Markov-Perfect Industry Dynamics: A Framework for Empirical Work
Author(s): Richard Ericson and Ariel Pakes
Source: *The Review of Economic Studies*, Vol. 62, No. 1 (Jan., 1995), pp. 53-82
Published by: Oxford University Press
Stable URL: http://www.jstor.org/stable/2297841

Review of Economic Studies (1995) **62**, 53-82 0034-6527/95/00030053$02.00

Markov-Perfect Industry Dynamics: A Framework for Empirical Work

RICHARD ERICSON
Columbia University

and

ARIEL PAKES
Yale University and NBER

First version received March 1990; *final version accepted July* 1994 (*Eds.*)

This paper provides a model of firm and industry dynamics that allows for entry, exit and firm-specific uncertainty generating variability in the fortunes of firms. It focuses on the impact of uncertainty arising from investment in research and exploration-type processes. It analyses the behaviour of individual firms exploring profit opportunities in an evolving market place and derives optimal policies, including exit, in this environment. Then it adds an entry process and aggregates the optimal behaviour of all firms, including potential entrants, into a rational expectations, Markov-perfect industry equilibrium, and proves ergodicity of the equilibrium process. Numerical examples are used to illustrate the more detailed characteristics of the stochastic process generating industry structures that result from this equilibrium.

I. INTRODUCTION

A salient feature of firm-level data is the great variability in the fate of similar firms over time. Manifestations of this variability include simultaneous entry and exit in an industry, simultaneous firm-level job creation and destruction, and variability in growth rates, found in the analysis of firm and establishment level panel data sets. These indications of differences in outcome paths among firms persist even after one controls for the firm's entry date, location, and industry, and therefore for time, location, and industry specific differences in economic environments. Moreover they tend to be associated with a remarkable degree of heterogeneity among firms in the same industry in both levels and movements over time in the variables that we typically want to analyse (industry output shares, investment, productivity, etc.).[1] We provide a model of industry behaviour which, because it incorporates idiosyncratic or firm-specific sources of uncertainty, can generate the variability in the fortunes of firms observed in these data.

There is a policy, as well as a descriptive, need for such a model. In a world where firms differ, policy and environmental changes are likely to have different impacts, and lead to different responses, in different firms. Since these responses are frequently nonlinear

1. Partly due to increased data availability, there has been a resurgence in the analysis of firm level panels over the last decade; see Evans (1987*a*, *b*), Dunne *et al.* (1988), Pakes-Ericson (1990), Davis-Haltwinger (1992), and the literature cited in those articles. These articles also contain references to the extensive empirical literature on the nature, extent, and implications of the variation in performance paths among firms.

functions of the changing variable (entry and exit reflecting an extreme nonlinearity), any analysis of their effects, even if only an analysis of their aggregate impacts (say on industry supply or productivity), requires both the underlying distribution of firms by the source of response heterogeneity, and the (equilibrium) response of that distribution to the given policy or environmental change.[2] Of course, policy issues are often more directly related to the heterogeneity in the distribution of responses per se, as in, for example, the analysis of the effects of a policy or an environmental change on job turnover, on market structure, or on default probabilities. In these situations the whole focus is on characteristics of the distribution of the response heterogeneity, and hence the need for a structural model that allows for idiosyncratic uncertainty becomes even stronger.

Models of industry dynamics allowing for firm heterogeneity and/or idiosyncratic shocks have begun to appear in the literature, beginning with the models of intrinsic, initially unknown and unchanging "types" which are slowly revealed through economic activity, by Jovanovic (1982) and Lippmann–Rumelt (1982).[3] Another class of models emphasizes the sunk cost nature of initial investments whose relative profitabilities change over time in response to outcomes of some exogenous process; see Dixit (1989) and Lambson (1992). Both classes of models deal with large, perfectly competitive industries. In addition there are game-theoretic duopoly models exploring possible characteristics of alternative dynamic equilibria; see Maskin–Tirole (1988), Rosenthal–Spady (1989), Budd *et al.* (1993), and Cabral–Riordan (1992), and in the technology "race" literature see Vickers (1986), Beath *et al.* (1987), and Dutta *et al.* (1993). Finally, there is a growing body of literature emphasizing the need for such models in empirical work (e.g. Thomas (1990); Olley–Pakes (1991)), and beginning to implement them in policy analysis (Berry–Pakes (1993), Hopenhayn–Rogerson (1993)).

The purpose of this paper is to provide a model which allows for heterogeneity and idiosyncratic shocks, and which is general enough to serve as a framework for empirical work. In Section II the model of an industry and its equilibrium are presented. The industry model is based upon a stochastic model of the entry and growth of a firm through the active exploration of its economic environment. The firm invests to enhances its capability to earn profits in an environment characterized by substantial competitive pressure from both within and outside the industry. The stochastic outcome of a firm's investment, the success of other firms in the industry, and competitive pressure from outside the industry (both in the market and through entry) determine the "success" of the firm, i.e. its profitability and value. If success is limited, a deterioration in the profitability of the firm can lead to a situation in which it is optimal to abandon the whole undertaking. This endogenizes exit behaviour, and provides a natural way of accounting for selection in the process of determining the evolution of the industry.

We close the model by showing the existence of a Markov-perfect Nash equilibrium in the investment, entry, and exit decisions of each firm. Firms maximize their present discounted value given expectations about the evolution of their competition. At equilibrium those expectations are fully consistent with the process generated by the optimal

2. See Geweke (1985), and Pakes–McGuire (forthcoming) for related discussions and numerical examples. The importance of explicitly accounting for heterogeneity in response patterns when analysing the aggregate impacts of changes comes out clearly in the recent empirical work that uses disaggregate data, e.g. Thomas (1990) or Olley–Pakes (1991).

3. These models were a natural dynamic extension of the static models of industry equilibrium with heterogeneity among agents; see Lucas (1978), Kihlstrom–Laffont (1978), and the summary in Brock–Evans (1985). Hopenhayn (1992) provides a hybrid model in which perfectly competitive firms are subject to exogenous productivity shocks, but do not engage in Bayesian learning as they know the distribution of those shocks. See further discussion in Section III below.

decisions of all firms within or entering the industry. Thus we show the existence of a rational expectations equilibrium with a finite number of heterogeneous agents subject to idiosyncratic shocks. These results are presented and discussed in Section III, where we further characterize that equilibrium as an ergodic stochastic process, and note the implications of that result for interpreting the observed dynamics of industry equilibria. Section III concludes by relating our results to those of a number of other dynamic industry models.

Our model is general enough to encompass a wide variety of specific models of competition. The answers to many questions of interest, however, depend on details of the functional forms which determine the fine structure of any application. Hence we have, elsewhere, developed a computational algorithm which computes and characterizes the equilibria associated with the different functional forms that can be fed into our model (see Pakes–McGuire, (forthcoming)). Section IV uses this algorithm to compute and analyse a particular example: a Cournot–Nash, homogeneous product, version of our model in which firms are differentiated with respect to their efficiency of production. These efficiencies evolve with the outcomes both of a research and exploration process and of an aggregate process which shifts the costs of factors of production to the industry. We then provide a brief comparison of these results to the results from a differentiated product version of our model used as the example in Pakes–McGuire (forthcoming). Section V concludes with a summary and discussion of potential extensions, focusing primarily on steps that would allow us to make more intensive use of the model in interpreting data. Finally, proofs of the more important of our results are gathered in a technical appendix.

II. AN INDUSTRY MODEL

A. Overview

The active force in our model is an entrepreneur/firm exploring a speculative idea, a perceived profit opportunity in some industry.[4] To learn the value of the opportunity, a firm must invest to enter the industry and then in developing and, possibily, in exploiting it. Investment to enter is a sunk cost, perhaps partially recoverable if there is some scrap value realizable on exit. The quantity of investment, together with parameters describing the evolution of the market and the competition, determine the distribution of outcomes from the exploratory activities of an active firm in each period.

Favourable outcomes from its own investment activity tend to move the firm towards "better" states; states in which its idea can be embodied in a good or service likely to be marketed more profitably. Favourable outcomes of direct competitors, or advances in alternatives to the industry's products, tend to move the firm toward less profitable states. Indeed, a firm whose investment activity generated a string of relatively unsuccessful outcomes may well find itself in a situation in which its idea is not perceived to be worth developing further, so that the enterprise is best liquidated and its salvageable resources committed to an alternate use. Hence the model generates exit as a natural outcome of an evolutionary process.

The opportunity (technology) provided by this industry is open to all, so that the only distinction among firms is their achieved state of "success" (index of efficiency), $\omega \in \mathbb{Z}$, in exploiting it.[5] The state, ω, of each firm within the industry is measured relative to an alternative which reflects the strength of competition outside the industry. It changes

4. We do not explore the nature of the firm, but take it to be a unitary maximizing agent.
5. $\mathbb{Z}$ is the ordered set of all integers; $\mathbb{Z}_+$ is the set of non-negative integers.

over time as a result of autonomous factors which shift the demand and/or cost parameters of all firms producing in the industry. Therefore higher ω indicates that the firm is in a stronger (more profitable) position relative to both other firms in, and competition from outside, the industry. There is a set of states, $\Omega^e \subset \Omega$, at which new firms may enter the industry after making a sufficient investment. We denote the industry structure at any point of time by $s = \{s_\omega\}_{\omega \in \mathbb{Z}} \in \mathbb{Z}_+^\infty$; s provides the number of firms at each possible ω state. The state couple, (ω, s), determines the entire distribution of the firm's current and future profits, and hence the firm's viability as an enterprise.

Precisely how the state (ω, s) affects payoffs to the firm depends on the nature of competition and thus the associated type of within-period market equilibrium. That will determine the "strength" of the competition faced, and hence which states are "better" for the firm. For our theoretical results we do not need to be precise about the nature of the equilibrium in the spot market for current output. We will require that it generates a complete preorder, $\succeq$, over s, which unambiguously defines the strength of the competition. That is, we assume that, no matter what the firm's ω, current profits are (weakly) decreasing in s in the sense of $\succeq$. Also, conditional on any s, current profits are (weakly) increasing in (the natural order of) ω. Hence many models of the interaction among firms (including price taking "competitive" models) abide by our assumptions.[6]

The state (ω, s) changes as a result of the outcomes of the firm's own investment and development efforts, the outcomes of the efforts of other firms operating in the same market, and with changes in the overall market environment, i.e. in demand, input costs, and science and technology, in which it is embedded. The firm's own level of investment, denoted by $x_t \in \mathbb{R}_+$, is chosen to maximize the expected present discounted value of profits as a function of all information available at t. We assume this information to include the history of all past states and of the firm's own past investment decisions, i.e. $\{(\omega_{t'}, s_{t'}), x_{t'}\}_{t'<t}$; the current state, (ω_t, s_t); and the probability laws governing the evolution of that state over time, including the law governing the impact of the firm's own investment on that evolutionary pattern. Of course, those probabilities are determined, in part, by the investment decisions of all firms in or entering the industry. We assume that the firm does not directly observe the investments of its competitors, and hence cannot make decisions based on them.[7]

The dynamics of the model are thus generated by the stochastic outcomes of the firm's investments and the outcome of an exogenous process reflecting improvements made by competition outside the industry. Outcomes of this exogenous stochastic process generate a correlated non-positive stochastic shift in all the firm's ω's, reflecting, for example, increases in the quality of goods outside the industry that vie for the consumer's dollar (and/or increases in factor costs). It is, therefore, a source of continuous dynamic competitive pressure that forces all firms in the industry to struggle to maintain profits and survive. It can also induce a positive correlation in the profits of different firms in the same industry, a phenomenon we often observe in data.[8] Also, it is assumed that the outcomes of the exogenous process generating increases in the knowledge stock outside the industry are embodied in the new generations of potential entrants to this industry; otherwise entry would eventually die out, and with it the industry. That is, the new

6. Two illustrative examples involving Nash equilibrium in prices and quantities are developed and numerically analysed in Pakes-McGuire (forthcoming) and Section IV.

7. Hence this is a game with imperfectly observable actions or in the terminology of Maskin-Tirole (1993) a "game of moral hazard with simple type spaces."

8. Without the exogenous process, any outcome which leads to an increase in profits for one firm would necessarily reduce the profits of its rivals.

generation of entrants brings with it knowledge which was not available to previous generations.

A new entrant incurs a sunk cost of entry, x^e, and then takes a full period to set up the specific fixed capital with which it enters. The precise state of entry depends on the "quality" or "efficiency" of the entering firm, i.e. on how "good" its idea or innovation is relative to the achieved standards of the industry. This we assume to be unknowable ex-ante; an idea must be tried, and time, money and effort invested, before competitiveness can be precisely known. Hence there is only a common knowledge distribution, $P(\omega^0)$, over the potential entry states, Ω^e, indicating the uncertainty of both entrants and incumbents as to the competitiveness of potential entrants.[9]

B. The assumptions

The opportunity presented to each firm by the industry is defined by model primitives, which are common knowledge to all actual and potential participants:

$$\{A(\omega, s), p(\omega'|\omega, :), q_\omega(\hat{s}'|s), [m(s), P(\omega^0), \{x^e_m\}^\infty_{m=1}], \phi, c(\omega), \beta\}_{(\omega,s)\in\Omega\times S}.$$

We describe these objects, then present the assumptions required for our general model.

The state space is $\Omega\times S\subset\mathbb{Z}\times\mathbb{Z}^\infty_+$, where S is a set of counting measures on $\mathbb{Z}$. The structure of the industry, that is s, the list which counts the number of firms in each state ω, is just such a measure. The function $A(\omega, s)$ gives the payoff or profits of a firm from its current production and sales activities. It is a reduced form, reflecting the equilibrium of the industry spot market, and its detailed characteristics can vary from example to example. $p(\omega'|\omega, x)$ is a firm's transition function: it gives the probability of shifting into state ω', conditional on being in state ω and investing amount $x\in\mathbb{R}_+$. $q_\omega(\hat{s}'|s)$ provides the firm's beliefs about the transition probabilities for the other firms in, or entering, the industry, given that it is in state ω. Here $\hat{s}\equiv s-e_\omega$, where e_ω is a vector with one in the ω-th place and zero elsewhere; $\hat{s}$ is a measure providing the location of the firm's competitors. Thus next period's industry structure will be $s'=\hat{s}'+e_{\omega'}$, where $\hat{s}'$ includes any new entrants and ω' is the new state achieved by the firm in question.

The triple $[m(s), P(\omega^0), \{x^e_m\}^\infty_{m=1}]$ characterizes the conditions of entry into the industry. The number of entrants stimulated by any structure (state of competition), s, is given by the function $m(s)$. The initial investment required to begin the process of entry is x^e_m, which may depend on the number of firms simultaneously entering. Finally, the state, ω^0, at which a new firm enters the industry is determined by the probability distribution $P(\cdot)$ with support $[\text{supp}\,(P)]\Omega^e$.

The parameter ϕ gives the opportunity cost of being in the industry; it is the amount recoverable on exit. The function $c(\omega)$ gives the unit cost of activity level x, so that investment activity costs $c(\omega)\cdot x$, and current net revenues or profits are given by:

$$R(\omega, s; x)=A(\omega, s)-c(\omega)\cdot x. \tag{1}$$

Finally, β is the common discount factor of all the agents in the model.

We use the following assumptions for our general results.

A.0 $\omega\in\Omega\subset\mathbb{Z}; s\in S\subset\mathbb{Z}^\infty_+$, with $\succeq$ a complete pre-order on S.
A.1 $\beta\in(0, 1); \phi\in\mathbb{R}$.

9. We assume Ω^e is bounded above, implicitly limiting the progress that can be made in the area/niche of this industry while remaining outside the industry. We also show that there is a lower bound below which rational entry would never occur.

A.2 $\forall \omega, c(\omega) \in [\underline{c}, \infty), \underline{c} > 0$.

A.3 $\forall s \in S$, $\lim_{\omega \to \infty} A(\omega, s) = \bar{A} < \infty$ and $\lim_{\omega \to -\infty} A(\omega, s) < (1-\beta)\phi$. $A(\cdot)$ is non-decreasing in ω for all s, and is non-increasing in s, ordered by $\succeq$, for each ω. Finally, $\forall \omega, \forall s \in \hat{S}_n(\omega), A(\omega, s) \leq (1-\beta)\phi + o(\frac{1}{n})$, where $\hat{S}_n(\omega) \equiv \{s \in S \mid \sum_{\omega' \geq \omega} s_{\omega'} \geq n\}$.

A.4 $\forall \omega \in \Omega, \forall x \geq 0, p(\cdot \mid \omega, x)$ is formed from the convolution of two distributions with finite connected support: $\pi(\cdot \mid \omega, x)$ with $\operatorname{supp}(\pi) = \{\omega' \mid \omega' = \omega + \tau, \tau = 0, \ldots, k_1\}$; $p_0 = \{p_\eta\}_{-k_2}^{0}$ with $\operatorname{supp}(p_0) \subset \{\omega' \mid \omega' = \omega + \eta, \eta = -k_2, \ldots, 0\}$. $\pi(\cdot \mid \omega, x)$ is stochastically increasing, continuous in x, $\partial\pi/\partial x\,(\omega \mid \omega, x) < 0$, $\partial\pi/\partial x\,(\omega'/\omega, x) > 0$ and concave at each $\omega' \in \{\omega+1, \ldots, \omega+n\}$, and $\pi(\omega' \mid \omega, 0) = \begin{cases} 1 & \text{if } \omega' = \omega \\ 0 & \text{otherwise} \end{cases}$.

A.5 $m(s)$ firms enter in each period, $m: S \to \mathbb{Z}_+$. Each entrant pays $x^e_m > \beta\phi$, non-decreasing in the number of entrants, m. The entry process is completed at the beginning of the succeeding period, when each entrant becomes an incumbent at some state $\omega^0 \in \Omega^e \subset \Omega$ with probability $P(\omega^0) = \sum_{\eta=-k_2}^{0} p_\eta \cdot \pi^e(\omega^0 - \eta)$. Ω^e is a compact connected set.

A.6 There exists a regular Markov transition kernel, $Q: \mathbb{Z}_+^\infty \times \mathbb{Z}_+^\infty \to [0, 1]$, i.e.:

$$\forall B \subset S, \forall s \in S, \textstyle\sum_{s' \in B} Q(s' \mid s) = \operatorname{Prob}\{s_{t+1} \in B \mid s_t = s\},$$

with range $S(s) \equiv \{s' \mid Q(s' \mid s) > 0\} \neq \emptyset$, such that the functions $q_\omega(\hat{s}' \mid s) \equiv \sum_\eta q_\omega(\hat{s}' \mid s, \eta) p_0(\eta)$ are the consistent marginal transition probabilities derived from it for $\hat{s} = s - e_\omega$. The stochastic kernels Q and q_ω have the Feller property, i.e. each maps the space of continuous functions on S, $C(S)$, into itself.

A.7 (a) There exists a constant $M < \infty$, such that, for all $s \in S$, $m(s) \leq M$.
(b) The set of potential feasible industry structures, $S \subset \mathbb{Z}_+^\infty$, is compact.

(A.3) gives the consequences of spot market competition. Whatever the structural model that lies behind $A(\omega, s)$, we require it to have the property that if we increase the number of competitors with ω's at least as large as the firm's own ω then, eventually, the firm's profits will fall to less than $(1-\beta)\phi$, the annuity value of the recoverable assets obtained by the firm when it exists. Similarly we require that no matter the competition inside the industry, there is sufficient competition from outside that a firm whose ω drops low enough will eventually find its profits to be less than $(1-\beta)\phi$.

(A.4) implies that $\omega' = \omega + \tau + \eta$, where the realization of τ is determined by the outcome of the firm's expenditures and has a distribution given by $\pi(\cdot \mid \omega, x)$, while the realization of η is determined by the outcome of the process defining the outside alternative, and has a distribution given by p_0. Consequently $p\{\omega' = z \mid \omega, x, \eta\} \equiv \pi(z - \eta \mid \omega, x)$ and $p\{\omega' = z \mid \omega, x\} \equiv \sum_{\eta'} \pi(z - \eta' \mid \omega, x) p_{\eta'}$. Similarly the distribution of both entering states (in A.5) and of the likely locations of one's competitors (in A.6) are also obtained by first conditioning on η. Note that if $x = 0$ the firm's ω cannot improve, and will, in fact, stochastically decay with negative realizations of η. The assumptions on the derivatives of $\pi(\cdot)$ are only used to insure the uniqueness of the firms choice of level of investment; only continuity and stochastical monotonicity are fundamental.

(A.5) describes the entry process, incorporating the impact of the negative drift on firms engaged in the process of entry. It is essentially a free-entry assumption.[10] It also indicates that the real sunk cost of entry is $x^e - \beta\phi$, as any entrant could recover ϕ next

10. There are many possible entry assumptions that could be inserted here without affecting the general nature of the theoretical results. We did not delve further into this both because the free entry assumption seemed the natural place to start, and because so little is known about empirically relevant alternatives.

period by immediately exiting after becoming an incumbent. The last two assumptions are auxiliary in the sense that they are used to restrict agents' perceptions, and then are shown to be natural consequences of an equilibrium given those perceptions.

C. The incumbent's decision

An incumbent firm makes decisions to maximize the expected present value of net cash flows. At any time t and state (ω_t, s_t), it must decide to continue or to exit the industry, and if it stays in operation, it must decide how much to invest. It thus solves[11]

$$W_t(\omega_t, s_t) \equiv \max\left\{ \sup_{\{x_\tau, \chi_\tau\}_{\tau=t}^{\infty}} E_t\left[\sum_{\tau=t}^{\infty} \beta^{\tau-t} R(\omega_\tau, s_\tau; x_\tau)\chi_\tau + (\chi_{\tau-1} - \chi_\tau)\phi \mid (\omega, s)\right], \phi \right\}, \quad (2)$$

where χ_τ is the continuation decision [$\chi_\tau = 1 \Rightarrow$ continue; $\chi_\tau = 0 \Rightarrow$ exit], and $x_\tau \geqq 0$ is the amount to be invested, in period τ. Clearly, $\chi_\tau = 0$ implies that, for all $\sigma \geqq \tau$, $\chi_\sigma = x_\sigma = 0$. For any given $\{x_\tau, \chi_\tau\}$, the distribution used to form the expectation in (2) can be derived from the firm's perception of the Markov transition kernel for its competitors, $\{q_\omega(\hat{s}'|s)\}$, and the controlled Markov process governing the evolution of the firm's own state, $\{p(\omega'|\omega, x)\}$.[12]

In any state, the incumbent firm compares the expected present discounted value of remaining in the industry, assuming optimal future decisions, to the opportunity cost of remaining, ϕ. If the latter is larger, it exits, foregoing $R(\omega, s; 0)$ and all potential future earnings in the industry. If not, it invests $x \geqq 0$, receives $R(\omega, s; x)$, and retains the option of further activity in the industry starting in a new state (ω', s') next period.

This formulation has an inherently stationary Markovian structure. That is, the current state, (ω_t, s_t), and the current decision, x_t and χ_t, are sufficient to completely determine its dynamics, i.e. the evolution to the next state, (ω_{t+1}, s_{t+1}). This implies that the optimal investment strategy, if it exists, can be chosen from the class of stationary Markov strategies, vastly simplifying its analysis.[13] Thus we are justified in writing $x(\omega, s)$ and $\chi(\omega, s)$; that is, both the investment and shutdown decisions are stationary functions of only the current state (ω, s).

This together with boundedness implies that if a solution exists to the entrepreneur's problem it must satisfy the Bellman equation

$$V(\omega, s) = \max\left[\sup_{x \geqq 0} \left\{ R(\omega, s; x) + \beta \cdot \sum_{\eta'} \sum_{s'} \sum_{\omega'} V(\omega', s') p(\omega'|\omega, x, \eta') q_\omega(\hat{s}'|s, \eta') p_{\eta'} \right\}, \phi \right], \quad (3)$$

as can readily be seen by substitution. In any state the optimal policy thus involves first choosing a level of investment that maximizes the expression in braces on the r.h.s. of (3). This requires selecting an investment level equalizing current marginal costs with the marginal change in the expected present value of the states that might be realized next

11. See Chapter 9 of Stokey, Lucas, and Prescott (1989) for more detail on setting up related intertemporal optimization problems.

12. This distribution can be explicitly written using the Chapman–Kolmogorov equation. See Doob (1953, p. 88).

13. This is a standard result of the literature on optimization in a Markovian environment. See, for example, Dynkin and Yushkevich (1975, p. 148), or Stokey, Lucas, Prescott (1989, Chapter 9.1). Of course it involves non-trivial behavioural assumptions about firms' responses to the industry's strategic situation; see Mashkin–Tirole (1993).

period. When the expected future value generated by optimal investment is less than or equal to the opportunity cost of the entire enterprise, ϕ, then the optimal decision is to liquidate the enterprise.

D. The entrant's decision

An entrant faces a similar optimization problem, with the added uncertainty as to where he will be, once in the industry. Entry decisions are taken at the beginning of each period, and the process of entry takes a full period (A.5); firms deciding to enter in period t become incumbents at the beginning of period $t+1$. Attempted entry is successful upon payment of the sunk cost, x_m^e, which depends on the number of firms, m, entering at t. As an incumbent at some ω^0, the new firm at $t+1$ invests (or exits) to solve (3), i.e. to generate the maximal value, $V(\omega^0, s_{t+1})$, where s_{t+1} includes all entrants from the preceding period. Any potential entrant must evaluate this expected value of optimal behaviour in the industry, labeled $V^e(s, m)$, relative to the cost of entry, x_m^e, both of which depend on the number of new firms entering in that period. Note that this is an expectation over all the states $\omega^0 \in \Omega^e$ at which the firm might enter, and is the same ex-ante for all potential entrants.

Assumptions (A.5) and (A.6) imply that

$$V^e(s, m) \equiv \beta \cdot \sum_{\eta'} \sum_{s'} \sum_{\hat{\omega}_m} \sum_{\omega^0} V(\omega^0, s' + e_{\omega^0} + \hat{\omega}_m) \cdot \pi^e(\omega^0 - \eta')$$
$$\times \prod_{j=1}^{m-1} \pi^e(\omega_j^0 - \eta') \cdot q^0(s'|s, \eta') \cdot p_{\eta'} \geqq \phi, \tag{4}$$

where $\hat{\omega}_m \equiv \sum_{j=1}^{m-1} e_{\omega_j^0}$, and $q^0(\cdot|\cdot)$ is the marginal of $Q(\cdot|\cdot)$ for incumbents only.[14] The given firm enters at ω^0 with probability $P(\omega^0)$. The other $m-1$ entrants come in each at their own ω_j^0 according to the same probability distribution, adding the vector of entrants, $\hat{\omega}_m$, to the old incumbent's new stucture s'.

If $V^e(s, m) \leqq x_m^e$ for all $m \geqq 1$, then no entry can optimally take place: the expected value of being in the industry at some ω^0 cannot justify the sunk cost of even one entrant. We assume that, in each period, ex ante identical firms decide to enter sequentially until the expected value of entry falls sufficiently to render further entry unprofitable. That occurs when $V^e(s, m+1) - x_{m+1}^e \leqq 0 < V^e(s, m) - x_m^e$; m is then the number of new firms that rationally enter. Formally, the number of entrants into any industry structure s is thus given by the function:

$$m(s) \equiv \begin{cases} 0 & \text{if } V^e(s, m) \leqq x_m^e \quad \text{for all } m \geqq 1, \\ \min \{m \in \mathbb{Z}_+ | V^e(s, m) > x_m^e,\ V^e(s, m+1) \leqq x_{m+1}^e\} & \text{otherwise.} \end{cases} \tag{5}$$

E. The equilibrium

We study the dynamic equilibrium of the industry arising from the competitive interaction of firms both within and entering the industry. All firms know the structure of the industry, s, their place in it, ω, and the likely impact of their own investment. Firms also have beliefs, $q_{\omega}(\cdot)$, about how the structure of the industry, and hence the states of its competitors, will change. The industry is said to be in dynamic equilibrium when the process generating the change in industry structure is accurately reflected in the beliefs of each of the firms

14. $q^0(\cdot)$ is given by a multinomial distribution from the $|s|$ independent transitions with probabilities $p(\cdot|\cdot, \cdot)$, ignoring the entrants induced by the structure s. See the Remark in Section E below.

entering or active in the industry. Thus the equilibrium is one of "rational expectations", where optimal decisions are based on the true distribution of future states generated by the optimal behaviour of all incumbents and potential entrants.

Formally, we define an *equilibrium* for this industry as the 6-tuple,

$$[\{V(\omega, s), x(\omega, s), \chi(\omega, s), Q(s'|s), m(s)\}_{(\omega,s)\in\Sigma}, s^0], \tag{6}$$

with $\sum\equiv\Omega\times S$ and $\Omega\equiv(0,\ldots,K)$, $K<\infty$, such that

6.a $\forall(\omega, s)\in\sum$, $V(\omega, s)$ satisfies (3):

$$V(\omega, s)=\max\,[R(\omega, s; x(\omega, s))+\beta\cdot\{\textstyle\sum_{\eta'}\sum_{\omega'}\sum_{s'} V(\omega', \hat{s}'+e_{\omega'})$$

$$\times q_{\omega}(\hat{s}'|s, \eta')p[\omega'|\omega, x(\omega, s), \eta']p_{\eta'}\}, \phi]$$

6.b $\forall(\omega, s)\in\sum$, $x(\omega, s)$ and $\chi(\omega, s)$ solve (3) and satisfy:

$$\{-c(\omega)+\beta\cdot\textstyle\sum_{\eta'}\sum_{\omega'}\hat{V}(\omega'|\omega, s, \eta')\cdot p_x(\omega'|\omega, s, \eta')\cdot p_{\eta'}\}\cdot x(\omega, s)=0,$$

$$\{V(\omega, s)-\phi\}\cdot[\chi(\omega, s)-1]=0,$$

where $\hat{V}(\omega'|\omega, s, \eta')\equiv\sum_{\hat{s}'} V(\omega', \hat{s}'+e_{\omega'})q_{\omega}(\hat{s}'|s, \eta')$;

6.c $\forall(s', s)\in S\times S$, $Q(s'|s)\equiv\sum_{\eta}p_{\eta}\cdot Q_{\eta}(s'|s)$, with

$$Q_{\eta}(s'|s)\equiv\textstyle\sum_{Y\in\mathscr{Y}(s'|s)}\prod_{j=0}^{K} m_{\eta}(y_{0j},\ldots, y_{Kj}|s_j)\cdot m_{\eta}^{e}(y_{0j},\ldots, y_{Kj}|m(s)),$$

where $Y\equiv[y_{ij}]\in\mathbb{Z}_+^{(K+1)^2}$, y_{ij} is the number of firms shifting to s_i' from s_j, $\mathscr{Y}(s'|s)\equiv\{Y\in\mathbb{Z}_+^{(K+1)^2}|\,Y\cdot e=s', e\cdot Y=(m(s), s)\}$, $m_{\eta}(y_j|s_j)$ is the multinomial probability of $y_j=(y_{0j},\ldots, y_{Kj})$ firms out of s_j going to the states $i=0,\ldots, K$, conditional on η, and $m_{\eta}^{e}(y_j|m(s))$ is the same for the $m(s)$ new entrants;[15]

6.d $\forall s\in S$, equation (5) determines the number of entrants, $m(s)$: $\forall t$, $m_t>0$ if and only if $x_1^e\leqq V^e(s_t, 1)$ [defined in (4)], where $m_t=m(s_t)\equiv\min\{m\in\mathbb{Z}_+|\,x_m^e\leqq V^e(s_t, m),\ V^e(s_t, m+1)<x_{m+1}^e\}$;

6.e There is an exogenously given initial state, $s^0\in S$. ||

Remark. The definition assumes that the number of states can be bounded above and below as proved in Proposition 1. The optimal policy, $\{x(\omega, s), \chi(\omega, s)\}$, and (A.4) together define Markov transition probabilities from each active state l, to each feasible state j, conditional on each possible value of η as $\pi(j-\eta|l, x(l, s))\equiv p_{jl}(\eta, s)$. For every s, equilibrium defines a matrix of transition probabilities for incumbents as $P(s)\equiv\sum_{\eta'}p_{\eta'}P(n', s)$ where $P(\eta, s)\equiv[p_{jl}(\eta, s)]_{j,l=0}^{K}$. These transition probabilities, together with the distribution of incumbents along the rows of this matrix and the entry rule, determine $Q_{\eta}(s'|s)$ in (6.c). To actually compute $Q_{\eta}(s'|s)$, note that the multinomial theorem implies that the s_j firms in state $\omega=j$ allocate themselves among the $K+1$ possible states, relabeled $\{0,\ldots, K\}$, with probabilities, conditional on progress (η) outside the

15. Y is a matrix summarizing one way that the vector s' might have been generated, and $\mathscr{Y}(s'|s)$ is the set of all feasible such matrices. A row i of Y shows the numbers of firms moving into state i, while a column of Y, y_j, shows the allocation of firms in state j among new period's states, $i=0,\ldots, K$. The first column shows the allocation of new entrants among the states within $\Omega^e\subset[0, K]\subset\mathbb{Z}_+$.

industry given by

$$m_\eta(y_j|s_j) \equiv \left\{\frac{s_j!}{(y_{0,j})! \cdot (y_{1,j})! \cdot \cdots \cdot (y_{K,j})!}\right\} \cdot \prod_{l=0}^{K} [p_{jl}(\eta)]^{y_{jl}}, \tag{7}$$

A similar expression gives the distribution of the $m(s)$ entrants over the states in $\Omega^e \subset \{0, \ldots, K\}$, with conditional probabilities $p_{j0}(\eta)$ given by $\pi^e(\omega^0 - \eta)$. Thus $Q_\eta(s'|s)$, as defined in (6.c), is the probability that optimal investment strategies will generate a shift in the structure from s to s' conditional on the outside competition making a positive advance of η (all incumbents and entrants drift downward by as much as η if their investment efforts fail to yield a counteracting advance). It follows that $Q(s'|s)$ is the unconditional probability that $s_{t+1} = s'$ when $s_t = s$. Finally, $q_\omega(\hat{s}'|s, \eta)$ is just the conditional marginal distribution over the competing firms:

$$q_\omega(\hat{s}'|s, \eta) = \textstyle\sum_{\omega'} Q_\eta(\hat{s}' + e_{\omega'}|s). \tag{8}$$

In equilibrium, all firms optimize with respect to a given distribution of future states (industry structures), $Q(\cdot|s)$, and their optimal decisions generate industry transitions with precisely the distribution used in their optimization (6.c). $Q(\cdot|s)$ is derived by aggregating the incumbent firms' transition probabilities, $p(\omega'|\omega, x(\omega, s))$, where $x(\omega, s)$ is the optimal investment strategy, with the distribution of the m new entrants, $\prod_{j=1}^{m} P(\omega_j^0)$, where investment, entry and exit are all optimal given the individual state and industry structure, and that state and structure evolve according to the anticipated distribution. The dependence of current market returns, $A(\omega, s)$, on structure s (A.3) insures that the spot market for current output clears.

This equilibrium is also a Nash equilibrium in investment strategies defined for all (ω, s)-nodes in the game tree. By assumption, firms take the distribution of outcomes of others' decisions as fixed, thereby choosing their exit and investment decisions independently of others in the industry. As the optimal strategies and transition probabilities are functions only of payoff-relevant states, $(\omega, s) \in \sum$, the equilibrium is a *Markov-Perfect Nash Equilibrium* in the sense of Maskin–Tirole (1988, 1993). Agents solve dynamic programming problems that are interdependent only through those variables, so their investment strategies, $x(\omega, s)$, remain optimal at every state, regardless of how that state was reached, against the optimal decisions of all other agents.[16]

At the heart of this dynamic equilibrium is a (time-homogeneous) Markov process, $(S, Q(\cdot|\cdot), s^0)$, on the space of industry structures (counting measures of firms in the industry), S, defined Q, a transition kernel determing the distribution of s_{t+1} conditional on all alternative possible values of s_t, and by s^0, the initial state (see Section III.C below). A realization of this process is a unique sequence $\{s_t\}_{t=0}^{\infty}$ where $s_0 = s^0$ and s_t is a realization from the distribution $Q(\cdot|s_{t-1})$. Associated with each such realization of this process are the sequences: $\{m_t\}$, the optimal entry process derived from (6.d); $\{\underline{\underline{\omega}}_t\}$, the highest exit states defined by $\underline{\underline{\omega}}_t = \max\{\omega|\chi(\omega, s_t) = 0\}$; and $\{f_t\}$, the number of firms that exit in period t, $f_t \equiv \sum_{\omega=0}^{\underline{\underline{\omega}}_t} s_{\omega,t}$. The notion of equilibrium guarantees that the distribution of these sequences is generated by the optimal investment strategies of both incumbents and potential entrants and that the spot market for current output always equilibrates.

All decisions within a period are understood to be taken simultaneously, based on common knowledge of the industry structure, s_t, the number of entrants that this structure

16. This is an immediate consequence of the dynamic programming formulation and the consistency of all firms' problems at equilibrium.

will call forth, $m_t = m(s_t)$, the exit states that the structure generates, $\{\omega | \omega \leqq \underline{\underline{\omega}}_t(s_t)\}$, and the distribution of future states that will arise from that structure, $Q(\cdot | s_t)$. While m_t new firms are entering, incumbents either rationally exit $[\chi(\omega, s_t) = 0]$ or invest $x(\omega, s_t) \geqq 0$ generating their transition probabilities which at equilibrium will collectively, when combined with the distribution of new entrants, precisely coincide with those given by the common knowledge distribution $Q(s'|s_t)$. This yields the new industry structure at the beginning of the next period in which again entry, exit, and investment decisions will be made. To close the model we need to show that these decisions can be consistently taken, i.e. that such a stochastic dynamic equilibrium exists.

III. RESULTS

A. Characterizing optimal agent behaviour

The primary agent in this model is an incumbent firm. The first result shows that an optimal solution exists to the decision problem (2), giving well-defined investment and exit decisions and a well-defined value to the firm (3), and then characterizes the optimal policies. Entrants are distinguished only in the initial period of their entry; thereafter they are incumbents. Here the only question that needs to be answered is how many find it profitable to sink x_m^e in order to enter the industry. Our second result shows that it is finite in any period, and indeed will be zero if competition within the industry is sufficiently strong. These results imply that the state space S is compact, as assumed (A.7.b) for some of the results characterizing incumbent behaviour. They also allow us to show the consistency of our assumptions about the industry structure transition probabilities (A.6), setting the stage for a proof of existence of equilibrium.

Proposition 1. *Consider the firm's decision problem* (2). *Under assumptions* (A.0) *through* (A.7):

(a) *There exist* (i) *a unique* $V(\omega, s)$, $V: \mathbb{Z} \times \mathbb{Z}_+^\infty \to \mathbb{R}_+$, *monotonic increasing in* ω, *uniformly bounded, and satisfying* (3); (ii) *an* $\bar{x} < \infty$ *and a unique optimal investment policy* (*function*), $x(\omega, s)$, $x: \mathbb{Z} \times \mathbb{Z}_+^\infty \to \mathbb{R}_+$, *with* $x(\omega, s) \leqq \bar{x}$; *and* (iii) *an optimal termination policy* $\chi(\omega, s)$, $\chi: \mathbb{Z} \times \mathbb{Z}_+^\infty \to \{0, 1\}$; *solving* (2) [*or* (3)] *for* $\forall (\omega, s) \in \mathbb{Z} \times \mathbb{Z}_+^\infty$.

(b) *There exist two finite boundaries in* $\mathbb{Z} \times \mathbb{Z}_+^\infty$, $\underline{\omega}(s)$ *and* $\bar{\omega}(s)$, *such that* $x(\omega, s) = 0$ *if* $(\omega, s) \in C \equiv C_l \cup C_u$, *where* $C_l \equiv \{(\omega, s) | \omega < \underline{\omega}(s)\}$ *and* $C_u \equiv \{(\omega, s) | \omega > \bar{\omega}(s)\}$, *and there exists a finite lower bound* $\underline{\underline{\omega}}(s) \in \mathbb{Z}$ *such that* $\chi(\omega, s) = 0$ *if and only if* $(\omega, s) \in \{(\omega, s) | \omega \leqq \underline{\underline{\omega}}(s)\} = L$. *Further,* $\inf_s \underline{\underline{\omega}}(s) > -\infty$, and $\sup_s \bar{\omega}(s) < \infty$.

(c) *There exists a random variable,* $T: \mathbb{Z} \times \mathbb{Z}_+^\infty \to \mathbb{Z}_+$, $T(\omega_0, s_0) = \inf \{t \geqq 0 | (\omega_0, s_0) = (\omega_0, s_0)$ *and* $(\omega_t, s_t) \in L\}$, *associating each initial state,* (ω_0, s_0), *with the first time,* t, *such that* $\chi_t \equiv \chi(\omega_t, s_t) = 0$, *where* (ω_t, s_t) *is the state achieved in period* t *under the optimal policy* $\{x(\omega, s), \chi(\omega, s)\}$. $T(\omega_0, s_0) < \infty$, a.s. *and is stochastically increasing in* ω.

Proof. See Appendix. ||

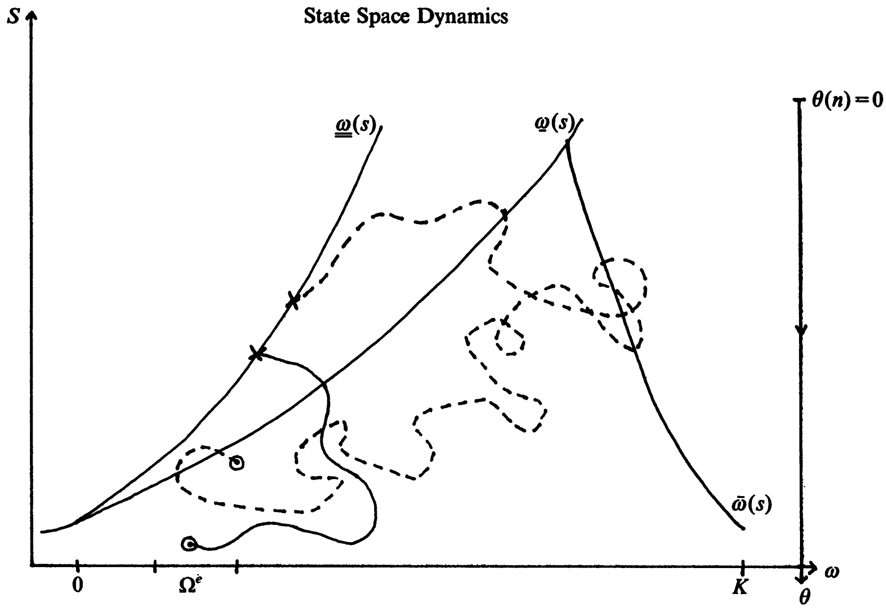
An incumbent firm in state ω facing an industry structure s has an expected present discounted value of $V(\omega, s)$. When $V(\omega, s) = \phi$, it will optimally exit the industry. This is the case at all (ω, s) with $\omega \leqq \underline{\underline{\omega}}(s)$. Hence we will never observe a firm with an efficiency less than $\underline{\underline{\omega}} = \min \{\underline{\underline{\omega}}(s) | s \in S\}$. When $V(\omega, s) > \phi$, the firm pursues an optimal investment policy, $x(\omega, s)\in[0, \bar{x}]$, earning a current cash flow of $R(\omega, s)=A(\omega, s)-c(\omega)x(\omega, s)$. Part (b) of the proposition proves the existence of boundaries $\underline{\omega}(s)$, and $\bar{\omega}(s)$, such that $x(\underline{\omega}(s)-\tau, s)=x(\bar{\omega}(s)+\tau, s)=0$, for all $\tau\geqq 1$. Since ω cannot increase in value without some investment (A.4), and the distribution of increments to ω has finite support, an immediate consequence of this optimal behaviour is that we will never find a firm at ω-states higher than $\bar{\bar{\omega}}=\max\{\bar{\omega}(s)+k_1|s\in S\}$. Thus (A.4), (A.7), and the first Proposition imply that the relevant set of states is the compact, connected interval $\{\underline{\underline{\omega}}, \ldots, \bar{\bar{\omega}}\}\subset\mathbb{Z}$; the compactness of Ω in our definition of equilibrium (6) is satisfied and, by relabelling, we can set $\Omega=\{0, 1, \ldots, K\}$. The space of admissible structures, then, is no greater than $(K+1)$-dimensional: $S\subset\mathbb{Z}_+^{K+1}\subset\mathbb{Z}_+^{\infty}$.

The results in (a) to (c) provide a fairly detailed characterization of incumbent behaviour. Part (a) guarantees that incumbent behaviour is well defined and shows that the valuation of optimal behaviour satisfies the natural monotonicity property in ω; greater success gives a higher value. Part (b) gives two types of "coasting" states, C_u and C_l, in which the firm neither invests in, nor exits from, the industry. Coasting in "successful" states, C_u, reflects the optimal response to a situation in which the expected marginal gain to further advance is outweighed by the marginal cost of further investment, $c(\omega)$. Recall that the return to investing is an increase in the probability of transiting to higher ω. The value of these increments is given by the "slope" of the value function. Since the value function is bounded that slope must eventually becomes less than the marginal cost of (even zero) investment. There are also states in which $A(\omega, s)$ is *low*, $x(\omega, s)$ goes to zero, and yet the firm does not leave the industry. Indeed the firm can choose to stay in the industry even in situations where it is optimal to shut down current production (possibly incurring a fixed cost for mothballing its plant). In these cases fixed costs are incurred, and exit values are foregone, because of the likelihood that an improved future condition $(s_{t+1}\prec s_t)$ will lead to a situation where it pays to produce and invest again.

There is, however, a limit to such lower coasting. When $(\omega, s)\in C$, $E(\Delta\omega|\omega, s)<0$ as $x(\omega, s)=0$, and hence ω drifts lower with probability $\sum_{\eta<0}p_\eta$ (A.4). This will reduce the value of the enterprise, $V(\omega', s')$, unless there is a countervailing shift in s so that $s'\prec s$. Indeed, without a random "improvement" in s, parts (b) and (c) insure that the firm will enter a true "liquidation state", $(\omega, s)\in L$, where $V(\omega, s)=\phi$ indicates the optimality of exit from the industry. That this occurs in finite time with probability one, despite the possibility of exogenous improvements, is the principal content of part (c).

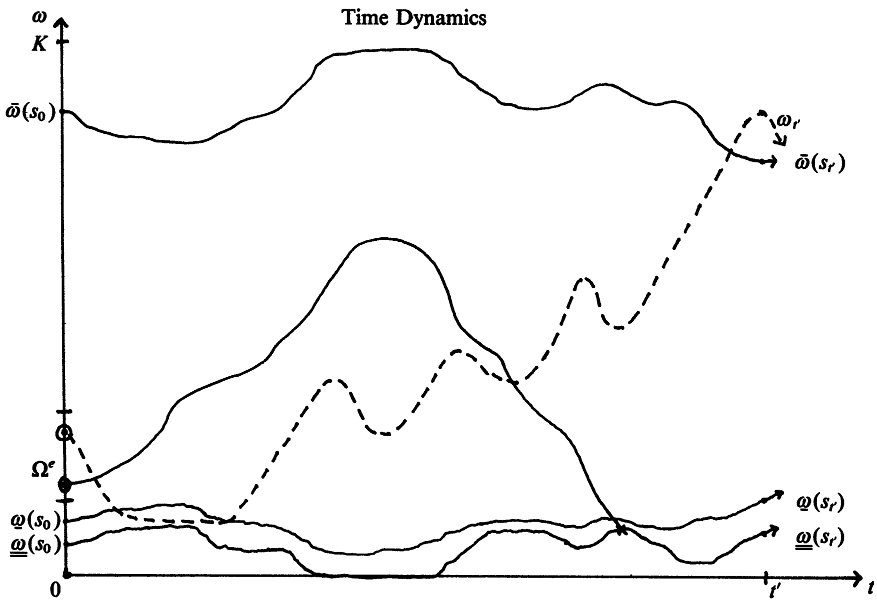
Proposition 1 characterizes firm behaviour in an industry in which active exploration and learning through investment is required for survival. We know that eventually all firms will die, but the life cycle of the firm can include a variety of behaviours, including periods of active struggle and learning $(x_t>0)$, with its successes $(\omega_{t+1}>\omega_t)$ and failures $(\omega_{t+1}\leqq\omega_t)$, periods of coasting on the successful outcomes of past efforts wherein no exploratory investment takes place but profits are derived from previous development, and, possibly, periods of coasting wherein a firm earns no profits and its current prospects warrant no further investment, but there is some probability that the market will "improve" $(s_{t+1}\prec s_t)$, which deters the firm from exit. Due to outside competition (p_0) and entry (ω_m) the state is inexorably moving in a direction unfavourable to the firm. Only through active investment $(x>0)$ can the firm hope to counteract this pressure. Yet, despite its best efforts, the firm must eventually succumb and liquidate, even though phenomenal profits may have been earned between birth and death. This situation is schematically illustrated, along with several possible sample paths for a firm, in Figure 1.

Despite the finite life of firms, it might be possible for entry rates to generate an unboundedly growing industry. It might be possible for either a countable set of firms to



⊙ —entry
× —exit
—sample paths

(a)



⊙ —entry
× —exit
—sample paths

(b)

FIGURE 1

decide to enter at some $s[V(\omega, s) > x^e_m, \forall m \geqq 1]$, or for there to be a steady excess of entrants over exitors, thus violating (A.7).[17] To bound the size of the industry, we provide a direct proof of the fact that $V(\omega, s)$ can be made arbitrarily close to ϕ for all (ω, s) by increasing the number of active firms in the industry. This will imply that $m(s)$ is finite for all $s \in \mathbb{Z}^{K+1}_+$ and that there exists an $N < \infty$ such that

$$S \equiv \{s \in \mathbb{Z}^{K+1}_+ \,||s| \equiv \textstyle\sum_{\omega=0}^{K} s_\omega \leqq N\}, \tag{9}$$

i.e. S is compact. Hence (A.7) is justified and w.l.o.g. we can normalize the full state space to $\sum = \{(\omega, s) \in \mathbb{Z} \times \mathbb{Z}^{K+1}_+ | \omega \in \Omega \subset \mathbb{Z}_+, s \in S \subset \mathbb{Z}^{K+1}_+\}$, where $\Omega = \{0, \ldots, K\}$. This is a key step both in showing the existence of an equilibrium and in computing it.

To prove this, we fix ω and an arbitrary structure s and consider a sequence of industry structures that increases the number of firms at ω, i.e. $\{s_n(\omega)\}_{n=1}^{\infty}$, where $s_n(\omega) = s + n \cdot e_\omega$. The following proposition shows that no matter which ω and s we fix, as n increases, the value to being in the industry at that ω falls to the exit value. Consequently enough entry will, in fact, choke off further entry, and there can never be more than a finite number of firms at any ω.

Proposition 2. *Let $s_n(\omega) \equiv s + n \cdot e_\omega$. Under Assumptions* (A. 0) *to* (A.6), *for all $\omega \in \Omega$, and all $s \in \mathbb{Z}^{\infty}_+ : \lim_{n\to\infty} V(\omega, s_n(\omega)) = \phi$, i.e. $\forall \varepsilon > 0\ \exists n_\varepsilon$ such that $n \geqq n_\varepsilon$ implies $V(\omega, s_n(\omega)) < \phi + \varepsilon$.*

Corollary 2.1. *There exists an $M < \infty$ such that, $\forall m \geqq M$, $V^e(s, m) \leqq x^e_m$, $\forall s \in S$.*

Corollary 2.2. *There exists an $N < \infty$ such that $V^e(1, s) < x^e_1$, i.e. $m(s) = 0$, for all $s \in \hat{S}_n(1)$ with $n \geqq N$.*

B. Existence of equilibrium

First note that assumptions (A.6) and (A.7) need no longer be imposed; they were made merely to facilitate analysis of a single firm in the industry. They are a consequence of the more basic assumptions, and our definitions of equilibrium transitions and entry decisions (6.c–d). (A.7) was shown to hold in the corollaries to Proposition 2, while (A.6) follows from the following proposition.

Proposition 3. *Under assumptions* (A.0)–(A.5), *assumption* (A.6) *holds with $Q(\cdot|\cdot)$ defined using* (6.c) *and* (7), *when $q_\omega(\hat{s}'|s)$ is defined by equation* (8).

We can now prove the existence of a rational expectations equilibrium for this model of active exploration and learning through investment. This closes the model by showing that the assumptions on the industry structure and its evolution used to determine optimal behaviour are in fact consistent with that behaviour. To do so, we show that given $Q(\cdot|s)$, as defined in (6.c), the optimal decisions of incumbents solving (3) and entrants satisfying (4) generate transition probabilities which aggregate to form $Q(\cdot|s)$. This requires a fixed point argument that is outlined in the Appendix. In essence, it involves showing that

17. We note that if $V(\cdot)$ were isotone to $\succeq$ on S [i.e. $\forall \omega$, $s_1 \succ s_2 \Rightarrow V(\omega, s_1) \prec V(\omega, s_2)$], then new entrants would increase s' driving $V(\cdot, s')$, and hence $V^e(s, m)$, down, eventually choking off entry. Unfortunately, the subtleties generated by the interactions among agents (particularly in entry deterrence), imply that it is not in general true that $V(\cdot)$ is isotone in s, so that one cannot use this fact to stop entry (or induce exit) as the number of firms in the industry grows. See, for example, the discussion of Figure 2 in Section IV.

investment, entry and exit decisions depend continuously on the distribution of future states, which in turn depends continuously on those decisions. The continuous compound function maps a compact, convex space of probability distributions into itself, and hence has a fixed point: a rationally anticipated Markov transition function $Q(\cdot \mid s)$.

Theorem 1. *Under Assumptions* (A.0)–(A.5) *there exists an equilibrium* (6), *satisfying conditions* (6.a–e).

This theorem shows both existence of equilibrium and that the preceding results for a firm in the industry are valid at equilibrium. Due to the autonomous structure of the model the equilibrium is characterized by stationary valuation of states, stationary optimal investment strategies, and stationary Markov transition probabilities. Yet the sequence of states for any firm, and, indeed, the sequence of (almost surely finite) structures for the entire industry, are truly random realizations from an underlying stochastic structure. This structure is determined by the specific values of the parameters of the model, and by our equilibrium conditions. We now turn to its analysis.

C. *Equilibrium dynamics*

This dynamic equilibrium is characterized by a remarkable degree of flux. Active firms are truly heterogeneous, distinguished by their "state of success," ω, and have truly idiosyncratic outcomes to even identical investment decisions. Multiple rank reversals (according to criteria such as sales, profitability, employment) are possible during the life of any collection of firms (cohort), as is simultaneous entry and exit ($\underline{\underline{\omega}}(s) < \omega^0 \in \Omega^e$). All firms die in finite time (a.s.), yet new firms continually enter to try their fortune in the evolving industry. Thus the structure of the industry can change dramatically over time, although it must remain finite (Corollary 2.2). In view of this continual change, the question of characterization of the "average" structure of the industry and its relation to the industry's long-run evolution arises.

Among the things that we would like to know are whether the industry structure settles down into some recurrent pattern and, if so, the characteristics of that pattern. For example, does the industry survive forever, or might it fade away as fewer and fewer firms enter while old firms exit one after another? If the industry does survive, is there a sense in which we can speak of a long-run average number of firms, or structure, for the industry? What determines these and other characteristics of the process defined by the industry equilibrium, and how do they change in response to perturbations of various environmental and policy parameters? This section proves a result which lies at the heart of our ability to answer these questions: the ergodicity of the stochastic process defined by the industry equilibrium.[18] Some direct implications of this ergodicity will be noted outright, but answers to many of the more interesting questions about the nature of the ergodic distribution will depend on the detailed characteristics of functional forms in our model. We begin to explore some of these in an example in Section IV.

Before turning to a formal analysis, we would like to emphasize two points on its relevance. First, one of the advantages of an explicit dynamic model such as ours is that it allows us to study the distribution of the entire sequence of structures that the industry passes through and not just some notion of a limit structure. Our focus here on long run

18. Here we use ergodicity in the wide sense: a stochastic process is said to be *ergodic* if it converges to a stationary ergodic process. See Halmos (1956) or Friedman (1970).

averages stems from the fact that, at least in the absence of a specific empirical example with a particular value for s_0, if one wants to investigate the effect of a policy or environmental change on the (distribution of the possible) structures of the industry, a natural place to start is to investigate the effects of these changes on the time-average of the structures the industry will pass through. This leads us immediately to the question of whether there is a time-average, in particular one that is independent of initial conditions, to which all sequences converge. Second, for these limiting results to be appropriate, our behavioural assumptions might have to provide an adequate approximation to those prevalent in the industry over fairly long time periods.

Formally, the evolution of the equilibrium structure of the industry, s_t, is given by

$$s_t = (I[\{\omega > \underline{\underline{\omega}}(s_{t-1})\}] \cdot s_{t-1})' + \omega_{m(s_{t-1})}, \tag{10}$$

where $I[\{\omega > \underline{\underline{\omega}}(s_{t-1})\}]$ is a diagonal matrix whose diagonal elements are either unity [if $\omega > \underline{\underline{\omega}}(s_{t-1})$] or zero, $\omega_{m(s_{t-1})}$ is the realization of the counting measure giving the location of firms paying their entry fee in $t-1$, and 'prime' indicates a realization from the distribution $q^0(\cdot\,|\,\cdot)$.[19] Here equilibrium transition probabilities, entry, and exit are defined in (6.b–d). By Proposition 2, the state space, S, for this stochastic process is compact, and hence finite. Let $Q(s, s')$ be the stationary transition matrix of the equilibrium transition probability function $Q(s'|s)$ defined in (6.c). Then $s \equiv \{s_t\}_{t=0}^{\infty}$ is a Markov process with stationary transitions given by the $|S| \times |S|$-matrix Q and with distribution [sample path probabilities]

$$P_{s^0}\{s_t = \bar{s}_t \text{ for } t = 0, \ldots, n\} = e_{s^0} \cdot \textstyle\prod_{t=0}^{n-1} Q(\bar{s}_t, \bar{s}_{t+1})$$

for a specific path $\bar{s} = (\bar{s}_1, \bar{s}_2, \ldots)$ when the process begins in state s^0. Similarly P_v is the distribution of this Markov process when the initial state has probability v_s of being in state (having structure) s. Therefore, the distribution of industry structures evolving from an initial s^0 after n periods can be written

$$\mu_n(s^0)\ [\mu_n(v)] = e_{s^0} Q^n\ [v \cdot Q^n] \in \Delta^S, \tag{11}$$

where Q^N is the n-th iterate (power) of Q and Δ^S is the $(|S|-1)$-dimensional simplex. That is, $\mu_n(v)$ is an $|S|$-vector whose elements $\mu_{n,s}$ give the probabiality that the structure of the industry, with initial distribution v, is in state s after n periods.

This notation enables us to formulate our principal result on industry equilibrium dynamics: the evolution of the industry is ergodic in that the stochastic process defined by the industry equilibrium possesses a unique limiting distribution of structures.

Theorem 2. *Under Assumptions* (A.0) *through* (A.5) *at equilibrium* (6):

(a) *The stochastic process $s = \{s_t\}_{t=0}^{\infty} \in (S^{\infty}, \otimes\mathcal{S})$ with initial state s^0 is Markov with stationary transitions $Q(s, s')$ and distributions P_{s^0}, where $\mathcal{S}$ is the σ-field of all subsets of S.*

(b) *The state space, S, contains a unique, positive recurrent communicating class $R \subset S$.*

(c) *There exists a unique, invariant probability measure, μ^*, on S such that $\mu_s^* = [mQ(s, s)]^{-1}$ for $s \in R$, and $\mu_s^* = 0$ for $s \in S \backslash R$, where $mQ(s, s')$ is the P_s-expectation of the time of first reaching state s'.*

(d) $\forall s \in S$, $\mu_n(s) \to_{n\to\infty} \mu^*$.

19. $q^0(\cdot)$ is defined in note 16 above. Also see the Remark in Section II.E.

Corollary. *P_{μ^*} is the distribution of a stationary, ergodic Markov process with transition Q, i.e. $\mu^* Q = \mu^*$.*

Ergodicity of the equilibrium process generating industry structures has a number of empirical implications. First, it implies that the industry structure evolves in a non-degenerate, though increasingly regular, way over time, so that there never is a "limit" structure of the industry. Indeed, all viable industry structures, that is all structures in the recurrent class $R \subset S$, are realized infinitely often. Thus, just as there is continual flux in the relative position of firms in the industry, there is continual change in the industry structures that those firms comprise, showing that there is much less of a relationship between structure and behaviour, and indeed between structure and the welfare properties of the resulting equilibrium, than traditional models assume.[20] A given industry structure generates investment, exit, and entry decisions as optimal responses to the valuation of the opportunity presented by the industry. The idiosyncratic outcomes of these investment decisions, together with the evolution of the state of competition from outside the industry, determine the structure of the industry at the beginning of the next period, a structure that is only probabilistically related to the structure which generated it. Though all firms eventually die, entrants replenish the population of active firms, and hence the industry of this model lives forever, eventually going through all the epochs determined by its recurrent states and its transition kernel.

Another consequence of ergodicity is that, after some time, a certain stochastic regularity will appear in the evolution of the industry. If the initial structure is transient, $s^0 \in S \backslash R$, then a finite (a.s.) time will be spent shifting to some recurrent structure, $s \in R$. Thereafter, the portion of time spent in any state $\underline{s} \in R$ will approach the invariant probability of that state, $\mu^*_{\underline{s}}: \lim_{T \to \infty} 1/T \sum I_{\underline{s}} = \mu^*_{\underline{s}}$. Thus the structure of the industry, s_t, while shifting randomly in response to the idiosyncratic outcomes of optimal decisions by firms, will spend more time near "natural" states, with a "natural" number of incumbents, entrants and exits. What is "natural" will depend on the values of the underlying parameters of the industry, $\theta = \{A(\cdot), c(\cdot), \phi, \beta, \hat{\omega}, \pi(\cdot), p_0, P, x^e_m, \Omega^e\}$, and will be reflected in the mass of the invariant measure over the set of recurrent structures. Thus, over time, structures that are natural or normal for this industry will reveal themselves as more likely by their more frequent occurrence: time averages will approximate state averages, i.e. the ergodic distribution, μ^*.

A final consequence of ergodicity is that the influence of any initial situation systematically fades, becoming irrelevant for the future evolution of the industry. As Theorem 2.d indicates, the actual distribution over industry structures, μ_n, evolving from any initial structure, s^0, (or distribution over structures, ν), converges to the unique invariant distribution, μ^*, hence losing any information that it contained about the initial condition of the industry. Indeed, a strong Markov property (Freedman (1983), § 1.3) holds in this class of models; the future is independent of the past conditional on any measurable (Markov time) event. Thus two possible histories for the industry with different initial conditions (structures), once they intersect in any state, as they must with probability one, have identical distributions over future sample paths conditional on that intersection.

These ergodic characteristics of the model differentiate it from other stochastic dynamic equilibrium models currently in the literature. Models of competitive industries have a continuum of infinitesimal firms leading to a deterministic limit structure. In

20. This is clearly illustrated in the results of simulations discussed in Section IV below and in Pakes–McGuire (forthcoming).

+Jovanovic (1982) there is no entry or exit and a fixed distribution of active "good" firms in the limit; in Hopenhayn (1992) the fixed limiting distribution incorporates entry, exit and changing firm productivities due to continuing exogenous shocks. The Hopenhayn model is ergodic, while learning models lack ergodicity as there is a time-invariant parameter generating observable sequences that differentiates among firms. A simple nonparametric test based on a ϕ-mixing condition can be used to test for ergodic vs. nonergodic models (Pakes–Ericson (1990)). Lambson (1992) generates firm heterogeneity without idiosyncratic shocks through hysteresis of entry/investment decisions. There the competitive industry equilibrium process may be ergodic if the exogenous market environment process is first-order Markov.

The game-theoretic models of dynamic industry evolution are more specialized, all assuming a duopoly structure. They generally focus on characterizing the optimal strategies in a Markov-perfect equilibrium, and make more detailed progress by assuming specific functional forms. Most of these models (Maskin–Tirole (1988); Rosenthal–Spady (1989); Beggs-Klemperer (1992); Cabral–Riordan (1992)) generate a deterministic evolution to some fixed structure, perhaps with different firms at different times. Rosenthal–Spady and Cabral–Riordan allow exit and entry to preserve duopoly despite the outcome of competition. Only Maskin–Tirole can generate a non-trivial ergodic process (an Edgeworth cycle) from the impact of randomized strategies. The most developed of these dynamic duopoly models is that of Budd *et al.* (1993) which explores the optimal Markovian control of a one-dimensional diffusion process of market state (share), building on an earlier model of Harris (1988). Since the model has arbitrary boundary behaviour on a compact interval, rather than endogenous entry and exit, the authors do not focus on the ergodic distribution over industry states. Rather, four determinants of Markov perfect equilibrium strategies are uncovered, and related to existing results in the dynamic duopoly literature. Despite the vast structural differences from our model (continuous time and state space, etc.), it seems that similar factors drive optimal investment behaviour in it.[21] Finally, there are a number of repeated "technology race" models of industry evolution, focused on whether the industry exhibits growing dominance of one firm or alternating (technological) leadership (Vickers (1986); Beath *et al.* (1987); Dutta *et al.* (1993)). These models find Nash equilibria with respect to a finite sequence of fixed technological innovations, thus exhibiting simple dynamics. Only the Dutta *et al.* model is stochastic, but its dynamics end after a second innovation (a refinement).

All of these models are distinguished from ours by the simplicity of their dynamics, particularly in the limit (steady state); none allows simultaneously for an endogenously determined industrial structure with finitely-lived, heterogeneous firms, endogenous entry and exit, and dynamics whose only regularity is imposed by ergodicity of the limiting distribution. Thus they do not provide as comprehensive or flexible a framework for applied work as does the present model. Applications to date include an analysis of merger activity (Gowrisankaran (1994); Berry–Pakes (1993)), an analysis of the interaction between for- and non-profit institutions in the Hospital industry (Gowrisankaran–Town (1994)), and an analysis of the evolution of productivity in the telecommunications equipment industry (Olley–Pakes (1991)).

The general characterization in Theorem 2 still leaves many questions about the finer structure of the equilibrium stochastic process. For example, does the unique stationary

21. Their "end point" effects (relief from effort) are immediate near and at coasting states, and the "self-reinforcing joint cost" effects seem evident in our simulations (Section IV and Pakes–McGuire (forthcoming)). We have not looked for the "joint profit" or "pattern of profit" effects, but believe that they could be found.

ergodic distribution, to which the time-average of the industry structures eventually converges (see 2.d), possess a large number of small firms or a small number of large firms? Are the industry structures of the recurrent class "similar", so that one can think of the industry's structure "settling down" after some finite number of periods? Or does this recurrent class contain very diverse structures, so that no matter how long the time period elapsed since the "start up" of the industry we will still observe the industry structure undergoing distinct evolutionary patterns? To the extent that the recurrent class contains quite divergent industry structures, do the sample paths through these structures typically cycle, and if so, with what periodicity, or are there Poisson-type events that cause relatively quick and sharp discontinuities in the industry structure? Which structures of the recurrent class generate large amounts of simultaneous entry and exit, and which generate periods of high investment? Finally, and perhaps most importantly, how long will it generally take before the industry's structure enters the recurrent class, and through what type of sample paths does an industry typically pass before its recurrent pattern becomes evident? We have begun to explore such questions, and how their answers change in different policy or environmental settings, in some numerical examples. Some answers appear highly sensitive to precise functional forms or parameter values, while others seem more robust to these detailed assumptions. We turn to one such example now, and compare it to others that we have computed elsewhere.

IV. AN EXAMPLE

As an example, we consider a homogeneous product market having producers with different, but constant, marginal costs. Marginal costs, θ_ω, are determined by a firm specific efficiency index and a common factor price index. Let $-\eta$ be the logarithm of the factor price index and τ that of the firm's efficiency index; then $\omega \equiv \tau + \eta$ and $\theta_\omega = \exp(-\omega)$. Firms' R&D investments are directed at improving their efficiency of production (increasing their τ). Factor prices $(-\eta)$ are a non-decreasing stochastic process generating the correlated negative drift in the state of the firms in the industry.

The spot market equilibrium in this market is assumed to be Nash in quantities. Letting q_i be firm i's output, $Q = \sum q_i$, and f be the fixed cost of production, the profits of our classic Cournot oligopolists are given by $A_i = p(Q)q_i - \theta_i q_i - f$ where $p(Q) = D - Q$. It is straightforward to show that the unique Nash equilibrium for this problem gives quantities and price as $q_i^* = \max\{0, p^* - \theta_i\}$ and $p^* = [D + \sum_{i=1}^{n^*} \theta_i]/(n^*+1)$, where n^* is the number of firms with $q^* > 0$. Current profits can therefore be written as $A(\omega, s) = [p^*(s) - \theta_\omega]^2 - f_\omega$, where $p^*(s) = [D + \sum_{\omega \geq \omega^*} s_\omega \cdot \theta_\omega]/(n^*+1)$, and $\omega^* = \min\{\omega | q_\omega > 0\}$.[22]

To complete the specification we assume:

$$\theta_\omega = \gamma e^{-\omega}, \text{ with } \pi(\omega'|\omega, x) = \pi(\omega' - \omega | x) \equiv \pi(\tau + \eta | x),$$

$$\pi(\tau|x) = \begin{cases} ax/(1+ax) \text{ that } \tau = 1 \\ 1/(1+ax) \text{ that } \tau = 0 \end{cases}, \quad p_0 = \begin{cases} 1-\delta \text{ that } \eta = 0 \\ \delta \text{ that } \eta = -1 \end{cases}, \quad c_\omega = c,$$

$$\Omega^e = \{\omega^0 - 1, \omega^0\}, \; x_1^e = x^e \text{ and } x_m^e = \infty \text{ for } m > 1, \text{ and } \omega^e = \omega^0 + \eta.$$

Transitions in ω are determined by the difference between the increment in efficiency of

22. This current profit function is, in many senses, an extreme alternative to the profit function used in the example of our model analysed in Pakes–McGuire (forthcoming). It considers a differentiated product industry in which all firms have the *same* (constant) marginal costs but are differentiated by the quality of the product they produce; a quality which increases with successful research activity. In that example the spot market equilibrium was assumed to be Nash in prices.

production generated by the outcomes of the firm's own research activity (τ), and the increment in the factor price index ($-\eta$). τ can either increase by 1 or stay the same. The probability of τ increasing is an increasing function of investment, and the cost of a unit of investment is independent of ω. η either decreases by one or stays the same, but here the probabilities are given by an exogenous process. In each period there is at most one entrant who pays a setup cost of x^e and enters in the following year at state ω^0 if the cost of production has not increased in the interim, and at ω^0-1 if it did ($\eta=-1$).[23]

It is easy to see that this specification (together with an appropriate choice for β and ϕ) satisfies all of (A.0) through (A.5), although (A.3) perhaps requires checking. Note that

$$A(\omega, s)=\max\left\{\left[\frac{D+\sum_{k\geqq\omega^*} s_k\cdot \gamma e^{-k}}{|s^*|+1}-\gamma e^{-\omega}\right]^2-f, -f\right\}, \qquad \text{where } |s^*|=\textstyle\sum_{k\geqq\omega^*} s_k.$$

Clearly $\bar{A}=(D+\gamma)^2$, and $A(\omega, s)$ is increasing in ω and decreasing in s with the natural vector pre-order. In particular, $A(\omega, s)\downarrow -f$ as s_k increases at any $k\geqq\omega$, or as ω falls for any s. Note that if $A(\omega, s)=-f$ then marginal cost is greater than price and the firm is not active in the spot market. The same firm can, however, still be a participant in the industry. That is, plants will be mothballed without being dismantled if there is sufficient hope that the environment will improve to the extent that it will pay to bring the plant back on line in the future.

As this example satisfies our assumptions, all of the results of Section III hold and we have a well-defined dynamic equilibrium that generates an ergodic Markov process in industry structures. To obtain the more detailed results, we substitute the specification given above into the computational algorithm developed specifically for this model in Pakes–McGuire (forthcoming), initialize the various parameters, and let that algorithm calculate the policy functions for all (potential and active) firms. This allows the generation of statistics that describe the industry structures, and the welfare implications, of the Markov–Perfect Nash (MPN) equilibrium. We also calculate the optimal policies for both a social planner and a multi-plant monopolist (or perfectly colluding cartel) faced with the same cost and demand primitives as those generating the MPN equilibria, and then generate the descriptive statistics and welfare measurements that emanate from the equilibria obtained from these institutional environments. The colluder makes all decisions (investment, quantities marketed, entry, and exit) to maximize the expected discounted value of the total profits earned in the market. Similarly the planner maximizes consumer surplus.[24]

Some of the results from these computations are listed in Table 1. All descriptive statistics are obtained from simulation runs starting with one firm at the entry state [i.e. $s_0=e_{\omega^0}$], and then using the computed policies to simulate from that point. Panel A and B provide descriptive statistics from a 10,000 period simulation run. Panel C provides the distribution of expected discounted values from 100 independent simulation runs of 100 periods each.

Panel A indicates that this is an industry which is most often a duopoly, though in a significant fraction (about a quarter) of the periods only one firm is active. Note that "monopoly" positions here are built up solely from successful past research; a firm which

23. Note that by assuming that the period of time at which we actually observe new data points is larger than the decision period of the model, this specification could allow for both many entrants and for richer conditional distributions for the changes in ω per data period, while still maintaining the computational advantages available when there are single step transitions.

24. There is a question of whether there is a feasible set of institutional arrangements which could lead to an industry which follows either a colluder's or a planner's dictates.

TABLE 1

*Simulated quantities from a homogeneous product model**

A. % of Simulated Periods with	MPN	Colluder	Planner
1 firm active	27·9	92·4	98·3
2 firms active	70·8	7·6	1·7
3 firms active	1·2	0	0
4 firms active	0·1	0	0
Entry and Exit	16·5	5·4	1·2
Entry or Exit	20·4	10·0	2·1
B. Average (standard deviation of)			
Price	1·79 (0·35)	2·22 (0·36)	**
Total investment	1·05 (0·41)	0·68 (0·29)	0·84 (0·41)
Entry	0·19	0·08	0·02
Number active	1·74 (0·48)	1·08 (0·27)	1·02 (0·13)
C. Welfare Runs (average and, in parenthesis, standard deviation of)			
1. Discounted Consumer Benefits	27·4	6·6	**
	(6·4)	(5·5)	
2. Discounted Net Cash Flow	11·6	22·5	—
	(5·4)	(8·5)	
3. Discounted Entry-Exit Fees	2·5	1·0	—
	(1·0)	(1·0)	
4. Discounted Welfare	36·5	28·1	58·8
	(11·9)	(14·1)	

* All runs are based on the specification described in the text with the following parameter values $D=4$, $f=0{\cdot}2$, $x^e=0{\cdot}4$, $w^0=4$, $\phi=0{\cdot}2$, $c=1$, $S=0{\cdot}7$, $a=3$, $\beta=0{\cdot}925$. Panels A and B are obtained from a run which starts with one firm entering the industry, goes 10,000 periods, and then calculates the appropriate descriptive statistics. Panel C is obtained by doing 100 runs, each starting with one entrant and each lasting 100 periods. The appropriate discounted values are taken from each run, and then their averages and standard deviations across runs are computed.

** The welfare result for the planner can be read off the value function which is computed exactly. The planner sets price equal to the marginal cost of the minimum cost producer. This minimum marginal cost averaged 0·15 with a standard deviation of 0·39.

is efficient enough will deter entry (a "persistence-of-dominance" effect). Of course, as noted in our theoretical results, even the most efficient of firms will eventually decay and be taken over by more successful competitors. Consequently it is not the same two or three firms that are active in all of the periods. Indeed this industry exhibits substantial entry and exit; there is entry in about 19% of the periods. Moreover entry and exit are positively correlated, a fact which is consistent with the time series evidence in many (though not all) industries (see Dunne *et al.* (1988)), and which very clearly brings out the need for allowing for idiosyncratic sources of uncertainty.

Figure 2 provides a section of the optimal investment policy surface. The vertical axis gives the investment of a firm as a function of its own ω (ω_1) and the ω of a competitor (ω_2) when no other firms are active. From the figure it is clear that the firm starts investing at $\omega=3$ or 4 depending on the value of its competitor's ω. Thereafter investment is an initially increasing and then decreasing function of the firm's own ω.

In another paper (Ericson–Pressman (1989)) we note that the investment function *must* be an initially increasing and then decreasing function of ω for a monopolist with functional forms for the primitives similar to the ones used here. The intuition behind this result follows from the form of the value function. As investment increases the probability of increments to ω, it will increase when those increments result in larger increments to

Optimal Investment: Firm 1

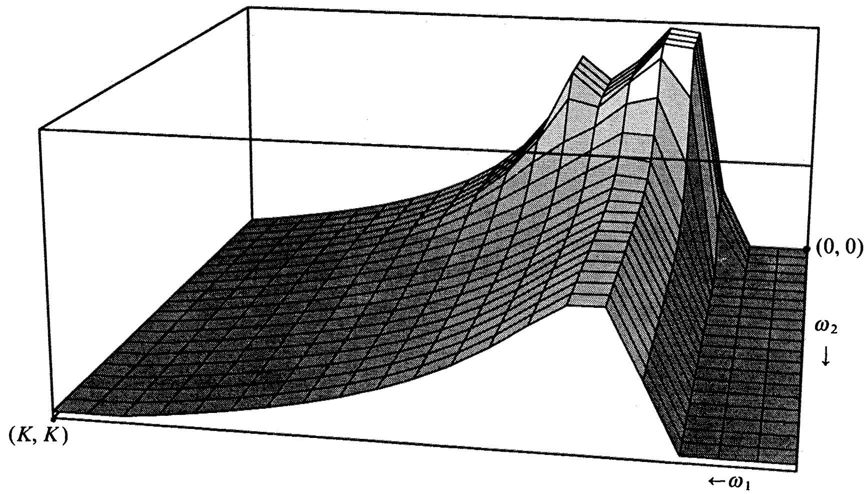


FIGURE 2

the value function. Thus an initially convex and then concave value function will generate an initially increasing and then decreasing policy function. Indeed, the value function for all of our models has this property as it is bounded from both below and above. In the case of a monopolist with a simple enough profit function we can show, in addition, that the value function has only one point of inflection. Once we allow for free entry and consider sections of the value function that hold competitors ω's fixed, then there need not be only one inflection point, but the initial convexity and eventual concavity of the value function are maintained.

This implies that new entrants begin with a relatively low level of investment. As a result most entrants will never actually overcome the negative drift imposed by advances of their competitors both inside and outside the industry, and die at early ages. This generates high mortality rates in an initial "learning" period, and a large fraction of entrants whose realized discounted value of returns from participating in the industry are negative. On the other hand the few new entrants who do get a good sequence of initial draws begin to increase their profits and invest more, thereby increasing the probability that they develop even further. Of course the successful firms will eventually pass over an inflection point of the value function, and decrease their investment, at which point their expected increment in ω will fall. However once their ω falls back to near the inflection point their investment will pick up again, so that an initially successful firm will tend to be productive for a long period of time. This, in turn, implies that both the lifetime and the realized value distributions from our model tend to be very skewed (see also Pakes-McGuire (forthcoming); similar, in fact, to the life spans and value distributions reported in the empirical literature.[25]

Figure 2 also shows how the subtleties generated by the interactions among agents can destroy any simple generalizations on the form of the value function. Consider

25. The literature on lifespan distribution is extensive; see Dunne *et al.* (1988), Pakes-Ericson (1990), and the literature cited in those articles. There is less on value distributions, but a more substantial literature on both the distribution of sales and profits, and persistence in the process generating the sales and profits of different firms. Also see Evans (1987*a*, *b*), Hall (1987), and Mueller (1986), and the literature referred to therein.

any one of the sections in which ω_2 is low ($\omega_2 \leqq 5$) and follow the investment pattern of the first firm as its ω increases. As before it is initially increasing until about $\omega = 5$, and then it decreases, but at $\omega = 8$ we see a surge of investment, which heads back down after $\omega = 9$. The reason for the increase in investment at $\omega = 8$ is to deter entry. It works out that a potential entrant finds it profitable to enter if there is one firm in the industry at $\omega = 7$, but not if there is one firm in the industry at $\omega = 9$. This surge in investment destroys the simple characterization of the investment function that we get if there are no potential competitors as in Ericson–Pressman (1989).

Coming back to the first panel of Table 1, it is clear that both the planner, and the colluder, tend to generate equilibria with fewer firms than does the Markov Perfect Nash solution. Indeed, given that the optimal policy for both the planner and the colluder is to have only one firm actually produce output in any period, the firm with the lowest θ_ω, it is somewhat surprising that either of these two institutional structures ever find it optimal to have more than one firm active. They do because it is sometimes optimal for them to run parallel R&D efforts (see Nelson (1960)), and then only use the most efficient production technique developed. Still the logic behind the fact that both the colluder and the planner have less entry and generate less investment (panel B) than does the MPN solution is clear enough; entry and investment decisions in the MPN solution depend on the expected incremental cash flow going to the entrant and to the investor, and some of this cash flow is taken away from (other) incumbents. Both the colluder and the planner internalize the losses to incumbents and hence invest less (Mankiw–Whinston (1986)). This result is similar to that of the differential products example, although there the planner had distinctly more entry and investment than the colluder due to the impact of product variety on consumer surplus.

There are several other interesting aspects of the numerical results that are similar to those obtained from the differentiated products case. First, note that though the colluder generates an industry structure that looks much more like the planner than does the industry structure from the MPN solution, the welfare generated by the MPN solution is much higher and hence closer to that generated by the planner. The big difference between the welfare results in the homogeneous and differentiated product cases is that in the differentiated product example the welfare from the MPN solution was generally within 2–3% of the welfare that a planner could generate, even when equilibrium typically involved only two firms active, leaving little room for improvement over the "free market". In the homogeneous product case, at least with parameters typically generating only one or two active firms, the difference between the welfare generated by the planner and that generated by the MPN solution seems to be much more substantial (on the order of 40%).

More generally, in simulations we have consistently been surprised by the extent to which institutional structures which generate "similar market structures" (similar numbers of firms active, similar shares for the largest firms, similar entry and exit, etc.) can have very different welfare implications, and institutional arrangements which lead to very different market structures can generate very similar welfare results. Also we have found surprisingly high standard deviations for the welfare results from any given institutional structure. In this example, the average difference in total welfare between the MPN and the Colluder's solution is less than the standard deviation of the welfare results from either of them. This should make us wary about generalizing from case study attempts to compare different institutional arrangements, even when the case studies have a "laboratory perfect" comparison to make in the sense that the other primitives of the model are the same in the two institutional arrangements being compared.

V. CONCLUDING REMARKS

We noted in the Introduction that models of firm and industry behaviour that allowed for idiosyncratic, or firm-specific, uncertainties and entry and exit were required in order to account for many of the phenomena exhibited in firm-level data sets. These phenomena include: simultaneous entry and exit; strikingly different outcome paths from similar initial conditions, investment strategies, and exogenous events; and industry structures that never seem to remain stable. We also noted that the need for models which can account for such phenomena is not merely descriptive, but indeed lies at the heart of our ability to analyse many of the impacts of policy and environmental changes.

This paper has provided one possible model of firm and industry dynamics that can account for these empirical phenomena. The focus has been on the basic logic and implications of the model in a framework that is general enough to accommodate primitives that could be thought appropriate for a broad number of industries in which research and exploration processes are important. Even at this level of generality, however, the model is rich enough to both generate empirically testable implications (Pakes–Ericson (1990)), and to suggest nonparametric procedures for correcting for selection (induced by entry and exit) and simultaneity (induced by endogenous input demands) problems when analysing firm's responses to policy and environmental changes (Olley–Pakes (1991)).

However, many of the more detailed issues that one might want to analyse with the model depend on the finer properties of the primitives of our model, θ, and are currently buried in the relationship between those primitives and the nature of the equilibrium process generating industry dynamics. For both policy and descriptive purposes we will ultimately be interested in the relationship between each primitive and the recurrent class of industry structures, the ergodic distribution on that class, and the nature of the transition process into that class. This would enable us to analyse how a change in either a policy variable (such as an R&D tax credit, or a tariff) or in the external environment (such as a technical change that increased the effectiveness of external competition, or a shift in the structure of demand), affect the nature of the equilibrium process generating industry supply, productivity, shut downs, default probabilities, job creation and destruction at the firm level, etc.

There are at least three (related) ways of proceeding to the more detailed analysis required to unravel these relationships. In order of (what we believe to be) increasing difficulty, they are: simulation based on assumed functional forms and particular parameter values for all of the primitives (see Section IV), comparative dynamics within parametric classes, and simulation based on estimated functional forms. We are pursuing all three of these in related research.

APPENDIX[26]

Proof of Proposition 1. (a) The only assertion not immediate from standard results is the monotonicity of $V(\omega, s)$ in ω. Existence, uniqueness and boundedness follow from the properties of the linear operator T, $T: l_\infty(\mathbb{Z}\times\mathbb{Z}^\infty)\to l_\infty(\mathbb{Z}\times\mathbb{Z}^\infty)$,

$$Tu(\sigma)\equiv\max\left\{\max_x\,(A_\sigma - c_\sigma x + \beta\cdot\textstyle\sum_{\sigma'\in\Sigma} u_{\sigma'}p(\sigma'|\sigma, x)),\ \phi\right\}, \tag{A1}$$

where $p(\sigma'|\sigma, x)\equiv\sum_{\eta'} q_\omega(\hat{s}'|s, \eta')\cdot p(\omega'|\omega, x, \eta')\cdot p_{\eta'}$.

26. Space limitations preclude full proofs. They can be found in Ericson–Pakes (1992).

Let $\omega_1 \geqq \omega_2$. By the contraction property of the linear operator T, $V(\omega, s) = \lim_{n\to\infty} V^n(\omega, s)$ where $V^n(\omega, s) \equiv TV^{n-1}(\omega, s) \equiv T^n A(\omega, s)$. The proof follows by induction from (A3). Let monotonicity hold at step n.

$$
\begin{aligned}
V^{n+1}(\omega_1, s) - V^{n+1}(\omega_2, s) &= TV^n(\omega_1, s) - TV^n(\omega_2, s)\\
&\geqq A(\omega_1, s) - A(\omega_2, s) + \beta \cdot \textstyle\sum_{\eta'} \{\sum_{\omega_1'} [\sum_{\omega_2'} V^n(\omega_1', \bar{s}' + e_{\omega_1'} + e_{\omega_2'}) q_{\omega_1\omega_2}\\
&\quad \times (\bar{s}'|s, \eta') p(\omega_2'|\omega_2, x_1, \eta') - \textstyle\sum_{\bar{s}'} \sum_{\omega_2'} V^n(\omega_2', \bar{s}' + e_{\omega_1'} + e_{\omega_1'}) q_{\omega_1\omega_2}\\
&\quad \times (\bar{s}'|s, \eta') p(\omega_2'|\omega_2, x_1, \eta')] \cdot p(\omega_1'|\omega_1, x_1, \eta')\} p_{\eta'}\\
&= \textstyle\sum_{\eta'} \sum_{\omega_1'} \sum_{\omega_2'} \sum_{\bar{s}'} [V^n(\omega_2', \bar{s}' + e_{\omega_1'} + e_{\omega_2'}) - V^n(\omega_1', \bar{s}' + e_{\omega_1'} + e_{\omega_2'})]\\
&\quad \times q_{\omega_1\omega_2}(\bar{s}'|s, \eta') \cdot p(\omega_1'|\omega_1, x_1, \eta') \cdot p(\omega_2'|\omega_2, x_1, \eta') \cdot p_{\eta'} \geqq 0, \qquad \text{(A2)}
\end{aligned}
$$

where $x_i \equiv x^n(\omega_i, s)$; $\hat{s}_i$, ω_i' are similarly defined; $\bar{s} \equiv s - e_{\omega_1} - e_{\omega_2}$; $q_{\omega_1\omega_2}$ is the marginal derived from either q_{ω_i}: $q_{\omega_1\omega_2}(\bar{s}'|s, \eta') \equiv \sum_{\omega_j} q_{\omega_i}(\bar{s}' + e_{\omega_j}|s, \eta')$, $i \neq j$; and prime indicates next period's (random) realization of the variable. The first inequality in (A2) is due to the use of $x(\omega_1, s)$ at (ω_2, s) and substitution of the appropriate marginal probabilities. The second follows from the monotonicity of $A(\omega, s)$, $c(\omega)$, $V^n(\omega, s)$ in ω, since s' and the associated probabilities must be identical at both ω, as they arise from the same s and their firms invest identically. The first step of the induction follows from an identical argument with $A(\omega, s)$ in place of $V^n(\omega, s)$.

(b) From the first-order conditions given in equation (6.b), we know that $x(\omega, s) > 0$ iff $G(\omega, s) \equiv \beta \cdot \sum_{\eta'} \sum_{\omega'} \hat{V}(\omega'|\omega, s, \eta') \cdot p_x(\omega'|\omega, x(\omega, s), \eta') \cdot p_{\eta'} > c(\omega)$, where $\hat{V}(\omega'|\omega, s, \eta') \equiv \sum_{\hat{s}'} V(\omega', \hat{s}' + e_{\omega'}) q_\omega(\hat{s}'|s, \eta')$. Part (a) shows that $V(\cdot) \in [0, \bar{V}]$ and that, by monotonicity in ω, $\forall s$, $\lim_{\omega\to-\infty} V(\omega, s) = \phi$ and $\lim_{\omega\to\infty} V(\omega, s) = \bar{V}$. Therefore $\forall s$ $\lim_{\omega\to\pm\infty} G(\omega, s) = 0$, as the support of $\sum_{\eta'} p_x(\cdot) \cdot p_{\eta'}$ is finite ($k_1 + k_2 + 1$ elements) and $\sum_{\omega'} \sum_{\eta'} p_x(\omega'|\cdot) \cdot p_{\eta'} = 0$. Letting $\bar{p} = \max_{\omega'} \{\sum_{\eta'} p_x(\omega'|\cdot) p_{\eta'}\}$, $G(\omega, s) < \bar{p} \cdot [\hat{V}(\omega + k_1|\cdot) - \hat{V}(\omega - k_2|\cdot)] = \bar{p} \cdot \varepsilon_\omega \downarrow 0$ as $\omega \to \pm\infty$. Define $\underline{\omega}(s) := \min \{\omega | G(\omega, s) > c(\omega)\}$ and $\bar{\omega}(s) := \max \{\omega \geqq | G(\omega, s) > c(\omega)\}$. Clearly $\underline{\omega}(s)$ and $\bar{\omega}(s)$ are finite, for otherwise V cannot remain bounded. Further, $x(\omega, s) = 0$ for $(\omega, s) \in C \equiv C_l \cup C_u$, where $C_l \equiv \{(\omega, s) | \omega < \underline{\omega}(s)\}$ and $C_u \equiv \{(\omega, s) | \omega > \bar{\omega}(s)\}$.

For a finite termination policy we need to show that, for each s, there exists an $\underline{\underline{\omega}}(s) > -\infty$ such that $V(\omega, s) = \phi$ for all $\omega \leqq \underline{\underline{\omega}}(s)$. When $x(\omega, s) = 0$,

$$V(\omega, s) = A(\omega, s) + \beta(1 - \textstyle\sum_{-k_2}^{-1} p_\eta) Q_{\omega 0}(s, \cdot) V(\omega, \cdot) + \beta \sum_{-k_2}^{-1} p_\eta Q_{\omega\eta}(s, \cdot) V(\omega + \eta, \cdot), \qquad \text{(A3)}$$

where $Q_{\omega\eta}(s, \cdot)$ is the s-th row of the finite-dimensional stochastic matrix representing $q_\omega(\hat{s}'|s, \eta)$ and $V(\omega, \cdot)$ is the column vector of firm values at ω for each s. Let $\hat{\omega}(s) = \min \{\omega | A(\omega, s) \geqq (1 - \beta)\phi\}$ and $\omega^* = \max_\omega [\{\omega < \hat{\omega}(s) | \forall s, x(\omega, s) = 0\}] > -\infty$, as there are only finitely many $s \in S$ (A.7.b). Then for all $\omega \leqq \omega^*$ equation (A3) holds, so we can write in matrix notation

$$V_\omega = A_\omega + \beta Q_{\omega 0} \cdot V_\omega - \beta \textstyle\sum_\eta p_\eta \cdot Q_{\omega\eta} \cdot \Delta_\eta V_\omega \qquad \text{(A4)}$$

where $Q_{\omega\eta}$ is the stochastic transition matrix for each exogenous shock η, V_ω is the vector of values at ω, and $\Delta_\eta V_\omega \equiv V_\omega - V_{\omega-\eta}$. Solving (A4) we get

$$[I - \beta Q_{\omega 0}] \cdot V_\omega = A_\omega - \beta \cdot \textstyle\sum_\eta p_\eta \cdot Q_{\omega\eta} \cdot \Delta_\eta V_\omega.$$

$$
\begin{aligned}
\boldsymbol{\phi} \leqq V_\omega &= [I - \beta Q_{\omega 0}]^{-1} A_\omega - \beta [I - \beta Q_{\omega 0}]^{-1} \textstyle\sum_\eta p_\eta \cdot Q_{\omega\eta} \cdot \Delta_\eta V_\omega\\
&\leqq [I - \beta Q_{\omega 0}]^{-1} (1 - \beta) \boldsymbol{\phi} - \beta [I - \beta Q_{\omega 0}]^{-1} \textstyle\sum_\eta p_\eta \cdot Q_{\omega\eta} \cdot \Delta_\eta V_\omega \leqq \boldsymbol{\phi} \qquad \text{(A5)}
\end{aligned}
$$

where $\boldsymbol{\phi}$ is a column vector with ϕ for each structure s. Hence $\forall s \in S$, $-\infty < \omega \leqq \omega^*$, $V(\omega, s) = \phi$. For each s, let $\underline{\underline{\omega}}(s) = \max \{\omega | V(\omega, s) = \phi\}$, and let $L = \{(\omega, s) | V(\omega, s) = \phi\}$. Then $\chi(\omega, s) = 0$ on L and $\chi(\omega, s) = 1$ elsewhere.

(c) The proof follows from demonstrating that all states $(\omega, s) \notin L$ are transient and hence will never be reached after some finite (random) time, while all $(\omega, s) \in L$ are recurrent, indeed absorbing, i.e. Prob $\{\exists \tau > t | (\omega_t, s_t) \in L, (\omega_\tau, s_\tau) \notin L\} = 0$. Since the probability of reaching L in finite time is strictly positive, Doob (1953), Chapter V.3, implies the existence of the a.s. finite stopping time $T(\omega_0, s_0)$.

The stochastic monotonicity of $T(\cdot)$ is shown by a coupling argument. Consider $\omega_2 > \omega_1$ and initial states (ω_2, s_0) and (ω_1, s_0). Denote (for this argument only) the underlying measure space by $\{U, \Sigma, P\}$ with elements u. Let $\omega_t^i(u)$ be the sample path arising from initial ω_i at $u \in U$. For each $u \in U$ define the stopping time $\tau(u) = \min \{t \geqq 0 | \omega_t^1(u) = \omega_t^2(u)\}$, and the new sequence

$$\omega_t^*(u)=\begin{cases}\omega_t^2(u) & \text{if } \omega_t^2(u)-\omega_t^1(u)>0 \quad \text{for all } t\geqq\tau(u),\\ \omega_t^1(u) & \text{otherwise.}\end{cases}$$

Note that: (a) the random sequence $\{\omega_t^*, s_t\}\geqq\{\omega_t^1, s_t\}$ with probability one; (b) as the random sequence $\{\omega_t^i, s_t\}$ is a Markov process and a stopping time is Markov, the distribution of $\{\omega_t^*, s_t\}$ is the same as that of $\{\omega_t^2, s_t\}$. Property (a), the monotonicity of $V(\cdot)$, and the stopping rule imply that $T(\omega_t^*, s_t)\geqq T(\omega_t^1, s_t)$ a.s. Property (b) and the continuous mapping theorem (Billingsley (1968)) imply that the distribution of $T(\omega_t^2, s_t)$ is the same as that of $T(\omega_t^*, s_t)$. Since the latter stochastically dominates $T(\omega_t^1, s_t)$, the proof is complete. ||

Proof of Proposition 2. Writing s_n for $s_n(\omega_0)$, letting $P^t_{(\omega_0,s_n)}((\omega, s)|\{x^*, \chi^*\})$ be the probability of reaching (ω, s) in t steps from (ω_0, s_n) under optimal investment and shutdown policies $\{x^*, \chi^*\}$, and $I_L(\cdot)$ be the indicator function of the shutdown states,

$$\begin{aligned}\phi\leqq V(\omega_0, s_n)&=\textstyle\sum_{t=0}^{\infty}\beta^t\sum_{\omega=0}^{K}\sum_s (R(\omega, s; x(\omega, s))[1-I_L(\omega, s)]+\phi I_L(\omega, s))P^t_{(\omega_0,s_n)}((\omega, s)|\{x^*, \chi^*\})\\ &\leqq\textstyle\sum_{t=0}^{\infty}\beta^t\sum_{\omega=0}^{K}\sum_s (A(\omega, s)[1-I_L(\omega, s)]+\phi I_L(\omega, s))P^t_{(\omega_0,s_n)}((\omega, s)|\{x^*, \chi^*\})\\ &\leqq\textstyle\sum_{t=0}^{\infty}\beta^t\sum_{\omega=0}^{K}\sum_s [A(\omega, s)\vee(1-\beta)\phi\,]P^t_{(\omega_0,s_n)}((\omega, s)|\{x^*, \chi^*\})\end{aligned}$$

where the first inequality is due to ignoring the cost of the optimal investment generating the transition probabilities, and the second from using $(1-\beta)\phi$ in place of $A(\omega, s)$ whenever it is larger. Let $p_\omega(s_n, t, e_t)$ be the probability that a firm starting at (ω_0, s_n) will have $\omega_t\geqq\omega$, conditional on a particular t-period sequence, e_t, of realizations of the exogenous process and the decision structure $\{x^*, \chi^*\}$. By (A3),

$$A(\omega, s)\leqq A_\omega(n)\equiv\sup_{\{s|\Sigma_{\omega^*\geqq\omega}s_{\omega^*}\geqq n\}}A(\omega, s)=(1-\beta)\phi+\theta(n)$$

for $s\in S_n(\omega)$, where $\theta(n)$ is monotone-decreasing to zero in its argument. Hence, for any of the n firms starting at ω_0, we can write

$$\begin{aligned}&\phi\leqq V(\omega_0, s_n)\\ &\leqq\textstyle\sum_{t=0}^{\infty}\beta^t\left[\sum_{e_t}\sum_\omega p_\omega(s_n, t, e_t)\sum_{k=0}^{n-1}A_\omega(k+1)\binom{n-1}{k}[p_\omega(\cdot)]^k[1-p_\omega(\cdot)]^{n-k-1}P(e_t)\right],\\ &\leqq\phi+\textstyle\sum_{t=0}^{\infty}\beta^t\left[\sum_{e_t}\sum_\omega p_\omega(\cdot)\sum_{k=0}^{n-1}\theta(k+1)\binom{n-1}{k}[p_\omega(\cdot)]^k[1-p_\omega(\cdot)]^{n-k-1}P(e_t)\right]\end{aligned}\tag{A6}$$

where $\binom{n}{k}$ is the number of k-combinations of n objects, and $P(e_t)$ is the probability of the realization, e_t, of the exogenous process. Let $f(n, t)$ be the function in the large square brackets in (A6). Clearly $f(n, t)\leqq\bar{A}$ and $\sum_{t=0}^{\infty}\beta^t\cdot\bar{A}=(1-\beta)^{-1}\bar{A}<\infty$. Hence, by the Lebesgue Dominated Convergence Theorem for sums, it suffices to show that $\forall t$, $\lim_{n\to\infty}f(n, t)=0$, for which it further suffices that $\forall\omega$,

$$p_\omega(\cdot)\textstyle\sum_{k=0}^{n-1}\theta(k+1)\binom{n-1}{k}[p_\omega(\cdot)]^k[1-p_\omega(\cdot)]^{n-k-1}\xrightarrow[n\to\infty]{}0\tag{A7}$$

for a.e. e_t. Now note that $(k+1)/(n-1)=\text{argmax}_p\,\{p^{k+1}(1-p)^{n-k-1}\}$. Thus, $\forall N\leqq n-1$,

$$\begin{aligned}&\textstyle\sum_{k=0}^{n-1}\theta(k+1)\binom{n-1}{k}[p_\omega(\cdot)]^{k+1}[1-p_\omega(\cdot)]^{n-k-1}\\ &\leqq\theta\cdot\textstyle\sum_{k=0}^{N-1}\binom{n-1}{k}\left(\frac{k+1}{n-1}\right)^{k+1}\left(\frac{n-k-2}{n-1}\right)^{n-k-1}+\theta(N+1)\\ &=\theta\cdot\textstyle\sum_{k=0}^{N-1}\frac{(n-1)\dots(n-k)}{(n-1)^{k+1}}\cdot\frac{(k+1)^{k+1}}{k!}\cdot\left(\frac{n-k-2}{n-1}\right)^{n-k-1}+\theta(N+1)\\ &\leqq\theta\cdot\left(\textstyle\sum_{k=0}^{N-1}\frac{(k+1)^{k+1}}{k!}\right)\cdot\frac{1}{n-1}+\theta(N+1).\end{aligned}$$

Now fix $\varepsilon>0$ and let n_1 be the minimum n such that $\theta(n+1)\leqq\varepsilon/2$ and n_2 be the smallest $n\geqq n_1-1$ such that

$$\theta\cdot\textstyle\sum_{k=0}^{n_1-1}\frac{(k+1)^{k+1}}{k!}\cdot\frac{1}{n_2-1}\leqq\frac{\varepsilon}{2}.$$

Hence (A7) holds, so that for $n\geqq n_2$, $V(\omega_0, s_n)\leqq\varepsilon$ as required. ||

Proof of Corollary 1. We use Proposition 2 to show that for any s, $\exists M<\infty$ such that $\forall m \geqq M$, $V^e(s, m) < \varepsilon \leqq x^e_M$. We do so here only for $\Omega^e = \{\omega^0 - k_2, \ldots, \omega^0\}$: all entry occurs at a single $\omega^0 + \eta$; the general case with distributed entry is an immediate consequence. Then $V^e(s, m) = \beta \sum_\eta \{\sum_{s'} V(\omega^0 + \eta, m \cdot e_{\omega^0+\eta} + s') q^0(s'|s, \eta)\} p_\eta$, and, for each of a finite number of η's and each s', $V(\cdot) \leqq \phi + \varepsilon$ for $m \geqq M$ by Proposition 2, giving the desired result. ||

Proof of Corollary 2. For the industry to remain finite we must show that all entry ceases with a sufficiently large number of firms. In Lemma 1 we show that there can never be more than a finite number of firms at any $\omega \in \{1, K\}$ without their all desiring to exit the industry immediately. Letting N_ω be that number for each ω, $N = \sum_{\omega=1}^K N_\omega$ is so large that for $\forall s \in \hat{S}_n(1)$, $\forall n \geqq N$, $V^e(s, m) < x^e_1$, $\forall m \geqq 1$.

Lemma 1. *For each* ω, $\exists N_\omega$ *such that* $\forall n \geqq N_\omega$, $V(\omega, s + n \cdot e_\omega) = \phi$.

Proof. Proposition 2 gives an n_ω such that for all $n \geqq n_\omega$ $V(\omega, s + n \cdot e_\omega) < \phi + \varepsilon$. We now show that by increasing n sufficiently we can drive the continuation value, $V^c(\cdot)$ [the expression within braces in equation (3)], below ϕ. At each ω there are two cases to consider: (a) $x(\omega \cdot) = 0$ and (b) $x(\omega, \cdot) > 0$.

Case (a): $x(\omega, \cdot) = 0$ implies $V^c(\omega, s + n \cdot e_\omega) \leqq A(\omega, s + n \cdot e_\omega) + \beta \cdot V(\omega, s + n \cdot e_\omega)$. But Assumption (A.3) says that $\exists n^*_\omega$ such that $A(\omega, s + n^*_\omega \cdot e_\omega) < (1-\beta)\phi - \varepsilon$. Hence, for $n \geqq \min\{n_\omega, n^*_\omega\}$, $V^c(\omega, s + n \cdot e_\omega) < (1-\beta)\phi - \varepsilon + \beta\phi + \beta\varepsilon < \phi$.

Case (b): $x(\omega, \cdot) > 0$ implies that there is a positive probability of advancing to any $\omega' \in \{\omega - k_2, \ldots, \omega + k_1\}$. Hence there exists an n^* such that with probability $1 - \varepsilon_1$ there are at least $n_{\omega+k_1}$ firms at $\omega + k_1$. Therefore, letting $\bar{V} = \sup_\Sigma V(\omega, s)$,

$$\begin{aligned} V^c(\omega, s + n^* e_\omega) &\leqq A(\omega, s + n^* e_\omega) + \beta(1-\varepsilon_1)(\phi + \varepsilon) + \beta\varepsilon_1 \bar{V} \\ &\leqq (1-\beta)\phi - \varepsilon + \beta(1-\varepsilon_1)(\phi+\varepsilon) + \beta\varepsilon_1\bar{V} \\ &< \phi - \varepsilon - \beta\varepsilon_1\phi + \beta\varepsilon_1\bar{V} + \beta(1-\varepsilon_1)\varepsilon = \phi + (\bar{V} - \phi)\beta\varepsilon_1 - (1 - \beta(1-\varepsilon_1))\varepsilon, \end{aligned}$$

where ε comes from (A3) as in case (a) [$n^* \geqq n^*_\omega$]. Hence we need only choose $n > n^*$ so large that $(\hat{V} - \phi)\beta\varepsilon_1 < (1 - \beta(1-\varepsilon_1))\varepsilon$. ||

Proof of Proposition 3. See Ericson–Pakes (1992). ||

Proof of Theorem 1. For existence we need to show the mutual consistency of four fundamental mappings: (i) $V: \Omega \times S \to [\phi, \bar{V}] \subset \mathbb{R}$; (ii) $x: \Omega \times S \to [0, \bar{x}] \subset \mathbb{R}_+$; (iii) $\mathscr{L}: S \to \Delta^S$; and (iv) $V^e: \hat{M} \times S \to [\phi, \hat{V}] \subset \mathbb{R}$; where $\hat{M} \equiv \{0, 1, \ldots, M\}$ is the set of numbers of potential entrants, Δ^S is the set of probability measures with support in the finite set S, i.e. a simplex of dimension $|S| - 1$, and $\mathscr{L}$ is the conditional probability distribution generated by the Markov transition kernel, Q. Given Q characterizing the behaviour of the industry structure, s (6.c), individual firm optimization generates an x (6.b) which solves equation (6.a) yielding both an optimal valuation of ω-states and industry structures, V, and optimal exit from that structure. V together with Q then generate the value of entering the industry, V^e (4), that determines the number of new entrants, $m(s)$ (6.d). The optimal investment, exit, and entry decisions of firms in turn define (see (A4), (A5)) a transition probability function, $\mathscr{L}$, for the industry structure through equations (6.c) and (7). An equilibrium will exist iff the resulting $\mathscr{L}$ is the same as that which determined the optimal valuation and investment functions of firms in the industry. We use a fixed-point argument to show that there exists such a $\mathscr{L}$, and hence appropriate Q, V, x, and V^e functions also exist, all satisfying the required properties (6.a–d).

Each of these mappings, V, x, $\mathscr{P}$, V^e, can be represented by a point in a compact subset of real Euclidean space: $V \in [\phi, \bar{V}]^{\Omega \times S}$, $x \in [0, \bar{x}]^{\Omega \times S}$, $\mathscr{L} \in (\Delta^S)^S$, and $V^e \in [\phi, \bar{V}]^{\hat{M} \times S}$. Define a mapping $\zeta: (\Delta^S)^S \to [0, \bar{x}]^{\Omega \times S} \times [\phi, \bar{V}]^{\Omega \times S} \times [\phi, \bar{V}]^{\hat{M} \times S}$, which takes a market structure transition function into an optimal investment policy and optimal valuation function for any firm in the industry, and an optimal valuation for any firm considering entry. It is generated by the solution to the Bellman equation (3) for a given transition probability function for industry structures and by equation (4). Define ψ. $[0, \bar{x}]^{\Omega \times S} \times [\phi, \bar{V}]^{\Omega \times S} \times [\phi, \bar{V}]^{\hat{M} \times S} \to (\Delta^S)^S$, a mapping which takes an optimal investment policy and state and entry valuations into a market-structure transition function. It is determined by equations (6.c) and (7). Finally, define the mapping $\mathscr{V}: (\Delta^S)^S \to (\Delta^S)^S$ by the composition $\mathscr{V} = \psi \circ \zeta$.

Lemma 2. ζ *is a continuous function.*

Proof. As argued in Ericson–Pakes (1992), ζ is continuous as the composition and product of continuous functions. ||

Lemma 3. *ψ is a continuous function.*

Proof. See Ericson-Pakes (1992) for the proof that ψ is a continuous function, as the composition of continuous functions. ||

Lemma 4. *There exists $\mathscr{L}^* \in (\Delta^S)^S$ such that $\mathscr{L}^* = \mathscr{V}(\mathscr{L}^*)$.*

Proof. The function $\mathscr{V}$ is continuous as the composition of two continuous functions, and $(\Delta^S)^S$ is clearly convex and is compact, so the result follows from Brouwer's Fixed Point Theorem. ||

Thus there exists a $\mathscr{L}$ such that the V and x functions satisfying equations (6.a) and (6.b) generate the transition kernel Q satisfying equations (6.c) and (7). The remaining condition (6.d) is an immediate consequence of preceding Propositions, while (6.e) is an arbitrary initial condition. ||

Proof Theorem 2. (a) This is an immediate consequence of Proposition 3 when we define the elements of the matrix Q to be given by the equilibrium transition kernel: $Q(s, s') \equiv Q(s'|s)$, incorporating optimal exit and entry, as well as investment, decisions. The Kolmogorov consistency theorem insures that the measure P_s is uniquely given by:

$$P_s\{\xi_t = s_t \text{ for } t=0, 1, \ldots, \tau\} = e_s \cdot \prod_{t=0}^{\tau-1} Q(s_t, s_{t+1}). \tag{A8}$$

(b) The existence of a unique positive recurrent communicating class will be shown through a series of lemmata.

Lemma 5. *There exists a positive recurrent communicating class, $R \subset S$.*

Proof. This is immediate as S is compact (finite). ||

Lemma 6. *There exists an $\bar{s}$ such that, $\forall s \in S$, $s \to \bar{s}$, i.e. $\exists n_s \geqq 1$ such that $P_s\{\xi_{n_s} = \bar{s}\} > 0$.*

Proof. Let $\bar{s} \equiv (0, \ldots, 0, N, 0, \ldots, 0)$ where $N > 0$ is a finite number of firms at $\omega^0 = \min \Omega^e$. We will show in two stages that there exists a finite trajectory, $\{s_0, s_1, \ldots, s_T\}$, with positive P_s-probability such that $s_0 = s$ and $s_T = \bar{s}$.

(i) For all s let s' be defined as follows: $s'_K = 0$, $s'_\omega = s_{\omega+1}$ for all $\omega \neq \omega^0$, $\omega \geqq \underline{\omega}(s)$, and $s'_{\omega^0} = s_{\omega^0+1} + m(s)$. Thus competition of all firms outside the industry inexorably advances, while the investments of all active firms fail to yield any success. Then (see Assumptions (A.4) and (A.5))

$$Q(s, s') = p_{-1} \cdot \prod_{\omega \geqq \underline{\omega}(s)} [\pi(\omega|\omega, x(\omega, s))]^{|s_\omega|} \cdot P(\omega^0)^{|m(s)|} > 0,$$

as must be any finite product of these transition and entry probabilities. Repeat until all active firms have dropped (at some τ_1) to ω^0 or lower:

$$s_{\tau_1} = (n_0, n_1, \ldots, n_{\omega^0}, 0, \ldots, 0\}.$$

This occurs in finite time as the initial industry structure is finite (Corollary 2).

(ii) For all $s \in \{s | s_\omega = 0 \ \forall \omega > \omega^0\}$ let s' be defined as follows: $s'_\omega = 0$, $\omega > \omega^0$; $s'_{\omega^0} = s_{\omega^0} + m(s)$; $s'_{\omega^0-1} = 0$; $s'_\omega = s_{\omega+1}$, $\omega < \omega^0 - 1$. Again outside competition advances, while all active inside firms, except those at ω^0, fail to generate any success with their investment. Firms at ω^0 succeed in holding their own. Again such a transition has strictly positive probability:

$$Q(s, s') = p_{-1} \cdot [\pi(\omega^0 + 1|\omega^0, x(\omega^0, s))]^{|s_{\omega^0}|}$$
$$\cdot \prod_{\omega \in W} [\pi(\omega|\omega, x(\omega, s))]^{|s_\omega|} \cdot P(\omega^0)^{|m(s)|} > 0,$$

where $W = \{\omega \in \Omega | \underline{\omega}(s) \leqq \omega < \omega^0\}$. Repeat until all firms below ω^0 have exited the industry. Again finiteness of the industry insures that this will occur in finite time τ_2. This yields, at $T = \tau_1 + \tau_2$, $\bar{s} \equiv s_T = (0, \ldots, 0, N, 0, \ldots, 0)$, where $N = \text{argmin}_n \{m(0, \ldots, 0, n, 0, \ldots, 0) = 0\} = n_{\omega^0} + \sum_{t=\tau_1+1}^{\tau_2} m(s_t)$. ||

Lemma 7. *$\bar{s} \in R$.*

Proof. This is immediate as R is positive recurrent. ||

Lemma 8. *Let $\hat{s}\in S$ be any recurrent state. Then $\hat{s}\in R$, R is the* only *recurrent class, and $s\notin R$ implies that s is transient.*

Proof. That $\hat{s}\in R$ must hold follows from Theorem 1.55, Freedman (1983). Hence R is unique and any $s\notin R$ must be transient. ||

(c) This is an immediate consequence of the existence of a single positive recurrent class: see Freeman (1983), Theorems 1.81, 1.88.

(d) $\mu_n = vQ^n$ (13) and hence converges iff the matrix Q^n does so. By Freedman (1983, Theorem 1.68), $\lim_{n\to\infty} Q^n(s, s') = 0$ if s' is transient ($s'\notin R$), and by Theorem 1.69(c), if s' is recurrent ($s'\in R$) then

$$\lim_{n\to\infty} Q^n(s, s') = \frac{\varphi Q(s, s')}{mQ(s', s')}$$

where $\varphi Q(s, s') \equiv P_s\{\xi_n = s' \text{ for some } n \geqq 0\}$, P_s is defined in (A8), and $mQ(s', s')$ is defined above. Notice that, for all n, vQ^n is a probability measure. Hence μ_n converges to some probability measure, $\lim \mu_n = \pi$ (say). Now notice that $\pi Q = (\lim vQ^n)Q = v \lim Q^n \cdot Q = v \lim Q^n = \lim vQ^n = \pi$ so that π is an invariant probability measure for Q. Howevert, by part (c) above, μ^* is the only (unique!) invariant probability measure, and therefore $\pi = \mu^*$. ||

Proof of Corollary 3. That $\mu^* Q = \mu^*$ was shown in Theorem 2. That P_{μ^*} (A8) is stationary is an immediate consequence of the fact. Let μ_t be the t-th period distribution starting from μ^*:

$$\mu_t = \mu_{t-1}Q = \cdots = \mu^* Q^t = \mu^* Q^{t-1} = \cdots = \mu^* Q = \mu^*. \qquad \|$$

Acknowledgements. This research was supported by the NSF (grant SES88-21722). Earlier versions have been presented in seminars at NBER, Ohio State, Brown, Columbia and Yale Universities. We would like to thank the editor and referees for numerous valuable comments on earlier versions, and Paul McGuire for invaluable programming and computational assistance. The usual caveat applies.

REFERENCES

BEATH, J., KATSOULACOS, Y. and ULPH, D. (1987), "Sequential Product Innovation and Industry Evolution", *Economic Journal, Conference Papers*, **97**, 32–43.

BEGGS, A. and KLEMPERER, P. (1992), "Multi-Period Competition with Switching Costs", *Econometrica*, **60**, 651–666.

BERRY, S. and PAKES, A. (1993), "Some Applications and Limitations of Recent Advances in Empirical Industrial Organization: Merger Analysis", *American Economic Review, Papers and Proceedings*, **83**, 247–252.

BROCK, W. A. and EVANS, D. S. (1985) *Federal Regulation of Small Business* (New York: Holmes and Meier).

BUDD, C., HARRIS, C. J. and VICKERS, J. S. (1993), "A Model of the Evolution of Duopoly: Does the Asymmetry between Firms tend to Increase or Decrease?", *Review of Economic Studies*, **60**, 543–573.

CABRAL, L. and RIORDAN, M. (1992), "The Learning Curve, Market Dominance and Predatory Pricing" (ISP Discussion Paper No. 39, Boston University, December).

DAVIS, S. and HALTWANGER, J. (1992), "Gross Job Creation, Gross Job Destruction, and Employment Reallocation", *Quarterly Journal of Economics*, **107**, 819–864.

DIXIT, A. (1989), "Hysteresis, Import Penetration and Exchange rate Pass-Through", *Quarterly Journal of Economics*, **104**, 205–228.

DOOB, J. L. (1953) *Stochastic Processes* (New York: John Wiley & Sons).

DUNNE, T., ROBERTS, M. and SAMUELSON, L. (1988), "Dynamic Patterns of Firm Entry, Exit, and Growth", *Rand Journal of Economics*, **19**, 495–515.

DUTTA, P. K., LACH, S. and RUSTICHINI, A. (1991), "Innovation and Product Differentiation" (manuscript, Columbia University, November).

DYNKIN, E. B. and YUSHKEVICH, A. A. (1975) *Controllable Markov Processes and Applications* (Moscow: Nauka). (In Russian.)

ERICSON, R. E. and PAKES, A. (1992), "An Alternative Theory of Firm and Industry Dynamics" (Cowles Foundation Discussion Paper No. 1041, Yale University, December 1992).

ERICSON, R. E. and PRESMAN, E. L. (1987), "A Stochastic Model of Optimal Investment in New Technology under Competition", in V. I. Arkin and P. K. Katyshev (eds.) *Mathematical Modeling of Control Processes under Uncertainty* (Moscow, USSR: CEMI) 89–100. (In Russian.)

EVANS, D. S. (1987*a*), "The Relationship between Firm Size, Growth and Age: Estimates for 100 Manufacturing Industries", *Journal of Industrial Economics*, **35**, 567–581.

EVANS, D. S. (1987*b*), "Tests of Alternative series of Firm Growth", *Journal of Political Economy*, **94**, 657-674.

FREEDMAN, D. (1983) *Markov Chains* (New York: Springer-Verlag).

FRIEDMAN, A. (1970) *Introduction to Ergodic Theory* (New York: Van Nostrand Reinhold).

GEWECKE, J. (1985), "Macroeconomic Modeling and the Theory of the Representative Agent", *American Economic Review*, **75**, 206-210.

GOWRISANKARAN, G. (1994), "Endogenous Horizontal Mergers in Infinitely Repeated Markov-Perfect Games" (manuscript, Yale University, June).

GOWRISANKARAN, G. and TOWN, R. (1994), "Dynamic Equilibrium in the Hospital Industry" (manuscript, Yale University, June).

HALL, B. (1987), "The Relationship between Firm Size, Growth and Age: Estimates for 100 Manufacturing Industries", *Journal of Industrial Economics*, **35**, 583-606.

HALMOS, P. (1956) *Ergodic Theory* (New York: Chelsea).

HOPENHAYN, H. A. (1992), "Entry, Exit, and Firm Dynamics in the Long Run", *Econometrica*, **60**, 1127-1150.

HOPENHAYN, H. A. and ROGERSON, R. (1993), "Job Turnover and Policy Analysis: A General Equilibrium Analysis", *Journal of Political Economy*, **101**, 915-938.

HARRIS, C. J. (1988), "Dynamic Competition for Market Share: An Undiscounted Model" (Nuffield College Discussion Paper No. 30, Oxford University).

JOVANOVIC, B. (1982), "Selection and the Evolution of Industry", *Econometrica*, **50**, 649-670.

KIHLSTROM, R. E. and LAFFONT, J. J. (1979), "A General Equilibrium Theory of Firm Formation Based on Risk Aversion", *Journal of Political Economy*, **87**, 19-48.

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

LIPPMAN, S. A. and RUMMELT, R. P. (1982), "Uncertain Imitability: An Analysis of Interfirm Differences Under Competition", *Bell Journal of Economics*, **13**, 418-438.

LUCAS, R. E. (1978), "On the Size Distribution of Business Firms", *Bell Journal of Economics*, **9**, 508-523.

MANKIW, N. G. and WHINSTON, M. (1986), "Free Entry and Social Inefficiencies", *Rand Journal of Economics*, **17**, 49-58.

MASKIN, E. and TIROLE, J. (1988), "A Theory of Dynamic Oligopoly: I & II", *Econometrica*, **56**, 549-600.

MASKIN, E. and TIROLE, J. (1993), "Markov Perfect Equilibria" (manuscript, Harvard University, November).

MUELLER, D. (1986) *Profits in the Long Run* (Cambridge, UK: Cambridge University Press).

NELSON, R. R. (1961), "Uncertainty, Learning and the Economics of Parallel R&D", *Review of Economics and Statistics*, **43**, 351-364.

OLLEY, S. and PAKES, A. (1991), "The Dynamics of Productivity in the Telecommunications Equipment Industry" (manuscript, Yale University, January).

PAKES, A. and ERICSON, R. E. (1990), "Empirical Implications of Alternative Models of Firm Dynamics" (manuscript, Yale University, February).

PAKES, A. and MCGUIRE, P. (1994), "Computing Markov Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model", *Rand Journal of Economics* (forthcoming).

ROSENTHAL, R. W. and SPADY, R. H. (1989), "Duopoly with both Ruin and Entry", *Canadian Journal of Economics*, **22**, 834-851.

STOKEY, N. L. and LUCAS, R. E. (with E. C. Prescott) (1989) *Recursive Methods in Economic Dynamics* (Cambridge, MA: Harvard University Press).

THOMAS, L. G. (1990), "Regulation and Firm Size: FDA Impacts on Innovation", *Rand Journal of Economics*, **21**, 497-518.

VICKERS, J. S. (1986), "The Evolution of Market Structure when there is a Sequence of Innovations", *Journal of Industrial Economics*, **35**, 1-12.