Social Systems Research Institute

University of Wisconsin – Madison.

SSRI Workshop Series

EMPIRICAL IMPLICATIONS OF ALTERNATIVE MODELS OF FIRM DYNAMICS

Ariel Pakes
Richard Ericson

8803

December 1987

*Discussions with Buz Brock, Gary Chamberlain, Art Goldberger, and Boyan Jovanovic have been very helpful. Steve Berry provided first rate programming assistance. This research was funded by the NSF (grant SES-8520261) and the Graduate School of the University of Wisconsin. All errors remain, of course, our own responsibility.

Abstract

This paper considers two models for analyzing the dynamics of firm behavior that allow for idiosyncratic (or firm-specific) sources of uncertainty, and discrete outcomes (exit and/or entry). Models with these characteristics are needed for the structural econometric analysis of several economic phenomena, including the behavior of capital markets when there are significant failure probabilities, and the analysis of productivity movements in industries with large amounts of entry and exit. In addition, these models provide a means of correcting for the self-selection induced by liquidation decisions in empirical studies of firms responses to alternative policy and environmental changes. It is shown that both models have nonparametric implications - implications that depend only on basic behavioral assumptions and mild regularity conditions on the functional forms of interest - that can be taken directly to data. This circumvents the need for the computationally difficult, and functional-form specific, estimation algorithms that have been needed for analyzing stochastic control models with discrete outcomes in the past. One difference between the two models corresponds to the distinction between heterogeneity and an ergodic form of state-dependence (a form in which the effect of being in a state in a particular period erodes away as time from that period lapses). So we develop a test for this difference based on ϕ-mixing conditions. The paper concludes by checking for the implications of the two models on an eight-year panel of Wisconsin firms. We find one model to be consistent with the data for manufacturing, and the other to be consistent with the data for retail trade.

1. Introduction

In this, and in a companion piece (see Ericson and Pakes, 1987), we consider structural econometric models for analyzing the dynamics of firm behavior that allow for idiosyncratic, or firm-specific, uncertainty, and discrete events (exit and/or entry). Our reason for providing an empirical framework with these features are twofold. First, the nature of uncertainty, and its relationship to exit and/or entry, is at the heart of several issues we, as economists, try to analyze. Examples include the analysis of capital markets when there are diverse possible outcome paths and significant failure probabilities; the evolution of the size distribution of the firms in an industry; and the analysis of industry supply (or productivity) changes when more efficient firms thrive and grow, and less efficient contract and, in the extreme case, exit. The second reason for studying models that allow for uncertainty and exit is that some allowance has to be made for these phenomena before we can get an accurate empirical picture of firms' responses to any policy or environmental change. Table 1 illustrates why this is so.

The table provides information on the fraction of firms operating in Wisconsin in 1978 that were liquidated by 1986 (more details on the data will be given in Section 5). Firms are classified as liquidated only if they physically closed down (changes of ownership are treated separately). If we were to use these data to build a panel of firms to follow the impact of some (say) policy change, we would, at least traditionally, start from the 1978 cross-section and then construct the panel by eliminating those firms not in operation over the entire eight-year period. Column 5 shows that this procedure would lose a third of the firms due to liquidations, and column 6 shows that this third would account for about a fifth of the jobs in 1978. If we

Table 1. Liquidation in the 1978/86 Wisconsin Panel[a]

Sector	1 # Firms Active in 1978	2 % of all Firms in 1978	3 Employ-ment in 1978	4 % of all Employ-ment in 1978	5 % of Firms Active in 1978 Liquidated by 1986	6 % of 1978 Employment in firms Liquidated by 1986	7 % of 1978 Employment in firms with > 50 Employees	8 % of 1978 Firms with > 50 Employees Liquidated by 1986
Wholesale	7,251	17	85,135	8	29.5	16.0	35	10.5
Retail	22,568	51	316,498	30	39.5	26.0	45	17.0
Manu-facturing	6,987	16	550,200	52	24.0	13.0	87	13.0
Eating and Drinking	7,466	17	103,192	10	44.5	29.5	36	18.5
Total	44,272	100%	1,055,205	100%	36.5	19.0	65	14.5
Substitute "transferred out" for "liquidated" in colums 5, 6, and 8.					8.5	11.1		10.5
Substitute "either transferred out or liquidated" for "liquidated" in columns 5, 6, and 8					45.0	30.1		25.0

[a] If a firm ever undergoes a change in legal status (a change in ownership) it will not be counted as a liquidation thereafter (even though the resulting firm may have liquidated). Firms in the construction and service sectors in 1978 have been excluding from this sample. These firms accounted for about 340,000 jobs.

decided to consider only the larger of the 1978 firms, say those with more than 50 employees (and as column 7 shows, this is a selection which, by itself, omits over a third of the 1978 jobs), liquidation would be somewhat less prevalent, but would still cause an attrition rate of about 15 percent. The last two rows of the table give an indication of the extent of changes in ownership in this data (this includes mergers and acquisitions). To the extent that the pre and post change firms cannot be spliced together, changes in ownership also generate attrition. It is a relatively more important source of attrition among larger firms, but even if we confine ourselves to firms with over 50 employees, and assume that all the changes in ownership result in attrition, changes of ownership would still only account for 40 percent of total attrition (liquidation accounts for the rest). Note that, when taken together, liquidations and changes of ownership would cause the attrition of almost half the firms in the 1978 sample, and of about a quarter of those with more than 50 employees.

If liquidation decisions were independent of the economic phenomena typically being investigated, then the omission of the liquidated firms from the sample might lead to an imprecise, but would not lead to an inconsistent, description of the phenomena of interest. This is, however, hardly likely. Firms terminate their activities when they perceive adverse changes in the distribution of their future profit streams. The phenomena we typically want to investigate involve the actual profitability (and productivity) changes resulting from alternative policy and environmental changes. If there is any relationship at all between perceptions and realizations we will, by eliminating those firms which liquidate, omit precisely those firms for whom the events in question are likely to have had a particularly negative impact. That is, we will tend to omit one tail of the distribution of responses we set out to study.[1]

To control for the selection induced by the liquidation process we need a model that explains why firms operating in similar environments develop differently - a model with idiosyncratic outcomes that allows for exit. At least two such models are currently available, and each will, no doubt, prove more useful in approximating the characteristics of different industries in different time periods. This paper provides a simple set of procedures which enable the researcher to determine whether either of them might be relevant for the problem at hand.

The first model considered here is a model with passive or Bayesian learning. Firms are endowed at birth with an unknown value of a time-invariant profitability parameter which determines the distribution of its profits thereafter. Past profit realizations contain information on the value of the parameter which determines the distribution of possible future profit streams, and this fact is used by management to form a probability distribution over future net cash flows (see Jovanovic, 1982). The second, or active learning, model assumes the firm knows the current value of the parameter that determines the distribution of its profits, but that the value of that profitability parameter changes over time in response to the stochastic outcomes of the firm's own investments, and those of other actors in the same market (see Ericson and Pakes, 1987). In both models firms act so as to maximize the expected discounted value of future net cash flow, and in both cases optimal behavior generates a set of stopping states; i.e. outcomes which, if realized, would induce the firm to exit. Moreover, both models are 'complete' in the sense that if we were willing to append a set of precise functional form assumptions to them, they would produce frameworks rich enough to take directly to data.

The strategy of appending precise functional form assumptions and then using their implications to structure the data, is the strategy taken in all of the recent econometric literature on analyzing stochastic control models involving discrete outcomes (see Miller, 1984; Wolpin, 1984; Pakes, 1986; and Rust, 1987). Its success depends upon, among other diverse factors, the extent of prior information on the relevance of alternative assumptions. We eschew it here because there is not a great deal of a priori information on either which of the models (if any) is appropriate for different data sets or on the relevance of alternative functional form assumptions. Moreover, just as in all the previous literature on discrete choice optimal stochastic control models, were we to estimate fully parametric versions of these models we would have to build a different estimation algorithm for each form estimated. This makes it difficult, if not impossible, to examine the robustness of the major empirical results to changes in the specification of the model.

The alternative strategy we choose is to look for empirical implications of the different models that depend only on the models' basic behavioral assumptions, and some mild regularity conditions on the relevant functional forms. Precisely because these 'nonparametric' implications have to be valid for a variety of functional forms, they cannot require functional form specific estimation and testing algorithms. Consequently, there are computationally simple ways of checking whether they are consistent with the data. Therefore, in addition to not being dependent on particular functional form assumptions, our strategy is easy to implement. On the other hand, the nonparametric procedures provided here do not produce precise values for alternative response parameters. Their goals are only to: 1) provide a low cost way of obtaining (we hope reliable) information on which of the alternative models seems relevant for the problem at hand, and 2) to provide a reduced form empirical characterization of

the data which is easily interpretable and can be used to indicate which of the different ad hoc procedures for correcting for the selection problem induced by the liquidation process is more appropriate.

One of the nonparametric differences between the two models corresponds to the distinction between heterogeneity and state dependence that has played so large a role in labor econometrics (see Heckman, 1983; Chamberlain, 1984; and Heckman and Singer; 1984). In particular the passive learning model implies that the stochastic process generating the size of a firm is characterized by a generalized form of heterogeneity, while the active learning model implies that this stochastic process is generated by a quite general form of state dependence. Theory restricts the state dependence in the active learning model to have ergodic characteristics; i.e. the effect of being in a state in a particular period erodes away as time from that period lapses. So we develop a test for the distinction between heterogeneity and ergodic forms of state dependence based on ϕ-mixing conditions. The test is simple, intuitive, and seems to be able to distinguish between the two models on panel data sets the size of the ones used here (these follow about 400 observations over eight years).

In particular, we find both the ϕ-mixing test, and an analysis of the evolution of the size distribution of firms in a cohort, suggest that one model is consistent with the data for manufacturing, while the other seems consistent with the data for retail trade. The importance of this result is twofold. First the different models have distinctly different implications for the manner and the extent to which firm-specific uncertainties get resolved over time, and hence for the way in which issues related to these uncertainties ought to be analyzed. Second, the two models imply different determinants for the probability of liquidation, and hence different procedures for correcting for

liquidation induced attrition in the analysis of firm's responses to alternative policy and environmental changes.

Section 2 of the paper provides the passive learning model and then derives its nonparametric implications. Section 3 does the same for the active learning model. In section 4 we develop appropriate estimation and testing procedures. Section 5 begins with a description of the Wisconsin panel, and then examines various subsets of it for the implications of the two models. Brief concluding remarks close the paper.

Notation

The distribution of any random variable, say x, conditional on any event, say z, is denoted $P_x(\cdot \mid z)$, and its density (with respect to the implied dominating measure) by $p_x(\cdot \mid z)$. Superscripts denote the vector of all prior realizations of a process, and subscripts denote a particular value, so $x^t = (x_1, \ldots, x_t)$. Weak vector inequalities are interpreted element by element, but a strong vector inequality means only that at least one of the element by element inequalities is strong. Z will be used for the generic set, z for a member of that set, and diag[x] for a diagonal matrix with x on the principal diagonal. Lemmas, theorems, examples etc. will be numbered in one consecutive ordering within each section. They are referred to in the following sections with a section prescript.

Section 2. Passive Learning.

This section considers models in which each firm is endowed with a time-invariant characteristic which determines the distribution of its profits, but whose value is not known to management at the time the firm begins operation. Models of industries composed of firms which learn about an unknown profitability parameter have been provided by Jovanovic (1982) and Lippman and Rumelt (1982). Following Jovanovic (1982), we consider a Bayesian learning process. At entry the firm believes the value of its characteristic, say θ, is a random draw from some known distribution. Each period the firm is in operation it obtains a realization from the distribution of profits conditional on the true value of its θ. These realizations are used to compute a sequence of posterior distributions. The posterior available in each period is used as a basis for decision-making in that period. The decisions of interest are whether to produce at all and, if so, at what scale. If the firm does decide not to produce it sells off its assets and exits, never to reappear again. Note that in this model learning is passive in the sense that information is obtained as a costless byproduct of operating. Perhaps the clearest analogy is to the operation of a retail outlet. The outlet learns whether its neighborhood will support its product, and, if so, at which scale of operation.

Jovanovic (1982) focuses on establishing the existence of a perfect foresight equilibrium for a homogeneous product industry composed of firms which operate in this manner. We focus on the implications of the learning process on the evolution of cohorts of firms, where cohorts are defined by entry dates. In particular we shall look for empirical implications that rely on the nature of the learning process, and only some mild regularity conditions on the form of the profit function and the underlying distributions of interest.

Later we compare these implications to data in an attempt to identify those sectors in which this form of learning process seems relevant.

2.1 The Model

It will be assumed that each entrant is endowed with a value of θ which, in turn, determines the distribution of a payoff relevant random variable η, say $P_\eta(\cdot \mid \theta)$. To motivate our assumptions, consider the example of a homogeneous product industry of price-takers whose production efficiencies are subject to random perturbations so that profits in period t are $\pi_t = \alpha_t \eta_t F(\ell_t) - \omega_t' \ell_t$ where; ℓ_t is a vector of input quantities, ω_t provides their prices, $F(\cdot)$ is a concave production function, $\{\eta_j\}$ is a sequence of independent and identically distributed (i.i.d.) random variables, and α_t is the product price. Assume η_t is known at the time ℓ_t is chosen. Then

$$\pi_t = \pi(\eta_t; \omega_t, p_t) = \max_{(\ell_t)} \{\alpha_t \eta_t F(\ell_t) - \omega_t' l_t\},$$

and $\pi(\eta; \omega_t, p_t)$ is an increasing function of η. In a perfect foresight equilibrium future prices will be known, so that if θ were also known the distribution of future profits could be calculated directly from $P_\eta(\cdot \mid \theta)$. Since management does not know θ it is assumed to summarize its beliefs about that parameter in terms of a probability distribution over the possible values of θ. At entry, management only knows that θ is a random draw from $G_0(\theta)$. The first period produces an η which management uses, together with Bayes law, to update its prior $[G_0(\theta)]$ and form a posterior which is then used to make second period decisions. If the firm stays in operation, this updating process continues and decisions are made on the basis of the sequence of updated posteriors.

As the example illustrates, the model will require at least four primitives; a sequence of random variables, a class of distributions for those random variables indexed by θ, a prior distribution for θ, and a payoff function.

Before introducing these primitives we need a way of comparing distribution functions; i.e. we need an interpretation for the statement that one value of θ is 'better than' another. We shall assume that the family of distributions formed from different values of θ can be ordered in the likelihood ratio sense defined below. This ensures that higher realizations of the payoff relevant η lead to Bayesian posteriors for θ that assign larger probability to higher values of θ (see below, and Milgrom 1981).

1. Definition (likelihood ratio ordering, or $\gtrsim_{\ell r}$)

Let $P_1(\cdot)$ and $P_2(\cdot)$ be two distributions possessing densities $p_1(\cdot)$ and $p_2(\cdot)$ (with respect to some dominating measure), and with support, Z^k, a compact subset of $\mathbb{R}^k$, k-dimensional Euclidean space. We will say that P_1 likelihood ratio dominates P_2, in the strong sense, and write $P_1 \gtrsim_{\ell r} P_2$, if and only if,

$$p_1(z_1)p_2(z_2) - p_2(z_1)p_1(z_2) > 0,$$

whenever $z_1 > z_2$, and $p_1(z_1)$ or $p_2(z_2) > 0$, $z_1, z_2 \in Z^k$. If weak inequalities replace the strong inequalities in this definition, we will say that P_1 likelihood ratio dominates P_2 in the weak sense, and write $P_1 \gtrsim_{\ell rw} P_2$. [ ]

If $P_1 \gtrsim_{\ell r} P_2$ then, for any two possible values of z, the ratio of the probabilities of a larger to the smaller z value is always higher for P_1; i.e., P_1 is more likely to have generated the higher z value. The following lemma points out that $\gtrsim_{\ell r}$ is a stronger criteria for ordering distribution functions than the more familiar first order stochastic dominance criteria.

2. Lemma (likelihood ratios and stochastic dominance).

Say P_1 stochastically dominates P_2, and write $P_1 \gtrsim_s P_2$, if and only if

for every nondecreasing nonconstant function, $h(\cdot)$, such that $\int h(\zeta)P_1(d\zeta) < \infty$,

$$\int h(\zeta)P_1(d\zeta) > \int h(\zeta)P_2(d\zeta).$$

Then,

$$P_1 \gtrsim_{\ell r} P_2, \text{ implies, } P_1 \gtrsim_s P_2.$$

If weak inequalities replace the strong inequalities in this definition we say that P_1 stochastically dominates P_2 in the weak sense, and write $P_1 \gtrsim_{sw} P_2$. $P_1 \gtrsim_{\ell rw} P_2$ implies $P_1 \gtrsim_{sw} P_2$.

Proof See Ross (1982), Appendix 1, 3.1, and 4.1. [ ]

Assumption 3 provides the primitives of the passive learning model and endows them with some regularity conditions. It generalizes the assumptions used in our example. In particular the example assumed that conditional on a $\theta \epsilon \Theta$, the sequence of payoff relevant random variables, $\{\eta_t\}$, are independently and identically distributed (i.i.d.) over time. Then the joint distribution of the sequence $\{\eta_t\}$ conditional on a $\theta \epsilon \Theta$ is entirely described by the single distribution, $P_\eta(\cdot|\theta)$. Though the i.i.d. case is easy to deal with, it produces a host of very strong empirical implications which are a result of the i.i.d. assumption and not of the logic of the passive learning model per se. We, therefore, allow for dependence in the stochastic process generating $\{\eta_t\}$ conditional on θ. In (3.i) we assume only that the marginal distribution of η_t conditional on θ is stationary (does not depend on time), and that the conditional distribution of η_t (conditional on past η-realizations) satisfies the condition that higher past values of η are at least as likely to lead to higher future values of η. (3.ii) insures that higher values of θ are better in the ℓr-sense; i.e. it insures that for any t, higher values of the vector

$\eta^t = (\eta_1,\ldots,\eta_t)$ are more likely to be generated by larger θ values. (3.iv) provides the profit and size functions. It is important that both be increasing in η.[2]

3. Assumption (primitives of the model)

(i) (η_t) is a sequence of payoff relevant random variables (a stochastic process) whose joint distribution, say $\underline{P}(\theta)$, is indexed by a $\theta\epsilon\Theta$, where Θ is a compact subset of $\mathbb{R}_+$. The marginal distribution of η_t is stationary and is denoted by $P_\eta(\cdot|\theta)$, while its conditional distribution satisfies a weak ℓr-ordering in realizations of η^{t-1}, say n^{t-1}: i.e.

$$P_{\eta_t}(\cdot|n_1^{t-1},\ \theta) \gtrsim_{\ell rw} P_{\eta_t}(\cdot|n_2^{t-1},\ \theta)$$

whenever $n_1^{t-1} \geqslant n_2^{t-1}$.

(ii) The family of distributions

$$\mathbb{P} = (\underline{P}(\theta):\theta\epsilon\Theta),$$

have marginal distributions with support N (a compact subset of $\mathbb{R}_+$) and densities with respect to some dominating measure. Further, these distributions satisfy an ℓr-ordering in θ; i.e., provided $\theta > \theta'$ we have, for every t

$$P_{\eta t}(\cdot|\theta) \gtrsim_{\ell r} P_{\eta t}(\cdot|\theta').$$

(iii) $G_0(\cdot)$ is a prior probability distribution with density $g_0(\cdot)$ on Θ.

(iv) $\pi(\cdot)$ and $S(\cdot)$ are continuous increasing functions from N into $\mathbb{R}_+$. $\pi(\cdot)$ provides the payoff to, and $S(\cdot)$ the size of, the firm. [ ]

Our behavioral assumption is that management acts so as to maximize the expected discounted value of future net cash flow conditional on current information, where the conditional distribution of future net cash flows are formed, in a Bayesian fashion, from; the family of η processes ($\int P$), the prior for θ [$G_0(\cdot)$], and past realizations of η, say $n^t = (n_1,\ldots,n_t)$. The next assumption provides these conditional distributions.

4. Assumption [posterior distributions]

Let J_t contain all information available in period t. Then

$$\Pr(\theta \leq z | J_t) = p_{\eta t}(n^t | \theta \leq z) G_0(z) \Big/ \int p_{\eta t}(n^t | \zeta) G_0(d\zeta) \equiv P_\theta(z | n^t),$$

for $z \epsilon \theta$. Moreover $P_\theta(\cdot|n^t)$ has a density, $p_\theta(\cdot|n^t)$, with respect to the G_0 measure (for $n^t \epsilon N^t$, and all t). [ ]

Lemma 5 states that, under the ℓr-ordering assumptions, higher past η realizations lead to more favorable posteriors for θ. It follows directly from Bayes law and assumption (3.ii).[3]

5. Lemma (monotonicity of posteriors)

For any t, and n_1^t, $n_2^t \epsilon N^t$ with $n_1^t > n_2^t$,

$$P_\theta(\cdot|n_1^t) \underset{\sim}{\geq}_{\ell r} P_\theta(\cdot|n_2^t). \qquad [\ ]$$

Now consider the decision problem facing the owners of a firm which has been in existence t periods and has had η realizations of n^t. The owners must choose whether to continue in operation over the coming period, or close down and sell the firm at the value, Φ. If the owners decide to operate the firm they will obtain the profits over the coming period, plus the option of

keeping the firm in operation over subsequent periods should they desire to do so.[4]

Assume, temporarily, the existence of a bounded function, say $V_{t+1}(n^{t+1})$, from N^{t+1} into $\mathbb{R}$, which provides the value of continuing in operation from period t+1 given a realization of η^{t+1} equal to n^{t+1}. Then, letting $\beta\epsilon(0,1)$ be the discount factor, we have

$$(6) \qquad V_t(n^t) = E[\pi(\eta_{t+1})|n^t] + \beta E[\max\{\Phi, V_{t+1}(\eta^{t+1})\}|n^t],$$

where for any $h(\cdot)$, the expectation $E[h(\eta^{t+1})|n^t] \equiv \int h(\zeta,n^t)P_{\eta_t}(d\zeta|n^t)$.

Given (6) the optimal strategy of the owner is straightforward. Operate the firm if and only if $V_t(n^t) \geq \Phi$. Theorem 7 insures that the value function in (6) exists and then provides some of its properties.

7. Theorem (existence and montonicity of the value function)

At each t there exists a unique $V_t(\cdot):N^t \to \mathbb{R}_+$ which provides the value of continuing in operation assuming optimal behavior in each future period. $V_t(\cdot)$ is bounded, satisfies (6), and is nondecreasing in n^t; i.e., if $n_1^t \geq n_2^t$, then $V_t(n_1^t) \geq V_t(n_2^t)$ [for $n^t\epsilon N^t$, and all t]

Proof See Appendix I. [ ]

Note that Theorem 7 depends only on Assumption 3. It does not depend on: the precise functional form (or even the curvature) of the profit function (so the production function could display regions of increasing returns); on the form of $G_0(\cdot)$; or on the family $\mathbb{P}$ provided that it satisfy the monotone likelihood ratio properties in (3) (in particular the posteriors for θ need not possess simple sufficient statistics, nor need they be weakly continuous

in their arguments). We now move on to consider the empirical implications of the passive learning model and we shall focus on implications which require only the assumptions reviewed above.

2.2 Empirical Implications of Passive Learning.

Throughout we shall focus on the empirical implications of the passive learning model that are true at each age (that model also has limit properties as age grows large, but it is hard to use these as a basis for empirical analysis without further, a priori, information). We begin by deriving the implications of the passive learning model on the evolution of the size distribution of firms.

The theorem that underlies our results on the evolution of the size distribution is the economist's (far more palatable) version of the Darwinian dictum of "survival of the fittest." It states that as age passes the θ-distribution of the surviving firms improves (in the stochastic dominance sense). This is a result of self-selection. As time passes firms with lower θ's are more likely to draw lower η's and self-liquidate.

8. Theorem (the evolution of the θ-distribution)

Let $A^t = \{n^t = (n_1,\ldots,n_t): V_1(n_1) \geq \Phi,\ldots,V_t(n^t) \geq \Phi\}$, and

$$\chi_t(n^t) = \begin{bmatrix} 1 & \text{if } n^t \in A^t \\ 0 & \text{if } n^t \notin A^t \end{bmatrix} .$$

Then a firm is still operating in period t if and only if $\chi_t = 1$. Further, for every $z \in \Theta$ and all t let

$$P_\theta(z|t) \equiv \Pr\{\theta \leq z | \chi_t = 1\}.$$

Then

$$P_\theta(\cdot|t+1) \gtrsim_{sw} P_\theta(\cdot|t).$$

Proof Take an arbitrary (z,t). Then, by Bayes law,

$$P_\theta(z|t) = \Pr\{\chi_t=1|\theta\leqslant z\}G_0(z)/\Pr\{\chi_t=1\}$$

$$= [{}_{\theta\leqslant z}\textstyle\int \Pr\{\chi_t=1|\theta\}G_0(d\theta)]\Big/[{}_{\theta}\textstyle\int \Pr\{\chi_t=1|\theta\}G_0(d\theta)].$$

We must show that $P_\theta(z|t-1) \geqslant P_\theta(z|t)$. For this it suffices that

$$(8.1)\qquad \frac{{}_{\theta}\textstyle\int \Pr\{\chi_t=1|\theta\}G_0(d\theta)}{{}_{\theta}\textstyle\int \Pr\{\chi_{t-1}=1|\theta\}G_0(d\theta)} \geqslant \frac{{}_{\theta\leqslant z}\textstyle\int \Pr\{\chi_t=1|\theta\}G_0(d\theta)}{{}_{\theta\leqslant z}\textstyle\int \Pr\{\chi_{t-1}=1|\theta\}G_0(d\theta)}.$$

Using the fact that

$$\Pr\{\chi_t=1|\theta\} \equiv \Pr\{\chi_t=1|\chi_{t-1}=1,\theta\}\Pr\{\chi_{t-1}=1|\theta\},$$

and letting

$$Q_1(d\theta) = \Pr\{\chi_{t-1}=1 \mid \theta\}G_0(d\theta)\Big/{}_{\theta}\textstyle\int \Pr\{\chi_{t-1}=1 \mid \theta\}G_0(d\theta), \text{ and}$$

$$(8.2)\qquad Q_2(d\theta) = \begin{bmatrix} 0 & \text{for } \theta > z \\ \Pr\{\chi_{t-1}=1 \mid \theta\}G_0(d\theta)\Big/{}_{\theta\leqslant z}\textstyle\int \Pr\{\chi_{t-1}=1 \mid \theta\}G_0(d\theta), & \text{otherwise} \end{bmatrix}$$

(8.1) can be rewritten as

$$(8.3)\qquad {}_{\theta}\textstyle\int \Pr\{\chi_t=1 \mid \chi_{t-1}=1,\theta\}Q_1(d\theta) \geqslant {}_{\theta}\textstyle\int \Pr\{\chi_t=1 \mid \chi_{t-1}=1,\theta\}Q_2(d\theta).$$

Since (8.2) implies $Q_1(\cdot) \gtrsim_{sw} Q_2(\cdot)$, (8.3) will be true provided $\Pr\{\chi_t=1 \mid \chi_{t-1}=1,\theta\}$ is nondecreasing in θ. To see that this is indeed the case write

$$\Pr\{\chi_t=1 \mid \chi_{t-1}=1,\theta\} = \textstyle\int \Pr\{\chi_t=1 \mid n^{t-1},\theta\}P_{\eta^{t-1}}(dn^{t-1} \mid n^{t-1}\epsilon A^{t-1},\theta).$$

Then, taking $\theta \geqslant \theta'$

$$\int \Pr(\chi_t=1 \mid n^{t-1},\theta) P_{\eta t-1}(dn^{t-1} \mid n^{t-1} \epsilon A^{t-1},\theta) \geq$$
$$\int \Pr(\chi_t=1 \mid n^{t-1},\theta') P_{\eta t-1}(dn^{t-1} \mid n^{t-1} \epsilon A^{t-1},\theta) \geq$$
$$\int \Pr(\chi_t=1 \mid n^{t-1},\theta') P_{\eta t-1}(dn^{t-1} \mid n^{t-1} \epsilon A^{t-1},\theta'),$$

where the first inequality follows from the monotonicity of $V(\cdot)$ and the fact that $P_{\eta t}(\cdot \mid n^{t-1},\theta)$ is stochastically increasing in θ, and the second from (3.1) and the fact that if $P_{\eta t}(\cdot \mid \theta) \underset{\sim}{\geq}_{\ell r} P_{\eta t}(\cdot \mid \theta')$, then, for any $A \epsilon N^t$, $P_{\eta t}(\cdot \mid \eta^t \epsilon A,\theta) \underset{\sim}{\geq}_{\ell r} P_{\eta}(\cdot \mid \eta^t \epsilon A,\theta')$ (see Ross, 1982, appendix I). [ ]

Our first empirical implication of the passive learning model is a direct corollary of Theorem 8. Since size is an increasing function of η, and η is stochastically increasing in θ, the fact that the θ distribution of the surviving firms is stochastically increasing over time implies that the size distribution of surviving firms ought to be stochastically increasing in time.

9. Corollary (The evolution of the size distribution.)

Let χ_t be defined as in theorem 8, recall that $S_t = S(\eta_t)$, and for all z and t define

$$P_s(z \mid t) = \Pr(S_t \leq z \mid \chi_t=1).$$

Then, provided $t \geq t'$

$$P_s(\cdot \mid t) \underset{\sim}{\geq}_{sw} P_s(\cdot \mid t'). \qquad [ ]$$

There are many ways of employing Corollary 9 to identify industries that might abide by the passive learning model. The simplest is to plot the size distribution for different ages and compare them; the proportion of the sample greater than any given size should increase in age. More generally the corollary implies that if $h(\cdot)$ is any increasing function, then whenever $t \leq t'$

$$\bar{h}(t) = \int h(\zeta)P_s(d\zeta \mid t) \leq \int h(\zeta)P_s(d\zeta \mid t') = \bar{h}(t').$$

So we could take the sample analogue of $\bar{h}(t)$ [the sample mean of h(S)], and investigate whether it increases in age. We come back to these points below. Note also that Theorem (8) and Corollary (9) imply that each sequence of distribution functions, $\{P_\theta(\cdot|t)\}$, and $\{P_s(\cdot|t)\}$, converges (pointwise), to a well-defined limiting distribution, say $P_\theta(\cdot|\infty)$ and $P_s(\cdot|\infty)$.

Implications of the passive learning model that specify a monotonic relationship between two or more observables are particularly useful since they can be checked against data without imposing undue functional form restrictions. Though the literature on the passive learning model seems to have missed Corollary 9, it has associated at least three other monotonic relationships with passive learning. These are that:

i) the hazard rate is nonincreasing in current size; i.e., that $\Pr\{\chi_t=0 \mid \chi_{t-1}=1, S_{t-1}=s_{t-1}\}$ is nonincreasing in s_{t-1} for all t;

ii) the hazard rate is nondecreasing in age (usually, but not always, conditional on size);

iii) and that the variance in growth rates (again usually conditional on size) is nonincreasing in age

(these implications are discussed in Jovanovic, 1982; Evans, 1987a and 1987b; and Dunne, Roberts and Samuelson, 1987).

The next example shows that of these three only the first survives our search for nonparametric implications of the passive learning model (the example assumes, as did Jovanovic, 1982, that the distribution of $\{\eta_t\}$ conditional on θ is i.i.d.). It is true, however, that the first implication, that is that hazard rates are nonincreasing in size at a given age, is consistent

with the data from every empirical study we are aware of [Churchill, 1955; Wedervang, 1965; Evans 1987a and 1987b; Dunne, Roberts, and Samuelson, 1987]. However, most other models that allow for mortality, including the active learning model of Ericson and Pakes (1987), also imply mortality rates that decrease in size for a given age. Therefore, this property fails to distinguish among the alternative models, and we do not pay further attention to it in this paper.

As to the other implications, the fact that the passive learning model does not imply that either hazard rates, or the variance in growth rates, decline in age (at least not without further ad hoc assumptions) is somewhat disconcerting. Decreasing hazards and decreasing variances in growth rates have both been associated with the passive learning model in the past, and, in addition, have been shown to be fairly robust features of the data. On the other hand, the intuition underlying our counterexample is clear enough. For many functional forms it will take take time to accumulate the information necessary to ensure that exit is optimal, and this fact generates an initial increasing portion to the hazard function (actually the example generalizes this intuition and generates a hazard function which oscillates over age). As to differences in the variance in growth rates over age, these will depend upon, among other factors, the relative variances of η conditional on θ for different values of θ. If θ-values which are more likely to induce exit are associated with low variances, the observed variance in growth rates may well increase over age.

10. Example

Let $\pi_t = \pi \cdot \eta_t$, with $\{\eta_t\}$ i.i.d. conditional on θ,

$$\eta_t = \begin{cases} 1 & \text{with probability } \theta \\ 0 & \text{otherwise} \end{cases} \quad ; \text{ and } \quad \theta = \begin{cases} \delta & \text{with probability } \ell \\ 0 & \text{otherwise} \end{cases} .$$

The posterior for θ in this problem depends only on the couple (x_t,t), where $x_t = \max[n_1,\ldots,n_t]$. Consequently the value function in (6) has the simple form,

$$V_t(n^t) = V(x_t,t).$$

x_t is either 0 or 1. If $x_t=1$ management knows that $\theta=\delta$ and a direct calculation shows

$$V(1,t) = \pi\delta/(1-\beta)>\Phi,$$

where the inequality is by assumption. This inequality ensures that if $x_t=1$ management will never drop out. If $x_t=0$ the firm continues in operation if and only if $V(0,t)\geq\Phi$. It is easy to show that $\Pr(x_{t+1}=1 \mid x_t=0,t) = \Pr(\eta_{t+1}=1 \mid x_t=0,t)$ decreases in t, and converges to zero. This ensures that $V(0,t)$ decreases in t and converges to zero. Clearly then, there exists a unique t^* such that $V(0,t)\geq\Phi$ if and only if $t\leq t^*$. Let $S(\eta_t=1)=S$, $S(\eta_t=0)=0$, $H(t,S_t)$ be the hazard rate for firms of size S_t in period t, and $H(t)$ be the unconditional hazard. Straightforward calculations show that for

	$H(t,S_t=0)$	,	$H(t,S_t=S)$,		$H(t)$
$t<t^*$	0	,	0	,	0
$t=t^*$	$[(1-\delta)^{t^*}\ell+(1-\ell)]/[(1-\delta)\ell+(1-\ell)]$,		0	,	$(1-\delta)^{t^*}\ell+(1-\ell)$
$t>t^*$	0	,	0	,	0

So neither the conditional, nor the unconditional, hazard declines in age. This simply reflects the fact that for many possible assumptions on the relevant

functional forms it will take time to gather the information required to decide whether exit is optimal.

Next we consider the variance in growth rates. Provided $t > t^*$, any firm that is active has $\theta = \delta$, and $V(S_{t+1}-S_t \mid S_t)=V(S_{t+1} \mid \theta=\delta) = S^2\delta(1-\delta)$, regardless of S_t. If $t < t^*$ and $S_t = S$, then θ still is δ with probability one, and $V(S_{t+1}-S_t \mid S_t)$ is still given by the above formulae. So the variance in growth rates conditioned on $S_t = S$ is constant over age. However, if $t < t^*$, and $S_t = 0$, then θ can equal either δ or 0 with positive probability, and the variance in the growth rate is $[\delta S^2(1-\ell)(1-\delta)\ell]/[(1-\ell) + (1-\delta)\ell]^2$. Thus

$$V(S_{t+1}-S_t \mid S_t = 0, t>t^*)/V(S_{t+1}-S_t \mid S_t=0,\ t<t^*) = \frac{[(1-\ell)+(1-\delta)\ell]^2}{(1-\ell)\ell\cdot\delta},$$

which can be made as large as we like by choosing δ or ℓ small enough. The variance in growth rates need not decline in age. Whether or not they do will depend upon whether growth rates associated with high θ's are more variant than growth rates associated with low θ's, an issue which the basic passive learning model is silent on.

To see how this example generalizes, consider the case where θ has a beta prior distribution with parameters (r,s), i.e., $G_0(\cdot) = B(r,s)$, so that θ can take any value between zero and one. The posterior in this case is another beta with parameters $r + \sum^t \eta_i$ and $s + t - \sum^t \eta_i$, so that the sum, $x_t = \sum^t \eta_i$, and t, can be used as sufficient statistics. (Note that x_t is a nonnegative integer.) Using an argument analogous to that given above we find that for any fixed x, V(x,t) declines to zero with t. Thus for each x there exists a $t^*(x)$ such that $V(x,t) \gtrless \phi$ according as $t \lessgtr t^*(x)$ [see Figure 1]. Both the mortality, and the hazard rate will be zero for a value of t such that $t^*(x) < t < t^*(x+1)$ (for x = 1,2,...). Moreover it can be shown that $t^*(x+1)$ cannot equal $t^*(x)+1$ for consecutive values of x. That is, the hazard function will usually have a zero

between any two positive portions, making it oscillate over age. For $t = t^*(x)$ the hazard and mortality rates will be determined by the precise form of the prior. One such sequence of hazard rates is given in the bottom part of Figure 1. Similar pictures could be drawn for the variance in growth rates. [ ]

This example illustrates that if we are interested in other nonparametric implications of the passive learning model we should look beyond the implications of passive learning on the pattern of either the hazard or the variance in growth rates. It is, therefore, fortunate that the passive learning model has some very distinctive implications on the underlying structure of the conditional probabilities generating growth and mortality.

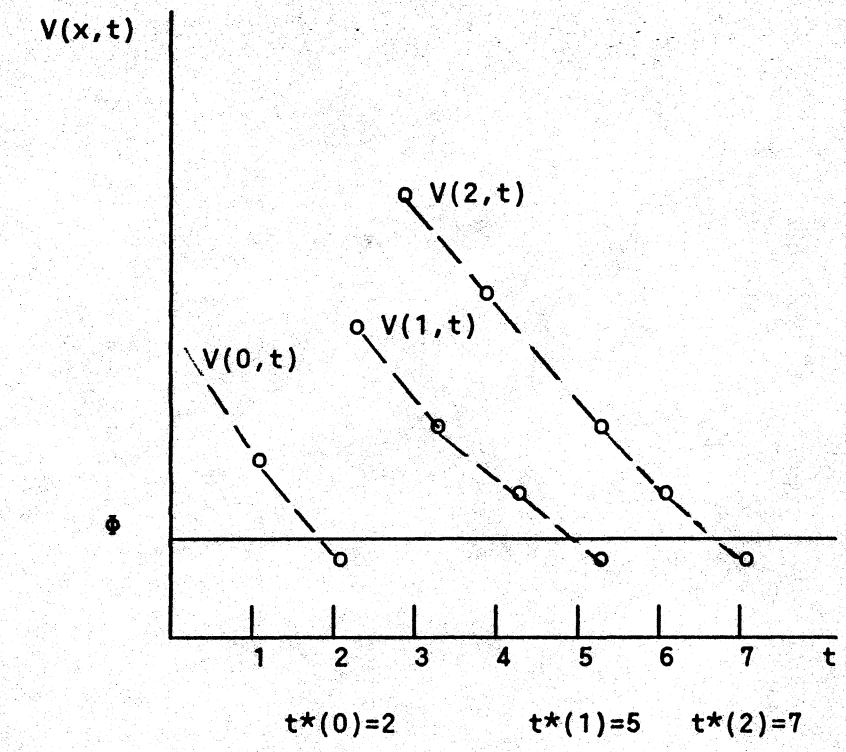
These implications stem primarily from the fact that θ is time-invariant. As a result, early realizations of η contain information about the parameter that determines the distribution of its future values; and this will be true no matter the time that elapses in the interim. Put differently, the dependence in the joint distribution of η_t and η_1 does not erode away as t grows large. This is seen most clearly in the special case where, conditional on θ, the $\{\eta_t\}$ are an i.i.d. process. In this case, as can easily be verified, for any n'

$$P_{\eta_t}(\cdot \mid \eta_k = n') = P_{\eta}(\cdot \mid n'),$$

which is independent of t and k. This strong invariance property is destroyed when we allow θ to index the more general family of stochastic processes permitted in (3). In the general case we have, for any $z \epsilon N$,

$$P_{\eta_t}(z \mid \eta_k = n') = \int P_{\eta_t}(z \mid \eta_k = n', \theta) P_{\theta}(d\theta \mid \eta_k = n'),$$

and since $P_{\eta_t}(z \mid \eta_k = n', \theta)$ can depend upon t and k, so can $P_{\eta_t}(z \mid \eta_k = n')$. However, the passive learning model does imply that the dependence in this



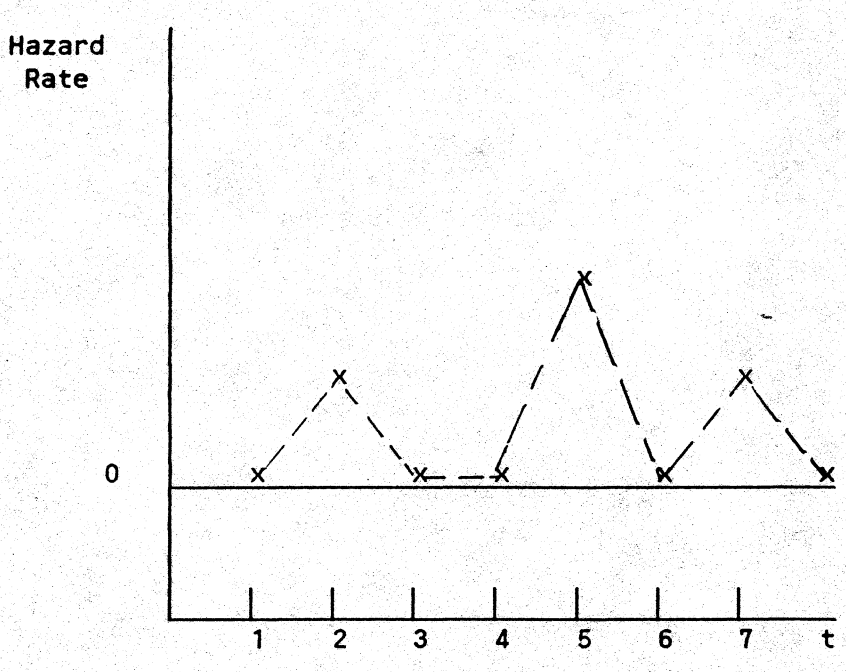


Figure 1: A Beta/Binomial Example

latter distribution has two sources, one of which will not erode away as t grows large. Though the dependence in the process generating η_t conditional on θ (in the integrand) may erode away with t (it will if the process generating η_t is ergodic), the dependence that results from the effect of the realization of η_k on the posterior for θ will not.

This argument can be formalized and then used to produce a test for the passive learning model based on differences between the marginal distribution of $S_t = S(\eta_t)$, and the distribution of S_t conditional on S_1. Actually we can do better than this and produce tests based on a comparison of the distribution of S_t conditional on $S_{t-1},\ldots,S_{t-k}$ to the distribution of S_t conditional on $S_{t-1},\ldots,S_{t-k}$, and S_1, for any $k \geq 0$. With a positive k this test is likely to be more powerful against alternatives in which the value of the parameter determining the firm's distribution of profits evolves in a Markovian fashion over time (and one such alternative is the active learning model considered in the next section).

Our test is a direct implication of the following theorem. The theorem states that if we choose any group of years for which there is information on past realizations of η, and derive the family of posterior distributions for θ conditional on possible η-realizations in those years, then members of the family with higher past η-realizations will stochastically dominate those with lower η-realizations.

11. Theorem (conditional distributions for η_t)

Let t and k be positive integers with $t \geq k$, and $(i_1,\ldots,i_k)$ be any selection of k distinct elements from $\{1,\ldots,t-1\}$. Then if $\underset{\sim}{n}_1' = (n^1_{i_1},\ldots,n^1_{i_k})$ and $\underset{\sim}{n}_2' = (n^2_{i_1},\ldots,n^2_{i_k})$ are arbitrary $(i_1,\ldots i_k)$ histories of η satisfying $\underset{\sim}{n}_1 > \underset{\sim}{n}_2$, and $\bar{x}_t$ is defined as in (8),

$$P_{\eta_t}(\cdot \mid \underset{\sim}{n}_1, \chi_t=1) \geq_s P_{\eta_t}(\cdot \mid \underset{\sim}{n}_2, \chi_t=1).$$

Proof. For any $z \in N$ and $\underset{\sim}{n} = \underset{\sim}{n}_1$ or $\underset{\sim}{n}_2$,

$$(11.1) \quad P_{\eta_t}(z \mid \underset{\sim}{n}, \chi_t=1) = \int P_{\eta_t}(z \mid \underset{\sim}{n}, \chi_t=1, \theta)\, P_\theta(d\theta \mid \underset{\sim}{n}, \chi_t=1),$$

where

$$P_{\eta_t}(z \mid \underset{\sim}{n}, \chi_t=1, \theta) = \int P_{\eta_t}(z \mid \eta^{t-1}, \theta) P_{\eta_{t-1}}(d\eta^{t-1} \mid \underset{\sim}{n}, \chi_t=1, \theta)$$

Now use Bayes law to show that for $\eta^{t-1} \geq \eta_*^{t-1}$,

$$p(\eta^{t-1} \mid \underset{\sim}{n}_1, \theta) p(\eta_*^{t-1} \mid \underset{\sim}{n}_2, \theta) - p(\eta^{t-1} \mid \underset{\sim}{n}_2, \theta) p(\eta_*^{t-1} \mid \underset{\sim}{n}_1, \theta)$$

$$= \kappa \left[p(\underset{\sim}{n}_1 \mid \eta^{t-1}, \theta)\, p(\underset{\sim}{n}_2 \mid \eta_*^{t-1}, \theta) - p(\underset{\sim}{n}_1 \mid \eta_*^{t-1}, \theta) p(\underset{\sim}{n}_2 \mid \eta^{t-1}, \theta) \right] \geq 0,$$

where the inequality is a trivial consequence of $\underset{\sim}{n}$ being determined by η^{t-1}. Since conditioning on $\eta^t \in A_t = \{\eta^t: \chi_t(\eta^t) = 1\}$ does not affect the ℓr-ordering, we have

$$(11.2) \quad P_{\eta_t}(\cdot \mid \underset{\sim}{n}_1, \chi_t=1, \theta) \geq_{\ell rw} P_{\eta_t}(\cdot \mid \underset{\sim}{n}_2, \chi_t=1, \theta).$$

Given (11.1), (11.2) and lemma 2, the theorem requires only that $P_\theta(\cdot \mid \underset{\sim}{n}_1) \geq \ell r\, P_\theta(\cdot \mid \underset{\sim}{n}_2)$. But by lemma 4, this condition is satisfied provided

$$P_{\underset{\sim}{n}}(\cdot \mid \theta_1) \geq_{\ell r} P_{\underset{\sim}{n}}(\cdot \mid \theta_2),$$

whenever $\theta_1 \geq \theta_2$. Take any $\underset{\sim}{n}_1 > \underset{\sim}{n}_2$, then

$$p(\underset{\sim}{n}_1 \mid \theta_1)\, p(\underset{\sim}{n}_2 \mid \theta_2) - p(\underset{\sim}{n}_1 \mid \theta_2)\, p(\underset{\sim}{n}_2 \mid \theta_1)$$

$$= \int \left[p_{\eta t}(d\eta, \underset{\sim}{n}_1 \mid \theta_1)\, p_{\eta t}(d\eta, \underset{\sim}{n}_2 \mid \theta_2) - p_{\eta t}(d\eta, \underset{\sim}{n}_2 \mid \theta_1) p_{\eta t}(d\eta, \underset{\sim}{n}_1 \mid \theta_2) \right] > 0$$

where the integral runs over those η_t whose indices are in $\{1,\ldots,t-1\}$ but are not in $\{i_1,\ldots,i_\kappa\}$, and the inequality results from (3.ii). [ ]

The empirical implication of theorem (11) that we will be using is that it implies that for any $k \geq 0$, and any $(n_{t-1},\ldots,n_{t-k}) \in N^{k-1}$,

$$(12) \qquad P_{\eta_t}(\cdot \mid n_{t-1},\ldots n_{t-k},n_1,\chi_t=1) \gtrsim_s P_{\eta_t}(\cdot \mid \eta_{t-1},\ldots,n_{t-k},n_1',\chi_t=1),$$

whenever $n_1 > n_1'$. Corollary (13) is an immediate implication of (12).

13. Corollary

Let t and k be nonnegative integers with t>k, and let χ_t be defined as in Theorem 8. Then

$$E[S_t \mid S_{t-1}=s_{t-1},\ldots,S_{t-k}=s_{t-k},\ S_1=s_1,\ \chi_t=1]$$

is strictly increasing in s_1 for almost every $(s_{t-1},\ldots,s_{t-k})$. [ ]

That is, expected future size conditional on k past sizes and survival will be strictly increasing in the initial size. This is because the parameter which determines the conditional distribution of the payoff relevant η is time-invariant. In models in which these conditional distributions depend on a parameter which evolves over time in response to, say, the outcomes of a firm's exploratory investment, corollary (13) will not necessarily be true. We turn to these types of models now.

Section 3. Active Learning

This section considers the empirical implications of a model (originally developed by Ericson and Pakes, 1987), in which firms can invest to improve the value of a parameter, say ω, which determines the distribution of its profits. In the active (in contrast to the passive) learning model, management is assumed to know its current value of ω (and hence the actual profit distribution it faces), and makes current production decision based on it. On the other hand ω itself evolves over time in response to the outcomes of the firm's own investment process, and the investments of other firms operating in related markets. These outcomes are stochastic; in the active learning model the firm is investing to explore and develop alternative market niches which may, or may not, prove profitable.

In this model the distribution of futures states is determined entirely by the current state and the optimal investment policy. It is, therefore, independent of the age of the firm per se. Startup is treated as the appearance of an idea which, given current market conditions, appears worth exploring. Formally it is an initial location on the ω-axis. If the idea requires substantial successful development before it can generate noticeable profits, the initial ω is associated with a distribution of profits which is degenerate (or nearly so) at zero. Successful investment will enable the idea to be embodied in a more profitable marketable good or service. Unsuccessful exploration may well convince the entrepreneur that the whole idea is not worth pursuing and lead to liquidation.

Ericson and Pakes begin with a three-dimensional state vector and then show how, under certain conditions, the three dimensions can be collapsed into two; one providing the outcomes of the firm's own investments relative to those of

its competitors, and one providing the strength of the market per se (a factor which can be affected by exogenous shifts in demand and supply conditions). The latter has a role similar to that of output price in the passive learning model (its value is the same for all firms at a given point in time), and we shall, for expositional simplicity, ignore it here also. We provide a brief description of the active learning model focussing on those results needed to compare its empirical implications to those from the passive learning model. Again we consider only those empirical implications that are nonparametric in the sense that they require only mild regularity conditions on the relevant functional forms.[5]

The Active Learning Model

We will assume that the state space is countable and index it by the integers so that $\omega \epsilon Z$. Each firm operating in period t is endowed with an ω_t. Higher values of ω are better in the sense that the distribution of the payoff relevant η is stochastically increasing in ω. Management has three choices to make in each period, and they are made to maximize the expected discounted value of future net cash flows. First the firm must decide whether to operate at all. If it decides against it receives a liquidation value of Φ and exits never to reappear again. If the firm does operate management must decide on both a level of current input demand, and an amount of exploratory investment, say x_t. Given a realization of η, current input choices will determine current operating profits, say $\pi(\eta_t)$. Current cash flows are

$$R(\eta_t,\omega_t,x_t) = \pi(\eta_t)-c(\omega_t)x_t$$

where $c(\cdot)>0$, and can be decreasing in ω to reflect the possibility that more profitable firms may find it easier to raise finance capital. Increases in

current investment decrease current cash flow but make higher values of ω_{t+1}, and hence higher future profits, more likely. In particular, let $\tau_{t+1}=\omega_{t+1}-\omega_t$, and J_t be the information available to management at t. Then we assume that for $z\epsilon Z$,

$$P_\tau(\tau_{t+1}\leq z|J_t) = P_\tau(z_t|x_t),$$

where $P_\tau(\cdot\,|\,x_t)$ is stochastically increasing in x. Hence, to formalize the firms decision problem we will require the following primitives.[6]

1. Assumption (primitives of the active learning model)

i) $\mathcal{P}_\eta = (P_\eta(\cdot\,|\,\omega):\omega\epsilon Z)$, is a family of distribution functions indexed by ω. The family has support, N, a compact subset of Z containing zero, and exhibits a weak first order stochastic dominance ordering in ω, i.e.

$$P_\eta(\cdot\,|\,\omega) \gtrsim_{SW} P_\eta(\cdot\,|\,\omega')$$

whenever $\omega>\omega'$. It is assumed that $\lim_{\omega\to-\infty} P_\eta(0\,|\,\omega)=1$. (This, together with the assumption that $\pi(0)=0$, insures that for small enough ω payoffs are zero with probability one.)

ii) $\mathcal{P}_\tau = (P_\tau(\cdot\,|\,x):x\epsilon\mathcal{R}_+)$ is a family of distributions with support T, a compact subset of Z, exhibiting a weak first order stochastic dominance ordering in x, i.e.

$$P_\tau(\cdot\,|\,x) \gtrsim_{SW} P_\tau(\cdot\,|\,x')$$

whenever $x>x'$, and satisfying the condition that

$$P_\tau(0|0) = 1,$$

so that the firm's product cannot be improved without some investment. The

family of densities $(p_\tau(\cdot \mid x): x \epsilon]R_+)$, is (pointwise) differentiable in x with derivatives which are decreasing in x for $\tau>0$, and increasing in x for $\tau<0$ (this insures that the investment problem is concave and therefore has a unique solution), and both $p_\tau(0 \mid x)$ and $p_\tau(-1 \mid x)$ are strictly positive for all x less than any finite upper bound (these are technical conditions whose roles are explained in more detail below).

iii) $\pi(\cdot)$ and $S(\cdot)$ are increasing functions of η, and $c(\cdot)$ is a non-increasing function of ω, into $]R_+$. $\pi(\cdot)$ provides the profits, and $S(\cdot)$ provides the size, of the firm; while $c(\cdot)$ provides the cost of a unit of x. $\pi(0)=0$, and $c(\cdot)$ is bounded away from zero. [ ]

We now consider management's choice of policies. Letting ω_0 be the initial state and χ_τ be the indicator function which takes the value one if the firm is active in period τ and zero elsewhere, a policy, say d, is a sequence of functions mapping available information into operating and investment decisions, that is

$$d = \{\chi_0(J_0),\ x_0(J_0),\ \chi_1(J_1),\ x_1(J_1),\ldots\},$$

with $\chi_\tau=\chi_t(J_\tau)$, $\chi_\tau=0$ implying $\chi_{t+\tau}=0$ for $t \epsilon Z_+$, $x_\tau = x_\tau(J_\tau)$, and $J_\tau = \{\omega_\tau, x_{\tau-1}, x_{\tau-1}, \omega_{\tau-1}, \ldots, \omega_0\}$. Recall that $R(\eta_\tau, \omega_\tau, x_\tau, \chi_\tau) = \pi(\eta_\tau) - c(\omega_\tau)x_\tau$ if $\chi_\tau=1$ and zero otherwise, so the expected discounted value of net cash flows given the policy d is

$$V_d(\omega_0) = E_d \{\sum \beta^\tau R(\eta_\tau, \omega_\tau, x_\tau, \chi_\tau) + \Phi(\chi_{\tau-1} - \chi_\tau)] \mid \omega_0\}$$

where $\beta \epsilon (0,1)$ is a discount factor, and the expectation is taken assuming that the d-policy is followed. Note that (1) implies that $R(\cdot)$ is bounded, and let

$$V(\omega) = \sup_{d} V_d(\omega)$$

for each ω. A policy d^* will be optimal if $V_{d^*}(\omega) = V(\omega)$ for all ω. If an optimal policy exists management chooses it, in which case the expected discounted value of future net cash flow is $V(\omega)$. Management will operate the firm if and only if $V(\omega)>\Phi$, the liquidation value. The following theorem combines the results from Ericson and Pakes (1987) that are used in our derivation of the empirical implications of their model. The theorem is followed by diagrammatic and verbal expositions of its contents.

2. Theorem (properties of the active learning model).

A unique optimal policy and associated value function exist and they have the following characteristics:

i) $V(\omega)$ is bounded and nondecreasing in ω.

ii) The optimal policy, $x^*_\tau(J_\tau)$ is bounded, depends only on current ω, and is stationary, i.e. for all τ

$$x^*_\tau(J_\tau) = x^*_\tau(\omega_\tau) = x^*(\omega_\tau) \leq \bar{x} < \infty.$$

iii) There exists a couple, $(\underline{\underline{\omega}},\bar{\bar{\omega}})$ with, $-\infty < \underline{\underline{\omega}} \leq \bar{\bar{\omega}} < \infty$, such that

$$x^*(\omega)=0 \quad \text{if} \quad \omega \notin \{\omega' : \underline{\underline{\omega}} \leq \omega' \leq \bar{\bar{\omega}}\}.$$

iv) There exists a second couple $(\underline{\omega},\bar{\omega})$, with $-\infty < \underline{\omega} \leq \underline{\underline{\omega}} \leq \bar{\bar{\omega}} \leq \bar{\omega} < \infty$, such that

$$V(\omega)>\Phi \quad \text{if and only if} \quad \omega>\underline{\omega},$$

and

$$\inf_t \left[\inf_{\omega_0 \leq \bar{\omega}} \Pr(\omega_t \leq \bar{\omega} \mid \omega_0) \right] = 1. \qquad [\ ]$$

Parts (i) and (ii) of this theorem ensure that both the value function and investment policy are stationary functions of ω, the value function being increasing in ω. Figure 2 illustrates this with one special case developed in Ericson and Pakes. In the figure $A(\omega) = \int \pi(\eta) P_{\eta}(d\eta \mid \omega)$, provides expected profits conditional on ω. The value of ω below which a firm exits, i.e. the $\underline{\omega}$ in (2.iv), is determined by the point at which $V(\omega)$ equals Φ. In this example $\underline{\omega} = \underline{\underline{\omega}}$, that value of ω below which a firm stops investment. So positive investment occurs at $\underline{\omega}+1$, even though profits at that point are zero with probability one. The incentive for the investment is that it makes higher values of ω_{t+1}, and hence higher future profits, more likely. The monetary value of an increase in ω is $V(\omega_{t+1}) - V(\omega_t)$. Since $V(\omega)$ is bounded, after some point increases in ω cannot bring with it much of a change in $V(\cdot)$. It follows that, after some ω, it will not be in the firm's interest to invest at all. The ω at which this occurs is the $\bar{\bar{\omega}}$ of (2.iii). If $\omega > \bar{\bar{\omega}}$, no investment takes place and this insures (see 1.ii) that the firm's ω does not increase (in fact it will stochastically deteriorate as other firms gradually develop goods and services that obsolete the product of this firm). Let τ^* be the largest value of τ that has positive probability when $x=\bar{x}$ (recall that $\bar{x}=\max x_*(\omega)$, and that τ^* is finite by virtue of 1.ii). Then firms with $\omega_t < \bar{\bar{\omega}}$ have $\omega_{t+1} \leq \bar{\bar{\omega}}+\tau^* = \bar{\omega}$, and since firms with $\bar{\bar{\omega}} < \omega_t \leq \bar{\omega}$ have $\omega_{t+1} \leq \omega_t$, if $\omega_t \leq \bar{\omega}$ so must be ω_{t+1}. This explains the second statement in 2.iv; that is, if $\omega_0 \leq \bar{\omega}$, then, with probability one, so will be the entire sequence $\{\omega_t\}_{t=0}^{\infty}$.

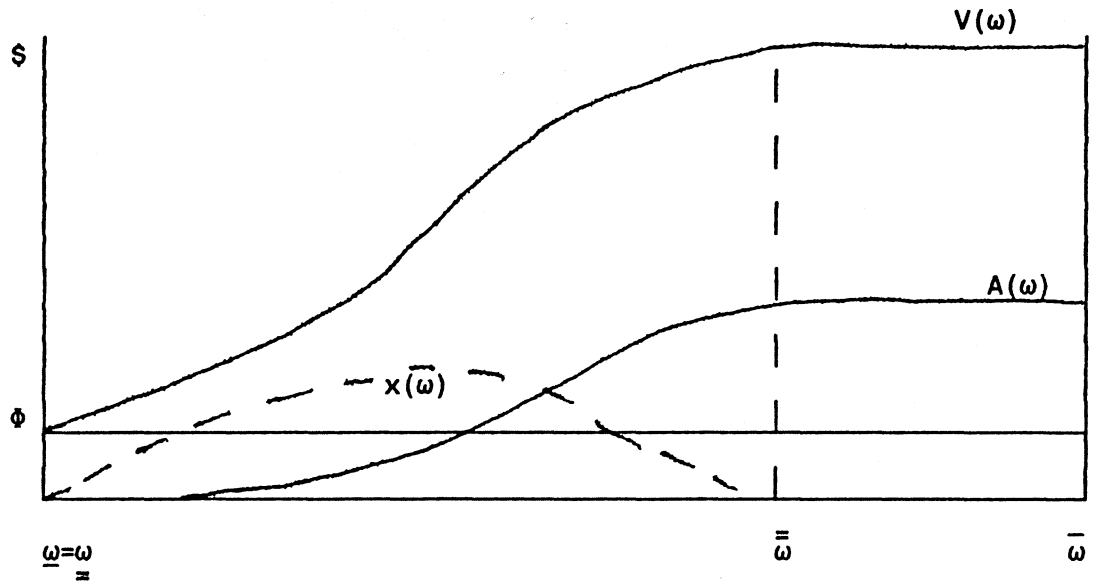


Figure 2: Policies in the Active Learning Model

Since all values of $\omega \leq \bar{\omega}$ induce permanent exit, there is no need to distinguish among them. It is, therefore, convenient to transform the state space by the map $f(\cdot)$, where

$$f(\omega) = \begin{cases} 0 & \text{for } \omega \leq \underline{\omega} \\ \omega - \underline{\omega} & \text{elsewhere.} \end{cases}$$

Let $K = \bar{\omega} - \underline{\omega}$, so that if $f(\omega_t) \leq K$, so is $f(\omega_{t+1})$. We shall work only with values of $f(\omega)$ in what follows. At the risk of some notational confusion, then, we also label its values by ω.

With this understanding, theorem 2.2, implies that the sequence $\{\omega_t\}$ together with any $\omega_0 \leq K$ is a finite state Markov chain on $\Omega = \{0,1,\ldots,K\}$. Its 'zero' or 'death' state is absorbing, so the transition matrix for the chain is given by $\underline{P}$, where

$$\underline{P} = [p_{i,j}] = \begin{bmatrix} 1,0\ldots,0 \\ p_{i,j} \end{bmatrix}$$

and for $0 < i \leq K$ (3)

$$p_{i,j} = \begin{cases} p_\tau(\tau=j-i \mid x^*(i)), & \text{for } K \geq j > 0 \\ \sum_{\tau \leq -i} p_\tau(\tau=j-i \mid x^*(i)), & \text{for } j=0. \end{cases}$$

Two remarks are in order here. First, recall that realizations of ω are not observable. Realizations of $\{S_t\}$ are, but $S(\eta_t)=\bar{S}(\omega_t)+U(\eta_t)$, where $\bar{S}(\omega_t) = \int S(\eta_t)P(d\eta \mid \omega_t)$, and $U(\eta_t) = S(\eta_t)-\bar{S}(\omega_t)$. Since the distribution of $U(\eta_t)$ is also determined by ω_t, and $\{\omega_t\}$ is a Markov process, S_t is a sum of two Markov processes. But a process which is a sum of Markov processes is not, in general, Markov. So the observable $\{S_t\}$ process is not Markov.

The second point to note concerns the mortality of firms. Assumption (1.iii) insures that exists a finite n*, such that for n>n*

$$\min_{i \in \Omega} \{p^n_{i,0} : i \in \Omega\} \geq \epsilon > 0,$$

where $p^n_{i,j} = \Pr(\omega_{t+n} = j \mid \omega_t = i)$. Since $p_{0,0} = 1$, this implies that all states but 0 are 'transient'. That is, no matter its initial ω, a firm will, with probability one, reach zero in finite time and stay there. Firms, like people, eventually die.

Since the passive learning model implies that firms can survive forever there is a sense in which this latter result differentiates the active from the passive learning model. However, in order to make empirical use of this distinction we would require a very long time series of data. On the other hand the passive learning model did have the additional implication that the size distribution of surviving firms ought to be stochastically increasing in any finite range of ages (corollary 2.9). To investigate the properties of the survivor distribution in the active learning model we require some additional notation.

Let

$$Q^L = \{q \epsilon R_+^L : \sum q_i = 1\}$$

be an L-dimensional simplex, so that any $q \epsilon Q^{K+1}$ can be regarded as a density on Ω. Note that a potential entrant with an $\omega=0$ would not enter, so that the initial distribution of the ω in a cohort is a $p^0 \epsilon Q^K$. Similarly an ω distribution for the survivors in a cohort is a $q \epsilon Q^K$. To obtain the properties of this distribution we require the operator Γ: $Q^K \to Q^K$, which produces the density of survivors at t+1 from any $q \epsilon Q^k$ at t, i.e.

$$(\Gamma q)_j = \sum_{i=1}^{k} q_i p_{ij} \; / \; [1 - \sum_{i=1}^{k} q_i p_{i,0}]$$

or, in matrix notation,

$$\Gamma(q) = q' \underline{P} (q' \underline{P} e)^{-1}$$

where e is a column vector of ones. Then $\Gamma^t(p^0)$ provides the ω distribution of survivors at age t from a cohort with initial distribution p^0. Theorem 4, and the explanation which follows it, are a direct consequence of the results in Ericson and Pakes (1987).

4. Theorem (the distribution of survivors)

i) For any initial ω-distribution (any $p^0 \epsilon Q^k$), $\lim_{t \to \infty} \Gamma^t(p^0) = p^*$, where p^* is the unique solution to $\Gamma(p^*)=p^*$.

ii) If $P_s(\cdot \mid t, p_0)$ provides the size distribution of firms surviving until period t from a cohort with an initial ω-distribution of p^0, then $P_s(\cdot \mid t, p^0)$ converges (pointwise) to $P_s^*(\cdot)$, where, for all z,

$$P_s^*(z) = \sum_j P_\eta(\eta \leq S^{-1}(z) \mid \omega = j) p_j^*. \qquad [\;]$$

(4.i) states that the ω-distribution of the surviving firms converges to an invariant distribution (invariant to both the initial distribution, and to the passage of time) whose density is given by p^*. (4.ii) provides the analogous limit property for the size distribution of the surviving firms. The Ericson Pakes paper actually goes one step further than this and shows that, given some additional regularity conditions on the location of p_0 and on the transition probabilities, there will be a finite t^*, such that for any p^0

$$(5) \qquad P_s(\cdot \mid t+1, p^0) \underset{\sim SW}{\geq} P_s(\cdot \mid t, p^0)$$

provided $t>t^*$. That is, not only does the size-distribution of surviving firms converge to an invariant distribution, but after some t^* the convergence will be 'monotone' and the size distribution of surviving firms will stochastically increase from period to period (just as in the passive learning model).

Still, however, the empirical implications of the active learning model on the evolution of the size-distributions of surviving firms are weaker than those of the passive learning model. In particular the active learning model does not predict that the size distribution will be stochastically increasing at each age. On the other hand, the active learning model does not bar this event from occurring, and it can predict that the size distribution will be stochastically increasing at later ages.

There is, however, at least one set of observable implications which differentiate between the two models more sharply. Recall that in the passive learning model the parameter that determines the distribution of profits is time invariant. This induces a dependence between the initial size of a firm and the size at any future date. Indeed as equation (2.12) shows, the passive learning model implies the stronger result that the conditional distribution of size at t, conditional on the immediate past sizes and the initial size,

will always be strictly increasing in the initial size. In the active learning model the parameter determining the firm's profitability distribution, i.e. ω, evolves over time. Later year size realizations are governed by a different value of ω than those from earlier years and, as time passes, the dependence between the later and earlier values of ω, and therefore of size, dies out. This is also true for the conditional distribution of S_t; i.e. the distribution of S_t conditional on immediate past values of S should gradually become independent of initial year sizes. Moreover, since the dependence of ω_t on its history is only through the value of ω_{t-1}, we might expect that if we condition on immediate past sizes the dependence on initial size will die out relatively quickly. Indeed, in the extreme case where $S_t = \bar{S}(\omega_t)$, so that sales is a deterministic function of ω_t, the conditional distribution of S_t depends only on S_{t-1}. In this case a three year panel is enough to differentiate the active from the passive learning model.

When there is noise in the relationship between ω_t and size, we must base our distinction between the active and the passive learning model on a more formal property of the stochastic process generating size conditional on survival (ϕ-mixing). Let $\{S_t^a\}_{t=1}^{\infty}$ be that process (it is described formally in Appendix 2). Then, the active learning model implies that as τ grows large the distribution of $(S^a_{X+\tau}, S^a_{X+\tau+1},\ldots)$ becomes, roughly speaking, independent of realizations of $(S^a_1,\ldots,S^a_X)$. More precisely, we have lemma 6 (which is proved in Appendix 2 and its implications (explained immediately after presentation of the lemma).

6. Lemma (ϕ-mixing of the $\{S_t^a\}$ process).

Let $\{S_t^a\}_{t=1}^{\infty}$ be the stochastic process formed from the distribution of sales

conditional on survival and any initial $\omega_0 \epsilon \{1,2,\ldots,K\}$, and M_x^y be the σ-algebra generated by possible realizations of S_x^a, $S_{x+1}^a,\ldots,S_y^a$. Then $\{S_t^a\}$ ϕ-mixes at a geometric rate, i.e.

$$\sup\{\, |P(E_2|E_1)-P(E_2)|\,,\ E_1 \text{ with } P(E_1) > 0 \text{ and } E_1 \epsilon M_1^x, E_2 \epsilon M_{x+\tau}^{\infty}\} \leqslant A\phi^{\tau}$$

with $\phi < 1$. [ ]

Lemma 6 states that any dependence between size realizations that occur after $x+\tau$, and size realizations that occur before x, goes down geometrically in τ. It implies that for $k \geqslant 0$

$$(7) \qquad \sum_z |p_s(z \mid s_{t-1},\ldots,s_{t-k},s_1,\chi_t=1) - p_s(z \mid s_{t-1},\ldots,s_{t-k},\chi_t=1)| \leqslant A_k\phi^t$$

for some $\phi<1$, on a set of $(s_{t-1},\ldots s_{t-k})$ with probability one. That is by choosing k sufficiently large we can make the conditional distribution of S_t, conditional on $s_{t-1},\ldots,s_{t-k}$, as close as we like to being independent of s_1. Note that equation (2.13) insures that this is <u>not</u> the case in the passive learning model. The next corollary is an immediate implication of (6) and (7).

<u>8. Corollary</u>

For any $k \geqslant 0$

$$E_{(s_1)} \Big| E[S_t \mid S_{t-1}=s_{t-1},\ldots,S_{t-k}=s_{t-k},S_1=s_1,\chi_t=1] - E[S_t \mid S_{t-1}=s_{t-1},\ldots,S_{t-k}=s_{t-k},\chi_t=1) \Big| \leqslant A_k\phi^t$$

on a set of $(s_{t-1},\ldots s_{t-k})$ with probability one. [ ]

Recall that corollary (2.14) insures that in the passive learning model the

conditional expectation of S_t, conditional on any realization, (s_{t-1}, s_{t-2},...s_{t-k},s_1) and survival until t, is strictly increasing in s_1. Hence corollary (8) differentiates the active from the passive learning model. The distinction between the two models is particularly striking in the special case where $S_t = \bar{S}(\omega_t)$, in which case A_k=0 for k>1. We now consider the econometric techniques needed to bring this distinction to data.

Section 3: Estimation and Testing

There are two nonparametric implications of the models we are considering that will be investigated empirically. The first is whether the size distribution of surviving firms is stochastically increasing in age; or whether, for all t

$$P_s(\cdot|t) \underset{\sim}{\geq}_{SW} P_s(\cdot|t-1). \qquad (1)$$

The passive learning model implies it must, while the active learning model implies it may, but need not - at least in the early ages. The second question posed of the data is whether, for different values of k,

$$E[S_t|S_{t-1} = s_{t-1}, \ldots, S_{t-k} = s_{t-k}, S_1 = s_1, \chi_t = 1] \qquad (2)$$

is strictly increasing in s_1. Again the passive learning model says it must be. But here there is a sharper contrast with the implications of the active learning model. The active learning model implies that, for t large enough, the regression function in (2) cannot depend on s_1. To check whether (1) seems consistent with the data, we will simply plot and compare the size distribution at different ages. It is more difficult to present a pictoral representation of the regression function in (2). Our analysis of its properties must, therefore, be somewhat more formal.

This section develops an intuitive nonparametric estimator for (2), and then considers tests of whether or not it is increasing in s_1. Indeed, since both models imply that the regression function is nondecreasing in s_1, we employ a two-part testing sequence. We first test whether (2) is weakly increasing in s_1. If this were not the case we would doubt whether either of our models provided an adequate approximation to the process generating the data being analyzed. If, on the other hand, the hypothesis of weak monotonicity

is acceptable, we move on to test the null of whether the regression function does not depend on s_1 against the alternative of it being strictly increasing in that variable. Acceptance of both null hypotheses is interpreted as support for the active learning model, while acceptance of only the first is interpreted as support for passive learning.

To obtain our estimator of the regression function we define J positive numbers, say $\{\bar{\sigma}_j\}_{j=1}^{J}$, and use them to break $\mathbb{R}_+$ into cells, as in figure 3. We then define the function $\sigma(\cdot)$: $\mathbb{R}_+ \to [1,\ldots,J]$ which assigns to each S_t the number of the cell it falls into, i.e. for j=1,...,J,

$$\sigma_t = \sigma(S_t)=j, \quad \text{if and only if,} \quad \bar{\sigma}_{j-1} < S_t \leq \bar{\sigma}_j , \tag{3a}$$

where it is understood that $\bar{\sigma}_0 = 0$, and $\bar{\sigma}_J = \infty$.

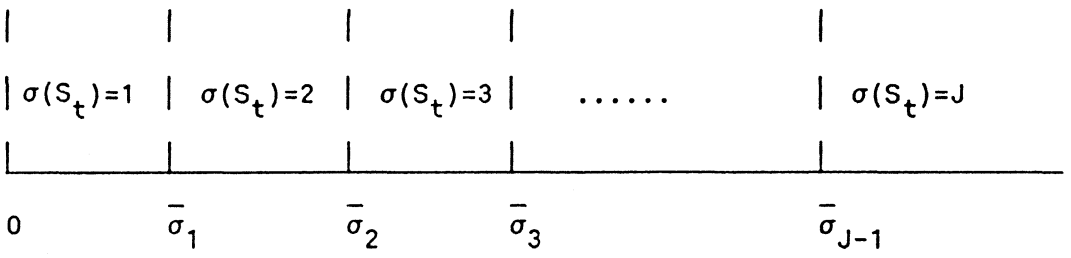


Figure 3: The Function, $\sigma(S_t)$.

Similarly for k < t define the function $\sigma^k(\cdot)$: $\mathbb{R}_+^{k+1} \to [1,\ldots,J]^{k+1}$, by

$$\sigma^k(S^{t-1}) = \{\sigma(S_{t-1}), \sigma(S_{t-2}), \ldots, \sigma(S_{t-k}), \sigma(S_1)\}, \tag{3b}.$$

In the empirical analysis we treat all values of S that fall into the same cell as equivalent (for the theoretical properties of the test statistics we require that the cell or 'band' width go to zero at an appropriate rate). For our purposes, then, a $\{S_{t-1}, S_{t-2},\ldots,S_{t-k},S_1\}$ history of a firm which

survives until period t is one of the J^{k+1} possible values of $\sigma^k(S^{t-1})$. Each of these values is a k+1 dimensional cell, and we denote the set of such cells by $\{\sigma_p^k; p=1,\ldots,J^{k+1}\}$. Our testing procedure is based on estimating the mean and the variance of the regression function in (2) in the intervals defined by these cells.

More precisely let μ^k and V^k denote the vectors

$$\mu^k = [\ \mu_p^k = E(S_t|\sigma^k(S^{t-1}) = \sigma_p^k)\ ],$$

and (4)

$$V^k = [\ V_p^k = Var(S_t|\sigma^k(S^{t-1}) = \sigma_p^k)\ ].$$

Further, let $\hat{\mu}^k$ and $\hat{V}^k$ be the sample analogues of μ^k and V^k (that is the vector of cell means and within cell variances) from a randomly drawn sample from the population of interest. $\sqrt{N}^k$ will denote the vector containing the square root of the number of firms falling into each cell. Then, provided that the size realizations of the firms in the population are independent of one another, the central limit theorem and the law of large numbers imply that

$$diag[\sqrt{N}^k](\mu^k - \hat{\mu}^k) \rightsquigarrow N(0, diag[V^k])$$

while (5)

$$diag[\hat{V}^k] \xrightarrow{P} diag[V^k],$$

where diag[x] denotes a diagonal matrix with x on the principal diagonal, $\rightsquigarrow$ reads converges in distribution, $\xrightarrow{P}$ denotes convergence in probability, and $N(\cdot,\cdot)$ denotes the multivariate normal distribution.

Now consider possible values of $\sigma^* = [\sigma(S_{t-1}),\ldots,\sigma(S_{t-k})]$. The test for weak monotonicity of the regression function in S_1 is a test of whether, for all $\sigma^* \in [1,\ldots,J]^k$

$$\mu(\sigma^*,\ \sigma(S_1) = \sigma_1) \geqslant \mu(\sigma^*,\ \sigma(S_1) = \sigma_2)$$

whenever $\sigma_1 \geqslant \sigma_2$. Similarly the test of whether the realization of S_1 does not effect the regression function is a test of whether for $\sigma^* \in [1,\ldots,J]^K$,

$$\mu(\sigma^*,\sigma(S_1) = \sigma_1) = \mu(\sigma^*,\sigma(S_1) = \sigma_2)$$

whenever $\sigma_1 \neq \sigma_2$.

More formally assume that, for each σ^*, the vector μ^k is ordered by the associated values of $\sigma(S_1)$. Then each of the weak monotonicity constraints can be represented as a linear inequality constraint of the form $r'\mu^k \geqslant 0$, when $r' = [0,\ldots 0,-1,1,0,\ldots 0]$. Gathering all such constraints into the matrix R, the null hypothesis of weak monotonicity is written as

$$H_0^M:\ R\mu^k \equiv r \geqslant 0, \qquad (6).$$

Note that R is of full row rank, say C. We want a test of (6) under the maintained hypothesis that $r \in \jmath R^C$.

To this end we consider the following two estimators for r,

$$\hat{r} = R\hat{\mu} \qquad (7a)$$

and

$$\hat{r}^M = \arg\min_{r\geqslant 0}\ [(r-\hat{r})'R[\hat{V}^k]^{-1}R'(r-\hat{r})], \qquad (7b).$$

$\hat{r}$ is an 'unconstrained' estimator of r obtained from substituting sample for population means. $\hat{r}^M$ is a 'constrained' estimator, an estimator forced to satisfy the inequality constraint in (6). Subject to that constraint, it is obtained by minimizing a quadratic form in $(r-\hat{r})$, where the weighting matrix, $R[\hat{V}^k]^{-1}R'$, is chosen to be the variance-covariance of $\hat{r}$ under the null that $R\mu^k = 0$.

Since the quadratic form in (7b) is nonnegative and equal to zero if $r = \hat{r}$, if $\hat{r} \geq 0$, $\hat{r} = \hat{r}^M$. Figure (4) illustrates possible solutions for $\hat{r}^M$ in the case where C = 2. The ellipsoids represents sets of r which produce a constant $(r-\hat{r})'R[\hat{V}^k]^{-1} R'(r-\hat{r})$ value.

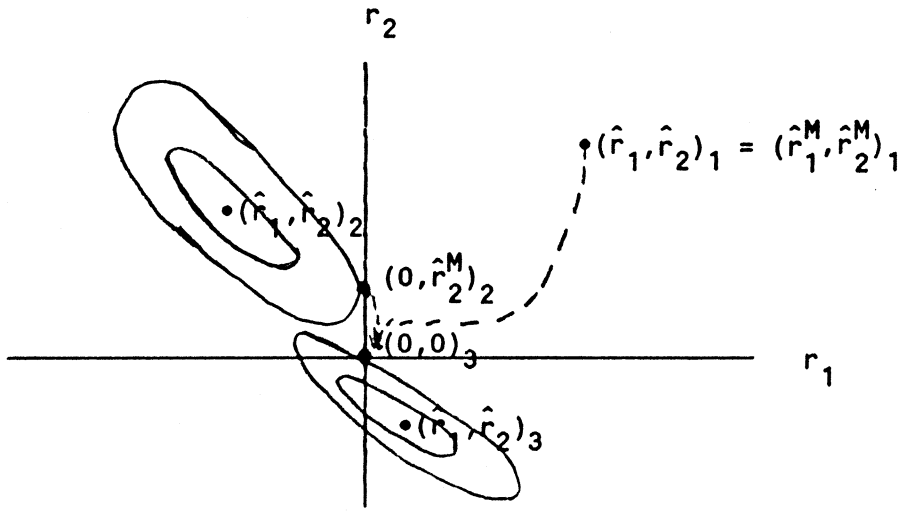


Figure 4. Constrained and Unconstrained Estimates of r

If

$$\chi^2_M = \min_{r \geq 0} (r-\hat{r})'R[\hat{V}^k]^{-1}R'(r-\hat{r}), \tag{8}$$

then large realized values of this statistic are evidence against H^0_M. Indeed, Barlow, Bartholemew, Bremner, and Brunk (1972) have shown that for all $a \geq 0$

$$T_M(a) = Pr\{\chi^2_M > a \mid r=0\} \rightarrow \sum_{c=0}^{C} W(c)Pr\{\chi^2_c > a\} \tag{9a}$$

as sample size grow large, where

$$W(c) = Pr(\hat{r}^M \text{ has exactly c zero components } | r=0), \qquad (9b)$$

and χ^2_c denotes a chi-square deviate with precisely c degrees of freedom (c=0,...,C). Thus, if χ^0_M is the realized value of χ^2_M, $T_M[\chi^0_M]$ provides the "p-value" (or the probability of type I error) of a test that would reject the null if $\chi^2_M = \chi^0_M$ when the true value of r was zero. The p-value when r is any value greater than zero cannot be larger.[7]

Unfortunately the orthant probabilities, that is the values of $(W(c))_{c=0}^{C}$ needed to obtain (9a), are difficult to calculate. As a result we obtain simulated estimates of their values, say $\hat{W}_c$, and provide a simulated estimate of $T_M(\cdot)$ say $\hat{T}_M$, where

$$\hat{T}_M[a] = \sum_{c=0}^{C} \hat{W}_c \, Pr(\chi^2_c > a) = \hat{W}'X \qquad (10)$$

and

$$\hat{W}' = [\hat{W}_0,\ldots,\hat{W}_C], \text{ whereas } X' = [P_r(\chi^1_0>a),\ldots,Pr(\chi^2_C>a)].$$

Since the $\hat{W}_c$ can be regarded as cell means from repeated draws from a multinomial distribution (where NSIM, the number of simulations, is the number of draws), the variance of $\hat{T}_M[a]$ about its expected value of $T_M[a]$ can be obtained from the formla for the variance of a multinomial as;

$$Var[\hat{T}_M(a)] = X'[\text{diag } W - WW']X(NSIM)^{-1}$$

So, along with $\hat{T}_M(a)$, we provide an estimate of its variance obtained from substituting the simulated for the actual values of W in this variance formula.

We now move on to the test of the null hypothesis that the regression function in (2) does not depend on s_1 conditional on it being nondecreasing in that

variable. That is we consider a test of

$$H_Z^0: \quad R\mu^k = r = 0$$

under the maintained hypothesis given by H_M^0 in (6). Once again $\hat{r}^M$ in (7b) will serve as our estimate of r given H_M^0, while under H_Z^0 the estimate of r is zero (thus, in figure 4, the ellipsoids bring us from $\hat{r}$ to the estimator which abides by H_M^0, while the dashed lines bring us from the latter to the estimator which abides by H_Z^0). A measure of the distance between the estimator obtained conditional on the null and the estimator which is only constrained to satisfy the maintained hypothesis is given by

$$\chi_Z^2 = \hat{r}^{M\prime} R[V^k]^{-1} R'\hat{r}^M, \qquad (11).$$

Once again, for all a>0

$$T_Z(a) = \Pr(\chi_Z^2 > a \mid r=0) \rightarrow \sum_{c=0}^{C} \tilde{W}(c) \Pr(\chi_c^2 > a) \qquad (12a)$$

as sample size grows large, where, in this case

$$\tilde{W}(c) = \Pr(\hat{r}^M \text{ has exactly c positive components} \mid r=0) \qquad (12b)$$

and χ_c^2 is defined as above (c=0,1,...,C). Letting χ_Z^{0} be the observed value of χ_Z^2, we will provide estimates of $T_Z[\chi_Z^0]$, say $\hat{T}_Z[\chi_Z^0]$ (obtained from simulating the $\tilde{W}(c)$), together with an estimate of the variance of $\hat{T}_Z[\chi_Z^0]$.

It is useful to compare this sequence of tests, that is the test for weak monotonicity under an unconditional maintained hypothesis coupled with the test of the hypothesis that s_1 has no effect on the regression function conditional on the maintained that any effect is nondecreasing, to the more familiar direct test of whether s_1 has no effect on the regression function conditional on an

unconstrained maintained hypothesis. One test of the latter would check whether a measure of the distance between $\hat{r}$ and 0, say

$$\chi_T^2 = \hat{r}'R[\hat{V}^k]^{-1}R'\hat{r}$$

is close to zero. Under the unconstrained maintained hypothesis χ_T^2 has the familiar chi-square distribution with C degrees of freedom. Since the properties of Lagrange multipliers insure that

$$[\hat{r}-\hat{r}^M]'R[V^k]^{-1}R'\hat{r}^M = 0,$$

we have from (7) and (11),

$$\chi_T^2 = \chi_M^2 + \chi_Z^2$$

with probability one. That is the observed value for the test of no effect of s_1 conditional on an unconstrained maintained, say χ_T^0, will be just the sum of χ_M^0 and χ_Z^0. For comparison, our tables will also provide the p-value of χ_T^0, $T_T[\chi_T^0]$ (these can be found in standard tables).

Section 5. The Data and the Empirical Results

The data used in this study were obtained from the Wisconsin Department of Industry Labor and Human Relations' (DILHR's) records for unemployment insurance (UI) coverage. The records for the years between 1978 and 1986 (inclusive) were linked together by UI account number by David Neuendorf and Ron Shaffer (see Neuendorf and Shaffer, 1987).[8]

Any private employer hiring at least one worker and paying at least $1,500 in a quarter is required to file information on the number of workers, wages, and UI tax contributions to DILHR. For the purposes of our analysis the first time it does so is treated as the 'birth' of the firm. Size in that, and in subsequent, years is measured by the number of employees.

The unit used to match observations over time was the UI account number. When a new business changes ownership or legal status, DILHR freezes its current account and either creates a new account, or, in the case of an acquisition, merges the employment information into another account. When this occurs the old account has a successor code, and a new account, if created, will have a predecessor code. New accounts which were a result of a change in legal status (and therefore had a predecessor code) were separated out and not treated as a part of a birth cohort in this analysis. Analogously we use the successor code to distinguish between attrition due to liquidation, and attrition due to mergers (and other changes in legal status). A major advantage of this type of data is that it can distinguish between these two sources of 'exit'.

Tables 2 and 3 provide information on the evolution of the size distribution of the surviving firms from the 1979 birth cohort in retail and in manufacturing, respectively (recall, from table 1, that these two sectors account for

Table 2: Evolution of Size Distribution Over Age Retail: 1979 Cohort
(Entries are proportion of active firms with employment ≥ X)

	Age								Cross Section[a]	
X	1	2	3	4	5	6	7	8	1978	1986
1	67.0	73.3	76.8	77.7	78.2	80.0	80.3	83.9	85.5	82.5
2	47.6	52.0	57.5	57.8	58.7	62.9	63.1	66.0	72.5	74.8
3	34.6	40.5	42.9	45.7	47.9	51.0	50.5	53.6	61.3	64.1
4	26.4	33.7	34.9	36.5	37.7	41.0	40.1	43.7	52.2	55.2
5	22.3	28.2	29.4	30.7	32.5	35.0	34.2	38.3	45.0	47.8
10	11.1	12.7	13.7	14.7	17.2	17.5	18.9	21.7	25.5	27.5
15	6.7	7.5	8.7	10.1	10.0	9.6	10.8	14.6	16.9	18.6
20	5.3	6.2	6.4	7.0	7.5	7.9	8.2	9.8	12.2	13.6
25	4.2	4.9	5.4	5.6	5.7	6.0	6.7	7.3	9.3	10.6
30	3.1	3.7	3.7	4.5	4.6	5.3	6.3	6.2	7.2	8.5
50	1.0	1.0	1.5	1.8	1.8	2.1	2.2	3.2	3.3	4.2
Count	1180	973	816	713	610	571	539	464	22,568	23,435
Mean	5.42	5.98	6.41	6.87	7.14	7.71	7.85	8.80	14.02	15.23
Mortality Rate	17.54	13.31	8.73	8.73	3.31	2.71	6.27	[60.67[b]]		
Hazard Rate	17.54	16.14	12.62	14.45	6.39	5.60	13.73			
Number Subsequently "Transferring Out"[c]	95	72	59	5	3	1	0			

Notes to table 2:

a. Size distribution of all firms active in 1978 (1986) regardless of birth cohort
b. Mortality rate over the eight year period.
c. These are firms active at the relevant age but who "transferred out", due to a change in legal status, at some point thereafter. They are not included in the size distribution calculations at that age.

Table 3: Evolution of Size Distribution Over Age Manufacturing: 1979 Cohort
(Entries are proportion of active firms with employment ≥ X)

	Age								Cross Section[a]	
X	1	2	3	4	5	6	7	8	1978	1986
1	71.9	78.6	80.9	86.9	86.3	89.5	90.2	90.9	93.3	92.1
2	49.9	61.2	65.1	71.7	73.1	80.8	82.9	82.5	87.2	89.0
3	38.8	49.6	55.7	63.6	64.3	69.8	75.0	72.7	80.7	78.4
4	32.4	40.6	44.3	53.0	55.0	62.8	66.5	65.0	74.9	72.9
5	25.1	33.7	38.8	45.0	47.8	55.2	57.9	57.8	70.6	68.4
10	8.6	18.1	20.9	21.7	23.1	31.4	31.4	34.4	54.1	51.3
15	4.3	9.1	9.5	13.1	15.4	19.2	18.9	22.7	43.9	40.3
20	3.1	3.3	6.0	9.1	9.3	12.2	12.8	15.6	36.8	34.1
25	2.8	2.2	3.4	4.6	6.0	9.3	9.8	12.3	31.2	29.3
30	2.1	1.5	2.1	2.0	5.0	7.6	8.5	9.7	28.0	25.9
50	.6	.7	.9	1.5	2.8	4.1	6.1	6.5	19.6	18.3
Count	327	276	235	198	182	172	164	154	6,987	7,789
Mean	4.92	6.27	7.09	8.10	8.79	10.79	12.38	13.34	73.81	61.70
Mortality Rate	15.60	12.54	11.31	4.89	3.06	2.45	3.06	[52.91[b]]		
Hazard Rate	15.60	14.86	15.74	8.08	5.49	4.65	6.10			
Number Subsequently "Transferring Out"[c]	13	11	10	1	0	0	0			

Notes to table 3:

Notes a,b, and c, are identical to the same notes in Table 2.

80 percent of the employment in our sample). The row labelled 'count' gives the number of firms active in the column age. The row labelled transferring out provides the number of firms which were active in the column year but transferred out (due to a change in legal status) before 1986. This source of attrition accounts for about 8% of the 1979 cohort in retail trade, and about 4% in manufacturing. This should be compared to the extent of liquidation (the figures given in the row labelled mortality rates). Over 60% of the 1979 birth cohort in retail liquidated before 1986, and the analogous figure in manufacturing was over 50%. Since liquidation was quantitatively so much more important a source of attrition in these data, we simply omitted those firms who subsequently changed ownership from the analysis.[9]

The passive learning model implies that the proportion of surviving firms with size greater than any X, or the numbers in each row of the body of the tables, should increase with age (i.e., as we move from left to right on the table). We have 'squared off' the adjacent transitions which do not satisfy this condition. On the whole, the consistency of the data with the hypothesis is quite striking -- particularly in retail. Of the seventy-seven possible adjacent transitions, only six are decreasing, and none of them indicate a fall of more than 1.0%. In manufacturing there are nine transitions which decrease; two fall by more than 1.5%, and two more by .6%. Given the possibilities for reporting and recording errors in this type of data (see Neuendorf and Shaffer, 1987), if the null were true, we would not find these results to be 'surprising'. That is, to us these results are quite consistent with the implications of passive learning -- indeed amazingly so for retail trade. Note also that, in both sectors, the means are strictly increasing in age.

There are also some interesting contrasts in the evolution of the size distribution between the two sectors. The size distribution in the initial year is not much different between the two sectors; indeed if anything the initial size distribution is slightly 'larger' in retail trade (retail has the larger initial year mean, 5.4 vs. 4.9, and a higher percentage of the firms in the largest size classes). However, by age eight this ordering has turned around. That is, by age eight the size distribution for manufacturing is stochastically larger (even in the strict sense) than that in retail (the means are 13.3 vs 8.8, and manufacturing has over twice the fraction of firms with 50 or more employees). The size distribution is stochastically increasing in age in both sectors, but it is increasing at a much more rapid rate in manufacturing.

Moreover, the age eight distribution in retail is quite close to the cross-sectional distribution of all retail firms active in 1978 (or 1986, see the last two columns of the table). Both have about 3% of their firms with more than 50 employees (though the cross-sectional distribution still has the larger mean, 14 vs. 9). In contrast, the age eight distribution in manufacturing is much smaller than the 1978 cross-sectional distribution in that sector. In manufacturing the cross-sectional distribution has more than three times the fraction of firms with more than 50 employees (19.6 vs. 6.5), and a mean which is almost six times that from the age eight distribution (73.8 vs. 13.3). If we were to think of the cross-sectional distribution as an approximation to the limit distribution (even though formally it is not), then we might conclude that by age eight the retail cohort had almost reached it, but the manufacturing cohort was still nowhere near its limit distribution. Indeed, if we also assumed that eight years was enough time to form a fairly precise posterior about a time invariant profitability parameter, then we would conclude that

the data from retail was supportive of the passive learning model, but the data from manufacturing was not.

A more formal check of the consistency of the data with the two models can be derived from an analysis of the regression for size at age eight on size in the immediate preceding periods, and size at age one. Both models imply that this function will be weakly increasing in initial size, but the passive learning model implies that it will be strictly increasing in that variable, and the active learning model implies that it will not.

Tables 4 and 5 provide some evidence on the relevant hypothesis. Because there were less than half the number of entering firms annually in manufacturing, we aggregated the 1979 and 1980 manufacturing cohorts and examined the regression for expected sales at age seven of the aggregated cohort. The cell size cutoffs were set at the beginning of the analysis and not changed thereafter. For the weak monotonicity, and the zero conditional on monotonicity, restrictions, we have presented two sets of 'p-values' for each observed value of the test statistic. The first column provides the simulated estimates of the true p-values as explained in section 4 (the estimated standard errors of these estimates appear in parenthesis below their values). The second column provides the p-value that would be obtained if each 'orthant' had equal probability. In this case

$$\Pr\{\chi^2_M > a\} = [\sum_{c=0}^{C} \binom{C}{c} \Pr\{\chi^2_c > a\}]/2^C,$$

which can be easily calculated (here $\binom{y}{x}$ is notation for the numbers of combinations of y elements taken x at a time). The orthants would have equal probability if the constraints being tested were independent of one another - which they are not. On the other hand the figure in column (2) is trivial to

Table 4. Tests for Mean Independence of the Distribution of S_t Conditional on $S_{t-1}, \ldots, S_{t-k}$, from S_1

Data: Retail, 1979 Cohort and t=8,[a]

Size Cutoffs: 2,5,10,25,50, $+\infty$

k	Weak Monotonicity				Zero Conditional on Montonicity				Unconditional Zero		
	C	χ^0_M	p-values[b] (1)	(2)	C	χ^0_z	p-values[b] (1)	(2)	Df	χ^0_T	p-value
1	17	1.1	1.00 (.00)	.99	17	37.2	.00 (.00)	0	17	38.2	.00
2	22	6.5	.88 (.03)	.80	22	23.9	.00 (.00)	.02	17	30.4	.11
3	25	11.5	.66 (.05)	.52	25	28.0	.00 (.00)	.01	25	39.5	.03
4	22	19.1	.05 (.01)	.05	22	19.1	.04 (.01)	.08	22	38.2	.02
5	19	17.6	.05 (.01)	.07	19	13.6	.12 (.02)	.19	19	31.2	.04

[a] Cohort dimensions: number in cohort = 1,275; number of firms reaching age eight = 464.

[b] The value in column (1) is a simulated estimate of the true p-value and the value just below it is the standard error of this estimate. Ten simulation draws were used to calculate the estimates of the orthant probabilities. The value in column (2) is obtained by assuming each orthant has equal probability (see the explanation in the text).

Table 5. Tests for Mean Independence of the Distribution of S_t Conditional on $S_{t-1}, \ldots, S_{t-k}$, from S_1

Data: Manufacturing, Combined 1979 and 1980 Cohorts for t = 7.[a]

Size Cutoffs: 2,5,10,25,50, + ∞

k	Weak Monotonicity C	χ^0_M	p-values[b] (1)	(2)	Zero Conditional on Monotonicity C	χ^0_z	p-values[b] (1)	(2)	Unconditional Zero Df	χ^0_T	p-value[b]
1	16	8.0	.54 (.06)	.44	16	3.5	.57 (.07)	.86	16	11.5	.78
2	25	17.6	.19 (.03)	.17	25	5.8	.79 (.03)	.91	25	23.6	.55
3	23	14.3	.28 (.05)	.27	23	4.9	.81 (.06)	.92	23	19.3	.67
4	15	10.1	.13 (.02)	.24	15	5.9	.54 (.03)	.59	15	16.0	.39

[a] Firm dimensions: number born in cohorts = 737, number of firms reaching age seven = 353.

[b] See note b to Table 4.

calculate and, if it tended to be very close to the figure in column (1), one might use it as a preliminary indication of the true p-value in situations where a 'close guess' might do (a strategy we actually followed). A comparison of column (1) to column (2) therefore provides some indication of just how close the guess would be for problems with structures similar to ours. The answer seems to be, quite close.

Note first that there is no evidence against weak monotonicity in either retail or manufacturing. So both data sets seem to be consistent with the hypothesis that the regression function is nondecreasing in s_1, just as both our models predict. There the similarity in the test results on the two data sets ends. In retail it is clear that if we condition on one lagged value of S, that is on realizations of S_7, and then vary s_1, firms with larger s_1 have larger expected sales at age 8. There is really no doubt about this point as the p-value of the test statistic is essentially zero, so we would reject the null at any traditional significance level. The same is true if we condition on s_7 and s_6; or on s_7, s_6 and s_5; or even on s_7, s_6, s_5 and s_4; and then vary s_1. In all these cases realizations of S_1 have an independent effect on the expectation of sales at age eight. This dependence only starts to become insignificant at five percent significance levels when we condition on five past sales realizations. However, this might well be a result of the possibility that, with our limited amount of data, a fifth order nonparametric autoregression would provide an adequate approximation to the expectation for size generated from any stochastic process - (ϕ-mixing or not; we come back to this point below).[10]

The results for the test of zero conditional on weak monotonicity are strikingly different in manufacturing. Table 5 indicates that, in manufacturing, once we condition on a single lagged value of S, i.e. a realization of

S_6, any differences in s_1 do not effect the expected size at age seven. This time there is little doubt about accepting the null as the p-value is well above .5. Moreover, the same results obtain if we condition instead on s_6 and s_5; or on s_6, s_5, and s_4; or s_6, s_5, s_4 and s_3.

Tables 6 and 7 push the nonparametric analysis one step further and ask what order of Markov process provides an adequate nonparametric fit to the (expectation from the) stochastic process generating size conditional on survival in retail and in manufacturing. The tests in these tables follow a pattern analogous to that in tables 4 and 5. That is, we first test whether first year size, size in the first two years, ..., have a nondecreasing effect conditional on the variables left in the regression function; and then test whether we can accept a zero effect conditional on any of the existing effects being nondecreasing. Again the results are quite clear. We never reject weak monotonicity. In retail we need a fifth order nonparametric Markov process to adequately approximate the data. Recall that this is precisely the same 'k' we needed before we could accept the null that the conditional regression function for size, conditional on s_{t-1}, ..., s_{t-k}, did not depend on s_1. In contrast, in manufacturing a third order nonparametric Markov process seems to provide an adequate fit to the data. That is, in manufacturing there is a distinction between the orders needed for the ϕ-mixing and the Markov tests (compare tables 7 and 5). Table 5 says that conditional on realizations of S_6 realizations of S_1 do not affect the regression function. Table 7 says that realizations of S_5, and of S_4, do. The active learning model explains this difference by allowing the parameter that determines the size distribution to evolve over time in a 'smooth' fashion, so that its value in year 5 will tend to be closer to its

Table 6. Markov Tests for Properties of Retail Regression Function for Size at Age Eight

Data: Retail, 1979 Cohort[a]

Size Cutoffs: 2,5,10,25,50, + ∞[b]

Markov Order for Tests	Weak Monotonicity C	χ^0_M	p-value[c] (1)	(2)	Markov Conditional on Monotonicity C	χ^0_z	p-value[c] (1)	(2)	Unconditional Markov Df	χ^0_T	p-value
7 → 6	13	9.5	.13	.20 (.03)	13	5.0	.48	.58 (.05)	13	14.5	.34
7 → 5	23	18.3	.16	.11 (.02)	23	5.9	.64	.87 (.05)	23	24.2	.40
7 → 4	32	18.3	.47	.32 (.05)	32	96	.00	.00 (.00)	32	114	.00
7 → 3	38	18.7	.56	.48 (.05)	38	100	.00	.00 (.00)	32	118	.00
7 → 2	43	19.7	.76	.56 (.03)	43	107	.00	.00 (.00)	43	121	.00
7 → 1	48	20.1	.92	.67 (.01)	48	149	.00	.00 (.00)	48	169	.00

[a]Cohort Dimensions: number in cohort = 1275; number of firms reaching age eight = 465; number in cells with ≥ 2 = 291.

[b]Cell Dimensions: possible number = 279,936; number populated 228; number with ≥ 2 observations = 54.

[c]See note b, Table 4.

Table 7. Tests for Properties of Manufacturing Regression Function for Size at Age Seven

Data: Manufacturing, Combined 1979 and 1980 Cohorts[a]

Size Cutoffs: 2,5,10,25,50, + ∞[b]

Markov Order for Tests	Weak Monotonicity C	χ^2_M	p-value[c] (1)	(2)	Markov Conditional on Monotonicity C	χ^2_z	p-value[c] (1)	(2)	Unconditional Markov Df	χ^2_τ	p-value
6 → 5	9	11.9	.02	.04 (.01)	9	2.0	.65	.75 (.10)	9	14.0	.12
6 → 4	15	13.3	.09	.10 (.02)	15	11.7	.07	.16 (.02)	15	25.1	.05
6 → 3	25	15.5	.24	.27 (.05)	25	17.6	.11	.17 (.03)	25	33.1	.13
6 → 2	31	16.1	.42	.42 (.04)	31	61.3	.00	.00 (.00)	31	77.4	.00
6 → 1	37	16.3	.66	.59 (.04)	37	76.0	.00	.00 (.00)	37	92.3	.00

[a]Cohort Dimensions: number of firms = 737; number of firms reaching age seven = 353; number in cells with ≥ 2 = 179.

[b]Cell Dimensions: possible number = 46,656; number populated 217; number with ≥ 2 observations 43.

[c]See note b, Table 4.

value in year 7, and therefore have a more distinct effect on the regression function for S_7, then its value in year 1 will.[11]

Section 6. Concluding Remarks

Our empirical results can be summarized quite succinctly. The nonparametric implications of the active learning model are consistent with the data in manufacturing, while the nonparametric implications of the passive learning model definitely are not. On the other hand, the nonparametric implications of the passive learning model seem consistent with the data in retail trade, while those from the active learning model do not. These distinctions ought to effect the type of models we use to analyze phenomena that are tightly tied to firm-specific uncertainty and differences in output paths among firms within an industry; phenomena such as the behavior of capital markets when there are significant failure probabilities, or the evolution of the size distribution of output among firms within an industry.

They also ought to effect how we account for liquidation induced attrition in the analysis of longitudinal firm-level data (with or without a detailed model of liquidation). As an example, consider the following excerpt from Davis, Gallman, and Hutchins, "Productivity in American Whaling: The New Bedford Fleet in the Nineteenth Century."

> "The age of the vessel (entered as age and age squared) also captures the effects of more than a single set of factors. Elements of wear and tear that influenced productivity, a technical characteristic that one might hope to capture in the age variable, are confounded with the consequence of qualitative differences among survivors; ineffective vessels

> were transferred by their owners to other activities, were condemned at an early age, or were destroyed in service."
>
> Davis, Gallman and Hutchins (1987) p.26.

This quotation illustrates how even one of the most traditional of variables (age), in one of the most traditional of settings (productivity analysis), can have its "structural" effects (as a measure of the likely extent of physical deterioration) confounded by the self-selection process induced by the endogeneity of the liquidation decision (it also demonstrates a remarkable understanding of the environment generating the data). Davis, Gallman and Hutchins (1987) do indeed find a significant positive first order effect of age on vessel productivity. Our models would allow one to separate out the structural coefficients by adding equations to account for the selection process. If the active learning model were relevant then the selection equations should be based on productivity realizations in the immediately preceding periods, but if the passive learning model were, then both age, and earlier years' productivity, will also determine the selection probabilities. The simple nonparametric procedures detailed in the previous sections ought to provide guidance as to which of the alternatives seems relevant.

The results in section 5 also illustrate two more technical points. First, it is possible to develop computationally simple, yet rigorously correct, checks for the relevance of alternative stochastic control models with discreteness in their choice sets that are independent of precise assumptions on the functional forms of interest. When possible, we think that the nonparametric implications of stochastic control models should be checked against data before

sinking alot of resources into building an estimation algorithm for a particular parametric form of a model. Finally, if we are willing to use either theory (as we have done), or ad hoc argumentation to restrict our use of state dependence to refer to state dependence in ergodic processes, then there is a natural test for the distinction between state dependence and heterogeneity based on ϕ-mixing. Heterogeneity implies that initial and later years realizations of a process will be dependent -- no matter the time that elapses in the interim -- whereas state dependence, in an ergodic setting, implies that they will not be. The distinction can often be made more powerful by conditioning on years just prior to the current year, and doing so made the tests quite effective in distinguishing among the alternatives on our, moderately sized, eight-year panels.

Footnotes

1. See Heckman and Robb, 1985, and the literature cited there, for a discussion of related issues in labor economics.

2. Two points should be noted here. First we are ignoring the effect (on both $\pi(\cdot)$ and $S(\cdot)$) of random variables which have the same value for different individuals at the same point in time, but differ in value over time (this would have occurred in our example if prices had varied over time). At the cost of complicating the notation we could add a price process to our problem without changing any of our major results (though some modifications would have to be made to the procedure that matches the model to data; see below.) Second, it should be noted that the interpretation of $\pi(\cdot)$ and $S(\cdot)$ as mappings from realizations of η, would only be appropriate for our example if η were realized before input decisions were made (Marschak and Andrews, 1944). In this case both output and inputs can be determined from η_t, and the size measure can be either output produced or inputs purchased. The extreme alternative is to assume there is no within-period adjustment to η (Zellner, Kmenta, and Dreze, 1966), in which case inputs are chosen to maximize $\alpha_{t+1} E_{(t)} \eta_{t+1} F(\ell_{t+1}) - w_t \ell_{t+1}$, where E_t provides expectations conditional on current information (and will be defined more precisely below). In this case $\pi(\cdot)$ and $S(\cdot)$ would be interpreted as mappings from $E_t \eta_{t+1}$ to $E_t \pi_{t+1}$, and input demand in period t+1 respectively. There are, of course, intermediate cases where within period adjustment is either partial, or more costly (the appropriate characterization is likely to depend upon the characteristics of the industry being studied). We shall discuss the various alternatives in more detail in the empirical section, but for now suffice it to note that the results we focus attention on do not depend on the timing of the input decision.

3. The following counterexample shows that this would not be the case if we were to assume only a weaker first order stochastic dominance ordering. Let $\theta = (\theta_1,\theta_2)$ with $\theta_2 > \theta_1$, and consider the following family of densities (with respect to a counting measure): $p(\eta = 2 \mid \theta_2) = p(\eta = 4 \mid \theta_2) = 1/2$, and $p(\eta = 1 \mid \theta_1) = p(\eta = 3 \mid \theta_1) = 1/2$. Clearly, $P_\eta(\cdot \mid \theta_2) \gtrsim_s P_\eta(\cdot \mid \theta_1)$. However if $\eta_1 = 2$, the posterior is $\theta = \theta_2$ with probability one, whereas if $\eta_1 = 3$, the posterior is $\theta = \theta_1$ with probability one; i.e., the posterior for $\eta = 2$ dominates the posterior for $\eta = 3$.

4. The assumptions that Φ is the same known value for all agents, and is constant over time, are made for expositional convenience. What is required is that Φ not increase too rapidly with n^t. More precisely, if $V_t(n^t)$ is the value of continuing in operation at t given that $\eta^t=n^t$ (a more precise definition of this function is given below), then what we need is that $V_t(n^t) - \Phi_t(n^t)$ be nondecreasing in n^t. Of course, the actual behavior of "exit values" is an empirical question. If the process generating the exit we are modelling is indeed a liquidation process, and not a process generated by changes of ownership and continued operation of the firm in a different guise, the assumptions we require ought not to be problematic.

5. Ericson and Pakes (1987) also consider the more detailed theoretical implications delivered by particular parametric examples. The parametric families investigated were those that seemed suitable for the econometric specification of estimable forms of the active learning model.

6. Just as in our description of the passive learning model we will assume here, for expositional simplicity, that input choices are made after the realization of η, and that liquidation values are a constant Φ. Further the formulation presented here assumes that the conditional distribution of τ does not

depend on ω, an assumption not required for our results (see Ericson and Pakes, 1987).

7. The reader interested in more detail on the testing procedures used in this action should consult Barlow et al (1972), or the more recent econometric literature on testing subject to inequality constraints which begins with the work of Gouriéroux, Holly, and Monfort (1982). Golberger's (1987) exposition is particularly clear.

8. We are grateful to them for granting us access to their data, and for graciously answering our subsequent queries. More detail on the data can be found in the appendix of Neuendorf and Shaffer (1987). Though multiestablishment firms have a choice as to whether to report as a single, or as multiple, units, we have, where possible, merged the establishments of multiestablishment firms. This should therefore, be thought of as firm-level data.

9. However, as one might guess from the figures, we get very similar results when we leave these firms in until the year they transfer out.

10. We have been motivating our two-part testing sequence as a way of providing additional information on the relevance of alternative models. Inequality tests were originally motivated as providing more powerful ways of testing a given null. Table 4 also illustrates this point. Take, for example, the case where k=2. The p-value in column 2 for acceptance of the null that realizations of S_1 do not matter under the maintained hypothesis that any effect of S_1 is non-decreasing, is zero; but the p-value for the test that S_1 does not matter under the unconstrained maintained hypothesis (the unconditional zero columns) is a traditionally acceptable .11.

11. Footnote 2 discussed the possibility that input decisions are either wholly, or partially, made before the realization of η, and concluded by

asserting that the various alternatives would not affect the results we focus on. Table 7 insures this is so for the very special, but important, case which Jovanovic's (1982) original article was based on. His assumptions were a special case of the following ones; the process generating $\{\eta_t\}$ conditional on θ was i.i.d., the posterior for θ had sufficient statistics (x_t, t) with $x_t = f_t(x_{t-1}, \eta_t)$ for some $f_t(\cdot)$, and that no input could be adjusted after any information about η_t was available. In this case, if input quantities were our size measure, size in period t is determined by (x_{t-1}, t) and for a given t, there is a 1:1 correspondence between S_{t-1} and x_{t-2}. So size is a first order Markov process. This conclusion would be destroyed if some, say costly, adjustments could be made after η were realized, or if there were any dependence in the process generating $\{\eta_t\}$ conditional on θ. However, if Jovanovic's restrictions were true, the passive learning model would satisfy the constraint that the regression for S_t conditional on $S_{t-1}, \ldots, S_{t-k}$ does not depend on S_1 provided $k \geq 1$; i.e., it would satisfy the constraint used to test for the active learning model. On the other hand Table 7 mades it clear that the stochastic process generating size is not first order Markov, so the special case discussed by Jovanovic (1982) is not relevant.

References

Barlow R.E., D.J. Bartholemew, J.M. Bremner, & H.D. Brunk (1972); Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression, New York; John Wiley & Sons.

Billingsley P. (1968); Convergence of Probability Measures, New York; John Wiley & Sons.

Billingsley P. (1979); Probability and Measure, New York; John Wiley & Sons.

Chamberlain G. (1985); "Heterogeneity, Omitted Variable Bias, and Duration Dependence", in J. Heckman and B. Singer (ed.); Longitudinal Analysis of Labor Market Data, Cambridge University Press.

Churchill B. (1955); "Age and Life Expectancy of Business Firms", Survey of Current Business, Vol. 25, pp. 15-20.

Davis L., R. Gallman, & T. Hutchins (1987); "Productivity in American Whaling: The New Bedford Fleet in the Nineteenth Century", Department of Economics, the California Institute of Technology.

Dunne T., M. Roberts, & L. Samuelson (1987); "Plant Failure and Employment Growth in the U.S. Manufacturing Sector", Department of Economics, The Pennsylvania State University.

Ericson R., & A. Pakes (1987); "An Alternative Theory of Firm Dynamics", Department of Economics, Columbia University.

Evans D. (1987a); "The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries", Journal of Industrial Economics, Vol. 35, pp. .

Evans D. (1987b); "Tests of Alternative Theories of Firm Growth", Journal of Political Economy, Vol. 95, pp. 657-674.

Goldberger A. (1987); "One-sided and Inequality Tests for a Pair of Means", Social Systems Research Institute (No. 8629), University of Wisconsin.

Gouriéroux C., A. Holly, & A. Monfort (1982); "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints", Econometrica, Vol. 50, pp. 63-80.

Heckman J. (1981); "Statistical Models for Discrete Panel Data", in C. Manski, & D. McFadden (ed.); Structural Analysis of Discrete Data with Econometric Applications, pp. 114-178.

Heckman J., & B. Singer (1985); "Social Science Duration Analysis", in J. Heckman, and B. Singer (ed); Longitudinal Analysis of Labor Market Data, Cambridge University Press.

Heckman J., and R. Robb (1985); "Alternative Methods for Evaluating the Impact of Interventions", in J. Heckman and B. Singer (ed.), Longitudinal Analysis of Labor Market Data, Cambridge University Press, 1985.

Jovanovic B. (1982); "Selection and the Evolution of Industry", Econometrica, 50, pp. 649-70.

Lippman S., & R. Rumelt (1982); "Uncertain Immitability: An Analysis of Interfirm Differences in Efficiency Under Competition", Bell Journal of Economics, Vol. 13, pp. 418-438.

Marschak J., & W.H. Andrews Jr. (1944); "Random Simultaneous Equations and the Theory of Production", Econometrica, Vol. 12, pp. 143-205.

Milgrom P. (1981); "Good News and Bad News: Representation Theorems and Applications", the Bell Journal of Economics, Vol. 12, pp. 380-391.

Miller R. (1984); "Job Matching and Occupational Choice", the Journal of Political Economy, Vol. 92, pp. 390-409.

Neuendorf D., & R. Shaffer (1987); "Private Sector Economic Changes in Dane County: 1978-1986", Department of Agricultural Economics, University of Wisconsin.

Pakes A. (1986); "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks", Econometrica, Vol. 54, pp. 755-784.

Ross S. (1983); Introduction to Stochastic Dynamic Programming, New York and London, Academic Press.

Rust J. (1987); "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", Econometrica, Vol. 55, pp. 999-1034.

Wedervang F. (1965); The Development of a Population of Industrial Firms, Scandinavian University Books.

Wolpin K. (1984); "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality", the Journal of Political Economy, Vol. 92, pp. 852-874.

Zellner A., J. Kmenta, & J. Dreze (1966); "Specification and Estimation of Cobb-Douglas Production Functions", Econometrica, Vol. 34, pp. 784-95.

Appendix I (Proof of Theorem 2.7)

The proof proceeds as follows. First it considers the finite horizon problem in which a firm which remains active until period T must liquidate for Φ dollars at T+1. For this problem the value of continuing in operation from period t (as a function of past η-realizations) will be denoted by $V_t^T(\cdot)$: $N^t \to \mathbb{R}_+$, and the resulting stopping function by $\chi_t^T(\cdot)$: $N^t \to [0,1]$. $V_t^T(\cdot)$ can be determined by backward recursion from the terminal year and a stopping policy which dictates liquidation if and only if the value of continuing in operation is less than Φ. The implied stopping function, $\chi_t^T(n^t)$, is one if and only if $n^t \in A_t^T = (n^t: V_1^T(n_1^t) \geq \Phi, V_2^T(n_1^t, n_2^t) \geq \Phi, \ldots, V_t^T(n^t) \geq \Phi)$. As T increases $V_t^T(\cdot)$ converges (pointwise) to a limit function, $V_t(\cdot)$. This limit function is bounded, monotonic (in each component of) n^t, and satisfies the Bellman condition, i.e. equation 6, in the text. The proof concludes by showing that $V_t(\cdot)$, and the associated limit stopping policy, $\chi_t(\cdot)$, are indeed the solution to the infinite horizon problem.

A1 Lemma $P_{\eta_{t+1}}(\cdot \mid n_1^t) \underset{\sim s}{\geq} P_{\eta_{t+1}}(\cdot \mid n_2^t)$ whenever $n_1^t \geq n_2^t$ $(n_1^t, n_2^t \in N^t$, and all t)

Proof Take any $z \in N$. Then $P_{\eta_{t+1}}(z \mid n^t) = \int P_{\eta_{t+1}}(z \mid n^t, \theta) P_\theta(d\theta \mid n^t)$. $P_{\eta_{t+1}}(z \mid n^t, \theta)$ is nonincreasing in n^t by (3.i) and strictly decreasing in θ by (3.ii), while $P_\theta(\cdot \mid n^t)$ is stochastically increasing in n^t by (4). [ ]

A2 Lemma Fix any T then, $V_t^T(n_1^t) \geq V_t^T(n_2^t)$ whenever $n_1^t \geq n_2^t$ $(n_1^t, n_2^t, \in N^t$, and $t \leq T)$.

Proof The proof is by backward induction on t. Note that

$$V_T^T(n^T) = \int \pi(\zeta) P_{\eta_{T+1}}(d\zeta \mid n^T) + \beta\Phi$$

which is nondecreasing in n^T by virtue of the monotonicity of $\pi(\cdot)$, and A1. Now assume monotonicity at t+1. Then if $n_1^t \geqslant n_2^t$

$$V_t^T(n_1^t) = \int \pi(\zeta) P_{\eta_{t+1}}(d\zeta \mid n_1^t) + \beta \int \max[\Phi, V_{t+1}^T(\zeta, n_1^t)] P_{\eta_{t+1}}(d\zeta \mid n_1^t)$$

$$\geqslant \int \pi(\zeta) P_{\eta_{t+1}}(d\zeta \mid n_2^t) + \beta \int \max[\Phi, V_{t+1}^T(\zeta, n_1^t)] P_{\eta_{t+1}}(d\zeta \mid n_2^t)$$

$$\geqslant \int \pi(\zeta) P_{\eta_{t+1}}(d\zeta \mid n_2^t) + \beta \int \max[\Phi, V_{t+1}^T(\zeta, n_2^t)] P_{\eta_{t+1}}(d\zeta \mid n_2^t) = V_t^T(n_2^t),$$

where the inequalities are due to A1, the monotonicity of $\pi(\cdot)$, and the hypothesis of the inductive argument. [ ]

A3 Lemma Fix T. Then, $V_t^{T+1}(n^t) \geqslant V_t^T(n^t)$ ($n^t \in N^t$, and $t \leqslant T$)

Proof. The proof is again by backward induction on t. For the initial condition of the inductive argument, note that

$$V_T^{T+1}(n^T) = \int \pi(\zeta) P_{\eta_{t+1}}(d\zeta \mid n^T) + \beta \int \max[\Phi, V_{T+1}^{T+1}(\zeta, n^T)] P_{\eta_{t+1}}(d\zeta \mid n^T)$$

$$\geqslant \int \pi(\zeta) P_{\eta_{t+1}}(d\zeta \mid n^T) + \beta\Phi = V^T(n^T).$$

Assuming the condition is true for a = t+1 we have

$$V_t^{T+1}(n^t) = \int \pi(\zeta) P_{\eta_{t+1}}(d\zeta \mid n^t) + \beta \int \max[\Phi, V_{t+1}^{T+1}(\zeta, n^t)] P_{\eta_{t+1}}(d\zeta \mid n^t)$$

$$\geqslant \int \pi(\zeta) P_{\eta_{t+1}}(d\zeta \mid n^t) + \beta \int \max[\Phi, V_{T+1}^{T}(\zeta, n^t)] P_{\eta_{t+1}}(d\zeta \mid n^t) = V_t^T(n^t). \quad [ ]$$

Proof of Theorem 2.7

Lemma A3 insures that for each (t, n^t) the limit,

$$V_t(n^t) = \lim_{T \to \infty} V_t^T(n^t),$$

exists. Let $\sup_{\eta\in N}\pi(\eta) = \bar{\pi}$ [$\bar{\pi}$ exists and is finite by virtue of the compactness of N and the continuity of $\pi(\cdot)$]. It is straightforward to show that $V_t^T(\cdot)$ is bounded, uniformly over t, by the constant function $(1-\beta)^{-1}\max[\bar{\pi},\Phi]$. Since boundedness and (weak) monotonicity are preserved by limit functions, this insures that $V_t(n^t)$ is monotonic and bounded. Also

$$\lim_{T\to\infty} V_t^T(n^t) = \int\pi(\zeta)P_{\eta_{t+1}}(d\zeta|n^t) + \beta\lim_{T\to\infty}\int\max[\Phi,V_{t+1}^T(\zeta,n^t)P_{\eta_{t+1}}(d\zeta|n^t)$$

$$= \int\pi(\zeta)P_{\eta_{t+1}}(d\zeta|n_1^t) + \beta\int\max[\Phi,V_{t+1}(\zeta,n^t)]P_{\eta_{t+1}}(d\zeta|n^t),$$

because
$$\lim_{T\to\infty}\int\max[\phi,V_{t+1}^{T+1}(\zeta,n^t)]P_{\eta_{t+1}}(d\zeta|n^t)$$

$$= \int\lim_{T\to\infty}\max[\phi,V_{t+1}^{T+1}(\zeta,n^t)]P_{\eta_{t+1}}(d\zeta|n^t)$$

by the Lebesque dominated convergence theorem, since $\{\max[\phi,V^{T+1}(\zeta,n^t)\}$ is dominated by $\max(\Phi,(1-\beta)^{-1}\bar{\pi})$ which is integrable with respect to $P_{\eta_{t+1}}(\cdot|n^t)$.

We have shown that if $V_t(\cdot)$, and the associated stopping policy, were optimal, then they would satisfy the conditions of the theorem. What remains is to show that they are indeed optional. To see this assume, to the contrary, that there exists an alternative stopping policy, say $\{\chi_\tau^*(\cdot)\}_{\tau=0}^{\infty}$, where $\chi_\tau^*(n^\tau)$ is one if a firm with η realizations of n^τ is in operation in period τ and zero otherwise, which generates a value function, say $V_t^*(n^t)$, which satisfies, for at least one (t,n^t),

(A4) $$V_t^*(n^t) - V_t(n^t) \geq \epsilon > 0.$$

Note that for any arbitrary T,

$$V_t^*(n^t) \leq E_t\{\sum_{\tau=0}^{T} \beta^\tau [\chi^*_{t+\tau}(\eta^{t+\tau})\pi(\eta_{t+\tau}) + \{\chi^*_{t+\tau}(\eta^{t+\tau}) - \chi^*_{t+\tau-1}(\eta^{t+\tau-1})\}\Phi] + \beta^T\bar{\pi}/(1-\beta)$$

(A5)

$$\equiv V_t^{*T}(n^t) + \beta^T\bar{\pi}/(1-\beta) \leq V_t^T(n^t) + \beta^T\bar{\pi}(1-\beta)^{-1} \leq V_t(n^t) + \beta^T\bar{\pi}(1-\beta),$$

where $V_t^{*T}(\cdot)$ is the value function that arises when the policy $\{\chi_\tau^*(\cdot)\}$ is followed for a T-horizon problem. The first inequality follows from the fact that current returns are bounded by $\bar{\pi}$, the second from the fact that $V_t^T(n^t)$ is the optimum for the T horizon problem, and the third is from A3. Provided T is chosen to be greater than $-\ell n[\epsilon(1-\beta)/\bar{\pi}]$ / $-\ell n\beta$, equations A4 and A5 contradict one other. [ ]

Appendix 2.

This appendix proves the following Lemma.

3.6 Lemma.

Let $\{S_t^a\}$ be the stochastic process generated from the distribution of size conditional on survival and any $\omega_0 \epsilon \{1,2,\ldots,K\}$. The sample space for this process consists of all possible sequences of elements from the finite set $\underline{S} = \{S: S(\eta),\ \eta \epsilon N\}$. Its probability measure is obtained from the family $\}P_\eta = \{P_\eta(\cdot|\omega),\ \omega \epsilon [1,2,\ldots,K]\}$ and the Markov transition matrix for ω_{t+1} conditional on ω_t and survival until t+1, say $Q=[q_{ij}]$. Q is derived from $P = [p_{i,j}]$ by dividing its i^{th} row by $1-p_{i,0}$ (for $i=1,\ldots,K,$) and then deleting its first row and column. Let M_x^y denote the σ-algebra generated by $S_x^a,\ldots,S_y^a$. Then

$$\sup\{|P(E_2|E_1) - P(E_2)|,\ E_1 \epsilon M_1^x,\ E_2 \epsilon M_{x+\tau}^\infty\} \leq A\phi^\tau$$

with $\phi < 1$.

Proof

Let $S_x^y = (S_x^a,\ldots,S_y^a)$, and s^r be the generic element in $\underline{S}^r$, $r = 1,2,\ldots$. Then it suffices to show that for any $E_1 \epsilon M_1^x$ and any $E_2 \epsilon M_{x+\tau}^{x+\tau+r}$

$$(A2.1) \quad |P\{(S_{x+\tau}^{x+\tau+r} \epsilon E_2) n (S_1^x \epsilon E_1)\} - P\{S_{x+\tau}^{x+\tau+r} \epsilon E_2\} P\{S_1^x \epsilon E_1\}| \leq A\phi^\tau P\{S_1^x \epsilon E_1\}$$

(Billingsley, 1967, section 20). But the left hand side of (A2.1) is

$$\leq \sum |P\{(S_{x+\tau}^{x+\tau+r} = s^r) n (S_1^x = s^x)\} - P\{S_{x+\tau}^{x+\tau+r} = s^r\} P\{S_1^x = s^x\}|$$

when the summation extends on $s^r \epsilon E_2$ and $s^x \epsilon E_1$. Thus it suffices to show for all E_1 such that $P(E_1) > 0$

$$(A2.2)\qquad \sum \left| P\{S_{x+\tau}^{x+\tau+r}=s^r \mid S_1^x=s^x\} - P\{S_{x+\tau}^{x+\tau+r}=s^r\} \right| \leq A\phi^\tau$$

Let $S_t^a=\bar{S}_t+U_t$ where $\bar{S}_t=\int S(\eta_t)P(d\eta \mid \omega_t)$ and $U_t=S_t^a-\bar{S}_t$, and note that both $\bar{S}_t$ and U_t can take on only a finite set of values, say values in the sets $\bar{S}$ and U, respectively. Finally let

$$\bar{S}_x^y = \{\bar{S}_x,\ldots,\bar{S}_y\}, \quad U_x^y = \{U_x,\ldots,U_y\}$$

and $\bar{s}^x$ and u^x be the generic elements in $\bar{S}^x$ and U^x. Then each element in the sum in (A2.2) can be written as

$$(A2.3)\qquad \sum_{u^r} P\{U_{x+\tau}^{x+\tau+r}=u^r \mid \bar{S}_{x+\tau}^{x+\tau+r}=\bar{s}^r-u^r\}P\{\bar{S}_{x+\tau}^{x+\tau+r}=\bar{s}^r-u^r \mid S_1^x=s^x\} - P\{S_{x+\tau}^{x+\tau+r}=s^r\}$$

$$= \sum_{u^r} \Big[\prod_{i=x+\tau}^{x+\tau+r} P(U_i=u_i^r \mid \bar{S}_i=\bar{s}_i^r-u_i^r)\Big] \left[P\{\bar{S}_{x+\tau}^{x+\tau+r}=\bar{s}^r-u^r \mid S_1^x=s^x\} - P\{\bar{S}_{x=\tau}^{x+\tau+r}=\bar{s}^r-u^r\}\right],$$

where the equality follows from the fact that the distribution of U_t conditional on all information prior to t, depends only on $\bar{S}_t$. Consider each of the expressions in the latter square brackets separately. Letting $(i_1,\ldots,i_r)$ be the unique values of ω that lead to $\bar{S}_{x+\tau}^{x+\tau+r}=\bar{s}^r-u^r$ those expressions can be written as

$$(A2.4)\qquad \sum_{\bar{s}^x-u^x} P(\omega_{x+\tau+r}=i_r,\ldots,\omega_{x+\tau}=i_1 \mid \bar{S}_1^x=\bar{s}^x-u^x)P(\bar{S}_1^x=\bar{s}^x-u^x \mid S_1^x=s^x) - P\{\bar{S}_{x+\tau}^{x+\tau+r}=\bar{s}^r-u^r\}$$

Assumption (1.ii) guarantees that Q, the Markov transition matrix for the survivor process, is irreducible aperiodic. It therefore has a unique invariant distribution say q*. Moreover, for any i,j the n period transition probability, q_{ij}^n satisfies

$$| q_{ij}^{n} - q_{i}^{*} | \leq A\phi^{n},$$

for some $\phi < 1$. (On these latter points see Billingsley, 1979, section 1.8). Consequently if $(j_1,\ldots,j_x)$ indexes the values of ω that lead to $\bar{S}_1^x$, the absolute value of (A2.4) is less than or equal to

$$q_{i_r i_{r-1}} \cdots q_{i_2,i_1} \Big| \sum_{j_x} q_{j_x,i_1}^{\tau} P(j_x \mid S_1^x = s^x) - q_{i_1}^{*} \Big|$$

(A2.5)

$$\leq q_{i_r i_{r-1}},\ldots q_{i_2 i_1} A\phi^{\tau}.$$

To complete the proof of the proposition, substitute (A2.5) into (A2.4), the result into (A2.3), and the result of that into (A2.2). [ ]