2-1-1990

Empirical Implications of Alternative Models of Firm Dynamics

Ariel Pakes

Richard Ericson

ECONOMIC GROWTH CENTER

YALE UNIVERSITY

Box 1987, Yale Station
New Haven, Connecticut 06520

CENTER DISCUSSION PAPER NO. 594

EMPIRICAL IMPLICATIONS OF ALTERNATIVE MODELS OF FIRM DYNAMICS

Ariel Pakes
Yale University

Richard Ericson
Columbia University

February 1990

Abstract

This paper considers two models for analyzing the dynamics of firm behavior that allow for heterogeneity among firms, idiosyncratic (or firm specific) sources of uncertainty, and discrete events (exit and entry). Models with these characteristics are needed for empirical analysis of the causes and effects of the dispersion in the distribution of outcome paths among firms, and for correcting for the self–selection induced by liquidation in the empirical analysis of firms responses to alternative policy and environmental changes. It is shown that the two models have different nonparametric implications, and that these are rich enough to enable the construction of both testing, and selection correction, procedures that are both, fully consistent with the theoretical model, and easy to implement. 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 on manufacturing, and the other to be consistent with the data for retail trade.

I.   Introduction

The paper considers the empirical implications of two models for the dynamics of firm behavior that allow for heterogeneity among firms, idiosyncratic (or firm–specific) sources of uncertainty, and discrete events (exit and/or entry). Our reasons for investigating the empirical implications of models with these features are twofold. First many phenomena of interest are intricately tied up with the nature of firm specific differences in outcome paths, and a detailed analysis of these phenomena requires a choice among models that generate, or at least allow for, such idiosyncratic differences. Examples include the analysis of; default probabilities, of the extent of job turnover generated by the growth and contraction of individual firms within larger aggregates, and of changes in market structure (or in the size distribution of firms in an industry). The second reason for studying such models is that without a framework for empirical analysis that allows for the (appropriate) source of firm specific differences in outcome paths, we will often be unable to obtain a meaningful picture of firms' responses to any policy or environmental change. Table 1 illustrates one reason 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 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,

1

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

Sector	1 #Firms Active in 1978	2 % of all Firms in 1978	3 Employment in 1978	4 % of all Employment 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	55	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 column 5, 6, and 8.					8.5	11.1		10.5
Substitute "either transferred out or liquidated" for "liquidated" in column 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.

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).[1] 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.

To either control for the selection problems induced by the liquidation process, or to analyze any of the phenomena that involve the determinants (or the effects) of the characteristics of the distribution of sample outcome paths, we need a model that explains

---

[1]Changes of ownership, however, become the major source of attrition in data sets constructed from information on (the typically much larger) publically traded firms, see Hall (1988).

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 to the researcher to determine whether either of them might be relevant for the problem at hand.[2]

The first model considered here is a model with passive 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 model is a model of active exploration. It assumes that 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, 1989).

_____

[2]For a review of the economics literature on models of firm behavior that allow for idiosyncratic uncertainty, see Ericson and Pakes (1989). There is also a large related econometric and statistical literature on the analysis of selection bias in models of individual and household (in contrast to firm) behavior; for details and references see the articles in Heckman and Singer (1985); in particular Heckman and Robb (1985). Our analytic approach to both the selection, and to the model choice, problem is more similar to that of the recent econometric literature on analyzing stochastic control models with discrete outcomes (see below), in that all our assumptions will be placed on the primitives of the environment underlying the agent's optimal control problem (and not on the form of the function determining current choices), and all implications will be derived from the characteristics of the solution to that agent's control problem (the underlying models used here also ensure the consistency of the interactions among firms in the sense of proving the existence of a dynamic stochastic equilbrium for the industry). We differ from the control literature in that our assumptions do not restrict the primitives of the analysis to particular parametric families of functional forms. In this respect our analysis is most similar of that of Frydman and Singer (1985), which, though it does not begin with an agents control problem, does derive tests for the qualitative characteristics of various dynamic processes.

Both models are traditional in that a firm is treated as a single decision maker faced with a set of profit opportunities who acts 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 assumption. 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 interpretable regularity conditions on the relevant functional forms. Precisely because these 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 being less 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 more limited. They are meant only to provide an interpretable characterization of the data which: 1) aids in distinguishing which, if either, of the alternative models seems relevant for the problem at hand, and 2) acts as a basis for building a procedure for correcting for the selection problem induced by the liquidation process when one of the models seems appropriate.

One of the 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, 1981; 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 model with active exploration 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, and 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).

The data analyzed are obtained from the unemployment insurance (UI) account records of firms. UI account data have the advantages that they are available for all firms (regardless of their size and of whether or not the firm is in manufacturing), and that they allow the researcher to both distinguish between attrition due to liquidation and attrition due to changes in ownership and to assign a birth date to the firm (see section 5). These characteristics of the UI account data make it particularly useful for the analysis of the initial stages of the development of firms in different sectors of the economy. 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.

Section 2 of the paper outlines the passive learning model and then derives its nonparametric implications. Section 3 does the same for the model with active exploration. 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. Section 6 summarizes and considers further implications of the empirical results.

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 a particular value, so $x^t = (x_1,...,x_t)$. 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, and z for a member of that set. Lemmas, theorem, examples, etc. will be numbered in one consecutive ordering with each section. They are referred to in the following sections with a section prescript.

Section 2. Innate Ability Differences of Unknown Value.

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 competitive industry of firms who learn about their ability parameters in a Bayesian fashion. There are a large number of ex ante identical potential entrants each of whom believes the value of its parameter, say θ, is a random draw from some known distribution. Entrants pay a sunk cost of entry. Each period the firm is in operation thereafter 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. In this model the firm is a value of θ, and the role of management is to make inferences on the likelihood of alternative possible values of θ and then determine optimal behavior conditional on those inferences and the state of the industry. One possible 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 competitive industry composed of firms which operate in this manner. We focus on the implications of the nature 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 only on the learning process and interpretable regularity conditions on the functional forms 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 sequence of payoff relevant random variables, say $\{\eta_t\}$. 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 $\pi_t = \alpha_{1t} \eta_t F(\ell_t) - \alpha_{2t}\ell_t$ where ; ℓ_t is a vector of input quantities, $F(\cdot)$ is a concave production function, $\alpha_t = (\alpha_{1t}, \alpha_{2t})$ is a price vector, and $\{\eta_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables, with distribution, conditional on θ, given by $P_\eta(\cdot \mid \theta)$. Assume η_t is known at the time ℓ_t is chosen. Then

$$\pi_t = \pi(\eta_t; \alpha_t) = \max_{\ell_t} \{\alpha_{1t}\eta_t F(\ell_t) - \alpha_{2t}\ell_t\},$$

where $\pi(\eta; \alpha_t)$ 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 form $G_0(\theta)$. 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 payoff relevant random variables (a stochastic process), a family of distribution functions for those sequences (one for each value of θ), a prior distribution for θ, and a function giving the payoffs for different realizations of η. The example assumed that conditional on a particular value of θ the sequence, $\{\eta_t\}$, was i.i.d. over time. Though the i.i.d. assumption simplifies the

analysis, it produces a host of very strong empirical implications that are entirely a result of the i.i.d assumption and not of the logic of the passive learning model per se. Our assumptions will, therefore, allow for dependence in the stochastic process generating $\{\eta_t\}$ conditional on θ.

This necessitates defining an ordering over stochastic processes; that is we need an interpretation for the statement that one value of θ generates a better stochastic process than another. The ordering we define below is designed to insure that the family of Bayesian posteriors for both θ, and for future realizations of η, generated by alternative past histories of η, say of $n^t = (n_1...,n_t)$, will be partially ordered (in the stochastic dominance sense) by the natural partial order of n^t. That is, if $n_1^t \geq n_2^t$ (component by component), then the posteriors for both θ, and for future values of η, generated from the n_1^t history will stochastically dominate those generated by the n_2^t history. To be more precise we need the following definitions.

1. <u>Definitions</u> (likelihood ratio ordering, or $\succsim_{\ell r}$, and stochastic dominance, or $\succsim_s$)

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 strictly likelihood ratio dominates P_2 and write $P_1 \succsim_{\ell r} P_2$, if and only if,

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

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 weakly likelihood ratio dominates P_2, and write $P_1 \succsim_{\ell rw} P_2$. Similarly, say P_1 weakly stochastically dominates P_2, and write $P_1 \succsim_{sw} P_2$, if and only if for every increasing function, $h(\cdot)$, such that $\int h(\zeta)P_1(d\zeta) < \infty$,

$$\int h(\zeta) P_1(d\zeta) \geq \int h(\zeta) P_2(d\zeta)$$

If $P_1 \succsim_{sw} P_2$, but $P_2 \not\succsim_{sw} P_1$ we say that P_1 strictly stochastically dominates P_2, and write $P_1 \succsim_s P_2$. [ ]

Note that if $P_1 \succsim_{lr} P_2$ 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. This insures that the Bayesian posterior for θ generated from a higher z value will lr dominate the posterior generated from a lower z value (no matter the prior for θ, see, for eg. Milgrom, 1983).[3]   Also $P_1 \geq_{lr} P_2$ implies $P_1 \geq_s P_2$ when the underlying state space is totally ordered (as is the case when univariate random variables are being compared), however neither ordering need imply the other when the state space is only partially ordered (as is the case when the relevant random variables are vector valued; on these points see Whitt, 1980 and 1982).

Assumption 2 provides the primitives of the passive learning model and endows them with some regularity conditions. (2.i) formalizes our stochastic assumptions. Higher values of θ are assumed to provide better, in the lr sense, distributions for the vector η^t. Moreover conditional on a $\theta \in \theta$, the natural partial order of the histories of η, i.e. of n^{t-1}, is assumed to partially lr order the distributions of future η (histories which are larger in every past year are at least as likely to generate larger future values of η).[4] (2.ii) provides the prior distribution

---

[3]The following counterexample shows that the posteriors for θ need not be ordered in η realizations if we were to assume only a first order stochastic dominance ordering of $P_\eta(\cdot \,|\, \theta)$ in θ. Let $\theta = (\theta_1, \theta_2)$ with $\theta_2 > \theta_1$ and consider the following family of densities (with respect to the counting measure): $p(\eta = 2 \,|\, \theta_2) = p(\eta = 4 \,|\, \theta_2) = 1/2$, and $p(\eta = 1 \,|\, \theta_1) = p(\eta = 3 \,|\, \theta_1) = 1/2$. Clearly $P_\eta(\cdot \,|\, \theta_2) \geq_s P_\eta(\cdot \,|\, \theta_1)$. However, if $\eta = 2$ the posterior is $\theta = \theta_2$ with probability one, whereas if $\eta = 3$, the posterior is $\theta = \theta_1$ with probability one; i.e., the posterior for $\eta = 2$ dominates the posterior for $\eta = 3$.

[4]We have shown that a sufficient condition for (2.ib) is that conditional on any $\theta \in \theta$, all

for θ, while (2.iii) provides the profit and size functions. It is important that both be increasing in η for all possible realizations of the price vector. [5]

2. <u>Assumption</u> (primitives of the model)

(i) $\{\eta_t\}$ is a sequence of payoff relevant random variables (a stochastic process) whose distribution, say $\underline{P}(\theta)$, is indexed by a $\theta \in \Theta$, where Θ is a compact subset of $\mathbb{R}_+$. The family

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

have finite dimensional distributions, $P_\eta{}^t(\cdot \mid \theta)$, which have support in $\mathbb{N}^t$ (a compact subset of $\mathbb{R}^t$), and densities with respect to some dominating measure. These densities obey a likelihood ration ordering in θ, i.e.

$$P_\eta t(\cdot \mid \theta_1) \underset{\ell r}{\gtrsim} P_\eta t(\cdot \mid \theta_2) \tag{a}$$

---

finite–dimensional distributions for η^t are totally positive of order two as defined by Karlin and Rinott (1980), or Whitt (1982). The weaker condition given in (2.1b) becomes equivalent to total positivity if the stochastic process generating $\{\eta_t\}$ conditional on a $\theta \in \Theta$ is a first order Markov process.

[5]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 come back to some of the alternatives below, but for now suffice it to note that the results we focus attention on do not depend on the timing of the input decision.

whenever $\theta_1 \geq \theta_2$; and have conditional distributions for future values which satisfy a weak ℓr ordering in the (componentwise) partial ordering of past realizations, of n^{t-1}, i.e. for any θ

$$P_{\eta_t}(\cdot \mid n_1^{t-1}, \theta) \underset{\sim}{\geq} \ell r w \ P_{\eta_t}(\cdot \mid n_2^{t-1}, \theta) \qquad \text{(b)}$$

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

(ii) $G_0(\cdot)$ is a (nondegenerate) prior probability distribution, with density $g_0(\cdot)$ on θ.

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

To complete the specification of the model, we need assumptions on the evolution of prices, and on the behavior of management. Jovanovic (1982) assumes constant input prices, and 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 ($\mathbb{P}$), past realizations of η, say $n^t = (n_1,...,n_t)$, and the prior for θ [$G_0(\theta)$]. The implied posteriors for θ are provided in assumption 3 below. Section 5 notes the modification to our empirical implications that would be required if different prices prevailed in different periods.

3. Assumption [posterior distributions]

Let J_t contain all information available to management in period t. Then management's beliefs about its own value of θ are summarized by the (Bayes) posterior distribution

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

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

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.[6]

Assume, temporarily, the existence of a bounded function, say $V_{t+1}(n^{t+1}; \underline{\alpha})$, from $\mathbb{N}^{t+1}$ into $\mathbb{R}$, which, for a given equilibrium price sequence, $\underline{\alpha}$, 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

$$(4) \quad V_t(n^t; \underline{\alpha}) = E[\pi(\eta_{t+1}; \underline{\alpha}) \mid n^t] + \beta E[\max\{\Phi, V_{t+1}(\eta^{t+1}; \underline{\alpha})\} \mid n^t],$$

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

Given (4) the optimal liquidation policy is straightforward. Operate the firm if and only if $V_t(n^t; \underline{\alpha}) \geq \Phi$. Theorem 5 insures that the value function in (4) exists and then provides some of its properties.

---

[6]The assumptions that Φ is the same know value for all agents, and is constant over time, are made for expositional convenience. What is required is that Φ not increase as rapidly with n^t as does the value of continuing in operation at t. Of course, the actual behavior of "exit values" is an empirical question. If the process generating the exit we are modeling is indeed a liquidation process, and not a process generated by changes of ownership and continued operation of the firm in a different guise (see the discussion of the data in section 5) this assumption ought not to be problematic.

5. Theorem (existence and monotonicity of the value function)

At each t there exists a unique $V_t(\cdot): \mathbb{N}^t \to \mathbb{R}_+$ which provides the value of continuing in operation assuming optimal behavior in each future period. $V_t(\cdot)$ is bounded, satisfies (4), and is nondecreasing in n^t; i.e., if $n_1^t \geq n_2^t$, then $V_t(n_1^t; \underline{\alpha}) \geq V_t(n_2^t; \underline{\alpha})$ [for $n^t \in N^t$, and all t].

Proof See Pakes and Ericson, 1987 Appendix I.   [ ]

Note that Theorem 5 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 (2) (in particular the posteriors for θ need not posses simple sufficient statistics). We now move on to consider the empirical implications of the passive learning model given only the assumptions reviewed in this section.

2.2 Empirical Implications of Passive Learning.[7]

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 increases the θ–distribution of the surviving firms improves (in the stochastic dominance sense). This is a result of self (in contrast to natural) selection. As time passes firms with lower θ's are more likely to draw lower η's and liquidate.

---

[7]We have focused the analysis on empirical implications that we thought could be investigated with currently available panel data sets. These are characteristically quite short in the time dimension, and this fact persuaded us to stay away from implications that stemmed from the familiar convergence properties of Bayesian learning models for finite dimensional parameter vectors. That is, in this model survivors will, in the limit, know there true value of θ exactly.

16

6. <u>Theorem</u> (the evolution of the θ-distribution)

Let $A^t = \{n^t = (n_1, ..., n_t): V_1(n_1; \underline{\alpha}) > \pmb{\$}, ..., V_t(n^t; \underline{\alpha}) > \pmb{\$}\}$, and

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

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

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

Then

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

<u>Proof</u> Take an arbitrary (z,t). Then, by Bayes law,

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

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

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

(6.1)
$$\frac{_\theta \int Pr\{\chi_t = 1 | \theta\} G_0(d\theta)}{_\theta \int Pr\{\chi_{t-1} = 1 | \theta\} G_0(d\theta)} \ge \frac{_{\theta \le z} \int Pr\{\chi_t = 1 | \theta\} G_0(d\theta)}{_{\theta \le z} \int Pr\{\chi_{t-1} = 1 | \theta\} G_0(d\theta)}.$$

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 defining

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

(6.2)

$$Q_2(d\theta) = \begin{bmatrix} 0 & \text{for } \theta > z \\ Pr\{\chi_{t-1}=1\,|\,\theta\}G_0(d\theta)/\int_{\theta \leq z}Pr\{\chi_{t-1}=1\,|\,\theta\}G_0(d\theta), & \text{otherwise} \end{bmatrix},$$

(6.1) can be rewritten as

(6.3) $$\int_\theta Pr\{\chi_t = 1\,|\,\chi_{t-1}=1,\theta\}Q_1(d\theta) \geq \int_\theta Pr\{\chi_{t-1}=1\,|\,\theta\}Q_2(d\theta).$$

Since (6.2) implies $Q_1(\cdot) \gtrsim_{sw} Q_2(\cdot)$, (6.3) will be true provided $Pr\{\chi_t=1\,|\,\chi_{t-1}=1,\theta\}$ is nondecreasing in θ. To see that this is indeed the case write

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

Then, taking $\theta \geq \theta'$

$$\int Pr\{\chi_t=1\,|\,n^{t-1},\theta\}P_\eta^{t-1}\{dn^{t-1}\,|\,n^{t-1}\epsilon A^{t-1},\theta\} \geq$$

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

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

where the first inequality follows from the monotonicity of $V(\cdot)$ and the fact that for any n^{t-1}, (2.i) insures that $P_\eta^t(\cdot\,|\,n^{t-1},\theta)$ is stochastically increasing in θ, and the second from the fact that if $P_{\eta_t}(\cdot\,|\,\theta) \gtrsim_{\ell r} P_\eta^t(\cdot\,|\,\theta')$, then, for any $A\epsilon\mathbb{N}^t, P_{\eta_t}(\cdot\,|\,\eta^t\epsilon A,\theta) \gtrsim_{\ell r} P_\eta^t(\cdot\,|\,\eta^t\epsilon A,\theta')$ and recursive application of (2.i). $[\,]$

Since θ is not observable we look for the implications of Theorem 6 on the observable

size variables. The easiest case to consider is when price is constant over time, and the marginal distribution of η conditional on θ is either constant, or stochastically increasing, over age. Then the twin facts that the distribution of η is stochastically increasing in θ, and that the θ distribution of the survivors is stochastically increasing over time, imply that the size distribution of the surviving firms is stochastically increasing over time.

7. <u>Corollary</u> (the evolution of the size distribution)

Assume that $\alpha(t) = \alpha$ for all t, and that the marginal distribution of η conditional on θ is either constant, or stochastically increasing, over age. Let χ_t be defined as in Theorem 6 and for all z and t define

$$P_S(z|t) = Pr\{S_t \le z | \chi_t = 1\}.$$

Then, provided $t \ge t'$

$$P_S(\cdot|t) \gtrsim_{sw} P_S(\cdot|t'). \qquad\qquad []$$

As explained in the empirical section we can, with a little more attention to details, allow prices to vary over time and derive a modified version of Corollary 7 with anlogous testable implications provided $S(\eta_t;\underline{\alpha}) = f_1(\alpha_t) f_2(\eta_t)$ (as would be the case, for example, if η represented neutral efficiency differences among firms, the production function was homothetic in its inputs, and size were measured either by output produced or by employees hired).

The refutable implication of Corollary 7 (or its modification) that we will take to data is that it implies that the proportion of the sample of surviving firms with S(t) greater than any given number should be increasing in age (more generally the mean of the survivor distribution of any increasing function of size should be increasing over age). Note also that

Theorem 6 and Corollary 7 imply that each sequence of distribution functions, $\{P_\theta(\cdot\,|t)\}$, and $\{P_S(\cdot\,|t)\}$, converges (pointwise) to a well defined limiting distribution.

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 7, 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

$\Pr\{\chi_t=0\,|\,\chi_{t-1}=1, S_{t-1}=s_{t-1}\}$ is nonincreasing in s_{t-1};

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, 1988 and 1989).

Appendix 1 provides an example which shows that of these three only the first survives our search for nonparametric implication 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, both persists and 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, 1988 and 1989]. However, most other models that allow for mortality, including Ericson and Pakes' (1989) model of active exploration, 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, though the fact that the passive learning model does not necessarily imply that either hazard rates, or the variance in growth rates, decline in age is somewhat disconcerting, the intuition underlying our counter example is clear enough. For

many functional forms it will 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). Differences in the variance in growth rates over age, 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. Thus 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, for any n' and z

$$P_{\eta_t}(z|\eta_k = n') = {}_{\theta}\!\int P_{\eta}(z|\theta)P_{\theta}(d\theta|\eta_k = n')$$
$$= {}_{\theta}\!\int P_{\eta}(z|\theta)P_{\eta}(n'|\theta)g_0(\theta)d\theta / {}_{\theta}\!\int P_{\eta}(n|\theta)g_0(\theta)d\theta;$$

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 (2). In the general case we have, for any $z \in \mathbb{N}$,

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

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 latter distribution has two sources, one of which will __not__ erode away as t grows large. That is, 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 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 $S_{t-1}, \ldots, S_{t-k}$, and 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 model of active exploration considered in the next section).

Theorem 8 defines a notion of dependence for the posterior for θ which leads directly to our test.

8. __Theorem__ (posteriors for θ)

Fix any $(t,k) \in \mathbb{Z}_+^2$ with $t > k+1$, and any set of realizations $(S_t = s_t, \ldots, S_{t-k} = s_{t-k})$. Then, letting χ_t be defined as in Theorem 6, the family of posteriors

$$\left\{ P_{\theta}(\cdot \mid s_t, \ldots, s_{t-k}, S_1 = s_1, \chi_t = 1) \right\},$$

is ℓr ordered in s_1.

Proof. A simple proof can be obtained by induction on k. To this end assume that for all possible $(s_t,...,s_{t-k},s_1,s_1',\theta,\theta')$ with $s_1 > s_1'$ and $\theta > \theta'$

$$p(\theta|s_t,...,s_{t-k},s_1,\chi_t = 1)\, p(\theta'|s_t,...,s_{t-k},s_1',\chi_t = 1)$$

(8.1)
$$-p(\theta|s_t,...,s_{t-k},s_1',\chi_t = 1)\, p(\theta'|s_t,...,s_{t-k},s_1,\,\chi_t = 1) > 0$$

By definition

(8.2)
$$p(\theta|s_t,...,s_{t-k+1},s_1,\chi_t = 1) = \int p(\theta|s_t,...,s_{t-k},s_1,\chi_t = 1)\, p(s_{t-k}|s_t,...,s_{t-k+1},s_1,\chi_t = 1)ds_{t-k}$$

where, for notational simplicity, we have assumed that the density of S is with respect to Lebesque measure. Consequently

$$p(\theta|s_t,...,s_{t-k+1},s_1,\,\chi_t = 1)\, p(\theta'|s_t,...,s_{t-k+1},s_1',\chi_t = 1)$$

(8.3)
$$-p(\theta'|s_t,...,s_{t-k+1},s_1,\chi_t = 1)\, p(\theta|s_t,...,s_{t-k+1},s_1',\chi_t = 1)$$
$$= \int I(s_t,...,s_{t-k},s_1,s_1',\theta,\theta')ds_{t-k}$$

with $I(\cdot)$ derived directly from (8.2). Inspecting $I(\cdot)$ we find that its sign is equal to the sign of (8.1) for every s_{t-k}. Since the latter is positive by assumption, the integral in (8.3) is positive, which completes the inductive step of the argument. For the initial condition of the inductive argument we require

$$p(\theta|s_t,...,s_2,s_1,\chi_t = 1)\, p(\theta'|s_t,...,s_2,s_1',\chi_t = 1)$$

(8.4)
$$-p(\theta|s_t,...,s_2,s_1',\chi_t = 1)\, p(\theta'|s_t,...,s_2,s_1,\chi_t = 1) > 0,$$

for all possible $s_t,...s_2,s_1',s_1,\theta,\theta'$ with $s_1 > s_1'$ and $\theta > \theta'$ [by possible we mean $\chi_t(s_t,...,s_2,s_1) = 1$, which insures that $\chi_t(s_t,...,s_2,s_1) = 1$, and $\theta,\theta' \in \theta^2$]. Since $\chi_t(\cdot)$ is measureable with

respect to the σ–algebra generated from s^t, its presence in (8.4) is redundant, and the required inequality follows directly from (2.i) and the definitions in (3) and (1).   [ ]

Corollary (9) is a direct implication of (8), (2.i), and the fact that for univariate distributions likelihood ratio dominance implies stochastic dominance.

9. Corollary

Assume that conditional on θ and past history, the distribution of η_t is Markov of order less than or equal to k. Then

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

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

Though we expect the assumption that conditional on θ and past history the distribution of η_t is Markov of order at most k to be a reasonable approximation, without it the expectation of S_t conditional on a particular realization $(s_{t-1}, \dots s_{t-k})$ and $S_1 = s_1$ will, in general, only depend on s_1 and need not be increasing in that variable.[8] As a result the empirical section will provide test results for both monotonity of the regression function in s_1, and for dependence of the regression function on s_1.

This dependence occurs because the parameter which determines the conditional

---

[8]This is one of several subtilities that result from the fact that our observations are censored by the survival process. Firms with a higher s_1 will survive on the basis of lower subsequent realizations. As a result, by conditioning also on survival, we are conditioning on "better" subsets of the state space for years near the first when we start with lower s_1 values. This possible positive dependence between s_t and the sets of past values of S that we are not conditioning on, conditional on any particular value of θ, can reverse the ordering of the expectation of S_t conditional on alternative realizations of S_1 given in (9).

distribution of size 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, the distribution of S_t conditional on a realization $(s_{t-1},...,s_{t-k})$ and $S_1 = s_1$ need not depend on s_1. We turn to such a model now.

Section 3. Active Exploration

This section considers the empirical implications of a model, orginally developed by Ericson and Pakes (revised 1989), in which firms can invest to improve the value of a parameter, say ω, which determines the distribution of its profits. In this model firms are assumed to know the current value of their ω and make all current decisions based on it. On the other hand ω itself evolves over times in response to the outcomes of the firm's own investments, and to the investment's of other firms operating in related markets. These outcomes are stochastic; in the model with active exploration the firm invests to explore and develop alternative market niches which may, or may not, prove profitable.

Ericson and Pakes begin with a special case in which the firm's profits depends only on the difference between the firm's own level of development and an aggregate index of the state of the market. Startup is treated as the appearance of an idea at some location on the ω axis. If the idea requires substantial successful development before it can generate profits, the initial ω is associated with a distribution of profits which is degenerate, or nearly so, at zero. In each period in which the firm decides to remain active it determines a quanitity of investment to maximize the expected discounted value of future net cash flows. Successful investment enables the idea to be embodied in a more profitable marketable good or service. Unsuccessful may well convince the entrepreneur that the whole idea is not worth pursuing and lead to liquidation.

Their general model allows the firm's profits to also depend explicitly on the levels of development of all other firms active in the market, endogenizes entry, and then solves for a

Markov perfect Nash equilibrium in investment strategies. Here we suffice with a brief description of the special case, as it makes it particularly easy to illustrate the empirical implications we are after (though, as will become intuitively clear, those implications do not depend on the simplifying assumptions used in the special case). Just as in our discussion of passive learning, we begin with a brief description of the model, then move on to a listing of the assumptions we make on the primitives of that model, and conclude with the empirical implications of those assumptions.

The Active Exploration Model

We will assume that the state space is countable and index it by the integers so that $\omega \in \mathbb{Z}$. Each firm operating in period t is endowed with an w_t. Higher values of w are better in the sense that the distribution of the payoff relevant η is stochastically increasing in w. 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

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

where $c(\cdot) > 0$, and can be decreasing in w 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 w_{t+1}, and hence higher future profits, more likely. In particular, let $\tau_{t+1} = w_{t+1} - w_t$, and J_t be the information available to management at t. Then we assume that for $x \in \mathbb{Z}$,

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

where $P_\tau(\cdot | x_t)$ is stochastically increasing in x. Hence, to formalize the firm's decision problem we will require the following primitives.[9]

1. <u>Assumption</u> (primitives of the active exploration model)

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

$$P_\eta(\cdot | w) \gtrsim_{sw} P_\eta(\cdot | w')$$

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

   ii) $\mathbb{P}_\tau = \{P_\tau(\cdot | x): x \in \mathbb{R}_+\}$ is a family of distributions with support T, a compact subset of $\mathbb{Z}$, 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

---

[9]We omit aspects of the model which do not impact on the empirical implications we are after (this includes allowing for price, and other time–specific firm–invariant processes). Also, just as we did in the last section, we assume here, for expositional simplicity, that input choices are made after the realization of η, and that liquidation values are a constant Φ. Finally we work with a countable state space and an assumption that current investment does not affect the distribution of profits until the following year, to minimize the need for technical detail (in particular measureability conditions are satisfied trivially in this framework, and are therefore ignored throughout).

cannot be improved without some investment. The family of densities $\{p_\tau(\cdot\,|x): x\in\mathbb{R}_+\}$, 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 $p_\tau(0|x)$ and $p_\tau(-1|x)$ are strictly positive for all x less than any finite upper bound (it is always possible for the competition to advance as much or more than the firm in question does).

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

We now consider management's choice of policies. Letting w_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 (J) into operating (χ) and investment (x) decisions, that is

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

with $\chi_\tau = \chi_\tau(J_\tau)$ and $\chi_\tau = 0$ implying $\chi_{t+\tau} = 0$ for $t\in\mathbb{Z}_+$, $x_\tau = x_\tau(J_\tau)$, and $J_\tau = \{w_\tau, \chi_{\tau-1}, x_{\tau-1}, w_{\tau-1},..., w_0\}$. Recall that $R(\eta_\tau, w_\tau, x_\tau, \chi_\tau) = \pi(\eta_\tau)-c(w_\tau)x_\tau$ if $\chi_\tau=1$ and zero otherwise, so the expected discounted value of net cash flows given the policy d is

$$V_d(w_0) = E_d\left\{\Sigma\beta^T R(\eta_\tau, w_\tau, x_\tau, \chi_\tau) + \Phi(\chi_{\tau-1}-\chi_\tau)]\,|\,w_0\right\}$$

where $\beta\in(0,1)$ 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

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

for each w. A policy d^* will be optimal if $V_d^*(w) = V(w)$ for all w. If an optimal policy exists management chooses it, in which case the expected discounted value of future net cash flow is $V(w)$. Management will operate the firm if and only if $V(w) > \Phi$, the liquidation value. The following theorem combines the results from Ericson and Pakes (1989) 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.   <u>Theorem</u> (properties of the active exploration model).

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

i) $V(w)$ is bounded and nondecreasing in w.

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

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

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

$$x^*(w) = 0 \text{ if } w \notin \{w' : \underline{w} \leq w' \leq \overline{\overline{w}}\}.$$

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

$$V(w) > \Phi \text{ is and only if } w > \underline{w},$$

and

$$\inf_{t} \left[\inf_{w_0 \leq \overline{w}} \Pr\{ w_t \leq \overline{w} \mid w_0 \} \right] = 1.$$

<u>Proof</u>: See Ericson and Pakes, 1989   []

    Parts (i) and (ii) of this theorem ensure that both the value function and investment policy are stationary functions of w, the value function being increasing in w. Figure 1 illustrates this with one special case. In the figure 2, $\overline{\pi}(w) = \int \pi(\eta) P_\eta(d\eta \mid w)$, provides expected profits conditional on w. The value of w below which a firm exits, i.e. the $\underline{w}$ in (2.iv), is determined by the point at which $V(w)$ equals Φ. In this example $\underline{w} = \underline{w}$, that value of w below which a firm stops investment. So positive investment occurs at $\underline{w} + 1$, even though profits at that point are zero with probability one. The incentive for the investment is that it makes higher values of w_{t+1}, and hence higher future profits, more likely. The monetary value of an increase in w is $V(w+1) - V(w)$. Since $V(w)$ is bounded, after some point increases in w cannot bring with it much of a change in $V(\cdot)$. It follows that, after some w, it will not be in the firm's interest to invest at all. The w at which this occurs is the $\overline{\overline{w}}$ of (2.iii). If $w > \overline{\overline{w}}$, no investment takes place and this insures (see 1.ii) that the firm's w 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 in T. Then firms with $w_t < \overline{\overline{w}}$ have $w_{t+1} \leq \overline{\overline{w}} + \tau^* = \overline{w}$, and firms with $\overline{\overline{w}} < w_t \leq \overline{w}$ have $w_{t+1} \leq w_t$. So if $w_t \leq \overline{w}$, so must be w_{t+1}. This explains the second statement in 2.iv; that is, if $w_0 \leq \overline{w}$, then, with probability one, so will be the entire sequence $\{w_t\}_{t=0}^{\infty}$.
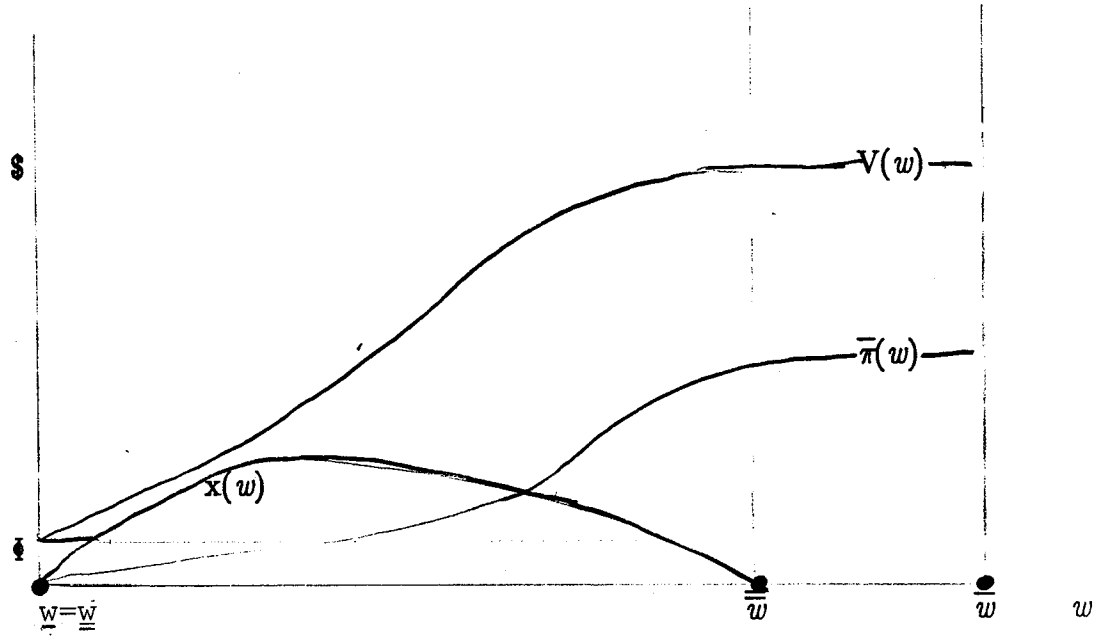
Figure 1: <u>Policies in the Active Exploration Model</u>

Since all values of $w \leq \underline{w}$ 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

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

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

With this understanding, theorem 2 implies that the sequence, $\{w_t\}$, together with any $w_0 \leq K$, is a finite state Markov chain on $\Omega = \{0,1,...,K\}$. Its 'zero' or 'death' state is absorbing, so the transition matrix for the chain is given by $\underline{P}$, where

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

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

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

Two remarks are in order here. First, recall that realizations of w are not observable. Realizations of $\{S_t\}$ are, but $S(\eta_t) = \overline{S}(w_t) + U(\eta_t)$, where $\overline{S}(w_t) = \int S(\eta_t)\, P(d\eta_t \,|\, w_t)$, and $U(\eta_t) = S(\eta_t) - \overline{S}(w_t)$. Since the distribution of $U(\eta_t)$ is also determined by w_t, and $\{w_t\}$ 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 there exists a finite n^*, such that for $n > n^*$

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

where $p^n_{i,j} = \Pr\{w_{t+n} = j \,|\, w_t = i\}$. Since $p_{0,0} = 1$, this implies that all states but zero are 'transient'. That is, no matter its initial w, 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 model with active exploration 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, so we now consider whether the model with active exploration has this implication.

Let $P_w(\cdot \,|\, t, p^0)$ and $P_S(\cdot \,|\, t, p^0)$ be the w and S distributions of the period t survivors from a cohort which started with an initial distribution $p^0 \in Q^K$(the $K{-}1$ dimensional simplex

generated by the set of possible densities on the finite set of integers, $\{1,...,K\}$). Ericson and Pakes show that as t grows large each of these two distributions converge (pointwise) to a unique invariant distribution, say $P_\omega^*(\cdot)$ and $P_S^*(\cdot)$ (these are invariant to both p^o and to the passage of time). In fact, one can go further than this and show that, given some additional regularity conditions on the locations of p^o and on the transition probabilities, there will be a finite t^* such that for $t \geq t^*$ the convergence will be "monotone", and the sequence of survivor distributions will be stochastically increasing over age (just as is implied by the passive learning model). However, the model with active exploration does not, at least does not necessarily, generate the stronger implication that is generated by the passive learning model; that is, that the size distribution must stochastically increase from each age. On the other hand, the active exploration model does not bar this event from occurring (it does occur in the illustrative example carried along in Ericson and Pakes), and, as noted, it can actually predict that the size distribution will be stochastically increasing at later ages.

There is, however, at least one set of observable implications which differentiates 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 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 depend on (indeed we expect it to be strictly increasing in) the initial size. In the active learning model the parameter determining the firm's profitability distribution, i.e. w, evolves over time. Later year size realizations are governed by a different value of w than those from earlier years and, as time passes, the dependence between the later and earlier values of w, 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 w_t on its history is only through the value of w_{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 $S_t = \bar{S}(w_t)$ so that sales is a deterministic function of w_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 w_t and size we must base our distinction between the active and the passive learning model on more detailed properities of the stochastic process generating size conditional on survival. Let $\{S_t^a\}_{t=1}^{\infty}$ be that process (it is described in more detail below). Then, the active learning model implies that as τ grows large the distribution of $(S_{x+\tau}^a, S_{x+\tau+1}^a, ...)$ becomes, roughly speaking, independent of realizations of $(S_1^a, ..., S_x^a)$. More precisely, we have lemma 4 and its implications (explained immediately after the presentation of the lemma).

4. <u>Lemma</u> (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 $w_0 \in \{1,2,...,K\}$, and M_x^y be the σ–algebra generated by possible realizations of $S_x^a, S_{x+1}^a, ..., S_y^a$. Then $\{S_t^a\}$ ϕ–mixes at a geometric rate, i.e.

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

with $\phi < 1$.

<u>Proof</u>. Let Q be formed from $\underline{P}$ (in 3) by dividing its ith row by $1-p_{i,0}$ (for i=1,...,k) and deleting its first row and column. Q is the Markov chain for the process generating the w–states conditional on survival. Note that (2.i) insures that Q is irreducible aperiodic and hence possesses a unique invariant distribution, say $q^* = [q_1^*, ..., q_k^*]$ (see, for eg., Billingsley, 1979, theorem 8.6). The probability space for the process generating sales conditional on survival is then $(\underline{S}^{\infty}, S, \mathbb{P})$, where $\underline{S}^{\infty}$ is the set of sequences on $\underline{S} = \{s: s = S(\eta), \eta \in \mathbb{N}\}$, S is the

σ–algebra generated by $\underline{S}^{\infty}$, and, for any $w_0 \epsilon \Omega$ and $A \epsilon S$, $\mathbb{P}_{w_0}\{A\}$ is calculated from Q and the family of probability distributions for η conditional on w, $\mathbb{P}_{\eta}$.

A monotone class argument identical to that given in Billingsley (1968, section 20) implies that it suffices to prove

$$|P(E_2|E_1) - P(E_2)| \leq A\phi^{\tau}$$

for any $E_1 \epsilon M^x_1$ such that $P(E_1) > 0$, $E_2 \epsilon M^{x+\tau+\ell}_{x+\tau}$, and $(x,\ell) \in \mathbb{Z}^2_+$. To this end, fix any (E_1, E_2, x, ℓ), let $i_t \epsilon \{1,...,K\}$ index realizations of w_t, and $\underset{\sim}{i}^2$ index realizations of $\underset{\sim}{w}^2 = (w_{x+\tau},...,w_{x+\tau+\ell})$. Then

$$P(E_2|E_1) = \sum_{\underset{\sim}{i}^2, i_x} P(E_2|E_1, \underset{\sim}{i}^2, i_x) \, P(\underset{\sim}{i}_2, i_x | E_1)$$

$$= \sum_{\underset{\sim}{i}^2, i_x} P(E_2|\underset{\sim}{i}^2) \, P(\underset{\sim}{i}^2|i_x) \, P(i_x|E_1)$$

since

$$P(E_2|E_1, \underset{\sim}{i}^2, i_x) = P(E_2|\underset{\sim}{i}^2), \quad \text{and} \quad P(\underset{\sim}{i}_2|i_x, E_1) = P(\underset{\sim}{i}_2|i_x).$$

Use $Q = [q_{i,j}]$ to evaluate $P(\underset{\sim}{i}_2|i_x)$ and substitute to obtain

$$P(E_2|E_1) = \sum P(E_2|i_{x+\tau+\ell},...,i_{x+\tau}) q_{i_{x+\tau+\ell-1}, i_{x+\tau+\ell}} ... q_{i_{x+\tau}, i_{x+\tau+1}} q^{\tau}_{i_x, i_{x+\tau}} P(i_x|E_1),$$

where the summation extends over $i_x, i_{x+\tau},...,i_{x+\tau+\ell}$, and $q^{\tau}_{i_x, i_{x+\tau}} = P(\omega_{x+\tau} = i_{x+\tau} | \omega_x = i_x)$.

Now note that

$$P(E_2|i_{x+\tau}) = \sum P(E_2|i_{x+\tau+\ell},...,i_{x+\tau}) \, q_{i_{x+\tau+\ell-1},i_{x+\tau+\ell}} \cdots q_{i_{x+\tau},i_{x+\tau+1}},$$

where the summation extends over $i_{x+\tau+1},...,i_{x+\tau+\ell}$, so that

$$P(E_2|E_1) = \sum_{i_x,i_{x+\tau}} P(E_2|i_{x+\tau}) \, q^\tau_{i_x,i_{x+\tau}} \, P(i_x|E_1),$$

while

$$P(E_2) = \sum_{i_{x+\tau}} P(E_2|i_{x+\tau}) \, q^*_{i_{x+\tau}} = \sum_{i_x,i_{x+\tau}} P(E_2|i_{x+\tau}) q^*_{i_{x+\tau}} \, P(i_x|E_1)$$

Thus

$$|P(E_2|E_1) - P(E_2)| \leq \sum_{i_{x+\tau},i_x} P(E_2|i_{x+\tau}) |q^\tau_{i_x,i_{x+\tau}} - q^*_{i_{x+\tau}}| \, P(i_x|E_1)$$

$$\leq \phi^\tau \sum_{i_{x+\tau}} P(E_2|i_{x+\tau}) \sum_{i_x} P(i_x|E_1) \leq K\phi^\tau,$$

with $\phi < 1$, where the second inequality follows directly from the fact that the Markov chain converges exponentially to its unique invariant distribution; that is from the fact that for any τ, $\max_{(i_x,i_{x+y})} |q^\tau_{i_x,i_{x+y}} - q^*_{i_{x+y}}| \leq \phi^\tau$ (see, for e.g. Billingsley, 1979, Theorem 8.7)   []

Lemma 4 implies that for any fixed k the distribution of $(S_t,...,S_{t-k})$ conditional on S_1 and survival until period t, eventually (as t grows large) becomes independent of S_1 (more precisely any "dependence" on S_1 goes down at a geometric rate). This generates the following corollary.

5. Corollary.

Fix any $k \geq 0$, and any realization, $S_{t-1} = s_{t-1}, \ldots, S_{t-k} = s_{t-k}$. Then,

$$| E[S_t | s_{t-1}, \ldots, s_{t-k}, \chi_t = 1, s_1] - E[S_t | s_{t-1}, \ldots, s_{t-k}, \chi_t = 1] | \leq A_k \phi^t$$

on a set of $(s_{t-1}, \ldots, s_{t-k}, s_1)$ that has probability one. []

Recall that in the passive learning model the conditional expectation of S_t, conditional on any realization, $(s_{t-1}, \ldots, s_{t-k})$, $S_1 = s_1$ and survival until t, was expected to be strictly increasing in s_1. Hence corollary (1) differentiates the active from the passive learning model. The distinction between the two models is particularly striking in the special case where $S_t = \overline{S}(\omega_t)$, in which case $A_k = 0$ for $k > 1$. We now consider the econometric techniques needed to bring this distinction to data.

Section 4: 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

$$P_s(\cdot | t) \gtrsim_{sw} P_s(\cdot | t-1). \tag{1}$$

The assumptions of the passive learning model imply that it must, while the active exploration model implies it might, but need not — at least in the early ages. The second question posed of the data is whether, for different values of k,

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

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 exploration model. The model with active exploration 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 distributions at different ages. The fact that to examine whether the regression function in (2) is strictly increasing in s_1 we have to condition also on alternate realizations of $(S_{t-1}, ..., S_{t-k})$ makes it more difficult to rely on a pictoral analysis of that hypothesis. As a result we now develop a set of more formal test statistics.

Indeed, since both models suggest that this 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 the relevance of either of our models for the data at hand. If, on the other hand, the hypothesis of weak monotoncity 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 exploration model, while acceptance of only the first is interpreted as support for passive learning (for comparison we will also provide a test of the null that the regression function does not depend on s_1 against an unconstrained alternative). Note that this testing procedure ignores the fact that the active learning model implies that the dependence of the regression function in (2) on s_1 only disappears in the limit. So we should be careful about rejecting the null on the basis of small differences that are very precisely estimated.

Verbally, our procedure consists of allocating the data into alternative groups of cells, where a cell is defined to include all firms with similar sales histories. We then take two groups of firms in the same $(s_{t-1}, ..., s_{t-k})$ cell, but in different s_1 cells, and compare their average values of s_t. If the passive learning model is correct, the group with the larger s_1 values should have a larger average s_t.

More formally, let the function $\sigma^k(\cdot): R^k_+ \rightarrow [1,...,J]^k$ define the required cells by setting

$$\sigma^k(s^{t-1}) = \{\sigma(s_{t-1}),...,\sigma(s_{t-k}),\sigma(s_1)\}$$

where for all τ,

$$\sigma(s_\tau) = j, \quad \text{if and only if} \quad , \underline{\sigma}_{j-1} \leq s_\tau < \underline{\sigma}_j$$

(3)

and $\underline{\sigma}_0 = 0$ and $\underline{\sigma}_J = \infty$. Here the intervals $[\underline{\sigma}_0, \underline{\sigma}_1),...,[\underline{\sigma}_{j-1}, \sigma_J)$, partition the positive orthant into J intervals, and the function $\sigma(s_\tau)$ defines which of these J intervals the realization of S_τ falls into. Similarly, the function $\sigma^k(s^{t-1})$ defines which sequence of intervals the realizations of $S_1, S_{t-k},...,S_{t-1}$ fell into. The J^k possible values of $\sigma^k(\cdot)$, will be denoted by σ^k_p, for p = $1,...,J^k$.

Our testing procedure is based on estimates of the mean and the variance of S_t conditional on the different possible values of σ^k_p. The vectors of population means and variances will be denoted by μ^k and V^k where

$$\mu^k = [\mu^k_p = E\{S_t \mid \sigma^k(S^{t-1}) = \sigma^k_p, \chi_t = 1\}]$$

and

(4)

$$V^k = [V^k_p = V\{S_t \mid \sigma^k(S^{t-1}) = \sigma^k_p, \chi_t = 1\}].$$

In terms of these means our two null hypothesis are as follows. Fix any of the possible values of $[\sigma(s_{t-1}),...,\sigma(s_{t-k})]$, say σ^*. Then the null hypothesis of weak monotonicity can be written as

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

whenever $\sigma_1 \geq \sigma_2$, and the null that the realization of S_1 does not effect the mean of S_t conditional on $(s_{t-1},...,s_{t-k})$, is

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

whatever the value of the couple (σ_1,σ_2).

We adopt the convention that for each σ^*, the vector μ^k is ordered by the associated value of $\sigma(S_1)$ so that each of the weak monotonicity constraints is an inequality constraint of the form, $1'\mu^k \geq 0$, where $1' = [0,0,...,0,-1,1,0,...,0]$. Gathering all such constraints into the matrix R, the null hypothesis of weak monotonicity can be written as

$$H_M^0: \quad R\mu^k \equiv r \geq 0, \tag{5}.$$

Note that R is of full row rank, say C. We test weak monotonicity by testing for (5) under the (unconstrained) maintained hypothesis that $r \in \mathbb{R}^c$. Similarly our test of the null that the realization of S_1 has no effect on the conditional mean of S_t given that any effect is nonnegative will be a test of

$$H_Z^0: \quad r = 0, \tag{6},$$

conditional on the maintained hypothesis in (5).[10]

---

[10]Formally the monotonicity properties that we are building tests for are conditional on a given value of $(S_{t-1},...,S_{t-k})$. The tests themselves, however, condition only on $(S_{t-1},...,S_{t-k})$ being a member of a particular interval. Since the monotonicity property is hypothesized to hold for every $(s_{t-1},...,s_{t-k})$, it holds over any average of realizations of $(S_{t-1},...S_{t-k})$ in a given interval, provided the same averaging procedure is used to compare different values of s_1. Our null hypothesis uses the distribution of $(S_{t-1},...,S_{t-k})$, conditional on $(S_{t-1},...,S_{t-k})$ being a member of the interval, survival, and the realization of S_1, to average. Under the

To obtain tests of these hypothesis we construct estimates of the vector μ (and therefore of r) that satisfy, first just the maintained, and then also the null, hypothesis, and then ask whether any difference between the two estimates can reasonably be attributed to sampling error. To this end let $\tilde{\mu}^k$ and $\tilde{V}^k$ be the sample analogues of μ^k and V^k (the vector of the sample's cell means, and the sample's within cell variances), and N^k be a vector containing the number of observations in each cell. Then, letting $N=\Sigma N(p)$ and assuming that $N(p)$ converges in probability to a constant, say $\delta(p)$, as N grows large ($p=1,...,J^k$), the exchangeability of the observations within each cell guarantees that[11]

$$\text{diag}[\sqrt{N^k}] \, (\tilde{\mu}^k - \mu^k) \sim> \, N(0, \, \text{diag}[V^k])$$

---

strict, (or limiting) form of the mixing condition (lemma 4), this distribuiton will not vary with s_1. Under the alternative, however, it will. Any bias caused by this dependence is a consequence of the finite sample bias in the (nonoverlapping) kernel regression function estimator implicit in our testing procedure, and will, therefore, go to zero with the size of the band width used to generate the cell intervals in (3).

[11]For $| \leq p \leq J^k$, let $S^p = \{S^p_i, i \geq 1\}$ be the sequence of observations with sales history in cell p. By exchangeable within each cell we mean that for each p_0 and each permutation of a finite number of elements from $S^{p(o)}$, say π, $S = [S^p, p = 1,...,J^k]$ has the same distribution as S^*, where

$$S^*{}^p_i = \begin{cases} S^p_i & \text{for } p \neq p_0, \\ S^p_{\pi(i)} & \text{for } p = p_0. \end{cases}$$

Verbally this condition states that the model does not distinquish among observations with the same sales history. Note that it allows for dependence between the sales outcomes of any two observations, and for different distributions for observations with different sales histories. The within cell exchangeability, together with finite variances, quarantee (5). To see this, note first that these conditions imply conditional independence among the cells, conditional on the random distribution which (from de Finetti's theorem) generates the within cell sequences as i.i.d. draws (see Aldous, 1983, Theorem 3.8 and it corollary), and then apply standard results on central limit theorems for exchangeable sequences (see, for eg., Aldous, 1983, 2.25, and 2.27).

while $\qquad$ (7)

$$\text{diag}[\tilde{V}^k] \overset{P}{\to} \text{diag}[V^k]$$

where $\text{diag}[x]$ denotes a diagonal matrix with x on the principal diagonal, $N(\cdot, \cdot)$ denotes a multivariate normal distribution, $\overset{P}{\to}$ denotes convergence in probability, and $\sim>$ denotes convergence in distribution.

We obtain our test of H_M^0 (in 5) by comparing

$$\tilde{r} = R\tilde{\mu},\qquad (8a)$$

and

$$\tilde{r}_M = \underset{r \geq 0}{\text{argmin}} \{(r-\tilde{r})' R[\tilde{V}^k]^{-1} R'(r-\tilde{r})\},\qquad (8b).$$

$\tilde{r}$ is an unconstrained estimate of r obtained from substituting sample for population means in the formula for r. $\tilde{r}_M$ is an estimator of r that is forced to satisfy the inequality constraint in (5). Subject to that constraint it is chosen to minimize a quadratic form in $(r-\tilde{r})$. This quadratic form uses $R[\tilde{V}^k]^{-1} R'$, the estimated variance–covariance matrix of $\tilde{r}$ under the hypothesis that $R\mu^k = 0$, as its weighting matrix. Given (8) our test of whether $\tilde{r}$ and $\tilde{r}_M$ are significantly different from one another can rely on distributional results for inequality constrained estimators that date back at least to Barlow (1959).[12] Define

$$\chi_M^2 = \underset{r \geq 0}{\min}\{(r-\tilde{r})' R[\tilde{V}^k]^{-1} R'(r-\tilde{r})\},\qquad (9),$$

so that large realized values of this statistic are evidence against the null. Then, as is shown

---

[12]For a more recent exposition see Barlow, Bartholemew, Bremner, and Brunk, 1972, and the econometric literature which started with the article of Gourieroux, Holly, and Monfort, 1982. More recent contributions include Goldberger (1987) and Wolak (1989).

by Barlow (1959), for all $a \geq 0$

$$T_M(a) = \Pr\{\chi_M^2 \geq a \mid r=0\} \rightarrow \sum_{c=0}^{C} W(c) \Pr\{\chi_c^2 \geq a\}$$

as N grows large, where (10)

$$W(c) = \Pr\{\tilde{r}_M \text{ has exactly c zero components} \mid r=0\},$$

and χ_c^2 denotes a χ^2 deviate with c degrees of freedom (c=0,...,C). Thus if χ_M^0 is the realized value of χ_M^2, $T_M[\chi_M^0]$ provides the probability of type I error (the p–value) of a test that would reject the null if $\chi_M^2 = \chi_M^0$ when the true value of r was zero. The p–value when r is any value greater than zero is smaller.

Unfortunately, the orthant probabilities, that is the values of the W(c) needed to calculate (10), cannot be expressed as easy to evaluate functions (this will generally be true provided, as in our case, the weighting matrix in 6b is not diagonal). As a result we obtain simulated estimates of the W(c), say $\tilde{W}(c)$, and use them to provide a simulated estimate of $T_M(\cdot)$, say $\tilde{T}_M$, where

$$\tilde{T}_M(a) = \Sigma \tilde{W}(c) \Pr\{\chi_c^2 \geq a\} = \tilde{W}'X$$

and (11)

$$\tilde{W}' = [\tilde{W}(0),...,\tilde{W}(C)], \quad \text{and} \quad X' = [\Pr\{\chi_0^2 \geq a\},...,\Pr\{\chi_C^2 \geq a\}].$$

The $\tilde{W}(c)$ are obtained by taking NSIM draws from a C dimensional normal distribution with mean 0 and variance–covariance $R[\tilde{V^k}]^{-1}R'$, passing each through the estimation subroutine, and then simply counting up the number of times the simulated draws for r produce an $\tilde{r}_M$ with precisely c zero components (for c = 0,...C). Note that the $\tilde{W}(c)$ are, in fact, cell means from repeated draws from a multinominal distribution (where NSIM is the number of draws), so the variance of $\tilde{T}_M(a)$ about its expected value of $T_M(a)$ can be obtained from the formula for the variance of the multinomial as

$$\text{Var}\,[\tilde{T}_M(a)] = X'\{\text{diag}\,[W] - WW'\}X[\text{NSIM}]^{-1}, \qquad (12).$$

Along with the point estimate of $T_M(a)$ given by (11), we provide an estimate of its variance obtained from substituting the simulated for the actual values of W in (12). For comparison we also provide the estimate of $T_M(a)$ that would be obtained if we were to incorrectly assume that $R[V^k]^{-1}R'$ was an identity matrix, since this would allow us to calculate the W(c) directly from a simple combinatorial formula (C choose c), and one might like to see if this simplification produces a significantly biased estimate of $T_M(a)$.[13]

The test statistic for the null hypothesis in (6), that is for the hypothesis that the regression function does not depend on s_i, conditional on the maintained hypothesis that the regression function is nondecreasing in that variable , is based on the difference between $\tilde{r}$, the estimate of r which satisfies the nonegativity constraint, and zero. Use of the estimate of the variance–covariance of r to weight the difference between these two estimators generates the test statistic

---

[13]Note that we need only simulate the weights for, and not the entire distribution of, the test statistic. Use of the available information on the vector X in this way makes it easy to obtain fairly precise estimate of the value of the test statistic with a small number of simulation draws (see the next section).

$$\chi_Z^2 = \tilde{r}'_M R[\tilde{V}^k]^{-1} R' \tilde{r}_M, \qquad (13).$$

The results on testing subject to inequality constraint imply that conditional on the null in (6), χ_Z^2 also distributes as a weighted average of χ^2 deviates. That is for any 'a'

$$T_Z(a) = \Pr\{\chi_Z^2 \geq a \,|\, r=0\} \rightarrow \Sigma W^*(c) \Pr\{\chi_C^2 \geq a\},$$

as sample size grows large, where, in this case, $\qquad (14)$

$$W^*(c) = \Pr\{\tilde{r}_M \text{ has exactly c positive components} \,|\, r=0\}$$

and χ_C^2 is defined as above (c=0,...,C). We obtain estimates of the $W^*(c)$ from the same procedure used to obtain the estimates of $\tilde{W}(c)$ used in (10), and use these estimates to generate an estimated value of $T_Z(\chi_Z^0)$ and its standard error (the p–value for the test of r=0 conditional on the maintained that r$\geq$0).

We also 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 test of whether s_1 has no affect on the regression conditional on an unconstrained maintained hypothesis. A test of the null in (6) conditional on an unconstrained maintained, say χ_T^2, can be built from standard tests for the equality of means, that is

$$\chi_T^2 = \tilde{r}' R[\tilde{V}^k]^{-1} R' \tilde{r}, \qquad (15),$$

and has the familiar χ^2 limit distribution with C degrees of freedom. Note that since the

properties of Lagrange multipliers insure that

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

(13), (9), and (15) insure that

$$\chi^2_T = \chi^2_M + \chi^2_Z,$$

with probability one, as will be illustrated in the empirical results which we now turn to.

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).

Note that the data covers firms in all (not just the manufacturing) sector. Also,

as explained below, panels created from UI account numbers can distinguish between the two

major sources of attrition in firm level panel data sets discussed in the introduction; attrition

due to liquidation, and attrition due to changes in ownership.[14]

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

---

[14]We are grateful to D. Nuendorf and R. Shaffer 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). The UI account rules allow multiestablishment firms to choose to report as a single, or as multiple, units. For consistency the establishments of multiestablishment firms that reported separately have been merged into single observations in this data. The data should therefore be thought of as firm—level data.

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 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).

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 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 that 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. These figures 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. (However, almost identical empirical results are obtained if we include the firms in the analysis until the year before they transfer out.)

The passive learning model together with the assumption that prices are constant over time implies that the proportion of surviving firms with size greater than any X, or the

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

X	Age								Cross Section [a]	
	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"	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"	13	11	10	1	0	0	0	0		

Notes to table 3:

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

numbers in each row in 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 transition which do not satisfy this condition. On the whole, the consistency of the data with the hypothesis is quite striking — particularly in retail trade. 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.

We now pause briefly to consider whether movements of prices (generated, in part, by changing demand conditions) are likely to provide an alternative explanation for the observed changes in the size distribution of survivors illustrated in tables 2 and 3. As noted in section 2, if prices are not constant over time, but the production function is homothetic, then the measure of size for firms active in period t, $S(t)$, will be $f_1(t)f_2(\eta_t)$; where $f_1(t)$ is a function of prices, $f_2(\cdot)$ is an increasing function, and, as before, the distribution of η is stochastically increasing in θ. The passive learning model implies that the distribution of, θ, and hence of $f_2(\cdot)$, among the survivors of a cohort, will be stochastically increasing in age. To analyze the implications this result on the evolution of the observed size distribution of survivors under the assumption that prices can vary over time but that the production function is homothetic, we need a consistent estimate of $f_1(t)$, say $\hat{f}_1(t)$. Given such an estimate we can use it to construct $\hat{S}(t) = \hat{f}_1(t)^{-1}S(t)$, and then examine whether the survivor distribution of the $\hat{S}(t)$ = $f_2(\eta_t)$ is stochastically increasing in age. The model suggests at least one simple way of estimating $\{f_1(t)\}$. Firms who are old enough should know their value of θ almost exactly. Thus provided the marginal distribution of η conditional on θ for these firms does not vary from year to year, a comparison of the mean sizes of the older firms over time should provide a consistent estimate of $\{f_1(t)\}$ (up to a normalization). Though we do not know the age of

the firms that were alive in the first year of our sample (1978), an analysis of the 1986 cross section indicates that the vast majority of the firms with more than 50 employees are firms older than age eight, so we estimated $f_1(t)$ by comparing the mean size of the largest firms in different periods.

As is well known this was a period of contraction for manufacturing in Wisconsin. The contraction was accompanied by a continual "fall" (in the first order stochastic dominance sense) in the cross sectional size distribution of active firms over the period (compare, for example, the last two columns of Table 3). In particular the average size of firms with more than fifty employees fell by about 14% over the period. This implies that were we to plot the implied distribution of $\hat{S}(t)$ (as defined in the last paragraph) for our cohort over time, it would be stochastically increasing at an even faster rate than the distribution of $S(t)$ is. Put differently the size distribution of survivors was increasing during a period when both the size of the market they were in, and the size of the other incumbents in that market, were falling. The estimates of $\{f_1(t)\}$ for retail trade did not vary much over the period (there was an increase of under 6% over the eighth year period). As a result our estimate of the evolution of the distribution of $\hat{S}(t)$ among survivors in retail was very close to that for $S(t)$ plotted in Table 2. We conclude that it is unlikely that price variation over the period underlies the stochastically increasing evolution of the size distribution of survivors observed in Tables 2 and 3.

Further insights into the nature of the evolution of the size distribution of survivors can be gleaned from a comparison of the empirical distributions of $S(t+1)$ conditional on alternative possible values of $S(t)$ at different ages (that is, from the Markov transition matrices at different ages). Table 4 plots these distributions for the retail cohort between ages one and two (4A), and between ages six and seven (4B); while Tables (5A) and (5B) provide the analogous information for the manufacturing cohort. The figures in brackets beside the entries in the table are consistent estimates of the variance of the transition probabilities provided in the table, while the last column lists the number of firms underlying

Table 4. Markov Transition Matrices: Retail Trade[*]

Size in t \ Size in t+1	Shutdown	1-5	6-10	11-15	16-25	26+	N (row)
			4A: t = 1				
1-5	.20(.01)	.73(.02)	.06(.01)	.01(.10)	0.0	0.0	917
6-10	.11(.03)	.18(.03)	.54(.04)	.14(.03)	.03(.02)	0.0	132
11-15	.14(.05)	.04(.03)	.37(.07)	.37(.07)	.10(.04)	0.0	52
16-25	.10(.06)	0.0	.13(.06)	.20(.07)	.30(.08)	.27(.08)	30
26+	.06(.03)	0.0	0.0	.04(.03)	.12(.05)	.78(.06)	49
			4B: t = 6				
1-5	.05(.01)	.90(.02)	.04(.01)	0.0	0.0	0.0	371
6-10	.09(.03)	.19(.04)	.62(.05)	.08(.03)	.01(.01)	.01(.01)	100
11-15	.04(.03)	0.0	.09(.04)	.71(.07)	.13(.05)	.02(.02)	45
16-25	.05(.05)	.05(.05)	.05(.05)	.10(.07)	.66(.10)	.10(.07)	21
26+	0.0	0.0	0.0	.06(.04)	.03(.03)	.91(.05)	34

[*]The (i,j) element of these matrices is the fraction of firms with age t size class equal to i that have age t+1 size class equal to j. Numbers in brackets beside entries are there estimated standard errors. N is the number of firms in age t size class i.

Table 5. Markov Transition Matrices: Manufacturing

Size in t / Size in t+1	Shutdown	1-5	6-10	11-15	16-25	26+	N (row)
4A: t = 1							
1-5	.16(.02)	.69(.03)	.09(.02)	.05(.01)	.01(.01)	0.0	245
6-10	.11(.04)	.24(.06)	.35(.06)	.19(.05)	.11(.04)	0.0	54
11-15	.14(.09)	.07(.07)	.14(.09)	.21(.11)	.36(.13)	.07(.07)	14
16-25	0.0	0.0	0.0	0.0	.8(.18)	.2(.18)	5
26+	.33(.16)	0.0	0.0	0.0	.22(.14)	.44(.13)	9
4B: t = 6							
1-5	.08(.03)	.81(.05)	.10(.03)	.01(.01)	0.0	0.0	77
6-10	.05(.03)	.15(.06)	.76(.07)	.05(.03)	0.0	0.0	41
11-15	0.0	0.0	.19(.09)	.62(.11)	.19(.09)	0.0	21
16-25	0.0	.06(.06)	.06(.06)	.24(.08)	.53(.12)	.12(.08)	17
26+	0.0	0.0	0.0	.06(.06)	.06(.06)	.88(.08)	16

*See the footnote for Table 4.

the empirical distributions given in each of the rows.[15] The striking fact evidenced by these tables is the increase in the diagonal elements of the transition matrices (the 'squared off' numbers) as we move from the age one transition to the age six transition (plots of the intervening matrices show that these diagonal elements, that is that the fraction of firms staying in the same size class, increases from each transition to the next). There is alot more movement among size classes in the early, than in the later, ages; just as one might expect from a learning process if, by the later ages, firm's have already acquired a fairly precise posterior on the true value of their ability parameters.[16] Also it seems that, conditional on survival, the transitions in the later ages are not as weighted towards increases as they are among the early ages (and looking back to Tables 2 and 3 we find that the increase in the size distribution of survivors from period to period seems more dramatic in the earlier, than in the later, ages).

So far we have noted only the similarities in the evolution of the manufacturing and retail cohorts pointed out by Tables 2 to 5. A more detailed look at those tables, however, uncovers some revealing contrast. Looking first at Tables 2 and 3, it is clear that 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 firms in the largest size classes).

---

[15]We thank an editor for suggesting plotting the transition matrices. The variances of the entries in the table were calculated from the multinomial formula. Note that the elements within each row are negatively correlated, with covariance consistently estimated by the negative of the multiple of the two row entries divided by N.

[16]We also plotted the transition matrices generated by the twelve size classes in Tables 2 and 3. Though these were more cumbersome, they illustrated the same results; the number of firms staying in the same size class increased continually from transition to transition culminating in an increase of 50% in retail trade (from .41 to .62) and of 40% in manufacturing (from .36 to .5). Were we to have presented the transition matrix from age seven to eight rather than the transition from age six to seven, the change in the transition matrices from the early to the later ages emphasized here would not have been as striking (though still quite evident). The age seven to eight transition was the only transition in which the diagonal elements of the transition matrix actually fell from the previous tranistion, and we attribute this to the fact that the data indicate that 1985 to 1986 period seems to be more turbulent than the years preceeding it (look, for example, at the jump in the hazard rate in that year recorded in both Tables 2 and 3).

However, by age eight the 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 (and this would be true to an even greater extent if we were to make the adjustments for price variation detailed above).

Moreover, though the transition matrices indicate that there is much more jumping from size class to size class in the early ages in both sectors, it is also quite clear that the diagonal elements of these transition matrices are always lower in manufacturing. That is the size class of the manufacturing firms do not seem to stabilize to the extent that those of retail firms do (and the difference seems particularly marked, at least in the later ages, in the larger size classes). In fact, if we go back to Table 2 we see that by age eight the size distribution of the retail cohort is quite close to the cross—sectional distribution of all retail firms active in 1978 (or in 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 means, 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). Thus, if we were to think of the cross—sectional distribution as an approximation to the limit distribution (though this is not true, the cross—section is dominated by older firms), 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 be strictly increasing in that variable, and the active learning model implies that it will not.

Tables 6 and 7 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 parentheses below their values). The second column provides the p–values that would be obtained if the components of the estimator of the vector of constraints being tested had mutually independent distributions under the null. In this case the weights required for the calculation of the limit distribution (see equation 4.7) have an analytic form, so there is no need for simulation. Though the independence assumption is wrong in our (and probably in most) cases, it does provide an easily calculable approximation to the non–analytic true p–value which might be of use in (at least) the preliminary stages of analysis if the approximation produced numbers that were sufficiently close to those we are after. Comparing columns (1) and (2) in the next four tables, it is clear that in the cases where the true p–values were low (say less than .10), so that there was some chance of rejecting the null, the approximation did produce a value within .05 of the value we were after.

Note first that none of the tests reject weak monotonicity at traditional levels of significance. So both the retail and the manufacturing data are 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

Table 6. Tests for Mean Independence of the Distribution of S_t Conditional on $S_{t-1}, ..., S_{t-k}$, 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_M^0	p-values[b] (1)	(2)	C	χ_z^0	p-values[b] (1)	(2)	Df	χ_T^0	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 7. 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, $+ \infty$

k	Weak Monotonicity				Zero Conditional on Monotonicity				Unconditional Zero		
	C	X_M^0	p-values[b] (1)	(2)	C	X_z^0	p-values[b] (1)	(2)	Df	X_T^0	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.

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 average 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 <u>reject</u> 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).[17]

The results for the test of zero conditional on weak monotonicity are strikingly different in manufacturing. Table 7 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 <u>accepting</u> 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.

Table 8 and 9 push the nonparametric analysis one step further and ask what order of Markov process provides an adequate nonparametric fit to the (expectation from the)

---

[17]Two points are worth noting here. First the p—values for the tests of zero conditional on monotonicity should be treated with caution as no adjustment has been made for the sequential nature of the testing procedure. As a result the probability of type I error conditional on the null and observing the test statistics given in the table are larger than the p—values reported in these columns. Second 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 unconditonal zero columns) is a tradtionally acceptable .11.

Table 8. 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			Markov Conditional on Montonicity			Unconditional Markov		
	C	X_H^0	p-value[c] (1) (2)	C	X_z^0	p-value[c] (1) (2)	Df	X_T^0	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 9. 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			Markov Conditional on Monotonicity			Unconditional Markov		
	C	X_M^0	p-value[c] (1) (2)	C	X_z^0	p-value[c] (1) (2)	Df	X_τ^0	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.

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 6 and 7. That is, we first test whether first year size, size in the first two years,..., have a nondecreasing effect conditional on the sizes in all the intermediate years, 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 we need a third order nonparametric Markov process 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 9 and 7). Table 7 implies that conditional on realizations of S_6 realizations of S_1 do not affect the regression function. Table 9 says that realizations of S_5, and of S_4, do. The active exploration 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 value in year 7, and therefore, have a more distinct effect on the regression function for S_7, than its value in year 1 will.[18]

---

[18]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

Section 6. A <u>Summary</u> <u>and</u> <u>Some</u> <u>Further</u> <u>Implications</u> <u>of</u> <u>the</u> <u>Empirical</u> <u>Results</u>.

Our empirical results can be summarized quite succinctly. The size distribution of the firms surviving from a cohort tends to stochastically increase from age to age in both retail trade and in manufacturing, but the rate of increase is much more rapid in manufacturing. Also, the variance in future size conditional on current size tends to decline over age in both sectors, but it tends to be smaller in retail trade at all ages. Both these results suggest the importance of learning and selection in determine the dynamic patterns of firms behavior, but the nature of the learning process seems to be different in the two sectors. Indeed formal tests indicate that, in retail trade, size in the initial years is closely tied to size in all subsequent years (even after we condition on intermediate sizes), but that the sizes of the surviving manufacturing firms quickly lose their dependence on their early year sizes (at least if we condition on size in at least one immediately preceding period).

Going back to the theoretical section of this paper, we find that, for manufacturing these empirical results are consistent with the implications of the active exploration model, but not with those of the passive learning model. Conversely, the empirical results for retail trade are consistent with the implications of the passive learning model but not with those of the model with active exploration. The importance of this distinction is that the two different models imply different frameworks for the analysis of phenomena that depend upon the sources of firm—specific uncertainties and, consequently, for the analysis of how different factors impact on the distribution of outcome paths among firms within an industry; phenomena such as the behavior of capital markets when there are significant failure probabilities, and the causes of job turnover (or of the changes in industry structure) generated by the growth and contraction of individual firms within industry aggregates (see the discussion in Jovanovic, 1982, and in Ericson and Pakes, 1989).

$S_{t-1},..., S_{t-k}$ 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 makes it clear that the stochastic process generating size is not first order Markov, so the special case discussed by Jovanovic (1982) is not relevant.

The nonparametric results ought also to effect how we account for liquidation induced attrition in the analysis of longitudinal firm–level data. As an example of the importance of such corrections, 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 familiar 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 great deal of understanding of the environment generating the data). Davis, Gallman and Hutchins (1987) do indeed find a significant <u>positive</u> first order effect of age on vessel productivity.

It is worth noting that the nonparametric implications we used to test for the relevance of alternative models are rich enough to enable the development of estimation procedures that can separate out the structural production function coefficients from the effects of the selection induced by liquidation, in examples such as this one. To see this assume that output is a parametric function of inputs, say $f(x_t, \beta)$, and an additive disturbance, say ϵ_t, whose value is not known when input decisions are made. Then the expectation of output (y_t), conditional on the current value of inputs (x_t), survival until period $t(\chi_t = 1)$, and the information set available in $t-1(J_{t-1})$ is a sum of two functions; the structural production function, and the expectation of the disturbance conditional on $\chi_t = 1$ and J_{t-1}, i.e.

$$E[y_t | x_t, \chi_t=1, J_{t-1}] = f(x_t, \beta) + E[\epsilon_t | \chi_t=1, J_{t-1}].$$

Now note that both models imply that the decision as to whether to operate the firm in year t is determined by information available in t–1 (i.e. χ_t is measurable with respect to J_{t-1}), so the last term depends only on variables in J_{t-1} (though on different variables in each of the two models). This implies that none of the determinants of $E[y_t | x_t, \chi_t, J_{t-1}]$ are determinants of both, $f(\cdot;\beta)$ and $E[\epsilon_t | \chi_t=1, J_{t-1}]$. As a result, once we determine which of the dynamic models are relevant for the data at hand, and therefore what variables determine $E[\epsilon_t | \chi_t=1, J_{t-1}]$, we can, under mild regularity conditions, obtain a (root n) consistent asymptotically normal estimator for β_0 (the true value of β) by minimizing a distance between y_t and sum of $f(x_t, \beta)$ and a non parametric estimator for the 'nuisance' function $E[\epsilon_t | \chi_{t-1}=1, J_{t-1}]$ (for details see Robinson, 1988). Note that, just as was the case for our testing procedures, this method of correcting for the selection process induced by liquidation behavior does not require either the precise functional forms of the relevant dynamic processes, nor the (computationally difficult) solution to the problem of finding the optimal stopping states as a function of the parameters of the model. On the other hand, these selection correction procedures are fully consistent with the structural economic models generating them, and can, therefore, be directly integrated into the analysis of any issue provided that the framework used in that analysis is consistent with one of these models.

References

Aldous D.J., (1983); "Exchangeability and Related Topics" pp 2–199 in Springer Lecture Notes in Mathematics, No. 117 Ecole d'Ete de Probabilites de Saint–Flour XIII.

Bartholemew, D.J. (1959); "A Test of Homogeneity for Ordered Aternatives", Biometrika, 46, pp 36–48.

Barlow, R.E., D.J., Bartholemew, J.Wm. 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., Roberts, M., and L. Samuelson (1988); "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries", Rand Journal Vol. 19 #4, pp. 495–515.

Dunne, T., Roberts, M. and L. Samuelson (1989); "The Growth and Failure of U.S. Manufacturing Plants", Q.J.E., Vol. 104, #4 pp. 671–698.

Ericson, R., & A. Pakes (Revised 1989); "An Alternative Theory of Firm and Industry Dynamics," Discussion Paper No. 445, 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 of Economy, Vol. 95, pp. 657–674.

Frydman, H., and B. Singer (1985); "Assessing Qualitative Features of Longitudinal Data", pp. 308–24 in J. Heckman and B. Singer. (ed), op. cit.

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

Gourieroux, 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.

Hall, B. (1988); "The Effect of Takeover Activity on Corporate Research and Development" in A. Auerbach (ed) <u>Corporate Takeovers: Causes and Consequences</u>, Chicago University Press.

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

Heckman, J., & B. Singer (1985); "Social Science Duration Analysis," in J. Heckman, and B. Singer (ed); op. cit.

Heckman, J., and R. Robb (1985); "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer (ed.), op. cit.

Jovanovic, B. (1982); "Selection and the Evolution of Industry," <u>Econometrica</u>, 50, pp. 649–70.

Karlin, S. and Rinott, Y. (1980); "Classes of Orderings of Measures and Related Correlation Inequalities I: Multivariate Totally Postive Distributions", <u>Journal of Multivariate Analysis</u> 10, 467–498.

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

Marschak, J., & W.H. Andrews Jr. (1944); "Random Simultaneous Equations and the Theory of Production," <u>Econometrica</u>, 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.

Pakes, A., and R. Ericson (1987); "Empirical Implications of Alternative Models of Firm Dynamics", SSRI working paper 8803, University of Wisconsin.

Robinson, P. (1988); "Root–N Consistent Semiparametric Regression," Econometrica, Vol. 56, pp. 931–954.

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.

Whitt, W. (1980); "Uniform Conditional Stochastic Order", Journal of Applied Probability, 17, pp. 112–123.

Whitt, W. (1982); "Multivariate Monotone Likelihood Ratio and Uniform Conditonal Stochastic Order", Journal of Applied Probability, 19, 695–701.

Wolak, F.A. (1989); "Testing Inequality Constraints in Linear Econometric Models", Econometric Theory, pp 1–35.

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: Example 1

This example shows that the passive learning model does not necessarily imply that either hazard rates, or the variance in growth rates, are nonincreasing in age (and that this is true whether or not we condition on size).

<u>Example</u>

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

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

The posterior for θ in this problem depends only on the couple (x_t, t), where $x_t = \max [n_1, ..., n_t]$. 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 $\Pr\{x_{t+1} = 1 | x_t = 0, t\} = \Pr\{\eta_{t+1} = 1 | x_t = 0, t\}$ 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 the 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 $V(S_{t+1}-S_t|S_t) = V(S_{t+1}|\theta=\delta) = S^2\delta(1-\delta)$, regardless of S_t. If $t < t^*$ and $S_t = S$, then θ still is δ with probability one, and $V(S_{t+1}-S|S_t)$ 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

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

which can be made as large as we like by choosing 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 $r + \overset{t}{\Sigma}\eta_i$ and $s+t-\Sigma\eta_i$, so that the sum, $x_t = \overset{t}{\Sigma}\eta_i$, 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) $\overset{\geq}{\underset{<}{}}$ Φ according as $\lessgtr t^*(x)$ [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. []
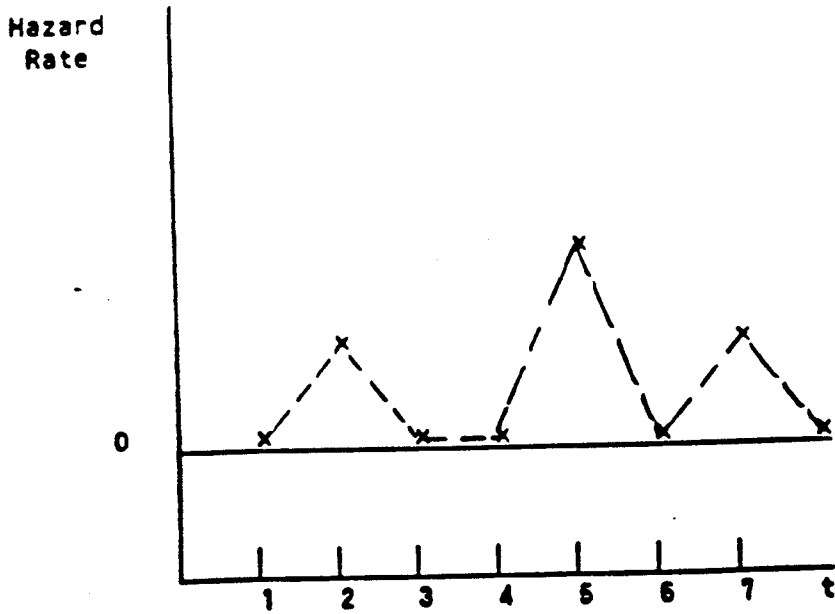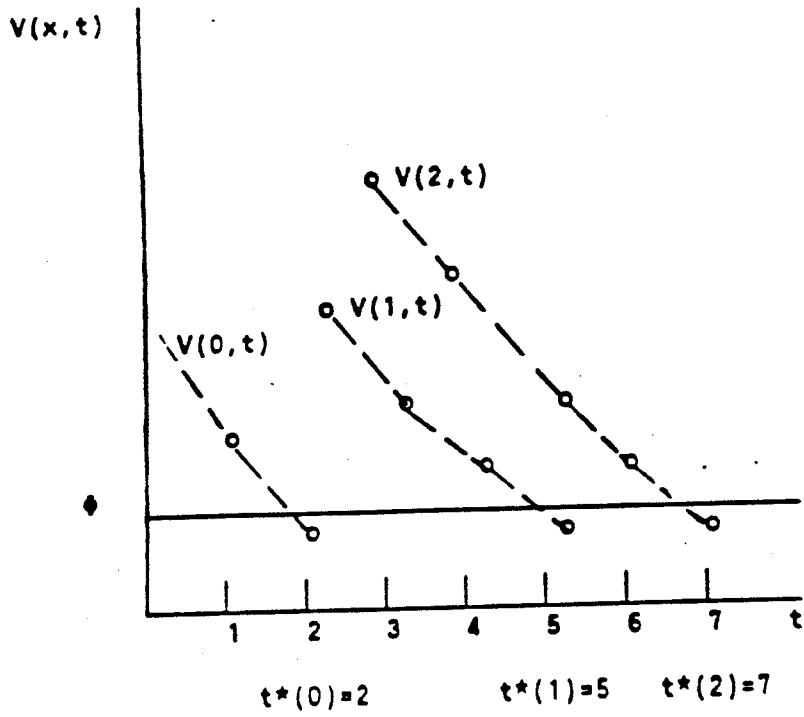
Figure 2: A Beta/Binomial Example