorem asserts that the equilibrium number of firms never falls short of the optimal number by more than one. It is not difficult to construct examples where the equilibrium number of firms greatly exceeds the optimal number.

Theorem: For all $h \in H$, $y_h^e \geq y_h^* - 1$.

Proof: First note that, for all $h \in H$ and for all $y_h > 1$,

$$\begin{aligned} S_h(y_h) - S_h(y_h - 1) &= \int_{Q_{y_h - 1}}^{Q_{y_h}} P_h(s)ds - y_h c_h(q_{y_h}) + (y_h - 1)c_h(q_{y_h - 1}) \\ &= P_h(Q_{y_h-1})q_{y_h-1} - c_h(q_{y_h-1}) \\ &\quad - \left[P_h(Q_{y_h-1})q_{y_h-1} - \int_{Q_{y_h-1}}^{Q_{y_h}} P_h(s)ds + y_h \left[c_h(q_{y_h}) - c_h(q_{y_h-1})\right] \right] \\ &\equiv \pi_h(y_h - 1) - M_h^{y_h}. \end{aligned}$$

[6] The second part of A4—namely, $\lim_{y_h \to \infty} Q_{y_h} < \infty$—holds, for example, if $P(Q) < c'(0)$ for some Q.

Assumptions A4-A6 imply $M_h^{y_h} > 0$, because

$$\begin{aligned}
M_h^{y_h} &\equiv P_h(Q_{y_h-1})q_{y_h-1} - \int_{Q_{y_h-1}}^{Q_{y_h}} P_h(s)ds + y_h\left[c_h(q_{y_h}) - c_h(q_{y_h-1})\right] \\
&\geq P_h(Q_{y_h-1})q_{y_h-1} - P_h(Q_{y_h-1})\left[Q_{y_h} - Q_{y_h-1}\right] + y_h\left[c_h(q_{y_h}) - c_h(q_{y_h-1})\right] \\
&\geq P_h(Q_{y_h-1})q_{y_h-1} - P_h(Q_{y_h-1})\left[Q_{y_h} - Q_{y_h-1}\right] + y_h c_h'(q_{y_h-1})\left[q_{y_h} - q_{y_h-1}\right] \\
&= \left[P_h(Q_{y_h-1}) - c_h'(q_{y_h-1})\right] y_h\left[q_{y_h-1} - q_{y_h}\right] \\
&> 0.
\end{aligned}$$

The first line is definitional, the second is from A4 and $P' < 0$, the third is from $c'' \geq 0$, the fourth is from $Q_y/y = q_y$ (symmetry), and the fifth is from A5 and A6.

Consider the T-period truncation of the model. Let $\mathbf{y}^*(T)$ be the optimal stochastic sequence of numbers of firms for the T-period truncation of the model; that is, $\mathbf{y}^*(T)$ maximizes

$$W_T(\mathbf{y}) \equiv \sum_{\tau=1}^{T}\sum_{h\in H_\tau}\rho_h\left[S_h(y_h) - \max\{0, y_h - y_{h^{-1}}\}\xi_h + \max\{0, y_{h^{-1}} - y_h\}\chi_h\right].$$

The strategy of proof is as follows. First show, by backward induction, that $y_h^e(T) \geq y_h^*(T) - 1$ for all $h \in H_t$, all $t \leq T$, and all T. Then limiting arguments as $T \to \infty$ establish $y_h^e \geq y_h^* - 1$ for all $h \in H$.

Consider an arbitrary T and arbitrary $h \in H_T$. If $y_h^* \leq 1$, then it is trivial that $X_h^T \geq N_h^T \geq y_h^* - 1$. If $y_h^* \geq 2$ then let $\mathbf{y}$ be the truncated path that is identical to $\mathbf{y}^*$ except that $y_h = y_h^* - 1$. By the optimality of $\mathbf{y}^*$,

$$\begin{aligned}
W_T(\mathbf{y}^*) - W_T(\mathbf{y}) &= \delta^{T-1}\left[S_h(y_h^*) - S_h(y_h^* - 1) - \gamma_h\right] \\
&= \delta^{T-1}\left[\pi_h(y_h^* - 1) - M_h^{y_h^*} - \gamma_h\right] \\
&\geq 0
\end{aligned}$$

where $\gamma_h = \xi_h$ if $y_h^* > y_{h^{-1}}^*$ and $\gamma_h = \chi_h$ otherwise. Since $M_h^{y_h^*} > 0$, this implies $\pi_h(y_h^* - 1) > \gamma_h$. Thus, $X_h^T \geq y_h^* - 1$ and, furthermore, if $y_h^* > y_{h^{-1}}^*$ then $N_h^T \geq y_h^* - 1$.

Now consider $h \in H_t$ for $t < T$ and make the induction hypothesis that, for all $g > h$ in the truncated model, $X_g^T \geq y_g^* - 1$ and, furthermore, that if $y_g^* > y_{g^{-1}}^*$ then $N_g^T \geq y_g^* - 1$.

If $y_h^* \leq 1$, then it is trivial that $X_h^T \geq N_h^T \geq y_h^* - 1$. If $y_h^* \geq 2$ then consider the path $\mathbf{y}$ that is identical to $\mathbf{y}^*$ except $y_h = y_h^* - 1$ and $y_g = \min\left\{X_g^T, \max\left\{N_g^T, y_{g^{-1}}\right\}\right\}$ for $g > h$; in words, $\mathbf{y}$ follows the optimal path until h, exhibits one less firm at h, and then is induced by the equilibrium strategies thereafter. Define three sets:

$$A = \left\{g \geq h \mid y_f = y_h^* - 1 \text{ and } y_f^* = y_h^* \ \forall \text{ f} \in [h, g]\right\}$$

$$B = \left\{g \notin A \mid g^{-1} \in A \text{ and } y_g^* \geq y_h^*\right\}$$

$$C = \left\{g \notin A \mid g^{-1} \in A \text{ and } y_g^* < y_h^*\right\}.$$

The first set, A, is the set of histories exhibiting no change in either path from the number of firms at h. The second and third sets together are the set of "first times" that either of the paths exhibits entry or exit; B is the set of such "first times" that exhibits no exit along the optimal path (and hence exhibits entry or exit along the path $\mathbf{y}$) and C is the set of such "first times" that exhibits exit along the optimal path.

If $g \in B$ then either $y_g^* > y_h^* = y_{g^{-1}}^*$ or $y_g^* = y_h^* = y_{g^{-1}}^*$. If $y_g^* > y_{g^{-1}}^*$ then, by the induction hypothesis, $N_g^T \geq y_g^* - 1 > y_{g^{-1}}$; thus $V_g^T(y_g) \geq \xi_g$ (because, intuitively, there is entry along the path $\mathbf{y}$ induced by equilibrium strategies). If $y_g^* = y_{g^{-1}}^*$ and $g \in B$ then $y_g \neq y_{g^{-1}} = y_g^* - 1$. Since, by the induction hypothesis, $y_g^* - 1 \leq X_g^T$, it must be that $y_g > y_{g^{-1}}$. Thus $y_g = N_g^T$ and $V_g^T(y_g) \geq \xi_g$. Thus $V_g^T(y_g) \geq \xi_g$ for all $g \in B$. Furthermore, it is trivial that $V_g^T(y_g) \geq \chi_g$ for all $g \in C$.

Now define the path $\mathbf{y}'$ for the truncated model by $y_g' = y_h^* - 1$ for $g \in A$ and $y_g' = y_g^*$ otherwise. By the optimality of $\mathbf{y}^*$,

$$\begin{aligned} & W_T(\mathbf{y}^*) - W_T(\mathbf{y}') \\ = & \sum_{\tau=t}^{T} \delta^{\tau-t} \left[\sum_{g \in A_\tau} \rho(g \mid h)\pi_g(y_h^* - 1) - M_g^{y_h^*} + \sum_{g \in B_\tau} \rho(g \mid h)\xi_g + \sum_{g \in C_\tau} \rho(g \mid h)\chi_g\right] - \gamma_h \\ \geq & \ 0 \end{aligned}$$

where $\gamma_h = \xi_h$ if $y_h^* > y_{h^{-1}}^*$, $\gamma_h = \chi_h$ otherwise, and where $A_\tau \equiv A \cap H_\tau$, $B_\tau \equiv B \cap H_\tau$, and $C_\tau \equiv C \cap H_\tau$. Since $M_g^{y_h^*} > 0$ for all g, this implies

$$\sum_{\tau=t}^{T} \delta^{\tau-t} \sum_{g \in A_\tau} \rho(g \mid h) \pi_g(y_h^* - 1)$$
$$\geq \gamma_h - \sum_{\tau=t}^{T} \delta^{\tau-t} \sum_{g \in B_\tau} \rho(g \mid h) \xi_g - \sum_{\tau=t}^{T} \delta^{\tau-t} \sum_{g \in C_\tau} \rho(g \mid h) \chi_g$$

By definition

$$\begin{aligned} V_h^T(y_h^* - 1) &= \sum_{\tau=t}^{T} \delta^{\tau-t} \left[\sum_{g \in A_\tau} \rho(g \mid h) \pi_g(y_h^* - 1) + \sum_{g \in B_\tau \cup C_\tau} \rho(g \mid h) V_g^T(y_g) \right] \\ &\geq \sum_{\tau=t}^{T} \delta^{\tau-t} \left[\sum_{g \in A_\tau} \rho(g \mid h) \pi_g(y_h^* - 1) + \sum_{g \in B_\tau} \rho(g \mid h) \xi_g + \sum_{g \in C_\tau} \rho(g \mid h) \chi_g \right] \end{aligned}$$

where $y_g = \min\left\{X_g^T, \max\left\{N_g^T, y_{g^{-1}}\right\}\right\}$. The previous two inequalities together imply

$$V_h^T(y_h^* - 1) \geq \gamma_h.$$

Thus, $X_h^T \geq y_h^* - 1$ and, furthermore, if $y_h^* > y_{h^{-1}}^*$ then $N_h^T \geq y_h^* - 1$. Thus, by induction, for all $h \in H_t$ and $t \leq T$, $X_h^T \geq y_h^* - 1$ and, furthermore, if $y_h^* > y_{h^{-1}}^*$ then $N_h^T \geq y_h^* - 1$. With this result in hand it follows immediately that, for the same initial conditions, the equilibrium path satisfying $y_h^e = \min\left\{X_h^T, \max\left\{N_h^T, y_{h^{-1}}^e\right\}\right\}$ will always satisfy $y_h^e \geq y_h^* - 1$ in the truncated model. Intuitively, when there is entry along the optimal path there will be at least $N_h^T \geq y_h^* - 1$ firms on the equilibrium path (because $y_h^e \geq N_h^T$ for all h). Thus $y_h^e < y_h^* - 1$ could only arise if there is exit along the equilibrium path; but if there is exit at h along the equilibrium path then $y_h^e = X_h^T \geq y_h^* - 1$.

Thus, for all T, the result holds for the truncated model. Taking limits (using diagonalization arguments as described in Section 7) then establishes the result. **Q.E.D.**

By its very construction, our infinite-horizon equilibrium is the limit of finite-horizon equilibria. As is well known, an infinite-horizon game may also exhibit equilibria that are not related in any direct way to finite-horizon equilibria. This very common feature of dynamic games is sometimes referred to as the failure of lower hemi-continuity of the equilibrium correspondence at infinity. (See, for example, Fudenberg and Tirole (1991).) In most studies of both theory and application of dynamic games, attention is restricted to equilibria that are limits of their finite-horizon counterparts. This is even true of the class of linear-quadratic games, widely used in economics and in particular in strategic models of macroeconomics. In the linear-quadratic framework, there is a unique equilibrium in the infinite-horizon case with strategies that are linear functions of the state. This equilibrium is easy to compute and characterize, and is always the limit of the unique (Markov) equilibrium of the finite-horizon version of the same game. While virtually all studies of linear-quadratic games limit consideration to the linear equilibrium, the infinite-horizon game often has other equilibria that are typically not related to the finite-horizon equilibrium. (See Basar and Olsder (1999) for further details on this point for linear-quadratic games.[7]) The following example—due to an anonymous contributor—is a reminder that restricting attention to limit equilibria is not without loss of generality in our framework either. In particular, the Mankiw-Whinston tendency for excessive entry need not hold.

Suppose the same Cournot game is repeated in all periods, where inverse demand is $P = 1 - Q$ and production costs are zero. Thus per-firm operating profit in a period with y active firms is $\pi(y) = \left(\frac{1}{y+1}\right)^2$. Suppose $\xi = 2$, $\chi = 1.7$, and $\delta = .99$. It is straightforward to establish that a social planner would maximize discounted producer and consumer surplus (subject to Cournot behavior) by having three firms enter in the first period and allowing no further entry or exit. There exist equilibria, however, that exhibit no entry at all along the equilibrium path. One of the simplest is defined as follows. Given a_{t-1}, let ι be the index of the active firm at time $t-1$ with the largest index; formally, $\iota = \max_{i \in \{i | a^i_{t-1}=1\}} i$. (If there are infinitely many active firms then $\iota = \infty$.} Let $\#a_{t-1}$ be the number of active firms in period $t-1$ and set $\#a_0 = 0$. For each i, define firm i's strategy as follows: for each t, $a^i_{t-1} = 1$ if $\#a_{t-1} = 6$ and $a^i_{t-1} = 1$, or if $\#a_{t-1} \notin \{0, 6, \infty\}$ and $i \in \{\iota + 1, ..., \iota + 6\}$, or if $\#a_{t-1} = \infty$ and $i \in \{1, ..., 6\}$; $a^i_{t-1} = 0$ otherwise. Intuitively, since $\pi(6)/(1-\delta) > \xi > \chi > \pi(7)/(1-\delta)$, there are subgame perfect Nash equilibria exhibiting entry by six firms and no further entry or exit. These

[7]Another well-known example is the resource extraction model analysed by Levhari and Mirman (1980). They also restrict consideration to the unique infinite-horizon equilibrium (also linear in the stock) that they obtain as a limit of the finite-horizon equilibrium while other infinite-horizon equilbria may exist even in their simple model.

equilibria can be used as "threats" to dissuade all entry. Specifically, given the strategies of the other firms, if a firm ever enters it will enjoy monopoly profit for one period but will be joined thereafter by six new firms; it will thus find it optimal to exit without having recouped its entry cost.

9 Concluding Remarks

This paper has developed a general, stochastic, infinite-horizon model of entry and exit with imperfect competition. It should be applicable to many traditional industrial organization questions, incorporating as it does an explicitly dynamic framework and requiring an integer number of firms. Some examples of applications were presented in the paper, where it was shown that high sunk costs can actually reduce average industry concentration over time, that the lumpiness of the technology that results in imperfect competition can explain entry when current profits are negative, and that the amount of entry may not be positively correlated with the length of the product cycle. Going beyond examples, a more general treatment of the welfare properties of equilibrium entry and exit was developed. In particular, a tendency toward excessive entry and insufficient exit when there is imperfect competition (a la Mankiw and Whinston (1986)) was shown to hold for all equilibria obtained as limits of finite-horizon equilibria using a last-in-first-out exit rule. On the other hand, a counterexample showed that this tendency need not be exhibited by equilibria that are not limits of finite-horizon equilibria.

References

[1] Amir, R. (1996), "Continuous Stochastic Games of Capital Accumulation with Convex Transitions," *Games and Economic Behavior, 15, 111-131.*

[2] Amir, R., and V. E. Lambson (2000), "On the Effects of Entry in Cournot Markets," *Review of Economic Studies*, 67, 235-254.

[3] Basar, T. and G. Olsder (1999), *Dynamic Noncooperative Game Theory*, Academic Press, New-York.

[4] Bentolila, S. and G. Bertola (1990), "How Bad is Eurosclerosis?" *Review of Economic Studies*, 57, 381-402.

[5] Bernheim, D. (1984), "Strategic Deterrence of Sequential Entry into an Industry," *The Rand Journal of Economics*, 15, 1-11.

[6] Demsetz, H. (1974), "Two Systems of Belief about Monopoly," in Harvey Goldschmid et al., eds., *Industrial Concentration: The New Learning*, Boston: Little, Brown.

[7] Dewey, D. (1976), "Industrial Concentration and the Rate of Profit: Some Neglected Theory," *Journal of Law and Economics*, 19, 67-78.

[8] Dixit, A. (1989), "Entry and Exit Decisions Under Uncertainty," *Journal of Political Economy*, 97, 620-638.

[9] Ericson, R. and A. Pakes (1995), "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62, 53-82.

[10] Fudenberg, D. and J. Tirole (1985), "Preemption and Rent Equalization in the Adoption of New Technology," *Review of Economic Studies, 52, 383-402.*

[11] Fudenberg, D. and J. Tirole (1987), "Understanding Rent Dissipation: On the Use of Game Theory in Industrial Organization," *American Economic Review*, 77, 176-183.

[12] Fudenberg, D. and J. Tirole (1991), GameTheory, MIT Press.

[13] Hopenhayn, H. (1992), "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60, 1127-1150.

[14] Jovanovic, B. (1982), "Selection and the Evolution of Industry," *Econometrica*, 50, 649-670.

[15] Klepper, S. (1996), "Entry, Exit, Growth, and Innovation over the Product Cycle," *American Economic Review*, 86, 562-583.

[16] Lambson, V. E. (1987), "Is the Concentration-Profit Correlation Partly an Artifact of Lumpy Technology?" *American Economic Review*, 77, 731-733.

[17] Lambson, V. E. (1991), "Industry Evolution with Sunk Costs and Uncertain Market Conditions," *International Journal of Industrial Organization*, 6, 263-271.

[18] Lambson, V. E. (1992), "Competitive Profits in the Long Run," *Review of Economic Studies*, 59, 125-142.

[19] La Manna, M. (1986), "Why Are Profits Correlated with Concentration?" *Economics Letters*, 22, 81-85.

[20] Levhari, D. and L. Mirman (1980), The great fish war: An example using a dynamic Cournot-Nash solution, *Bell Journal of Economics*, 322-344.

[21] Mankiw, N. G. and M. D. Whinston (1986), "Free Entry and Social Inefficiency," *The Rand Journal of Economics*, 17, 48-58.

[22] Rob, R. (1991), "Learning and Capacity Expansion Under Demand Uncertainty," *Review of Economic Studies*, 58, 655-675.

[23] Weiss, L. (1974), "The Concentration-Profits Relationship and Antitrust," in Harvey Goldschmid et al., eds., *Industrial Concentration: The New Learning*, Boston: Little, Brown.