Competitive Diffusion

Jovanovic, Boyan and Glenn M. MacDonald

Working Paper No. 253
October 1990

Competitive Diffusion

Boyan Jovanovic and Glenn M. MacDonald

Rochester Center for Economic Research
Working Paper No. 253

October 1990

COMPETITIVE DIFFUSION†

Boyan Jovanovic* and Glenn M. MacDonald**

University of Rochester

Working Paper No. 253

Revised
October 1990

†This research was supported by the C.V. Starr Center for Applied Economics and the National Science Foundation (Jovanovic), and the Lynde & Harry Bradley Foundation (MacDonald). For thoughtful comments on an earlier version of this paper, the authors would like to thank Rob Porter and the referees; Yaw Nyarko and Phil Reny also deserve special mention. Paul Gomme performed the calculations in Sec. 3(c).

*Department of Economics, New York University

**Centre for Decision Sciences and Econometrics, University of Western Ontario; Center for Economic Research, University of Rochester; and Economics Research Center/NORC, University of Chicago

ABSTRACT

The paper models a competitive industry in which both the discovery and spread of new technological know-how are endogenous. Allowing for aggregate shocks and informational linkages, general results are provided concerning the evolution and long run behaviour of the distributions of knowledge and other observables; these include firm output, growth and rank within the industry, and the time path of the product price and industry output. Further, equilibrium does not coincide with a social optimum, with both innovative and imitation activities possibly being too low from the planner's perspective. Several special cases are analysed, one yielding S-shaped diffusion of new technology; this is used to organize data on the spread of diesel locomotion in the U.S. Railroad industry.

1. INTRODUCTION

This paper models a competitive industry in which the discovery of new technology and its spread are endogenous. The model is motivated by two kinds of evidence. The first is the observation (Solow, 1957) that the better part of growth in economic activity cannot be explained by increasing quantities of factors producing output subject to a fixed technology. It has been argued widely that a theory of the development and spread of new technology is needed to resolve this anomally; see Solow (1988), and for a contrasting view, Becker (1988). The second type of evidence is in Figure 1. The top panel displays the fraction of shipments of bits of dynamic random access memory (DRAM) by chip density and over time. Low density (1 kilobyte) chips are displaced by those with higher density (4k), which are then overtaken by those with yet greater density (16, then 64 and then 256 k). In the meantime, the total quantity of bits delivered explodes, and price falls dramatically. The bit industry displays the waves of change and improvement stressed by Schumpeter (1934, 1939). Such data call for a theory in which new developments occur periodically and do not spread instantaneously.

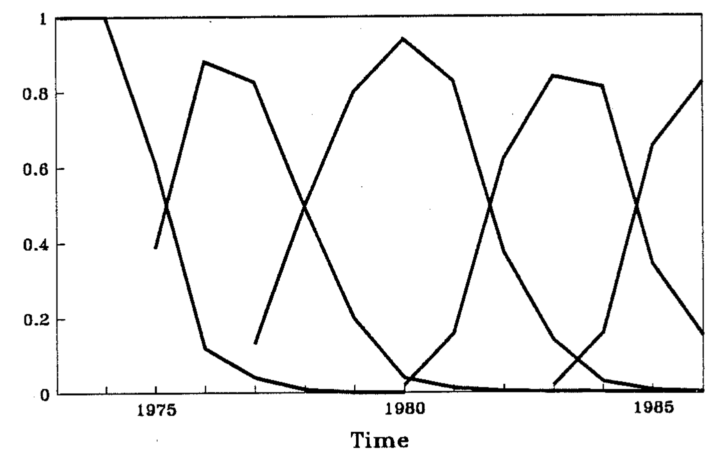
The theory has three main primitives: product demand, cost, and a learning technology; all three are allowed to vary with calendar time and on exogenous "aggregate" shocks. In addition, however, the learning technology can depend on the current (cross-firm) distribution of technological knowledge -- this will allow informational linkages to affect the spread of knowhow.[1]

Section 2 sets out the model, in which there is a fixed population of measure-zero firms. Theorem 1 asserts the existence of a unique, symmetric equilibrium in Markov strategies. Next, Lemma 2 shows that the distribution of knowledge over firms improves over time, occurring because each firm's knowledge either moves forward or stands still, but does not regress or depreciate. Theorem 2 then states that information gathering must eventually

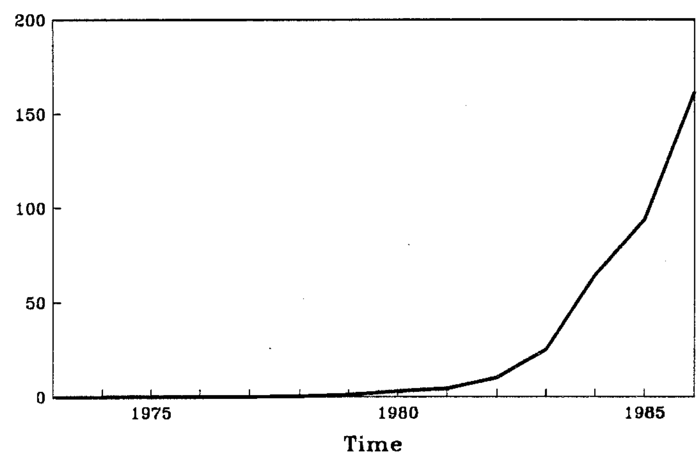
[1]The other primitives are initial conditions and a transition law for aggregate shocks, and an initial distribution of knowledge.

Figure 1

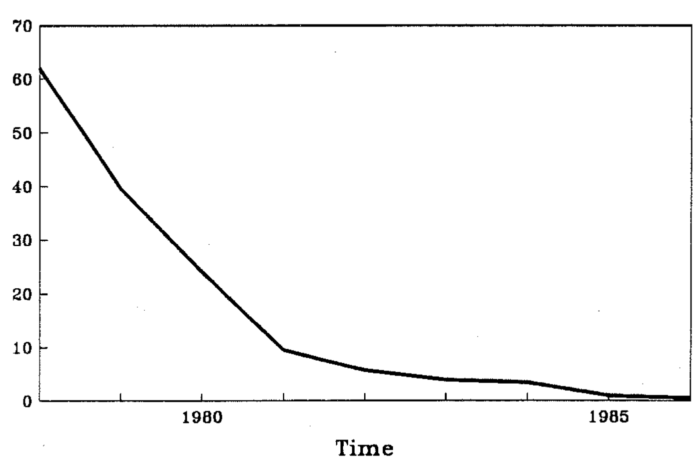
Diffusion of Density in the
DRAM Industry



Quantity of Bits
(Billions)



Price per Bit
(Real 1978 Millicents)



Source: *Icecap Report*

cease-- the distribution of knowledge over firms approaches a limiting long-run distribution.[2]

Describing this distribution is easiest when the environment is stationary. For that case Theorem 3 shows that the set of equilibrium long-run distributions of knowledge has a simple representation. With this result in hand, Theorem 4 asks under what conditions the population of firms will remain heterogeneous forever. The answer is that heterogeneity will not survive forever if a firm can easily acquire better technology through either innovation or imitation. If innovation is easy, all firms will end up discovering improved technology for themselves; if imitation is cheap, followers will employ it to catch up.[3]

Knowledge is typically not observed, but some actions are. Under mild conditions, the distribution of actions will have positive variance if and only if the distribution of knowledge does; this correspondence carries over to the limiting distributions as well. In fact, the distribution of output over firms does not collapse as an industry matures. This was found when Stigler's "survivor test" was applied to over a hundred industries (Stigler 1958, Saving 1961). Evidently, in most industries the best technology is hard to come by through invention, and the best "state of the art" technique is hard to imitate. Were this not so, firm sizes would differ less than they do.

Theorem 6 states that when there is no learning by doing, the monotonic improvement in the distribution of knowledge will lead to a monotonic decline in price and a monotonic increase in industry output, holding constant exogenous shocks to demand and cost. This prediction squares with Gort and Klepper's (1982) findings about the time path of industry price and quantity for over twenty products for which data are available since birth.

Theorems 8 and 9 deal with two fundamental properties of the evolution of firm size. First, the corrolary to Theorem 8 asserts that firms having superior knowledge today are less

---

[2]This is one way to interpret the result (Scherer 1980, p. 56, tables 3.3) that among the top 100 corporations in the U.S., the amount of turnover seems to have decreased over the past century. The correspondence with the model is weak, however, because these corporations' activities are spread over many markets and are continually expanding into new ones.

[3]Empirically, the time-path of inequality as industries mature has received little attention. One exception is Hart and Prais (1956). In a sample of all companies quoted on the London Stock Exchange from the mid-nineteenth to the mid-twentieth century, they found that inequality in valuations of firms first increased and then decreased over the hundred-year period.

likely to learn anything new. If the firm's output decision is increasing in its knowledge, this also means that while firms that are large today will probably be large tomorrow, they are less likely to grow much. That is, there is positive persistence in firm size in the *levels*. On the other hand, Theorem 9 asserts that smaller firms will grow faster than larger firms. In a similar vein, Theorem 10 shows that the expected gain in rank is larger for smaller firms.[4]

Finally, Theorem 11 states that relative to what a surplus-maximizing social planner would prefer, equilibrium information-gathering activities (including imitation) are too low in that the planner's payoff would be raised by a one-shot increase in these activities. This result may surprise those who think, as did Schumpeter, that the best way to encourage invention is to *limit* imitation through, say, a patent system. A planner who can control invention and imitative effort directly can do better by exploiting the fact that current imitation makes imitation easier later.

Section 3 specializes the model to permit only two kinds of information gathering: innovation and imitation. Innovation yields a firm new technological know-how as a function of its present knowledge and effort; imitation yields the firm improvements as a function of its effort and others' knowledge. Attention then shifts to (i) a numerical example that illustrates and supplements the theorems; and (ii) a specific case focusing on the diffusion of a single new technology, heterogeneity in firm outputs, and the time paths of industry aggregate R & D expenditures. Both (i) and (ii) model imitation like a generalized contagion: the extent of imitation depends positively on the number of firms that seek to imitate a new technology and how hard they try, and on the number of firms that already have it. When imitation is easy, S-shaped diffusion of technology is the likely outcome, for roughly the same reasons as S-shaped diffusion of a disease occurs in contagion models. Indeed, such diffusion will explain the wave-like behavior depicted in the top panel of figure 1. Finally, (iii), the specific

---

[4]These predictions also match some data: In levels, firm output exhibits positive serial correlation, (Gort 1963) but growth rates of smaller firms tend to exceed those of larger firms (Evans, 1987, Hall 1987). This is so even after controlling for sample selection, which is not an issue in the model because it has no entry or exit. Entry and Exit are studied in Jovanovic and MacDonald (1990). Note that not all samples show a tendency for small firms to grow faster -- see Singh and Whittington (1975).

model is put to use organizing data on the diffusion of diesel technology in the U.S. railroad industry.

Section 4 contains concluding remarks; proofs are in the Appendix.

2. GENERAL MODEL

The situation is as follows. At any date, the behavior of consumers implies demand for a homogeneous good. A group of firms may supply this product, behaving as perfect competitors in product and factor markets. A firm's technological knowledge may change in response to what the firm does, to what its competitors know, and to what is known more generally (in other industries, science, etc.). Thus, given what it knows and how easily it can use various sources of new information, the firm must allocate resources to production and information-gathering activities. This allocation problem is the center of attention in what follows. What determines this choice? And what about the time-series and cross-section properties of the distribution of knowledge across firms, and of related entities such as the distribution of output over firms?

a) Structure

Time, t, is discrete and the horizon is infinite: $t \in T \equiv \{0,1,...\}$.[5]

---

[5]The following notation will be used. First, whenever time is an argument of a function, or some element of a set is to be interpreted as being associated with period t, t will appear as a subscript.

Next, for any real valued function f with domain Y,

$$\|f\| \equiv \sup_{y \in Y} |f(\cdot)|.$$

When some arguments of f are to be held fixed, they will be displayed as arguments. That is, write $y = (y^0, y^1)$, then for fixed y^0:

$$\|f(y^0, \cdot)\| \equiv \sup_{(y^0, y^1) \in Y} |f(y^0, \cdot)|.$$

Finally, the statement "f is Lipschitz in y^0 with Lipschitz constant k" will mean: $\forall\, y^0, y^{0'}$, $\|f(y^{0'}, \cdot) - f(y^0, \cdot)\| \leq k\|y^{0'} - y^0\|$.

Outside-the-industry entities follow a Markov process $\{X_t\}$ with compact state space $K \subset \mathbb{R}^k$, $k < \infty$. A realization of X_t, denoted by x_t, is observed by all agents before they act at t, and comprises exogenous factors influencing product demand, firm costs, or possibilities for learning new production technologies.

Consumer behavior is summarized by the continuous inverse market demand function $D{:}T \times [0,\bar{Q}] \times K \rightarrow [\underline{D},\bar{D}]$, where $0 < \underline{D} \leq \bar{D} < \infty$ and $\bar{Q} < \infty$. $Q \in [0,\bar{Q}]$ represents total industry output of the commodity, assumed nonstorable. The function D is nonincreasing and Lipschitz in Q with Lipschitz constant d. The vector x is an argument of D that allows for economy-wide shocks, changes in tastes, endowments or incomes of consumers and variation in the prices of other goods; in particular, declining demand due to introduction of new products elsewhere may be included. In general D will depend nontrivially on only a subvector of x.

Define a firm-specific state variable $\theta \in \Theta$, where $\Theta \equiv [0,\bar{\theta}] \subset \mathbb{R}$. Let $\mathscr{B}$ be the class of Borel sets in Θ, with typical element b, and $\mathscr{M}$ the set of probability measures on $\mathscr{B}$ with generic element m, endowed with the topology of weak convergence (Billingsley (1968)).[6]

The description of firms, strategies and payoff to an arbitrary strategy now follows. At any t, θ indexes technological know-how, and firms differ only to the extent that θ varies among them.[7] Any heterogeneity at t is captured by $\nu_t \in \mathscr{M}$; ν_0 is exogenous.

In general, a strategy would be a function specifying a feasible action for each opportunity at which the firm might act. These opportunities would be indexed by the prior history of the game. Here it is assumed that the vector $s_t \equiv (t, \theta_t, x_t, \nu_t)$ suffices as a description of history. This implies that the firm cannot condition its behavior on any

---

[6]The sup norm $\|\cdot\|$ defined in footnote 3, as applied to $\mathscr{M}$, does *not* coincide with the Prohorov metric (Billingsley). This distinction arises only in the proof of Theorem 1, in which all topologies are taken to be compatible with $\|\cdot\|$.

It will prove convenient to order probability measures in $\mathscr{M}$ according to whether the associated distribution functions may be ranked by first order stochastic dominance. To that end, let $\succ$ be the reflexive partial order on $\mathscr{M} \times \mathscr{M}$ defined by $m \succ m' \Leftrightarrow \forall\, \theta, m([0,\theta]) \leq m'([0,\theta])$.

[7]Since θ is a scalar, costly-to-adjust capital inputs are ignored here. Also, the model may be interpreted in terms of *product* instead of *process* innovation in a manner similar to Spence (1984), fn. 2.

elements of history apart from s_t, so that, the "state space" is simply $S \equiv T \times \Theta \times K \times \mathcal{M}$, with generic element s.

A firm's action at t is a vector $\alpha_t = (\alpha_{it})_{i=1}^{i=l}$, also written (q, α_{-q}). q represents the firm's output and α_{-q} is interpreted as comprising the levels of other activities influencing the firm's technological know-how: $R\&D$, measures of the extent of attempts to learn from others, and so on. It is assumed that $(q, \alpha_{-q}) \in [0,\bar{Q}] \times A_{-q} \equiv A$, where A_{-q} is a convex, compact subset of $\mathbb{R}^{l-1}$, $2 \leq l < \infty$. Let $\underset{\sim}{0}_q \equiv (q,0,...,0) \in A$ and $\underset{\sim}{0} \equiv (0,...,0)$, with length clear from the context.

A strategy for a firm is, therefore, a function a: $S \to A$.

How does the firm evaluate a given strategy? First, consider the return a firm would earn from an arbitrary sequence of actions, $\{\alpha_t\}$, product prices $\{p_t\}$ and states $\{s_t\}$. It is assumed that this return is the present value of profits:

$$\sum_{t=0}^{\infty} \beta^t [p_t q_t - c_t(\alpha_t, \theta_t, x_t)]. \tag{1}$$

In (1), $\beta \in (0,1)$ is a discount factor and c: $T \times A \times \Theta \times K \to \mathbb{R}_+$ is the cost of taking action α at t given θ and x. The function c is i) continuous for each t; ii) strictly decreasing and Lipschitz in θ if $q \neq 0$; iii) strictly increasing and convex in α, and equal to zero if $\alpha = \underset{\sim}{0}$; and iv) continuously differentiable in α with the gradient $\nabla_\alpha c$ satisfying a) $\forall x$ and for some open interval $I \subset \Theta$, $q = 0$ and $\theta \in I$ imply $\partial c(\alpha,\theta,x)/\partial q < \bar{D}$; and b) $\forall(\alpha,\alpha')$, $\|\nabla_\alpha c(\alpha,\cdot) - \nabla_\alpha c(\alpha',\cdot)\| \geq c_{\alpha\alpha}\|\alpha-\alpha'\|$, $c_{\alpha\alpha} \in (0, \infty)$.

This structure guarantees that (i) the firm's maximal action is unique and solves first order conditions; (ii) a higher θ corresponds to a better technology; (iii) there are some θ for which production must occur in any equilibrium; and (iv) actions are not overly responsive to the state variables. Only those components of x *not* referring to technological knowledge are allowed to enter c nontrivially, so that θ represents fully the firm's knowledge at t. Any other contributions to the firm's knowledge--for example, those having origins outside the industry--are embedded in the transition function for θ given below.[8]

---

[8]This specification permits technology-specific factors provided there are no costs of adjustment and the evolution of the prices of factors is independent of the extent of their use in this industry, in which case these prices may be subsumed in x.

The product price (p_t) and state vector (s_t) evolve as follows. First, suppose there is a function P: $T \times K \times \mathcal{M} \to [\underline{D}, \bar{D}]$ such that

$$p_t = P_t(x_t, v_t). \tag{2}$$

Next, assume there is a function Φ: $T \times K \times \mathcal{M} \to \mathcal{M}$, and that

$$v_{t+1} = \Phi_t(x_t, v_t), \tag{3}$$

with v_0 given. Since Φ depends on t, any sequence $\{v_t\}$ can be represented this way. Consistency of P and Φ with the equilibrium behavior of firms is required as part of the definition of equilibrium below. Finally, θ_t follows a Markov process on Θ, with transition function Ψ: $\mathcal{B} \times T \times A \times \Theta \times K \times \mathcal{M} \to [0,1]$, written $\Psi_t(b,\alpha,\theta,x,v)$. Note that given α and s, $\Psi \in \mathcal{M}$. Ψ is assumed to satisfy i) $\forall s$ with $\theta < \bar{\theta}$, $\max_{\alpha} \Psi_t[(\theta,\bar{\theta}], \alpha, \theta, x\ v] \in (0,1)$; ii) If $\theta \in (0,\bar{\theta}]$, $\forall s_t$, $\forall \alpha$, $\Psi_t[[0,\theta), \alpha, \theta, x, v] = 0$; iii) If $\theta \geq \theta'$, $\alpha \geq \alpha'$ and $v \succ v'$, then $\forall (x, t)$, $\Psi_t(\cdot,\alpha,\theta,x,v) \succ \Psi_t(\cdot,\alpha',\theta',x,v')$; iv) $\forall (t, b, \alpha)$, Ψ is Lipschitz in θ, x and v with Lipschitz constants ψ_θ, ψ_x and ψ_v; v) $\forall (b,s)$ Ψ is differentiable with respect to α, with the gradient $\nabla_\alpha \Psi$ a) uniformly bounded by ψ_α, and b) Lipschitz in α, θ, x and v with Lipschitz constants $\psi_{\alpha\alpha}$, $\psi_{\alpha\theta}$, $\psi_{\alpha x}$ and $\psi_{\alpha v}$; vi) $\forall (\theta', s)$, $\Psi_t[[0,\theta'], \alpha, \cdot]$ is convex in α; and vii) Ψ is continuous in v.

Conditions (i) and (ii) require first that if there is anything to learn, learning is not impossible but cannot be guaranteed, and then that the firm not implement any technology having index lower than that which it currently knows. (Given (iii) and that c is never increasing in θ, the latter is included solely as a notational convenience.) (iii) imposes that learning is never impeded by increases in either θ or α, or by other firms knowing more.

Given $s_0 = (\theta_0, x_0, v_0)$, Ψ, χ, P and Φ, an arbitrary strategy a induces a probability measure over sequences $\{s_t\}_1^\infty$. Let $E_0(\cdot \mid \theta_0, x_0, v_0, a)$ be the associated conditional expectation operator. Then, for any (θ_0, x_0, v_0), the firm's evaluation of the arbitrary strategy a is the expected discounted value of profits

$$E_0\left\{\sum_{t=0}^{\infty} \beta^t \Big[P_t(x_t,v_t)q_t(\theta_t,x_t,v_t) - c_t[a_t(\theta_t,x_t,v_t),\theta_t,x_t]\Big] \mid \theta_0, x_0, v_0, a\right\}. \tag{4}$$

Below, the definition of equilibrium will require that the firm select a strategy that maximizes (4). Since (1) is bounded by $\bar{D}\bar{Q}/(1-\beta)$, a maximal policy, written $\alpha_t(\theta,x,v)$, obeys:

$$\forall s,\ \alpha_t(\theta,x,\nu) \in \underset{\alpha\in A}{argmax} \left\{P_t(x,\nu)q - c_t(\alpha,\theta,x) + \beta \int V_{t+1}[\theta',x',\Phi_t(x,\nu)]\Psi_t(d\theta',\alpha,\theta,x,\nu)\chi_t(dx',x)\right\}, \quad (5)$$

where V: $S \to \mathbb{R}$ satisfies:

$$\forall s,\ V_t(\theta,x,\nu) = \max_{\alpha\in A} \left\{P_t(x,\nu)q - c_t(\alpha,\theta,x) + \beta \int V_{t+1}[\theta',x',\Phi_t(x,\nu)]\Psi_t(d\theta',\alpha,\theta,x,\nu)\chi_t(dx',x)\right\}. \quad (6)$$

b) **Equilibrium**

Two sets of conditions hold in equilibrium. First, given P and Φ, firms are to select a maximal strategy. Secondly, P and Φ should agree with the behavior of firms as a group. Formally, an equilibrium is a set of functions $\{V,\alpha,P,\Phi\}$ such that (5) and (6) hold, and, $\forall(t,x,\nu)$ both

$$P_t(x,\nu) = D_t\left[\int q_t(\theta,x,\nu)\nu(d\theta),x\right] \quad (7)$$

and

$$\Phi_t(x,\nu) = \int \Psi_t[\cdot,\alpha_t(\theta,x,\nu),\theta,x,\nu]\nu(d\theta). \quad (8)$$

Conditions (5) and (6) ensure the maximality of each firm's strategy; (7) imposes product market clearing; and (8) requires the distribution of technology in use at $t+1$ to agree with that in use at t coupled with what was learned during that period.

This definition demands that equilibrium be *symmetric*: All firms must select the same strategy. Their actions at t will, of course, generally differ when ν_t is not a point mass.[9]

[9]The imposition of symmetry is particularly convenient if ν_t has atoms. It does not yield nonexistence of equilibrium because there are no fixed costs; compare Mas Colell (1984). Next, by substitution of (7) and (8) into (6), the model can be represented as a normal form game with the set of players described by ν_0, the strategy space being functions from S to A, and payoffs obtained from (5) and (6). Since each firm's policy is required to be maximal given any $s \in S$, the equilibrium strategies form a perfect equilibrium in the derived game. This game has much in common with an anonymous discounted sequential game; see Jovanovic and Rosenthal (1988). The difference is that, unlike Jovanovic and Rosenthal, the present structure allows for the "aggregate shocks" X_t but places additional structure on payoffs. Existence of equilibrium in a general anonymous sequential game with aggregate shocks has been demonstrated by Bergin and Bernhardt (1989). The additional structure imposed here yields existence of a *unique symmetric* equilibrium.

Also, if D, Ψ, χ and c do not depend on t, the environment will be referred to as *stationary*. When this is the case, V, Φ, P, and α do not depend on t and the subscript t will be dropped.

c) **Remarks**

The model has many interpretations. First, suppose that at t all firms freely learn others' technological know-how and may implement at $t+1$ or later what they have learned. In this case the motivation for *R&D* is purely the return to utilizing superior knowledge one period earlier than others do. Empirically, the length of the period would be the duration of effective protection (patent, or other) new knowledge enjoys. Second, suppose knowledge can be sold and transferred costlessly, say through a simple announcement of ingredients. Then the equilibrium price of information is zero, and given that other firms sell what they know, no firm is harmed by doing likewise, in which case announcement by all is part of an equilibrium strategy. Third, a patent system might prohibit a firm from using a technology unless it has invented it itself, perhaps simultaneously with other firms. Under this specification Ψ is such that $\theta_{t+1} = \theta_t$ unless *R&D* yields the firm a technology characterized by θ outside the support of ν_t. Finally, inventions arriving from outside the industry can be included by structuring Ψ so that, for particular x, learning certain θ requires minimal effort; for example, $K = \Theta \times K' \subset \mathbb{R}^{k-1}$ could be assumed, and the first component of x interpreted as knowledge available to any firm at low cost.

d) **Results**

d.1. Technical Properties

The first proposition provides a condition sufficient to guarantee existence of an unique equilibrium.

Theorem 1: *Assume* $max(d,\psi_\alpha) \leq \left[(1\text{-}\beta)c_{\alpha\alpha} - \psi_{\alpha\alpha}\bar{D}\bar{Q}\left[1+\frac{\beta\psi_\alpha^2\psi_\nu}{\psi_{\alpha\alpha}(1-\psi_\nu)(1-\beta)}\right]\right] \times \left[1-\beta+\beta\psi_\alpha\bar{Q}\right]^{-1}$.

Then, an equilibrium exists and is unique.

Observe that the parameter restriction in Theorem 1 is always met if $c_{\alpha\alpha}$ is sufficiently large, or d and ψ_α sufficiently small.[10]

Lemma 1: *In any equilibrium (i) V, α, P and Φ are continuous, and ii) V is increasing in θ.*

d.2. The Sequence of Distributions of Technologies in Use

The first result states that the distribution of technology increases over time in the sense of first order stochastic dominance.

Lemma 2: $\forall t, x$ *and* ν, $\Phi_t(x,\nu) \succ \nu$.

Lemma 2 follows from condition (ii) on Ψ, which requires any firm not to implement any technology inferior to the one it currently uses.

To proceed, let $\tilde{x} \equiv \{x_t\}_0^\infty$ be an arbitrary sequence. The equilibrium law of motion Φ then generates a deterministic sequence $\{\nu_t\}_0^\infty$ via (3). Theorems 2-4 focus on the properties of this sequence. Theorem 2 states that from any initial distribution of know-how and for any evolution of aggregate shocks, $\{\nu_t\}$ converges.

Theorem 2: *For each* $(\nu_0,\tilde{x})$, *there exists* $\nu^* \in \mathcal{M}$ *such that* $\nu_t \to \nu^*$.

Intuitively, Theorem 2 holds because the distribution of knowledge is always improving (Lemma 2) but cannot advance beyond the distribution under which all firms have technology $\bar{\theta}$. Thus knowledge cannot advance indefinitely.

---

[10]This restriction guarantees satisfaction of the hypotheses of the Contraction Mapping Fixed Point Theorem (Kolmogorov and Fomin, p. 66), which yields uniqueness along with existence, and will be assumed in all that follows. Strictly, the contraction mapping argument gives existence and uniqueness of Φ and P. Existence and uniqueness of V and α then follow from standard arguments (similar to those exposited in Stokey and Lucas (Ch. 8)) utilizing properties of the operator defined by the right hand side of (6).

In all that follows, stationarity will be assumed. While many of the propositions have nonstationary analogues, they are much easier to state and prove under stationarity. Theorem 2 states that, for large t, ν_t is close to ν^*, and may even equal ν^*. Thus further properties of $\{\nu_t\}$ may be obtained from those of ν^*. Results on ν^* follow.

When the environment is stationary, Φ does not depend on t. Let

$$\mathcal{M}^* \equiv \bigcap_{x \in K} \{\nu \in \mathcal{M} | \Phi(x,\nu) = \nu\}$$

be the set of measures such that $\nu_t \in \mathcal{M}^*$ implies $\nu_{t'} = \nu_t$ for all $t' \geq t$, for all x. $\mathcal{M}^*$ is nonempty since it contains ν satisfying $\nu(\{\theta\}) = 1$.

Given x_0, let $\tilde{\chi}$ be the probability measure on $\mathcal{B}^K \times \mathcal{B}^K \times ...$ consistent with χ (Ash, Sec. 2.7). That is, $\tilde{\chi}$ is a probability measure over sequences whose first element is x_0. Also, let $\chi^\tau(\cdot\,|x_t)$ be the τ-fold iterate of χ; i.e. $\forall b \in \mathcal{B}^K$, $\chi^\tau(b|x_t) = Pr(X_{t+\tau} \in b|x_t)$. Finally let K' be a countable dense subset of K.

Theorem 3: Assume that for each (ν_0, x_0):

(*i*) $\tilde{\chi}(\{\tilde{x}\,|\,\exists t \text{ such that } x_t \notin K'\}) = 0$;

and

(*ii*) *For each* $b \in \mathcal{B}^K$ *having nonempty interior,* $\exists\ \varepsilon > 0$, *and* $\hat{t} < \infty$ *such that*

$$\min_{\substack{x \in K \\ \tau \in \{1,...,\hat{t}\}}} \chi^\tau(b|x) \geq \varepsilon.$$

Then $\tilde{\chi}(\{\tilde{x}\,|\,\nu^* \notin \mathcal{M}^*\}) = 0$.

Informally, the argument underlying Theorem 3 is as follows. Theorem 2 indicates that the distribution of knowledge ν_t settles at ν^*. There are two ways ν_t can come to rest. One is that the coupling of firm actions and possibilities for learning are such that no learning will occur irrespective of aggregate shocks: $\nu^* \in \mathcal{M}^*$. The other way for ν_t to settle is that aggregate shocks can conspire to prevent learning even though learning would go on for certain other values of x_t; a trivial example is that in which $\nu_0 \notin \mathcal{M}^*$, learning is prohibitively costly if $x_t = x'$, and $\tilde{x} = \{x',x',...\}$. The proof shows that such conspiracy of events is "unlikely" and therefore that the set of sequences $\tilde{x}$ for which the associated ν^* is an element of $\mathcal{M}^*$ has full measure.

By asserting that $\nu^* \in \mathcal{M}^*$ almost surely, Theorem 3 raises the possibility of obtaining features of ν^* from those of $\mathcal{M}^*$. The next result displays a useful characterization of $\mathcal{M}^*$ when information acquisition requires some action beyond production of output.

Lemma 3: *Assume that* $\forall s, \forall q, \Psi(\{\theta\}, \underset{\sim}{0}_q, \theta, x, \nu) = 1$. *Then* $\mathcal{M}^* = \{\nu \in \mathcal{M} | \forall x, \forall \theta \in$ *supp* $\nu, \exists\, q$ *such that* $\alpha(\theta,x,\nu) = \underset{\sim}{0}_q\}$.

Since some effort is required if learning is to occur, without such effort there can be no learning, and no growth of knowledge. On the other hand, if the distribution of knowledge does not evolve, the set of θ's for which continued learning is maximal must have measure 0. Moreover, firms knowing such θ necessarily have technological know-how distinct from most others, for otherwise (since α is continuous) many others with similar knowledge would also find it to their advantage to learn as well, causing the distribution of knowledge to evolve.

Given Lemma 3 and Theorem 3, properties of ν^* can be obtained from the requirement that $\alpha(\theta, x, \nu^*) = \underset{\sim}{0}_q$ for θ in the support of ν^*. Theorem 4 first shows that in general $\nu_t \neq \nu^*$, so that ν^* approximates ν_t closely only for large t. Second, the Theorem establishes some properties of ν^*. In particular, sufficient conditions are given for (i) all possible knowledge eventually to be acquired by nearly all firms; (ii) some knowledge to remain undiscovered by many firms; and (iii) some knowledge to remain undiscovered and nontrivial heterogeneity in technological know-how to persist.

Theorem 4 uses the following notation. For given x and ν, let $\bar{P}(x,\nu)$ and $\bar{q}(\theta,x,\nu)$ solve

$$\bar{q}(\theta,x,\nu) = \underset{q}{argmax} \left\{ \bar{P}(x,\nu)q - c(\underset{\sim}{0}_q, \theta, x) \right\}$$

and

$$\bar{P}(x,\nu) = D[\int \bar{q}(\theta,x,\nu)\nu(d\theta), x].$$

Let $\bar{V}(\theta,x,\nu)$ be the expected present value of profits obtained by a firm selecting $\bar{q}(\theta,x,\nu)$ at each date when $\nu_t = \nu$ for all t.

Theorem 4: *Along with the conditions in Theorem 3, assume (i)* $\forall s, \forall q, \Psi(\{\theta\}, \underset{\sim}{0}_q, \theta, x, \nu) = 1$; *(ii)* $c(\alpha,\cdot) = c^0(q,\cdot) + c^1(\alpha_{-q},\cdot)$; *(iii)* $\alpha_{-q} = \alpha'_{-q}$ *implies* $\Psi(\cdot,\alpha,\cdot) = \Psi(\cdot,\alpha',\cdot)$; *and (iv)* (ν_0, x_0) *is such that* $\exists\theta \in$ *supp* ν_0 *such that some component of* $-\nabla_{\alpha_{-q}} c^1(\underset{\sim}{0},\theta,x_0) + \beta \int \bar{V}(\theta',x',\nu_0)\nabla_{\alpha_{-q}} \Psi(d\theta',\underset{\sim}{0},\theta,x_0,\nu_0)\chi(dx',x_0)$ *is strictly positive. Then, except for a set of measure 0 in* K^∞, *(a)* $\forall t, \nu_t \notin \mathscr{M}^*$; *(b)(i) if* $\forall\, \theta \in [0,\bar{\theta}), \forall\, (\nu, x)$, *in some component* $\nabla_{\alpha_{-q}} c^1(\underset{\sim}{0},\theta,x)$ *equals zero and* $\nabla_{\alpha_{-q}} \Psi[(\theta, \bar{\theta}], \underset{\sim}{0}_q, \theta, x, \nu]$ *is strictly positive, then* $\nu^*(\{\bar{\theta}\}) = 1$; *(b)(ii) if* $\forall\theta, \forall x, \nabla_{\alpha_{-q}} c^1(\underset{\sim}{0},\theta,x) > \underset{\sim}{0}$, *then* $\nu^*(\{\bar{\theta}\}) < 1$; *(b) (iii) assuming, in addition to the condition in (b)(ii), that* $\forall\alpha, \forall\, \theta \in$ *supp* $\nu_0, \forall\theta' \in(\theta,\bar{\theta})$, $\Psi[(\theta,\bar{\theta}], \alpha, \theta, x_0, \nu_0] > 0$ *implies* $\Psi[(\theta',\bar{\theta}], \alpha, \theta, x_0, \nu_0] > 0$, *then supp* ν^* *contains at least two points.*

Condition (i) states that learning requires effort. Conditions (ii) and (iii) separate production from information gathering.[11] The final condition is a restriction on primitives guaranteeing (by virtue of Lemma 3) $\nu_0 \notin \mathscr{M}^*$.

The argument underlying part (a) is as follows. If learning comes to a halt at date t', (i.e. $\nu_{t'} \in \mathscr{M}^*$) it is easy to show that efforts to learn would not pay at $t'-1$; this is largely because the distribution of others' knowledge is more helpful at t' than at $t'-1$. Thus learning must in fact halt at $t'-1$: $\nu_{t'-1} \in \mathscr{M}^*$. Repeating the argument yields no learning at $t=0$, contradicting the assumption $\nu_0 \notin \mathscr{M}^*$. Thus $\nu_{t'} \notin \mathscr{M}^*$.

Part (b) follows because while there must be some learning at every date, it must gradually vanish, so that at large t some firms must be on the margin about whether to try to learn. In part (i) this cannot happen unless all firms know $\bar{\theta}$; in part (ii) it cannot if all firms know $\bar{\theta}$. Part (iii) also precludes all firms knowing $\bar{\theta}$, but also requires learning to be varied enough so that for all firms to know some other θ cannot be the ultimate outcome either.

---

[11]These conditions do not rule out learning by doing. They merely insist that if production might generate new knowledge, some aspect of noting what the information is and putting it to use is both required and not literally costless. Imitation too is unlikely to be costless. Mansfield, Schwartz and Wagner (1981) find that imitating something is about 70% as costly as inventing it, and Evenson and Kislev (1973) find that to imitate successfully, the imitator must do some research himself.

d.3. Sequences of Distributions of Observables

Identifying technologies indexed by θ may be straightforward; the diesel/steam locomotive case studied below is an example. In general, however, focusing directly on technology may prove extremely difficult. Therefore, implications about observables such as output, R&D expenditures, or net revenue are now developed.

Let $f: A \times \Theta \times K \times \mathscr{M} \to \mathbb{R}^n$, $1 \leq n < \infty$, be continuous. (f might simply be $f = \alpha$, or given any P, $f = P(x,\nu)q - c(\alpha,\theta,x)$). Continuity of policies, and of continuous transformations of policies, follows by lemma 1.

Given equilibrium $\alpha(\cdot)$, define a probability measure ν^f on the class of Borel sets of $\mathbb{R}^n$, $\mathscr{B}^{\mathbb{R}^n}$, by: $\forall\, b \in \mathscr{B}^{\mathbb{R}^n}$, $x \in K$ and $\nu \in \mathscr{M}$, $\nu^f(b,x,\nu) = \nu\left[\{\theta \mid f[\alpha(\theta,x,\nu),\theta,x,\nu] \in b\}\right]$. Also, let $\nu^{f*}(\cdot,x) \equiv \nu^f(\cdot,x,\nu^*)$.

Theorem 5: $\forall\, x,\ \nu_t \to \nu^* \Longrightarrow \nu^f(\cdot,x,\nu_t) \to \nu^{f*}(\cdot,x)$.

While cumbersome, the dependence of ν^{f*} on x cannot in general be eliminated. An example may prove helpful. Suppose that x is a scalar entering the model only via D and that $f = q$. Then, even if growth in θ comes to a halt, the distribution of output will vary in response to current shocks to demand, and hence ν^{f*} will fluctuate endlessly.[12]

The limiting measure ν^{f*} is no more directly observable than is ν^*; interest attaches to ν^{f*} to the extent that it is informative about ν^f_t. A result analogous to Theorem 4 can be stated; it provides conditions under which heterogeneity in observed actions will exist for large t, and others under which homogeneous actions will prevail. For example, when whatever f describes is one to one with technological know how, the distribution of observables is degenerate if and only if the distribution of technological know how is.

---

[12]The stationary distribution of f is

$$\int_K \nu^f(\cdot,x,\nu^*)\hat{\chi}(dx),$$

where $\hat{\chi}(\cdot) = \int_K \chi(\cdot,x)\hat{\chi}(dx)$.

d.4. The Evolution of Price and Output

Lemma 2 states that the distribution of technology in use improves in the sense of first order stochastic dominance. Under the conditions below, this implies that for fixed x, price will decrease and industry output increase over time.

Theorem 6: Assume i) Ψ does not depend on q; and ii) $c(\alpha,\cdot) = c^0(q,\cdot) + c^1(\alpha_{-q},\cdot)$, where $\partial c^0/\partial q$ is decreasing in θ. Then $\forall\, x \in K$, $P[x,\Phi(x,\nu)] \leq P(x,\nu)$.

Corollary 1: $\forall\, x \in K$, $\int q[\theta,x,\Phi(x,\nu)]\, \Phi(x,\nu)(d\theta) \geq \int q(\theta,x,\nu)\nu(d\theta)$.

A related result removes the conditioning on x as follows. Each $\tilde{x}$ implies sequences $\{p_t\}$ and $\{Q_t\}$; thus $\tilde{\chi}$ induces a measure on the spaces of such sequences, and (unconditional) cumulative probability distributions F^p_t and F^Q_t on price and quantity at t.

Corollary 2: F^p_t (F^Q_t) is stochastically decreasing (increasing) in t provided x_t is i.i.d.

The next result focuses on the time path of the distribution of "normalized" output, again taking x as given. Assuming $q(\cdot) > 0$, let $\tilde{q}(\theta,x,\nu) \equiv q(\theta,x,\nu)/ \min_{\theta \in \, supp\, \nu} [q(\theta,x,\nu)]$.

Theorem 7: In addition to the hypotheses of Theorem 6, assume that $c^0(q,\theta,x) = \bar{c}^0(q)\hat{c}^0(\theta,x)$, where $\bar{c}^0$ is homogeneous of degree strictly larger than 1 and $\bar{c}^{0\prime}(0) = 0$. Then for fixed x, the distribution of $\tilde{q}$ increases over time in the sense of first order stochastic dominance.

The conditions in Theorem 7 ensure that the scaling factor in the denominator of $\tilde{q}$ is bounded away from zero, without which $\tilde{q}$ is undefined, and then impose that the relative outputs of firms knowing distinct θ be independent of the product price. If $\hat{c}^0(\theta,x) = \hat{c}^0_1(\theta)\hat{c}^0_2(x)$ is imposed, the conditioning on x may be dropped; if x_t is i.i.d., a statement analagous to Corollary 2 can be made.

d.5. Cross-Section Implications

(i) Learning

If the costs of, and prospects for, learning do not depend on the firm's knowledge or its rate of production, firms having higher θ learn less in the sense of first order stochastic dominance.

Define a probability measure Ψ^{Δ}: $\mathscr{B} \times \Theta \times K \times \mathscr{M} \longrightarrow [0,1]$ pointwise as follows: $\forall(b,\theta,x,v)$, $\Psi^{\Delta}(b,\theta,x,v) = \Psi[(b + \{\theta\}) \cap \Theta, \alpha(\theta,x,v),\theta,x,v]$; that is, Ψ^{Δ} is a probability measure on learning taking current θ as the reference point. Further, let Ψ^0: $A \times K \times \mathscr{M} \longrightarrow [0,1]$, Ψ^1: $\mathscr{B} \times K \times \mathscr{M} \longrightarrow [0,1]$ and, I: $\mathscr{B} \times \Theta \longrightarrow \{0,1\}$ where $I(b,\theta) = 1$ if and only if $\theta \in b$.

Theorem 8: Assume (i) $\forall(b,\alpha,\theta,x,v)$, $\Psi(b,\alpha,\theta,x,v) = \Psi^0(\alpha,x,v)\, \Psi^1[b \cap (\theta,\bar{\theta}],x,v] + I(b,\theta)\, \{\Psi^0(\alpha,x,v)\, \Psi^1([0,\theta], x,v) + [1\text{-}\Psi^0(\alpha,x,v)]\}$, *where* $\alpha_{-q} = \alpha'_{-q}$ *implies* $\Psi^0(\alpha,\cdot) = \Psi^0(\alpha',\cdot)$; *and (ii)* $\forall(\alpha,\theta,x)$, $c(\alpha,\theta,x) = c^0(q,\theta,x) + c^1(\alpha_{-q},x)$. *Then* $\forall\ (x,v)$, $\theta^1 > \theta^0 \Longrightarrow \Psi^{\Delta}(\cdot,\theta^0,x,v) \succ \Psi^{\Delta}(\cdot,\theta^1,x,v)$.

Corollary: $\theta^1 > \theta^0 \Longrightarrow \Psi[(\theta^0,\bar{\theta}], \alpha(\theta^0,x,v), \theta^0,x,v] \geq \Psi[(\theta^1,\bar{\theta}], \alpha(\theta^1,x,v), \theta^1,x,v]$.

The conditions in the Theorem have two effects. First, the separability restriction on Ψ makes learning a two-step process in which the likelihood of learning anything at all (i.e. Ψ^0) is influenced by α and θ, but given that some discovery has been made, just what is learned is independent of α and put to use if it exceeds the existing θ. The second effect is to separate production from learning. Under these conditions, learning a superior technology is no easier for firms that are already ahead and since there is less for them to learn, they invest fewer resources in learning, and so are less successful.

(ii) Growth and Firm Size

Given θ_t, the growth of firm output is a random variable $q[\theta_{t+1}, x_{t+1}, \Phi(x_t,v_t)]/q(\theta_t,x_t,v_t)$ with distribution function $G(\cdot,\theta_t,x,v)$.

Theorem 9: Under the hypotheses of Theorems 6-8, and if $\forall(\theta,x)$, $\hat{c}^0(\theta,x) = \hat{c}_1^0(\theta)\hat{c}_2^0(x)$, *then* $\forall(\theta,x,v)$, *G is decreasing in* θ *in the sense of first order stochastic dominance.*

Since $q(\cdot)$ is monotone in θ, the result may also be stated as conditional on q, x and v; that is, defined that way, G is stochastically decreasing in q.

(iii) Turnover in the Size Distribution of Firms

Define $d:\Theta \times K \times \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ pointwise by $\forall(\theta,x,v,v')$,

$$d(\theta,x,v,v') = v'([0,\theta])\Psi[\{\theta\},\alpha(\theta,x,v),\theta,x,v] + \int_{(\theta,\bar{\theta}]} v'([0,\theta'])\Psi[d\theta',\alpha(\theta,x,v), \theta,x,v]-v([0,\theta]).$$

$d(\theta,x,v_t,v_{t+1})$ is the expected change in rank between t and $t+1$ for firms currently knowing θ.

Theorem 10: Assume the conditions in Theorem 8, along with: $\forall(a,x,v)$ *such that* $\theta^0 \le \theta^1$ *implies* $\forall(x,v)$, $\Psi^0[a(\theta^0,x,v),x,v] \ge \Psi^0[a(\theta^1,x,v),x,v]$,

$$v((\theta^0,\theta^1]) \ge \Psi\,[[0,\theta^0], a(\theta^0,x,v),\theta^0,x,v] \times \int_{[0,\theta^1]} \Psi[(\theta^0,\theta^1], a(\theta,x,v), \theta,x,v]v(d\theta).$$

Then $d(\theta,x,v_t,v_{t+1})$ *is declining in* θ.

Whether a firm's rank advances or falls back depends on what those who know less learn, as well as learning by those knowing more. The (sufficient) condition in the Theorem limits how fast those who know less learn new technology.

d.6. Equilibrium and Surplus Maximization

The presence of v as an argument in Ψ represents an externality. Thus a "social optimum" is unlikely to coincide with equilibrium. While a social optimum is hard to analyse fully, the next result shows that there is a sense in which a social planner would prefer information gathering efforts beyond those taken in equilibrium.[13]

Given (x_0,ν_0), any strategy a, followed by all firms, generates a probability measure over sequences $\{x_t, \nu_t\}_1^\infty$. Let $\hat{E}(\cdot\,|x_0,\nu_0, a)$ denote the associated expectation operator, and $Q_t(x,\nu) = \int q_t(\theta,x,\nu)\nu(d\theta)$. Surplus generated by strategy a is

$$\hat{W}(a) \equiv \hat{E}\left\{\sum_{t=0}^{\infty} \beta^t \left[\int_0^{Q(x_t,\nu_t)} D(z,x_t)dz - \int c[a(\theta,x_t,\nu_t),\theta,x_t]\nu_t(d\theta)\right] \middle| x_0,\nu_0,a\right\}.$$

Assume $\hat{W}$ reflects the planner's preferences over common strategies a.

For fixed t', suppose that $\alpha_{t'} \in Int\, A$ on a set of θ's of positive $\nu_{t'}$ - measure, and let a' be a feasible strategy differing from α only at date t', the period t' difference being that the last l-1 components (i.e., other than q) of a' exceed those of α, for all θ for which $\alpha_{t'} \in Int\, A$. Thus a' involves a one time increase in pure information gathering activities for a positive measure of firms.

Theorem 11: *Assume that i)* $c(\alpha,\theta,x) = c^0(q,\cdot) + c^1(\alpha_{-q},\cdot)$, *where* $\alpha_i = 0$ *implies* $\partial c/\partial \alpha_i = 0$; *ii)* Ψ *is independent of* q; *iii)* $\forall(\alpha,\theta,\hat{\theta},x,x',\nu)$ *and any strictly increasing function* $\xi\colon \Theta \to \mathbb{R}$,

$$\frac{d}{d\varepsilon} \int \xi(\theta')\, \Psi\left[d\theta', \alpha, \theta, x', \int \Psi[\cdot, \varepsilon a^1 + (1\text{-}\varepsilon)a^2, \bar{\theta}, x, \nu]\nu(d\bar{\theta})\right] > 0$$

at $\varepsilon = 0$, *where* a^1 *and* a^2 *are feasible strategies such that* $a^1_{-q} \geq a^2_{-q}$; *and iv)* $\forall(\alpha,\theta,x,\nu,\nu')$,

$$\int \Psi([0,\theta'],\alpha,\bar{\theta},x,\varepsilon\nu + (1\text{-}\varepsilon)\nu')[\varepsilon\nu(d\bar{\theta}) + (1\text{-}\varepsilon)\nu'(d\bar{\theta})]$$

is differentiable with respect to ε, *at* ε=0. *Then, at* ε=0, *for any* t',

$$\frac{d}{d\varepsilon}\hat{W}[\varepsilon a' + (1\text{-}\varepsilon)\alpha] > 0.$$

[13]The social optimum for a specific case is given in Section 3.(c) below.

Corollary: *If Ψ is independent of v, equilibrium and the planner's optimum coincide.*

Condition (ii) and the separability in condition (i) eliminate learning-by-dong. This allows the planner to evaluate information-gathering activities without regard to their immediate impact on consumer's surplus. Condition (iii) guarantees that even small improvements in the distribution of knowledge are of some use in learning. The other restrictions validate the variational method of proof.

This result is usefully thought of in the innovation/imitation setting described below. There, innovation is an information gathering activity in which success depends on θ but is independent of v, while imitative success depends on v but not θ. Theorem 11 indicates that a planner would prefer that *both* innovative and imitative effort be raised at the same date relative to the equilibrium outcome. That imitative effort, in particular, is too low from the planner's viewpoint follows because in making their imitation decisions, competitive firms do not account for the fact that they too may be imitated later. Put differently, in comparison to equilibrium, the planner would prefer both that there be more new discoveries, and that they spread faster.

The result obviously does not say how optimum compares to equilibrium. However, it does suggest that surplus-maximization and active discouragement of imitation -- say by a patent system where patents are long-lived -- are likely inconsistent. Also, without further structure, it does not state that if the planner could make a small adjustment to equilibrium at some t, greater information gathering would be a *maximal* adjustment; it merely asserts that for the planner, such an adjustment is better than no adjustment at all; in particular equilibrium is not a social planner's optimum.

3. INNOVATION AND IMITATION

This Section distinguishes innovation from imitation. The force of this distinction is that innovative success depends on what the firm knows and on how hard it tries to learn, whereas imitative success responds *not* to what the firm knows, but rather to what *others* know together with how hard the firm tries to learn from them.

a) **More Structure**

First, the vector of actions α is restricted to length three and written $\alpha = (q,\eta,\mu)$, with $\eta \in [0,1]$ and $\mu \in [0,1]$. As before, q is output; η and μ stand for innovative and imitative effort.

Second, possibilities for learning--the probability measure Ψ--are required (along with the restrictions imposed earlier) to agree with the distribution function τ defined by:

$\forall\ (\theta', \alpha, \theta, x, \nu) \in \Theta \times A \times \Theta \times K \times \mathcal{M},$

$$\tau(\theta', \alpha, \theta, x, \nu) = \begin{cases} 0 & \theta' < \theta \\ [1-\eta + \eta\, N(\theta', \theta, x)]\ [1-\mu + \mu M(\theta', x, \nu)] & \theta \leq \theta'. \end{cases}$$

Here N is a distribution function on Θ given θ and x, representing innovation possibilities; M is also a distribution function on Θ, but takes ν and x as given, and $M(\cdot, x,\nu) \succ \nu$. η is the probability with which the firm gets a draw from N, the distribution of new technological know how; this draw is interpreted as an innovation. Similarly, with probability μ a draw from M is obtained, interpreted as the firm's learning the technique known by some other firm, say via reverse engineering. Given θ_t, θ_{t+1} is then the maximum of θ_t and what (if anything) is learned from innovative and imitative effort.[13]

Finally, $c(\alpha,\theta,x) = c^0(q,\theta,x) + c^1(\eta,\mu,x)$; that is, production and information gathering are separate activities, and θ influences production cost only. Given the previous restriction,

---

[13]The distribution relevant for imitation is taken to be M rather than $F(=\nu([0,\theta']))$ to capture the notion that the firm may direct its imitative efforts towards others it sees as likely to have superior technological knowledge. For instance, the decision about which firm to try to imitate could be based on a noisy signal of other firms' know how. This captures the intuition of Mansfield et al. (1977) and others who suggest that more significant technological advances would be imitated more quickly.

Also, since no individual action can influence aggregate behavior, no firm has an incentive to take any "evasive" action to thwart imitation. If the firm knows θ, there are always many firms that have this information as well. Evasive action by one firm is costly and does not lower the probability that θ will be learned by others from those that also know θ currently.

the correct interpretation of the absence of θ in c^1 is that higher θ does not make it any less difficult to fabricate some new technology or learn from others. Rather, θ operates by influencing the distribution of what innovations are discovered, via N, and the set of new ideas put to use.

To eliminate clutter, assume that i) any firm's policy involves $\alpha \in Int\, A$; ii) α and V are continuously differentiable with respect to θ; and iii) τ is continuously differentiable with respect to θ' for $\theta' \in (\theta, \bar{\theta}]$. It is trivial to provide conditions under which this will occur; let $n(\cdot, \theta, x)$ and $m(\cdot, v, x)$ denote the densities associated with N and M. First order conditions characterizing $\alpha(\cdot)$ are:

$$P(x,v) - \frac{\partial}{\partial q} c^0[\alpha(\theta,x,v), \theta, x] = 0, \tag{9a}$$

$$- \frac{\partial}{\partial \eta} c^1[\alpha(\theta,x,v),x] + \beta \int V[\theta',x',\Phi(x,v)] \frac{\partial}{\partial \eta} \Psi[d\theta',\alpha(\theta,x,v),\theta,x,v]\, \chi(dx',x) = 0 \tag{9b}$$

and

$$- \frac{\partial}{\partial \mu} c^1[\alpha(\theta,x,v),x] + \beta \int V[\theta',x',\Phi(x,v)] \frac{\partial}{\partial \mu} \Psi[d\theta',\alpha(\theta,x,v),\theta,x,v]\, \chi(dx',x) = 0. \tag{9c}$$

b) The Effect of Knowledge on Innovative and Imitative Activity

Given an additional restriction, imposed below, the structure set out in subsection a) satisfies the hypotheses of Theorem 8: firms with access to more technological know-how are less likely to learn more. The issue addressed here is whether anything be said about the underlying innovative and imitative behaviour that implements this fundamental outcome. This question is of interest because, for example, the empirical literature on R & D expenditures supposes some relatively simple relationship between effort devoted to R & D and other observables such as firm size.

While clear results are available under strong conditions, the central conclusion is that there are forces at work rendering it impossible to produce simple results more generally, either in cross-section or time-series experiments. The formal analysis follows.

Since $\eta(\theta,x,v) > 0$ and $\mu(\theta,x,v) > 0$ for some $\theta < \bar{\theta}$, and $\eta(\bar{\theta},x,v) = \mu(\bar{\theta},x,v) = 0$, η and μ must be declining in θ for some θ. The result to follow provides conditions under which η and μ are declining in θ throughout.[15]

Theorem 12: Assume i) $\forall\, x$, *$N(\cdot,\theta,x)$ does not depend on θ; and ii)* $\forall(\theta,x,v)$,

$$- \frac{\partial^2}{\partial\eta\partial\mu} c^1[\eta(\theta,x,v),\mu(\theta,x,v),x,] + \beta \int \Big\{ V[\theta,x',\Phi(x,v)][N(\theta',\theta,x) - 1][M(\theta',x,v) - 1]$$

$$+ \int_{(\theta,\bar{\theta}]} V[\theta',x',\Phi(x,v)] \Big[n(\theta',\theta,x)[M(\theta',x,v) - 1] + m(\theta',x,v)[N(\theta',\theta,x) - 1] \Big] d\theta' \Big\}$$

$$\chi(dx',x) \geq 0.$$

Then η and μ are declining in their first argument.

The first condition states that having greater technological know-how does not improve the distribution from which innovations are drawn. Obviously, significant violation of this restriction may allow η, in particular, to be non-monotone in θ. The second condition limits the size of "cross effects"; if reducing imitative activity raises the marginal return to innovation sufficiently, for example, both will not respond to greater θ in the same direction (although the degree to which greater innovative activity can compensate for reduced imitation, for example, is limited by Theorem 8.)

This cross effect is a basic source of ambiguity in results on η and μ. Greater imitation makes it less likely that whatever is learned from innovation will be put to use, and conversely. In this sense innovation and imitation are substitutable. At the same time, greater θ reduces the firm's incentive to try to learn by any means, and for this reason innovation and imitation are complementary. The interplay of these two forces works against clear-cut results on the policies consistent with the basic result in Theorem 8; the situation is even more

---

[15]The conditions are not stated solely in terms of primitives; however, doing so is neither difficult, nor, in this case, very helpful. That is, the restriction in question involves c^1, α, V and Φ. To state it in terms of primitives, index α, V and Φ by c^1, then restate the condition in terms of c^1 and the indexed functions.

difficult when N depends on θ nontrivially. The same kind of considerations interfere with time series results on how η and μ vary as $\{v_t\}$ unfolds. The theory is likely to be of much greater use in organizing evidence on the kind of result displayed in Theorems 8, 9 and 10.[16]

c) **Example**

The theory set out above contains a variety of testable restrictions, and in a specific example will now illustrate them.[17] Let

$$D(Q) = d_0 - d_1 Q,$$
$$\Theta = \{\theta_0, \theta_1, \theta_2\}, 0 < \theta_0 < \theta_1 < \theta_2,$$
$$c^0(q,\theta) + c^1(\eta,\mu) = c_q\, q^2/(2\theta) + c_\eta\, \eta^2/2 + c_\mu\, \mu^2/2,$$
$$M = F,$$
$$v_0(\{\theta_0\}) = 1,$$

and N be consistent with the markov transition matrix

$$\begin{array}{cc} & \theta_{t+1} \\ & \begin{array}{ccc} \theta_0 & \theta_1 & \theta_2 \end{array} \\ \theta_t \begin{array}{c} \theta_0 \\ \theta_1 \\ \theta_2 \end{array} & \left[\begin{array}{ccc} 1-\delta_{01}-\delta_{02} & \delta_{01} & \delta_{02} \\ 0 & 1-\delta_{12} & \delta_{12} \\ 0 & 0 & 1 \end{array}\right] \end{array} .$$

The parameters are d_0, d_1, c_q, c_η, c_μ, δ_{01}, δ_{02} and δ_{12}.

All firms start with technology θ_0. Given $\theta=\theta_0$, innovation yields $\theta=\theta_1$ (θ_2) with probability $\eta\delta_{01}$ ($\eta\delta_{02}$). If $\theta=\theta_1$, innovation produces $\theta=\theta_2$ with probability $\eta\delta_{12}$. (The case studied in subsection a) imposes $\delta_{02} = \delta_{12}$.) Imitation involves random sampling of firms,

---

[16]The evidence on the relationship between firm size and the likelihood of adoption of new technologies is very mixed; see Rose and Joskow (1988) and the references therein. Evidently, in terms of the present model, $N(\cdot,\theta,\cdot)$ depends on θ nontrivially in many cases.

[17]In all that follows, the aggregate shocks (X) will be suppressed since they play no role in the results derived.

yielding, at t, $\theta=\theta_1$ with probability $\mu\nu_t(\{\theta_1\})$ and $\theta=\theta_2$ with probability $\mu\nu_t(\{\theta_2\})$ regardless of θ. Let $\nu^i \equiv \nu(\{\theta_i\})$, $i \in \{0,1,2\}$.[17]

Table 1 lists the parameters values used in the example.

Table 1

Parameter Values

$\beta = .98$	$\theta_0 = 1$
$d_0 = 2$	$\theta_1 = 5$
$d_1 = 2.5$	$\theta_2 = 15$
$c_q = 1$	$\delta_{01} = .05$
$c_\eta = 1$	$\delta_{02} = .01$
$c_\mu = .666$	$\delta_{12} = .05$

The parameters chosen have no particular significance, except for $c_\mu/c_\eta = .66$, approximately in line with Mansfield et. al. (1981).

Figure 2 displays evolutions of the distribution of knowledge, innovative and imitative effort, and price. "Low" tech, "medium" tech and "high" tech refer to θ_0, θ_1 and θ_2. Appendix Section A.16 provides the underlying values of ν_t^i, $\eta(\theta_0,\nu_t)$, $\mu(\theta_0,\nu_t)$, $\eta(\theta_1,\nu_t)$, $\mu(\theta_1,\nu_t)$ and p_t. These figures, taken together with the parameters, yield the evolution of various objects of interest; for example, aggregate R&D expenditures, the distribution of output, etc.

Perhaps the most striking feature is the evolution of imitative effort. As was indicated above, the interaction of scale and substitution effects renders the time path of imitative behaviour a highly irregular one.

For comparison, Figure 3 displays corresponding entities in a full social optimum; i.e. a complete solution to the planner's problem. As might be anticipated, the optimum evolves very quickly relative to equilibrium, and resources devoted to information gathering vanish

---

[17]These parameters used below do not guarantee $\eta \in [0,1]$ and $\mu \in [0,1]$; however, all probabilities just mentioned are so restricted.

Figure 2

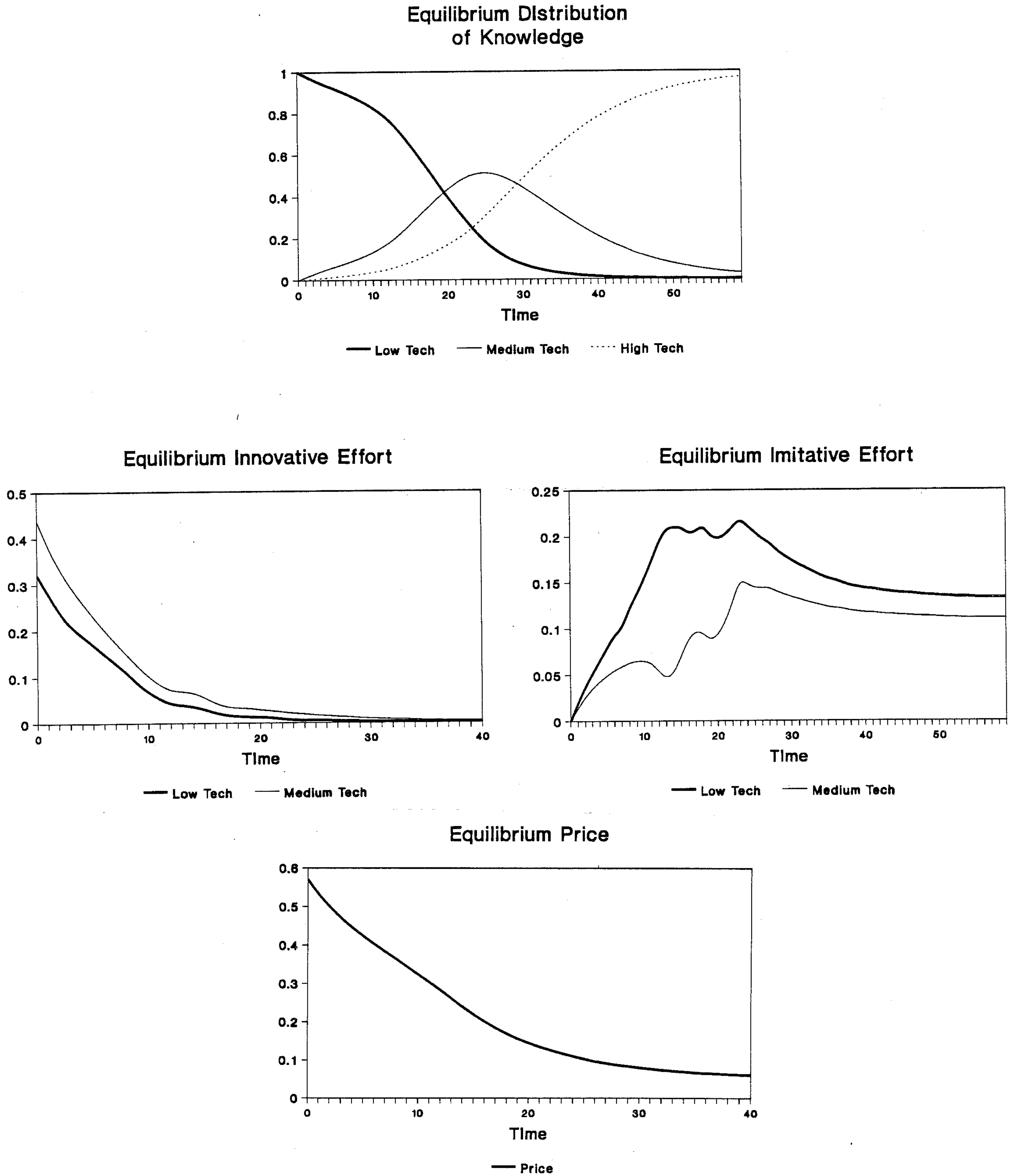
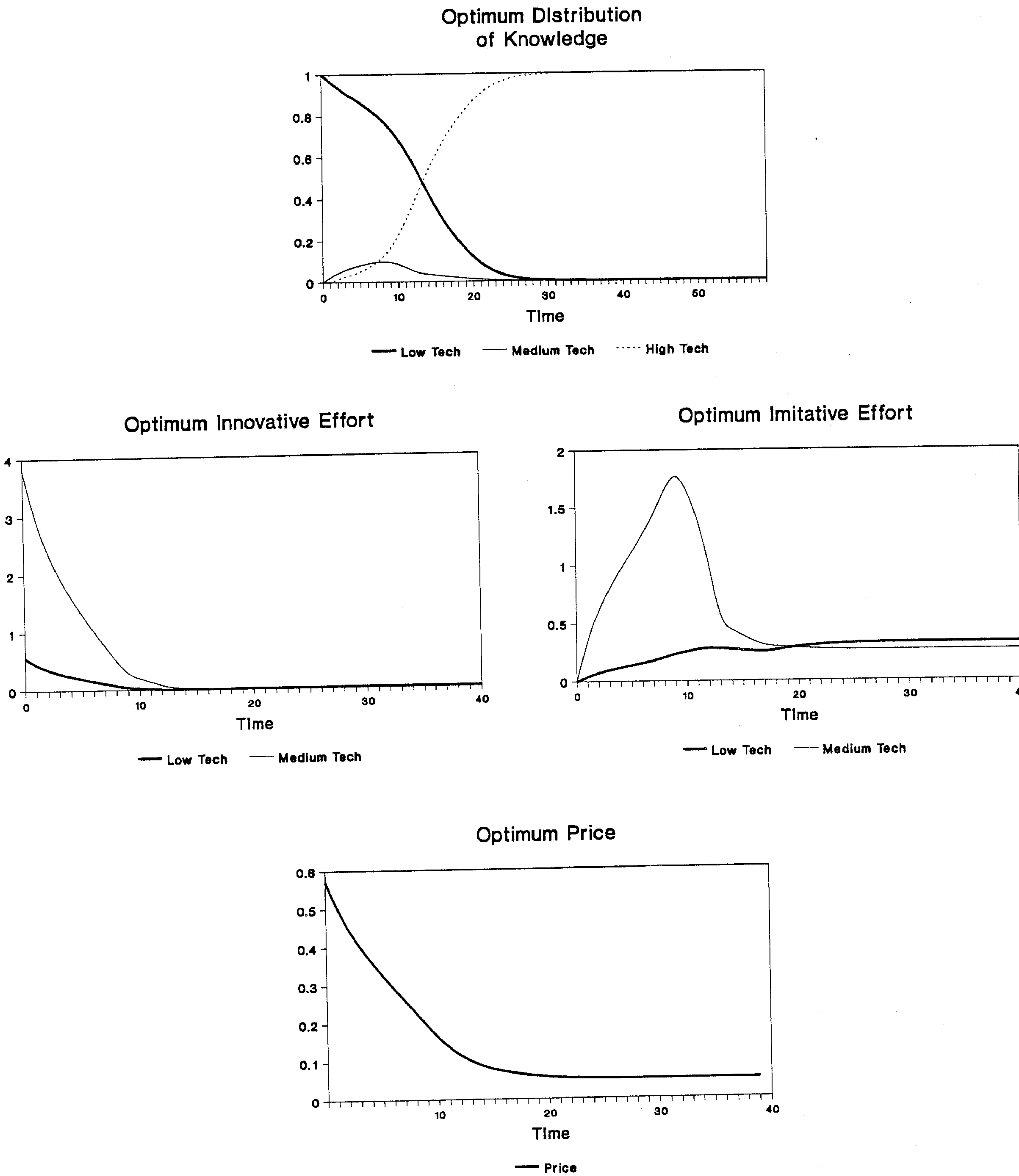

Equilibrium Distribution
of Knowledge
1
0.8
0.6
0.4
0.2
0
0 10 20 30 40 50
Time
Low Tech
Medium Tech
High Tech
Equilibrium Innovative Effort
0.5
0.4
0.3
0.2
0.1
0
0 10 20 30 40
Time
Low Tech
Medium Tech
Equilibrium Imitative Effort
0.25
0.2
0.15
0.1
0.05
0
0 10 20 30 40 50
Time
Low Tech
Medium Tech
Equilibrium Price
0.6
0.5
0.4
0.3
0.2
0.1
0
0 10 20 30 40
Time
Price

Figure 3


Optimum Distribution
of Knowledge
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
50
Time
Low Tech
Medium Tech
High Tech
Optimum Innovative Effort
4
3
2
1
0
0
10
20
30
40
Time
Low Tech
Medium Tech
Optimum Imitative Effort
2
1.5
1
0.5
0
0
10
20
30
40
Time
Low Tech
Medium Tech
Optimum Price
0.6
0.5
0.4
0.3
0.2
0.1
0
10
20
30
40
Time
Price

sooner. Medium tech makes a very brief appearance as the large information gathering efforts yield rapid diffusion of high tech.

(d) Two-State Example

The next example is like the previous one, except that here it is possible to proceed analytically because $\Theta = \{\theta_0, \theta_1\}$, and $d_1 = 0$. Thus there is but one technological improvement to be learned, and $p_t = d_0$, all t. Since θ_2 is absent, δ_{01} may be written δ. Again let $v^i \equiv v(\{\theta_i\})$. The final difference is that $M(\theta_0, \theta_0, v) = 1 - (v^1)^{1/2}$; i.e. the probability of imitation is $\mu(v^1)^{1/2}$. In contrast to the previous example, since $(v^1)^{1/2} > v^1$, the imitator's efforts are directed towards firms that already have the more advanced technology.

Since firms knowing θ_2 need neither innovate nor imitate, a complete description of equilibrium is given by

$$\Phi(v)(\{\theta_1\}) = v^1 + (1 - v^1)\,[\eta(\theta_0, v)\delta + \mu(\theta_0, v)\,(v^1)^{1/2} - \eta(\theta_0, v)\,\mu(\theta_0, v)\,\delta(v^1)^{1/2}] \quad (10a)$$

$$v_0(\{\theta_0\}) = 1, \quad (10b)$$

$$V(\theta_0, v) = \theta_0 d_0^2/(2c_q) - c_\eta\,\eta(\theta_0, v_t)^2/2 - c_\mu\,\mu(\theta_0, v_t)^2/2$$
$$+ \beta\Big\{[1 - \eta(\theta_0, v)\delta][1 - \mu(\theta_0, v)(v^1)^{1/2}]\;V[\theta_0, \Phi(v)]$$
$$+ \Big[\eta(\theta_0, v)\delta + \mu(\theta_0, v)(v^1)^{1/2} - \eta(\theta_0, v)\mu(\theta_0, v)\delta(v^1)^{1/2}\Big] V[\theta_1, \Phi(v)]\Big\}, \quad (11a)$$

$$V(\theta_1, v) = \theta_1 d_0^2/[2c_q(1 - \beta)], \quad (11b)$$

$$- c_\eta\,\eta(\theta_0, v) + \beta\,\delta[1 - \mu(\theta_0, v^1)(v^1)^{1/2}]\,\Big[V[\theta_1, \Phi(v)] - V[\theta_0, \Phi(v)]\Big] = 0, \quad (12a)$$

$$- c_\eta\,\mu(\theta_0, v) + \beta[1 - \eta(\theta_0, v)\delta](v^1)^{1/2}\Big[V[\theta_1, \Phi(v)] - V[\theta_0, \Phi(v)]\Big] = 0, \quad (12b)$$

and

$$c_\eta\, c_\mu - \Big[\beta\,\delta(v^1)^{1/2}\,\Big[V[\theta_1, \Phi(v)] - V[\theta_0, \Phi(v)]\Big]\Big]^2 \geq 0. \quad (12c)$$

(10) corresponds to (8) above, (11) to (6), and (12) to (5).

From (12) it is immediate that if $(v^1)^{1/2}\ \Delta(v) \equiv (v^1)^{1/2} \left[V[\theta_1, \Phi(v)] - V[\theta_0, \Phi(v)]\right]$ is rising in v^1, $\eta(\theta_0, v)$ declines and $\mu(\theta_0, v)$ rises as $\{v_t\}$ unfolds. The intuition is as follows: $\Delta(v)$ is the net benefit from learning θ_1, which declines as v_t evolves, this decline occurring because the value of knowing only θ_0 increases as imitation becomes easier with rising v^1. Falling $\Delta(v)$ works to reduce both innovation and imitation. However, that imitation is becoming easier makes it more likely (for fixed η and μ) that it will be imitation, not innovation, that yields knowledge of θ_1. This effect further encourages a substitution of imitation for innovation. When $(v^1)^{1/2}\ \Delta(v)$ is rising in v^1, the substitution of imitation for innovation dominates the influence of declining net benefit to learning, so that imitation must rise, and innovation fall.

That $(v^1)^{1/2}\ \Delta(v)$ is rising in v^1 must hold "on average", since $v_0^1\ \Delta(v_0) = 0$ (ie. $v_0^1 = 0$) and $\lim_{t\to\infty} v_t^1\ \Delta(v_t) > 0$. In what follows, assume that $(v^1)^{1/2}\ \Delta(v)$ is invariably increasing in v^1. This restriction leads to the simple results put to use in the subsection following.[19]

The variety of behaviors possible in industries in which technological advance is endogenous can be illustrated by two polar cases--pure imitation and pure innovation. Pure imitation results when innovation is very difficult: $\delta \cong 0$. In this case -- similar in spirit, as well as in implications, to Schumpeter's (1934) model--innovation is rare, and nearly all diffusion of new technology is imitative. Pure innovation is obtained by introducing the parameter ξ into M via $M(\theta_0, \theta_0, v) = 1 - \xi(v^1)^{1/2}$, and then considering $\xi \cong 0$; ξ will be suppressed until it is used below. Pure innovation is related to the situation studied in the R & D literature(for example, Scherer (1967), Kamien and Schwartz (1972), Telser (1982) and Nelson (1982)) in that diffusion of any new technology is solely a consequence of independent efforts to implement it; however, this example improves matters by specifying production and

---

[19]That it is indeed possible to satisfy this condition has been checked numerically; indeed, it appears to be difficult to produce a case where it is not satisfied. In an earlier version of this paper (available from the authors) it was shown that this restriction is satisfied whenever a condition on μ and η holds. A variety of comparative dynamics proposition are also provided for the pure models.

innovation as simultaneous, ongoing processes instead of requiring innovation to precede production.

Three basic questions can be answered in each pure model: (i) What are the characteristics of diffusion of new technology? (ii) How does the distribution of firm outputs evolve? (iii) What is the time path of aggregate R & D expenditures?

Diffusion of New Technology

In the two-state setting, diffusion is fully captured by $\{v_t^1\}$. Under pure imitation,

$$v_{t+1}^1 \cong v_t^1 + (1 - v_t^1)\, \mu(\theta_0, v_t)\, (v_t^1)^{1/2},$$

where

$$\mu(\theta_0, v_t) \equiv \beta(v_t^1)^{1/2}\, \Delta(v_t) / c_\mu.$$

Observe that firms knowing only θ_0 are more likely to learn θ_1 as time passes; i.e. $\mu(\theta_0,v_t)(v_t^1)^{1/2}$ increases as v_t evolves. Using the expression for μ, diffusion is given by

$$v_{t+1}^1 - v_t^1 = \beta(1 - v_t^1)\, v_t^1\, \Delta(v_t) / c_\mu,$$

which is small both early on and for large t, and achieves a maximum before a majority of firms have learned θ_1, since Δ is declining in v^1.[20] In this sense pure imitation must result in the "S-shaped" diffusion familiar at least since the work of Griliches (1957, 1963) and Mansfield (1963).

Under pure innovation, the absence of effective opportunity to learn from others renders $\Delta(v)$ nearly constant; it follows that $\eta(\theta_0,v_t)$ is nearly constant as v_t evolves, in which case diffusion is concave and most rapid at the outset.

Evolution of the Distribution of Firm Outputs

Since the price of output is constant, firm output given θ is also constant; thus define $q^0 \equiv q(\theta_0,v)$ and $q^1 \equiv (\theta_1,v)$.

---

[20]Obviously this result is dependent on the particular structure assumed for M. Indeed, that $(v^1)^{1/2}\, \Delta(v)$ is rising in v^1 implies that maximal diffusion occurs for $v_t^1 \in (1/3, 1/2)$.

Average output at t is given by $v_t^1 q^1 + (1-v_t^1)q^0 = q^0 + v_t^1(q^1 - q^0)$; i.e. a simple rescaling of v_t^1. It follows that the time path of average output displays the same qualitative features as diffusion.

The variance of output at t is proportional to $v_t^1(1-v_t^1)$. Thus, given the diffusions discussed above, under pure imitation heterogeneity in output rises slowly and is driven out quickly. Pure innovation rapidly yields heterogeneity, which then dies off slowly.

Note that *endogenous* imitative activities, spillovers, etc. occur only when the distribution of knowledge has positive variance, here equivalent to variance in the distribution of output. In particular, attempts to uncover the extent of spillovers empirically by focusing on mean output instead of its variance are doomed since variance drives the spillovers, but mean and variance are not even consistently positively correlated in this model. For example, in the 3-state example studied in Section 3(c), the time-series correlation of average output and output variance is -.66. Aggregate imitation expenditures are positively correlated (.77) with the variance of output, but negatively and weakly correlated (-.19) with average output.

The Time Path of Aggregate R & D Expenditures

Assuming R & D expenditures comprise both the costs of innovation and imitation (evaluation of others' methods of production, reverse engineering, etc.), industry expenditures on R & D are $(1-v_t^1)[c_\eta \eta(\theta_0,v_t)^2/2 + c_\mu \mu(\theta_0,v_t^2)^2/2]$.

Under pure imitation, industry expenditures are proportional to $(1-v_t^1)v_t^1\Delta(v_t)^2 = (v_{t+1}^1 - v_t^1)\Delta(v_t)$, which peaks before maximal diffusion.

In contrast, pure innovation yields a strictly declining aggregate expenditure path, proportional to $(1-v_t^1)$.

Using the expressions given above, these results may be cast as follows: pure imitation predicts a rising hazard of new adoption, a rising ratio of output variance to diffusion, and a rising ratio of diffusion to aggregate R & D expenditures; pure innovation yields a constant hazard, a rising ratio of output variance to diffusion, and a constant ratio of diffusion to R & D expenditures.

e) **Death of the Steam Engine**

The displacement of steam locomotives by diesels is a phenomenon to which the two-state model usefully applies.[21] The twentieth century has seen a host of innovations in the railroad industry, but they plausibly are dwarfed by the switch from steam to diesel.

The first usable diesel locomotive was invented by Rudolf Diesel in 1912 (Schmookler, 1966). Diesel locomotives were first used in the U.S. in 1925, and by 1968 they had displaced steam engines entirely.[22] The top panel of Figure 4 displays data for the U.S., 1925-67, on the fraction of all locomotives that were diesels. Evidently, the distribution of technology increases over time in the sense of first order stochastic dominance (here equivalent to v_t^1 rising over time).

Modelling imitation as involving spillovers implies that the likelihood of switching technologies reflects the degree to which new technology is currently utilized. Most starkly, in the pure imitation model, the hazard is strictly rising. If pure imitation is an adequate description of the railroad industry, the hazard $h_t \equiv (v_{t+1}^1 - v_1^t)/(1 - v_t^1)$, should be increasing over time. Figure 4 also displays $\{h_t\}$, which certainly increases throughout most of its range; indeed the declining portion of $\{h_t\}$ is precisely that period during which less than 1% of all locomotives were stream-driven, in which case $1 - v_t^1 \leq .01$, and erratic fluctuation in the h_t series is to be anticipated. The pure imitation framework has the stronger implication that $\mu(\theta_0, v_t)$ is rising as v_t evolves. The bottom panel of Figure 4 depicts $\mu_t \equiv h_t/(v_t^1)^{1/2}$ which, like $\{h_t\}$, is generally rising and falls only during the period when $1 - v_t^1$ is very close to zero.

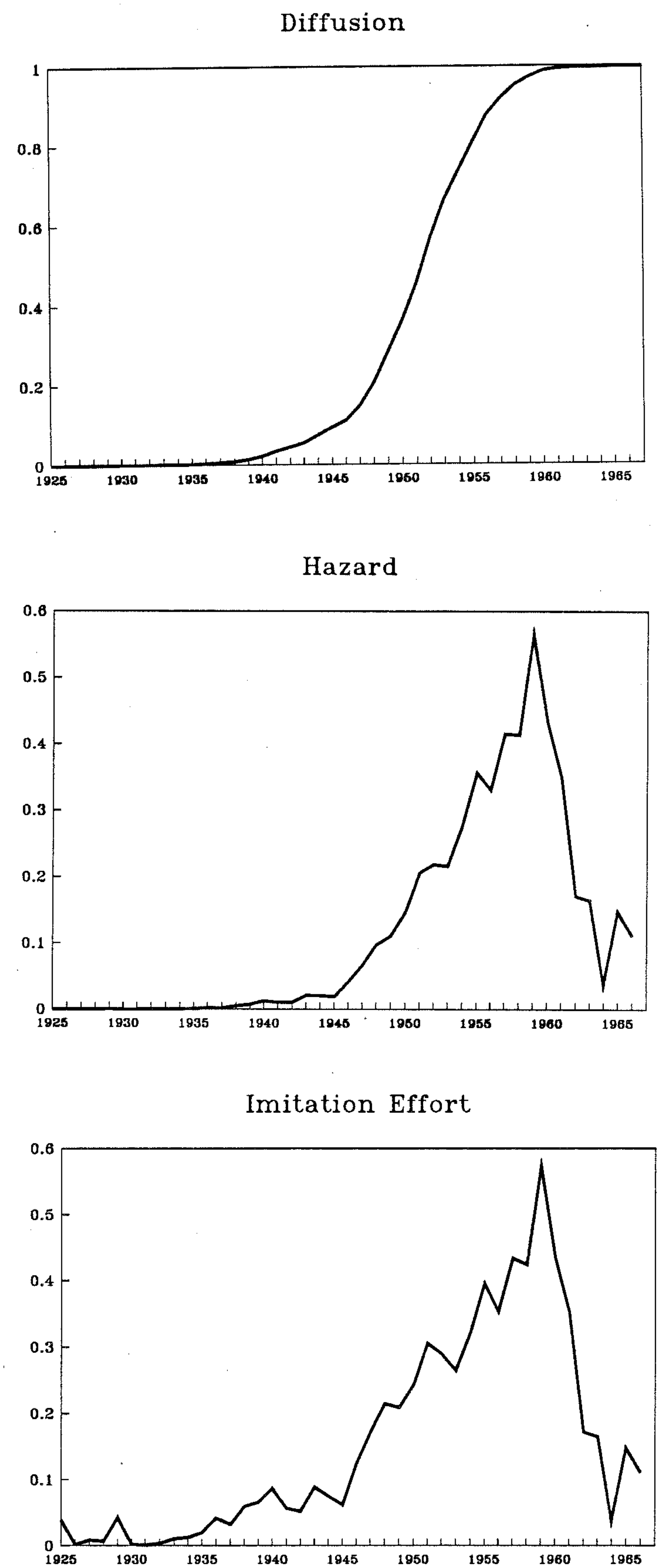
While these illustrative calculations do not prove that an informational model is indeed generating the data, it is worth noting first that a key implication of the main alternative--namely a vintage capital model--fails here for two reasons. First, new stream

---

[21]In the earlier version, a similar exercise, but exemplifying pure innovation, was carried out using data on mechanized loading techniques in U.S. underground coal mines.

[22]A few electric and "other" locomotives are ignored in what follows; as a group they never amounted to more than 2% of the total.

Figure 4


Diffusion
1
0.8
0.6
0.4
0.2
0
1925
1930
1935
1940
1945
1950
1955
1960
1965
Hazard
0.6
0.5
0.4
0.3
0.2
0.1
0
1925
1930
1935
1940
1945
1950
1955
1960
1965
Imitation Effort
0.6
0.5
0.4
0.3
0.2
0.1
0
1925
1930
1935
1940
1945
1950
1955
1960
1965

locomotives were *produced* long after the introduction of diesels; see Interstate Commerce Commission (1950, Tables A-4 and A-5). And second, there appears to be no evidence that the age distribution of steam locomotives was bell-shaped in 1925; a bell-shape is key for a vintage explanation of the S-shaped diffusion displayed in Fig. 4.[23]

Nor did the case that the substitution of diesels for steam locomotives merely reflect the cheapening of oil relative to coal over time. In fact, over the period during which the primary displacement of steam engines occurred (approximately, 1940-60), the ratio of coal to oil prices *fell* by about 13%; see U.S. Bureau of the Census, *Historical Statics of the United States*, series M93-106 and M138-142.

4. CONCLUDING REMARKS AND RELATED EVIDENCE

This paper has analyzed competition among firms that differ only in their productive knowledge. All lags in the diffusion of technology stem from informational barriers between decision units that are labelled "firms" here. There are good reasons why the informational unit may indeed be the firm: patents, for instance, are granted to firms, and information-sharing mechanisms such as patent-swapping arrangements and licensing deals are made among firms. On these grounds, it is natural to think of the firm as the owner of a piece of information, information that other *firms* can try to acquire.[24]

On the other hand, further theoretical and empirical considerations suggest that the informational unit is not the firm, but perhaps the plant. Holmstrom (1982), among others, has highlighted the incentive problems that may arise within the firm -- problems that may deter a plant manager from sharing his technological know-how with his peers. And on the empirical side, Mansfield (1963) has shown that the spread of a new technology within the firm can take almost as long as its spread within the industry. To explain diffusion lags within the firm the model must interpret them as resulting from informational barriers inside the firm.

---

[23]Furthermore, Mansfield (1963, p. 355) finds that, contrary to the vintage hypothesis, the process of dieselization was no quicker in firms who had older steam engines on hand prior to starting dieselization.

[24]Prescott and Visscher (1980) elaborate on this point.

Now if plants are treated as decision units, then the concept of the firm is lost, and the model's predictions are about plants, not firms. But whether the informational unit is the firm or the plant, this paper has related the nature of discovery and imitation to some key properties of the evolution of an industry. This is the paper's positive contribution.

APPENDIX

A.1 Proof of Theorem 1

The proof employs the following result.

Lemma: Let (i) Z be an arbitrary set; (ii) v_1 *and* v_2 *be positive measures on a* σ*-algebra of subsets of Z; and (iii)* f_1 *and* f_2 *be positive, real-valued, measurable functions on Z, bounded by* $\bar{f}$*. If* $v_i(Z) \leq \bar{v} < \infty$*, for* $i = 1, 2$*, then*

$$\left| \int f_1 dv_1 - \int f_2 dv_2 \right| \leq \bar{v} \, \| f_1 - f_2 \| + \bar{f} \, \| v_1 - v_2 \|.$$

Proof of Lemma:

$$\int f_1 dv_1 - \int f_2 dv_2 = \int (f_1 - f_2) dv_2 + \int f_1 dv_1 - \int f_1 dv_2.$$

For the first term on the right-hand side,

$$\int (f_1 - f_2) dv_2 \leq \left| \int (f_1 - f_2) dv_2 \right| \leq \| f_1 - f_2 \| \int dv_2 \leq \bar{v} \, \| f_1 - f_2 \|.$$

Turning to the second term

$$\int f_1 dv_1 - \int f_1 dv_2 = \int f_1 d(v_1 - v_2) = \int f_1 d(v_1 - v_2)^+ - \int f_1 d(v_1 - v_2)^- ,$$

where $(v_1 - v_2)^+$ and $(v_1 - v_2)^-$ are the positive and negative parts of $v_1 - v_2$.

Since $\int f_1 d(v_1 - v_2)^+$ and $\int f_1 d(v_1 - v_2)^-$ are both positive,

$$\left| \int f_1 d(v_1 - v_2) \right| \leq max \left[\int f_1 d(v_1 - v_2)^+, \int f_1 d(v_1 - v_2)^- \right] \leq \bar{f} \, \| v_1 - v_2 \|,$$

since f_1 is positive. Thus

$$\left| \int f_1 dv_1 - \int f_2 dv_2 \right| = \left| \int (f_1 - f_2) dv_2 + \int f_1 d(v_1 - v_2) \right|$$

$$\leq \left| \int (f_1 - f_2) dv_2 \right| + \left| \int f_1 d(v_1 - v_2) \right|$$

$$\leq \bar{v} \, \| f_1 - f_2 \| + \bar{f} \, \| v_1 - v_2 \|,$$

completing the proof of the lemma.

The proof of Theorem 1 follows. For any strategy a, let

$$\xi_{1t}(x,v) \equiv D_t \left[\int q_t(\theta, x, v) v(d\theta), x \right],$$

$$\xi_{2t}(x,v) = \int \Psi_t[\cdot, a_t(\theta, x, v), \theta, x, v] \, v(d\theta)$$

and

$$\xi_t(x,v) = (\xi_{1t}(x, v), \xi_{2t}(x, v)).$$

Let $(\Xi, \|\cdot\|)$ be the space of functions that are also continuous on $K \times \mathcal{M}$, with the topology on $\mathcal{M}$ being weak convergence. Then $(\Xi, \|\cdot\|)$ is a complete metric space.

For fixed ξ, let V^{ξ} be the unique fixed point of the operator defined by the right-hand side of (6), and α^{ξ} the policy that uniquely attains V^{ξ}. Existence of equilibrium is equivalent to existence of $\xi \in \Xi$ such that: $\forall (t, x, v)$,

$$\xi_t(x,v) = \left[(D_t[\int q_t^{\xi}(\theta,x,v) v(d\theta),x], \int \Psi_t[\cdot, \alpha_t^{\xi}(\theta,x,v),\theta,x,v] v(d\theta) \right]. \tag{A.1.1}$$

Let $R(\xi)$ be the map from Ξ to Ξ defined by the right hand side of (A.1.1). Then, it remains to prove the existence of $\xi \in \Xi$ such that $\xi = R(\xi)$. To do so, it is shown that (a) R preserves continuity of ξ, and (b) R is a contraction mapping, in which case R has an *unique* fixed point. For (b), it suffices to show that $\forall(\xi,\xi')\in \Xi\times\Xi$, $\|R(\xi) - R(\xi')\| \le \gamma\|\xi - \xi'\|$ for some $\gamma\in(0,1)$. (b) is proved first.

Evaluated at (t,x,v), for any $(\xi,\xi')\in \Xi^2$,

$$R(\xi) - R(\xi') \le \sup_t \max\left\{d\int(q_t^{\xi} - q_t^{\xi'})v(d\theta), \psi_\alpha\int(\alpha_t^{\xi} - \alpha_t^{\xi'})v(d\theta)\right\}$$

$$\le \sup_t \max(d,\psi_\alpha)\int(\alpha_t^{\xi} - \alpha_t^{\xi'})v(d\theta) \quad \le \max(d,\psi_\alpha)\|\alpha^{\xi} - \alpha^{\xi'}\|.$$

Thus , it suffices to show that for some $\gamma\in(0, 1)$, and all (ξ, ξ'), $\in \Xi$

$$\|\alpha^{\xi} - \alpha^{\xi'}\| \le \frac{\gamma}{\max(d,\psi_\alpha)}\|\xi - \xi'\|. \tag{A.1.2}$$

To demonstrate (A.1.2), let $B^{\xi}_{t+1}(\theta',x,v) \equiv \int V^{\xi}_{t+1}[\theta',x',\xi_{2t}(x,v)]\chi(dx',x)$. Then

$$\int V_{t+1}[\theta',x',\xi_{2t}(x,v)]\Psi_t(d\theta',\alpha_t^{\xi},\theta,x,v)\chi(dx',x) = \int_{[\theta,\bar{\theta}]} B^{\xi}_{t+1}\Psi_t(d\theta',\alpha^{\xi},\theta,x,v)$$

$$= B^{\xi}_{t+1}(\bar{\theta},x,v) - B^{\xi}_{t+1}(\theta,x,v)\Psi_t[[0,\theta],\alpha^{\xi},\theta,x,v] - \int_{(\theta,\bar{\theta}]} \Psi_t[[0,\theta'],\alpha^{\xi},\theta',x,v]B^{\xi}_{t+1}(d\theta',x,v)$$

$$= B^{\xi}_{t+1}(\bar{\theta},xv) - \int_{[\theta,\bar{\theta}]} \Psi_t[[0,\theta], \alpha^{\xi},\theta',x,v]B^{\xi}_{t+1}(d\theta',x,v),$$

where the second equality follows from an integration by parts. Thus the first order condition for maximal α^{ξ} is: $\forall\ (\theta,t,x,v)$,

$$\xi_{1t}I - \nabla_\alpha c_t - \beta\int_{[\theta,\bar{\theta}]} \nabla_\alpha\Psi_t B^{\xi}_{t+1}(d\theta',\cdot) = 0. \tag{A.1.3}$$

where I is an $n\times 1$ vector with 1 as its first element and zeros elsewhere. (This argument assumes $\alpha\in Int\ A$. A Lipschitz condition will first be shown for regions of $S\times\Xi$ where α is interior. Subsequently, the argument is extended to allow $\alpha\in Bndy\ A$.)

For any $s\in S$, let (c_t,Ψ_t) and (c'_t,Ψ'_t) indicate that c and Ψ are evaluated using α^{ξ} and $\alpha^{\xi'}$ respectively, (A.1.3) implies

$$(\xi'_{1t} - \xi_{1t})I + \nabla_\alpha c_t - \nabla_\alpha c'_t + \beta\left[\int_{[\theta,\bar{\theta}]} \nabla_\alpha \Psi_t B^{\xi}_{t+1}(d\theta',\cdot) - \int_{[\theta,\bar{\theta}]} \nabla_\alpha \Psi'_t B^{\xi'}_{t+1}(d\theta',\cdot)\right] = 0.$$

Since B^{ξ} is measurable as a function of θ', $\|\nabla_\alpha \Psi_t\| \leq \psi_\alpha$ has been assumed, and $B \leq \bar{D}\,\bar{Q}\,/(1 - \beta)$ must hold, the Lemma gives

$$\|\nabla_\alpha c_t - \nabla_\alpha c'_t\| \leq \|\xi - \xi'\| + \beta\left[\psi_\alpha \| B^{\xi} - B^{\xi'}\| + \frac{\bar{D}\bar{Q}}{1-\beta}\|\nabla_\alpha \Psi - \nabla_\alpha \Psi'\|\right],$$

where the supreme is taken over that subset of S on which α^{ξ} and $\alpha^{\xi'}$ are both interior. Since it has been assumed that $\|\nabla_\alpha c - \nabla_\alpha c'\| \geq c_{\alpha\alpha}\|\alpha^{\xi} - \alpha^{\xi'}\|$ and $\|\nabla_\alpha \Psi - \nabla_\alpha \Psi'\| \leq \psi_{\alpha\alpha}\|\alpha^{\xi} - \alpha^{\xi'}\|$, rearrangement gives

$$\left[c_{\alpha\alpha} - \psi_{\alpha\alpha}\frac{\bar{D}\bar{Q}}{1-\beta}\right]\|\alpha^{\xi}_t - \alpha^{\xi'}_t\| \leq \|\xi - \xi'\| + \beta\psi_\alpha\|B^{\xi} - B^{\xi'}\|,$$

or

$$\|\alpha^{\xi}_t - \alpha^{\xi'}_t\| \leq \frac{\|\xi - \xi'\| + \beta\psi_\alpha\|B^{\xi} - B^{\xi'}\|}{c_{\alpha\alpha} - \psi_{\alpha\alpha}\bar{D}\bar{Q}/(1-\beta)} \qquad \text{(A.1.4)}$$

where the condition in the Theorem guarantees that the denominator of the right-hand side is positive.

(A.1.4) holds on that subset of $S \times \Xi$ on which $\alpha \in Int\ A$. To extend inequality (A.1.4) to those s for which α^{ξ}_t or $\alpha^{\xi'}_t$ is not in $Int\ A$, it is most straightforward to argue componentwise in α_t, then fix s and use a triangle inequality on Ξ.

Taking the supremum over t in (A.1.4) gives

$$\|\alpha^{\xi} - \alpha^{\xi'}\| \leq \frac{\beta\psi_\alpha\|B^{\xi} - B^{\xi'}\| + \|\xi - \xi'\|}{c_{\alpha\alpha} - \psi_{\alpha\alpha}\bar{D}\bar{Q}/(1-\beta)}$$

It remains to provide an expression relating $\|B^{\xi} - B^{\xi'}\|$ to $\|\xi - \xi'\|$.

To that end, let $\{v^a_t\}_0^\infty$ be the sequence of measures implied if all firms select the arbitrary strategy a. For fixed $\tilde{x}$, v^a_t, evolves according to

$$v^a_{t+1} = \int_\Theta \Psi_t(\cdot, a, \theta, x_t, v^a_t)\, v^a_t(d\theta),$$

$$v^a_0 = v_0.$$

Next, for any $\tau \geq t$, $\tilde{x}$ and θ_t, if a firm selects strategy a^2, while all others use a^1, the predictive probability distribution of θ_τ is that consistent with the measure $H(\cdot, a^1, a^2, \theta_t, \tilde{x})$, which evolves according to

$$H_t(\{\theta_t\}, a^1, a^2, \theta_t, \tilde{x}) = 1$$

$$H_{\tau+1}(\cdot, a^1, a^2, \theta_t, \tilde{x}) = \int_\Theta \Psi_\tau(\cdot\ a^2, \theta, x_\tau, v^{a^1}_\tau) H_\tau(d\theta, a^1, a^2, \theta_t, \tilde{x}).$$

Finally, let $\tilde{\chi}^t(\cdot, x_t)$ be the measure over sequences $\tilde{x}^t \equiv \{x_{t+\tau}\}^\infty_{\tau=1}$ consistent with χ and conditional on x_t.

Since α^ξ is feasible given ξ', (6) and the definition of B give

$$\|B^\xi_t - B^{\xi'}_t\| \leq \sup_a \Big\{ \sum_{j>t} \beta^{j-t} \int_{K^\infty} \Big[\int_\Theta \pi^\xi_j(\theta_j, a, \tilde{x}) H_j(d\theta_j, \alpha^\xi, a, \theta_t, \tilde{x})$$

$$- \int_\Theta \pi^{\xi'}_j(\theta_j, a, \tilde{x}) H_j(d\theta_j, \alpha^{\xi'}, a, \theta_t, \tilde{x}) \Big] \Big\} \tilde{\chi}^t(d\tilde{x}^t, x_t),$$

where $\pi^\xi_j(\theta_j, a, \tilde{x})$ is the firm's period j return to following strategy a when others follow α^ξ, knowledge is θ_j, and $\tilde{x}$ is given; $\pi^{\xi'}$ is defined analogously.

Since $\forall\ \xi$, $\pi^\xi \leq \bar{D}\ \bar{Q}$, the Lemma yields

$$\|B^\xi - B^{\xi'}\| \leq \bar{D}\bar{Q} \sum_{j>t} \beta^{j-t} \| H_j(\cdot, \alpha^\xi, \cdot) - H_j(\cdot, \alpha^{\xi'}, \cdot)\| + \sum_{j>t} \beta^{j-t} \| \pi^\xi - \pi^{\xi'} \|$$

$$\leq \bar{D}\ \bar{Q} \sum_{j>t} \beta^{j-t} \| H_j(\cdot, \alpha^\xi, \cdot) - H_j(\cdot, \alpha^{\xi'}, \cdot)\| + \bar{Q}\ \frac{\|\xi_1 - \xi'_1\|}{1-\beta}. \qquad \text{(A.1.5)}$$

The definition of H, in conjunction with the Lemma, gives

$$\|H_j(\cdot, \alpha^\xi, \cdot) - H_j(\cdot, \alpha^{\xi'}, \cdot)\|$$

$$\leq \|H_{j-1}(\cdot, \alpha^\xi, \cdot) - H_{j-1}(\cdot, \alpha^{\xi'}, \cdot)\| + \|\Psi_j(\cdot, v^{\alpha^\xi}_j) - \Psi_j(\cdot, v^{\alpha^{\xi'}}_j)\|$$

$$\leq \sum_{s=t}^{j} \| \Psi_s(\cdot, v^{\alpha^\xi}_s) - \Psi_s(\cdot, v^{\alpha^{\xi'}}_s)\| \leq \psi_v \sum_{s=t}^{j} \| v^{\alpha^\xi}_s - v^{\alpha^{\xi'}}_s \|, \qquad \text{(A.1.6)}$$

by the Lipshitz property of Ψ. Using the Lemma again,

$$\| v_{t+1}^{\alpha^{\xi}} - v_{t+1}^{\alpha^{\xi'}})\| \leq \psi_v \| v_t^{\alpha^{\xi}} - v_t^{\alpha^{\xi'}})\| + \psi_\alpha \| \alpha^{\xi} - \alpha^{\xi'} \|,$$

implying

$$\| v_t^{\alpha^{\xi}} - v_t^{\alpha^{\xi'}})\| \leq \psi_\alpha \sum_{j=0}^{t} \psi_v^j \| \alpha^{\xi} - \alpha^{\xi'} \|.$$

Thus, substituting in (A.1.6) gives

$$\|H_j(\cdot, \alpha^{\xi}, \cdot) - H_j(\cdot, \alpha^{\xi'}, \cdot)\| \leq \left[\psi_\alpha \psi_v \sum_{s=t}^{j} \sum_{l=0}^{s} \psi_v^l\right] \|\alpha^{\xi} - \alpha^{\xi'}\|. \tag{A.1.7}$$

Then

$$\sum_{j>t} \beta^{j-t} \| H_j(\cdot, v^{\xi}, \cdot) - \mathrm{H}_j(\cdot, v^{\xi'}, \cdot)\| \leq \|\alpha^{\xi} - \alpha^{\xi'}\| \psi_\alpha \psi_v \sum_{j>t} \beta^{j-t} \left[\sum_{s=t}^{j} \sum_{l=0}^{s} \psi_v^l\right]$$

$$\leq \|\alpha^{\xi} - \alpha^{\xi'}\| \psi_\alpha \psi_v \sum_{j>t} (1+j-t)\beta^{j-t} \frac{1}{1-\psi_v} = \|\alpha^{\xi} - \alpha^{\xi'}\| \frac{\psi_\alpha \psi_v}{1-\psi_v} \frac{1}{(1-\beta)^2}. \tag{A.1.8}$$

Substitution in (A.1.5) then gives

$$\| B^{\xi} - B^{\xi'} \| \leq \bar{D}\bar{Q} \frac{\psi_\alpha \psi_\alpha}{(1-\psi_v)(1-\beta)^2} \| \alpha^{\xi} - \alpha^{\xi'} \| + \bar{Q} \| \frac{\xi_1 - \xi_1'}{(1-\beta)} \|$$

$$\leq \bar{D}\bar{Q} \frac{\psi_\alpha \psi_v}{(1-\psi_v)(1-\beta)^2} \| \alpha^{\xi} - \alpha^{\xi'} \| + \bar{Q} \| \frac{\xi - \xi'}{(1-\beta)} \|. \tag{A.1.9}$$

Thus, using (A.1.9), since $\| \alpha_t^{\xi} - \alpha_t^{\xi'} \| \leq \| \alpha^{\xi} - \alpha^{\xi'} \|$, (A.1.4) will hold if

$$\| \alpha^{\xi} - \alpha^{\xi'} \| \leq \frac{\| \xi - \xi' \|}{c_{\alpha\alpha} - \psi_{\alpha\alpha} \bar{D}\bar{Q} / (1-\beta)}$$

$$+ \frac{\beta\psi_{\alpha}}{c_{\alpha\alpha} - \psi_{\alpha\alpha}\bar{D}\bar{Q}/(1-\beta)} \left\{ \frac{\bar{D}\bar{Q}\psi_{\alpha}\psi_{v}}{(1-\psi_{v})(1-\beta)^2} \| \alpha^{\xi} - \alpha^{\xi'} \| + \frac{\bar{Q}}{(1-\beta)} \|\xi - \xi'\| \right\},$$

or

$$\| \alpha^{\xi} - \alpha^{\xi'} \| \leq \frac{1 + \frac{\beta\psi_{\alpha}\bar{Q}}{1-\beta}}{c_{\alpha\alpha} - \psi_{\alpha\alpha}\frac{\bar{D}\bar{Q}}{1-\beta}\left[1 + \frac{\beta\psi_{\alpha}^2\psi_{v}}{\psi_{\alpha\alpha}(1-\psi_{v})(1-\beta)}\right]} \| \xi - \xi' \|.$$

Thus referring to A.1.2., it suffices to require that

$$max(d,\psi_{\alpha}) \leq \frac{(1-\beta)c_{\alpha\alpha} - \psi_{\alpha\alpha}\bar{D}\bar{Q}\left[1 + \frac{\beta\psi_{\alpha}\psi_{v}}{1-\beta}\right]}{1-\beta + \beta\psi_{\alpha}\bar{Q}}.$$

This shows that R is a contraction mapping. That R indeed preserves continuity of ξ is shown as part of the proof of Lemma 1 (immediately following). ||

A.2 Proof of Lemma 1

The operator on the space of bounded functions from S to $\mathbb{R}$ defined by the right hand side of (6) preserves monotonicity in θ; thus V is increasing in θ, proving part (ii). Let ξ be as defined in (A.1.1). Let ξ^n be the sequence generated by iterating the right hand side of (A.1.1), beginning with an initial ξ^0. If it can be shown that R preserves continuity of ξ, then Theorem 1 proves that ξ^n converges strongly to the fixed point of (A.1.1). Now assume that ξ^0 is continuous. Since $(\Theta, \|\cdot\|)$ is a complete metric space, the fixed point of (A.1.1) will also be continuous if it can be shown that if ξ^n is continuous, then ξ^{n+1} is as well.

The latter task is accomplished by using Hildenbrand (1974, p. 51, property (38)). If ξ^n is continuous then V^{n+1} will be Lipschitz in θ if V^n is- - this occurs because c and Ψ are Lipschitz in θ. Moreover, given that $\beta < 1$, it easily follows that the Lipschitz constant on V

can be established independently of n. Now, on compact sets Lipschitz functions that converge pointwise also converge uniformly. Therefore, $V^n_{t+1}[\theta', x', \Phi^n_t(x,v_k)]$ converges uniformly on Θ as $v_k \rightarrow v$ weakly, so long as V^n and Φ^n_t are continuous; here $\Phi^n(x,v) \equiv \int \Psi(\cdot, \alpha^{\xi^n}, \theta, x, v) v(d\theta)$. Hence the n-th iterate of (A.1.1) preserves continuity of V^n; that is, property (38) of Hildenbrand implies that V^{n+1} is continuous. Now if it can be shown that Φ^{n+1} is also continuous, this will establish that the limit, V, is continuous.

To establish this, it is first shown that the policy maximal at the n-th iterate, α^{ξ^n}, is continuous. The policy α^{ξ^n} will also be continuous if the bracketed portion of the right hand side of (6) is concave in α when evaluated at equilibrium V. Since $-c_t$ is concave in α, it suffices to prove that $\forall s$,

$$\int_{\Theta} V^n_{t+1}[\theta', x', \Phi^n(x,v)]\Psi_t(d\theta', \alpha, \theta, x, v) \text{ is}$$

concave in α.

To that end, let $\bar{\alpha} \equiv \gamma\alpha^1 + (1-\gamma)\alpha^2$ for $\gamma \in (0,1)$ and α^1, $\alpha^2 \in A$. Let $\tilde{\Psi}$ be the distribution function implied by Ψ. It is enough to show that (suppressing θ, x, v and Φ) $\forall(\gamma, \alpha^1, \alpha^2)$:

$$\int_{\Theta} V^n(\theta', \cdot) d\tilde{\Psi}_t(\theta', \bar{\alpha}, \cdot) \geq \gamma \int_{\Theta} V^n(\theta', \cdot) d\tilde{\Psi}_t(\theta', \alpha^1, \cdot) + (1-\gamma) \int_{\Theta} V^n(\theta', \cdot) d\tilde{\Psi}_t(\theta', \alpha^2, \cdot). \qquad (*)$$

Since V^n is increasing in θ, (*) will hold if $\tilde{\Psi}_t(\cdot, \bar{\alpha})$ exceeds $\gamma\tilde{\Psi}_t(\cdot, \alpha^1, \cdot) + (1-\gamma)\tilde{\Psi}_t(\cdot, \alpha^2, \cdot)$ in the sense of first order stochastic dominance. This holds since $\tilde{\Psi}$ is assumed convex in α. This establishes that α^{ξ^n} is continuous. Next, the Lipschitz property of $\tilde{\Psi}$ in θ, and hence of V^n in θ, and the assumptions on c and Ψ involving the constants $c_{\alpha\alpha}$ and $\psi_{\alpha\alpha}$, imply that α^{ξ^n} is Lipschitz in θ, uniformly in n. This, Hildenbrand's property (38), and (A.1.1) imply that Φ^{n+1} is also continuous, as is $P^{n+1} \equiv D(\int q^{\xi^{n+1}}(\theta, x, v) v(d\theta), x)$. Hence V, α, P and Φ are continuous since continuity of these objects is preserved at each iterate of (A.1.1); moreover, R preserves continuity of ξ. ||

A.3 Proof of Lemma 2

Let $b = [0, \theta']$. Then

$$\Phi_t(x,v)*b) = \int_\Theta \Psi[b,\alpha_t(\theta,x.v),\theta,x,v]v_t(d\theta)$$
$$= \int_b \Psi[b,\alpha_t(\theta,x,v),\theta,x,v]v_t(d\theta)$$
$$\leq \int v_t(d\theta)$$
$$= v_t(b).$$

||

A.4 Proof of Theorem 2:

Consider the sequence of distribution functions $\{F_t\}$ corresponding to the sequence of probability measures $\{v_t\}$. By Lemma 1, for each $\theta \in [0,\bar{\theta}] \rightarrow [0,1]$. Define $F^*: \Theta \longrightarrow [0,1]$ pointwise by $F^*(\theta) = \lim_{t\rightarrow\infty} F_t(\theta)$. Note that F^* is increasing with $F^*(\theta) \leq F_t(\theta)$ for all θ and t.

Since $F^*(0) = 0$ and $F^*(\bar{\theta}) = 1, F^*$ is a distribution function if it is right continuous. Suppose not, then for some $\tilde{\theta} \in [0,\bar{\theta}]$ and sequence $\{\theta_n\}$, with $\theta_n \downarrow \tilde{\theta}$ there exists $\varepsilon < 0$ such that for all n, $F^*(\theta_n) > F^*(\tilde{\theta}) + \varepsilon$. Thus $F_t(\theta)_n > F^*(\theta_n) > F^*(\tilde{\theta})\ \varepsilon$. Since F_t is right continuous and ε is independent of n, $\theta_n \downarrow \tilde{\theta}$ yields $F_t(\tilde{\theta}) > F^*(\tilde{\theta}) + \varepsilon$; i.e. $\{F_t\}$ does not converge to F^* pointwise, a contradiction. Thus F^* is right continuous.

Next, for all $\theta \in \Theta$, $\theta' \in \Theta$, define $v[\theta,\theta']] = F^*(\theta)$. v can be extended uniquely to a measure $v^* \in \mathcal{M}$ (Ash, p. 24).

Finally, since $F_t(\theta) \rightarrow F^*(\theta)$ pointwise, it does so for all points of continuity of F^*, implying weak convergence of v_t to v^* (Billingsley, p. 18). ||

A.5 Proof of Theorem 3:

By the Monotone Class Theorem (Chung, 1968, p. 20, property (ix)) if $\tilde{\chi}(\{\tilde{x} \mid v^* \notin \mathcal{M}^*\}) > 0$, $\exists\ \varepsilon > 0$ such that $\sum_{x \in K'} \tilde{\chi}(\{\tilde{x} \mid \rho[v^*, \Phi(x,v^*)] > \varepsilon\}) > 0$, where $\rho: \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}_+$ is the Prohorov metric. Since K' is countable, $\exists \hat{x} \in K'$ such that

$$\tilde{\chi}(\{\tilde{x} \mid \rho[v^*, \Phi(\hat{x},v^*)] > \varepsilon) > 0. \qquad (*)$$

For any $x \in K$ and $\delta \in \mathbb{R}_+$, define $N_{\delta(x)} \equiv \{x' \in K \mid \|x' - x\| < \delta\}$. The proof uses the following lemma.

Lemma A.5.1 If (*) holds, $\exists \hat{\delta} > 0$ *such that* $\tilde{\chi}(\{\tilde{x} \mid \rho[v^*, \Phi(x,v^*)] > \varepsilon/2, x \in N_{\hat{\delta}}(\hat{x})\}) > 0.$

Proof of Lemma: For any $\delta \in \mathbb{R}_+$, define

$$B_\delta \equiv \{\tilde{x} \mid \rho[\nu^*, \Phi(\hat{x},\nu^*)] > \delta\}. \tag{*}$$

(*) yields $\tilde{\chi}(B_\varepsilon) > 0$.

Next, define $A_\delta \equiv \{\tilde{x} \mid \rho[\nu^*, \Phi(x,\nu^*) > \varepsilon/2, x \in N_\delta(\hat{x})\}$ and $A_0 \equiv \cup_\delta A_\delta$; observe that $\delta \to 0$ implies $A_\delta \uparrow A_0$. If $\tilde{\chi}(A_0) > 0$, the Monotone Class Theorem gives $\tilde{\chi}(A_{\hat{\delta}}) > 0$, some $\hat{\delta} > 0$, completing the argument. To show $\tilde{\chi}(A_0) > 0$, it suffices to prove $B_\varepsilon \subset A_0$, since $\tilde{\chi}(B_\varepsilon) > 0$.

Let $\tilde{x} \in B_\varepsilon$, so that $\rho[\nu^*, \Phi(\hat{x}, \nu^*)] > \varepsilon$. Note that (i) ρ is continuous in the topology of weak convergence; and (ii) Φ is Lipschitz in first argument, and therefore continuous in both the topology induced by the sup norm and the weak convergence topology. Thus $\exists \hat{\delta} > 0$ such that

$$x \in N_{\hat{\delta}}(\hat{x}) \Longrightarrow \rho[\nu^*, \Phi(x,\nu^*)] > \varepsilon/2;$$

that is, $\tilde{x} \in A_{\hat{\delta}}$. Since $A_{\hat{\delta}} \subset A_0$, $\tilde{x} \in A_0$. Thus $B_\varepsilon \subset A_0$. ||

Observe that the set of sequences $\tilde{x}$ such that $x_t \in N_{\hat{\delta}}(\hat{x})$ infinitely often has positive measure given assumption (ii), (*) and Lemma A.5.1.

To proceed, for any $\tilde{x}$ (with t^{th} component x_t) and associated $\{\nu_t\}$, the triangle inequality gives

$$\rho[\nu^*,\Phi(x_t,\nu^*)] \le \rho(\nu^*,\nu_t) + \rho[\nu_t,\Phi(x_t,\nu_t)] + \rho[\Phi(x_t,\nu_t),\Phi(x_t,\nu^*)]. \tag{**}$$

Let T' denote any infinite subset of T.

Lemma A.5.2.: $\forall x' \in K$ and $\eta \in \mathbb{R}_{++}$, $\exists \delta \in \mathbb{R}_{++}$ such that $\tilde{\chi}\{\tilde{x} \mid \exists T'$ such that $t \in T' \Longrightarrow$ both $\rho[\nu^*, \Phi(x_t,\nu^*)] \le \eta$ and $x_t \in N_\delta(x')\} = 1$.

Proof of Lemma: For each $x' \in K$, condition (ii) of the Theorem guarantees that $\tilde{\chi}(\{\tilde{x} \mid \exists T'$ such that $t \in T' \Longrightarrow x_t \in N_\delta(x')\} = 1$. For any $\tilde{x}$, denote the associated T' by $T'_{\tilde{x}}$.

Next, choose η, fix $\tilde{x}$ and consider (**). By Theorem 2, $\rho(v^*,v_t) \to 0$. Since $\{v_t\}$ is, therefore, a Cauchy sequence, and $v_{t+1} = \Phi(x_t,v_t)$, $\rho[v_t, \Phi(x_t, v_t)] \to 0$ as well. Let T'_η be the infinite set of dates for which the first two terms on the right hand side of (**) sum to less than $\eta/2$.

Focusing on the third term on the right hand side of (**), since $v_t \to v^*$ and Φ is continuous, for $t \in T'_{\tilde{x}}$ and t sufficiently large, $\rho[\Phi(x_t, v_t), \Phi(x_t, v^*)] \leq \eta/2$.

Altogether, let T' be the infinite subset of $T'_{\tilde{x}}$ for which the right hand side of (**) is less than η. Some such T' exists for all $\tilde{x}$ off a set of measure 0 in K^∞.

The proof concludes by selecting $\hat{x} = x'$, $\eta < \varepsilon/2$ and $\delta \leq \hat{\delta}$. Then Lemma A.5.1 and A.5.2 are contradictory. ||

A.6 Proof of Lemma 3

Let $\mathcal{M}' \equiv \{v \in \mathcal{M} | \forall x. \forall \theta \in \text{supp } v, \exists q \text{ such that } \alpha(\theta,x,v) = \underset{\sim}{0}_q\}$.

i) $\mathcal{M}' \subset \mathcal{M}^*$. Let $v \in \mathcal{M}'$, then $\forall(x,b)$, $\Phi(x,v)(b) = v(b)$, implying $v \in \mathcal{M}$

ii) $\mathcal{M}^* \subset \mathcal{M}'$. First, it can be shown that if for some x, v and open interval $I \subset \Theta$, $\inf_{\theta \in I} \|\alpha_{-q}(\theta,x,v)\| > 0$ holds, then $\forall \theta \in I$ there exists $\varepsilon > 0$ such that

$$\inf_{\theta' \in (\theta-\varepsilon.\theta]} \Psi[(\theta,\bar{\theta}],\ \alpha(\theta',x,v),\ \theta',v] > 0.$$

Now, let $v \in \mathcal{M}^*$. The measure space $(\Theta, \mathcal{B}v)$ can be decomposed into an atomless part and countable union of atoms; see Hildenbrand (p. 45). Let $\{\theta\}$ be an atom and suppose $v \notin \mathcal{M}'$; that is, for some x, $\alpha(\theta,x,v) \neq \underset{\sim}{0}_q$. Then

$$\begin{aligned}
\Phi(x,v)((\theta,\bar{\theta}]) &= v(\{\theta\})\Psi[(\theta,\bar{\theta}],\alpha(\theta,x,v),\theta,x,v] \\
&+ \int_{[0,\theta)} \Psi[[(\theta,\bar{\theta}],\alpha(\theta',x,v),\theta,x,v[v(d\theta') \\
&+ \int_{(\theta,\bar{\theta})} \Psi[(\theta,\bar{\theta}],\alpha(\theta',x,v),\theta',x,v]v(d\theta') \\
&\geq v(\{\theta\})\Psi[(\theta,\bar{\theta}],\ \alpha(\theta,x,v),\ \theta,v] + v[(\theta,\bar{\theta}]] \\
&> v[(\theta,\bar{\theta}]],
\end{aligned}$$

since $\Psi[(\theta,\bar{\theta}],\ \alpha(\theta,x,\nu),\theta,x,\nu] > 0$ is necessary for $\alpha(\theta,x,\nu) \neq \underset{\sim}{0}_q$. Thus, $\nu \notin \mathcal{M}^*$, a contradiction. Therefore $\nu \in \mathcal{M}$.

Suppose instead that $\theta \in$ supp ν is not an atom and $\nu \notin \mathcal{M}$; again for some x, $\alpha(\theta,x,\nu) \neq \underset{\sim}{0}_q$. Continuity of α in its first argument, in conjunction with $\theta \in$ supp ν, implies there is an interval I, with $\nu(I) > 0$, $\theta \in I$ and $\alpha(\theta',x,\nu) \neq \underset{\sim}{0}_q$ for all $\theta' \in I$. By the above, there is $\varepsilon > 0$ such that $\inf_{\theta' \in (\theta-\varepsilon.\theta]} \Psi[(\theta,\bar{\theta}],\alpha(\theta',x,\nu),\theta',x,\nu] > 0$. Thus

$$\begin{aligned}\Phi(x,\nu)[(\theta,\bar{\theta}]] &= \int_{[0,\theta-\varepsilon]} \Psi[(\theta,\bar{\theta}],\alpha(\theta',x,\nu),\theta',x,\nu]\nu(d\theta') \\ &+ \int_{(\theta-\varepsilon,\theta]} \Psi[(\theta,\bar{\theta}],\alpha(\theta',x,\nu),\theta',x,\nu]\nu(d\theta') + \nu[(\theta,\bar{\theta}]] \\ &\geq \nu[(\theta-\varepsilon,\theta]] \inf_{\theta' \in (\theta-\varepsilon,\theta]} \Psi[(\theta,\bar{\theta}],\alpha(\theta',x,\nu),\theta',x,\nu] + \nu[(\theta,\bar{\theta}]] > \nu[(\theta,\bar{\theta}]].\end{aligned}$$

Thus $\nu \notin \mathcal{M}^*$, again a contradiction. Altogether, $\nu \in \mathcal{M}$. ||

A.7 Proof of Theorem 4

First, using Lemma 3, note that condition (iv) implies $\nu_0 \notin \mathcal{M}^*$.

a) For some $t \in T\backslash\{0\}$, let $\nu_{t'} \in \mathcal{M}^*$. If it can be shown that $\nu_{t'-1} \in \mathcal{M}^*$ it will then follow inductively that $\nu_0 \in \mathcal{M}$, a contradiction.

To demonstrate that $\nu_{t'} \in \mathcal{M}^*$ implies $\nu_{t'-1} \in \mathcal{M}^*$, observe that restriction (i) on Ψ gives *supp* $\nu_{t'-1} \subset$ *supp* $\nu_{t'}$. Let $\theta \in$ *supp* $\nu_{t'-1}$. By Lemma 3, $\forall x$, $\alpha(\theta,x,\nu_{t'}) = \underset{\sim}{0}_q$, implying (using (i) and (ii) of the Theorem)

$$-c^1(\underset{\sim}{0}_q,\theta,x) + \beta \int_K V(\theta,x',\nu_{t'+1})\chi(dx',x)$$
$$\geq -c^1(\alpha_{-q},\theta,x) + \beta \int V(\theta',x',\nu_{t'+1})\Psi(d\theta',\alpha,\theta,x,\nu_{t'+1})\chi(dx',x).$$

Since (i) $\nu_{t'} = \nu_{t'+1}$ (by hypothesis); (ii) $\nu_{t'} \succ \nu_{t'-1}$ (Lemma 2); (iii) $\Psi(\cdot,\alpha,\theta,x,\nu_{t'}) \succ \Psi(\cdot,\alpha,\theta,x,\nu_{t'-1})$; and (iv) V is nondecreasing in θ, it follows that $\forall x$,

$$-c^1(\underset{\sim}{0}_q,\theta,x) + \beta \int_K V(\theta,x',\nu_{t'})\chi(dx',x)$$
$$\geq -c^1(\alpha_{-q},\theta,x) + \beta \int V(\theta',x',\nu_{t'})\Psi(d\theta',\alpha,\theta,x,\nu_{t'-1})\chi(dx',x).$$

Given uniqueness of α, this inequality implies $\alpha(\theta,x,\nu_{t'-1}) = \underset{\sim}{0}_q$. Since this holds $\forall\theta \in$ *supp* $\nu_{t'}$ and thus $\forall\ \theta \in supp\ \nu_{t'-1}$, Lemma 3 and condition (i) of the Theorem give $\nu_{t'-1} \in \mathcal{M}$.

b) First, except for a set of measure 0 in K^{∞}, $\nu^* \in Bd\,\mathcal{M}^*$. To see this observe that since α is Lipschitz in ν, it is also continuous in ν in the topology induced by the sup norm on $\mathcal{M}$. Since convergence in sup norm implies weak convergence, α is continuous in ν under weak convergence. Lemma 3 establishes the equivalence of $\alpha(\theta,x,\nu) = \underset{\sim}{0}_q$ and $\nu \in \mathcal{M}^*$. Thus, since α is continuous,every convergent sequence of elements in $\mathcal{M}^*$ has $\alpha = (q,\underset{\sim}{0}_q)$ as its limit; i.e the limit is $\mathcal{M}^*$. Thus $\mathcal{M}^*$ is closed. Next, if $\nu_0 \notin \mathcal{M}^*$, except for a set of measure 0 in K^{∞}, $\nu^* \notin Int\ \mathcal{M}^*$. This follows because every open ball containing ν^* contains $\nu_t \notin \mathcal{M}^*$ for t sufficiently large. Altogether, $\mathcal{M}^*$ being closed, $\nu^* \in \mathcal{M}^*$ and $\nu^* \notin Int$ yield $\nu^* \in Bd\ \mathcal{M}^*$.

Second, since $\nu^* \in \mathcal{M}^*$ almost surely, Lemma 3 and the differentiability assumed for c and Ψ imply: almost surely, $\forall x$, $\forall \theta \in supp\ \nu^*$,

$$-\nabla_{\alpha_{-q}} c^1(\underset{\sim}{0},\theta,x) + \beta \int_{K\times\Theta} V(\theta',x',\nu^*)\nabla_{\alpha_{-q}}\Psi\ (d\theta',\underset{\sim}{0},x',\nu^*)\chi(x',x) < \underset{\sim}{0}. \qquad (*)$$

Next, almost surely, $\exists\theta \in supp\ \nu^*$ and x such that some component of (*) holds as an equality. To see this, suppose that each component of (*) held as a strict inequality for all $\theta \in supp\ \nu^*$ and all x. By the assumed continuity of $\nabla_{\alpha}c$ and $\nabla_{\alpha}\Psi$, there is an open ball $B \subset \mathcal{M}^*$ containing ν^* such that $\forall\theta \in supp\ \nu^*$, $\forall x$, $\forall \nu \in B$, $\alpha(\theta,x,\nu) = \underset{\sim}{0}_q$. But then $\nu^* \notin Bd\,\mathcal{M}^*$. Therefore, $\exists\ \theta \in supp\ \nu^*$ and x such that some component of the condition holds as an equality.

(i) follows immediately. Unless $\nu^*(\{\bar{\theta}\}) = 1$, some component of (*) does not equal 0 for any $\theta \in supp\ \nu^*$ or x. Similarly, if $\nu^*(\{\bar{\theta}\}) = 1$ and the conditions of (ii) hold, (*) cannot hold as an equality for $\theta = \bar{\theta} = supp\ \nu^*$. Finally, in regard to (iii), (ii) gives $\nu^*(\{\bar{\theta})\} < 1$. Suppose, however, $\exists\tilde{\theta} < \bar{\theta}$ and $\nu^*(\{\tilde{\theta}\}) = 1$. Since $\nu_0 \notin \mathcal{M}^*$, $\exists b \subset supp\ \nu_0$ such that $\theta \in b$ implies $\alpha_{-q}(\theta,x_0,\nu_0) \neq \underset{\sim}{0}$. Let $b' = (\tilde{\theta},\bar{\theta}]$. Given $\tilde{\theta}$,$\nu^*(b') = 0$ must hold. But

$$\nu^*(b') > \int_b \Psi(b',\alpha,\theta,x_0,\nu_0)\nu_0(d\theta)$$
$$> 0$$

given the condition in part (iii), a contradiction. Thus $supp\ \nu^*$ must contain at least two points. ||

A.8 Proof of Theorem 5

For fixed $(\tilde{x}, v_0, x)$ and associated $\{v_t\}$ and v^*, let $\hat{f}_t(\theta) \equiv f[\alpha(\theta,x,v_t), \theta,x,v_t]$ and $\hat{f}^*(\theta) \equiv f[\alpha(\theta,x,v^*), \theta,x,v^*]$ Fix $\theta \in \Theta$ and let $\{\theta_t\}$ be a sequence such that $\theta_t \to \theta$. Since f and α are continuous and $v_t \to v^*$, $\hat{f}_t(\theta_t) \to \hat{f}(\theta)$. Since this is for any $\theta \in \Theta$, Theorem 5.5 of Billingsley gives that for each $x \in K$, $v_t^f(\cdot,x) \to v^{f^*}(\cdot,x)$. ||

A.9 Proof of Theorem 6

Assume $P[x, \Phi(x,v)] > P(x,v)$. Since under the stated conditions $q(\theta,x,v)$ solves $P(x,v) - \partial c/\partial q \leq 0$, with equality for $q > 0$, and $\partial^2 c/\partial q \partial \theta \leq 0$, it follows that $\forall\ \theta$, q is nondecreasing in θ and $q[\theta,x, \Phi(x,v)] \geq q(\theta,x,v)$. Thus,

$$\int_\Theta q(\theta,x,v)v(d\theta) \leq \int_\Theta q[\theta,x,\Phi(x,v)]v(d\theta) \leq \int_\Theta q[\theta,x,\Phi(x,v)]\Phi(x,v)(d\theta),$$

where the first inequality follows from the above, and the second from the monotonicity of q in θ and Lemma 2. Since D is nonincreasing in its first argument $P[x,\Phi(x,v)] \leq P(x,v)$ is implied, a contradiction. The corollary is immediate, again since D is nonincreasing in its first argument. ||

A.10 Proof of Theorem 7

By the conditions imposed in Theorem 6, along with $\tilde{c}^{0\prime}(0) = 0$, $q(\theta,x,v)$ is positive and increasing in θ and solves

$$P(x,v) - \tilde{c}^{0\prime}(q)\ \hat{c}^0(\theta,x) = 0. \qquad (*)$$

From property (i) of Ψ, $\mathit{inf\ supp}\ v_t = \mathit{inf\ supp}\ v_0 \equiv \underline{\theta}$. Thus $\forall(x,v)$, $\inf_\theta q(\theta,x,v) = q(\underline{\theta},x,v)$; moreover $\tilde{q}(\theta,x,v) = q(\theta,x,v)/q(\underline{\theta},x,v)$.

From (*),

$$\begin{aligned} q(\theta,x,v) &= \tilde{c}^{0\prime -1}[P(x,v)/\hat{c}^0(\theta,x)] \\ &= P(x,v)^\delta\, \tilde{c}^{0\prime -1}[1/\hat{c}^0(\theta,x)] \end{aligned}$$

for some δ, by homogeneity. Thus

$$\tilde{q}(\theta,x,v) = \frac{\tilde{c}^{0\,\prime\,-1}[1/\hat{c}^{0}(\theta,x)]}{\tilde{c}^{0\,\prime\,-1}[1/\hat{c}^{0}(\underline{\theta},x)]}.$$

That is, fixing x, $\tilde{q}(\cdot)$ is a fixed, positive, nondecreasing function of θ; in particular, $\tilde{q}(\cdot)$ does not vary with v.

Now, for any $\bar{q} \in \mathbb{R}_+$ and fixed x, let $\tilde{\theta} \equiv inf\,\{\theta \mid \tilde{q}(\theta,x,v) > \bar{q}\}$. Observe that $\tilde{\theta}$ is also independent of v. $\forall\, b\; C[0,\bar{Q}]$, let $v^{\tilde{q}}(b;\, x,v) \equiv v\{\theta \in \Theta \mid \tilde{q}(\theta,\, x,\, v) \in b\}$. Then

$$\begin{aligned}
v^{\tilde{q}}_{t+1}\Big[[0,q];\, x,\, v_t\Big] &= \int_{[0,\tilde{\theta}]} \tilde{q}(\theta,x,v_{t+1})v_{t+1}(d\theta) \\
&= \int_{[0,\tilde{\theta}]} \tilde{q}(\theta,x,v_t)v_{t+1}(d\theta) \\
&\geq \int_{[0,\tilde{\theta}]} \tilde{q}(\theta,x,v_t)v_t(d\theta) \\
&= v^{\tilde{q}}_t\Big[[0,\tilde{q}],\, x\Big].
\end{aligned}$$

||

A.11 Proof of Theorem 8

Let $\theta^0 < \theta^1$ and $\forall\, b \in \mathcal{B}$ write $\Psi^1[b \cap (\theta,\bar{\theta}],\, \cdot]$ as $\Psi^1_\theta(b,\cdot)$. The maximality of α implies

$$\begin{aligned}
&c^1[\alpha_{-q}(\theta^1,x,v),x] - c^1[\alpha_{-q}(\theta^0,x,v),x] \\
&\geq \{\Psi^0[\alpha(\theta^1,x,v),x] - \Psi^0[\alpha(\theta^0,x,v)]\} \\
&\times \beta\left\{\Big[\Psi^1([0,\theta^0],x,v) - 1\Big]\int_K V[\theta^0,x',\Phi(x,v)]\chi(dx',x)\right. \\
&\left. + \int_{K\times(\theta^0,\bar{\theta}]} V[\theta',x',\Phi(x,v)]\,\Psi^1_{\theta^0}(d\theta',x,v)\,\chi(d,x',x)\right\}. \qquad (*)
\end{aligned}$$

and

$$\begin{aligned}
&c^1[\alpha_{-q}(\theta^1,x,v),x] - c^1[\alpha_{-q}(\theta^0,x,v),x] \\
&\leq \{\Psi^0[\alpha(\theta^1,x,v),x] - \Psi^0[\alpha(\theta^0,x,v),x]\} \\
&\times \beta\left\{\Big[\Psi^1([0,\theta^1],x,v) - 1\Big]\int_K V[\theta^1,x',\Phi(x,v)]\chi(dx',x)\right.
\end{aligned}$$

$$+ \int_{K\times(\theta^1,\bar{\theta}]} V[\theta',x',\Phi(x,v)]\, \Psi^1_{\theta^1}(d\theta',x,v)\, \chi(dx',x)\Bigg\}; \qquad (**)$$

Combining (*) and (**), noting $\Psi^1_{\theta^1}((\theta^0,\theta^1],\cdot) = 0$, gives

$$\{\Psi^0[\alpha(\theta^1,x,v),x] - \Psi^0[\alpha(\theta^0,x,v),x]\}$$

$$\times \beta \left\{ \left[\Psi^1([0,\theta^1],x,v) - 1\right] \int_K V[\theta^1,x',\Phi(x,v)]\chi(dx',x) \right.$$

$$- \left[\Psi^1([0,\theta^0],x,v) - 1\right] \int_K V[\theta^0,x',\Phi(x,v)]\chi(dx',x)$$

$$\left. + \int_{K\times(\theta^0,\bar{\theta}]} V[\theta^1,x',\Phi(x,v)]\, [\Psi^1_{\theta^1}(d\theta',x,v) - \Psi^1_{\theta^0}(d\theta',x,v)]\chi(dx',x) \right\} \geq 0.$$

Since $\forall\ (x,v)$, $V(\theta^1,x,v) \geq V(\theta^0,x,v)$, and $\forall (b,x,v)$, $\Psi^1_{\theta^0}(b,x,v) \geq \Psi^1_{\theta^1}(b,x,v)$, the second factor in braces is negative. Therefore

$$\Psi^0[\alpha(\theta^1,x,v),x] \leq \Psi^0[\alpha(\theta^0,x,v),x]. \qquad (***)$$

Now, for any $\hat{\theta} \in \Theta$

$$\Psi^\Delta[(\hat{\theta},\bar{\theta}],\theta^0,x,v]$$
$$\equiv \Psi[(\theta^0+\hat{\theta},\bar{\theta}] \cap \Theta,\ \alpha(\theta^0,x,v),\ \theta^0,x,v]$$
$$= \Psi^0[(\theta^0,x,v),\ x]\ \Psi^1_{\theta^0}[(\theta^0+\hat{\theta},\bar{\theta}],x,v]$$
$$\geq \Psi^0[\alpha(\theta^1,x,v),x]\ \Psi^1_{\theta^1}[(\theta^0+\hat{\theta},\bar{\theta}],x,v]$$
$$\geq \Psi^0[\alpha(\theta^1,x,v),x]\ \Psi^1_{\theta^1}[(\theta^1+\hat{\theta},\bar{\theta}],x,v],$$
$$= \Psi^\Delta[(\hat{\theta},\bar{\theta}],\theta^1,x,v].$$

The penultimate in equality follows from (***) and $\forall\ (b,x,v), \Psi^1_{\theta^0}(b,x,v) \geq \Psi^1_{\theta^1}(b,x,v)$; the last follows from $(\theta^1+\hat{\theta},\bar{\theta}] \subset (\theta^0+\hat{\theta},\bar{\theta}]$. Thus, $\forall\ \hat{\theta} \in \Theta$, $\Psi^\Delta[(0,\hat{\theta}],\theta^0,x,v] \leq \Psi^\Delta[[0,\hat{\theta}],\theta^1,x,v]$.

The corollary sets $\hat{\theta}$=0. ||

A.12 Proof of Theorem 9

Proceeding as in the proof of Theorem 7, and removing conditioning on X_{t+1}, growth given θ', θ,x and v is

$$\int_K \left[q[\theta', x', \Phi(x, v)] / q[\theta, x, v]\right] \chi(dx', x) = \left[\frac{\hat{c}_1^0(\theta)}{\hat{c}_1^0(\theta')}\right]^{\delta_2} \int \left[\frac{P[x', \Phi(x, v)]}{P[x, v]}\right]^{\delta_1}$$

$$\left[\frac{\hat{c}_2^0(x)}{\hat{c}_2^0(x')}\right]^{\delta_2} \lambda'(dx', x)$$

for some $\delta_1 > 0$ and $\delta_2 > 0$, in which case the dependence of growth on θ and θ' is captured by the first factor on the right hand side. It thus suffices to show that the distribution of $\hat{c}_1^0(\theta) / \hat{c}_1^0(\theta')$ is stochastically declining in θ. For any $\xi \geq 1$, the probability that $\hat{c}_1^0(\theta) / \hat{c}_1^0(\theta') \leq \xi$ in equal to the probability that $\theta' \geq \hat{c}_1^{0^1}[\hat{c}_1^0(\theta) / \xi]$, or $\Psi^0[x(\theta, x, v), x, v]\, \Psi^1[(\hat{c}_1^{0^{-1}}[\hat{c}_1^0(\theta) / \xi], \bar{\theta}], x, v]$. The proof of Theorem 8 demonstrates that the first factor is declining in θ. The second factor is also declining in θ since the composition of $\hat{c}_1^{0^{-1}}$ and $\hat{c}_1^0$ is increasing.

A.13 Proof of Theorem 10

Let $\theta^0 \leq \theta^1$ and $\forall b \in \mathscr{B}$ write $\Psi^1[b \cap (\theta,\bar{\theta}],\cdot]$ as $\Psi^1_\theta(\cdot)$. Then, since $\Psi^1_\theta 1([\theta^0,\theta^1)) = 0$,

$$\begin{aligned}
&d(\theta^0,x,v_t,v_{t+1}) - d(\theta^1,x,v_t,v_{t+1}) \\
&= v_t([0,\theta^1]) - v_t([0,\theta^0]) + v_{t+1}([0,\theta^0])\Psi([0,\theta^0], \alpha(\theta^0,x,v_t),\theta^0,x,v_t) \\
&\quad - v_{t+1}([0,\theta^1])\Psi([0,\theta^1), \alpha(\theta^1,x,v_t),\theta,x,v_t) \\
&\quad + \int_{(\theta^0,\bar{\theta})} v_{t+1}([0,\theta']) \Big[\Psi^0[\alpha(\theta^0,x,v_t),x,v_t]\Psi^1_{\theta^0}(d\theta',x,v_t) \\
&\quad - \Psi^0[\alpha(\theta^1,x,v_t), x, v_t]\Psi^1_\theta 1(d\theta',x,v_t)\Big].
\end{aligned}$$

By Theorem 8, the final term on the right hand side is nonnegative. Also $\Psi([0,\theta^1],\cdot,\theta^1,\cdot) \geq \Psi([0,\theta^0],\cdot,\theta^0,\cdot)$. Thus

$$\begin{aligned}
&d(\theta^0,x,v_t,v_{t+1}) - d(,\theta^1,x_1,v_t,v_{t+1}) \\
&\geq v_t((\theta^0,\theta^1]) - \Psi([0,\theta^0], \alpha(\theta^0,x,v_t),\theta^0,x,v_t)\, v_{t+1}((\theta^0,\theta^1]) \\
&= v_t((\theta^0,\theta^1]) - \Psi([0,\theta^0], \alpha(\theta^0,x,v_t),\theta^0,x,v_t)
\end{aligned}$$

$$\times \int_{[0,\theta^1]} \Psi[(\theta^0,\theta^1],\ \alpha(\theta,x,v_t),\theta,x,v_t]v_t(d\theta)$$
$$\geq 0$$

by the condition in the Theorem. ||

A.14 Proof of Theorem 11

Fixing $\tilde{x}$, any strategy a, followed by all firms, is equivalent to the function ω: $T \times \Theta \times K^{\infty} \to A$ defined pointwise by $\omega_t(\theta,\tilde{x}) = a_t(\theta,x_t,v_t)$ where x_t is the t^{th} element in $\tilde{x}$, v_0 is given, and $v_{t+1} = \int_{\Theta} \Psi[\cdot, a(\theta,x_t,v_t),\theta,x_t,v_t]v_t(d\theta)$; in particular, let ω^e: $T \times \Theta \times K^{\infty} \to A$ be equivalent to the equilibrium strategy α. In this section, "strategy" will refer to functions from $T \times \Theta \times K^{\infty} \to A$. To emphasize the dependence of $\{v_t\}$ on ω and $\tilde{x}$, $\{v_t\}$ will be written $\{v_t(\omega,\tilde{x})\}$.

Writing $q_t(\theta,\tilde{x})$ for the first component of $\omega_t(\theta,\tilde{x})$, and $Q_t(\omega,\tilde{x}) = \int_{\Theta} q_t(\theta,\tilde{x})v_t(\omega,\tilde{x})(d\theta)$, period t surplus given ω and $\tilde{x}$ is

$$S_t(\omega,\tilde{x}) \equiv \int_0^{Q_t(x,\omega)} D(z,x_t)dz - \int_{\Theta} c[\omega(\theta,\tilde{x}),\theta,x_t]v_t(\omega,\tilde{x})(d\theta). \tag{A.11.1}$$

Recalling that $\tilde{\chi}(\cdot, x_0)$ is the probability measure on K^{∞} consistent with χ, given x_0, the social planner's expected surplus--$\hat{W}(a)$ in the text--is equivalent to

$$W(\omega) \equiv \int \sum_{t=0}^{\infty} \beta^t S_t(\omega,\tilde{x})\chi(d\tilde{x}, x_0). \tag{A.11.2}$$

Fix t', and let strategy ω' differ from ω^e only at t', and then only in that $\omega'_{-q} \geq \omega^e_{-q}$. For any $\varepsilon \in (0,1)$ the strategy $\varepsilon\omega' + (1-\varepsilon)\omega^e$ is feasible.

The following applies Theorem 1 of Luenberger (1969,p. 178). Assuming, temporarily, that the derivative exists,

$$\frac{d}{d\varepsilon}W[\varepsilon\omega' + (1-\varepsilon)\omega^e]\Big|_{\varepsilon=0} = \sum_{k=t'}^{\infty} \beta^k \int \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} S_k[\varepsilon\omega' + (1-\varepsilon)\omega^e,\tilde{x}]\chi(d\tilde{x},x_0). \tag{A.11.3}$$

Now

$$\frac{\partial S}{\partial \varepsilon}_{t'}[\varepsilon\omega' + (1-\varepsilon)\omega^e, \tilde{x}]\Big|_{\varepsilon=0} = -\int \nabla_\alpha c[\omega^e_{t'}(\theta,\tilde{x}),\theta,x_{t'}]'[\omega'_{t'}(\theta,\tilde{x}) - \omega^e_{t'}(\theta,\tilde{x})]\nu_{t'}(\omega^e_{t'},\tilde{x})(d\theta). \tag{A.11.4}$$

For $k > t'$,

$$\begin{aligned}\frac{\partial S}{\partial \varepsilon}_k[\varepsilon\omega' + (1-\varepsilon)\omega^e, \tilde{x}]\Big|_{\varepsilon=0} &= D_k[Q_k(\omega^e,\tilde{x}), x_k]\frac{d}{d\varepsilon}\int q^e_k(\theta,\tilde{x})\nu_k[\varepsilon\omega' + (1-\varepsilon)\omega^e, \tilde{x}](d\theta) \\ &\quad - \frac{d}{d\varepsilon}\int c[\omega^e_k(\theta,\tilde{x}), \theta, x_k]\nu_k[\varepsilon\omega' + (1-\varepsilon)\omega^e,\tilde{x}](d\theta), \\ &= \frac{d}{d\varepsilon}\int \pi_k(\omega^e,\omega^e,\theta,\tilde{x})\nu_k[\varepsilon\omega' + (1-\varepsilon)\omega^e, \tilde{x}](d\theta),\end{aligned} \tag{A.11.5}$$

and, in general, for any $(k,\omega^1,\omega^2,\theta,\tilde{x})$,

$$\pi_k(\omega^1,\omega^2,\theta,\tilde{x}) \equiv Q_k(\omega^1,\tilde{x})q^2_k(\theta,\tilde{x}) - c[\omega^2_k(\theta,\tilde{x}),\theta,x_k]$$

is defined as the net revenue earned at k by a firm following strategy ω^2 when all others select ω^1.

Substituting (A.11.4) and (A.11.5) in (A.11.3)

$$\begin{aligned}\frac{d}{d\varepsilon}W[\varepsilon\omega' + (1-\varepsilon)\omega^e]\Big|_{\varepsilon=0} &= \\ \int\Big\{-\beta^{t'}\int \nabla_\alpha\, c[\omega^e(\theta,\tilde{x})\theta,x_{t'}]'[\omega'_{t'}(\theta,\tilde{x}) - \omega^e_{t'}(\theta,\tilde{x})]\nu_{t'}(\omega^e,\tilde{x})d\theta \\ + \sum_{k>t}^{\infty}\beta^k\frac{d}{d\varepsilon}\int \pi_k(\omega^e,\omega^e,\theta,\tilde{x})\nu_k[\varepsilon\omega' + (1-\varepsilon)\omega^e,\tilde{x}](d\theta)\Big\}\chi(d\tilde{x}, x_0).\end{aligned} \tag{A.11.6}$$

Next, for any $\tau \geq t$, define $H_\tau(\cdot, \omega^1, \omega^2, \theta_t, \tilde{x})$ to be the probability measure on Θ at τ given θ_t and $\tilde{x}$, should a firm select ω^2 while all others utilise ω^1. Since

$$\nu_t(\omega,\tilde{x}) = \int H_t(\cdot,\omega,\omega,\theta,x)\nu_0(\omega,\tilde{x})(d\theta), \tag{A.11.7}$$

the maximum value of (4) that a firm may attain in equilibrium, given θ_0, is

$$\hat{V}(\theta_0) \equiv \max_{\omega}\left\{\sum_{t=0}^{\infty}\beta^t\int \pi_t(\omega^e,\omega,\theta,\tilde{x})H_t(d\theta,\omega^e,\omega,\theta_0,\tilde{x})\tilde{\chi}(d\tilde{x},x_0)\right\}. \tag{A.11.8}$$

Since a^e attains $\hat{V}(\theta_0)$ for each θ_0, it also attains $\int\hat{V}(\theta_0)\nu_0(\omega^e,\tilde{x})(d\theta_0)$. Thus using conditions (i) and

(ii) of the Theorem, the Gateaux derivative in the direction ω' must be zero:

$$0 = \frac{d}{d\varepsilon}\left\{\sum_{t=0}^{\infty} \beta^t \int \pi_t[\omega^e,\varepsilon\omega' + (1-\varepsilon)\omega^e,\theta,\tilde{x}]H_t[d\theta,\omega^e,\varepsilon\omega' + (1-\varepsilon)\omega^e, \theta_0,\tilde{x}]\nu_0(\omega^e,\tilde{x})(d\theta_0)\tilde{\chi}(d\tilde{x},x_0)\right\}.$$

Using (A.11.7),

$$0 = -\beta^{t'} \int \nabla_\alpha c[\omega^e(\theta,\tilde{x}),\theta,x_{t'}]'[\omega'_{t'}(\theta,\tilde{x}) - \omega^e_{t'}(\theta,\tilde{x})]\nu_{t'}(\omega^e,\tilde{x})(d\theta)$$

$$+ \sum_{k>t'} \beta^k \frac{d}{d\varepsilon} \int \pi_k(\omega^e,\omega^e,\theta,\tilde{x})H_k[d\theta,\omega^e,\varepsilon\omega' + (1-\varepsilon)\omega^e,\theta_0,\tilde{x}]\nu_0(\omega^e,\tilde{x})(d\theta_0)\tilde{\chi}(d\tilde{x},x_0). \quad \text{(A.11.9)}$$

Since (A.11.9) equals 0, it may be subtracted from the right hand side of (A.11.6) without altering its value. Thus

$$\frac{d}{d\varepsilon}W[\varepsilon\omega' + (1-\varepsilon)\omega^e]\Big|_{\varepsilon=0} = \sum_{k>t'} \beta^k \int \frac{d}{d\varepsilon}\Big\{\int \pi_k(\omega^e,\omega^e,\theta,\tilde{x})\nu_k[\varepsilon\omega' + (1-\varepsilon)\omega^e,\tilde{x}](d\theta)$$

$$- \int \pi_k[\omega^e,\omega^e,\theta,\tilde{x}]H_k[d\theta,\omega^e,\varepsilon\omega' + (1-\varepsilon)\omega^e,\theta_0,\tilde{x}]\nu_0(\omega^e,\tilde{x})(d\theta_0)\Big\}\chi(d\tilde{x},x_0). \quad \text{(A.11.10)}$$

Differentiation of (A.11.7) gives

$$\frac{d}{d\varepsilon}\nu_t[\varepsilon\,\omega' + (1-\varepsilon)\omega^e, \tilde{x}]\Big|_{\varepsilon=0}$$

$$= \int \Big\{\frac{d}{d\varepsilon}H_t[\cdot,\omega' + (1-\varepsilon)\omega^e,\omega^e,\theta,\tilde{x}] + \frac{d}{d\varepsilon}H_t[\cdot,\omega^e,\varepsilon\omega'+(1-\varepsilon)\omega^e,\theta,\tilde{x}]\Big\}\,\nu_0(\omega^e,\tilde{x})(d\theta).$$

Thus, since $\pi_k(\omega^e,\omega^e,\theta,\tilde{x})$ is independent of θ_0, (A.11.10) becomes

$$\frac{d}{d\varepsilon}W[\varepsilon\,\omega' + (1-\varepsilon)\omega^e]\Big|_{\varepsilon=0}$$

$$= \sum_{k>t'} \beta^k \int \frac{d}{d\varepsilon}\Big\{\int \pi_k(\omega^e,\omega^e,\theta,\tilde{x})H_k[d\theta,\varepsilon\omega'+(1-\varepsilon)\omega^e,\omega^e,\theta_0,\tilde{x}]\nu_0(\omega^e,\tilde{x})(d\theta_0)\Big\}\,\chi(d\tilde{x},x_0). \quad \text{(A.11.12)}$$

$$= \sum_{k>t'} \beta^k \int \frac{d}{d\varepsilon}\Big\{\max_{\omega} \int \pi_k(\omega^e,\omega,\theta,\tilde{x})H_k[d\theta,\varepsilon\omega'+(1-\varepsilon)\omega^e,\omega,\theta_0,\tilde{x}]\nu_0(\omega^e,\tilde{x})(d\theta_0)\Big\}\,\chi(d\tilde{x},x_0). \quad \text{(A.11.13)}$$

The second equality uses the envelope theorem and the interiority of ω^e implied by condition (i). Differentiability of W with respect to ε at $\varepsilon = 0$ follows if it can be shown that $\int \pi_k dH_k$ in (A.11.13) is, (a) differentiable for each k, and (b) bounded uniformly in k. Now (b) follows directly from (A.1.7); (a) follows because condition (iv) of the theorem, along with the definition of H_t preceeding (A.1.5), implies that H_{k+1} is differentiable if H_k is. Hence $dW/d\varepsilon$ exists at $\varepsilon = 0$.

Next, define $V^{\varepsilon}: T \times \Theta \times K \times K^{\infty} \longrightarrow \mathbb{R}$ pointwise by

$$V_t^{\varepsilon}(\theta,x,\tilde{x}) = \max_{\omega} \left\{\pi_t(\omega^e,\omega,\theta,\tilde{x}) + \beta\int V_{t+1}^{\varepsilon}(\theta',x',\tilde{x})\, \Psi\left[d\theta',\omega,\theta,x,\nu_t[\varepsilon\omega' + (1\text{-}\varepsilon)\, \omega^e,\tilde{x}]\right]\chi(x',x)\right\},$$

and note that both $\nu[\varepsilon\omega' + (1\text{-}\varepsilon)\omega^e,\tilde{x}] \succ \nu(\omega^e,\tilde{x})$ and $\Psi[\cdot,\omega,\theta,x,\nu[\varepsilon\omega' + (1\text{-}\varepsilon)\omega^e,\tilde{x}]] \succ \Psi[\cdot,\omega,\theta,x,\nu(\omega^e,\tilde{x})]$; thus V^{ε} is strictly increasing in ε, since $P(x,\nu) > \underline{D} > 0$. V^{ε} is the expected discounted value of profits, when profits are given by $\pi(\omega^e,\omega,\theta,\tilde{x})$, but learning is instead governed by $\Psi(\cdot,\omega,\theta,x,\nu[\varepsilon\omega' + (1\text{-}\varepsilon)\omega^e,\tilde{x}])$. Substitution in (A.11.12) gives

$$\frac{d}{d\varepsilon} W[\varepsilon\omega' + (1\text{-}\varepsilon)\omega^e] \Big|_{\varepsilon=0} = \beta^{t'+1} \int \frac{d}{d\varepsilon} V_{t'+1}(\theta',x',\tilde{x})$$

$$H_{t'+1}[d\theta',\varepsilon\omega' + (1\text{-}\varepsilon)\omega^e,\omega^e,\theta_0,\tilde{x}]\; \nu_0(\omega^e,\tilde{x})(d\theta_0)\tilde{\chi}(d\tilde{x}), \qquad \text{(A.11.14)}$$

$H_{t'+1}$ is independent of ε; i.e. learning at t' is a consequence of $\nu_{t'}$ and the firm's assumed choice $\omega^e_{t'}$. The remaining step in the proof yields a strictly positive lower bound for the derivative on the right hand side of (A.11.14).

Since $V^{\varepsilon}$is increasing in ε, and feasible actions are independent of ε,

$$V_{t'+1}^{\varepsilon}(\theta,x',\tilde{x}) - V_{t'+1}^{0}(\theta,x',\tilde{x}) \geq \beta \int V_{t'+2}^{0}(\theta',x',\tilde{x})\, [\Psi^{\varepsilon}(d\theta',\cdot) - \Psi^{0}(d\theta',\cdot)]\; \chi(x',x),$$

where $V^0 \equiv V^{\varepsilon}$ for $\varepsilon=0$, $\Psi^{\varepsilon}(\cdot) \equiv \Psi\left[\cdot,\omega^e,\theta,x,\nu[\varepsilon\omega' + (1\text{-}\varepsilon)\omega^e,\tilde{x}]\right]$ and $\Psi^0(\cdot) \equiv \Psi\left[\cdot,\omega^e,\theta,x,\nu(\omega^e,\tilde{x})\right]$. Using condition (iii) of the Theorem,

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int V_{t+2}^{0}(\theta',x',\tilde{x})[\Psi^{\varepsilon}(d\theta',\cdot) - \Psi^{0}(d\theta',\cdot)]\; \chi(x',x)$$

exists and is strictly positive. Thus, from (A.11.14), since $\frac{d}{d\varepsilon} W$ exists, it is also strictly positive.

The corollary is proved by first checking that if Ψ is independent of ν, $dW/d\varepsilon = 0$ at $\varepsilon=0$, where the variation allows arbitrary ω. Next, it is verified that if some ω^* yields $dW/d\varepsilon = 0$, the associated strategy is an equilibrium, say a^*. Finally, uniqueness of equilibrium gives $a^* = \alpha$. ||

A.15 Proof of Theorem 12

Since τ has density, say τ', on $(\theta,\bar{\theta}]$, the separability of production from information gathering implies:

$$\forall(\theta,x,v),\ (\eta(\theta,x,v),\ \mu(\theta,x,v)) = \underset{\eta,\mu}{argmax}\ -\ c^1(\eta,\mu,x)$$

$$+\ \beta \int_K \Big\{ V[\theta,x',\Phi(x,v)][1-\eta+\eta N(\theta,\theta,x)][1-\mu+\mu M(\theta,x,v)]$$

$$+ \int_{(\theta,\bar{\theta}]} V[\theta',x',\Phi(x,v)]\tau'(\theta',\eta,\mu,\theta,x,v)d\theta' \Big\}\chi(dx',x).$$

Total differentiation of the first order conditions associated with the maximization gives (after simplification):

$$\frac{\partial}{\partial\theta}\eta(\theta,x,v) \propto a_{\eta\mu}\, a_{\mu\theta} - a_{\mu\mu}\, a_{\eta\theta}$$

and

$$\frac{\partial}{\partial\theta}\mu(\theta,x,v) \propto a_{\eta\mu}\, a_{\eta\theta} - a_{\eta\eta}\, a_{\mu\theta}$$

where

$$a_{\eta\eta} = \frac{\partial^2 c^1}{\partial\eta^2} \le 0,$$

$$a_{\mu\mu} = \frac{\partial^2 c^1}{\partial\mu^2} \le 0,$$

$$a_{\eta\theta} = \beta \int_K \frac{\partial V}{\partial\theta}[\theta,x',\Phi(x,v)][N(\theta,\theta,x)-1][1-\mu + \mu M(\theta,x,v)]\chi(dx',x) \le 0,$$

$$a_{\mu\theta} = \beta \int_K \frac{\partial V}{\partial\theta}[\theta,x',\Phi(x,v)][M(\theta,x,v)-1][1-\eta + \eta N(\theta,\theta,x)]\chi(dx',x) \le 0,$$

and $a_{\eta\mu}$ is the expression in restriction (ii) of the Theorem, whose conclusion is immediate. ||

A.16 Evolutions in the 3-State Examples

a) Equilibrium

t	v_t^0	v_t^0	p_t	$\eta(\theta_0,v_t)$	$\mu(\theta_0,v_t)$	$\eta(\theta_1,v_t)$	$\mu(\theta_1,v_t)$
1.	1.000	0.000	0.571	0.323	0.000	0.439	0.000
2.	0.981	0.016	0.530	0.280	0.020	0.381	0.015
3.	0.964	0.030	0.497	0.246	0.037	0.335	0.026
4.	0.948	0.042	0.470	0.217	0.052	0.296	0.035
5.	0.933	0.054	0.446	0.191	0.066	0.262	0.043
6.	0.918	0.066	0.424	0.168	0.080	0.231	0.049
7.	0.903	0.077	0.404	0.146	0.093	0.203	0.055
8.	0.887	0.090	0.384	0.125	0.107	0.175	0.059
9.	0.870	0.103	0.364	0.105	0.122	0.149	0.063
10.	0.850	0.117	0.345	0.086	0.138	0.125	0.065
11.	0.828	0.134	0.325	0.069	0.155	0.103	0.065
12.	0.803	0.153	0.304	0.053	0.174	0.084	0.063
13.	0.773	0.175	0.283	0.042	0.195	0.072	0.055
14.	0.737	0.202	0.261	0.039	0.209	0.069	0.046
15.	0.694	0.234	0.239	0.037	0.210	0.067	0.051
16.	0.648	0.267	0.219	0.030	0.208	0.056	0.070
17.	0.599	0.302	0.201	0.022	0.204	0.043	0.088
18.	0.550	0.337	0.184	0.017	0.205	0.035	0.097
19.	0.498	0.371	0.170	0.015	0.212	0.033	0.096
20.	0.445	0.405	0.156	0.014	0.201	0.032	0.088
21.	0.395	0.436	0.145	0.013	0.197	0.030	0.092
22.	0.347	0.463	0.135	0.012	0.201	0.027	0.106
23.	0.302	0.485	0.126	0.010	0.209	0.024	0.128
24.	0.258	0.502	0.118	0.007	0.219	0.021	0.153
25.	0.215	0.512	0.110	0.007	0.212	0.020	0.149
26.	0.180	0.514	0.103	0.006	0.204	0.018	0.145
27.	0.150	0.509	0.097	0.006	0.198	0.016	0.144
28.	0.124	0.499	0.092	0.005	0.193	0.014	0.144
29.	0.103	0.484	0.088	0.004	0.186	0.013	0.140
30.	0.086	0.465	0.084	0.004	0.180	0.011	0.137
31.	0.072	0.443	0.080	0.004	0.174	0.010	0.134
32.	0.060	0.420	0.077	0.003	0.170	0.009	0.132
33.	0.051	0.395	0.074	0.003	0.165	0.008	0.129
34.	0.043	0.370	0.072	0.003	0.161	0.008	0.127
35.	0.036	0.345	0.070	0.003	0.158	0.007	0.125
36.	0.031	0.320	0.068	0.002	0.155	0.006	0.124
37.	0.026	0.295	0.066	0.002	0.152	0.006	0.122
38.	0.022	0.272	0.065	0.002	0.149	0.005	0.121
39.	0.019	0.250	0.063	0.002	0.147	0.005	0.119
40.	0.016	0.229	0.062	0.002	0.145	0.005	0.118
41.	0.014	0.209	0.061	0.002	0.144	0.005	0.118
42.	0.012	0.190	0.060	0.002	0.143	0.005	0.117
43.	0.010	0.172	0.059	0.002	0.141	0.004	0.116
44.	0.009	0.156	0.058	0.002	0.140	0.004	0.115
45.	0.008	0.141	0.058	0.002	0.139	0.004	0.115
46.	0.007	0.128	0.057	0.002	0.138	0.004	0.114
47.	0.006	0.115	0.056	0.002	0.138	0.004	0.114

a) Equilibrium (cont'd.)

t	v_t^0	v_t^0	p_t	$\eta(\theta_0,v_t)$	$\mu(\theta_0,v_t)$	$\eta(\theta_1,v_t)$	$\mu(\theta_1,v_t)$
48.	0.005	0.104	0.056	0.002	0.137	0.004	0.113
49.	0.004	0.093	0.056	0.002	0.136	0.004	0.113
50.	0.004	0.084	0.055	0.002	0.136	0.004	0.113
51.	0.003	0.075	0.055	0.002	0.135	0.004	0.112
52.	0.003	0.067	0.054	0.002	0.135	0.004	0.112
53.	0.002	0.060	0.054	0.002	0.134	0.004	0.112
54.	0.002	0.054	0.054	0.002	0.134	0.004	0.111
55.	0.002	0.048	0.054	0.002	0.134	0.004	0.111
56.	0.002	0.043	0.054	0.002	0.133	0.004	0.111
57.	0.001	0.039	0.053	0.002	0.133	0.003	0.111
58.	0.001	0.035	0.053	0.002	0.133	0.003	0.111
59.	0.001	0.031	0.053	0.002	0.133	0.003	0.111
60.	0.001	0.028	0.053	0.002	0.132	0.003	0.111

b) Optimum

t	v_t^0	v_t^0	p_t	$\eta(\theta_0,v_t)$	$\mu(\theta_0,v_t)$	$\eta(\theta_1,v_t)$	$\mu(\theta_1,v_t)$
1.	1.000	0.000	0.571	0.567	0.000	3.817	0.000
2.	0.966	0.028	0.502	0.436	0.046	2.943	0.376
3.	0.939	0.047	0.448	0.355	0.076	2.394	0.618
4.	0.915	0.061	0.402	0.293	0.100	1.976	0.805
5.	0.891	0.073	0.363	0.242	0.120	1.627	0.966
6.	0.866	0.083	0.327	0.197	0.139	1.320	1.115
7.	0.840	0.092	0.293	0.156	0.158	1.040	1.265
8.	0.810	0.098	0.260	0.118	0.180	0.774	1.434
9.	0.777	0.101	0.228	0.079	0.207	0.502	1.654
10.	0.737	0.098	0.195	0.044	0.238	0.269	1.821
11.	0.689	0.087	0.164	0.031	0.258	0.205	1.658
12.	0.633	0.070	0.137	0.017	0.276	0.137	1.418
13.	0.568	0.053	0.116	0.007	0.285	0.076	1.025
14.	0.498	0.041	0.100	0.000	0.282	0.030	0.493
15.	0.427	0.038	0.089	0.000	0.274	0.024	0.439
16.	0.361	0.033	0.080	0.000	0.265	0.019	0.387
17.	0.299	0.029	0.073	0.001	0.257	0.014	0.341
18.	0.246	0.024	0.068	0.001	0.250	0.011	0.304
19.	0.199	0.020	0.064	0.001	0.266	0.010	0.296
20.	0.157	0.017	0.061	0.002	0.280	0.009	0.288
21.	0.120	0.014	0.059	0.002	0.291	0.009	0.281
22.	0.089	0.011	0.057	0.003	0.301	0.008	0.275
23.	0.065	0.008	0.056	0.003	0.308	0.008	0.271
24.	0.046	0.006	0.054	0.003	0.314	0.007	0.267
25.	0.032	0.005	0.054	0.003	0.318	0.007	0.265
26.	0.022	0.004	0.053	0.003	0.321	0.007	0.263
27.	0.015	0.003	0.053	0.004	0.323	0.007	0.261
28.	0.010	0.002	0.053	0.004	0.324	0.007	0.261

b) Optimum (cont'd.)

t	v_t^0	v_t^0	p_t	$\eta(\theta_0,v_t)$	$\mu(\theta_0,v_t)$	$\eta(\theta_1,v_t)$	$\mu(\theta_1,v_t)$
29.	0.007	0.002	0.052	0.004	0.325	0.007	0.260
30.	0.005	0.001	0.052	0.004	0.326	0.006	0.260
31.	0.003	0.001	0.052	0.004	0.326	0.006	0.259
32.	0.002	0.001	0.052	0.004	0.326	0.006	0.259
33.	0.001	0.000	0.052	0.004	0.326	0.006	0.259
34.	0.001	0.000	0.052	0.004	0.326	0.006	0.259
35.	0.001	0.000	0.052	0.004	0.327	0.006	0.259
36.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
37.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
38.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
39.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
40.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
41.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
42.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
43.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
44.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
45.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
46.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
47.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
48.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
49.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
50.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
51.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
52.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
53.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
54.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
55.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
56.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
57.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
58.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
59.	0.000	0.000	0.052	0.004	0.327	0.006	0.259
60.	0.000	0.000	0.052	0.004	0.327	0.006	0.259

References

Ash, Robert B., *Real Analysis and Probability*, New York: Academic Press, 1972.

Becker, Gary S., "Family Economics and Macro Behavior," *American Economic Review*, 78 (March 1988), 1-13.

Bergin, D. and Bernhart, D., "Anonymous Sequential Games with Aggregate Uncertainty", November 1989.

Chari, V.V. and Hopenhayn, Hugo, "Vintage Human Capital, Growth, and the Diffusion of New Technology," Federal Reserve Bank of Minneapolis, Working Paper No. 375, November 1989.

Chung, K.L., *A Course in Probability Theory*. New York. Academic press, 1968.

Chow, Y.S., et al., *Great Expectations*: *The Theory of Optimal Stopping*, Boston: Houghton-Mifflin, 1971.

Evans, David, "Tests of Alternative Theories of Industry Evolution", *Journal of Political Economy* (August 1987).

Evenson, Robert and Yoav Kislev, "Research and Products -- ritz in Wheat and Maize", *Journal of Political Economy* (1973): 1309-29.

Gort, Michael, "Analysis of Stability and Change in Market Shares", *Journal of Political Economy* 71, no. 1 (February 1963): 51-63.

Gort, Michael, and Klepper, Steven, "Time-Paths in the Diffusion of Product Innovations", *Economic Journal* 92 (September 1982): 630-53.

Griliches, Zvi "Hybrid Corn: An Exploration in the Economics of Technical Change," *Econometrica* 25 (1957), 501-22.

Griliches, Zvi "The Sources of Measured Productivity Growth: U.S. Agriculture 1940-1960," *Journal of Political Economy* 71 (1963), 331-46.

Hall, Bronwyn, "The Relationship Between Firm Size and Firm Growth in the U.S. Manufacturing Sector", *Journal of Industrial Economics* 35, no. 4 (June 1987): 583-606.

Hart, P.E., and Prais, S.J., "The Analysis of Business Concentration: A Statistical Approach", *Journal of the Royal Statistical Society*, 119, pt. 2 (1956), 150-191.

Holmstrom, Bengt, "Moral Hazard in Teams", *Bell Journal of Economics* 12, no. 2 (Autumn 1982): 324-40.

Interstate Commerce Commission, Bureau of Transport Economics and Statistics, *Study of Railroad Motive Power*, Washington, D.C., 1950.

Jovanovic, B. and Rosenthal, R.W., "Anonymous Sequential Games", *Journal of Mathematical Economics* 17 (1988), 77-88.

Jovanovic, Boyan and MacDonald, Glenn M., "Competitive Diffusion," September 1988.

Jovanovic, Boyan and MacDonald, Glenn M., "The Life Cycle of a Competitive Industry: Theory and Measurement," Rochester Center for Economic Research, Working Paper 237, July 1990.

Kamien, Morton I. and Schwartz, N. L. "Timing of Innovations Under Rivalry," *Econometrica*, 40 (Jan. 1972), 43-60.

Luenberger, David G., *Optimization by Vector Space Methods*. New York: John Wiley & Sons, 1969.

Mansfield, Edwin, "Intrafirm Rates of Diffusion of an Innovation", *Review of Economics and Statistics* 65, no. 4 (November 1963): 348-49.

Mansfield, E. et al., "Social and Private Rates of Return from Industrial Innovations," *Quarterly Journal of Economics*, 16 (May 1977), 221-40.

Mansfield, Edwin. et al, "Imitation Costs and Patents: An Empirical Study", *Economic Journal* 91 (December 1981): 907-18.

Mas Colell, Andreu, "On a Theorem of Schmeidler", *Journal of Mathematical Economics* 13 (1984), 201-6.

Nelson, Richard R. "The Role of Knowledge in R & D Efficiency," *Quarterly Journal of Economics*, 97 (August 1982), 453-70.

Prescott, Edward and Visscher, M. "Organizational Capital," *Journal of Political Economy*, 88 (May 1980), 446-61.

Saving, Thomas, "Estimation of Optimum Size of Plant by the Survivor Technique", *Quarterly Journal of Economics* 75 (November 1961): 569-607.

Scherer, Frank, *Industrial Market Structure and Economic Performance*, Second Edition, Chicago: Rand McNally 1980.

Scherer, F. M. "Research and Development Resource Allocation Under Rivalry," *Quarterly Journal of Economics*, 81 (August 1967), 359-94.

Schmookler, J., *Invention and Economic Growth*, Cambridge: Harvard University Press, 1966.

Schumpeter, Josef, *The Theory of Economic Development.* Cambridge, Mass.: Harvard University Press, 1934.

Schumpeter, Josef, *Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalist Process*, 2 Vols. New York: McGraw Hill, 1939.

Singh, Ajit, and Whittington, Geoffrey, "The Size and Growth of Firms", *Review of Economic Studies* (1975): 15-26.

Solow, Robert M., "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, 39 (1957), 312-20.

Solow, Robert M., "Growth Theory and After," *American Economic Review*, 78 (June 1988), 307-17.

Spence, A.M., "Cost Reduction, Competition, and Industry Performance," *Econometrica*, 52 (January 1984), 101-21.

Stigler, George, "The Economies of Scale", *Journal of Law and Economics* 1 (October 1958): 54-71.

Telser, Lester G. "A Theory of Innovation and Its Effects," *Bell Journal of Economics*, 13 (Spring 1982), 69-92.