

filename = carpop.doc

MODELLING THE DYNAMICS OF INDUSTRY POPULATIONS*

P.A. Geroski

And

M. Mazzucato

London Business School

November, 1999

PRELIMINARY DRAFT: Please do not cite without the permission of the authors. Comments are most welcome.

* We are obliged to the ESRC for financial support, to Steve Klepper and Ken Simons who generously gave us some of their data, and to John Haltiwanger, participants at the 1999 CAED conference in the Hague and at a seminar at Hitotsubashi University for a number of useful observations and comments on early versions of the paper. The usual disclaimer applies however.

I. INTRODUCTION

All markets change, progressing through phases of birth, growth, maturity and decline. Different phases of market evolution are usually characterized by different market structures and different patterns of innovation. Particular firms appear and then disappear, sometimes growing to dominate their market for a time and sometimes making little impression on anyone. In general, markets adapt to the exogenous forces which drive change in one of two ways: through changes in the population of firms (“*selection*”), and through changes in the characteristics and competencies of the firms who populate the market at any time (“*adaptation*”). When the ability of firms to adapt to new circumstances is weak, most of the brunt of change will be recorded as changes in the population of firms operating in the market over time. Tracing movements in this population may, therefore, be an interesting and informative way to trace out patterns of market evolution.

This paper explores four models which might be used to account for industry population dynamics. Two of the models are standard textbook stories: the first argues that population size is driven by **market size**, while the second is the standard **negative feedback model of entry and exit**. The other two models are less standard. The third, a **contagion model of entry and exit**, suggests that waves of entry or exit are driven (or at least magnified) by fads or information cascades which, in effect, build up an overreaction to surprise shocks. The fourth and final model that we shall examine is a **density dependent** model of population growth derived from the work of organizational ecologists. It turns out that all four models are fairly closely related to each other, differing in features which are relatively easy to identify and, therefore, test. We shall apply all four models to an illustrative data base built up around about 100 years of data on the population of domestic car producers in the US.

The plan of the paper is as follows. We start in Section II with a brief discussion of why modelling the population dynamics of markets is interesting, and what we might learn from it. We then move on (in Section III) to outline the four models, and express them in a form which will enable us to examine them empirically without requiring a lot of data (we will need only data on the population of firms in each year and total sales). It is often useful to explore empirical models by applying them to particular data sets, and so, in Section IV, we make some illustrative calculations by applying the four models to data on the population of US car producers since 1894. It turns out that

the two simple textbook models are basically inconsistent with the data, spectacularly so in one case. Density dependence seems to play some role in accounting for population dynamics in this market, but it cannot account for what turns out to be the most stark feature of our data on car producers. The contagion model makes a useful contribution to explaining population dynamics in this industry, and the estimates generated using it are surprisingly robust. Finally, Section V concludes the essay by suggesting which of the results from our illustrative calculations might generalize to other sectors, and outlining the issues which lie ahead for further research.

II. THE CHALLENGE

A case for studying movements over time in the number of firms operating in a particular market can be made on the simple grounds that the patterns of population growth and decline that we typically observe in markets are interesting, and unusual enough to warrant a closer examination. The early colonization of a market often induces an S-shaped curve tracking population over time, and most markets undergo one or more periods of rationalization whose depth and timing are often a challenge to explain. More compelling is the argument that studying changes in population size provides useful information on changes in competitive conditions in particular markets. When firms operate independently and are of roughly equal size, then a population count gives a fair representation of how many alternatives consumers have to purchasing the products of any particular firm. It also provides a way of thinking about how difficult it is likely to be for these firms to get together and collude over prices.¹

The best reason for being interested in industry population dynamics, however, is that the rise and fall of markets is often mirrored by movements in the number of firms which inhabit them. The major phases of market evolution – birth, growth, maturity and decline – are usually identifiable by major changes in population size, and it seems reasonable to believe that any explanation of population dynamics will, as a matter of course, cast light on the deeper mechanics of market evolution. In fact, the diversity of the products offered to consumers also seems to vary over market life cycles, usually contemporaneously with the number of firms in the market. This is very marked in the early development of markets (where the number of product varieties and the number of firms is large), in the growth phase (when both drop, often precipitously) and, somewhat less clearly, in the mature phase (when market entry and product variety sometimes flower).² The association between entry and product proliferation arises because entrants are often an important source of new product innovations. Large, multi-product organizations – particularly those that are successful and well established -- are frequently slow to innovate, and often try to avoid making radical changes. This, of course, creates room for new firms to come in and take advantage of opportunities which incumbents either do not see or choose to spurn.³

The challenge, then, is to develop models which allow one to account for changes in industry population size over time. However, to insure that the exercise results in a model which is

useful, it is important to put some constraints on it. It is always possible to build a model which is driven by factors which are unobservable, but such models are valuable only as illustrations of what might be. To be useful, a model needs to be implementable and potentially refutable, and this means that the important exogenous variables must be relatively easy to observe. What is easy to observe depends on the nature of the data of course, but long histories of industries are rarely able to produce much more than an annual measure of population count and market size (plus a loose dating of major events, including major technology changes). In what follows, then, we shall work with models which require only this very minimum amount of information.

III. MODELS OF INDUSTRY POPULATION DYNAMICS

There are two paths that one might follow in trying to model movements in the population of firms in a particular market over time. One is to focus on particular periods of market evolution, or try to model the dynamics of particular events (such as a shakeout of producers).⁴ This strategy has the commendable virtue of recognizing that evolution might involve structural changes which introduce discontinuities, and it allows one to focus on the details of particular competitive episodes or discontinuities. However, if the nature of what happens in a particular competitive episode depends on when and why that episode starts, then it is impossible to analyse that episode in isolation from the events which preceded it. The second strategy starts from this observation, namely that the events which occur at any particular time in a market are part of an unfolding pattern of events which are difficult to appreciate if studied in isolation. This, in turn, inclines one to seek a single, all-encompassing model which might account for the sequence of events which one observes over time in the market.

There are now a number of models of industry evolution which provide attractive ways of thinking about how to account for movements in the number of firms (and other features of market structure) over time.⁵ Klepper, 1996, for example, develops a model in which increasing returns to process innovation and convex costs of growth cause leading firms in a market to develop an increasingly large competitive advantage over smaller rivals or entrant firms through their investments in process innovation. This transforms a market which was originally colonized by many firms introducing many product innovations into a thinly populated, highly concentrated oligopoly which displays relatively little product innovation. In much the same spirit, Jovanovic and MacDonald, 1994b, develop a model in which bursts of entry and the consequent rise in firm numbers are triggered by inventions developed outside the industry which firms try to implement. Fast movers establish a more or less permanent place in the industry while laggards enter later and then, in all likelihood, exit. Both models account for changes in the number of firms in a market over time in ingenious ways, but both do this by construction (both papers work backwards from a similar set of stylized facts). Further, both papers turn on patterns of technology change which are hard to observe. Neither yields a simple reduced form equation describing movements in an industry population which can be compared with sensible alternative hypotheses.

Thus, and without wishing to cast any aspersions on these two (and other similar) models, we propose to take a slightly different approach.⁶ Our goal in what follows is to construct a number of simple econometric models of industry population dynamics which are either nested or are otherwise relatively easy to compare against each other. Since it is typically very difficult to observe many features of industry evolution, we will focus on an extreme case, developing simple econometric models which explain movements in the number of firms present in a particular market at any given time, $N(t)$, only in terms of industry size, $S(t)$, and the past history of both $N(t)$ and $S(t)$. There are, no doubt, hundreds of slightly less simple models which one can develop that use slightly more information than this, but our guess is that most of these slightly less simple models will be rooted in one of the ones which we will consider here.

The models that we will examine are: a **market size model**, a **negative feedback model of entry and exit**, a **contagion model of entry and exit** and a **density dependent model** of population growth. Let us consider each in turn.

the market size model

The simplest textbook explanation of market structure is based on the insight that large markets are likely to host more firms than small markets, and, as a consequence, are likely to be less concentrated. Needless to say, the relationship between market size and market structure is not simple, since the number of firms and degree of concentration in markets of the same size depends (inter alia) on the degree of price competition and the level of endogenous sunk costs.⁷ It is easy to make this argument slightly more precise. If $\pi(N)$ denotes profits per firm and F fixed (and sunk) costs, then, in equilibrium, $\pi(N) = F$. Rearranging this yields

$$(1) \quad N(t) \equiv \theta S(t),$$

where $S(t)$ is the value of sales, $\theta = m/F > 0$ and m is the price cost-margin enjoyed by the average firm. Clearly, θ is lower the more price competitive the market is (i.e. the lower is m) and the higher are endogenous sunk costs, F . A low value of θ means that fewer firms will populate a market of given size $S(t)$ than would otherwise be the case.

Since it is probably unreasonable to assume that all of the observed data on $N(t)$ are equilibrium observations, a more general way to write down this model is to re-express (1) in error correction form as

$$(2) \quad \Delta N(t) = \theta_0 + \theta_1 N(t-1) + \theta_2 S(t-1) + \theta_3 \Delta N(t-1) + \theta_4 \Delta S(t-1) + \varepsilon(t),$$

where Δ is the first difference operator (i.e. $\Delta N(t) \equiv N(t) - N(t-1)$, and so on). The equilibrium number of firms at any time is $N^*(t) = (-\theta_2/\theta_1)S(t)$, and clearly varies directly with market size over time (we expect that $-\theta_2/\theta_1 > 0$).

The basic prediction of this model is that the size of the industry population at any time will be proportional to the size of the market. If markets become increasingly competitive over time (causing m to fall and F to rise as firms cut prices and, say, increase advertising), then the average sales per firm will rise even when total industry sales are constant, and the market will gradually become more concentrated. Firm numbers are likely to fall in this case.

the negative feedback model of entry and exit

The other simple textbook model which gives a prediction about the number of firms operating in a market at any time is the familiar story of entry: high profits attracts entry and entrants bid away these profits, eventually pushing the industry into a long run equilibrium with no excess profits and some number of firms. Similarly, whenever profits are below “normal” levels, exit occurs, and this depopulation of the industry will (sooner or later) raise profitability for the survivors back up to equilibrium (or “normal” levels). At it’s simplest, this feedback loop can be described using a two equation model. The first describes how changes in the number of firms respond to profitability:

$$(3) \quad \Delta N(t) = \alpha \{\pi^e(t) - F\} + \mu^1(t),$$

where $\pi^e(t)$ denotes expectations held at time t about post-entry profits. The assumption of rational expectations implies that $\pi^e(t) = \pi(t) + \mu^2(t)$, where $\pi(t)$ are realized profits and $\mu^2(t)$ is a classical residual. The second equation completes the loop, showing profits depending on cumulative past entry/exit:

$$(4) \quad \pi(t) = \pi^* - \beta N(t-1),$$

where $(\pi^* - \beta)$ is the level of monopoly profits. Putting this all together yields the familiar autoregression in $N(t)$ which is a feature of linear negative feedback processes like (3) and (4):

$$(5) \quad \Delta N(t) = \alpha_0 + \alpha_1 N(t-1) + \varepsilon(t),$$

where $\alpha_0 \equiv \alpha(\pi^* - F) > 0$, $\alpha_1 \equiv -\alpha\beta < 0$ and $\varepsilon(t) \equiv \mu^1(t) + \mu^2(t)$. In the long run, $\pi^* = F$ and $N^* = (-\alpha_0/\alpha_1)$.

There are two remarks worth making about this model. Notice first that as it stands, this model is one in which all motion is transitory: entry and exit cease when the market returns to equilibrium. However, in applying it to actual markets, we have to let that equilibrium vary over time and this will require us to allow for (let us say) secular (or equilibrium) changes in the number of firms. This, of course, introduces a second source of motion into the model. One possibility would be to allow π^* to be driven by market size and growth, an extension which could turn (5) into something very close to (2). Alternately, π^* might be driven by technological progress in the manner of either Klepper, 1996 or Jovanovic/MacDonald, 1994. Given our limited information, however, the closest that we can approach to this second kind of specification is to allow π^* to depend on a simple linear (or quadratic) time trend. The second observation is that if the entry feedback process does not operate (i.e. if either α or $\beta = 0$), then $N(t)$ follows a random walk over time. This might, for example, arise from a series of independent exogenous shocks which increase or decrease demand period by period, increasing or decreasing π^* . A random walk in $N(t)$ might also arise when a series of exogenously created invention opportunities are exploited by new entrant firms.

the contagion model of entry and exit

Entry and exit depend on expectations, but assessing the future is a tricky business. Although agents almost certainly try to anchor their expectations in “*the fundamentals*”, fads, fashions and information cascades often have profound effects on agents’ beliefs. The main characteristic of these phenomena is that they lead agents to act on the actions of others rather than directly basing their choices on their own evaluation of “*the fundamentals*”.⁸ There are several ways to model this phenomena. Perhaps the easiest is to modify (3). Suppose that $\Delta N^e(t) \equiv \alpha\{\pi^e(t) - F\}$ is the amount of

entry or exit that is expected to occur in period t based on an analysis of the fundamentals. However, the actual amount of entry or exit which occurs may exceed or fall short of this, causing a “surprise”, $\Delta N^s(t) \equiv \Delta N(t) - \Delta N^e(t)$. Agents who are willing to allow their beliefs to be affected by factors other than just the fundamentals may be influenced by this “surprise”, thinking that it contains useful information which is not fully taken account of in the evaluation of the fundamentals.⁹ If they respond to such surprises, then (3) becomes

$$(6) \quad \Delta N(t) = \alpha \{ \pi^e(t) - F \} + \psi \Delta N^s(t-1) + \mu^1(t),$$

where $\psi > 0$. Using (4) and repeating most of the calculations which lead to (5) yields

$$(7) \quad \Delta N(t) = \psi_0 + \psi_1 N(t-1) + \psi_2 \Delta N(t-1) + \varepsilon(t),$$

where $\psi_0 \equiv \alpha(1 - \psi)(\pi^* - F) > 0$, $\psi_1 \equiv \alpha\beta(\psi - 1) < 0$ and $\psi_2 \equiv \psi(1 - \alpha\beta) > 0$ if, as we presume, $\alpha\beta < 1$.

It is clear that the inclusion of a surprise term in (4) has no effect on the long run equilibrium number of firms; it does, however, substantially affect short run dynamics. If ψ is large, then any kind of unexpected rise in entry or exit in period t will be magnified over time as agents respond to the surprise by altering their expectations and, therefore, their behaviour. An initial unexpected rise in entry will, for example, inflate entry rates for subsequent years by gradually diminishing amounts. Obviously, if $\psi = 0$, then (7) collapses to (5). Note also that this model produces a nice interpretation of the $\Delta N(t-1)$ term which appears in the error correction model, (2). Finally, notice that if $\psi_1 = 0$, this model cannot be used to solve for an equilibrium number of firms. In this case, the model predicts that random shocks set off speculative bubbles, and that the number of firms present in the market at any time is just an accident: $N^*(t)$ depends the history of shocks to the market, and the strength of reactions to “surprises”.

the density dependence model

One of the more distinctive features of the typical evolution of industry populations is that growth rates typically rise and then fall in the early stages of market development, generating an S-curve. Organizational ecologists have developed an explanation of this phenomena which turns on

two concepts; *legitimation* and *competition*. In the very earliest phases of market evolution, both the various new product variants and the firms who make them are relatively unknown, and only the most daring, risk taking consumers are liable to patronize them. However, as the product (and the organizations which provide it) become more established, and consumers, suppliers and other interested parties get used to it (and come to value it), then (net) entry becomes easier and population growth rates rise. In due course, the market becomes established and, as (net) entry continues at relatively high levels, the market becomes congested. This competition for increasingly scarce consumers and resource suppliers eventually puts a break on further expansion, making (net) entry progressively more difficult and population growth rates fall.¹⁰

It is probably easiest to think of this model as a generalization of (5), which we will rewrite as:

$$(8) \quad \Delta N(t) = \rho_0 + \rho N(t-1) + \varepsilon(t).$$

If $\rho_0 = 0$, then the expected rate of growth of N is ρ , and the essence of the density dependent model is the assertion that population growth rates are affected by population density. Thus, we write

$$(9) \quad \rho = \rho_1 + \rho_2 N(t-1),$$

where $\rho_1 > 0$ but $\rho_2 < 0$. Combining (8) and (9) yields

$$(10) \quad \Delta N(t) = \rho_0 + \rho_1 N(t-1) + \rho_2 N(t-1)^2 + \varepsilon(t).$$

Notice that if $\rho_2 = 0$, (10) reduces to (5); however, in (5), $\alpha_1 < 0$, while in (10) $\rho_1 > 0$ which is a second way to distinguish the two models. It should be noted that this is a much simpler (i.e. more linear) model than is typically used by organizational ecologists, but more complex versions of the density dependent model can usually be approximated in such a way as to introduce an $N(t-1)^3$ term in (10). Needless to say, this provides an easy test of the adequacy of the linear density dependence model.¹¹

Two remarks are worth making about this model. First, (9) can be generalized to allow a whole range of factors to affect growth rates. If, for example, exogenous technology shocks made it

possible for new firms to overcome barriers to entry more easily, then population growth rates would rise. Alternately, changes in barriers to entry or exit are likely to directly affect new firm formation or mortality rates, and so affect the growth of the population. Given our limited information, it is possible to make only limited progress in this direction, mainly by including time trends in (9). Second, this model generates an S-curve, but observing that $N(t)$ follows an S-curve over time does not mean the model is right: the true model might, for example, be (1) and the S-curve in $N(t)$ might arise simply because $S(t)$ follows an S-curve over time. Further, the model generates an S-curve expansion path for firm numbers, but, since $N(t) > 0$ for all t , it cannot account for a fall in firm numbers which occurs over time unless (9) contains other factors which could cause negative population growth rates.

in summary

The four models of population dynamics which we have developed – (2), (5), (7) and (10) – are not exactly nested, but they are related in some rather obvious ways. Take (5) as the benchmark. If we adopt the null hypothesis that population size follows a random walk over time, then (5) is a natural alternative. The argument that contagions amplify unexpected movements in population size generalizes (5) to (7) with the addition of a term, $\Delta N(t-1)$, while the argument that population growth rates are density dependent generalizes (5) in a slightly different way with the addition of $N(t-1)^2$. Finally, the argument that market size drives population density supplements (5) with three terms, of which two, $S(t-1)$ and $\Delta S(t-1)$, distinguish it from the other four models considered here. Put the other way, (2), (7) and (10) all reduce to (5) and hence to the null hypothesis of a random walk with simple exclusion restrictions applied to one or more terms in each equation.

IV. THE POPULATION OF US CAR PRODUCERS

We are going to apply these four models to nearly 100 years of data on the population of domestic US car producers.¹² Figure I shows the basic times series that we will be working with. Most of the interesting dynamics in population size are crammed into the first 30 or so years of data, with a sharp rise in the number of producers (which peaks at around the time of the introduction of the Model-T) followed by a somewhat less precipitous fall. In fact, the last 80 or so years of the data describes a long, protracted phase of consolidation. The challenge in the data is to explain the sharp rise in numbers early in the history of this sector using a model that also accounts for the long and protracted decline in numbers. The interesting question is whether any of the four models we discussed earlier is up to meeting this challenge.

the number of firms

We start with the market size model. Figure II displays the number of passenger cars produced per year over our sample period, and describes a time path which is roughly S-shaped.¹³ However, a comparison between Figures I and II reveals that the market size model is completely inconsistent with the data: the market increased in size only after the consolidation phase was well underway (meaning that market size and population size are negatively correlated in this sector). In fact, the long drop in the number of producers shown on Figure I was mirrored by a steady rise in industry concentration, meaning that market concentration and market size were positively correlated over most of the sample period in this particular industry.¹⁴ This visual impression is confirmed by the econometrics. Co-integration tests, with and without linear and quadratic time trends, and with and without further lagged first differences in $N(t)$ or $S(t)$, always failed to find a statistically significant (much less a positive) co-integration co-efficient. Regression (1) on Table I shows estimates of the error correction model. The $S(t-1)$ and $\Delta S(t-1)$ terms were always insignificant (and were so in every variant of (1) that we tried), and the usual tests revealed that (2) can always be simplified to (7).

Regression (1) readily simplifies to the contagion model (7), but no further; i.e. it appears that the simple negative feedback model and the null hypothesis of a random walk are not consistent with the data. However, before exploring the contagion model a little more fully, it is worth spending some time on the negative feedback model. The comparison between (5) and the null

hypothesis resolves itself into the question of whether a unit root exists, and direct tests of this were slightly ambiguous.¹⁵ Regression (2) on Table I shows a simple estimate of (5), while (3) adds a quadratic time trend which allows π^* to move deterministically over time. The co-efficient on $N(t-1)$ in regression (2) is very small and insignificant, but it is significant in regression (3). Since it is clearly absurd to suppose that π^* is constant over time, regression (3) seems like a more reasonable choice. It predicts that $N^* = 335$ in the first year of the estimation period (1902), and then drops continuously thereafter, a computation which is obviously driven by the time trends. If we use this to help interpret the data, we are driven to the conclusion that the rise in population at the beginning of the period is a transitory adjustment to N^* (meaning that these years were not equilibrium observations). However, from about 1910 onwards, N^* and $N(t)$ are not far apart, meaning that these value of $N(t)$ are, in the main, equilibrium observations.

It is hard to draw very strong conclusions from all of this. As Figure I shows, $N(t-1)$ looks to be a rather better predictor of $N(t)$ than the mean of $N(t)$, and this (plus one of the unit root tests plus regression (2)) inclines one towards accepting the null hypothesis of a random walk. However, the other unit root test, regression (3) and the significance of $\Delta N(t-1)$ in regression (1) are clearly inconsistent with this, and incline one to reject the null. Further, in estimating regressions (2) and (3) over various sub-periods we noticed a tendency for the co-efficient on $N(t-1)$ to fall over time. This, of course, is consistent with the spirit of the density dependence model. What is clear, however, is that taken on it's own, the negative feedback model is inconsistent with the data. Our difficulty in rejecting the null of a random walk means that, at best, the negative feedback loop described by equations (3) – (5) is weak and tenuous. All of the useful action that we get out of regression (3) comes from the time trends which chart movements in π^* over time.¹⁶

The contagion model as we have written it in (7) is a very strong and robust generalisation of the negative feedback model. Regression (4) shows an estimate of it which includes a quadratic time trend, but much the same results emerged in virtually every variant of regression (4) that we ran. The $\Delta N(t-1)$ term is significant and attracts a co-efficient just less than 0.5.¹⁷ If $\psi = 0.5$ and there is a surprise increase in the number of firms of 10 in $t = 0$, then the “echoes” of this surprise over time will be 5, 2.5, 1.25, .625..., which adds up to 10 additional new firms over a period of 4-5 years and doubles the size of the original surprise. Needless to say, a surprise drop in N of 10 later on will reverse this.

The simple linear density dependent model is also obviously inconsistent with the data. The dynamics outlined in equations (8) – (10) describe an S-shaped approach to a steady state number of firms, $N^* = -\rho_1/\rho_2$ (if $\rho_0 = 0$). However, when (10) reaches N^* , it remains there: the simple linear density dependence model does not and cannot predict the long decline in firm numbers which is the major feature of our data.¹⁸ Regression (5) and (6) display estimates of the density dependent model, equation (10), with and without a quadratic time trend. In regression (5), the co-efficients on both $N(t-1)$ and $N(t-1)^2$ are not significant (and they are also “wrong” signed), while in (6) that on $N(t-1)^2$ is insignificant. The inclusion of an $N(t-1)^3$ term in these regressions had no appreciable impact on the estimates, meaning that there is no obvious evidence consistent with non-linear density dependence in the data. Since the $N(t-1)^2$ term is the distinctive feature of the density dependent model and it is not significant in these regressions, one has to conclude that the density dependent model is inconsistent with the data.

Nevertheless, the spirit of the density dependent hypothesis is that population growth rates vary with population density (and, presumably, other things). Regression (7) shows an extension of equation (9) in which ρ is determined by $N(t-1)$ and a time trend, while regression (8) extends this one step further by including a contagion term, $\Delta N(t-1)$. Surprisingly, regression (8) is probably our best description of the data. It does not simplify to any of the other regressions shown on Table I, and it is an acceptable simplification of regressions which also include quadratic time trends, higher order lags in $N(t-1)$ or $\Delta N(t-1)$ and terms in $S(t-1)$ and $\Delta S(t-1)$. It tells a fairly reasonable story. Population growth rates start off large and positive ($\rho \approx .33$) at the beginning of the sample and fall throughout, hovering above and below zero from about 1910 on (but remaining small throughout). This, of course, implies that N^* starts small and then reaches a maximum at around 1910. Much of this movement in ρ is driven by the time trend, and it is hard to claim that there are important traces of density dependence in the data: regression (8) suggests that competitive crowding starts to squeeze population numbers when $N(t)$ rises above 480, which is below the maximum $N(t)$ in the data. Finally, contagion effects exaggerate the short run dynamics in the manner discussed earlier (although these effects seem to be relatively less important in regression (8) than they were in regression (4)). In particular, the contagion effects help to accelerate the decline in producer numbers which begins around 1910, and they subside in importance over time as the effects of the shakeout in firm numbers becomes progressively smaller and more predictable.

the number of entrants and exitors

There is one obvious caveat to all of this, and that arises from the fact that $\Delta N(t)$ is a measure of net entry; that is, it equals entry less exit in period t . Most of the organizational ecology literature has been built up around the hypothesis that density affects entry and exit rates differently. Entry is thought likely to rise and then fall as density rises, while exit rates are expected to fall and then rise with density. Since it is more than conceivable that a whole range of interesting exogenous variables have different effects on entry and exit, it is worth decomposing $\Delta N(t)$ into its two constituent components and replicating what we have done thus far on each.

Unfortunately, our entry and exit data only go up to 1967, and this reduces our sample by 27 observations.¹⁹ Serious as this seems, it has little practical effect on the conclusions which we have drawn thus far: replicating all of the equations displayed on Table I (plus many of the unreported variants which we explored) produced no qualitative differences in the results, and in many cases parameter estimates were almost identical. Figure II displays the number of entrants and the number of exiting car producers over the sub-sample period, and it paints a picture similar to that shown on Figure I, with an initial burst of entry fuelling the large early rise in firm numbers. This is followed by a rather less sharp rise in exit, which occurs more gradually over a longer period of time than the wave of entry which preceded it. Entry and exit are positively correlated (.6086), and correlated with $\Delta N(t)$ by .223 and -.239 respectively. Both are highly positively correlated with firm numbers (.882 and .831 respectively).

The question is whether one is led to significantly different conclusions if one applies the four models discussed earlier to entry and exit rates separately. Consider first the entry equations (which explain the number of entrants who arrive each year). As before, the error correction model found absolutely no support in the data, with no significant correlations between market size or growth and entry being evident. The linear density dependence model also commanded no obvious support from the data: the co-efficients on $N(t-1)$ and $N^2(t-1)$ were significant but negative and positive respectively (which are the "wrong" signs), and, as before, their effects were dwarfed by those of the time trends. Much the same was true of the exit equations: the error correction model never generated significant co-efficients on $S(t-1)$ or $\Delta S(t-1)$; in the linear density dependence model, both estimated co-efficients were positive (the first should have been negative) and that on $N^2(t)$ was never significant. More positively, it was possible to reject the null hypothesis of a unit root in both

the entry and the exit equations, and both display clear auto-regressive patterns. Entry, but not exit, is positively correlated with $\Delta N(t-1)$, suggesting that contagion might be a possible explanation for the steep rises and falls in entry but not the more gradual wave of exit. This seems plausible, since the depopulation of this sector was played out over several decades.

in summary

Movements in the population of US car producers is not a random walk, and it is also not associated with current or near lagged movements in either market size or the size of the population of US car producers at any time. Further, transitory dynamics associated with negative feedback are, at best, an unimportant component of the total movement in population size which we observe. Our estimates indicate that there are two dominant drivers of $N(t)$: time trends (which we have interpreted as tracking movements in the equilibrium number of firms) and recent past “surprise” changes in the population (which we have interpreted as reflecting a speculative bubble). Looking once again at Figure I, this all seems reasonable: the early rise in numbers has all the hallmarks of a bubble, while the long secular decline seems likely to be a consequence of deep seated trends in cost and demand. Finally, entry and exit counts display slightly different patterns of variation over time, but there is enough common movement in both series to insure that movements in either are driven by basically the same forces as drive movements in the population as a whole.

V. CONCLUSIONS

In this paper, we have made a case for studying the population dynamics of particular industries, outlined four rather simple models which might be used in such an exploration and then applied these models to data taken from the US car industry. Although the number of producers is not the only interesting phenomena one might study in the context of industry evolution, the fact that population movements seem to be systematic and (broadly speaking) similar across many industries makes them worthy of attention. However, if the US car industry is at all typical, it seems clear that we will have to do rather more than exhume simple textbook models and apply them to the data. Neither the market size model nor the negative feedback model seem to provide any kind of a satisfactory account of the evolution of the US car industry, and the linear density dependence model also seems to be inconsistent with the data. Although it has turned out to be possible to contrive a model somewhat in the spirit of the density dependent model which seems to fit the data reasonably well, the sad conclusion is that the one robust and reliable model that we have examined seems to be the contagion model. This leaves one with the uncomfortable conclusion that much of the interesting dynamics in the data might be the consequence of bandwagon responses to “surprise” population movements.

There is clearly much to be done, both with this data well as on data from other industries. We suspect that economists will tire of running regressions on net firm numbers (or birth/death rates) sooner than organizational ecologists have, and our guess is that there is one feature of the data on cars which will prove to be quite robust and may lead to the development of much more satisfactory models. Essentially, the main interesting (i.e. slightly counter intuitive) feature of the data which we have been examining is that population numbers began to drop as the market expanded. This occurred partly through liquidations (of firms unable to compete with Ford, General Motors, Chrysler and the other early market leaders) but also partly through mergers and take-overs (this is particularly clear in the history of General Motors), meaning that the drop in firms numbers was caused by a rise in concentration which, one conjectures, was made possible by the growth and development of the market (allowing victorious firms to exploit economies of scale and scope). This pattern of a fall in firm numbers (and a rise in concentration) occurring against the backdrop of industry expansion is easy to discern in other sectors, and it seems to be part of a process by which some (but definitely not all) firms adapt more quickly than others to the new opportunities opened up by market growth. Thus, to explain the *selection* process which population dynamics reflect, one may

need to look more closely at differential *adaptation* rates. That is, the way forward may turn out to be a closer integration of selection models like density dependence with models of adaptation and pre-emption that have been a feature of the economics literature on strategic competition.

REFERENCES

- Banerjee, A. (1992) "A Simple Model of herd Behaviour", Quarterly Journal of Economics, 107, 797-817.
- Bikchandani, S. D. Hirschleifer and I. Welch (1998) "Learning from the Behavior of Others", Journal of Economic Perspectives, 12, 151-170.
- Bikchandani, S., D. Hirschleifer and I. Welch (1992) "A Theory of Fads, Fashion, Custom and Cultural Change as Information Cascades", Journal of Political Economy, 100, 992-1026.
- Camerer, C. and D. Lovo (1999) "Over Confidence and Excess Entry", American Economic Review, 89, 306-318
- Carroll G. and M. Hannan (eds) (1995) Organizations in Industry, Oxford University Press, Oxford.
- Carroll, G. and M. Hannan (1989) "Density Delay in the Evolution of Organizational Populations", Administrative Science Quarterly 34, 411-430.
- Carroll, G. and M. Hannan (1992) Dynamics of Organizational Populations, Oxford University Press, Oxford.
- Choi, J. (1997) "Herd Behavior, the Penguin Effect and the Suppression of Informational Diffusion", Rand Journal of Economics, 28, 407-425.
- Christensen, C. (1997) The Innovator's Dilemma, Harvard Business School Press, Boston, Mass.
- Gort, M. and S. Klepper (1982) "Time Paths in the Diffusion of Product Innovations", Economic Journal, 92, 630-653.
- Hannan, M. and J. Freeman (1984) "Structural Inertia and Organizational Change", American Sociological Review, 49, 149-164.
- Hannan, M. and J. Freeman (1989) Organizational Ecology, Harvard University Press, Cambridge, Mass.
- Hofbauer, J. and K. Sigmund (1988) The Theory of Evolution and Dynamical Systems, Cambridge University Press, Cambridge.
- Jovanovic, B. and G. MacDonald (1994a) "Competitive Diffusion", Journal of Political Economy, 102, 24-52.
- Jovanovic, B. and G. MacDonald (1994b) "The Life Cycle of a Competitive Industry", Journal of Political Economy, 102, 322-347.
- Kapur, S. (1995) "Technology Diffusion with Social Learning", Journal of Industrial Economics, 43, 173-195.
- Klepper, S. (1996) "Entry, Exit, Growth and Innovation over the Product Life Cycle", American Economic Review, 86, 562-583.

- Klepper, S. (1997) "Industry Life Cycles", Industrial and Corporate Change, 6 145-182.
- Klepper, S. and K. Simons (1997) "Technological Extinctions of Industrial Firms: An Inquiry into the Nature and Causes", Industrial and Corporate Change, 6, 379-460.
- Klepper, S. and K. Simons (1999) "Industry Shakeouts and Technological Change", mimeo, Carnegie Mellon University.
- Malerba, F. R. Nelson, L. Orsenigo and S. Winter (1999) "History Friendly Models of Industry Evolution", Industrial and Corporate Change, 8, 3-40.
- Mazzucato, M. and W. Semmler (1999) "Market Share Instability and Stock Price Volatility during the Industry Life Cycle: The US Automobile Industry", Journal of Evolutionary Economics, 9, 67-96.
- Nelson, R. (1995) "Recent Evolutionary Theorizing about Economic Change", Journal of Economic Literature, 33, 48-90.
- Pesaran, H. (1987) The Limits to Rational Expectations, Basil Blackwell, Oxford.
- Roughgarden, J. (1996) Theory of Population Genetics and Population Ecology, Prentice Hall, Upper Saddle River, NJ.
- Scherer, M. and T. Ross (1990) Industrial Market Structure and Economic Performance, 3ed edition, Houghton Mifflin, Boston.
- Silverberg, G., G. Dosi and L. Orsenigo (1988) "Innovation, Diversity and Diffusion: A Self Organizing Model", Economic Journal, 98, 1032 - 1054.
- Sutton, J. (1991) Sunk Costs and Market Structure, MIT Press, Cambridge, Mass.
- Sutton, J. (1998) Technology and Market Structure, MIT Press, Cambridge, Mass.
- Swaminathan, A. and G/. Carroll (1995) "Beer Brewing", in Carroll, G. and M. Hannan (eds), 1995.
- Utterback, J. (1994) Mastering the Dynamics of Innovation, Harvard Business School Press, Boston, Mass

TABLE 1: REGRESSIONS EXPLAINING CHANGES IN THE NUMBER OF US CAR PRODUCERS*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CNST	38.3 (2.88)	-.363 (.227)	52.9 (4.13)	46.5 (4.10)	-.93 (.52)	69.1 (4.27)	4.4 (2.91)	3.13 (2.08)
N (t-1)	-.126 (3.31)	-.016 (.908)	-.158 (4.00)	-.149 (4.30)	.028 (.440)	-.32 (2.95)	.508 (6.93)	.432 (5.84)
N (t-1) ²	-	-	-	-	-.0002 (.720)	-	-.009 (4.29)	-.0009 (4.49)
)N (t-1)	.42 (4.92)	-	-	.442 (5.19)	-	-	-	.27 (3.10)
)S (t-1)	.029 (.32)	-	-	-	-	-	-	-
S (t-1)	.034 (.512)		-	-	-	-	-	-
t	-1.19 (2.79)		-1.66 (4.27)	-1.39 (4.01)	-	-2.04 (4.52)	-	-
t ²	.008 (2.75)		.012 (4.20)	.01 (3.80)	-	.014 (4.51)	-	-
t*(t-1)							-.019 (8.30)	-.015 (5.65)
LLM	-342.2	-368.2	-359.5	-347.0	-367.9	-358.1	-345.9	-341.1
R ²	.305	.009	.178	.371	.014	.201	.444	.50

*N^(t) =number of car producers in t(mean = 49), and S, (t),=,normalized (to mean = 51) sales;
the dependent variable is)N (t-1); absolute values of t-statistics are given in brackets; all estimates of
standard errors are heterosceolastic – consistent.

NOTES

¹ However, when the size distribution of firms is skewed, some firms are likely to have much more power in the market than others, and a simple count will no longer be quite so informative. Further, observing the size of gross entry or exit flows from the market may also not be terribly informative if most entrants fail before they challenge large incumbents, or if they permanently occupy niche markets.

² For empirical work generally supporting (or at least documenting) these assertions, see Gort and Klepper, 1982, Utterback, 1994, Klepper, 1997, and others.

³ Christensen, 1997, discusses (and documents) some of the sources of “*incumbent inertia*” (i.e. the vulnerability of incumbents to innovative entrants), and contains references to the large managerial literature on this subject. In quite a different vein, Hannan and Freeman, 1984, and 1989, argue that organizations exist to insure the reliable replication and accountability of certain activities, and that successful firms are selected on that basis. The ability to reliably replicate a particular activity and the demands of accountability (which typically include the need for consensus and some degree of democratic decision making) are likely to make organizations more bureaucratic and less flexible or able to adapt to changed market circumstances.

⁴ See, for example, Klepper and Simons, 1997 and 1999, who examine the causes of shakeouts in four US industries.

⁵ In addition to the two papers discussed in the text, see Malerba et al, 1999, Jovanovic and MacDonald, 1994a, Silverberg et al, 1988, and others; Nelson, 1995, surveys some of this literature.

⁶ One obvious difference between the models of Klepper and Jovanovic/MacDonald and those which we will explore below is that the former place much emphasis on the evolution of differences between firms. Amongst other things, this difference makes these models better suited to explaining changes in industry concentration over time.

⁷ See Scherer and Ross, 1990, for a summary of much of the early work done in this area; see Sutton, 1991 and 1998, for recent work.

⁸ See Bikhchandani et al, 1992 and 1998, Bannerjee, 1992, Kapur, 1995, Choi, 1997, and others. The “surprise” model developed below is well known in the macroeconomics literature; see, for example, the survey of rational expectations models by Pesaran, 1987.

⁹ Camerer and Lovo, 1999, report a number of experiments which suggest that excess entry in many industries may be explained by the over confidence of agents who are unduly sanguine about their abilities, or the opportunities which they face.

¹⁰ Hannan and Freeman, 1989, is a classic exposition of the area of organizational ecology; Hofbauer and Sigmund, 1988, and Roughgarden, 1996, contain expositions of the population models used by ecologists; Hannon and Carroll, 1992, outlines the “legitimation and competition” argument summarized in the text, and their 1995 volume gives a number of applications (and further references to the large literature stimulated by their 1992 book).

¹¹ Most of the work done by organizational ecologists focuses on explaining entry and mortality rates separately (and not net movements in the population as we are doing here). Unfortunately, their models do not often aggregate easily into a simple prediction about movements in the population as a whole; see, for example, the work reported in Carroll and Hannan, 1992, Chapter 8.

¹² Data sources: The total number of passenger cars produced between 1900-1995 were taken from the 1996 edition of Wards Automotive Yearbook, published by Wards Communications. The number of US domestic car producers for the whole period and the number of entrants and exits between 1895 and 1967 were kindly given to us by Steven Klepper and Ken Simons. The population data runs from 1895-1995, but the first six years of data are missing and that plus the need to allow for up to two lags chops the first 8 observations off the sample. This data is very similar to that used by Mazzucato

and Semmler, 1999, and Klepper and Simons, 1997. The sample used by Carroll and Hannan (in Carroll and Hannan, 1995) appears to have many fringe and niche producers and also includes foreign producers in the US. Nevertheless, the over all pattern of population variation shown by their data seems qualitatively similar to that shown on Figure I below.

¹³ Regressing $S(t)$ on a constant, $S(t-1)$ and $S(t-1)^2$ yields parameter estimates (t-values) of: 1.144 (9.144) and -.0021 (1.77), with the constant = 1.52 (.579) and an $R^2 = .886$. We also used a measure of the value of car production and the same deflated by the consumer price index without reaching qualitatively different conclusions the logistic pattern of $S(t)$ over time, or about the usefulness of the market size model.

¹⁴ This may be a very general pattern: it has been observed in the US automobile tire industry (Jovanovic and MacDonald, 1994b), beer (Swaminathan and Carroll, 1995), television (Klepper and Simons, 1997) and other sectors. These correlations are consistent with the models of industry evolution of both Jovanovic/MacDonald and Klepper, but are hard to reconcile with the (basically cross section) predictions of Sutton, 1991, on the relationship between concentration and market size.

¹⁵ We used the weighted symmetric tau and the Dickey-Fuller tests, with and without a constant and time trend. Using the former, we were unable to reject the null hypothesis of a unit root, but the D-F test pointed to rejection of the null. The $S(t)$ series, by contrast, displayed clear evidence of a unit root.

¹⁶ The evolution of this sector has been characterized by extensive process innovation, the intensive drive to create and exploit scale economies in production, logistics and sourcing, and the exploitation of economies of scope (through extensive product ranges). In principle, developments of this type are likely to lead to fewer but larger establishments grouped in a smaller number of enterprises, a pattern which is consistent with the steady decline in the number of car producers which we see in our data. For a stimulating discussion of product and process innovation in this sector, see Klepper and Simons, 1997.

¹⁷ Generalizing regressions (2) or (3) by adding in a term $N(t-2)$, always yielded the same conclusion: the coefficient on $N(t-1)$ was positive, significant and equal in size but opposite in sign from the (also significant) coefficient on $N(t-2)$. This is, of course, consistent with the contagion model.

¹⁸ Needless to say, this is not necessarily true with the more complex non-linear density dependent models which organizational ecologists sometimes use. However, it is still sometimes difficult to account for the decline in population size using non-linear models, and Carroll and Hannan, 1989, have suggested that “*density delay*” (i.e. density conditions at the time of an organizations’s birth) may help to account for the decline from peak numbers.

¹⁹ It also eliminates from our sample the period in which Japanese and European car producers made major inroads into the share of US domestic car producers in the US car market, which makes it a doubly interesting sub-sample.