Equilibrium dynamics of entry and exit: industry-wide learning and endogenous heterogeneity

Takashi Kamihigashi
Kobe University, Kobe, Japan

Santanu Roy
Southern Methodist University, Dallas, Texas.

September 13, 2019

Abstract

Endogenous heterogeneity induced by the market explains important empirical regularities on industry dynamics. In a *deterministic* dynamic competitive industry where cost reducing investment generates industry-wide learning, we show that the market may create incentives for differences in entry and exit decisions and in age, size and performance of *ex ante* identical firms. Under verifiable conditions, entry occurs with delay, entry continues indefinitely, shake out of firms occurs and shake out continues indefinitely. Older firms are larger and less likely to exit (last in first out); exiting firms are younger and smaller. Market structure converges in the long run.

JEL: L11,D41,D92,O3

Keywords: Industry Dynamics, Dynamic Competitive Equilibrium, Industry-wide Learning, Investment, Cost Reduction, Entry, Exit, Shake out, Heterogeneity.

1 Introduction

It is well known that markets may offer incentives for participants to make significantly different choices even if their *ex ante* characteristics and possibilities are very similar. In competitive industries, the dynamic equilibrium process can create and magnify differences between firms by providing incentives for asymmetric market decisions by symmetric firms. Market induced differences in investment (in technological change, accumulation of knowledge and other forms of capital formation) can lead to sustained differences in scale, size, productivity and profitability of firms over time. Such differences may also be reflected in the diversity of entry and exit decisions of firms. The purpose of this paper is to explore the extent to which such market induced heterogeneity can explain some of the well known qualitative features of industry dynamics.

Over the past few decades, analysis of manufacturing census data and empirical studies of product life cycles for a wide range of industries have yielded a wealth of information regarding the dynamics of industries. It is now well understood that there is high dispersion in size and growth of firms as well as high rates of exit, entry and turnover of firms.[1] While entry peaks in the early phase of the product life cycle and shake-out or exit of firms is strong as the industry attains maturity[2], both entry and exit are fairly robust phenomenon and persist over long periods. Eventually, both entry and exit peter out and the industry structure stabilizes. Prices generally decline and industry size or output expands as the industry matures. Size and age are positively correlated; firms that enter earlier are more likely to grow faster, tend to be larger in size and have a greater chance of survival. Firms that exit the industry are likely to be smaller and younger than incumbents.[3]The key to explaining many of these empirical regularities is inter-firm heterogene-

[1]For example, Dunne, Roberts and Samuelson (1989)report rates of entry ranging from 30.7% to 42.7% and an equally dramatic exit rate ranging from 30.8% to 39% across manufacturing industries over a period of five years. See also, Davis and Haltiwanger (1992).

[2]Gort and Klepper (1982) and Klepper and Grady (1990) report that modal shakeout involved a decline of about 50% in the number of producers; in extreme cases like autos and tires the number of firms declined by about 90% in ten to fifteen years.

[3]For a nice summary of these and other regularities, see Klepper (1996).

ity. Existing models of industry dynamics generate heterogeneity among firms through two channels: heterogeneity in exogenous firm specific characteristics such as initial conditions prior to entry (including technological opportunities), and differences in realizations of random processes that affect the characteristics of firms. However, as mentioned above, heterogeneity among firms may emerge endogenously even if there are no differences in their initial conditions or dynamic opportunities, and even if differences among firms are not created by random shocks. The market may create incentives for identical firms with identical opportunities *to choose to be different* thereby creating heterogeneity in ex post characteristics. The specific aim of this paper is to examine whether market generated *endogenous* heterogeneity can explain the above mentioned empirical regularities on industry dynamics, and to relate this to the nature of technological and demand conditions in the industry.

In order to do this, we consider a *deterministic* model of a dynamic perfectly competitive industry where there is no uncertainty and all firms are *ex ante* identical. Firms are fully rational, forward looking and have equal opportunity to enter and exit the industry at any point of time. This implies that ex post heterogeneity among firms and the related dynamics of industry structure are entirely due to choices made by firms in response to incentives that arise from the equilibrating process of the market. Every period that a firm is active in the industry, it has the option of investing in firm-specific capital that improves its future productivity i.e., reduces its own production cost in the future. Firm specific capital formation may reflect, among other things, learning or accumulation of knowledge as well as organizational capital formation. In addition, aggregate investment activity in the industry also leads to accumulation of industry-wide capital that generates positive externality for all firms in the industry by reducing their future production cost. Industry-wide capital may reflect, among other things, the stock of industry-wide knowledge that is accessible by all firms including those that are not currently active. Firms have upward sloping marginal cost curves that are (weakly) decreasing in the stock of both firm-specific and industry-wide capital. Market demand is identical and independent over time. With a fairly general structure, we obtain a number of strong qualitative results that provide a theoretical foundation for the empirical regularities mentioned above.

First, prices decline and industry output expands over time. Despite the externality (so that the industry equilibrium path is not necessarily the first best) and the possible turnover of firms, the industry supply curve always expands (weakly) and the industry's marginal cost curve can only decline over time. Firms faces lower prices as they mature, and must have lower production cost through capital formation to remain profitable. Second, some firms enter later than others and indeed, entry may be significantly dispersed over time: the market may create incentives for identical firm to make different entry decisions. This is directly related to the externality created by industry-wide learning or capital formation. If production cost is independent of industry-wide capital, entry occurs only in the initial period; a firm entering later would face lower prices at each age of its active life in the industry (compared to an earlier entrant) and therefore, would be better off entering earlier. Positive externality from formation of industry-wide capital can help offset the price disadvantage faced by later entrants. Continued entry over time is most likely when the externality effect is strong and in addition, capital formation reduces the fixed cost of production more sharply than the marginal cost (i.e., the efficient scale contracts with capital formation), and if demand is relatively elastic. We outline verifiable sufficient condition on technology and demand under which entry occurs infinitely often. Third, while some firms remain active forever, others exit in finite time and the industry may exhibit significant shake-out of firms over time. In fact, even firms that enter in the same period may exit in different time periods. While this possibility has been shown earlier in a finite horizon version of this model with no externality (where all firms enter in the initial period)[4], we extend this to a more general framework where entry is not necessarily concentrated in the initial period and establish verifiable conditions for a stronger form of shake-out where exit continues to occur indefinitely. Shake-out is more likely to occur if demand is relatively price inelastic, and if the marginal cost curve shifts sufficiently with capital

[4] Petrakis and Roy (1999).

accumulation. Fourth, for any active firm, the optimal stock of (firm specific) capital next period is increasing in its current capital stock. This monotonicity property can be used to show that the cross section ordering of firms in terms of firm-specific capital, productivity, size and profitability (at any point of time) is preserved over the lifetime of active firms. In any period where exit occurs, firms that remain active in the industry have at least as large firm-specific capital stock, size and profit as the firms that exit. A firm that enters earlier holds at least much firm-specific capital and produces at least as much output as a firm that enters later (in every period for which they are both active in the industry). As long as there is some overlap between the time intervals over which they are active in the industry, a younger firm never exits later than an older firm and therefore, exiting firms are never older than incumbents that do not exit, and may often be younger. In other words, the equilibrium path is characterized by "last in first out" dynamics.[5] Finally, despite the fact that entry and exit may continue to occur indefinitely, the volume of entry and exit converges to zero in the long run. In other words, the market structure is convergent and there is no turnover of firms in the limiting state of the industry.

Our paper is most closely related to a small literature on characterization of deterministic competitive industry dynamics with endogenous entry, exit and technological change.[6] Petrakis, Rasmusen and Roy (1997) characterize the equilibrium path of a deterministic two-period competitive industry with entry, exit and *learning by doing,* where firms reduce future production cost through experience and there are no externalities. They show that even though all firms are initially identical, some firms may exit at the end of the

[5]A recent paper on oligopolistic industry dynamics with sunk cost and demand uncertainty by Abbring and Campbell (2010) *assumes* last in first out entry and exit dynamics to select a unique Markov-perfect equilibrium.

[6]Competitive (partial) equilibrium theory of investment and industry dynamics with price taking firms was first developed by Lucas and Prescott (1971) in a model with uncertainty; they did not allow for entry, exit or externalities. Hopenhayn (1990) establishes results on existence and social optimality of dynamic industry equilibrium with entry and exit (but no externalities) for a very general class of technology and stochastic shocks, both aggregate and firm specific.

first period; firms that do not exit overproduce in the first period (in order to learn), thus generating cross-section heterogeneity in size in the first period. If the cost function is convex, there is no delayed entry in their framework.[7] Closer to our framework is the one analyzed by Petrakis and Roy (1999); they study a finite horizon version of the model in this paper but without any externalities.[8] The latter implies that all entry occurs in the initial period and all active firms are of the same age; they focus on the possibility that some firms may exit before the terminal date, and that difference in planned length of stay may cause differences in investment and scale of firms. In this paper, we characterize conditions for a stronger form of shake-out where exit continues to occur indefinitely. More importantly, our paper shows that in the presence of industry-wide externalities, entry may be dispersed over time and may continue to occur indefinitely under certain conditions; because of dispersed entry, the equilibrium path can generate predictions about the age structure of firms that match empirical observations. We also provide predictions about the long run behavior of the industry (that was not possible in a finite horizon model), and a fuller characterization of changes in cross-section distribution of firm characteristics. The literature on dynamic oligopoly games with strategic investment capacity (and other forms of capital accumulation) by a fixed number of firms has shown the possibility of asymmetric equilibrium outcomes in a symmetric model, and endogenous heterogeneity in size and efficiency (see, among others, Flaherty, 1980, Maggi, 1996, Reynolds and Wilson, 2000, Besanko and Doraszelski, 2004). Our paper abstracts from strategic interaction by looking at a perfectly competitive industry; however, we allow for endogenous entry and exit of firms.

Finally, as mentioned above, there is a large literature on theoretical models of stochastic evolution and selection in competitive industries that explain some of the empirical regularities relating to industry dynamics. In this literature, heterogeneity among firms is primarily induced by differences in real-

[7] Stokey (1986) considers a dynamic industry with imperfect competition among a fixed number of firms, where aggregate output generates industry-wide learning by doing.

[8] See also, Bester and Petrakis (2003).

izations of random shocks or uncertain events affecting the characteristics and technological possibilities of the firm (though market incentives and market selection may magnify these differences). In contrast, our framework is fully deterministic. Closest to our paper, is the seminal paper by Hopenhayn (1992a) that generates entry and exit as part of the limiting behavior of a stochastic dynamic competitive industry with investment by firms and no externalities[9]. Bergin and Bernhardt (2008) study the affect of aggregate demand shocks in a similar framework. Jovanovic (1982) analyzes the dynamics of a competitive industry where firms are uncertain about their productivity and acquire noisy signals about their efficiency as they operate in the industry; incumbent firms afflicted by unfavorable signals conclude they are inefficient and exit the market to be replaced by new entrants; the efficient grow and survive, while the inefficient decline and fail (see also, Lippman and Rumelt, 1982, Pakes and Ericson, 1998). Jovanovic and Lach (1989) consider a model with learning by doing and stochastic diffusion of innovation where potential entrants can gain by learning from incumbent firms but all learning stops after entry; the model generates delayed entry and staggered exit. Jovanovic and MacDonald (1994) analyze a dynamic competitive industry where innovational opportunities fuel entry and failure to innovate, whose chances are exogenously specified, leads to exit. Klepper and Graddy (1990) discuss an evolutionary model where the number of potential entrants is limited, potential entrants differ in their initial cost and product qualities, receive new information over time which changes their cost and product quality in a stochastic fashion and no further updating of cost and quality occurs after entry.[10]In a somewhat different approach, Lambson (1991) analyzes a dynamic competitive model where firms make investments that entail sunk costs and the return on investment is affected by stochastic shocks; the equilibrium path generates high turnover of plants (see also, Dixit, 1989). Ericson and Pakes (1995) analyze the Markov perfect equilibria of a dynamic industry game with entry and exit where firms invest to improve future profitability and are affected by idiosyncratic shocks. They

[9]See also, Hoenhayn (1992b, 1993).
[10]See also, Klepper (1996).

establish the ergodicity of a rational expectations Markov-perfect equilibrium process for the industry; their paper has provided the foundation for a very large literature on computation based models of industry dynamics.

2 The Model and Assumptions

2.1 General Structure

Consider a perfectly competitive industry producing a homogenous good. Time is discrete and indexed by $t = 0, 1, 2, \ldots$. There is a continuum of price-taking firms that are free to enter and exit the industry in any period. All firms are *ex ante* identical.

Besides producing output, active firms may also invest in accumulation of firm specific capital (for instance, knowledge or organizational capital). We assume that firms cannot accumulate firm specific capital before entering the industry. We also assume that firms lose their stock of firm specific capital when they exit the industry and that the scrap value of firm-specific capital is zero. Distinct from firm specific capital, we also allow for industry-wide capital that accumulates through (industry-wide) learning resulting from investment activity in the industry over time. An entering firm can reap benefits from the current stock of industry-wide capital (even though it was not active in the industry when this kind of capital was being accumulated in the past).

To formally describe the model, let

$$\Lambda = \{(i, j) : i \in \mathbb{Z}_+, j \in \mathbb{Z}_+ \cup \{\infty\}, i \leq j\}, \tag{2.1}$$

$$\Lambda^t = \{(i, j) \in \Lambda : i \leq t \leq j\}, \quad t \in \mathbb{Z}_+, \tag{2.2}$$

$$I(j) = \{i \in \mathbb{Z}_+ : (i, j) \in \Lambda\}, \tag{2.3}$$

$$J(i) = \{j \in \mathbb{Z}_+ \cup \{\infty\} : (i, j) \in \Lambda\}. \tag{2.4}$$

If a firm never exits, then we also say that it exits in period ∞. With this convention, $J(i)$ is the set of periods in which an entrant in period i can exit, and Λ is the set of possible pairs of periods of entry and exit. We define a *firm* *of cohort* (i, j) as a firm that enters in (or more precisely, at the beginning of) period $i \in \mathbb{Z}_+$ and exits in (or more precisely, at the end of) period $j \in J(i)$.

For $(i, j) \in \Lambda$, let $Z(i, j)$ be the set of periods in which a firm of cohort (i, j) is active:

$$Z(i,j) = \begin{cases} \{i, \ldots, j\} & \text{if } j \in \mathbb{Z}_+, \\ \{i, i+1, \ldots\} & \text{if } j = \infty. \end{cases}$$

Let p_t and K_t denote the price and the stock of industry-wide capital in period t. Taking a sequence $\{p_t\}_{t=0}^{\infty}$ and a sequence $\{K_t\}_{t=0}^{\infty}$ as given, a firm of cohort (i, j) solves the following intertemporal profit maximization problem:

$$\Pi^{i,j} \equiv \max_{\{q_t, x_t, y_t\}_{t \in Z(i,j)}} \sum_{t \in Z(i,j)} \delta^{t-i}[p_t q_t - C(q_t, x_t, K_t) - \phi(y_t)] \tag{2.5}$$

$$\text{s.t.} \quad x_i = 0, \tag{2.6}$$

$$\forall t \in Z(i, j-1), \quad x_{t+1} = x_t + y_t, \tag{2.7}$$

$$q_t, y_t \geq 0, \tag{2.8}$$

where q_t is the firm's output in period t, x_t is its stock of firm-specific capital in period t, y_t is its investment in period t, δ is the discount factor (common to all firms), $C(\cdot, \cdot, \cdot)$ is the production cost function, and $\phi(\cdot)$ is the investment cost function.

The following assumption is imposed on the production cost function C.

Assumption 2.1. $C : \mathbb{R}^3_+ \to \mathbb{R}_{++}$ is twice continuously differentiable,[11] and on $\mathbb{R}^3_{++}$ the derivatives of $C(q, x, K)$ satisfy

$$C_q > 0, C_{qq} > 0 \text{ and } C_x < 0. \tag{2.9}$$

$$C_{xx} \geq 0; \;\; C_K \leq 0, C_{qq}C_{xx} - C_{qx}^2 \geq 0, C_{qx} \leq 0, C_{qK} \leq 0. \tag{2.10}$$

These inequalities imply that over the range of strictly positive output, the total production cost $C(q, x, K)$ is strictly increasing and strictly convex in output q, strictly decreasing and convex in firm specific capital x, decreasing in

[11] For $E \subset \mathbb{R}^n$ with $n \in \mathbb{N}$, we say that a function $f : E \to \mathbb{R}$ is continuously differentiable if f can be extended to a continuously differentiable function on an open set containing E.

industry-wide capital K, and convex in (q, x); furthermore, the marginal cost of production $C_q(q, x, K)$ is decreasing in x and K.[12]Thus, every active firm has an upward sloping marginal cost curve in each period that may depend on its stock of firm specific capital as well as the stock of industry-wide capital in that period; accumulation of both forms of capital shifts the marginal cost curve downwards, at least weakly. While we allow for the case where marginal cost is independent of one or both forms of capital, average cost is strictly decreasing in the stock of firm specific capital. Note that the fixed cost of production is given by $C(0, x, K)$; under our assumptions, it is decreasing in x in general, but in the special case where marginal cost is independent of x, our assumption implies that the fixed cost is strictly decreasing in x.Total production cost is decreasing in the stock of industry-wide capital, but we allow for the case where production cost is independent of industry-wide capital i.e., there are no externalities. Given any level of output, the reduction in production cost resulting from accumulation of firm specific capital diminishes as the stock of such capital increases i.e., we have diminishing returns to investment in firm specific capital. However, as far as the effect of industry-wide capital on production cost is concerned, we do not impose any convex structure; also,we do not impose any assumption regarding complementarity or substitutability of firm specific and industry-wide capital.

The next assumption specifies restrictions on the cost of investment ϕ:

Assumption 2.2. $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is twice continuously differentiable, strictly convex, and strictly increasing with $\phi(0) = 0$.

Indeed the strict convexity of ϕ in combination with Assumption 2.1 implies that the total intertemporal cost for a firm over its lifetime in the industry (including both production and investment costs) is strictly convex in its vector of outputs and investments. Therefore, for any $(i, j) \in \Lambda$, all firms of cohort (i, j) behave identically under Assumptions 2.1 and 2.2. Let $n^{i,j} \geq 0$ be the

[12]In this paper, "increasing" means "nondecreasing," and "decreasing" means "nonincreasing." Likewise, "positive" means "nonnegative," and "negative" means "nonpositive."

size of cohort (i,j). We define

$$\Lambda_+ = \{(i,j) \in \Lambda : n^{i,j} > 0\}.$$

We define Λ_+^t, $J(i)_+$, and $I(j)_+$ similarly by requiring $(i,j) \in \Lambda_+$ in (2.2)-(2.4). Let

$$n_t = \sum_{(i,j)\in\Lambda_+^t} n^{i,j}, \quad t \in \mathbb{Z}_+. \tag{2.11}$$

The total output of the industry is given by

$$Q_t = \sum_{(i,j)\in\Lambda_+^t} n^{i,j} q_t^{i,j}, \quad t \in \mathbb{Z}_+. \tag{2.12}$$

We assume that the demand curve for the good, denoted $D(p)$, is invariant overtime and satisfies:

Assumption 2.3. (i) $D : \mathbb{R}_{++} \to \mathbb{R}_+$ is continuously differentiable. (ii) $D'(p) < 0$ for all $p > 0$ such that $D(p) > 0$. (iii) $\int_0^\infty D^{-1}(Q)dQ < \infty$.

The first two parts of the above assumption are fairly standard restrictions on the demand function; note that we allow for a "choke" price. Part (iii) of Assumption 2.3 ensures that the total surplus generated in the industry is bounded above.

The market clearing condition is

$$D(p_t) = Q_t, \quad t \in \mathbb{Z}_+. \tag{2.13}$$

Each firm's investment adds to the stock of industry-wide capital, which accumulates according to

$$K_0 = 0, \tag{2.14}$$
$$K_{t+1} = K_t + \sum_{(i,j)\in\Lambda_+^t} n^{i,j} y_t^{i,j}, \quad t \in \mathbb{Z}_+. \tag{2.15}$$

We are now ready to define the concept of industry equilibrium.

Definition 2.1. An *industry equilibrium* consists of

(i) sequences $\{p_t\}_{t\in\mathbb{Z}_+}$, $\{n_t\}_{t\in\mathbb{Z}_+}$, $\{Q_t\}_{t\in\mathbb{Z}_+}$, and $\{K_t\}_{t\in\mathbb{Z}_+}$ in $\mathbb{R}_+$,

(ii) a set $\{n^{i,j}\}_{(i,j)\in\Lambda}$ with $n^{i,j} \geq 0$ for all $(i,j) \in \Lambda$,

(iii) a set $\{\{q_t^{i,j}, x_t^{i,j}, y_t^{i,j}\}_{t\in Z(i,j)}\}_{(i,j)\in\Lambda_+}$ with $q_t^{i,j}, x_t^{i,j}, y_t^{i,j} \geq 0$ for all $t \in Z(i,j)$ and $(i,j) \in \Lambda_+$, and

(iv) a set $\{\Pi^{i,j}\}_{(i,j)\in\Lambda}$ with $\Pi^{i,j} \in \mathbb{R}$ for all $(i,j) \in \Lambda$, such that

(a) (2.5) and (2.11)–(2.15) hold,

(b) $\Pi^{i,j} \leq 0$ for all $(i,j) \in \Lambda$, and $\Pi^{i,j} = 0$ for all $(i,j) \in \Lambda_+$,

(c) $\{q_t^{i,j}, x_t^{i,j}, y_t^{i,j}\}_{t\in Z(i,j)}$ solves (2.5)-(2.8) for each $(i,j) \in \Lambda_+$.

Among other things, condition (a) of the definition of industry equilibrium requires that the market clears every period, and that given the sequence of prices and industry-wide capital, all firms maximize their intertemporal net profit over the length of time for which they are active in the industry. Conditions (a) and (c) also imply that the sequence of industry output and industry-wide capital is identical to that resulting from the optimal investment and output behavior of active firms on the equilibrium path. Condition (b) of the definition ensures that no firm has an incentive to deviate from its entry and exit decision; all firms that enter the industry earn zero (intertemporal net) profit over their length of stay in the industry, and no firm can earn strictly positive profit by altering their entry, exit, investment or output decisions.

To ensure the existence of an industry equilibrium, we impose a few additional assumptions. For this purpose, we define

$$A(q,x,K) = \frac{C(q,x,K)}{q}, \quad q > 0, x, K \geq 0, \tag{2.16}$$

$$\underline{A}(x,K) = \inf_{q>0} A(q,x,K), \quad x, K \geq 0, \tag{2.17}$$

$$\underline{p} = \inf_{x,K\geq 0} \underline{A}(x,K), \tag{2.18}$$

$$\overline{p} = \underline{A}(0,0) < \infty. \tag{2.19}$$

$\underline{A}(x,K)$ is the minimum average production cost of a firm when its current stock of firm specific capital is x and that of industry-wide capital is K. Note that under our assumptions, the minimum average cost is strictly decreasing in x and decreasing in K. $\overline{p} = \underline{A}(0,0)$ is the minimum average cost of a firm that enters in the initial period. If the price exceeds $\overline{p}$ in any period, then a firm can enter the industry for one period and make strictly positive profit violating part (b) of the definition of industry equilibrium. Thus, $\overline{p} = \underline{A}(0,0)$ is an upper bound on equilibrium prices for all periods. $\underline{p}$, as defined above, is a lower bound on the average production cost of any firm in any time period. We will show that $\underline{p}$ is actually a lower bound on equilibrium prices for all periods.

Assumption 2.4. (i) $\lim_{q \downarrow 0} C_q(q,0,0) < \underline{p}$. (ii) There exists a strictly increasing continuous function $m : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$m(q) \leq \inf_{x,K \geq 0} C_q(q,x,K), \quad \overline{p} < \lim_{q \uparrow \infty} m(q).$$

Part (i) of Assumption 2.4 implies that the marginal cost at zero output for a firm is always lower than $\underline{p}$ (the lower bound on equilibrium prices); this ensures that the optimal outputs of firms are uniformly bounded away from zero over all periods. Satisfying this restriction implicitly requires that *the fixed cost of production is bounded away from zero over all levels of* (x,K). Thus, firms have upward sloping marginal cost curve and U-shaped average cost curve (that can shift downwards with increase in firm-specific and industry-wide capital). For an entrant in period 0, not only is the marginal cost at zero output lower than its current minimum average cost, it is is in fact required to be lower than the lowest possible minimum average cost that can be attained through capital accumulation. This assumption is clearly satisfied if the marginal cost at zero output is always zero. If the marginal cost at zero output is strictly positive, the assumption can be satisfied as long as the fixed cost is always somewhat large. Note that Assumption 2.4(i) also implies that

$$\underline{p} > 0$$

which ensures that equilibrium prices are bounded away from zero.

Part (ii) of Assumption 2.4 places an upper bound on the extent of dynamic scale economies. In particular, it requires that even though the marginal cost of production may fall with capital accumulation, all possible marginal cost curves are bounded below by an upward sloping curve; cost reduction does not allow the marginal cost curve to become flat in the limit (constant marginal cost and positive fixed cost would lead to a natural monopoly). Further, no matter how much capital is accumulated, the marginal cost of production exceeds $\overline{p}$, the upper bound on prices, if output exceeds a certain level. This allows us to derive a uniform upper bound on the output produced by any firm in any time period.

Assumption 2.5. $D(\overline{p}) > 0$.

Assumption 2.5 ensures that the market is active at the upper bound $\overline{p}$ of equilibrium prices.

For $p, x, K \geq 0$, we define

$$\pi(p, x, K) = \max_{q \geq 0}[pq - C(q, x, K)]. \tag{2.20}$$

Let $s(p, x, K)$ denote the unique solution to the maximization problem in (2.20). This is the optimal current output produced by an active firm whose stock of firm-specific capital is x, when the current price is p and the current stock of industry-wide capital is K; thus, $s(., x, K)$ is the current individual supply curve of such a firm. It is characterized by the following first order condition provided that $s(p, x, K) > 0$:

$$p = C_q(s(p, x, K), x, K).$$

Note from Assumption 2.4 that $s(p, x, K)$ is strictly positive if $p \geq \underline{p}$. Provided that $s(p, x, K) > 0$, the implicit function theorem implies that

$$s_p(p, x, K) = 1/C_{qq} > 0 \tag{2.21}$$

where C_{qq} is evaluated at $(s(p, x, K), x, K)$, and the inequality holds by Assumption 2.1. Let

$$\overline{s}(p) = \sup_{x,K \geq 0} s(p, x, K), \tag{2.22}$$

$$\overline{s}_p = \sup_{p \in [\underline{p}, \overline{p}], x, K \geq 0} s_p(p, x, K). \tag{2.23}$$

From Assumption 2.4, the marginal cost curves are bounded below by the upward sloping function $m(q)$ and therefore, $\overline{s}(p) \leq m^{-1}(p)$ for all $p \in [\underline{p}, \overline{p}]$. We define

$$\underline{q} = s(\underline{p}, 0, 0), \quad \overline{q} = \overline{s}(\overline{p}).$$

As mentioned earlier, $\overline{p}$ and $\underline{p}$ can be shown to be uniform bounds on equilibrium prices for all periods. Therefore, $\underline{q}$ and $\overline{q}$ are the uniform lower and upper bounds on the output of an individual firm in any period. By Assumption 2.4, we have $0 < \underline{q} < \overline{q} < \infty$.

We also assume that no matter how large the stock of capital, at any given price, the *slope* of the individual supply curve is bounded above:

Assumption 2.6. $\overline{s}_p < \infty$.

Finally, we make an assumption to ensure that independent of the period of entry, every entering firm that stays in the market for more than one period finds it optimal to make strictly positive investment in the first period of its active life. In particular, for a firm that has never invested, the marginal cost of investment is lower than the (discounted) marginal return next period (in terms of reduced production cost).

Assumption 2.7. For any $K \geq 0$, we have

$$\phi'(0) < -\delta C_x(\underline{q}, 0, K). \tag{2.24}$$

This assumption certainly holds if $\phi'(0) = 0$. Let

$$\underline{n} = \frac{D(\overline{p})}{\overline{q}}, \quad \overline{n} = \frac{D(\underline{p})}{\underline{q}}.$$

We define $\overline{y} > 0$ by

$$\phi(\overline{y}) = \frac{\delta(\overline{p} - \underline{p})\overline{q}}{1 - \delta}.$$

We also define

$$\underline{y}_0 = \min\{y \geq 0 : \exists K \in [0, \overline{n}\,\overline{y}], \phi'(y) = -\delta C_x(\underline{q}, y, K)\} > 0,$$

where the strict inequality holds by Assumption 2.7. We will show that (in Proposition 2.1) in any industry equilibrium, $\underline{n}$ is a lower bound on the measure of initial entrants that never exit the industry as well as the measure of active firms in any period, $\overline{n}$ is an upper bound on the measure of active firms in any period, $\overline{y}$ is an upper bound on the investment made by any firm in any period and $\underline{y}_0$ is a lower bound on the investment made by any firm in the initial period.

2.2 Existence and Bounds

We are now ready to state the result on existence of industry equilibrium and the exogenous bounds satisfied by the equilibrium values of endogenous variables such as prices, output, investment and the volume of active firms:

Proposition 2.1. *(i) There exists an industry equilibrium. (ii) Let*

$$\begin{aligned}\{\{p_t\}_{t\in\mathbb{Z}_+}, \{n_t\}_{t\in\mathbb{Z}_+}, \{Q_t\}_{t\in\mathbb{Z}_+}, \{K_t\}_{t\in\mathbb{Z}_+},& \\ \{n^{i,j}\}_{(i,j)\in\Lambda}, \{\{q_t^{i,j}, x_t^{i,j}, y_t^{i,j}\}_{t\in Z(i,j)}\}_{(i,j)\in\Lambda_+}, \{\Pi^{i,j}\}_{(i,j)\in\Lambda}\}& \quad (2.25)\end{aligned}$$

be an industry equilibrium. Then we have

$$\forall t \in \mathbb{Z}_+, \qquad n_t \in [\underline{n}, \overline{n}], \quad p_t \in [\underline{p}, \overline{p}], \quad Q_t \in [D(\overline{p}), D(\underline{p})], \qquad (2.26)$$

$$\forall t \in \mathbb{Z}_+, \forall (i,j) \in \Lambda_+^t, \qquad q_t^{i,j} \in [\underline{q}, \overline{q}], \quad y_t^{i,j} \leq \overline{y}. \qquad (2.27)$$

Furthermore, $, n^{0,\infty} \geq \underline{n}$, $y_0^{0,j} \geq \underline{y}_0$ *for any* $j \in J(0)_+$ *with* $j > 0$, $y_i^{i,j} > 0$ *for any* $(i,j) \in \Lambda_+$ *with* $i < j$, *and* $K_1 \leq \overline{n}\,\overline{y}$.

In the rest of the paper, we take an industry equilibrium (2.25) as given.

3 Aggregate Dynamics

We define the industry supply curve $S_t(p)$ in period t as follows:

$$S_t(p) = \sum_{(i,j)\in\Lambda_+^t} n^{i,j} s(p, x_t^{i,j}, K_t), \quad p \in \mathbb{R}_+, t \in \mathbb{Z}_+. \tag{3.1}$$

The individual supply curve of an active firm can only shift to the right over time (through firm specific and industry-wide capital accumulation). If no exit occurs at the end of the current period, the industry supply curve can only shift to the right next period (either because of entry or because of capital accumulation or both) so that the price next period cannot be higher than the current price (given that demand curve is stationary). On the other hand, if firms do exit at the end of the current period, every exiting firm must make non-negative profit in the current period (or else, they would be strictly better off exiting earlier) and therefore, the price next period cannot be higher than the current price (or else, the exiting firm could earn strictly positive profit by staying one more period in the industry without any additional investment). Thus, despite the turnover of firms on the industry equilibrium path, prices are decreasing over time and industry output expands over time. Note that declining prices and expanding industry output are among the strong empirical regularities of product life cycles.

Proposition 3.1. *Prices are decreasing over time and industry output is increasing over time i.e., for any $t \in \mathbb{Z}_+$, we have $p_{t+1} \le p_t$ and $Q_{t+1} \ge Q_t$.*

It is immediate from Proposition 3.1 that the industry supply curve never shifts to the left.

Since the sequences $\{p_t\}$, $\{K_t\}$, and $\{x_t^{i,\infty}\}$ with $(i,\infty) \in \Lambda_+$ are all monotone, their limits are well defined (some of them may be infinite) and denoted as follows:

$$p_\infty = \lim_{t\uparrow\infty} p_t, \quad K_\infty = \lim_{t\uparrow\infty} K_t \in (0,\infty], \quad x_\infty^{i,\infty} = \lim_{t\uparrow\infty} x_t^{i,\infty}.$$

For $t \in \mathbb{Z}_+$, let $n^{t,\oplus}$ be the sum of the sizes of all cohorts that enter in period

t i.e., the total mass of firms that enter in period t.Also, let $n^{\oplus,t}$ be the sum of the sizes of all cohorts that exit in period t i.e., the total mass of firms that exit in period t:

$$n^{t,\oplus} = \sum_{j \in J(t)_+} n^{t,j}, \quad n^{\oplus,t} = \sum_{i \in I(t)_+} n^{i,t}.$$

It can be shown that if some firms exit in a particular time period, new firms cannot enter the market in the next period; the argument is that if a firm entering next period can break even through some profile of actions, any firm that exits in the current period can make strictly positive intertemporal profit by staying on in the industry and modifying its action profile suitably. In other words, if entry occurs next period, no firm exits in the current period.

Lemma 3.1. *Shake-out or exit of firms is never followed by entry in the next period, and entry is never preceded by shake-out in the preceding time period. In particular for any $t \in \mathbb{Z}_+$, if $n^{\oplus,t} > 0$, then $n^{t+1,\oplus} = 0$; equivalently, if $n^{t+1,\oplus} > 0$, then $n^{\oplus,t} = 0$.*

Since entry is never preceded by exit in the previous period, and individual supply curves of firms can only shift to the right or remain unchanged, it follows that if entry occurs, the *industry* supply curve must necessarily shift to the right leading to a strict decline in price.

Proposition 3.2. *If entry occurs next period, then price must strictly decline between the current and the next period. In particular, if $n^{t+1,\oplus} > 0$ for some $t \in \mathbb{Z}_+$, then $p_{t+1} < p_t$.*

Further, since firms make strictly positive investment immediately after they enter the industry, entry is followed by expansion of industry-wide capital. Interestingly enough, it can shown that if firms exit the industry in a particular period, at least some of the incumbent firms that do not exit must make strictly positive investment in that period so that industry-wide capital expands; if none of the incumbent firms invest, industry supply curve would shift to the left leading to a strict increase in price that would contradict Proposition 3.1. Thus:

Proposition 3.3. *Both entry and exit are accompanied by a spurt of investment activity in the industry and a (strict) increase in the stock of industry-wide capital. In particular, for* $t \in \mathbb{Z}_+$*, if* $n^{t,\oplus} > 0$ *or* $n^{\oplus,t} > 0$*, then* $K_{t+1} > K_t$*.*

Our model is designed to bring out the role of investment in technological change as a key source of volatility in industry structure. Proposition 3.3 reflects this by relating turnover of firms to concurrent investment activity within the industry.

4 Age, Size and Capital Structure of Firms

Fix any sequence of prices $\{p_t\}$ and anticipated industry-wide capital stocks $\{K_t\}$. Given these prices and industry-wide capital stocks, the (optimal) continuation value of any firm in any period (that takes into account its option to exit at any point of time) depends only on its current level of firm-specific capital; let $v_t(x)$ denote the continuation value of a firm that is active in the industry at the beginning of period t with firm specific capital stock x. Let

$$g_t(x) = \max_{x' \geq x} \{\delta v_{t+1}(x') - \phi(x' - x)\}, \tag{4.1}$$

$$G_t(x) = \operatorname*{argmax}_{x' \geq x} \{\delta v_{t+1}(x') - \phi(x' - x)\}. \tag{4.2}$$

Here, $g_t(x)$ is the value of a firm that does not exit in the current period t and chooses its current investment in firm specific capital; this firm's optimal choice of current investment is given by $G_t(x)$. In the maximization problem on the right hand side of (4.1), the interaction between x, the current stock of firm specific capital, and x', the next period's stock of firm specific capital, depends on the curvature of ϕ, the cost of investment. As the feasible set of possible choice of x' is "increasing" in x, one can use the strict convexity of ϕ and standard techniques of monotone comparative dynamics, to show that:

Lemma 4.1. *Let* $x, z \geq 0$ *with* $x < z$*. Suppose that* $g_t(x) \geq 0$*. Let* $x' \in G_t(x)$ *and* $z' \in G_t(z)$*. Then* $x' < z'$*.*

The lemma indicates that for a firm with non-negative continuation value, a strict increase in the current stock of firm specific capital must lead to a strict increase in the next period's optimal capital stock. The intuition as follows: a strict increase in the current capital stock implies that one can reach any target level of next period's capital stock by making strictly smaller investment and therefore, at strictly lower current marginal cost of investment. The latter, in turn, makes it optimal to reach a strictly higher level of capital stock next period. This also implies that if a firm j currently holds strictly higher capital stock than some other currently active firm i, and the latter has non-negative continuation value (which implies that firm j has strictly positive continuation value), then the capital stock next period must also be necessarily strictly higher for firm i. Thus, the cross section ordering of capital stocks across active firms in any period is preserved over time (until exit).

If a firm exits the industry in a particular period, its continuation value is non-positive. Any firm that is active in the industry in that period and does not exit cannot hold lower capital stock (or else, its continuation value would be negative). Thus, at the point of exit, all incumbents that do not exit hold at least as much capital stock as largest exiting firm; this is independent of the age structure of these firms.

Lemma 4.2. *In any period where exit occurs, firms that remain active in the industry have at least as large firm-specific capital stock, size and profit as the firms that exit. In particular, let $(i, j), (i', j') \in \Lambda_+$ with $i \leq j' < j$. Then $x_{j'}^{i',j'} \leq x_{j'}^{i,j}$.*

Using these results, one can establish the following key proposition about the dynamic cohort structure:

Proposition 4.1. *As long as there is some overlap between the time intervals over which they are active in the industry, a younger firm never exits later than an older firm and therefore, exiting firms are never older than incumbents that do not exit. Formally, let $(i, j), (i', j') \in \Lambda_+$. Then one of the following must hold: (a) $i \leq i' \leq j' \leq j$; (b) $i' \leq i \leq j \leq j'$;(c) $j < i'$; or (d) $j' < i$.*

Furthermore, in case (a), we have

$$\forall t \in Z(i', j'), \quad x_t^{i',j'} \leq x_t^{i,j}, q_t^{i',j'} \leq q_t^{i,j}.$$

The proposition states that as long as there is some overlap between the time intervals over which two firms are active in the industry, then any firm that enters later must exit at or before the earliest time period in which a firm that entered earlier exits the industry. In other words, it is not possible that two firms are simultaneously active in the industry in some period, and the younger firm exits later than the older firm. Thus, when exit occurs in the industry, exiting firms cannot be older and may often be younger than incumbent firms. Further, a firm that enters earlier holds at least much firm-specific capital and produces at least as much output as a firm that enters later (in every period for which they are both active in the industry). Thus, if two firms are active in the industry at the same point of time, not only does the younger firm exit no later than the older firm, it is also weakly dominated by the older firm in size (output), firm specific capital stock and profit (because of higher production cost) in every period of its remaining stay in the industry. This prediction of our model neatly matches empirical regularities regarding age and size structure of firms over the product life cycle: younger firms are smaller and less likely to survive, exiting firms are younger and smaller than incumbents.

Our analysis indicates that if we observe smaller firms earning lower profit and exiting earlier than larger and older firms that stay on and make positive profit, and even if we find that exit is accompanied by price decline, we should not necessarily take this to be indicative of predation or anti-competitive behavior; these feature can arise naturally as consequences of competition and cost reducing investment by firms that have little or no market power.

Finally, we show that among the firms that never exit the industry, the investment made by a younger firm in each period is higher than the investment made an older firm:

Lemma 4.3. *Among firms that stay in the industry forever, a younger firm* *never invests less than an older firm. In particular, let $i, i' \in \mathbb{Z}_+$ with $i < i'$ be such that $(i, \infty), (i', \infty) \in \Lambda_+$. Then for all $t \in Z(i', \infty)$, we have $y_t^{i,\infty} \leq y_t^{i',\infty}$.*

5 Convergence of Industry Structure

One of the important empirical regularities established in the product life cycle literature is that despite prolonged volatility in industry structure and high turnover of firms for considerable length of time, industries appear to stabilize eventually with markets being dominated by relatively small number of large (and relatively old) firms. In this section, we characterize the long run convergence properties of industry structure implied by our theoretical model that qualitatively matches many of these empirical findings.

Define

$$\overline{d} = \max_{p \in [\underline{p}, \overline{p}]} |D'(p)|.$$

We have seen that the volume of active firms in the industry in any period, and therefore the volume of exiting and entering firms in any time period, are bounded above (uniformly over all time periods). We now establish an explicit upper bound on the total measure of firms that enter the industry over all time periods; this is then also a bound on the total measure of firms that exit the industry over all periods. The discussion in Section 3 indicates that entry in any time period is associated with expansion of industry supply curve and a decline in the market clearing price; as prices lie in a bounded interval, using other bounds on endogenous variables derived earlier, we can derive a bound on total entry over all periods.

Lemma 5.1. *We have*

$$\sum_{j \in \mathbb{Z}_+} n^{\oplus, j} \leq \sum_{i \in \mathbb{Z}_+} n^{i, \oplus} \leq \overline{n} + (\overline{d} + \overline{n}\, \overline{s}_p)(\overline{p} - \underline{p})/\underline{q}.$$

As the total sum of entry and exit over all periods is finite, the total volume of entry or exit occurring after any date t converges to zero as $t \to \infty$. In particular, there is no turnover of firms in the long run limit of the industry.

As a consequence, the volume of firms that are active in the industry converges over time i.e., the industry structure is convergent. This is consistent with empirical regularities about entry and exit petering out, and the industry structure stabilizing towards the end of the product life cycle. The next proposition sums up these results:

Proposition 5.1. *Industry structure converges in long run. In particular, the volume of active firms in the industry converges over time, while the volumes of entry and exit occurring in the industry converge to zero. More specifically ,*

$$\lim_{t\uparrow\infty} \sum_{i\in Z(t,\infty)} n^{i,\oplus} = 0, \lim_{t\uparrow\infty} \sum_{j\in Z(t,\infty)} n^{\oplus,j} = 0, \tag{5.1}$$

$$\lim_{t\uparrow\infty} n^{t,\oplus} = 0, \qquad \lim_{t\uparrow\infty} n^{\oplus,t} = 0, \tag{5.2}$$

$$n_\infty = \lim_{t\uparrow\infty} n_t = \sum_{i\in Z(0,\infty)} n^{i,\infty}. \tag{5.3}$$

6 Entry Dynamics

In this section, we discuss the nature of entry dynamics generated along the equilibrium path of the competitive industry. In particular, we will focus on the possibility of delayed entry, and of entry being dispersed over time as a result of industry-wide capital formation and the positive externality it creates for all firms including new entrants. To this end, we begin with a negative result about delayed entry that holds when there are no externalities i.e., production cost is independent of industry-wide capital.

Proposition 6.1. *Suppose that production cost does not depend on the stock of industry-wide capital (no externalities) i.e., $C(q,x,K)$ does not depend on K. Then entry occurs only in period 0.*

The above proposition indicates that if industry-wide capital has no effect on the production cost structure (only firm specific capital affects production cost), then no firm chooses to enter late. This extends the result established in Petrakis and Roy (1999) in the finite horizon case. The intuition behind the result is straightforward: since prices decline over time and industry-wide capital is irrelevant, a firm entering in the initial period faces a "better" sequence of prices over its active life in the industry compared to that faced by a firm entering later.

Empirical evidence indicates that entry is a robust phenomenon characterizing the life cycle of an industry; while the volume of entry is likely to peak early, significant amount of entry continues into the mature phase of the industry (though eventually it is reduced to a trickle and effectively vanishes). An important contribution of this paper is to show the possibility of delayed entry on the equilibrium path even though all firms are ex ante identical, have identical future technological possibilities and have equal opportunity of entering the industry at the initial date. Indeed, if such delayed entry does not occur, all active firms would be of identical age so that many of the results characterizing age, size and capital structure of firms in Section 4 would be vacuous.

Proposition 6.1 indicates that a necessary condition for delayed entry (or for entry to be dispersed over time) is the presence of externalities such as that due to the effect of industry-wide capital; if the latter grows over time and reduces production cost, a firm entering later may face a lower price but also enjoys the benefit of higher industry-wide capital stock at each age of its active life compared to a firm that entered earlier (in the time period at which it had the same age). It is therefore conceivable that the incentive to enter earlier due to declining prices is offset by rising industry-wide capital stocks so that late entrants may break even just as early entrants do. We want to show that such an equilibrium exists under a robust set of conditions and, in particular, provide verifiable conditions on demand and technology (exogenous industry characteristics) under which entry is necessarily dispersed over time (even though, as we have established in the previous section, entry vanishes in the long run).

Economic intuition suggests that if the effect of capital accumulation on the marginal cost curve of individual firms is relatively strong so that the supply curve and the production scale of incumbent firms expand at a fast rate for each unit of firm specific and industry-wide capital accumulation, prices are likely to fall sharply and the incentive for late entry is likely to be small (the price effect is likely to dominate the effect of expanding industry-wide capital). The same is true if the demand curve is relatively steep so that any shift in the industry supply curve leads to a sharp fall in the market clearing price. This also indicates that an equilibrium with delayed entry would be more likely if industry wide capital formation has a strong effect on the total production cost of firms (i.e., lowers the average cost curve significantly) but at the same time, capital accumulation does not shift the marginal cost curve or expand the individual supply curve at a fast rate, and if the demand curve is relatively flat or elastic.

To develop the economic intuition behind delayed entry of firms consider the following condition:

Condition 6.1. There exist functions $V : \mathbb{R}_+ \to \mathbb{R}_+$ and $F : \mathbb{R}_+^2 \to \mathbb{R}_+$ such that

$$\forall q, x, K \geq 0, \quad C(q, x, K) = V(q) + F(x, K).$$

where $F(x, K)$ is strictly decreasing in K, and V,F satisfy all restrictions implied by the assumptions on $C(q, x, K)$ in Section 2.

Under Condition 6.1, capital formation does not affect the marginal cost curve i.e., the individual supply function $s(p, x, K)$ does not depend on (x, K), is identical across all firms and remains unchanged over time; capital accumulation reduces only the fixed cost of production, shifting the average cost curve downwards along a given marginal cost curve.

First, note that if Condition 6.1 holds, then exit of firms would lead to an immediate leftward shift of the industry supply curve and therefore, an increase in the market price which (as we know from Proposition 3.1) violates a basic property of the equilibrium path. Therefore, no exit occurs on the equilibrium path. Second, as no exit occurs, if there is no entry of new firms after period zero,the industry supply curve remains unchanged over time so that prices are constant over time. However, in that case, a firm entering later benefits from higher industry-wide capital (resulting from investment by the initial entrants) but faces the same price at each age of its active life compared to a firm that enters in the initial period; as the latter breaks even in equilibrium, the former earns strictly positive net profit by entering later thus violating the definition of industry equilibrium. Therefore, some firms must enter later than the initial date. In fact, this argument can be extended to show that under Condition 6.1, delayed entry must occur not just once, but in fact *infinitely often*; for otherwise, prices would be constant after the last period in which entry occurs while industry wide capital expands immediately after that period (due to investment by the last round of entrants) creating strict incentive for a firm to enter after that date.

Proposition 6.2. *Under Condition 6.1, new firms enter the market infinitely often and no firm exits the market in finite time.*

Note that the argument for continued entry of firms over time under Condition 6.1 is independent of the nature of the demand curve. If capital formation reduces the variable or marginal cost of firms in addition to reducing the fixed cost so that their supply curves shift to the right, the industry supply curve can shift to the right even if no entry occurs and prices may fall over time. Delayed entry may still occur if the rate at which prices fall is more than compensated by expansion of industry-wide capital and this depends on a several factors including the slope of the demand curve. In the rest of this section we will outline a couple of general conditions for entry of new firms over time that allow for reduction of both fixed and marginal cost through capital accumulation.

At this stage, we introduce an additional assumption on the investment cost function ϕ:

Assumption 6.1. $\phi'(0) = 0$ i.e., the marginal cost of investment is zero at zero investment.

Assumption 6.1 is maintained in the rest of this paper.

Assumption 6.1 is a technical restriction that helps us develop conditions for delayed entry as well as exit of firms that are easy to understand and readily interpreted. It ensures that with the exception of the period in which a firm exits, a firm make strictly positive investment in every period of its active life in the industry. It can be used to show the firm specific capital stock of firms that do not exit in finite time tends to infinity in the long run; the stock of industry-wide capital too diverges to infinity. These and some other important implications of Assumption 6.1 are outlined in the following lemma:

Lemma 6.1. *$K_\infty = \infty$, $x_\infty^{i,\infty} = \infty$ for any $i \in I(\infty)_+$,*

$$p_\infty \leq \widetilde{p} = \lim_{K \to \infty} \underline{A}(0, K)$$

and further,

$$\frac{D(\widetilde{p})}{\overline{s}(\widetilde{p})} \leq n_\infty \leq \frac{D(\underline{p})}{\overline{s}(\underline{p})}.$$

Thus, under Assumption 6.1, we can not only predict the long run capital structure of the industry very precisely, but also obtain clear bounds on the volume of firms that are active in the long run.

Recall the definition of bounds $\overline{y}, \underline{y}_0$ on equilibrium investment by a firm (described in Section 2). The next proposition outlines a verifiable condition for delayed entry:

Proposition 6.3. *Let $\overline{t} \in Z(0, \infty)$ be such that $\delta^{\overline{t}} \phi(\overline{y}) \leq \phi(\underline{y}_0)$. Suppose that*

$$\forall K \in [0, \overline{t}\overline{n}\,\overline{y}], \quad \frac{D(\underline{A}(K/\underline{n}, K))}{s(\underline{A}(K/\underline{n}, K), 0, K)} < \frac{D(\widetilde{p})}{\overline{s}(\widetilde{p})}. \tag{6.1}$$

Then, some firms enter the industry after the initial period.

As firms discount the future, initial entrants must earn positive profits within a certain length of time in order to be able to cover the investment cost incurred initially. In Proposition 6.3, $\overline{t}$ is an upper bound on the length of time before which initial entrants must make positive profit at least in some period. In such a time period, the market price must exceed the minimum average cost of initial entrants and this yields an upper bound on the number of active firms in the industry in that period; this bound is given by the left hand side of the inequality in (6.1). Note that this bound depends on the stock of industry-wide capital. From Lemma 6.1, the right hand side of the inequality in (6.1) is a lower bound on the number of active firms in the long run. Delayed entry must occur if the former is smaller than the latter. The inequality only needs to hold for the range of industry-wide capital stocks that can be accumulated in the equilibrium path by period $\bar{t}$ (the upper bound of this range is given by $\bar{t}\bar{n}\,\overline{y}$).

Initial entrants that never exit must earn positive profit infinitely often. Further, the stock of industry wide capital becomes infinitely large in the long run. Using these two facts, the argument in the previous paragraph can be extended to show that if the inequality in (6.1) holds as long as industry-wide capital stock is large enough, then new firms must continue to enter the market infinitely often. The next proposition establishes this formally:

Proposition 6.4. *Suppose there exists $\underline{K} \geq 0$ such that that*

$$\forall K \geq \underline{K}, \quad \frac{D(\underline{A}(K/\underline{n}, K))}{s(\underline{A}(K/\underline{n}, K), 0, K)} < \frac{D(\widetilde{p})}{\overline{s}(\widetilde{p})}. \tag{6.2}$$

Then, entry never ceases and new firms enter the market infinitely often.

Proposition 6.4 provides a verifiable condition on demand and cost structure under which entry of new firms continues forever. In line with the economic intuition outlined earlier, condition (6.1) in Proposition 6.3 and condition (6.2) in Proposition 6.4 are more likely to hold if the demand curve is steeper and if the rate of expansion of the supply curve as a result of capital accumulation is smaller.

As the left hand side of the inequality in (6.2) converges to $\frac{D(\underline{p})}{\overline{s}(\underline{p})}$ as $K \to \infty$,(6.2) can hold only if

$$\widetilde{p} = \lim_{K\to\infty} \underline{A}(0, K) = \underline{p} = \lim_{x\to\infty, K\to\infty} \underline{A}(x, K)$$

i.e., sufficiently large industry-wide capital stock is a substitute for firm specific capital.

Condition (6.2) is easy to verify for specific functional forms. The appendix contains the details of an example where capital accumulation reduces both fixed and marginal cost of production and conditions 6.1 and (6.2) hold so that entry occurs infinitely often. In this example, the cost function is given by

$$C(q,x,K) = F\left[1+\frac{1}{1+x+K}\right] + q^2\left[1+\frac{1}{1+K}\right], F>0$$

and the demand is given by:

$$D(p) = p^{-\theta}, \theta > 1, p \in [\underline{p}, \overline{p}].$$

There are no further restrictions.

Our discussion in Section 3 indicates that firms that enter later are typically smaller, hold less of firm-specific capital, have higher production cost and are likely to exit earlier. Therefore, our results about entry occurring with delay is indicative of endogenous emergence of cross-section heterogeneity in size, capital and profits of *ex ante* identical firms. In particular, high dispersion of entry over time (as indicated by Propositions **??** and 6.4) is also likely to be associated with considerable heterogeneity among firms along the equilibrium path.

7 Shake-out

In this section, we discuss the possibility of exit or shake-out of firms on the industry equilibrium path. Established empirical regularities relating to product life cycle and results obtained from other studies of manufacturing industries indicate fairly high rates of shake-out of firms (particularly during the phase where the industry attains maturity). Further, significant exit continues over time even though the volume of exit is small in the long run. Our purpose is to understand the nature of exogenous industry conditions such as demand and technology under which the market provides incentives for some firms to exit from time to time, even when they have the same initial technological possibilities and opportunities as the firms that continue to be active forever.

Proposition 6.2 indicates that if capital accumulation reduces only the fixed cost of production, then no firm exits the industry. In other words, a necessary condition for shake out of firms is that capital accumulation should reduce the marginal cost of production and expand the individual supply curve. In what follows, we provide sufficient conditions for shake-out of firms on the equilibrium path.

As in the previous section, we will assume that Assumption 6.1 holds i.e., in each period, a firm's marginal cost of making a sufficiently small investment is negligible.

The next proposition outlines a verifiable condition under which some firms must exit in finite time.

Proposition 7.1. *Suppose that*

$$\frac{D(\underline{p})}{\overline{s}(\underline{p})} < \frac{D(\overline{p})}{s(\overline{p}, 0, 0)}. \tag{7.1}$$

Then some initial entrants exit in finite time.

Using Lemma 6.1, the left hand side of the inequality (7.1) indicates the number of active firms in the long run limit of the industry. The right hand side of inequality (7.1) is a lower bound on the number of firms in the initial period (as prices are bounded above by $\overline{p}$ and the supply curve of all firms in the initial period is $s(p, 0, 0)$). If the former bound is smaller than the latter, some firms must exit the industry. In fact. the difference between these two bounds provides a lower bound on the total volume of exit that must occur over time. Condition (7.1) is more likely to hold and therefore, shake-out is more likely to occur if demand is relatively price inelastic and if individual supply expands sharply i.e., the marginal cost curve shifts sufficiently with capital accumulation. In a finite horizon version of this model where there are no externalities, Petrakis and Roy (1999) provide a sufficient condition for the occurrence of shake-out; condition 7.1 in Proposition 7.1 is a somewhat different condition that is easy to interpret.

The appendix contains the details of an example where the cost function is given by:

$$C(q,x,K) = F + q^2\left(1 + \frac{1}{1+x+K}\right), F > 0,$$

and the demand function satisfies:

$$D(p) = p^{-\theta}, 0 < \theta < 1, p \in [\underline{p}, \overline{p}].$$

For this demand and cost functions, it can shown that:

$$\frac{D(\overline{p})}{s(\overline{p},0,0)} = 2^{\frac{1}{2}-\frac{3\theta}{2}}F^{-\frac{\theta+1}{2}} > 2^{-\theta}F^{-\frac{1+\theta}{2}} = \frac{D(\underline{p})}{\overline{s}(\underline{p})}$$

so that condition (7.1) holds. Further. a lower bound of the rate of shake-out (i.e., the total measure of exiting firms as a ratio of the measure of initial entrants) is given by

$$1 - 2^{\frac{1}{2}(\theta-1)}$$

which is decreasing in demand elasticity θ and is roughly 30% for θ close to 0.

A stronger version of condition 7.1 ensures that exit never ceases and the industry equilibrium path always exhibits shake-out of firms from time to time:

Proposition 7.2. *Suppose that there exists $\underline{K} \geq 0$ such that*

$$\forall K \geq \underline{K}, \quad \frac{D(\underline{p})}{\overline{s}(\underline{p})} < \frac{D(\underline{A}(0,K))}{s(\underline{A}(0,K), K/\underline{n}, K)}. \tag{7.2}$$

Then, firms exit the industry infinitely often i.e., shake-out continues to occur for an indefinitely long time horizon.

The left hand side of the inequality in condition (7.2) is identical to that in condition (7.1) in Proposition 7.1 and gives us an upper bound of the number of active firms in the long run limit of the industry. On the equilibrium path, the market price cannot exceed the minimum average cost for a new firm entering in that period. This allows us to derive a lower bound on the number of active firms in any period and this bound is given by the right hand side of the inequality in condition (7.2). It depends on the stock of industry-wide capital in that period. Under assumption 6.1, the industry-wide capital stock grows infinitely large in the long run. If the inequality in (7.2) holds, then after a certain length of time the number of active firms in the industry will always exceed that in the long run limit, so that firms must exit the industry infinitely often.

Note that the right hand side of the inequality in (6.2) converges to $\frac{D(\widetilde{p})}{\bar{s}(\widetilde{p})}$ as $K \to \infty$,so that like our condition (6.2) for entry to occur infinitely often, our condition for continued shake-out can be satisfied only if

$$\widetilde{p} = \lim_{K\to\infty} \underline{A}(0, K) = \underline{p} = \lim_{x\to\infty, K\to\infty} \underline{A}(x, K),$$

i.e., industry-wide capital accumulation can eventually compensate for lack of firm specific capital.

For the cost and demand functions specified above, condition (7.2) holds if:

$$\left(1 + \frac{1}{1+K}\right)^{\frac{1+\theta}{2}} < \left\{1 + \frac{1}{1 + K(2^{\theta}(\sqrt{2F})^{(1+\theta)} + 1)}\right\}, \forall K \geq 0,$$

which holds if θ and F are small (details are contained in the appendix).

The results derived in Section 3 indicate that at the point of exit, exiting firms are no larger and typically smaller i.e., hold less of firm specific capital, have higher production cost and earn lower profit than incumbents that do not exit (independent of the age structure of the firms). It follows that similar to delayed entry, shake-out of firms is associated with endogenous emergence of heterogeneity in size, capital and profits of *ex ante* identical firms. In particular, high dispersion of exit time across initial entrants as indicated by Proposition **??** is also likely to be associated with high cross-section heterogeneity among firms.

8 Conclusion

Exogenous differences among firms (in capabilities and opportunities) and stochastic shocks are not the only sources of observed heterogeneity among firms and changes in industry structure over time. A competitive market can create incentives for endogenous differences in the behavior of firms that is then reflected in the patterns of entry, exit, heterogeneity and changes in composition of firms in the industry. We develop a theoretical framework to understand the role of market forces in explaining some of the well known empirical regularities on industry dynamics. We analyze a fairly general *deterministic* model of a dynamic competitive industry with free entry and exit where all firms *ex ante* identical and engage in cost reducing investment that also generates industry-wide learning. In equilibrium, prices decline over time, industry output expands and identical firms make very different entry and exit decisions. As a result, firms may face very different profiles of prices and industry-wide knowledge over their active lifetimes in the industry. This generates cross-section differences in investment, output, firm-specific capital, size and flow profits of active firms. We develop verifiable sufficient conditions on the demand and cost structure under which entry of new firms is dispersed over time, and may continue to occur for an indefinite length of time. Continued entry is more likely if industry-wide learning has a strong effect on cost reduction, if capital formation reduces the fixed cost of production faster than the marginal cost, and if market demand is sufficiently elastic. We also characterize verifiable conditions under which firms choose to exit the industry over time; in fact, shake-out of firms may continue for an indefinite length of time. Shake-out is more likely if capital formation reduces the marginal cost of production more sharply than the fixed cost, and if market demand is sufficiently inelastic. At any point of time, an older firm holds higher stock of firm specific capital and therefore, has lower cost and higher current profit than a younger firm. Firms that exit are younger and smaller than incumbents (last in first out) and smaller firms are more likely to exit earlier. Despite the volatility that may be exhibited along the equilibrium path, industry structure converges in the long run.

Our model is simple in many ways and is not tailored to explain other well known empirical regularities on industry dynamics. The general structure makes it difficult to obtain sharp predictions on the time pattern of entry and exit to match the precise properties of observed product life cycles. However, our results are not inconsistent with any of the major findings about product life cycles. For instance, our condition for delayed entry in Proposition 6.3 can be simultaneously satisfied with the condition for exit to occur infinitely often in Proposition 7.2 and that would be perfectly consistent with a scenario of entry in the early stages of an industry followed by shakeout and a stable configuration as the industry matures. Sharper predictions require computation of dynamic equilibrium through numerical analysis; approximation of the infinite horizon equilibrium outcome by that in a finite horizon set-up is an important step in this direction. Some of our current research is related to these issues.

References

[1] Abbring, J. and J. Campbell, "Last-in first-out industry dynamics" , *Econometrica* 78 (2010), 1491-1527.

[2] Bergin, J. and D. Bernhardt, "Industry dynamics with stochastic demand," *RAND Journal of Economics* 39 (2008), 41-68.

[3] Besanko, D. and U. Doraszelski, "Capacity dynamics and endogenous asymmetries in firm size," *RAND Journal of Economics* 35 (2004), 23-49.

[4] Bester, H. and E. Petrakis, "Wages and productivity growth in a competitive industry", *Journal of Economic Theory* 109 (2003), 52-69.

[5] Davis, S. and J. Haltiwanger, "Gross job-creation, gross job destruction and employment reallocation", *Quarterly Journal of Economics* 107 (1992), 819-863.

[6] Dixit, A., "Entry and exit decisions under uncertainty", *Journal of Political Economy* 97 (1989), 620-38.

[7] Dunne, T., M. Roberts and L. Samuelson, "The growth and failure of U.S. manufacturing plants", *Quarterly Journal of Economics* 104 (1989), 671-698.

[8] Ericson, R. and A. Pakes, "Markov perfect industry dynamics: a framework for empirical work," *Review of Economic Studies* 62 (1995), 53-82.

[9] Flaherty, M.T., "Industry structure and cost reducing investment," *Econometrica* 48 (1980), 1187 - 1209.

[10] Gort, M. and S. Klepper, "Time paths in the diffusion of product innovations," *Economic Journal* 92 (1982), 630-53.

[11] Hopenhayn, H., "Entry, exit and firm dynamics in long run equilibrium," *Econometrica* 60 (1992a), 1127-150.

[12] Hopenhayn, H., "Exit, selection and the value of firms", *Journal of Economic Dynamics and Control* 16 (1992b), 621-53.

[13] Hopenhayn, H., "The shake-out", Working Paper # 33, Dept. of economics, University of Pompeu Fabra, 1993.

[14] Jovanovic, B., "Selection and the evolution of industry", *Econometrica* 50 (1982), 649-70.

[15] Jovanovic, B. and S. Lach, "Entry, exit and diffusion with learning by doing," *American Economic Review* 79 (1989), 690-99.

[16] Jovanovic, B. and G. MacDonald, "The life cycle of a competitive industry,"*Journal of Political Economy* 102 (1994), 322-47.

[17] Klepper, S., "Entry, exit, growth, and innovation over the product life cycle," *American Economic Review* 86 (1996), 562-83.

[18] Klepper, S. and E. Graddy, "The evolution of new industries and the determinants of market structure," *Rand Journal of Economics* 21 (1990), 27 -44.

[19] Lambson, V.E., "Industry evolution with sunk costs and uncertain market conditions," *International Journal of Industrial Organization* 9 (1991), 171-96.

[20] Lippman, S.A. and R.P. Rummelt, "Uncertain imitability: an analysis of interfirm differences in efficiency and competition", *Bell Journal of Economics* 13 (1982), 418-38

[21] Lucas, R. and E.C. Prescott, "Investment under uncertainty," *Econometrica* 39 (1971), 659-81.

[22] Maggi, G., "Endogenous leadership in a new market," *RAND Journal of Economics* 27 (1996), 641-59.

[23] Pakes, A. and R. Ericson, " Empirical implications of alternative models of firm dynamics", *Journal of Economic Theory* 79 (1998), 1-46.

[24] Petrakis, E. and S. Roy, "Cost-reducing investment, competition and industry dynamics", *International Economic Review* 40 (1999), 381 -401.

[25] Petrakis, E., E. Rasmusen and S. Roy, "The learning curve in a competitive industry," *Rand Journal of Economics* 28 (1997), 248-68.

[26] Reynolds, S. and B. Wilson, "Betrand-Edgeworth competition, demand uncertainty, and asymmetric outcomes," *Journal of Economic Theory* 92 (2000), 122-41.

[27] Stokey, N., "The dynamics of industry-wide learning, " in *Equilibrium Analysis: Essays in Honor of Kenneth J. Arrow, Volume II* (W. Heller, R. Starr and D. Starrett, Eds.), Cambridge University Press, 1986, 81-104.

Examples to be moved to the appendix:

1. (Entry) Let the cost function be given by:

$$C(q,x,K) = F\left[1+\frac{1}{1+x+K}\right] + q^2\left[1+\frac{1}{1+K}\right], F>0$$

It is easy to check that C satisfies all assumptions imposed in Section 2. The individual supply function is given by

$$s(p,x,K) = \frac{p}{2}\left(\frac{1+K}{2+K}\right)$$

so that

$$\overline{s}(p) = \sup_{x,K} s(p,x,K) = \frac{p}{2}$$

Further,

$$A(q,x,K) = \frac{F}{q}\left[1+\frac{1}{1+x+K}\right] + q\left[\frac{2+K}{1+K}\right],$$

so that

$$\underline{A}(x,K) = \min_q A(q,x,K) = 2\sqrt{F\left[1+\frac{1}{1+x+K}\right]\left[\frac{2+K}{1+K}\right]}$$

and

$$\underline{p} = \inf_{x,K} \underline{A}(x,K) = 2\sqrt{F}, \overline{p} = \underline{A}(0,0) = 4\sqrt{F}$$

Note that for this cost function,

$$\widetilde{p} = \lim_{K\to\infty} \underline{A}(0,K) = \underline{p} = 2\sqrt{F}$$

Also,

$$\overline{s}(\widetilde{p}) = \sqrt{F}.$$

The demand function satisfies:

$$D(p) = p^{-\theta}, \theta > 1, p \in [\underline{p}, \overline{p}].$$

It is easy to check that $\underline{q} = s(\underline{p}, 0, 0) = (\underline{p}/2)\left(\frac{1}{2}\right) = \frac{\sqrt{F}}{2}$, $\overline{q} = \overline{s}(\overline{p}) = \frac{\overline{p}}{2} = 2\sqrt{F}$ and therefore,

$$\underline{n} = \frac{D(\overline{p})}{\overline{q}} = \frac{(4\sqrt{F})^{-\theta}}{2\sqrt{F}} = 2^{-(1+2\theta)} F^{-(\frac{1+\theta}{2})}.$$

It follows that

$$\underline{A}(\frac{K}{\underline{n}}, K) = 2\sqrt{F\left[1 + \frac{1}{1 + \frac{K}{\underline{n}} + K}\right]\left[\frac{2+K}{1+K}\right]}$$

and

$$\frac{D(\underline{A}(\frac{K}{\underline{n}}, K))}{D(\widetilde{p})} = \left[\frac{\underline{A}(\frac{K}{\underline{n}}, K)}{\widetilde{p}}\right]^{-\theta} = \left[\left(1 + \frac{1}{1 + \frac{K}{\underline{n}} + K}\right)\left(\frac{2+K}{1+K}\right)\right]^{-\frac{\theta}{2}}.$$

Further,

$$\begin{aligned} s(\underline{A}(\frac{K}{\underline{n}}, K), 0, K) &= \frac{\underline{A}(\frac{K}{\underline{n}}, K)}{2}\left(\frac{1+K}{2+K}\right) \\ &= \left[F\left(1 + \frac{1}{1 + \frac{K}{\underline{n}} + K}\right)\left(\frac{1+K}{2+K}\right)\right]^{\frac{1}{2}} \end{aligned}$$

so that

$$\frac{s(\underline{A}(\frac{K}{\underline{n}}, K), 0, K)}{\overline{s}(\widetilde{p})} = \left[\left(1 + \frac{1}{1 + \frac{K}{\underline{n}} + K}\right)\left(\frac{1+K}{2+K}\right)\right]^{\frac{1}{2}}$$

Inequality (6.2) is satisfied if for all $K \geq 0$:

$$\left[\left(1+\frac{1}{1+\frac{K}{\underline{n}}+K}\right)\left(\frac{2+K}{1+K}\right)\right]^{-\frac{\theta}{2}} < \left[\left(1+\frac{1}{1+\frac{K}{\underline{n}}+K}\right)\left(\frac{1+K}{2+K}\right)\right]^{\frac{1}{2}}$$

which can be written as

$$\left[1+\frac{1}{1+\frac{K}{\underline{n}}+K}\right]^{\theta+1}\left[1+\frac{1}{1+K}\right]^{\theta-1} > 1$$

and this holds as $\theta > 1$.

2. (Exit) The cost function is given by:

$$C(q,x,K) = F + q^2[1+\frac{1}{1+x+K}], F > 0.$$

Note that C satisfies condition 6.1(i). We do not specify any functional form for ϕ other than requiring it to satisfy condition 6.1(i). It is easy to check that:

$$s(p,x,K) = \frac{p}{2}\left(\frac{1+x+K}{2+x+K}\right), \overline{s}(p) = \sup_{x,K} s(p,x,K) = \frac{p}{2}$$

$$A(q,x,K) = \frac{F}{q} + q[\frac{2+x+K}{1+x+K}]$$

$$\underline{A}(x,K) = \min_q A(q,x,K) = 2\sqrt{F\left(\frac{2+x+K}{1+x+K}\right)}$$

$$\underline{p} = \inf_{x,K} \underline{A}(x,K) = 2\sqrt{F}, \overline{p} = \underline{A}(0,0) = 2\sqrt{2F}$$

$$\underline{A}(0,K) = 2\sqrt{F\left(\frac{2+K}{1+K}\right)}$$

$$\underline{q} = s(\underline{p},0,0) = \frac{\sqrt{F}}{2}, \overline{q} = \sup_{x,K} s(\overline{p},x,K) = \frac{\overline{p}}{2} = \sqrt{2F}$$

$$s(\overline{p},0,0) = \frac{\overline{p}}{4} = \sqrt{\frac{F}{2}}$$

The demand function satisfies:

$$D(p) = p^{-\theta}, \theta > 0, p \in [\underline{p}, \overline{p}]$$

$$\underline{n} = \frac{D(\overline{p})}{\overline{q}} = \frac{(\overline{p})^{-\theta}}{\overline{q}} = \frac{(2\sqrt{2F})^{-\theta}}{\sqrt{2F}} = 2^{-\theta}(\sqrt{2F})^{-(1+\theta)}$$

$$\overline{s}(\underline{p}) = (\underline{p}/2) = \sqrt{F}$$

Thus,

$$\frac{D(\overline{p})}{s(\overline{p}, 0, 0)} = \frac{(2\sqrt{2F})^{-\theta}}{\sqrt{\frac{F}{2}}} = 2^{\frac{1}{2}-\frac{3\theta}{2}} F^{-\frac{\theta+1}{2}} > 2^{-\theta} F^{-\frac{1+\theta}{2}} = \frac{D(\underline{p})}{\overline{s}(\underline{p})}, \text{ as } 0 < \theta < 1,$$

so that condition (7.1) of Proposition 7.1 holds. Further,

$$s\left(\underline{A}(0, K), \frac{K}{\underline{n}}, K\right) = \left(\sqrt{F\left(\frac{2+K}{1+K}\right)}\right)\left\{\frac{1 + \frac{K}{\underline{n}} + K}{2 + \frac{K}{\underline{n}} + K}\right\}$$

$$\begin{aligned} \frac{\overline{s}(\underline{p})}{s\left(\underline{A}(0, K), \frac{K}{\underline{n}}, K\right)} &= \left(\sqrt{\frac{1+K}{2+K}}\right)\left\{1 + \frac{1}{1 + \frac{K}{\underline{n}} + K}\right\} \\ &= \left(\sqrt{\frac{1+K}{2+K}}\right)\left\{1 + \frac{1}{1 + K(2^{\theta}(\sqrt{2F})^{(1+\theta)} + 1)}\right\}. \end{aligned}$$

$$\frac{D(\underline{p})}{D(\underline{A}(0, K))} = \left(\frac{\underline{A}(0, K}{\underline{p}}\right)^{\theta} = \left(\frac{2+K}{1+K}\right)^{\frac{\theta}{2}}$$

Condition (7.2) in Proposition 7.2 requires:

$$\frac{D(\underline{p})}{D(\underline{A}(0, K))} < \frac{\overline{s}(\underline{p})}{s\left(\underline{A}(0, K), \frac{K}{\underline{n}}, K\right)}, \forall K \geq 0$$

i.e.,

$$\left(1+\frac{1}{1+K}\right)^{\frac{1+\theta}{2}} < \left\{1+\frac{1}{1+K(2^{\theta}(\sqrt{2F})^{(1+\theta)}+1)}\right\}, \forall K \geq 0$$

The inequality holds for θ, F small.