FUNDAMENTALS OR POPULATION DYNAMICS AND THE GEOGRAPHIC DISTRIBUTION OF U.S. BIOTECHNOLOGY ENTERPRISES, 1976-1989

Lynne G. Zucker
Michael R. Darby
Yusheng Peng

NBER WORKING PAPER SERIES

FUNDAMENTALS OR POPULATION
DYNAMICS AND THE GEOGRAPHIC
DISTRIBUTION OF U.S. BIOTECHNOLOGY
ENTERPRISES, 1976-1989

Lynne G. Zucker
Michael R. Darby
Yusheng Peng

ABSTRACT

Population ecology models are elegant in form and adequate in describing aggregate data, but poor in telling stories and predicting the location of growth. Fundamentals models emphasizing the variables central to resource mobilization, such as intellectual human capital, can predict where and when biotechnology enterprises emerge and agglomerate. Density dependence and previous founding dependence proxy many underlying processes; the legitimation and competition interpretation is more conjectural than empirically tenable. We argue and demonstrate for biotechnology that an alternative model based on the fundamentals related to resource reallocation and mobilization provides a stronger frame to explore industry formation. Fundamentals models outperform population ecology models in the estimations, while a combined model driven by fundamentals but incorporating weak population dynamics does best. In repeated dynamic simulations, the population ecology model predictions are essentially uncorrelated with the panel data on biotechnology entry by year and region while the combined model has correlation coefficients averaging above 0.8.

Lynne G. Zucker
Organizational Research Program
Institute for Social Science Research
University of California, Los Angeles
Los Angeles, CA 90095-1551
and NBER
zucker@nicco.sscnet.ucla.edu

Yusheng Peng
Department of Sociology
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
CHINA
yushengpeng@cuhk.edu.hk

Michael R. Darby
Anderson Graduate School of Management
University of California, Los Angeles
Los Angeles, CA 90095-1481
and NBER
michael.r.darby@anderson.ucla.edu

Fundamentals or Population Dynamics and the Geographic Distribution of U.S. Biotechnology Enterprises, 1976-1989

Since Michael Hannan and John Freeman published their first article on population ecology of organizations in *American Journal of Sociology* in 1977, their theory of internal population dynamics and density dependence has generated more empirical research than any other single perspective in organization theory. Time series data of labor unions, trade associations, newspapers, banks, life insurance companies and so on have been collected and analyzed using their framework. The vital rates and growth trajectory of each of these organizational populations are elegantly modeled by internal dynamics and reproduced by lagged density and lagged founding and/or mortality rates.

As useful as the population ecology model has been in describing and reproducing a population growth trajectory, we argue that it is simply post-factual description of data. The collection of time-series data implies that growth and concentration of organizations have already occurred. A central question is not asked: Why does the process occur in particular geographic locations? Using panel data in this article that allows us to build on previous work, we demonstrate a simple--but so far ignored--fact that the internal population dynamics central to the population ecology model cannot locate the geographic regions where a particular kind of organization is likely to grow and, over time, concentrate in numbers. We argue that this limitation is rooted in the conceptual weakness of the basic population ecology model: Lack of substantive variables and substantive stories. In contrast, we show that a model based on fundamentals can locate the geographic regions where growth and concentration are likely to occur over time by

identifying the substantive building blocks of the new organizations (see Zucker and Darby 1995; Zucker and Darby 1996b; Zucker, Darby, and Armstrong 1998; Zucker, Darby, and Brewer 1998).

The Population Dynamics Model

A fairly often observed growth trajectory of organization populations looks like this: the total number of organizations (such as automobile manufacturers, newspapers, labor unions, banks) grows slowly at the beginning, followed by explosive growth with stabilization at some ceiling, and then gradually declines. Organizational ecology models exploit this regularity by focusing on the endogenous population dynamics and predict organizational founding rates and death rates on the basis of contemporaneous organizational density. The argument is that the founding rate and death rate of organizations are quadratic functions of the existing stocks. Initial rises in density increase the founding rate and decrease the mortality rate; at high levels, rises in density decrease founding and increase the mortality rate (Hannan and Freeman 1987, 1988, 1989; Carroll and Hannan 1989a; Hannan and Carroll 1992; Hannan et al. 1995).

Theoretically, density dependence is interpreted as reflecting two opposing underlying processes: legitimation and competition. At the early evolutionary stage of an organizational population, density is hypothesized to reflect legitimation, i.e., the process of becoming routine and taken-for-granted. As population size grows large, however, competition for resources and markets sets in and is expected to overshadow the effects of legitimation. By using one observed variable to measure two underlying processes, Carroll and Hannan (1989a) claimed to have integrated institutional and ecological

2

theories of organization.

A similar argument is made for lagged founding and failing rates. An initial increase in last year's founding rate signals economic opportunities for entrepreneurs. But too big an increase in last year's founding rate suggests crowding and competition and thus deters new entry. An initial increase in last year's failing rate signals newly emptied organizational niches and thus invites new entry. But a large number of deaths signals a hostile environment (Delacroix and Carroll 1983; Carroll and Huo 1986).

Typical ecological birth models involve five variables: lagged stock, lagged stock squared, lagged birth, lagged birth squared, and sometimes period effects.[1] The reasoning of ecological models hinges on an inverted U-shape function and predicts a positive effect for the first-order term and a negative effect for the second-order term. For example, in models of labor union birth at the national level the authors include the number of national labor unions in the prior year and its square, the number of births in the prior year and its square, plus period effects (Hannan and Freeman 1987; Hannan and Freeman 1989:201-224).

Since its inception, organizational ecology has generated considerable momentum. This is partly because density data are easily available and partly because the model offers a kind of formal elegance. The ecological model is elegant because it is parsimonious: Using one key concept and a few parameters, it can describe fairly well the evolutionary process of many, if not all, organizational populations. Hannan and Carroll (1992) were able to dynamically simulate the population growth process by a few parameters, assuming that the organization foundings follow a Poisson or negative binomial process.

3

Conceptual and Empirical Issues in Population Dynamics

The population ecology approach has frequently received criticisms on both conceptual and empirical grounds. Conceptually, Zucker (1989) pointed out that the population ecology model does not directly measure either legitimation or competition. Rather, legitimation and competition are post-factual interpretations of population density dependence. Density estimates are proxies for other effects (see also Miner 1993). If density is an indirect indicator of legitimation, not an intervening process that relates density to founding and failure as Carroll and Hannan argue (1992:69), then a direct measure of institutional embeddedness (relations between the organizations and potential supporters in the environment) should account for density dependence, which is in fact what Baum and Oliver (1992) find in their study of day care centers. Since institutional embeddedness in the case of non-profit organizations is also a measure of access to environmental resources (funding, credentialing, and so on), "it is unlikely that increasing legitimacy accounts for more than a modest initial increase in the rate of founding in most organizational populations" (Zucker 1989:544).

Empirical support for the model has been somewhat uneven: About 75% of the studies of founding support the basic ecological model of birth, with even less support for the death model (Baum and Powell 1995; details on confirmation status are in Baum 1996:Table 4). Especially, population ecology models do not predict population declines well, as admitted by population ecologists (Hannan et al. 1995). Baum and Powell (1995) note that the density dependence model is ahistorical because organizational forms and dynamics of growth change greatly from early to late periods. Low population density at

4

a late evolutionary stage has a very different meaning, and involves a very different growth dynamic, from low density at an early evolutionary stage.

Most of these criticisms have been taken lightly by the proponents of population ecology, who regard them as quibbles around a basically successful paradigm (see Hannan and Carroll 1995; Carroll and Hannan 1989b). What we will do in the remainder of this paper is sketch an alternative theory of fundamentals based on location of resources and their mobilization or reallocation, identify areas in which fundamentals are expected to make predictions superior to population dynamics and density dependence, and then proceed with an empirical comparison that provides convincing evidence of the superior performance of the fundamentals model. We will demonstrate empirically that the population dynamics/density model makes significant, even striking, errors in prediction, especially in determining geographic location.

Fundamentals and Founding or Transforming Organizations

The fundamentals model selects substantive variables that identify the presence of resources needed to found or transform an organization. To start a new firm--or indeed to transform an existing firm--fundamental building-block resources are needed. These resources are often already "in use" by other persons or organizations (or subunits in the case of transformation) in which case the resources must be removed from the old use and redeployed.[2]

The human agency required to remove resources from one use and reallocate them to another use has already been recognized in resource mobilization theory (McCarthy and Zald 1977), a theory useful for understanding the initiation and growth of firms as well as

the non-profit sector. Resource mobilization involves real costs: Human time, energy, and financial resources are redirected, with possible increase in risk and certainly the loss of activities (or investments) foregone. For resource mobilization to take place, the benefits of the reallocation must be expected to outweigh the costs associated with it--at least for the persons taking the actions (Zucker and Kreft 1994; Zucker, Darby, Brewer, and Peng 1996).[3]

The central idea in the fundamentals model is one of relative resource importance and relative scarcity.[4] At the extreme, the value of a given resource is so high that the firm can not be founded or transformed without it; it is necessary. As that necessary resource increases in scarcity, whether because of prior decisions about investments in creating it (e.g., top research universities not located nearby) or because of relatively high estimated costs of mobilizing it (e.g., cost of personnel change to implement new technologies in incumbent organizations), the number of foundings and transformations will decrease.

An implication is that uneven distribution of a central and scarce resource, for example across different geographic areas, will produce a pattern of high density of foundings and transformations of firms where the resource is relatively plentiful and a low density or even absence of foundings and transformations of firms where the resource is relatively scarce. At the same time, the more valuable the resource, the greater the benefit of moving it; thus, in order for the geographic distribution of new foundings and transformations to remain uneven over time, the costs incurred in moving the resource must outweigh the benefits of moving it. Other resources, in addition to "necessary" resource(s) or in combination with each other, contribute to the probability of founding a

new firm or transforming an existing firm. Absent the ability to move many or all of these resources, an initial geographic differential in resource base will largely determine when and where the new foundings and transformations occur and hence where the new industry or industries will be located.

But the presence of these resources alone is not sufficient to determine the founding and transformation process: Institutions alter the ability of a person or persons to mobilize resources. Institutions constrain resource mobilization decisions by increasing the costs for reallocating a given resource or conversely facilitate resource mobilization decisions by, for example, subsidizing costs for reallocating a given resource. Institutional constraints affect the probability that resource mobilization will actually produce net positive returns; we focus here on constraints that are external to the firm such as whether there is a national market for human capital and financial capital, or balkanized local markets. If there are constraints on labor mobility, for example, the probability of founding a successful firm or successfully transforming an existing firm may differ significantly from one region to the next, whether the constraints are state-ordered as in China or driven by concerns of maintaining research productivity as in top scientists with productive labs (Peng, Zucker, and Darby 1997; Zucker, Darby, and Brewer 1998).[5]

As the resource reallocation process proceeds, the degree to which availability of each resource alters the count of new and transformed firms can be assessed. Institutional effects on the founding/transformation process can also be identified, though when institutional variation is low within a society such a test may require a cross-cultural, international design; we report such a design elsewhere in comparisons between the U.S.

7

and Japan (Darby and Zucker 1996; Zucker and Darby 1998).

The fundamentals model provides a more fruitful way of studying the vital rates of organizational populations than focusing exclusively on endogenous population dynamics as we will demonstrate empirically below. The effects of the lagged foundings variable only indicate that new foundings in this year are influenced by the same factors as in the past year. Density dependence mostly reflects geographic agglomeration produced by resources that are often geographically localized such as human capital, financial capital, and the product market.[6] Population ecology models tend to obscure the importance of these autocorrelated fundamental factors by misattributing their effects to internal population dynamics asserted to represent (unobserved, unmeasured) legitimacy and competition.

Predictive Power of Population Dynamics and Fundamentals: Hypotheses

Since endogenous population dynamics are essentially self-referential, population ecology models can only **describe** post-factually the growth trajectory of organizations but cannot **predict** when and where the growth and agglomeration will occur. Population ecology models cannot predict the spatial distribution of organizations. Where geographic agglomeration will occur has to be explained by location-specific exogenous variables.

Lack of geographical boundaries and location is one of the weaknesses in population ecology models pointed to by Lomi and Larsen (1996). Density dependence models simply state that agglomeration occurs where there is a large number of organizations. Therefore we expect that the basic population ecology model does not work well with panel data to predict where growth of the organizational population and

thus local agglomeration will occur because it does not have substantive variables to pinpoint the locations where growth and agglomeration are most likely.

As long as the fundamentals are well specified and measured, there will tend to be a reasonably strong correlation between the fundamentals and actual changes in the population dynamics, i.e. lagged density and lagged foundings. It is therefore reasonable to conceptualize population ecology variables as proxies of the underlying fundamentals. There can be some argument for the population dynamics approach simply based on significantly cheaper data collection: Think of the construction of right-side variables based on the left side numbers rather than collection of independent right-side variables. However, the difficulty for population ecology is exactly that "advantage": If fundamentals are collected and entered into the equation, then density and founding variables lose at least some of their impact. Generally, as autocorrelation in the series decreases, population dynamics perform increasingly less well as predictors, compared to the fundamentals. The common use of period effects in population ecology models is one "work-around" for this basic problem in the underlying theory. We can now state our first hypothesis:

> *Hypothesis 1:* Models that rely on the fundamentals will produce improved estimations of **when** foundings or transformations of firms will occur, since population dynamics are the result of autocorrelated fundamentals. We expect that this improvement will take two forms:
>
> > (a) Predicted timing will improve with fundamentals compared to population dynamics.

(b) Effects of population dynamics will be reduced in the presence

of the fundamentals.

We expect these differences generally to be small because estimated population dynamics can mimic the direct outcome of the effects of the fundamentals. We competitively test the models in a series of poisson regressions using data on the growth of the U.S. biotechnology industry.

But there is a more serious problem facing the population ecology paradigm: It expects starting or transforming a firm is equally likely across all geographic units that don't yet have one. Only where enough observations accumulate to identify the areas of substantial concentration will the population ecology models begin to predict those agglomerations will continue. The basic population ecology model can be expected to consistently undershoot the actual level of agglomeration in particular geographic locations. Providing initial data on the actual geographic distribution can be expected to improve the model only slightly, since each iteration will move the estimation further away from the actual agglomeration that tends to accumulate over time.

We can now state our second major hypothesis that has guided our empirical work:

Hypothesis 2: Models that rely on the fundamentals will produce improved estimations of **where** foundings or transformations of firms will occur, since population dynamics assume the same underlying processes are operating across all geographic subareas.

10

Differences between estimated and actual geographic distribution are expected to increase in population ecology estimates as the actual distribution of firms becomes more geographically uneven, even when initial data on location of firms is provided. In sharp contrast, the fundamentals model estimates of geographic location are expected to be close to the actual distribution, since the fundamentals drive the actual distribution of firms. We competitively test the models by conducting 100 simulations, and then compute three measures of fit between the estimated and actual data on the birth and geographic distribution of firms in the U.S. biotechnology industry: the total count of live firms, the standard deviations across regions, and the correlations across regions.

Birth of New Industries: The Case of U.S. Biotechnology

From the mid-1970s to the end of the 80s, the biotechnology industry grew in the U.S. from nonexistent to about seven hundred active firms. This rapid growth was based on a major breakthrough discovery in bioscience, the 1972 discovery by Stanford professor Stanley Cohen and University of California-San Francisco professor Herbert Boyer of the basic technique for recombinant DNA (reported in Cohen, Chang, Boyer, and Helling 1973). Today biotechnology refers principally to the application of genetic engineering based upon taking a gene from one organism and implanting it in another and production of the outcome of this process.[7]

Star Scientists: The Necessary and Scarce Resource in Biotech

The necessary and scarce resource in biotechnology is knowledge about the major gene splicing discovery and the discoveries that have cascaded from it, cataloged in a major scientific data base called GenBank. Especially initially, only a few "star" scientists

who had been involved in making the discoveries had access to this important intellectual human capital, because the discoveries were tacit and high in natural excludability, that is the degree to which other scientists could be excluded from making use of them (Zucker, Darby, and Brewer 1998).[8] High natural excludability explains both why discovering star scientists might be motivated to become involved with firms--they have the opportunity to earn extraordinary returns to their human capital--and why firms might be motivated to work with star scientists--these discovering scientists have very rare tacit knowledge.

We measure the degree of natural excludability by the extent to which scientists who have not worked in the genetic sequence area before can enter on their own or need to learn form doing bench-level science with a more experienced investigator. To quantify these largely invisible working relationships we use a co-publishing measure to provide a novel analysis of the diffusion of scientific knowledge: we measure scientific collaboration as who publishes together, and then collect related data on characteristics of the scientific research team.[9] We track each scientist who ever has a published article reporting a genetic sequence discovery listed in GenBank (1994): 84,461 individual authors entered GenBank from 1969 through 1993, publishing 66,070 articles, excluding unpublished sources and patents.

We operationalize a measure of natural excludability by distinguishing between "new" authors (those who have never published before in GenBank) and "old" authors (those whose publications have appeared previously in GenBank). If most new scientists are publishing with a least one old author, it suggests that natural excludability is high; if most new entrants to GenBank can do the research either by him/herself or with all new

authors, it suggests that natural excludability is low. Breakthrough discoveries seldom have only one type of co-publishing, so the issue is the balance between the two.[10] As we can see in Figure 1, new authors enter GenBank predominantly by publishing with old authors, with this mode accounting for 81 percent of entry from 1969 through 1992.[11] Excluding sole-authored articles, which may be dissertations for new authors and review articles by established authors, new authors write exclusively with other new authors 36 percent less frequently than old authors write exclusively with other old authors.[12]

[Figure 1 about here]

Local Effects of Stars: Institutional Policies and Geographic Mobility

At least in the case of biotechnology, few star scientists who knew how to do recombinant DNA were willing to abandon their university appointments and labs to pursue commercial applications of biotechnology. Thus, labor which we usually think of as mobile became a fixed factor and led to geographic concentration of the new industry around the universities with top bioscience departments.

In high technology industries, universities are important institutions, and their policies regarding the kinds of work that faculty can do outside the university will significantly shape the impact of that faculty on nearby industry. The primary pattern in the development of biotechnology involved a scientist-entrepreneur who remained on the faculty while establishing a business on the side--businesses which, where successful, resulted in millions or even billions of dollars for the professors who acquired early ownership stakes. Most universities in the U.S. increased local industrial growth by allowing faculty both to retain appointments at the university and to pursue private

13

commercial interests. We have shown elsewhere that the most common and most important source of intellectual capital for industry is the part-time, "linked" faculty member (Zucker, Darby, and Armstrong 1998; Zucker, Darby, and Torero 1997).[13]

Other Resources and Institutional Constraints

In addition to recombinant DNA, other techniques have played an important role in commercial biotechnology. Therefore, we develop two additional measures of intellectual human capital: the total number of top ranking university bioscience departments in the region and total number of faculty members receiving federal grants in the region.

Indicators of other aspects of the resource flow include the cost of capital (measured by the earnings-price ratio) and the general level of economic activity in a region (total employment). We have introduced several other measures of resources and institutional structure; we did try entering these other variables in the analyses reported here, and for simplicity kept the analyses as they are since the other variables were not significant (e.g., quality of labor in the region which is colinear with total employment).

Institutions may help or block resource reallocation, helping it in the case of national capital markets that can move capital efficiently to its best use or hindering it in the case of balkanized capital markets. To the extent that venture capital is a local resource, as implicit in the argument that venture capitalists want control and therefore want to invest close by, it will be deployed effectively only to the extent that the best investments happen to lie nearby. If venture capital is a national or international resource, then the capital is free to flow to the investments that are most likely to yield high returns.

14

The Data

Data has been collected in panel form for 14 years 1976-1989 and 183 regions [functional economic areas as defined by the U.S. Bureau of Economic Analysis (BEA) 1992b]. Frequently, the data are aggregated up to the region from the zip code or county level.[14] Lagged variables include data for 1975 in the unlagged form. See the Data Appendix for more details.

Firms: Our data set on firms was derived from a database purchased from the North Carolina Biotechnology Center (NCBC, 1992) which was cleaned and supplemented with information in *Bioscan* (1989-1993) and in Cetus Corp. (1988). We identified 752 distinct U.S. firms for which we could determine a zip code and a date of initial use of biotechnology. By the end of 1989, 22 of the 752 firms had died or merged into other firms.

Star Scientists: We identified a set of 327 star scientists based on their outstanding productivity through early 1990. The primary criterion for selection was the discovery of more than 40 genetic sequences as reported in GenBank (1990) through April 1990. However, 22 scientists were included based on writing 20 or more articles each reporting one or more genetic sequence discoveries.[15] In the 1990s sequence discovery has become routinized and is no longer such a useful measure of research success. These 327 stars were only 3/4 of one percent of the authors in GenBank (1990) but accounted for 17.3 percent of the published articles, almost 22 times as many articles as the average scientist.

We hand-collected the 4,061 articles authored by stars and listed in GenBank and recorded the institutional affiliation of the stars and their coauthors on each of these

articles, a project that took several years to complete (Zucker, Brewer, Oliver, and

Liebeskind 1993). Coauthors are called "**collaborators**" if they are not themselves a star.

Some data on the stars and collaborators who ever published in the U.S. is given on the

left side of Table 1 where the scientists are identified by the organization(s) with which

they were affiliated on their first-such publication. The higher citation rate for firm-

affiliated scientists is explored at length in Zucker and Darby (1996b).

Figure 2 illustrates the time pattern of growth in the numbers of stars and

collaborators who have ever published and the total number of firms using biotechnology

in the United States. There was a handful of stars who published articles reporting genetic

sequence discoveries before the 1973 breakthrough, but even after 1973 their number

increased gradually until taking off in 1980. The numbers of collaborators and firms

lagged behind the growth in stars by some years.

To identify those scientists clearly working at the edge of the science in a given

year, we term a star or collaborator as "**active**" if he or she has published three or more

sequence-discovery articles in the three-year moving window ending with that year. As

seen in the right side of Table 1, this stringent second screen provides an even more elite

definition of star scientists as well as identifying some very significant collaborators. We

count for each year the number of active stars and active collaborators who are affiliated

with an organization in each region.

[Figures 2 and 3 about here]

The locations of active stars and firms are both concentrated and highly correlated

geographically, particularly early in the period. Figure 3 illustrates this pattern for the

whole period by accumulating the number of stars who have ever been active in each region up to 1990 and plotting them together with the location of biotech-using firms as of early 1990.

Other Measures of Intellectual Human Capital. Some skeptical readers might also think that some simpler measures of regions' relevant academic resources would contain all the information which we have laboriously collected. We found two measures of regional scientific base which entered separately in regressions reported below, but none which were capable of eliminating the effects of the star scientists. One measure is a count of the number of "**top-quality universities**" in a region where top-quality is defined by having one or more "**biotech-relevant**" (biochemistry, cellular/molecular biology, and microbiology) departments with scholarly-quality reputational ratings of 4.0 or higher in the 1982 National Research Council survey (Jones, Lindzey, and Coggeshall 1982). There are 20 such universities in the United States.[16] Our second measure "**federal support**" is the total number (in hundreds) of faculty supported by 1979-80 federal grants to all universities in each region for biotech-relevant research.[17] These variables take on the same value for a given region in each year.

Other Variables. Using listings in Stanley E. Pratt (1982), we measure "**venture capital firms**" as the number of such firms in a region legally eligible to finance start-ups in each year up to 1981. For later years, the number of firms is fixed at the number in 1981 to avoid possible simultaneity problems once the major wave of biotech founding began.[18] (While great bookstores spring up around great universities, the former should not be counted as causing the latter.) Since entry of biotech firms would be expected to

occur where there is other economic activity, we also include **total employment** in all industries (in millions of persons) for each region and year. Finally, an increase in the (all-equity) cost of capital, as measured by the **earnings-price ratio** on the Standard & Poors 500 index would reduce the net present value of entry and so should have a negative impact on the birth of new firms or transformation of incumbent firms.

Empirical Analysis

Birth or Entry Model

In the early population ecology literature, hazard rate methods were used to model founding data (e.g., Carroll and Hannan 1989a), probably because event history analysis was fashionable at the time. But the appropriateness of applying hazard rate model to organizational founding data is dubious because it is not clear who or what is at risk of giving births to organizations (Zucker 1989). Later population ecologists realized that birth counts of organizations in a specified time interval and a particular region should be more appropriately modeled as a Poisson process or negative binomial process (e.g., Hannan and Carroll 1992; Hannan et el. 1995).[19]

We assume that births of new biotech enterprises in a region in a given year follow a Poisson process with a mean λ_{it}, which can be expressed as a multiplicative function of a set of explanatory variables. The logarithm of λ_{it} is a linear function of the explanatory variables. This process is consistent with the dependent variable which is a non-negative integer with significant mass at zero. In our estimates, we utilize the Wooldridge regression-based correction for the variance-covariance matrix estimates.[20] In functional form, the probability of y_{it} new biotech enterprises emerging in a BEA in a given year is

$$\Pr\{Y_{it} = y_{it}\} = \frac{e^{-\lambda_{it}} \lambda_{it}^{y_{it}}}{y_{it}!}$$

with $\quad \lambda_{it} = e^{x_{it}\alpha}$

in which x_{it} is a vector of covariates, either population dynamic variables or economic fundamental variables or both. Parameter vector α can be solved for via maximum likelihood method.

The **population ecology model** consists of foundings in the previous year and its square term, local stocks of biotech firms and its square term, national stocks in the previous year, and finally period effect measured by year. The squared national stocks, which is usually included in population ecology models, are dropped out of the model because its effect is nearly zero and insignificant. We include both local and national stocks in the regression because we believe that whereas national stocks may indicate legitimation and competition for markets, local density may indicate agglomeration effect. Our **fundamentals model** includes the crucial and scarce resource, number of active stars, as well as their active collaborators, plus their square terms, and other economic fundamental variables such as number of venture capital firms, log employment size, and earnings/price ratio. We also include the interactions between period dummy variable for 1986-89 and active stars and collaborators to capture the declining importance of intellectual capital due to the diffusion of rDNA techniques. **A mixed model or**

19

augmented fundamentals model that combines variables from the fundamentals and population dynamics models is also estimated. All the models are estimated using the LIMDEP package and results are presented in Tables 2a and 2b, with the Wooldridge correction.

In order to facilitate comparison between non-nested models, Bayesian Information Coefficients are calculated in addition to log likelihood ratio chi-squares (L^2). As developed by Raftery (1986),

$$BIC = L^2 - (df)ln(N),$$

where L^2 stands for the model chi-squares; "df" for degrees of freedom; and N for sample size. Models with negative BIC are preferable to the saturated model (raw data). A large negative BIC indicates a good model.

[Tables 2a and 2b about here]

Interpreting population ecology model: All the parameter estimates of the population ecology model come out with correct signs (Table 2a). The birth of new biotechnology enterprises in a BEA has an inverted U-shaped relationship with last year's founding rate and last year's local stock. The effect of local density can not be interpreted as legitimation since legitimacy is hardly localized in this case. Rather, we interpret it as the effect of geographic agglomeration. The concave curve suggests the existence of competition for locally available resources, such as intellectual capital and skilled labor etc. The national stock, which is supposed to reflect legitimation and competition for market,

20

enters negative and its square term positive and insignificant. If we believe in population ecology, then the negative impact of national total count would suggest that no legitimation process took place, with national competition emerging very early in the formative stage of U.S. biotechnology industry.

Interpreting the fundamentals model: The fundamentals model emphasizes the most important and scarce resource, intellectual human capital, operationalized as the role played by star scientists and their collaborators. Prior to 1986, star scientists played the main role in founding new biotech enterprises. The collaborators played insignificant roles. The situation reversed after the mid-80s, as indicated by the interaction effects with the period 1986-89. Star scientists became less important in commercialization after 1986, because the scarcity of their tacit knowledge declined rapidly with wide diffusion, even though the tacitness remained (refer to as Figure 1)

Other measures of intellectual human capital, top-ranking bioscience departments in universities and numbers of faculty receiving federal grants in the region, also enter significantly. Venture capital appears to be efficiently allocated nationally, rather than locally balkanized as is often assumed. The negative coefficient is due to the effects of two cities, New York and Chicago, that have many venture capital firms and relatively few biotechnology enterprises.

Interpreting the Mixed Model: The mixed model combines the population dynamics variables with the fundamentals variables. Comparing the regression coefficients of this model with the pure population ecology model reveals how much of the internal population dynamics actually reflect the effects of intellectual human capital and other

economic fundamentals, and vice versa. Controlling for the fundamentals variables has not changed the signs of lagged births, lagged local density, and their squared terms, but has greatly attenuated their magnitudes. Most notably, lagged density becomes insignificant. This suggests that the population dynamics variables are capturing the effects of intellectual human capital and other fundamentals variables. In other words, the effects of lagged density are eliminated and a large portion of the effects of lagged foundings are explained away by the fundamentals. There is also some effect of the population dynamics variables on the coefficients of the fundamentals models, but most of these disappear when the period effects are removed from the model in Table 2b. However, the effects of top ranking university bioscience departments is reduced, and both the squared term for active star scientists and the earnings/price ratio lose their significance.

To help visualize the reduction in the effects of population dynamics, Figures 4 and 5 present the density dependence function and the founding rate dependence function before and after controlling for the variables drawn from the fundamentals model. The functional forms of both lagged density and lagged foundings have inverse U-shape, as population ecologists predicted. After controlling for the fundamentals and population dynamics (mixed model), the curves are greatly flattened, especially for the density dependence function. The reduction in the magnitude of the U indicates that the lagged density dependence function in Figure 5 especially, and to a dramatic but still somewhat lesser extent the lagged founding rate dependence function in Figure 4, in large part reflect the geographic agglomeration of the fundamentals, especially intellectual human capital.

22

Even if the flattened curve can be attributed completely to the unmeasured effects of legitimation and competition, the magnitude of the effect is not substantively significant enough to be worth of the focus on theorizing about them in the population ecology literature.

[Figure 4 and 5 about here]

Comparing population dynamics and the fundamentals models: For the full data set (1976-89), the Chi-square of intellectual capital-centered fundamentals model is smaller than that of the population ecology model, indicating that the former is a better fitting model (Table 2a). But as the two models are not nested, statistical testing for the comparison is not possible. Because the fundamentals model uses twice as many parameters as the population dynamics model, sheer Chi-square is not a fair criterion for comparison. The Bayesian Information Coefficient adjusts for degrees of freedom by penalizing the excessive use of parameters. The BIC comparison indicates which model is more likely to be correct given the data. According to BIC, the mixed model is clearly the most preferable of the three. The fundamentals model, with a somewhat smaller BIC, seems to be slightly better than the ecology model.   Table 2b presents the same analyses with data restricted to the early period (1976-85). It tells basically the same story.

Death or Exit Model

Between 1976 and 1989 there were only 22 deaths (including 7 mergers and repossessions) out of 752 births. Because of the limited number of time intervals, we choose to model death as a simple discrete-time event (Allison 1982). Let p_t be the probability of dying in time t given that the firm has survived to time t, then

$$p_t = \frac{e^{x_t \alpha}}{1 + e^{x_t \alpha}}$$

In terms of the log odds of death versus surviving to next year, the equation becomes

$$\log\left(\frac{p_t}{1 - p_t}\right) = x_t \alpha$$

where age of the firm (with age at birth year as zero), age squared, year (measuring trend effect) and other covariates are included. The two models are estimated in a logistic regression procedure and the results are presented in Table 3. Because of the small number of deaths, only age, age squared, and year have a significant impact on mortality rate. Lagged founding and density variables and broadly delimited regional variables are insignificant (equation II, Table 3). With four more parameters in equation (II), the gain in Chi-square over equation (I) is minimal. Therefore, we selected the simple death model (equation I, Table 3) to use in the simulation with both the population dynamics model and the augmented or mixed fundamentals model. The significant positive coefficient on age and negative coefficient on age squared indicate that the mortality rate increases (at a decreasing rate) up to about age 7 and thereafter declines as the firm ages. However, generally, the mortality rate shows an increasing trend over the years.[21]

24

[Table 3 about here]

Dynamic Simulation

Dynamic simulation in population ecology literature refers to a computer-generated process aimed at duplicating the real process of population growth. It consists of at least two random processes: a Poisson birth process and a binomial dying process, the parameters of which are dictated by models estimated from data. It is called dynamic because the odds of birth or death depend upon the results of the previous year's trials.

One of the major attractions of population ecology model is the claim that it can reproduce, via computer simulation, population growth trajectories that closely resemble the actual historical trajectories of populations under research (Hannan and Carroll 1992, p.185). It is our contention, however, that dynamic simulation works well for population dynamics models only if we ignore geographic heterogeneity by dealing with single geographic region, or one region at a time with multiple regions, so that spatial distribution is not a question. The population dynamics model implicitly assumes geographic homogeneity because it does not contain any variables that identify variation in localized resources that may be available and competed for. Thus, the chances of growth in one location are as good as any other location. Only the pre-given and unexplained initial foundings or entry indicate where intensive activities are more likely to follow. The population ecology empirical literature therefore has focused on the **process** of geographic concentration of organizations. The question is not asked why geographic concentration has occurred in this particular location rather than others. Actually, the question is assumed away in the selection of locations.

With the panel data of regions in the U.S. by year, we address the following question: Where is the U.S. biotechnology industry likely to grow and concentrate? From the foregoing discussion of the fundamentals model and related analysis our answer is clear: Where the potential founders or initiators (star scientists and their collaborators) are located, where other intellectual human capital sources are available, and where institutions are supportive. In this section, we will compare the population dynamics model and augmented fundamentals model in dynamic simulation, examining the predictions of each model for the geographical distribution of biotechnology enterprises. Poisson birth and binomial death will be simulated simultaneously for each and every geographic region, assuming nil interactions among different regions except for the effect of national density.

We conducted three sets of simulations. The first set simulates the population dynamics birth model (equation I, Table 2a), starting with random Poisson births in 1976. Lagged birth and density (1975) are set to zero according to the data. Each and every new-born firm is subject to an individual random binomial dying process (equation I, Table 3) until it exits or survives through 1989. Lagged birth and density are updated each year according to birth and death counts in the previous years. In this simulation, the birth process is determined by internal dynamics, i.e., lagged birth and density, with trend as the only exogenous parameter. There is no exogenous factor that would differentiate some region as more likely to give birth to biotechnology enterprises than others. Montana or Arkansas, for example, would have the same probability of amassing a high density of biotech firms as California. Therefore, we hypothesize that the geographical distribution

26

and concentration of firms after 14 years of simulation by population ecology models would be a purely random process and have little resemblance to the actual data.

The second set of simulations is the same as the first set except that foundings in 1976 are taken from the data rather than randomly generated. There are 17 births recorded in 1976. The location of these initial births should provide some information to the population dynamics model as to where the agglomeration of biotechnology firms is likely to occur 14 years later. Our question is how much adding these initial foundings to the simulation improve the predictions of the population dynamics model. But we refrain from providing exogenous founding and density data after the first year even though it would obviously enhance the simulation results, because doing so would render the simulation non-dynamic and vitiate the attractiveness and cogency of the population ecology model and its associated claims.

The third set simulates the augmented or mixed fundamentals model (equation III, Table 2a). We elect to simulate the augmented fundamentals model rather than the pure fundamentals model because simulating the latter, not including any dynamic variables, would be a Markov process. We expected that the addition of variables based on the fundamentals variables would influence the time and location of initial foundings as well as steer the agglomeration process to probable locations. We expected the fundamentals model to produce a much more accurate geographical distribution of firms by the end of the 14 year period than does the dynamics/ecological model.

We performed 100 simulations for each of the three sets identified above. We present the results both longitudinally and cross-sectionally. Longitudinally, simulated

national growth trajectories alongside with actual data are plotted in Figure 6, panels a-c. In each case, 5 simulations are randomly selected out of the 100 simulations for each set to be used in the plot. The two population ecological models (with and without initial data) and the augmented fundamentals model generate reasonable national growth trajectories. Both ecological models slightly underestimate the total count for later years. Starting with randomly generated initial foundings (Figure 6a) or with actual initial birth data (Figure 6b) makes very little difference in the national growth trajectories.[22]

[Figure 6][Table 4]

While the population ecology model can generate reasonable national growth trajectory, its simulation can not produce a geographic distribution that remotely resembles the actual data. Table 4 compares the summary statistics of total counts, standard deviations, and correlation coefficients of actual and simulated cross-region data by the end of 1989. First, the population ecological model systematically under-produces the national total number of live firms, with an average total count from 100 simulations equal to about 687, falling short of actual national total (730) by 5%; it does not do much better with the initial data added. In contrast, the augmented fundamentals model generates a national total much closer to data. Second, the population ecology model underestimates spatial concentration as indicated by standard deviations, with average standard deviation from 100 simulations equal to about 5.1, less than half the size of the standard deviation over the actual data (11.8). In other words, ecological models predict a geographical distribution of biotech industry that is much more homogenous and much less agglomerated than it is in reality. The augmented or mixed fundamentals model,

28

however, produces a very realistic spatial dispersion of 11.2 compared to the actual of 11.8 (column 6, Table 4).

Finally, and most important, the geographic distribution of firms produced by the population ecology model does not match the actual data on geographic distribution at all, with an average correlation coefficient close to zero with the actual data on firm distribution (column 7, Table 4). This result is expected because without variables identified by the fundamentals model that differentiate one region from another as more favorable to industrial growth, the population ecology model randomly allocates firms to different regions. Even if we provide the initial founding data to the population ecology model, we get an average correlation coefficient of 0.148 with the actual data on firm location (column 8, Table 4). This is especially poor given the fact that the actual across-region correlation coefficient is 0.598 between the beginning (1976) year's founding and ending (1989) year's firm counts. Lacking substantive variables that identify the underlying mechanisms of industry origin and growth, internal dynamics of the population actually attenuate the effect of agglomeration. The augmented fundamentals model, on the other hand, can reproduce a geographic distribution of firms that corresponds closely to actual data, with an average correlation coefficient of 0.839 with the actual data on firm location (column 9, Table 4).

Conclusions and Implications

Population ecology models are elegant in form and adequate in describing data, but poor in telling stories, inadequate in predicting the location of growth. In the foregoing

analysis we advance two critiques of population dynamics and density dependence as developed in the population ecology theory and operationalized in empirical research, and at the same time develop a competing theory which we show to be a superior model of population processes in a number of different respects.

First, we argue that density dependence and previous founding dependence are proxy to many underlying processes; the legitimation and competition interpretation is more conjectural than empirically tenable. Second, we argue that an alternative model based on the fundamentals related to resource reallocation and mobilization provides a stronger frame within which to explore the empirical processes involved in industry formation and decline. In the case of most high technology industries, including the U.S. biotechnology industry, geographically localized intellectual human capital and institutional constraints on the mobilization of that capital are the forces underlying the internal population dynamics. Controlling for these variables greatly reduces the effects of lagged density and lagged foundings. To be fair, there is also some evidence that controlling for population dynamics reduces some of the effects of the variables measuring fundamentals, though not those of central importance. Population ecology misattributes the effects of economic fundamentals to density dependence and misinterprets density dependence as legitimation and competition.

The fundamentals model is not only more robust empirically, it also offers the advantage of integration with recent work in economics and in economic geography on the processes of industry growth and change, and the relationship of these processes to societal productivity. Recently, attention has been refocused on knowledge spillovers as

causes of both economic agglomeration and growth (Grossman and Helpman 1991, Romer 1986 and 1990). Zucker, Darby, and Armstrong (1998) demonstrate that in the case of biotechnology industry such localized knowledge "spillover" is actually a return to intellectual capital. Intellectual human capital provides a specific channel through which knowledge spillovers are embodied and returns to them captured.

Second, because of its self-referential nature, models based on population dynamics and density dependence are not able to predict where new organizations are likely to emerge and agglomerate. Only by adding substantive variables that measure the variables central to the resource mobilization process, such as intellectual human capital and other fundamentals that pinpoint the potential founders and favorable economic conditions, are we able to predict where biotechnology enterprises will emerge and agglomerate.

We conclude that much more will be gained in our understanding of organizational growth by examining the substantively meaningful structural variables, the variables that are central to resource mobilization, than if we continue to focus on internal growth dynamics. The internal population dynamics of organizations itself needs to be understood in light of well-specified and measured models of resource mobilization, including the actions of historically specific actors, and the related institutional processes that either encourage or inhibit the resource mobilization process.

Appendix A: Data

In this paper we used the data sets developed for the Project on "Intellectual Capital, Technology Transfer, and the Organization of Leading-Edge Industries: The Case of Biotechnology," Lynne G. Zucker, Marilynn B. Brewer, and Michael R. Darby, Principal Investigators. These data will be archived upon completion of the project in the Data Archives at the UCLA Institute for Social Science Research. The project has been housed at and supported in part by ISSR.

The data sets contain a variety of linkages, but are derived from distinct sources and are most easily discussed sequentially. Before doing so, however, it is important to understand a few salient features of the data sets.

First the data generally are in the form of panels: cross sections observed annually from 1975 or earlier through 1989 or later. The cross-sections are generally defined geographically. For this purpose, geocoding was originally done at the finest level possible of the zip-code, county, BEA- area, and national levels of geography. For the analyses reported in this paper, all data for finer levels of geography have been aggregated to the 183 BEA-defined functional economic areas in the United States. Every U.S. zip code and county is assigned to one of these areas.

Generally, natural geography exists only for institutions (universities, research institutes and hospitals, biotechnology firms, and venture capital firms), but a variety of economic data is collected at or can be aggregated to BEA level as well. Individual scientists are linked to locations through the institutional affiliations reported in their

publications in the article data set. The citations were also collected for articles if and only if they appeared in the article data set; so scientists are credited with citations only insofar as they are to the star scientists' 4,061 articles reporting gene sequence discoveries and published in major journals.

A.1. Scientist Data Set

As mentioned above, in the earlier stages of the project, GenBank was used to determine for each scientist the total number of genetic sequence discoveries reported in all articles reporting gene sequence discoveries up to 1990 for which that scientist was listed as an author.[23] On the basis of these total number counts, we identified 305 leading researchers whom we termed "stars." We used a cutoff of at least 41 sequence discoveries. However, total gene sequence discoveries is an imperfect measure of scientific output: many technological breakthroughs in the 1970s and 1980s made gene sequencing more efficient, and the discovery of sequences for some types of genes is either more difficult or more important. Accordingly, another 22 stars were identified based on frequency of publication of articles reporting gene sequences bringing the total to 327 stars.

Because research discoveries frequently occurred through teams, scientists who were coauthors with a star but who had not themselves met the star criteria were labeled "**collaborators**." In terms of distinct individuals, we captured data for 6,082 collaborators in addition to the 327 stars, or a total of 6,409 individuals worldwide. Each star and collaborator was assigned a unique scientist ID number for use in linking to the other data

files.

As discussed in Data Appendix A.2 below, we have the affiliations listed by each of these scientists on 4,061 journal articles in our article data set. We use the institutional affiliations to locate scientists: Generally a scientist is located for our analysis at the institution given in his or her last publication; however, scientists are said to be "born" in the year and at the institution of their first publication.

As discussed in the text, a star or collaborator is defined as "**active**" in any year in which he or she has published three or more articles in the three-year period ending with that year. The variables $ACSTAR_{it}$ and $ACCOLL_{it}$ are counts of the number of active stars or collaborators by year in each BEA area.

A.2. Article Data Set

Our article data set consists of all 4,061 articles in major journals listed in GenBank as reporting genetic sequence discoveries for which one or more of our 327 stars were listed as authors.[24] All of these articles were assigned unique article ID numbers and collected by hand. For each article, scientist ID numbers are used to identify the order of authorship and the institutional affiliation and location for each author on each article. This hand coding was necessary because, under the authorship traditions for these fields, the head of the laboratory who is often the most prestigious author frequently appears last. Our stars, for example, were first authors on 18.3 percent of the articles and last authors on 69.1 percent of the 4,031 articles remaining after excluding the 30 sole-authored articles.[25] Unfortunately, only first author affiliations are available in machine-

34

readable sources.

The resulting authorship data file contains 19,346 observations, approximately 5 authors for each of the 4,061 published articles. Each authorship observation gives the article ID number, the order of authorship, the scientist ID number of one of our stars and collaborators, and an institutional ID number for the authors affiliation which links him or her to a particular institution with a known zip code as of the publication date of the article.

A.3. Citation Data Set

We have collected data for 1982, 1987, and 1992, on the total number of citations to each of our 4,061 published articles listed in the Institute for Scientific Information's *Science Citation Index*. These citation counts are linked to the article and authorship data set by the article ID number.

A.4. University Data Set

Our university data set consists of all U.S. institutions listed as granting the Ph.D. degree in any field in the Higher Education General Information Survey (HEGIS), Institutional Characteristics, 1983-84.[26] Each university is assigned an institutional ID number, a university flag, and located by zip code based on the HEGIS address file.

Additional information was collected for those universities granting the Ph.D. degree in biochemistry, cellular/molecular biology, and/or microbiology which we define as **"biotech-relevant"** fields. All of the following additional variables are based on data in

the National Academy of Sciences study by Jones, Lindzey, and Coggeshall (1982).

FEDGRANT$_i$ is the total number of scientists in all biotech-relevant departments in BEA area i supported by 1979-80 grants from the NSF, NIH and the Alcohol, Drug Abuse and Mental Health Administration (ADAMHA).

R&DEXP$_i$ is the sum in thousands of dollars of total reported university expenditures for research and development in the biotech-relevant departments over all the universities listed in Jones et al. which are located in the BEA area. We note that reporting practices for these expenditures seemed inconsistent across universities, perhaps because they do not keep accounts as would have been required to accurately answer the survey.

We define university quality level based on the scholarly quality rating in the reputational survey in Jones et al. Reputational ratings were based on responses from approximately fifteen percent of the faculty in the fields studied. Since we were interested in identifying the very best programs, we considered only the highest rated of the biochemistry, cellular/molecular biology, and/or microbiology programs offered by a particular university. The number of universities in a BEA with one or more most highly rated programs (rated above 4) is our variable QUAL1.[27] Similarly, QUAL2, QUAL3, and QUAL4 represent the numbers of universities with rated programs of above 3 through 4, of above 2 through 3, and of 2 and below, respectively. That is, QUAL1 is the count of the very best universities in an area while QUAL4 is a count of those of the lowest quality to be rated at all.

A.5. Research Institute and Hospitals Data Set

For those U.S. research institutions and hospitals listed as affiliations in the article data set, we assigned an institutional ID number and an institute/hospital flag, and obtained an address including a zip code as required for geocoding. No additional information has been collected on these institutions.

A.6. Biotechnology Firm Data Set

The starting point for our firm data set covered the industry as of April 1990 and was purchased from the North Carolina Biotechnology Center (1992), a private firm which tracks the industry. This data set identified 1075 firms; some of which were duplicates or foreign and others of which had died or merged. Further, there were a significant number of missing firms which had died earlier. For these reasons, an intensive effort was made to supplement the NCBC data with information from *Bioscan* (1989-1993) and an industry data set provided by a firm in the industry which was also the ancestor of the *Bioscan* data set (Cetus Corp. 1988). Each of the firms was assigned an institutional ID number and an enterprise flag.

We combined these three sources to identify 752 distinct U.S. firms for which we could determine a zip code and a date of founding (or entry into biotechnology for subunits of preexisting firms). Of these, based on the financial information in the NCBC data supplemented where necessary and possible by our other sources, 512 are classified as entrants, 150 as incumbents, and another 90 could not be classified clearly into either

subcategory but were retained in the total biotech enterprises. Of the 90 "others," 18 were in the problematic category of joint ventures and the remaining had insufficient data to classify as independent firm or subunit. We also obtained for 52 of these 752 firms the date of their exit by death or merger through 1992.

Based on these data, we have developed a continuous series on the number of active new biotech enterprises NBE_{it} (and its subcomponents entrants NBF_{it} and incumbents NBS_{it}) and their births ($BNBE_{it}$, $BNBF_{it}$, and $BNBS_{it}$), all series being by year and BEA area.[28] The national stock of new biotech enterprises NBE_t is simply the annual aggregate of the BEA area values.

A.7. Venture-Capital Firm Data Set

We created a venture-capital firm data set by extracting from the Stanley Pratt (1982) directory the name, type, location, year of founding, and interest in funding biotech firms. This information was extracted for all venture capital firms which were legally permitted to finance start-ups. This latter requirement eliminated a number of firms which are chartered under government programs targeted at small and minority businesses.

We developed two alternative measures of the availability of venture capital: $VENTCAP_{it}$ which is the total number of eligible venture capital firms in BEA area i in year t and $VCBIO_{it}$ which counts only the subset of those firms indicating a specific interest in their Pratt listing in funding biotech firms. This approach accounts for the founding date of firms appearing in the 1982 Pratt directory, but not for those firms that may have either entered thereafter or existed in earlier years but exited before the

38

Directory was compiled.

A.8. Economic Data Set

This data set consist of three variables EMP_{it}, $EJOB_{it}$, and POP_{it} observed at the BEA area and year level, one variable $SIC2830_i$ which varies across BEA areas but not over time, and three national variables $UNEMP_t$, $E/PRATIO_t$, and $YEAR_t$ which vary only with the year.

Total employment EMP_{it}, average earnings per job $EJOB_{it}$, and total population POP_{it} are all reported by the BEA at the BEA area level from county level data in U.S. Department of Commerce (1992b):

EMP_{it} Table K, line 010

$EJOB_{it}$ Table V, line 290 (wage & salary disbursements, other labor income, and

proprietors income per job in thousands of current dollars) deflated by the

implicit price deflator for personal consumption expenditures.[29]

POP_{it} Table B, line 020

$SIC2830_i$ is the number in the BEA in 1980 of establishments producing goods with SIC code 2830 (Drugs, Total; Number of Establishments). This variable is aggregated to the BEA area level from county level data in U.S. Department of Commerce, Bureau of the Census, (1982).

The total unemployment rate $UNEMP_t$ and the S&P500 earnings-price ratio $E/PRATIO_t$ were taken from CITIBASE (1993), series LHUEM and FSEXP, respectively. The time trend $YEAR_t$ takes the value of 1976, 1977, etc.

Notes

1. The foregoing is not meant to imply the complete absence of substantive variables in empirical research done under the rubric of population ecology. Several innovators working within the paradigm have incorporated one or two local structural variables that they have found to improve their models (e.g., the political context as operationalized by Carroll and Huo 1986, Carroll and Delacroix 1982, and Delacroix and Carroll 1983), and occasionally the explanatory power of these substantive variables are directly compared to the core ecological variables (see the operationalization of the degree of organizational legitimacy in Baum and Oliver 1992). Other variables (e.g., organizational mass) have also been examined, but the focus has remained on lagged births and stock of firms (see Delacroix and Carroll 1983; Carroll and Huo 1986; Carroll 1987; Delacroix and Solt 1988; Tucker, Singh, and Meinhard 1990; Singh, Tucker, and Meinhard 1991).

2. There may not be a net effect of moving the resources. Starting a new firm, though it may appear to offer new jobs, may effectively be a reallocation of workers from one type of employment to another. To the extent that the reallocation leads to a change in local geographic concentration, then some localities may benefit from net gains while other localities suffer net losses. Empirical evidence of a general so-called "multiplier" effect is hard to find. However, if the reallocation involves moving labor and/or financial capital from a less productive to a *more* productive use, then there will be a net increase in productivity; if to a *less* productive use, a net decrease in productivity.

3. Central to the theory, but not varied in the research reported below, is the social identity of those making the resource mobilization decisions. Which resources are mobilized depend on decisions made by the person or persons making them. Reallocation decisions are based on personal assessment of expected value relative to cost and thus are strongly influenced by the identity of **who** can mobilize resources, with professional--or class--values altering the expected return from mobilizing particular resources and thus the decisions themselves. DiMaggio documents this for the Boston Brahmins' involvement in establishing high culture in Boston (1982a and b; 1988). Not examined in this paper, relative success of different firms can be used to assess the quality of the resource mobilization decisions made, and hence of the different strategies the firms adopted (see Zucker, Darby, and Armstrong 1998; Zucker and Darby 1998).

4. Scarcity is not simply the current presence or absence of a resource, but rather is a joint function of two underlying processes: (1) Effects of human agency operating through prior decisions about resource investments in particular assets, such as universities, and particular relationships, such as employing consultants in a certain scientific fields (Zucker and Darby 1996a); and (2) Estimated cost of reallocating the resources, including losses from diminished or abandoned prior use and the risks associated with their new use, often requiring both experts to make these assessments and leaders willing to make decisions under uncertainty about reallocation (on these points, see case study evidence in Zucker and Darby forthcoming).

41

5. Internal constraints include internal organizational convention ("organizational culture") that determines who has decision-making authority, for example giving power to reallocate resources to a person with a distinguished scientific background in one pharmaceutical company, while in another giving such decision-making power to a person trained in financial management (Zucker and Darby forthcoming).

6. In our model the local/national context is determined by how mobile the resources are and how wide the competitive field. In the case of biotech, we find no evidence that more local number of venture capital firms increases firm births (to the contrary they tend to reduce them through packaging larger firms); the point estimates are consistent with national competitive effects coexisting with continuing local agglomeration although these effects are not statistically significant (Zucker, Darby, and Brewer 1998).

7. The other basic technology is cell fusion (also termed monoclonal antibodies, MABs, or hybridomas) in which lymphocytes are fused with myeloma cells to create rapidly proliferating antibody-producing cells (see Sindelar 1992 and 1993 for more detail).

8. Scientists also took informal actions, especially when the information is highly valuable (measured as highly cited by other scientific articles), to slow its diffusion (Zucker, Darby, Brewer, and Peng 1996).

9. We build here on a novel empirical measure we developed in earlier research: "co-publishing," examining all scientists who publish together, to measure who the stars are

working with at the bench science level and which organizations are involved in the collaboration (by obtaining the organizational affiliation of all scientists). We have previously used our measure to examine reciprocal productivity effects of star scientists working with scientists in firms, effects of organizational boundaries as information envelopes slowing diffusion of scientific knowledge, and size and geography of scientific networks used by firms (Zucker, Darby, and Armstrong 1998; Zucker and Darby 1996b; Zucker, Darby, Brewer, and Peng 1996; Liebeskind et al. 1996; Zucker et al. 1993).

10. Obviously there can be bench-science transfers of tacit knowledge which result in no co-publication record, for example, when a student learns the technique as a research assistant without earning coauthorship or when a dissertation is published solo even where the chairman might in other circumstances have shared in authorship credit. Therefore, if new authors predominantly co-publish with old authors the case is strong for a significant degree of natural excludability.

11. Reports of publications for 1993 were incomplete in February 1994 so that year has been excluded from the figure and these calculations. In the incomplete reports for 1993, entry with old authors amounted to 83 percent of total entry.

12. Sole-authored articles account for only 6.5 percent of the authorships of new authors and 7.8 percent of the authorships of old authors over this period. Interestingly, new sole authors become more frequent later in the period as the value of the tacit knowledge declined as it

43

became more widespread (see also, Zucker, Darby, Brewer, and Peng 1996).

13. Other policies alter incentives to invent or to patent what is invented, by setting the percentage of royalties from patents paid to the professor who invents, or alter the probability that the patent will be licensed by a firm by establishing different policies including rules on exclusive licensing agreeements and conflict of interest policies that alter the rate and character of faculty participation.

14. The Bureau of Economic Analysis's (BEA's) functional economic areas divide all the counties in the United States into regions including one or more cities, their suburbs, and the rural counties most closely tied to the central city.

15. Scientists advised that some sequence discoveries are more difficult than others and thus merit an article reporting only one sequence. Therefore we included scientists with 20 or more discovery articles to avoid excluding scientists who specialized in more difficult problems.

16. The twenty universities were: Brandeis University, California Institute of Technology, Columbia University, Cornell University, Duke University, Harvard University, Johns Hopkins University, Massachusetts Institute of Technology, Rockefeller University, Stanford University, University of California-Berkeley, University of California-Los Angeles, University of California-San Diego, University of California-San Francisco, University of Chicago, University of Colorado at Denver, University of Pennsylvania, University of Washington (Seattle), University of Wisconsin-Madison, Yale University.

17. We also tried a measure of biotech-relevant research expenditures as reported by the universities, but this variable was too collinear with the federal support variable to enter separately and appeared to be less consistently measured across universities.

18. Instrumental variables would provide a more elegant approach to this problem if suitable instruments had been found.

19. The Poisson process is a special case of negative binomial process when the variance of the random variable equals its mean.

20. As discussed in Jerry Hausman, Bronwyn H. Hall, and Zvi Griliches (1984), the poisson process is the most appropriate statistical model for count data such as ours. In practice, overdispersion (possibly due to unobserved heterogeneity) frequently occurs. Given the problems with resort to the negative binomial (A. Colin Cameron and Pravin K. Trivedi 1990), Jeffrey M. Wooldridge (1991) developed a flexible and consistent method for correcting the poisson variance-covariance matrix estimates regardless of the underlying relationship between the mean and variance. We are indebted to Wooldridge and William Greene for advice in implementing the procedure in LIMDEP (Greene 1992, pp. 539-549).

21. Firm mortality does increase after 1990, but that is out of the data range considered here; due to the difficulty of locating each star and collaborator publishing, additional data collection needed to add years to our series was not feasible.

22. Note that these simulations produce quite smooth trajectories in comparison with single time-series simulations reported by others (e.g., Hannan and Carroll 1992). The smoothness is the effect of pooling individual simulations in 183 U.S. regions as defined by the BEA. Much random fluctuation in each individual region is evened out by the pooling.

23. We used GenBank Release 65.0 which combines data from DNA Data Bank of Japan (Mishima, Japan), EMBL Data Library (Heidelburg, Germany), and GenBank (Los Alamos, New Mexico). See Bilofsky and Burks (1988) for a description of the GenBank.

24. A small number of unpublished papers and articles appearing in proceedings volumes and obscure journals were excluded to permit the hand coding detailed below.

25. This positional tradition holds across national boundaries: As a percentage of articles co-authored by their fellow nationals, American stars are 16.4 percent of first authors and 71.2 percent of last authors, compared to 21.2 percent and 63.1 percent, respectively, for Japanese, and 19.7 percent and 69.2 percent for other nationalities.

26. See U.S. Department of Education, National Center for Education Statistics (1985).

27. The respondents were asked to rate programs using the following scale: 5 for distinguished, 4 strong, 3 good, 2 adequate, 1 marginal, and 0 not sufficient for

doctoral education. The reported scores are the averages among respondents.

28. Where a new enterprise enters the data set due to the merger of a NBF and another firm, we count it for the purposes of this paper as a continuation of the original NBF and not a new birth (the older NBF if two are involved). If enterprises already in the data set merge and one continues with the other(s) absorbed, the enterprise is counted as the continuing enterprise and not a new birth.

29. The annual data for the implicit price deflator for personal consumption expenditures were taken from U.S. Department of Commerce (1992a, p. 247, line 16) as updated in the July 1992 *Survey of Current Business*, (p. 92, line 16).

References

Allison, Paul D. 1982. "Discrete-Time Methods for the Analysis of Event Histories." *Sociological Methodology* 13:61-98.

Baum, Joel A.C. 1996. "Organizational Ecology." Pp. 77-114 in *Handbook of Organization Studies*, edited by S. R. Clegg, C. Hardy, and W. R. Nord. London: SAGE.

Baum, Joel A. C. and Christine Oliver. 1992. "Institutional Embeddedness and the Dynamics of Organizational Populations." *American Sociological Review* 57:540-559.

Baum, Joel A. C. and Walter W. Powell. 1995. "Cultivating an Institutional Ecology of Organizations: Comments on Hannan, Carroll, Dundon, and Torres." *American Sociological Review* 4:529-38.

Bilofsky, H.S., and C. Burks. 1988. "The GenBank (R) Genetic Sequence Data Bank." *Nucleic Acids Research* 16:1861-1864.

Cameron, A. Colin and Pravin K. Trivedi. 1990. "Regression-Based Tests for Overdispersion in the Poisson Model." *Journal of Econometrics* 46(December):347-364.

Carroll, Glenn R. 1987. *Publish and Perish: The Organizational Ecology of Newspaper Industries*. Greenwich, CT: JAI.

Carroll, Glenn R. and Jacques Delacroix. 1982. "Organizational Mortality in the Newspaper Industries of Argentina and Ireland: An Ecological Approach." *Administrative Science Quarterly* 27:169-198.

Carroll, Glenn R. and Yangchung Paul Huo. 1986. "Organizational Task and Industrial Environments in Ecological Perspective: Findings from the Local Newspaper Industry." *American Journal of Sociology* 91(January):838-873.

Carroll, Glenn R. and Michael T. Hannan. 1989a. "Density Dependence in the Evolution of Populations of Newspaper Organizations." *American Sociological Review* 54: 524-541.

Carroll, Glenn R. and Michael T. Hannan. 1989b. "On Using Institutional Theory in Studying Organizational Populations: Reply to Zucker." *American Sociological Review* 54:545-548.

Cohen, Stanley, A. Chang, Herbert Boyer, and R. Helling. 1973. "Construction of Biologically Functional Bacterial Plasmids *in vitro.*" *Proceedings of the National Academy of Sciences* 70:3240-3244.

Darby, Michael R. and Lynne G. Zucker. 1996. "Star Scientists, Institutions, and the Entry of Japanese Biotechnology Enterprises." National Bureau of Economic Research Working Paper No. 5795, October.

Delacroix, Jacques and Glenn R. Carroll. 1983. "Organizational Foundings: An Ecological Study of the Newspaper Industries of Argentina and Ireland." *Administrative Science Quarterly* 28:274-291.

Delacroix, Jacques and M. E. Solt. 1988. "Niche Formation and Foundings in the California Wine Industry." Pp. 53-70 in *Ecological Models of Organizations*, edited by G. R. Carroll. Cambridge, MA: Ballinger.

DiMaggio, Paul. 1982a. "Cultural Entrepreneurship in the Nineteenth-Century Boston, Part I: The Creation of an Organizational Base for High Culture in America," *Media, Culture, and Society* 4:33-50.

DiMaggio, Paul. 1982b. "Cultural Entrepreneurship in the Nineteenth-Century Boston, Part II: The Classification and Framing of American Art." *Media, Culture, and Society* 4:303-322.

DiMaggio, Paul. 1988. "Bureaucratizing the Muse: Public Funds and the Cultural Worker." *American Journal of Sociology* 94:208-210.

Greene, William H. 1992. *LIMDEP: User's Manual and Reference Guide, Version 6.0.* Bellport, NY: Econometric Software, Inc.

Grossman, Gene M. and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy.* Cambridge, MA: MIT.

Hannan, Michael T. and Glenn R. Carroll. 1992. *Dynamics of Organizational Populations: Density, Legitimation, and Competition.* New York: Oxford University.

Hannan, Michael T. and Glenn R. Carroll. 1995. "Theory Building and Cheap Talk about Legitimation: Reply to Baum and Powell." *Amercian Sociological Review* 60:539-544.

Hannan, Michael T., Glenn R. Carroll, Elizabeth A. Dundon, and John Charles Torres. 1995. "Organizational Evolution in Multinational Context: Entries of Automobile Manufacturers in Belgium, Britain, France, Germany, and Italy." *Amercian Sociological Review* 60:509-528.

Hannan, Michael T. and John Freeman. 1977. "The Population Ecology of Organizations," *American Journal of Sociology* 82:929-964.

Hannan, Michael T. and John Freeman. 1987. "The Ecology of Organizational Founding: American Labor Unions, 1836-1985." *American Journal of Sociology* 92:910-943.

Hannan, Michael T. and John Freeman. 1988. "The Ecology of Organizational Mortality: American Labor Unions, 1836-1985." *American Journal of Sociology* 94:25-52.

Hannan, Michael T. and John Freeman. 1989. *Organizational Ecology*. Cambridge, MA: Harvard University.

Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica* 52(July):909-938.

Liebeskind, Julia Porter, Amalya Lumerman Oliver, Lynne G. Zucker, and Marilynn B. Brewer. 1996. "Social Networks, Learning, and Flexibility: Sourcing Scientific Knowledge in New Biotechnology Firms." *Organization Science* 7(July/August):428-443.

Lomi, Alessandro and Erik R. Larsen. 1996. "Interacting Locally and Evolving Globally: A Computational Approach to the Dynamics of Organizational Populations." *Academy of Management Journal* 39:1287-1321.

McCarthy, John and Mayer Zald. 1977. "Resource Mobilization and Social Movements: A Partial Theory." *American Journal of Sociology* 82:1212-41.

Miner, Anne S. 1993. "Review of *Dynamics of Organizational Populations: Density, Competition, and Legitimation* by Michael T. Hannan and Glenn R.Carroll." *Academy of Management Review* 18:355-67.

Peng, Yusheng, Lynne G. Zucker, and Michael R. Darby. 1996. "Chinese Rural Industrial Productivity and Urban Spillovers." Working Paper No. 6202. Cambridge, MA: National Bureau of Economic Research.

Raftery, Adrian. 1986. "Choosing Models for Cross-Classifications." *American Sociological Review*. 51:146-147.

Romer, Paul M. 1986. "Increasing Returns and Long-run Growth." *Journal of Political Economy* 94:1002-1037.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98(5, part II):S71-S102.

Sindelar, Robert D. 1992. "Overview/Preview of Current and Future Recombinant DNA-Produced Pharmaceuticals." *Drug Topics* April 20, Supplement, pp. 3-16.

Sindelar, Robert D. 1993. "The Pharmacy of the Future." *Drug Topics* 137(9, May 21):66-84.

Singh, Jitendra V., David J. Tucker, and Agnes G. Meinhard. 1991. "Institutional Change and Ecological Dynamics." Pp. 390-422 in *The New Institutionalism in Organizational Analysis*, edited by W. W. Powell and P. J. DiMaggio. Chicago: University of Chicago.

Tucker, David J., Jitendra V. Singh, and Agnes G. Meinhard. 1990. "Organizational Form, Population Dynamics, and Institutional Change: The Founding Patterns of Voluntary Organizations." *Academy of Management Journal* 33:151-178.

Wooldridge, Jeffrey M. 1991. "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances." *Journal of Econometrics* 47(January):5-46.

Zucker, Lynne G. 1989. "Combining Institutional Theory and Population Ecology: No Legitimacy, No History," *American Sociological Review* 54:542-545.

Zucker, Lynne G., Marilynn B. Brewer, Amalya Oliver, and Julia Liebeskind. 1993. "Basic Science as Intellectual Capital in Firms: Information Dilemmas in rDNA Biotechnology Research." UCLA: Institute for Social Science Research Working Paper.

Zucker, Lynne G. and Michael R. Darby. 1995. "Virtuous Circles of Productivity: Star Bioscientists and the Institutional Transformation of Industry," National Bureau of Economic Research Working Paper No. 5342, November.

Zucker, Lynne G. and Michael R. Darby. 1996a. "Costly Information: Firm Transformation, Exit, or Persistent Failure." *American Behavioral Scientist* 89 (8, August): 959-974.

Zucker, Lynne G. and Michael R. Darby. 1996b. "Star Scientists and Institutional Transformation: Patterns of Invention and Innovation in the Formation of the Biotechnology Industry." *Proceedings of the National Academy of Sciences* 93 (23, November 12):12709-12716.

Zucker, Lynne G. and Michael R. Darby. 1998. "Capturing Technological Opportunity Via Japan's Star Scientists: Evidence from Japanese Firms' Biotech Patents and Products." National Bureau of Economic Research Working Paper No. 6360, January.

Zucker, Lynne G. and Michael R. Darby. Forthcoming. "Present at the Biotechnological Revolution: Transformation of Technical Identity for a Large Incumbent Pharmaceutical Firm." *Research Policy*.

Zucker, Lynne G., Michael R. Darby, and Jeff Armstrong. 1998. "Geographically Localized Knowledge: Spillovers or Markets?" *Economic Inquiry* 36 (1, January):65-86.

Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. 1998. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *American Economic Review* 88 (1, March):in press.

Zucker, Lynne G., Michael R. Darby, Marilynn B. Brewer, and Yusheng Peng. 1996. "Collaboration Structure and Information Dilemmas in Biotechnology: Organizational Boundaries as Trust Production." Pp. 90-113 in *Trust in Organizations*, edited by R. M. Kramer and T. R. Tyler. Thousand Oaks, CA: SAGE.

Zucker, Lynne G., Michael R. Darby, and Maximo Torero. 1997. "Labor Mobility from Academe to Commerce." National Bureau of Economic Research Working Paper No. 6050, May.

Zucker, Lynne G. and Ita G. G. Kreft. 1994. "The Evolution of Socially Contingent Rational Action: Effects of Labor Strikes on Change in Union Founding in the 1880s." Pp. 294-313 in *Evolutionary Dynamics of Organizations*, edited by J. A. C. Baum and J. V. Singh. Oxford, UK: Oxford University.

Data Sources

Bioscan. 1989-1993. Volumes 3-7.

Cetus Corp. 1988. "Biotechnology Company Data Base," predecessor source for *Bioscan.* Printout. Cetus Corp., 1988.

CITIBASE: Citibank Economic Database. 1993. Machine readable data base, 1946-June 1993. New York: Citibank, N.A.

GenBank, Release 65.0. 1990. Machine readable data base. Palo Alto, CA: IntelliGentics, Inc., September.

GenBank, Release 81.0. 1994. Machine readable data base. Bethesda, MD: National Center for Biotechnology Information, February 15.

Institute for Scientific Information. 1982, 1987, 1992. *Science Citation Index.* ISI Compact Disc Editions. Machine readable data base. Philadelphia: Institute for Scientific Information.

Jones, Lyle V., Gardner Lindzey, and Porter E. Coggeshall, eds. 1982. *An Assessment of Research-Doctorate Programs in the United States: Biological Sciences.* Washington: National Academy Press.

North Carolina Biotechnology Center. 1992. *North Carolina Biotechnology Center U.S. Companies Database.* Machine readable data base. Research Triangle Park, NC: North Carolina Biotechnology Center, April 16.

Pratt, Stanley E. 1982. *Guide to Venture Capital Sources,* 6th ed., Englewood Cliffs, NJ: Prentice-Hall, Inc.

U.S. Department of Commerce, Bureau of the Census. 1982. *County Business Patterns, 1980: U.S. Summary, State, and County Data,* machine readable data base, Washington: Bureau of the Census. [Ann Arbor, MI: Inter-University Consortium for Political and Social Research, 1986 (distributor)]

U.S. Department of Commerce, Economics and Statistics Administration, Bureau of Economic Analysis. 1992a. *National Income and Product Accounts of the United States, Volume 2, 1959-88,* Washington: U.S. Government Printing Office.

U.S. Department of Commerce, Economics and Statistics Administration, Bureau of Economic Analysis. 1992b. *Regional Economic Information System, Version 1.3,* CD-ROM. Machine readable data base. Washington: Bureau of Economic Analysis, May 5.

U.S. Department of Education, National Center for Education Statistics. circa 1985. *Higher Education General Information Survey (HEGIS), Institutional Characteristics, 1983-84.* Machine readable data base, ICPSR 8291. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Table 1: Distribution of Star Scientists and Collaborators Who Have Ever Published in the United States

Organization[a]	Full Data Set		Ever Active in U.S.[b]	
Type	Number of Citations[c]/ Scientists	Scientist/yr.	Number of Citations[c]/ Scientists	Scientist/yr.
Stars:				
University	158	85.5	108	110.8
Institute	44	63.0	26	98.7
Firm	5	143.7	1	694.3
Dual	0	n/a	0	n/a
Total	207		135	
Collaborators:				
University	2901	10.4	369	30.6
Institute	776	13.7	88	35.8
Firm	324	29.2	43	99.1
Dual	3	7.2	0	n/a
Total	4004		500	

Notes:

a. The organization type refers to the affiliation listed on their first publication with a U.S. affiliation.

b. Ever active in the U.S. means that in at least one three year period beginning 1974 or later and ending 1989 or earlier, the scientist was listed on at least three articles appearing in our data set of 4,061 articles which reported gene sequence discoveries and were published in major journals and that the affiliation listed in the last of the three articles was located in the United States.

c. Citation counts are for 1982, 1987, and 1992 for all articles in our data set (whenever published) for which the individual was listed as an author.

Table 2a__Poisson regressions of the birth of U.S. biotech-using firms by BEA and year (1976-1989): comparing organizational ecology model and fundamental models. (N = 2562)

VARIABLES	Population ecology (I)	Fundamentals (II)	Augmented fundamentals (III)
Constant	-1089.1***	-68.86	-1089.8***
	(170.1)	(40.43)	(250.4)
year	0.5497***	0.0338	0.5507***
	(.0860)	(.0203)	(.1268)
lagged births of biotech enterprises	0.5471***		0.3176***
	(.1100)		(.0791)
lagged births of biotech enterprises squared	-0.0359***		-0.0192**
	(.0096)		(.0059)
lagged local density	0.1344***		0.0493
	(.0188)		(.0253)
lagged local density squared	-0.0015***		-0.0008*
	(.0003)		(.0003)
lagged national density	-0.0099***		-0.0097***
	(.0013)		(.0022)
number of star scientists		0.3448***	0.2427**
		(.0841)	(.0925)
number of star scientists squared		-0.0129**	-0.0086
		(.0049)	(.0047)
number of collaborators		-0.0143	-0.0969
		(.0521)	(.0558)
number of collaborators squared		0.0011	0.0033*
		(.0015)	(.0015)
number of star scientists, 1986-89		-0.3167**	-0.2508
		(.1185)	(.1542)
number of star scientists squared, 1986-89		0.0102	0.0121
		(.0060)	(.0074)
number of collaborators, 1986-89		0.1203	0.1412
		(.0720)	(.0909)
number of collaborators squared, 1986-89		-0.0016	-0.0035
		(.0018)	(.0020)
top ranking universities		0.3330**	0.2857
		(.1132)	(.1550)
federal grant support		0.4825**	0.2951
		(.1526)	(.1767)
venture capital firms		-0.032***	-0.0299***
		(.0058)	(.0066)
logarithm of employment		0.7368***	0.6707***
		(.1071)	(.1088)
earnings/price ratio		0.0273	-0.0894
		(.0387)	(.0544)
Log likelihood ratio Chi-squares	1644.1	1497.7	1309.8
Degree of freedom	2555	2547	2542
Bayesian Information Coefficients	-18,409	-18,492	-18,641

Table 2b__Poisson regressions of the birth of U.S. biotech-using firms by BEA and year (1976-1985): comparing organizational ecology model and fundamental models. (N = 1830)

VARIABLES	Population ecology (I)	Fundamentals (II)	Augmented fundamentals (III)
Constant	-1095.7***	-274.82***	-1076.1***
	(159)	(67.61)	(267.7)
year	0.5531***	0.1374***	0.5436***
	(.0804)	(.0340)	(.1355)
lagged births of biotech enterprises	0.4210***		0.2543***
	(.1065)		(.0764)
lagged births of biotech enterprises squared	-0.0268**		-0.0136**
	(.0084)		(.0047)
lagged local density	0.2023***		0.0273
	(.0312)		(.0288)
lagged local density squared	-0.003***		-0.0012*
	(.0006)		(.0005)
lagged national density	-0.0105***		-0.0085***
	(.0013)		(.0026)
number of star scientists		0.2863**	0.2141*
		(.0919)	(.0911)
number of star scientists squared		-0.0113*	-0.0015
		(.0050)	(.0051)
number of collaborators		-0.0299	-0.0986
		(.0575)	(.0554)
number of collaborators squared		0.0011	0.0029*
		(.0017)	(.0014)
top ranking universities		0.5028***	0.4214*
		(.1354)	(.1723)
federal grant support		0.2133	0.2819
		(.1774)	(.1934)
venture capital firms		-0.0261***	-0.0309***
		(.0073)	(.0072)
logarithm of employment		0.8694***	0.802***
		(.1328)	(.1247)
earnings/price ratio		0.1045*	-0.0699
		(.0518)	(.0660)
Log likelihood ratio Chi-squares	1218.3	1014.9	911.5
Degree of freedom	1823	1819	1814
Bayesian Information Coefficients	-12,476	-12,649	-12,715

Note: Figures in the parentheses are Wooldridge-corrected standard errors. Asterisks indicate significance level.
* = significant at 0.05 level; ** = significant at 0.01 level; *** = significant at 0.001 level.

Table 3. Discrete-time mortality model of U.S. biotechnology enterprises by BEA and year (1976-1989).

VARIABLES	(I)	(II)
Constant	-1777.3***	-1661.8**
	(490.7)	(507.3)
year	0.8900***	0.8316**
	(.2468)	(.2552)
firm age	1.0392*	5.9393*
	(.4298)	(.4328)
firm age squared	-0.0721*	4.7973*
	(.0332)	(.0333)
lagged local density		0.0232
		(.0372)
lagged local density squared		-0.0001
		(.0004)
California		-0.3537
		(.8175)
East Coast (NY, NJ, MD, MA)		-0.3951
		(.7887)
Log likelihood ratio Chi-squares	230.95	228.28
Degree of freedom	3	7

Note: Figures in parentheses are standard errors. Asterisks indicate significance level. * = significant at 0.05 level; ** = significant at 0.01 level; *** = significant at 0.1 level.

Table 4. Comparing Geographic Distribution of U.S. Biotechnology Enterprises in 1989
Summary Statistics of 100 Simulations

	Total Count of Live Firms (data=730)			Standard Deviations across BEAs (data=11.8)			Pearson Correlation with Data (data=1.000)		
	Population ecology model	Population ecology model with initial data	Augmented fundamentals model	Population ecology model	Population ecology model with initial data	Augmented fundamentals model	Population ecology model	Population ecology model with initial data	Augmented fundamentals model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Maximum	741	732	751	8.4	7.3	12.2	0.449	0.522	0.898
Third quartile	698	697	729	6.1	5.9	11.5	0.035	0.191	0.863
Mean	687.8	687.5	723.3	5.0	5.1	11.2	0.013	0.148	0.839
Median	687	685.5	722	4.9	5.2	11.3	-0.007	0.122	0.845
First quartile	676	679	718	4.1	4.3	11.0	-0.045	0.064	0.819
Minimum	651	655	696	2.7	2.8	10.5	-0.094	-0.029	0.750

Figure 1: New authors of genetic sequence discovery articles by whether they enter *GenBank* writing with old authors

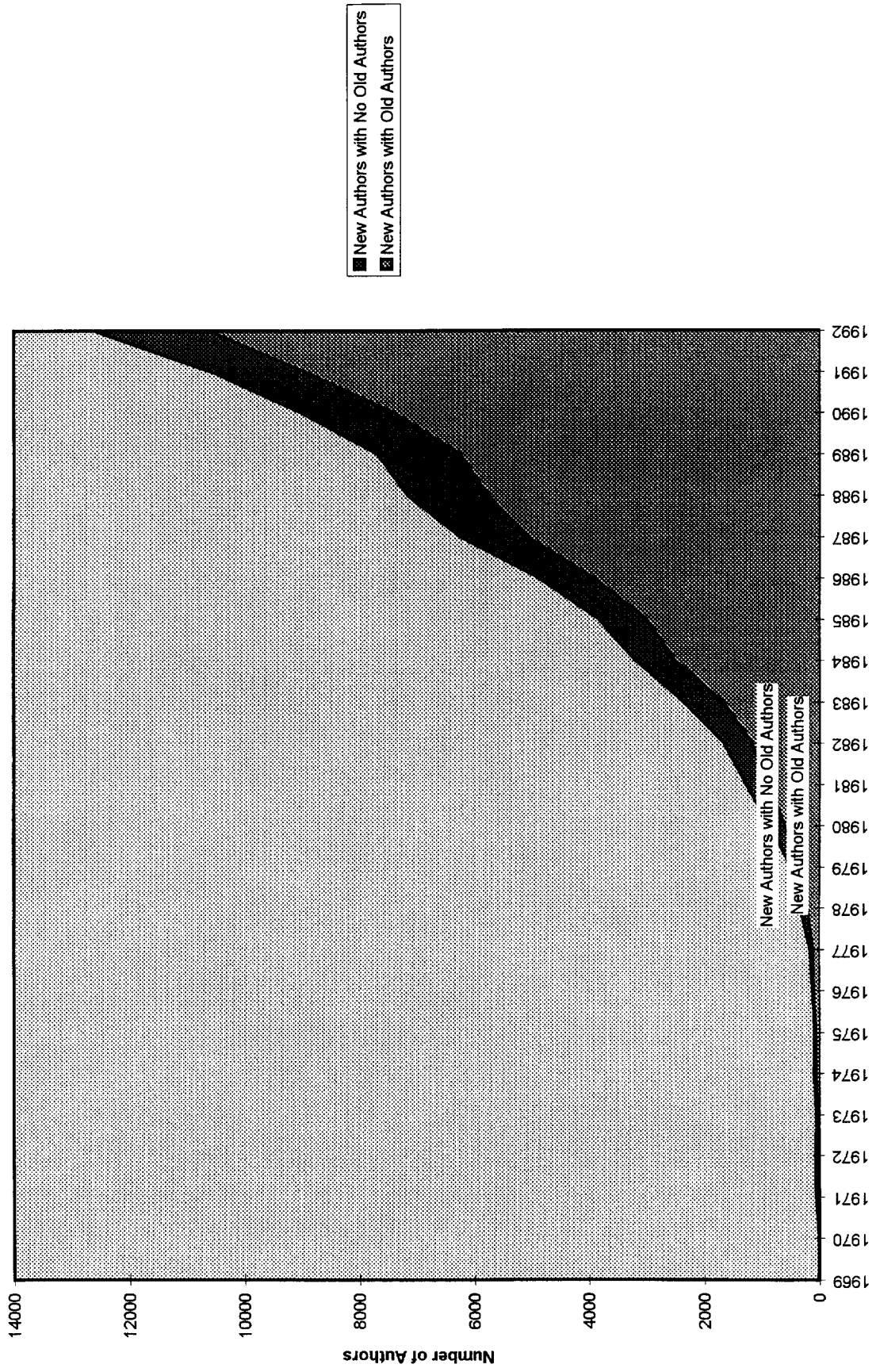- New Authors with No Old Authors
- New Authors with Old Authors

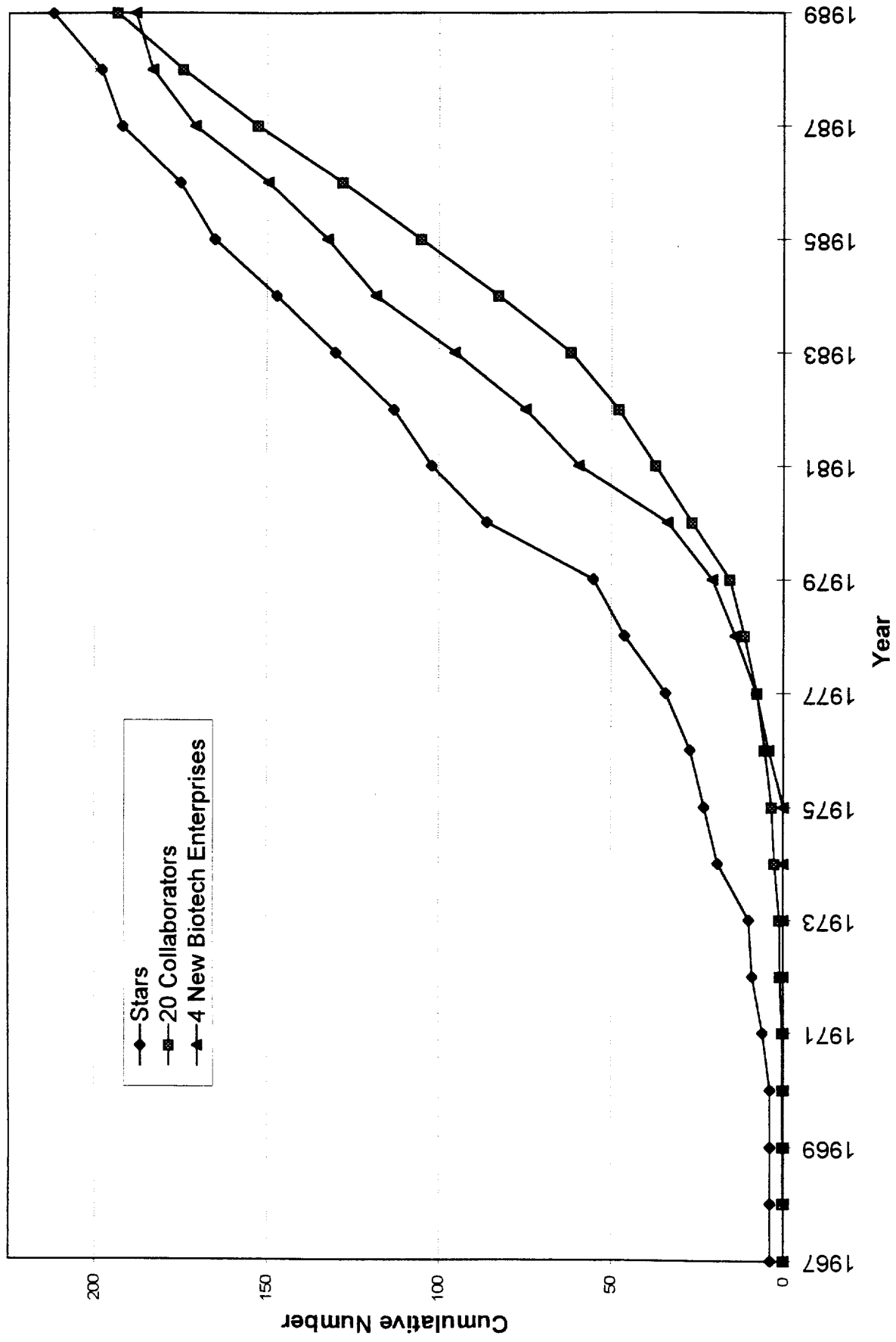Figure 2: Cumulative number of stars, collaborators, and new biotech enterprises in the U.S., 1967-1989
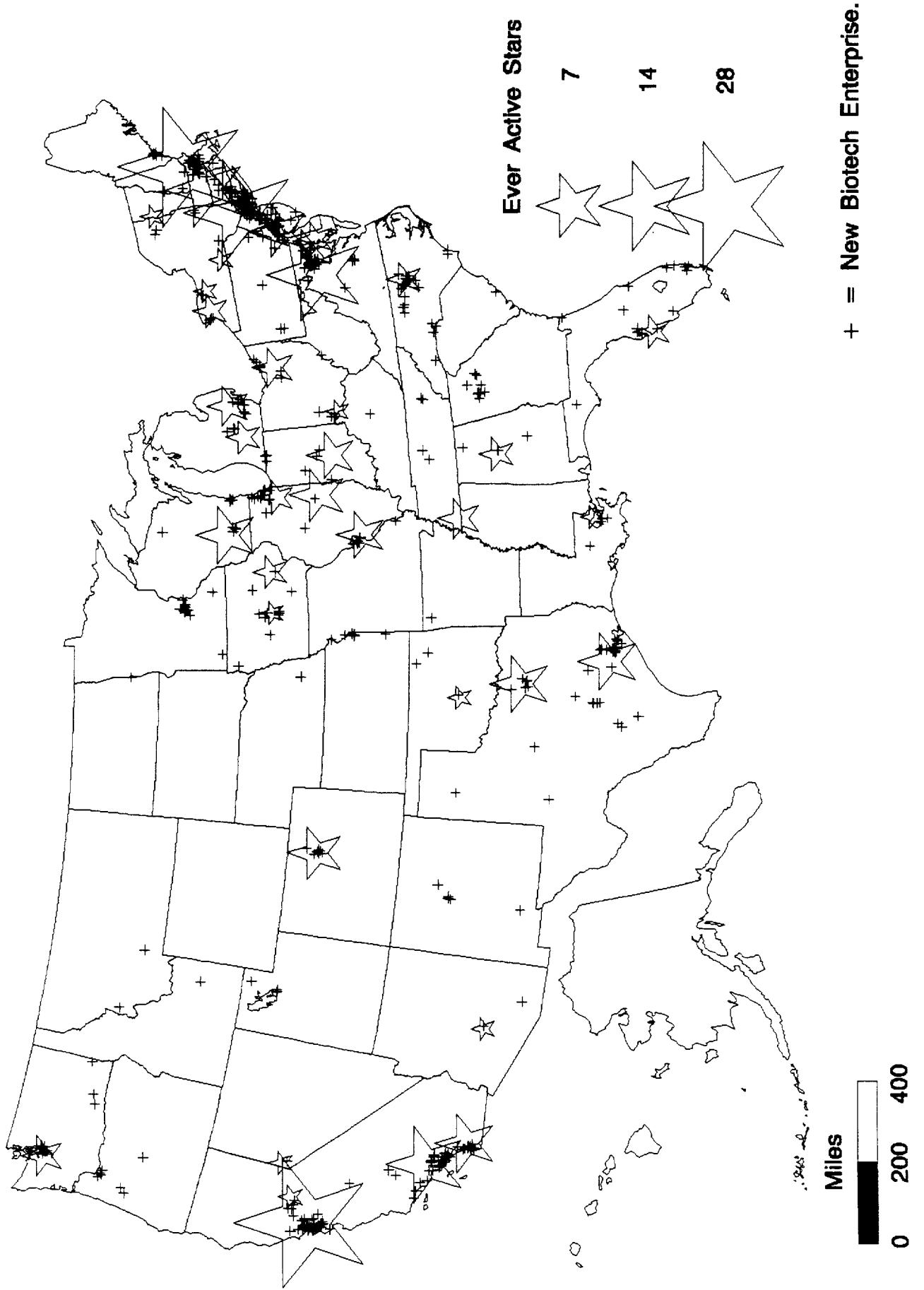
Figure 3: Cumulative number of ever-active stars and new biotechnology enterprises as of 1990

Ever Active Stars

7

14

28

+ = New Biotech Enterprise.

Miles

0   200   400

Figure 4: The function of dependence of births of new biotech enterprises on lagged births: Comparison of population ecology and mixed models
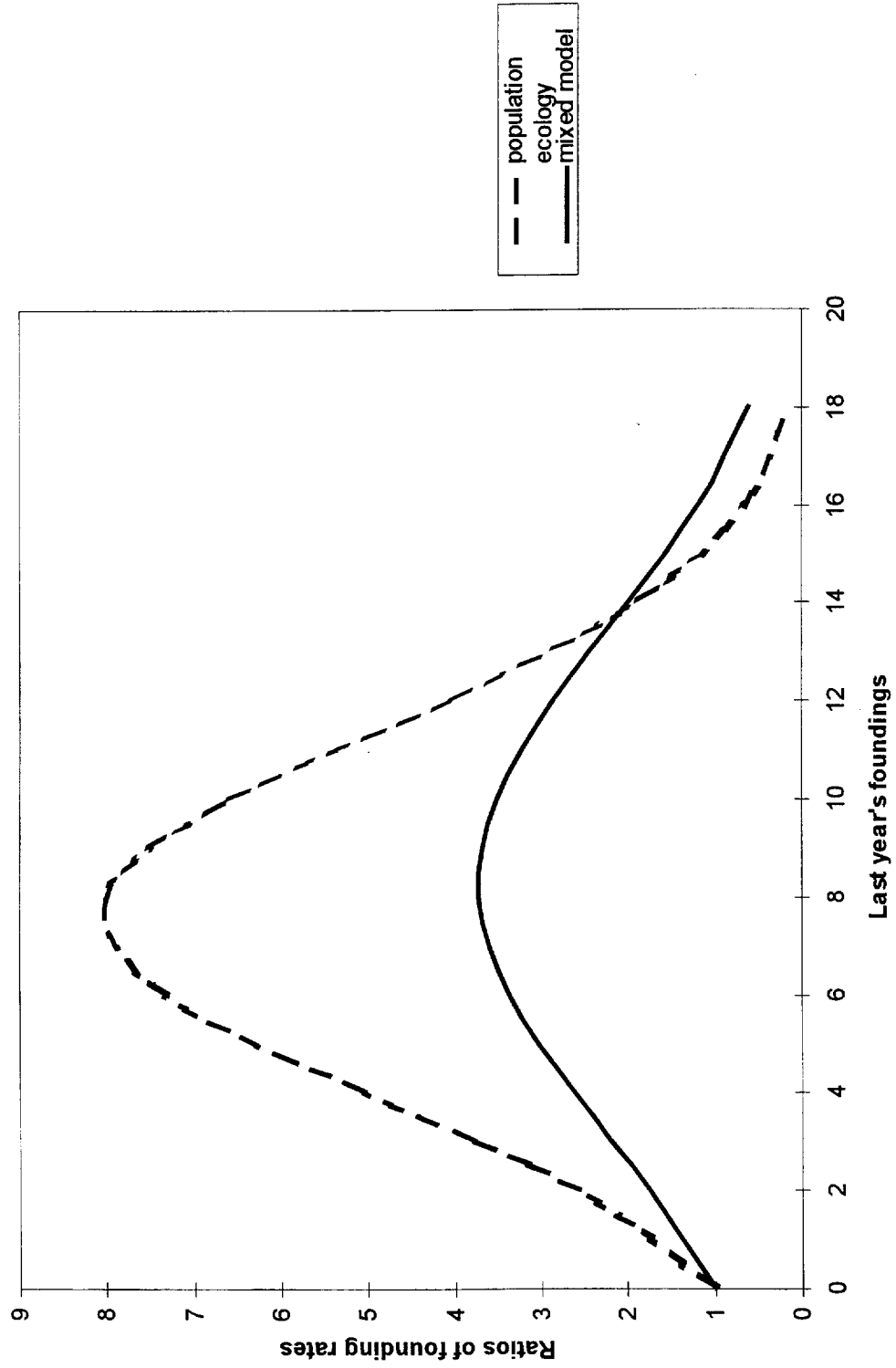
Figure 5: Density dependence function of births of new biotech enterprises: Comparison of population ecology and mixed models

population ecology
mixed model

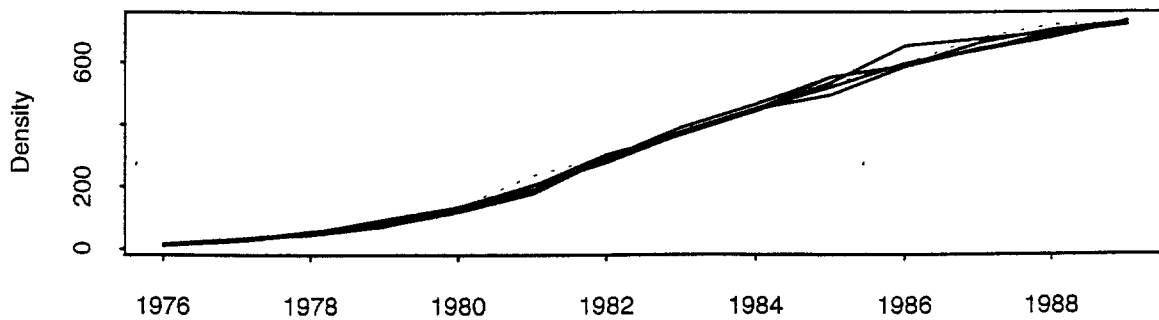Last year's density

Ratios of founding rates
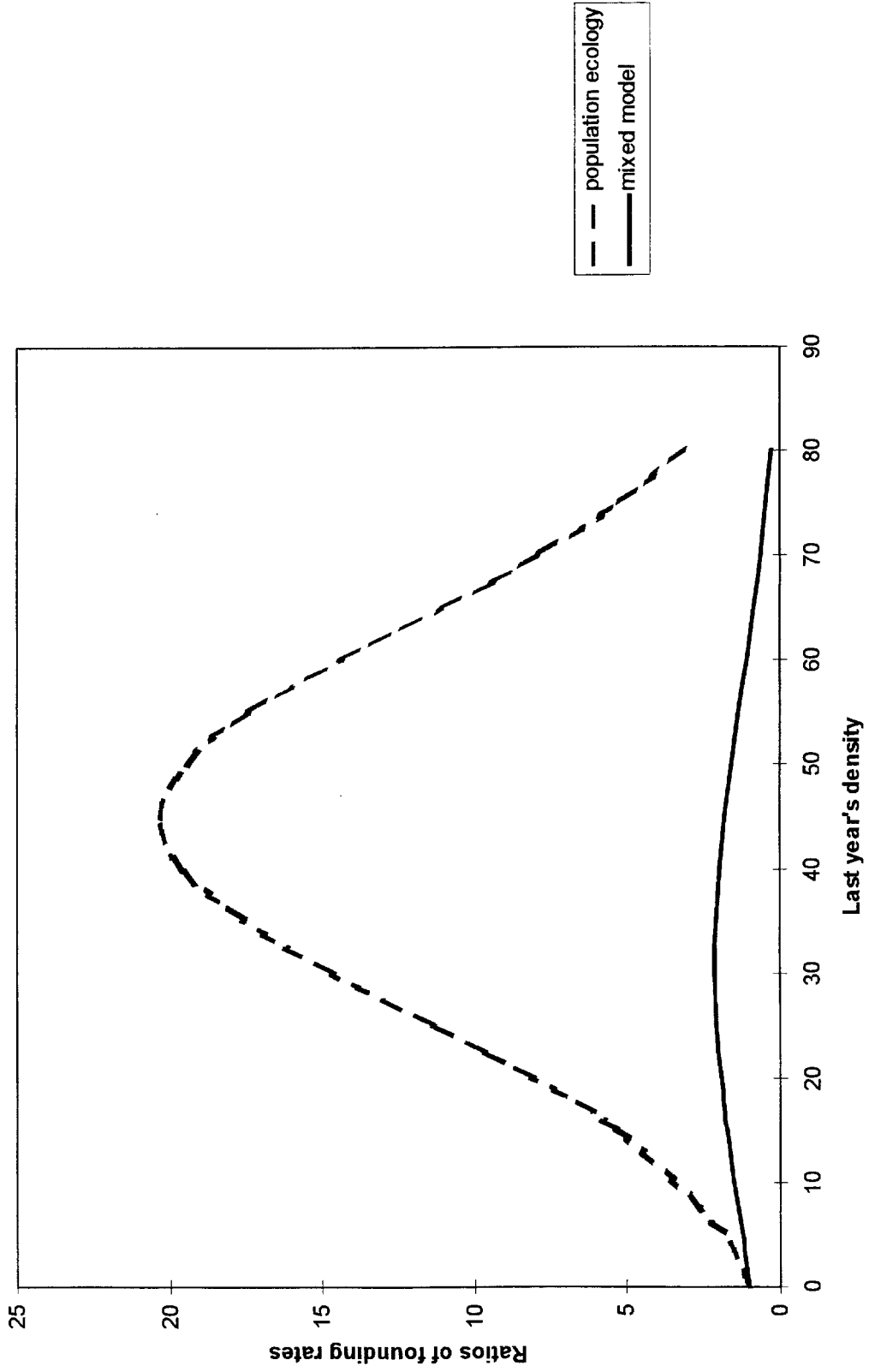
Figure 6: National totals from dynamic simulations of births of new biotech enterprises in the U.S., 1976-1989 (dotted lines represent actual values)
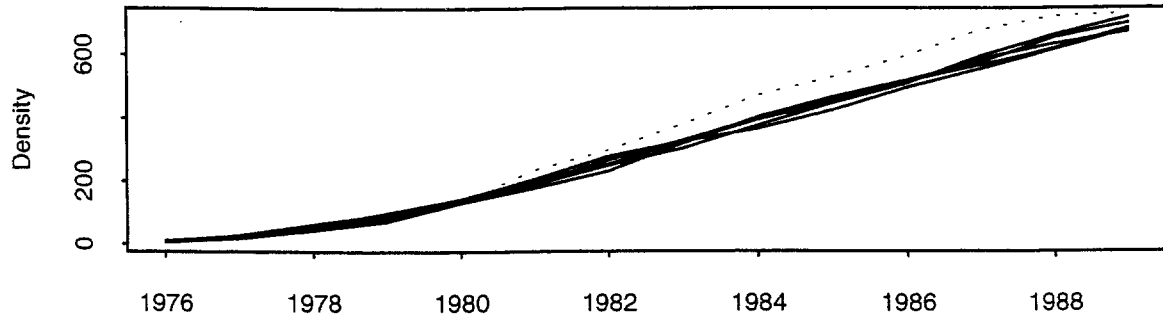
Plot a: Population ecology model without initial founding data, 5 simulations
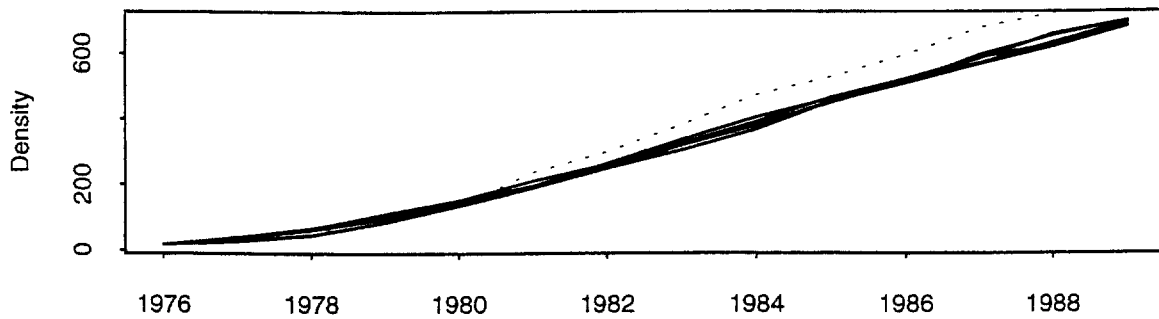
Plot b: Population ecology model with initial founding data, 5 simulations

Plot c: Augmented fundamentals model without initial founding data, 5 simulations