Geographic Concentration as a Dynamic Process[1]

Guy Dumais
Harvard University

Glenn Ellison
MIT and NBER

and

Edward L. Glaeser
Harvard University and NBER

October, 1997

[1]email: dumais1@fas.harvard.edu, gellison@mit.edu, glaeser@fas.harvard.edu. The first author thanks the Social Sciences and Humanities Research Council of Canada for financial support. The second and third authors thank the National Science Foundation for support through grants SBR-9515076 and SES-9601764, and the second author was also supported by a Sloan Research Fellowship. We also thank Joyce Cooper of the U.S. Census Bureau's Boston Research Data Center for a great deal of help, and Aaron Hantman for research assistance. The opinions and conclusions expressed in this paper are those of the authors and do not necessarily represent those of the U.S. Census Bureau.

Abstract

The degree of geographic concentration of individual manufacturing industries in the U.S. has declined only slightly in the last twenty years. At the same time, new plant births, plant expansions, contractions and closures have have shifted large quantities of employment across plants, firms, and locations. This paper uses data from the Census Bureau's Longitudinal Research Database to examine how relatively stable levels of geographic concentration emerge from this dynamic process. While industries' agglomeration levels tend to remain fairly constant, we find that there is greater variation in the locations of these agglomerations. We then decompose aggregate concentration changes into portions attributable to plant births, expansions, contractions, and closures, and find that the location choices of new firms and differences in growth rates have played the most significant role in reducing levels of geographic concentration, while plant closures have tended to reinforce agglomeration. Finally, we look at coagglomeration patterns to test three of Marshall's theories of industry agglomeration: (1) agglomeration saves transport costs by proximity to input suppliers or final consumers, (2) agglomeration allows for labor market pooling, and (3) agglomeration facilitates intellectual spillovers. While there is some truth behind all three theories, we find that industrial location is far more driven by labor mix than by any of the other explanatory variables.

1 Introduction

At this point, it is well known that industries are often geographically concentrated in particular states or metropolitan areas (*e.g.* Florence (1948), Hoover (1948), Fuchs (1962), Enright (1990), Krugman (1991), Ellison and Glaeser (1996)). Further, there is no shortage of theories explaining why this concentration may occur (von Thunen (1825), Marshall (1920), Krugman (1991)). This paper attempts to add a new dimension to existing descriptive work on geographic concentration and to improve our understanding of the relative importance of these theories by using data from the U.S. Census Bureau's Longitudinal Research Database (LRD) to examine geographic concentration as the outcome of a dynamic process in which new plants are constantly being born, existing plants are expanding and contracting at very different rates, and a substantial number of businesses are failing. Among our most notable conclusions are that there is a substantial degree of movement of industries over time even among geographically concentrated industries; that births of new firms and expansions of existing plants tend to reduce agglomeration, while differing rates of plant closures reinforce agglomeration levels; and that labor market explanations for geographic concentration seem to be important empirically.

Our investigation begins in section 3 with a review of (and expansion on) the most basic facts on the dynamics of agglomeration. Fuchs's (1962) monograph reported a fairly widespread decline in agglomeration between 1929 and 1954. Kim (1995) delves more deeply into the historical record and reports that on average agglomeration levels increased between 1860 and 1947 and declined thereafter, with the result that in many 2-digit industries the level of agglomeration today is quite similar to what it was in 1860. To this, we add a more detailed look at recent history using the index of Ellison and Glaeser (1997) to avoid certain measurement problems. We find that agglomeration levels have declined slightly in the past twenty years, but there is a striking degree of stability in which 3-digit manufacuring industries were agglomerated in 1972 and 1992.

The stability in the degree to which individual industries are geographically concentrated contrasts sharply with the dramatic changes in employment which exist at the plant level (Dunne, Roberts and Samuelson (1989 a,b), Davis and Haltiwanger (1991)). Given this plant-level turnover, the agglomeration of industries clearly can not be attributed solely to historical accidents which occurred when the industry was founded. Many views of the relative importance of historical accidents and external economies (or natural advantage) are, however, consistent with this turnover. At one extreme, favoring historical accidents, one could imagine that there is great turnover of employment at the plant level but none at the state-industry level; new plants may just be built next to the old ones they replace, and plant growth may come at the expense of competitors across the street. At the other extreme, there may be a lot of movement of agglomerated industries with old centers dying out while new ones emerge. In this case, the stability of agglomeration levels would suggest that agglomeration levels are determined by strong equilibrium forces. In section 4, we develop a framework for thinking about where in this range industries fall. We then use the LRD data to describe experience of various subsets of industries, with one of our primary conclusions being that there appears to be a substantial degree of movement even among highly agglomerated industries. Distant historical accidents may then be of limited importance as a determinant of where industries are located today.

The changes in employment across states and metropolitan areas are themselves the product of a variety of changes at the plant level. For example, there are a number of different ways in which new centers of activity in an industry may have developed: some areas may have been hotbeds of entrepeneurial startup activity; others may have been successful in attracting the new plant investments of established firms; others may have had one or a core of firms located there grow to a dominant position in their industry. In section 5, we develop a framework for decomposing agglomeration changes into portions attributable to various stages of the plant life-cycle (including births of new firms, new plant openings by existing firms, the expansion or contraction of existing plants and closures). We then discuss the extent to which events at the various stages of the plant life-cycle have tended to increase or decrease levels of geographic concentration in our data. Industry centers generate less then their proportionate share of new firm births. Existing plants also tend to grow more slowly in areas where an industry is concentrated. Each of these processes tends to reduce geographic concentration. We also find that rates of plant closure are lower in such areas, however, which tends to reinforce existing agglomerations. A complete theory of geographic concentration must then account for a variety of effects at different stages of a plant's life cycle.

Finally, in section 6, we focus directly on trying to explain why, on average, industrial concentration persists. We consider three basic theories (all discussed in Marshall (1920)): (1) firms locate near one another to decrease transportation costs, (2) firms locate near one another so that workers can move from one firm to another in the event of a firm specific downturn and (3) firms locate near one another because of intellectual spillovers. Because the theories all predict that firms will want to locate near other plants in the same industry, we have chosen to try to distinguish them by looking at how location choices are affected by the presence of plants in related industries.

We use data from a variety of sources to form variables reflecting factors which the various theories predict might be important determinants of location choices: we use input-output matrices to capture the interdependence of different industries due to supply/demand relations; tables of occupation by industry to determine the extent to which plants use te same mix of labor; and information on technology flows and cross-industry co-ownership patterns to get potential measures of opportunities for intellectual spillovers. We then look for the connection between these variables and the rates of plant births, expansions/contractions and closures at the metropolitan area level, controlling for initial area-industry employment and allowing for area-specific fixed effects.

Our results provide some support for each of the theories. New plants are more likely to be located in areas with more potential input suppliers and more potential downstream customers (with the latter effect being more robust), although the magnitude of this effect is fairly small. The co-ownership measure also tends to be important. However, both in terms of the economic magnitude of its effect and in its statistical significance, our measure of labor market suitability is the most important predictor of new plant locations. This effect is also more important in more volatile industries, which confirms an added implication of the Marshall theory. We should be careful to note, however, that industries with similar labor requirements may be similar in other ways as well, and thus part of our labor effect might very well be attributable to some form of intellectual spillovers instead.

2 Data

The main data used in this paper come from the U.S. Census Bureau's Longitudinal Research Database. The LRD is a longitudinal microdata file that links the information obtained from the manufacturing establishments included in the quinquennial Census of Manufactures (since 1963) and the Annual Survey of Manufactures (since 1972). McGuckin (1988) and Davis, Haltiwanger and Schuh (1996) provide a detailed account of the information found in this dataset. In this study, we focus the analysis on the five census years since 1972, i.e. 1972, 1977, 1982, 1987 and 1992, using 1972 as a base year. This provides between 312,000 and 370,000 observations in every census year, covering every manufacturing establishment in the U.S. We will briefly address some of the major features of this data set here; Appendix B provides a more detailed description of our variable construction.

One of the advantages of the LRD is that it makes it possible to follow a plant through the stages of its life-cycle. Our analysis will focus on a breakdown of employment changes into five categories: births of new firms, the opening of new plants by existing firms, the expansion or contraction of existing plants, plant closures and switches of plants between industries. We define a plant birth between time t and time $t+1$ as a plant that appears in the time $t+1$ census and either does not appear in the time t census or does so with zero reported employment. We obtain our first two categories of employment change by classifying the birth of an establishment that is not part of a firm owning other establishments covered by the Census of Manufactures as a "new firm birth," and the birth of an establishment which is owned by a firm that had establishments in previous censuses as an "old firm birth."[1] Using this distinction, an average of 87% of all newly-created plants in our four five-year intervals can be classified as new firm births. These plants are smaller on average than the plants opened by existing firms and account for about 50% of employment growth due to plant births. Our third category of employment change is the net expansion and contraction of plants which had positive employment in the industry at time t and which are still in operation with strictly positive employment at time $t+1$. Our fourth category, plant closures, consists of all changes attributable to plants which had positive employment in the industry at time t and have zero employment or do not appear in the time $t+1$ census. Finally, to make the employment changes add up to the total employment change we need a fifth category: switches. This category includes all employment losses or gains in an industry which are attributable to plants which were classified as belonging to the industry at time t being

[1]These new firm births could be connected to firms that existed in the previous census year but did not have a presence in manufacturing. We believe, however, that a large majority of these plants represent true entrepreneurship and the formation of a new firm, not just a new plant.

classified as belonging to another industry at $t+1$ and vice versa.[2] While some of these switches undoubtedly reflect real changes in the activity of plants, others are likely just reclassifications, and thus, we will only briefly discuss results on switches.[3]

We measure the level of economic activity in an industry in a given area by total employment in all manufacturing establishments excluding auxiliary units. At various points in the paper we use both states and metropolitan areas as geograpic units. There are 51 "states" including the District of Columbia and 307 metropolitan areas. States have the advantages of being all inclusive while metropolitan areas may be more meaningful economic units. At the industry level, we examine 134 manufacturing industries corresponding to 3-digit industries in the 1987 Standard Industrial Classification.[4] We, therefore, have 6,834 state-industry units of analysis, with four observations on each (one for each five-year interval between censuses) for a total of up to 27,336 observations when pooling observations from all four periods. Using metropolitan areas as the unit of analysis, we have 41,138 metropolitan area-industry (or city-industry) units of analysis in each time period, for a total of 164,552 total observations.

3 Preliminary facts

The geographic concentration of manufacturing industries displays a striking combination of large changes at the plant level and small changes in the aggregate. As several authors have noted, there is a great deal of turnover of manufacturing plants. For instance, 55% of all manufacturing employees in 1992 were working in plants that did not even exist as of 1972, and 73% of the plants that existed in 1972 closed in the next twenty years. There have also been substantial employment changes at existing plants. For example, at the 211 plants having between 950 and 1050 employees in 1972 which were still operating in 1992, employment in 1992 has a mean of

[2] There is some arbitrariness in how one allocates the growth plants which both change employment and switch industries. The convention we have adopted is to assign the full net expansion to the inital industry. For example, if a plant is listed as having 100 employees in industry A at time t and 80 employees in industry B at time $t+1$ we regard industry A as having lost 20 employees in a contraction and 80 in a switch and industry B as having gained 80 employees in a switch.

[3] See McGuckin and Peck (1992).

[4] The sample consists of all manufacturing industries except SIC's 211, 212, 213, 214, 237, and 381. In each of the industries we omitted there were large employment changes attributable to plants being reported as having shifted between the industry in question and a closely related industry, *e.g.* between the fur industry and the women's outerwear industry. We felt that these switches may well have been reclassifications rather than real changes, and because the industries in question were also fairly agglomerated, we worried that they could have a large impact on our results.

793 and a standard deviation of 678.

In addition to the generally high level of turnover, there are also substantial differences in turnover rates across geographical areas. In Arizona, only 18% of the manufacturing plants that were in operation in 1972 were still operating in 1992, less than half the rate in Wisconsin (37%). Looking at the birth process, we observe states with very high levels of new plant formation relative to others. For example in Nevada, the mean rate of employment change due to the birth of new plants in the four five-year periods we study was 44%, compared to 9% for Ohio. Of course, net growth rates of employment change also differ radically across states: Utah had an annualized mean growth rate of 3.2% while New York's manufacturing employment declined at the rate of 2.4% per year over the twenty year period.

Given this high turnover and the heterogeneity of experience in different areas, we think its interesting to note that there has not been much change in the geographic concentration of individual industries.[5] The Ellison and Glaeser index (herefter EG) of the degree to which industry i is geographically concentrated at time t is given by

$$\gamma_i \equiv \frac{G_{it}/(1-\sum_s s_{st}^2) - H_{it}}{1-H_{it}},$$

where s_{ist} is the share of industry i's time t employment located in state s, $G_{it} \equiv \sum_s (s_{ist} - s_{st})^2$, is a sum of squared deviations of the industry's employment shares s_{ist} from a measure, s_{st}, of the state's share of aggregate employment, and H_{it} is a Herfindahl-style measure of the plant-level concentration of employment in an industry: $H_{it} \equiv \sum_k e_{kt}^2 / (\sum_k e_{kt})^2$ where e_{kt} is the level of employment in the kth plant in industry i at time t.[6]

The first row of Table 1 reports the mean across 3-digit manufacturing industries of the EG index. As can be seen, the concentration of the mean industry has remained fairly constant between 1972 and 1982, and then fell by about 10% in the following decade.

In practice, changes in the value of the EG index over time are well approximated by the

[5]This message may seem a bit at odds with the emphasis of the previous literature on the trends in agglomeration levels. However, while Fuchs (1962) reports a sizable decrease in agglomeration between 1929 and 1954, the index he uses is problematic in that one would imagine that it will tend to decrease whenever the plant-level concentration of the industry decreases. Fuchs, in fact, notes that decreases in agglomeration were largest in the fastest growing industries and that slow-growing and shrinking industries on average saw their agglomeration level increase. Kim (1995) reports an increase in agglomeration up to 1947 followed by a decrease from 1947 to 1987 with the largest part of the decrease occurring between 1947 and 1967.

[6]In a departure from Ellison and Glaeser (1997), the measure s_{st} we use here is the unweighted arithmetic mean of the s_{ist} across the industries in our sample, i.e. $s_{st} = (1/I)\sum_i s_{ist}$ where I is the total number of industries.

difference between G_{it} and H_{it}. Ellison and Glaeser (1997) refer to G_{it} as the raw geographic concentration of employment in an industry. The subtraction of H_{it} is a correction which accounts for the fact that the G_{it} measure would be expected to be larger in industries consisting of fewer larger plants if locations were chosen completely at random. Because plant size distributions tend to change fairly slowly, the correction is less important in cross-time comparisons within a short time period than in cross-industry comparisons. The third and fourth rows of Table 1 give the means of G_{it} and H_{it} from 1972 to 1992. These show that the decline in the EG index is associated mainly with a decrease in raw geographic concentration rather than a change in the average size of plants. For this reason our initial discussion of concentration changes will focus on raw concentration.

Table 2 illustrates that the substantial differences which exist in concentration across industries change little over time. The correlation of EG indices measured five years apart is approximately 0.97, and the correlation between the 1972 and 1992 values is 0.92. Thus, the ongoing dynamic process somehow manages to maintain fairly stable levels of agglomeration. This stability is perhaps even more striking in Kim's (1995) calculation of Hoover's coefficient of regional localization for two-digit industries. The correlation between the 1860 and 1987 values of the localization coefficients he reports is 0.64.

With these contrasting facts as motivation, this paper has two complementary tasks: first, we seek to document the dynamic process that determines the overall concentration of industry, and second, we seek to assess whether various theories of geographic concentration are consistent with the observed dynamic patterns.

4 Agglomeration changes and the mobility of industries

The combination of turnover at the plant level and stability in agglomeration levels is compatible with two very different mobility patterns: agglomeration could be constant because new plants just replace old plants at the same location so state-industry employments do not change; or, alternatively, some equilibrating forces may keep agglomeration roughly constant even while the industries' locations are changing greatly. In this section, we develop a simple framework for thinking about how changes in agglomeration will result from a combination of the systematic growth and contraction of old industry centers and randomness in growth rates. We then use this framework to describe the movement of industries in our data.

4.1 A basic decomposition

Consider the following simple regression, in which we treat the change in state-industry employment shares (*e.g.* the share of industry i's employment located in state s) as a function of the growth of the state's average employment share (*i.e.* average across industries of the share of employment in state s) and the difference between initial state-industry share and the state's average employment share:

$$s_{is\,t+1} - s_{ist} = \hat{\alpha} + \hat{\beta}(s_{ist} - s_{st}) + \hat{\gamma}(s_{s\,t+1} - s_{st}) + \hat{\epsilon}_{ist}, \tag{1}$$

where as before s_{ist} is the share of industry i's employment in state s as of time t, s_{st} is the average of this variable for state s across industries, $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are estimated coefficients and $\hat{\epsilon}_{ist}$ is an estimated error term which is, by construction, orthogonal to each of the regressors. Note that this regression is specified so that each of the variables have mean zero and so that the two regressors are orthogonal. As a result, the OLS estimates will always be that $\hat{\alpha} = 0$ and $\hat{\gamma} = 1$.

In this section, we will analyze changes in agglomeration levels using the raw concentration component, G_{it} of the EG index of concentration. Write $G_t \equiv (1/I)\sum_i G_{it}$ for the mean of this variable across industries. Simple algebra reveals that

$$\begin{aligned}
G_{t+1} - G_t &= \frac{1}{I}\left[\sum_{is}(s_{is\,t+1} - s_{s\,t+1})^2 - \sum_{is}(s_{ist} - s_{st})^2\right] \\
&= \frac{1}{I}\left[\sum_{is}((1+\hat{\beta})s_{ist} - \hat{\beta}s_{st} + \hat{\gamma}(s_{s\,t+1} - s_{st}) + \hat{\epsilon}_{ist} - s_{s\,t+1})^2 - \sum_{is}(s_{ist} - s_{st})^2\right] \\
&= \frac{1}{I}\left[(2\hat{\beta} + \hat{\beta}^2)\sum_{is}(s_{ist} - s_{st})^2 + \sum_{is}\hat{\epsilon}_{ist}^2\right] \\
&= (2\hat{\beta} + \hat{\beta}^2)G_t + \frac{1}{I}\sum_{is}\hat{\epsilon}_{ist}^2.
\end{aligned}$$

This equation decomposes changes in concentration into the sum of two terms, which we will describe as the effects of "mean reversion" and of "randomness" (or dispersion) in the local employment process. The first term in the decomposition, $(2\hat{\beta} + \hat{\beta}^2)G_t$, depends on the extent to which net change in employment is correlated with the initial gap between the state-industry employment share and the state's mean employment share. When $\hat{\beta}$ is negative, current centers of the industry are tending to decline in importance and/or employment is tending to increase in areas which initially have a below average share of employment in the industry. In this case, we will describe the state-industry employment levels as displaying mean reversion, and the first term in the decomposition is the decrease in agglomeration attributable to this tendency. Conversely, when $\hat{\beta}$ is positive and industry centers are growing, the first term reflects a consequent increase in agglomeration.

The second term in the decomposition, $\sum_{is} \hat{\epsilon}_{ist}^2$, captures the effect of randomness in the growth (and decline) in state-industry employments. The magnitude of this component reflects the degree of heterogeneity in the experience of states which initially have similar concentrations of employment in a given industry. For example, it will be larger if some industry centers take off while others fail, and if some areas where the industry is not present are very successful in attracting new plants while others are not. Note that this term is always positive, so that the randomness of the employment process can always be thought of as tending to increase agglomeration levels. For the overall level of agglomeration to have remained roughly constant in U.S. manufacturing over the last 20 years, it must be that $\hat{\beta}$ is negative so that the agglomerating effect of random shocks has been counterbalanced by systematic mean reversion.

The role of mean reversion of state-industry employment and random shocks in maintaining a steady level of concentration over time is analagous to the classic discussion in statistics courses of the fact that for the distribution of people's heights to remain roughly the same over time it must be the case that children of tall parents are on average shorter than their parents and children of short parents are on average taller than their parents.

4.2 Evidence on U.S. manufacturing industries: 1972 - 1992

Table 3 contains parameter estimates for the state-industry employment change regression (1) for different subsamples of industries. Observations from all four periods have been pooled in each subsample. The first column gives the average EG index of each subgroup. The second gives the correlation between each state's shares of each industry's employment in 1972 and 1992. Note that these are quite high, though not as high as the 0.92 correlation between 1972 and 1992 values of the EG index. For the entire sample, the estimated coefficient $\hat{\beta}$ on the departure of the state-industry employment share from the average employment share is -.063. Hence, over the 20 year period, states in which an industry is initially overrepresented in an area would be expected to experience a decline in its excess employment of nearly one-fourth.

The sixth and seventh columns of the table present our decomposition of how much of the change in concentration is attributable to mean reversion and to dispersion in the employment change process. In the full sample, the mean reversion effect is sufficiently strong so as to produce a 12 percent decrease in agglomeration every five years, while the dispersion effect by itself would lead to a 9.6 percent increase. The effects almost balance each other and result in the net negative of -2.4 percent. We think of the magnitudes of the two separate effects as indicating that there is a substantial degree of industry mobility relative to the degree to which concentration levels have changed. This view constrasts with the emphasis of some previous authors (*e.g.* Krugman (1991a)) on historical accident.

The results for the various subsamples indicate that there are clear differences in the employment change patterns in different groups of industries. On the high concentration subsample (which contains the most agglomerated third of industries according to the 1972 EG index), we estimate a $\hat{\beta}$ which is smaller than that of the overall sample, although perhaps surprisingly there seems to be no more or less randomness in the growth process of this subsample (as measured by $\hat{\sigma}$) than in the full sample. The low concentration subsample (which contains the least concentrated third of the industries in the sample) is distinguished by having much stronger mean reversion; the estimated $\hat{\beta}$ of -0.116 indicates that on average areas where an industry was overrepresented saw their excess employment reduced by about 40% over the twenty year period.

The final four lines of the table are concerned with the behavior of four sets of industries with moderate to high concentration. We constructed these samples manually in an attempt to categorize the industries which appeared at the top of our concentration list. The four samples include the majority of the industries in our ”geographically concentrated” sample (as well as some others which are only slightly less concentrated). The one subsample which really stands out in the table is the set of textile industries (consisting of industries within SIC groups 22 and 23 for which the 1972 EG index was in the top 60). In these industries, there is much less mean reversion than in the other concentrated industries, and the degree of concentration has risen over time. In each of the other subsamples, raw concentration levels have declined. The behavior of the high technology subsample is fairly similar to the full sample with the one notable difference being that there is a larger random component in the employment changes.[7] (This randomness, however, is not sufficiently large so as to sustain the high initial level of agglomeration.) The set of industries where we imagined natural resources may be relevant to agglomeration (including several food, lumber, paper, petroleum and primary metals industries) appears to have much less randomness in the growth and decline of state-industries than do the others.[8] The pattern in the craft industries is quite similar to the pattern in the full sample.[9]

Table 4 reports separate decompositions of the decline in raw concentration into components stemming from mean reversion and from dispersion for each of the four five-year intervals in our sample. In all four periods, concentration declines, but the change has been most pronounced since 1982. The dispersion effect has been larger in the second half of the period than it was in the first, and thus the more rapid decline in agglomeration which has taken place in the last half of the period can be more than completely attributed to an increase in the rate at which old industry centers have tended to decline.

5 Agglomeration and the plant life cycle

A new geographic center of activity in an industry can develop in a number of different ways: the area can be a hotbed for startup firms; it may succeed in attracting the new plant investments of existing firms; a core of preexisting firms may grow into a position of prominence. In this section, we attempt to provide for the first time a systematic look at where in the life cycle of plant births, expansions and closures geographic concentration comes from. We begin by describing how changes in a measure of agglomeration may be decomposed into portions due to various life cycle events, and then use the decomposition to describe patterns found in various sets of U.S. manufacturing industries.

[7]The concentrated high technology sample consists of SICs 357, 365, 372, 376, 385 and 386.
[8]The concentrated natural resource sample consists of SICs 203, 207, 241, 242, 243, 261, 262, 291, 331 and 332.
[9]The concentrated crafts sample consists of industries 326, 328, 387, 391, 393 and 396.

5.1 A life cycle decomposition of agglomeration changes

Suppose that the data allow the employment change in a state-industry to be partitioned into portions attributable to one of J categories of events (*e.g.* to new plant births, plant closures, etc.):

$$E_{is\,t+1} - E_{ist} = \sum_j \Delta E^j_{ist},$$

where E_{ist} is the employment level of industry i in state s at time t, and ΔE^j_{ist} is the change in employment due to events of type j. Denote by ΔE^j_{it} the aggregate amount of employment change due to the jth growth process.

In the previous section, we used a simple raw concentration measure of geographic concentration. We justified this by noting that it is an empirical fact that the overall distribution of plant sizes has not changed much over the last twenty years, so that there would be little benefit to considering the more complicated EG index. Such an argument, however, is not applicable to the problem of decomposing concentration changes into portions attributable to various life-cycle stages. While the overall plant-level concentration of industries has remained roughly constant (e.g. using the plant Herfindahl measure), new firm births clearly tend reduce this concentration while plant closures tend to increase it. Hence the impact of births or closures on raw concentration and on the EG index may be quite different.

To obtain a tractable decomposition, we focus on a measure of agglomeration, $\tilde{\gamma}$, which approximates the EG index:

$$\tilde{\gamma}_{it} \equiv \frac{G_{it}}{1 - \sum_s s^2_{st}} - H_{it}.$$

(The approximation involves ignoring the $1 - H_{it}$ denominator of the EG index.[10]) Writing $G_t \equiv \frac{1}{I}\sum_i G_{it}$ for the average raw concentration across industries, we first define in equation (3) below a decompositon of the aggregate change in raw concentration in to portions attributable to events in each of the J categories, *i.e.* we define measures ΔG^j_t such that $G_{t+1} - G_t = \sum_{j=1}^{J} \Delta G^j_t$. Next in equation (4), we define a similar decomposition of changes in the average plant Herfindahl, $H_{t+1} - H_t = \sum_{j=1}^{J} \Delta H^j_t$. Our final measure of the portion of the change in the index $\tilde{\gamma} \equiv \frac{1}{I}\sum_i \tilde{\gamma}_{it}$

[10]Because the changes in H attributable to the various stages of the life cycle typically have a magnitude of about 0.001, this should have a very small impact on the changes in the agglomeration index attributed to the various stages.

attributable to the jth stage of the life cycle is then just

$$\Delta\tilde{\gamma}^j \equiv \frac{\Delta G_t^j}{1-\sum_s s_{st}^2} - \Delta H_t^j.$$

To decompose raw concentration changes, we first define a measure Δs_{ist}^j of the portion of the change in state s's share of employment in industry i which is due to the jth growth process by

$$\Delta s_{ist}^j \equiv \frac{\Delta E_{ist}^j - s_{ist}\Delta E_{it}^j}{E_{i\,t+1}}.$$

The numerator of the right hand side contains a difference of two terms: the actual employment change due to the events of the jth type and the employment change that would have resulted if events of the jth type had occurred in proportion to initial state-industry employment. The denominator is end-of-period industry employment. If, for example, new firm births in an industry occured proportionally to the initial state-industry employments (and there were no differences in the size of new firms across states), then there would be no changes in state-industry employment shares caused by new firm births, and $\Delta s_{ist}^{\text{new birth}}$ would be zero for each state. Because state-industry shares change nonlinearly with changes in each plant's employment, our partition of share changes into portions attributable to each of several factors is by necessity arbitrary. We believe, however, that this definition seems natural, and it satisfies the crucial adding up constraint: $s_{ist} - s_{st} = \sum_j \Delta s_{ist}^j$.

We then estimate for each of the J categories of employment changes a growth regression of the form

$$\Delta s_{ist}^j = \hat{\alpha}_j + \hat{\beta}_j(s_{ist} - s_{st}) + \hat{\gamma}_j(s_{s\,t+1} - s_{st}) + \hat{\epsilon}_{ist}^j. \tag{2}$$

The estimates from these regressions will satisfy $\sum_j \hat{\beta}_j = \hat{\beta}$, $\sum_j \hat{\gamma}_j = \hat{\gamma}$, and $\hat{\epsilon}_{ist} = \sum_j \hat{\epsilon}_{ist}^j$, where $\hat{\beta}, \hat{\gamma}$, and $\hat{\epsilon}_{ist}$ are the estimates from the employment change regression (1) of the previous section. The aggregate change in raw concentration is related to the parameters of these regressions by

$$\begin{aligned} G_{t+1} - G_t &= (2+\hat{\beta})\hat{\beta}G_t + \frac{1}{I}\sum_{is}\hat{\epsilon}_{ist}^2 \\ &= \sum_j \left(\hat{\beta}_j(2+\hat{\beta})G_t + \frac{1}{I}\sum_{is}\hat{\epsilon}_{ist}\hat{\epsilon}_{ist}^j\right) \\ &= \sum_j \Delta G_t^j, \end{aligned}$$

where the change in raw concentration due to events in the jth category is defined by

$$\Delta G_t^j \equiv \left(\hat{\beta}_j (2+\hat{\beta}) G_t + \frac{1}{I} \sum_{is} \hat{\epsilon}_{ist} \hat{\epsilon}_{ist}^j \right). \tag{3}$$

As motivation for this definition note that ΔG_t^j is a sum of four terms:

$$\Delta G_t^j = (2\hat{\beta}_j + \hat{\beta}_j^2) G_t + \frac{1}{I} \sum_{is} \hat{\epsilon}_{ist}^{j\ 2} + \hat{\beta}_j \sum_{k \neq j} \hat{\beta}_k G_t + \frac{1}{I} \sum_{is} \hat{\epsilon}_{ist}^j \sum_{k \neq j} \hat{\epsilon}_{ist}^k.$$

The first two terms reflect the change in concentration due to the mean reversion and randomness in employment changes of type j. The third and fourth terms are what we thought was a natural allocation of the additional agglomeration changes which result from correlations between employment changes due to events of type j and due to events of other types.

The decomposition of changes in the plant Herfindahl index into portions due to events at each stage of the life cycle is analogous with plant-industry employments taking the role of state-industry employments. Let e_{ikt} be the employment level in the kth plant in industry i at time t. We write $z_{ikt} \equiv e_{ikt} / \sum_k e_{ikt}$ for plant k's employment share within its industry so that our plant Herfindahl measure for industry i can then be written as $H_{it} = \sum_k z_{ikt}^t$. We write $H_t = \frac{1}{I} \sum_i H_{it}$ for the average of this measure across industries. We write $\Delta z_{ikt} \equiv z_{ikt+1} - z_{ikt}$ for the change in that share over time, with the convention that z_{ikt} is set to zero if plant k is not in industry i at time t. (For example, $\Delta z_{ikt} = -z_{ikt}$ if the k plant in industry i has switched to another industry at time $t+1$.)

Assume again that we have a partition of employment changes in each plant-industry into portions attributable to events in each of J categories:

$$\Delta e_{ikt} = \sum_j \Delta e_{ikt}^j.$$

As before, we define the portion of the change in each plant's share of employment to events in each category by

$$\Delta z_{ikt}^j = \frac{\Delta e_{ikt}^j - z_{ikt} \Delta E_{it}^j}{E_{it+1}}.$$

This definition again yields an allocation of share changes across the categories, *i.e.* $z_{ikt+1} - z_{ikt} = \sum_j \Delta z_{ikt}^j$. We then estimate separate employment change regressions for each of the J components of changes in the plant-industry employment shares,

$$\Delta z_{ikt}^j = \hat{\alpha}_j + \hat{\beta}_j \tilde{z}_{ikt} + \hat{\epsilon}_{ikt}^j,$$

where $\tilde{z}_{ikt} \equiv z_{ikt} - \frac{1}{N_i}$ for N_i is the number of plants which operate in the industry either in period t or in period $t+1$.

We show in Appendix A that the definition analagous to our definition of ΔG_t^j,

$$(4) \qquad \Delta H_t^j \equiv (2\hat{\beta}^j + \hat{\beta}^j \hat{\beta})(H_t - \frac{1}{I}\sum_i \frac{1}{N_i}) + \frac{1}{I}\sum_{ik} \hat{\epsilon}_{ikt}^{j\ 2} + \frac{1}{I}\sum_{ik} \hat{\epsilon}_{ikt}^j \left(\sum_{\ell \neq j} \hat{\epsilon}_{ikt}^\ell \right),$$

again provides a decomposition which satisfies

$$H_{t+1} - H_t = \sum_j \Delta H_t^j.$$

5.2 Evidence on the life cycle decomposition

In this section, we discuss how events at various stages of the plant life cycle have contributed to the geographic concentration of U.S. manufacturing industries. Our analysis of data from the LRD focuses on a description of employment changes as resulting from five categories of events: births of new firms, openings of new plants by existing firms, the growth or decline in employment at existing plants which continue to operate in the industry, plant closures, and switches of plants from one industry to another. Among our most general conclusions are the new firm births and expansions of existing plants have a deagglomerating effect while the plant closure process tends to reinforce concentration levels.

5.2.1 Overall patterns

Table 5 lists the coefficient estimates from regressing each component of employment change, Δs_{ist}^j, on the initial excess concentration of the industry in the state, $s_{ist} - s_{st}$, and on the overall growth of the state, $s_{st+1} - s_{st}$ as in equation (2). For the new firm birth and old firm birth regressions, the coefficient on $s_{ist} - s_{st}$ is negative indicating that there is mean reversion in employment shares: employment in new plants (especially those belonging to new firms) increases less than one-for-one with state-industry employment. The coefficient on initial excess employment is positive in the closure regression, indicating that plants are less likely to close in states which have a higher than expected share of the industry's employment. (Note that the dependent variable is negative in this regression.) For expansions and contractions the $\hat{\beta}$ coefficient is also negative, implying that growth rates are lower in states with a high initial concentration in the industry.

New firms are more likely to start away from current geographic centers of the industry, and growth is faster away from those centers, but the risks appear to be higher in the periphery and closures are also higher there.

Table 6 reports the portions of the change in the approximate Ellison-Glaeser index, $\tilde{\gamma}_t$ which we attribute to each of the life cycle stages. These changes are listed as a percentage of initial concentration in the set of industries, *i.e.* the table reports values of $100\Delta\tilde{\gamma}_t^j/\tilde{\gamma}_t$. Again, new firm births consistently have the effect of reducing the degree to which industries are geographically concentrated. On its own, the deagglomerating effect of new firm births can account for about three-fourths of the observed decline in geographic concentration over the last 20 years. While plants opened by preexisting firms are comparable to new firm births in their total employment, they have less of a deagglomerating effect — they only reduce geographic concentration in three of the four five year periods, and on average their effect is only a little more than one-third as large as the effect of new firm births.

Consistent with our previous observation that net expansions are lower in areas with an excess concentration of employment in an industry, we find that net expansions also tend to reduce geographic concentration. The magnitude of his effect is roughly comparable to that of new firm births. The one growth process which appears to reinforce geographic concentration is the closure process. Given that plant closures have not been explicitly discussed in the existing theoretical literature on geographic concentration, this result particularly merits attention.

When interpreting the results on expansions and closures, it should be kept in mind that the regression on which our decomposition is based does not have controls for differences in the age and size of plants. Plants in industry centers tend to be older and larger than the average plant in an industry, and hence one might expect them both to grow more slowly and to be less likely to close. We discuss how accounting for this may change one's view of the effect of these stages in section 5.2.4.

5.2.2 Changing patterns over time

One of our initial observation from Table 1 was that the trend toward industries being less geographically concentrated has been more pronounced in the second half of our sample than in the first. The average across industries of the value of the EG index was 0.039 in 1972, 0.038

in 1982, and 0.034 in 1992. From Table 6 we see that this change is largely attributable to the fact that differing plant closure rates ceased to reinforce initial concentrations and to an increased tendency for plant expansions to occur away from industry centers. The effect of new plant births has been fairly constant throughout the period we study.

5.2.3 Differences across industries

Table 7 reports the results of life-cycle decompositions of agglomeration changes for various subsets of industries. In each case we report the average of the effect in each of the four five-year periods.[11]

The first row of Table 7 repeats the average results from Table 6. The second and third rows look separately at the most geographically concentrated one-third and the least geographically concentrated one-third of industries (in terms of 1972 values of the EG index). Our main conclusion here is that the full sample results are also representative of what has happened within the highly geographically concentrated industries. In the nonlocalized industries, new firm births and expansions have not had a deagglomerating effect, and levels of geographic concentration have increased (albeit only slightly given that the base to which the percentages apply is very small.)

The final four rows of the table describe the behavior of the same four subsamples of the set of highly geographically concentrated industries which we discussed in section 4.2 in connection with Table 3. The life cycle pattern of the location of the high tech industries differs somewhat from the overall pattern. Both the openings of new plants by existing firms and the net expansions which have occured in existing plants have had stronger deagglomerating effects here than in the average industry. There also seems, however, to have been a greater difference between plant closure rates in and away from industry centers, so, in the aggregate, geographic concentration in these industries has decreased only a bit more than in the typical industry.

We noted earlier that the textile and apparel industries stand out for having become more geographically concentrated over the last twenty years. The largest part of this difference is attributable to net expansions acting to reinforce geographic concentration, with plant births (to new and old firms) also having less of a deagglomerating effect than in the full sample.

[11]Note that when measuring the concentration of industries within a subsample we still use the same measure s_{st} of state size as in the full sample. As a result, the decompositions within a subsample will not have the same adding up property as they do in the full sample.

5.2.4 Plant age and size effects

Plants located in industry centers are on average older and larger than the other plants in their industry. Our decompositions will therefore regard closures as reinforcing concentration even if the only reason why this is true is the general tendency of older, larger plants to be less likely to close. The effects of age and size on expansion will similarly make expansions deagglomerating, even if the net expansion is otherwise independent of a plant's location.

To help clarify why the expansion and closure processes have the effects they have on levels of geographic concentration, we present in Table 8 estimates of the regression

$$\frac{\Delta E_{ist}^j}{E_{it}} = \lambda_0 + \lambda_1 s_{ist} \frac{\Delta E_{it}^j}{E_{it}} + \lambda_2 s_{st} \frac{\Delta E_{it}^j}{E_{it}} + \lambda_4 AvgPlantSize + \lambda_5 FracAge04 + \lambda_6 FracAge59 + \epsilon_{ist},$$

where $AvgPlantSize$ is the average employment of plants in the state-industry at time t, and the last two variables are the fraction of plants in the state-industry which appear for the first time in the year t and year $t-5$ censuses, respectively. In this specification, a coefficient of one on $s_{ist}\frac{\Delta E_{it}^j}{E_{it}}$ would indicate that plant births, closures, etc., occur proportionally to initial employment in the state-industry (after controlling for age and size effects). At the other extreme, a coefficient of zero on $s_{ist}\frac{\Delta E_{it}^j}{E_{it}}$ would indicate that the events in question are independent of any departures of state-industry employment from overall state employment.

After adding the controls for plant age and size, plants in industry centers still appear to be less likely to close, but our previous conclusions about the effects of expansions seem to be entirely due to diffences in the age and size of plants in different areas. If anything, proportional expansion is greater in areas with a greater initial concentration of employment in an industry.

While the 0.63 coefficient on $s_{ist}\frac{\Delta E_{it}^j}{E_{it}}$ in the new births regression is less than one, it is also far from zero. Hence, while the fraction of new plant births located in an industry center tends to be less than the area's share of employment in the industry, it is much greater than would be predicted simply from the aggregate employment of the area. As a result, even though new plant births tend to reduce geographic concentration, they reduce it by far less than they would if the locations of new plants were not correlated with initial state-industry employment. In this sense, new firm births can also be regarded as causing geographic concentration, and examining them further will be an important part of our assessment of theories of geographic concentration.

Table 9 reports estimates from a similar regression where the coefficients on $s_{ist}\frac{\Delta E_{it}^j}{E_{it}}$ and

$s_{st}\frac{\Delta E_{it}^{j}}{E_{it}}$ were allowed to vary by two-digit industry. Note first that in the median two-digit industry new plant births are further from proportional to state-industry employment than our full sample estimates. The apparel industry (SIC 23) stands out for its concentration of new firm births in industry centers. The pattern of closure rates being lower in areas where an industry is concentrated appears to be very widespread and is most pronounced in the furniture and printing and publishing industries (SICs 25 and 27). The pattern of net expansions being greater in industry centers (after controlling for plant age and size effects) also seems to be very consistent, with the apparel industry standing out as the one clear exception.

6 Theories of industrial location

In this section, we use data from the LRD to assess the importance of three of Marshall's (1920) theories of industry agglomeration. While not explicitly dealing with dynamics, each of the theories identifies a benefit that firms may receive from locating near other firms in the same industry. Hence, they share a primary conclusion — in equilibrium industries will be geographically concentrated.

The approach we will take in trying to distingush between the theories is based on the observation that while their central conclusions are the same, the theories differ in their implications for which firms in other industries a firm may wish to be near. We thus focus on the cross-industry coagglomeration patterns in our data, controlling for the general tendency of plants to locate (or expand or not close when located) near others in the same industry and asking if in addition firms seem to locate or expand operations in areas where there is a great deal of activity in other industries which the theories suggest could provide coagglomeration benefits. While the tests will focus on coagglomeration patterns, our primary motivation is the belief that they will be informative also as to which forces provide the benefits of intra-industry agglomeration.

6.1 The theories

We first review quickly the theories of agglomeration in Marshall (1920) and the variables available to us which may be of use in testing them.

6.1.1 The presence of suppliers and customers

Marshall (1920) argues that transportation costs should induce plants to locate close to their inputs, close to their customers, or most likely at some point optimally trading off distance between inputs and customers. While quantifying this could involve a very complex measure involving the location of not just firms in particular area but also the location of all firms in neighboring areas, for simplicity, our measures of nearby supplier/customer presence will focus solely on their presence in a metropolitan area.

Our measure of input supplier presence for industry i in state s at time t is

$$Input_{ist} \equiv \sum_{j \neq i} I_{ji} \frac{E_{jst}}{E_{jt}},$$

and our measure of product customer presence is

$$Output_{ist} \equiv \sum_{j \neq i} O_{ij} \frac{E_{jst}}{E_{jt}},$$

where I_{ji} is the share of industry i's inputs that come from industry j, O_{ij} is the share of industry i's outputs that go to industry j, E_{jst} is industry j's employment in state s, and E_{st} is total employment in state s. These measures run from 0 (which means that no input supplying or output purchasing industries are in that state) to 1 (which is possible only if all of the input suppliers or output purchasers of the industry are located in the state).

One possible view is that the inputs or outputs that really matter are those that are not vertically integrated. Chinitz (1961) argued that the strength of New York City comes from the fact that input suppliers are themselves single-unit plants that are not tied to upstream or downstream firms and are available to either sell to or buy from new entrepreneurs. To examine this hypothesis, we will also look to see whether the presence of customers and suppliers who are not part of a multi-plant firm has a greater effect than the presence of customers or suppliers in general. The Chinitz (1961) view can also be examined by comparing the importance of suppliers in new firm births (where unconnected suppliers should be particularly important) and old firm births (which may have pre-existing supplier relationships).

As a potential indicator or where transportation costs may be a more important concern, we also obtained from the 1977 Census of Transportation the weight per dollar value of the industry's shipments. We call this variable $Weight_i$.

6.1.2 Labor market pooling

Marshall's second theory of industrial location is that firms locate near one another to shield workers from the vicissitudes of firm-specific shocks. Workers are willing to accept lower wages in locations where other firms stand by ready to hire them (see Diamond and Simon (1990) for evidence and Krugman (1991a) for a formalization). Rotemberg and Saloner (1991) present an alternative theory in which workers gain not because of insurance from shocks, but because multiple firms protect workers against ex post appropriation of investments in human capital. Both theories predict that plants that use the same type of workers will locate near one another.

To test this theory, we use a measure based on the similarity of the labor requirements of an industry with the composition of a state's labor pool outside of that industy. Specifically, we measure the suitability of the labor mix in an area with:

$$LaborMix_{ist} \equiv -\sum_{o}\left(L_{io} - \sum_{j\neq i}\frac{E_{jst}}{E_{st} - E_{ist}}L_{j0}\right)^{2},$$

where o indexes occupations and L_{io} is the fraction of industry i's employment in occupation o. The measure can thus be understood as a sum of squared deviations measure of the similarity between the occupation mix desired by industry i and an estimate of the composition of the area's labor force obtained by taking a weighted average of the average of the typical employment patterns of the other manufacturing industries located there. The variable is scaled so that better matches correspond to higher values, with the maximum of zero being achieved if industy i has exactly the same fraction of its employment in each of the 277 occupations as the average manufacturing industry in the area.

To assess the Marshallian version of the labor pooling hypothesis, we will interact this variable with $Closure_{it}$, a measure of risk in the industry. This measure is the rate of employment loss due to closures in the industry as a whole, calculated as $-\Delta E_{it}^{\text{closure}}/E_{it}$.

While our intention is for the $LaborMix_{it}$ variable to reflect labor market conditions, we should note that it could also capture other shared attributes of the industry, since industries that use the same type of workers could be similar along other dimensions. For example, industries that share workers may also be industries between which there is a greater possibility for intellectual spillovers.

6.1.3 Information spillovers

Firms may also locate where they are likely to learn from other firms. This learning can take the form of workers learning skills from one another (as Marshall argued) or industrial innovaters copying each other (as is reportedly the case in Silicon Valley, see Saxenian, 1994). Firms will group near one another either because of the gains from continued presence or because the idea leading to the opening of a new plant came from an existing concentration of employment in nearby plants.

Because it is difficult to observe and measure patterns of information spillovers, information spillover theories are the most difficult to assess empirically. In this paper, we employ two measures which we hope might reflect such spillovers. First, using a technology flow matrix constructed by Scherer (1984), we construct a variable measuring the extent to which different industries use each other's technological innovations. Scherer's matrix estimates the extent to which R&D activity in one industry flows out to benefit another industry either through a supplier-customer relation between these two industries or based on the likelihood that patented inventions obtained in one industry will find applications in the other industry. Specifically, we suppose that the information spillover benefits which a plant in industry i receives when locating in area s are captured by

$$Techflow_{ist} \equiv \sum_{j \neq i} T_{ji} \frac{E_{jst}}{E_{jt}},$$

where T_{ji} reflects the technology flows from industry j to industry i Scherer (1984) estimated from the R&D and patent data.

Since technology flows capture only one particular form of intellectual flows, we will also use a more indirect and tenuous measure based on co-ownership of plants across industries. The idea is that co-ownership occurs, at least in part, due to economies of scope such as those associated with the sharing of ideas within a firm. Using the LRD we have created a co-ownership matrix W_{ij} which captures the extent to which plants in industry i are integrated with plants in industry j It is equal to the fraction of industry i's employment which is contained in firms that also own plants in industry j. Clearly many factors other than information spillovers can lead firms to operate in multiple industries; in order to reduce the degree to which a co-ownership measure might reflect input-output relationships instead we set W_{ij} to zero whenever one industry in the pair accounts for more than 5% of all inputs used in the other industry at the national level.

As a measure of the degree to which area s has a lot of activity in industries which may provide information spillover benefits to firms in industry i we then define

$$Integration_{ist} \equiv \sum_{j \neq i} W_{ij} \frac{E_{jst}}{E_{jt}}.$$

As a potential inicator or where intellectual spillovers may be more important, we will also interact this variable with a measure of the fraction of the employment in the industry which is contained in occupations which we imagined would normally require a college degree, $College_i$.

6.2 Empirical framework

As discussed earlier, our approach to assessing the importance of the theories of agglomeration exploits the fact that the theories make different predictions about which pairs of industries will tend to coagglomerate. For example, labor market pooling predicts that firms should gain from locating near plants in other industries which employ workers with similar skills, while the technological spillover theory predicts that industries which are suited to cross-fertilization of ideas may coagglomerate. Because the different theories may also be more or less relevant at different stages of the plant life cycle, we will estimate in separate regressions whether employment changes at each stage follow the pattern predicted by the theory. We examine employment changes using two main regression specifications. The first is a linear growth regression which fits well with the previous decomposition:

$$\begin{aligned} \frac{\Delta E_{ist}^j}{E_{it}} &= (\lambda_0 + \lambda_1 s_{ist} + \lambda_2 s_{st} + \beta_0 Input_{ist} + \beta_1 Output_{ist} + \beta_2 LaborMix_{ist} \\ &\quad + \beta_3 TechFlow_{ist} + \beta_4 Integration_{ist}) \frac{\Delta E_{it}^j}{E_{it}} + \delta_i + \xi_s + \epsilon_{ist}. \end{aligned}$$

The dependent variable $\Delta E_{ist}^j / E_{it}$ is the change in industry i's employment in area s between time t and time $t+1$ due to events of type j, expressed as a proportion of initial employment in the industry. The right hand side variables are: two variables, s_{ist} and s_{st}, which control for state sizes and for the tendency of plants to locate near other plants in the same industry; the variables designed to reflect the coagglomeration preditions of the theories (interacted with the overall employment change in the industry due to the jth factor); and state and industry fixed effects. In the closure and expansion regressions, we also include as controls the log of the average plant size in the state-industry and the shares of initial employment in the state-industry in plants that are less than five and between five and ten years old, respectively. Previous work (e.g. Dunne, Roberts and Samuelson (1989a)) has shown that plant age and plant size are important correlates of the probability of closing and the potential for plant growth.

In our data, the linear growth regressions have the disadvantage of placing a great deal of weight on the number of large outliers which are present. For this reason, we will instead focus on an alternate specification which uses $\log(1+\Delta E^j_{ist})$ as the dependent variable.[12] This transformation improves the fit of the regression equation and the robustness of the results. We add one to the employment change in order not to discard any observations. The basic functional form is:

$$\begin{aligned}\log(1+\Delta E^j_{ist}) &= \lambda_0 + \lambda_1 \log((1+E_{ist})|\Delta E^j_{it}|/E_{it}) + \lambda_2 \log(s_{st}|\Delta E^j_{it}|) \\ &\quad +\beta_0 Input_{ist} + \beta_1 Output_{ist} + \beta_2 LaborMix_{ist} + \beta_3 TechFlow_{ist} \\ &\quad +\beta_4 Integration_{ist} + \delta_i + \xi_s + \epsilon_{ist}.\end{aligned}$$

To test whether the importance of each of the theories varies across industries in the expected manner, we also estimate the logarithmic specifications with a number of interactions as mentioned above.

In both specifications, we have tried to simplify the interpretation of the results by defining each of the variables so that the theoretical prediction is that the signs should be positive. We also divide all of our explanatory variables and our dependent variables by their standard deviations, so that the coefficients can be interpreted as measuring by how many standard deviations the component of employment change increases when the explanatory variable increases by one standard deviation.

A potential problem with any employment change regression is that if initial employment levels are in equilibrium, the benefits of further agglomeration will be exactly counterbalanced by whatever forces keep employment from concentrating at a single location. Hence, our tests could fail to find evidence of the benefits of agglomeration. Given that there are large random shocks to industries and that (some) plants are fairly long-lived, we feel that spatial distributions are likely to be sufficiently far from equilibrium to make our approach useful. We do feel though that it may be informative to look also at the cross-industry agglomeration patterns which exist in the initial employment levels. We, thus, include at the end of this section a regression like the

[12] In the case of closures we use $-\log(1-\Delta E^j_{ist})$ as the dependent variable.

logarithmic specification but with the logarithm of initial employment levels ($\log(1 + E_{ist})$)) as the dependent variable.

6.3 Results

In this section, we present the results obtained from estimating the regressions described above on the LRD data. In our base specification, we use metropolitan areas as the geographic unit of observation and estimate the regressions one the full pooled set of five year changes from 1972-1977, 1977-1982, 1982-1987 and 1987-1992. Our results strongly support the importance of labor market explanations for agglomeration, although we find some significant evidence for each of the theories.

Table 10 presents the coefficient estimates from our linear specification of employment growth. The births regressions have 164,552 observations, while the expansion/contraction and closure regressions only use 85,588 because we have left out state-industries with zero initial emploment (where the employment changes due to expansions/contractions and closures can only be zero.) All regressions include state, industry and year fixed effects, and we present standard errors which allow for correlation in the errors within a state-industry. In the lower part of the table, one can see that the control variables, $s_{ist}\Delta E^j_{it}/E_{it}$ and $s_{st}\Delta E^j_{it}/E_{it}$, are highly significant, and the estimates are quite similar to those reported in Table 8.

In the regression with employment changes due to new firm births as the dependent variable, most of the cross-industry variables have the expected positive signs, but the standard errors on most of the variables are sufficiently large so as to prevent us from drawing clear conclusions on which factors are and are not important. The $LaborMix$ variable is highly significant. The regression with employment changes due to plant openings by preexisting firms as the dependent variable suggests that proximity to downstream customers may be important for these firms, but, again, the results are inconclusive. In the closure regression, none of the explanatory variables are both significant and of the expected sign. None of the variables are significant in the expansion/contraction regression.

As we mentioned earlier, the linear specification might be expected to yield weak results given the large outliers which are present in the data. For this reason, we also estimated and will now focus on the log specification. The results, which are presented in Table 11, are much more consistent; most of the estimates are highly significant and have the expected signs.

The effects of the presence of input suppliers and customers are fairly modest. The effect of proximity to suppliers as captured by the $Input$ variable is insignificant in generating new firm births but appears to matter more in predicting the location of old firm births. The opposite appears to be true for our $Output$ measure of the presence of downstream customers, which is more important in explaining where new firms births are located. The magnitudes of these effects are fairly small; a one standard deviation increase in the present of input suppliers leads only to a 0.03 standard deviation increase in the presence of old firm births. In addition, in unreported regressions we did not find that these effects were any stronger in industries with higher transport costs (as measured by weight per dollar of shipment) or that the effects were stronger when the suppliers/customers were nonintegrated firms. Of course, part of the effect of these variables may be working through the initial level of employment in the state, and these regressions can only be interpreted as suggesting that input suppliers have only a negligible impact on new plant births over and above their effect on the "steady state" level of employment in the state.

Relative to the effect of input and output related industries, the importance of our $LaborMix$ measure of the suitability of an area's labor force is striking. New plants seem to have a very strong tendency to locate where there are other industries using the same kind of workers. A one standard deviation increase in the level of the labor mix variable causes a 0.18 standard deviation increase in the number of new plant births. This effect of labor mix is important for old firm plant births as well but somewhat smaller. A one standard deviation increase in the labor mix variable leads to an 0.13 standard deviation increase in the log of employment from new plant births. These effects appear to be somewhat stronger in industries with a higher closure rate, confirming the basic Marshallian hypothesis.

The effect of the $Integration$ variable, which is meant to be a proxy for some kinds of intellectual spillovers, is also quite significant. A one standard deviation increase in that variable leads to an 0.08 standard deviation increase in the log of new births. This effect is stronger for old plant births, and comparable to the effect of labor mix in those regressions. In addition, the effect of this integration variable is indeed stronger in those industries which employ more workers in occupations requiring a college education, suggesting that intellectual spillovers may be a more important determinant of location choices in idea-oriented industries. The effect of the $TechFlow$

measure of intellectual spillovers is much smaller, although it also is highly significant in the old firm birth regressions.

In the regression with employment changes due to plant closures as the dependent variable, only the $LaborMix$ and $TechFlow$ variables are significant and each is associated with more plant closings rather than fewer. While these are not effects which would tend to cause agglomeration, the $LaborMix$ result may not be inconsistent with Marshall's view of labor market pooling. While a geographically isolated firm which experiences a downturn in its demand or a negative productivity shock might be able to retain many of its workers at lower pay and exploit their occupation-specific human capital, one could imagine that a firm which shares a pooled labor market may be less able to retain its employees and be forced to shut down.

To provide a more complete picture of where the $LaborMix$ and $Integration$ results are coming from, Table 12 reports coefficient estimates from log regressions in which the coefficients on $LaborMix$ and $Integration$ are allowed to vary with the two-digit industry to which a three-digit industry belongs. Both variables seem to be very strong predictors of the locations of new and old firm births in the fabricated metals, industrial machinery, electonic and electrical equipment and instruments industries (SICs 34, 35, 36 and 38). This suggests that both the labor market pooling and intellectual spillovers hypotheses are particularly important for these more high technology industries. The variables are also both important for new firm births in the lumber and furniture industries (SICs 24 and 25), and $Integration$ is also important in the food and chemicals industries (SICs 20 and 28).

Table 13 repeats our log regressions at the state level. In both the new and old firm birth regressions, the presence of input suppliers now has a larger effect than does the presence of downstream customers. Taken together with the previous results, this suggests that suppliers are in general a more important consideration in location choice and that firms may only need to be reasonably close to suppliers while most of the benefits to locating from customers may come from locating in the same city. The integration variables are also, again, significant. By far, though, the dominant effect in these regressions is again the $LaborMix$ variable. The estimated magnitude of its effect here is larger than any of the effects in any of our previous regressions.

Finally, Table 14 reports estimates from regressions where the log of (one plus) the level of employment in a particular MSA-industry is the dependent variable. Since the presence of other industries is certainly a function of the location of one's own industry, these results are best interpreted as a reduced form establishing some stylized facts rather than establishing a causal link. The $Output$ variable is highly significant, as are the $Integration$ and $TechFlow$ variables. However, all of these variables are completely dominated by the importance of $LaborMix.$ A one standard deviation increase in this variable increases the level of steady state employment by 0.41 standard deviations.

7 Conclusion

This paper has argued that the concentration of industries is best viewed as a dynamic process in which the combination of plant births, closures and expansions/contractions act together to maintain a slow-changing level of industrial concentration. A primary finding of this paper is that the birth process, especially for new firms, acts to reduce concentration as the new plants are generally located away from established industry centers. This is partly balanced by the closure process which generally favors plants in agglomerated areas.

To examine these processes further and to understand the forces underlying agglomeration, we tested three different theories of industrial location. We found that the presence of input suppliers and customers is relatively unimportant in explaining why firms in different industries locate near one another. Intellectual spillovers appear to be somewhat more important, but the location process appears to be dominated by the labor mix of a particular area: plants do seem to locate near other industries when they share the same type of labor. This effect is quite large and suggests that labor market pooling is a dominant force in explaining the agglomeration of industry. Of course, this effect could potentially be occurring because industries with similar labor mixes share ideas as well as workers, and we leave further examination of that concern to later work.

References

Bartik, Timothy (1985): "Business Location Decisions in the United States: Estimates of the Effects of Unionization, Taxes, and Other Characteristics of States," *Journal of Business and Economic Statistics*, 1, 14–22.

Carlton, Dennis (1983): "The Location and Employment Decisions of New Firms: An Econometric Model with Discrete and Continuous Exogenous Variables," *Review of Economics and Statistics*, 65, 440–449.

Chinitz, Benjamin, (1961): "Contrasts in Agglomeration: New York and Pittsburg," *Papers and Proceedings, American Economic Review*, 279-289.

Creamer, Daniel (1943): "Shifts of Manufacturing Industries," in *Industrial Location and Natural Resources*. Washington, DC: US Government Printing Office.

Crihfield, John (1990): "Manufacturing Supply: A Long-Run Metropolitan View," *Regional Science and Urban Economics*, 20, 327–349.

Davis, Steven and Haltiwanger, John (1992): "Gross Job Creation, Gross Job Destruction and Employment Reallocation," *Quarterly Journal of Economics*, 107.

Davis, Steven, Haltiwanger, John and Schuh, Scott (1996): *Job Creation and Destruction*. Cambridge, Mass: MIT Press.

Diamond, Charles and Simon, Curtis (1990): "Industrial Specialization and the Returns to Labor," *Journal of Labor Economics*, 8.

Dunne, Timothy, Roberts, Mark and Samuelson, Larry (1989a): "The Growth and Failure of Manufacturing Plants in the U.S ," *Quarterly Journal of Economics*, 104.

Dunne, Timothy, Roberts, Mark and Samuelson, Larry (1989b): "Plant Turnover and Gross Employment Flows in the U.S. Manufacturing Sector," *Journal of Labor Economics*, 7

Ellison, Glenn and Glaeser, Edward L. (1997) "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach, " *Journal of Political Economy*, forthcoming.

Enright, Michael (1990): *Geographic Concentration and Industrial Organization*. Ph.D. thesis, Harvard University.

Florence, P. Sargant (1948): *Investment, Location and Size of Plant.* London: Cambridge University Press.

Fuchs, Victor (1962): *Changes in the Location of Manufacturing in the US Since 1929.* New

Haven: Yale University Press.

Glaeser, Edward L., Kallal, Hedi D., Scheinkman, Jose, and Shleifer, Andrei (1992): "Growth in Cities," *Journal of Political Economy*, 100, 1126–1152.

Henderson, J. Vernon (1988): *Urban Development: Theory, Fact and Illusion*. New York: Oxford University Press.

Holmes, Thomas (1995): "The Effects of State Policies on the Location of Industry: Evidence from State Borders," Federal Reserve Bank of Minneapolis Staff Report 205.

Hoover, E.M. (1948): *The Location of Economic Activity*. New York: McGraw Hill.

Jaffe, Adam, Trajtenberg, Manuel and Henderson, Rebecca (1993): "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108, 577–598.

Kim, Sukkoo (1995), "Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860-1987," *Quarterly Journal of Economics* 110, 881-908.

Krugman, Paul (1991a): *Geography and Trade*. Cambridge, Mass: MIT Press.

Krugman, Paul (1991b):"Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483–499.

Marshall, Alfred (1920): *Principles of Economics*. London: Macmillan.

McGuckin, Robert and Peck, SuZanne (1992): "Manufacturing Establishments Reclassified into New Industries: The Effect of Survey Design Rules" mimeo, Center for Economic Studies (U.S. Bureau of the Census).

Porter, Michael E. (1990): *The Competitive Advantage of Nations*. New York: The Free Press.

Saxenian, Annalee (1994): *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, Mass: Harvard University Press.

Scherer, Frederic et al. (1975): *The Economics of Multi-plant Operation: an International Comparisons Study*. Cambridge, Mass: Harvard University Press.

Scherer, Frederic M. (1984): "Using Linked Patent Data and R&D Data to Measure Technology Flows," in Griliches, Zvi (ed.) *R & D, Patents and Productivity*. Chicago, IL: The University of Chicago Press.

Schmenner, Roger, Huber, J., and Cook, R. (1987): "Geographic Differences and the Location of New Manufacturing Facilities," *Journal of Urban Economics*, 21, 83–104.

Appendix A

In this appendix we provide a more complete derivation of the life-cycle decomposition of changes in the average plant Herfindahl.

Let e_{ikt} be the employment level in the kth plant in industry i at time t. Define $z_{ikt} \equiv e_{ikt}/\sum_k e_{ikt}$, $\Delta z_{ikt} \equiv z_{ikt+1} - z_{ikt}$, and $\tilde{z}_{ikt} \equiv z_{ikt} - \frac{1}{N_i}$ where N_i is the number of plants which operate in the industry either in period t or in period $t+1$

If one runs a regression with employment changes within each plant-industry as the dependent variable on the sample of plant-industries which have positive employment either in period t or period $t+1$

$$(A1) \qquad z_{ikt+1} - z_{ikt} = \hat{\alpha} + \hat{\beta}\tilde{z}_{ikt} + \hat{\epsilon}_{ikt},$$

the estimates will always satisfy $\hat{\alpha} = 0$, $\sum_{ik} \hat{\epsilon}_{ikt} = 0$, and $\sum_{ik} \tilde{z}_{ikt}\hat{\epsilon}_{ikt} = 0$. (Note that we say plant-industry just to indicate that there's one observation on each plant for each industry to which it belongs in either time period of the pair; a plant which has switched industries between t and $t+1$ thus has its experiences recored in two observations in the regression.)

Changes in H_t are related to the regression coefficients by

$$H_{t+1} - H_t = (2\hat{\beta} + \hat{\beta}^2)\left(H_t - \frac{1}{I}\sum_i \frac{1}{N_i}\right) + \frac{1}{I}\sum_{ik}\hat{\epsilon}^2_{ikt}$$

To see this, we simply note that $\sum_k \tilde{z}^2_{ikt} = \sum_k (z_{ikt} - \frac{1}{N_i})^2 = H_{it} - \frac{1}{N_i}$ and then substiture the regression equation into the plant Herfindahl formula:

$$\begin{aligned} H_{t+1} - H_t &= \frac{1}{I}\sum_{ik} \tilde{z}^2_{ikt+1} - \tilde{z}^2_{ikt} \\ &= \frac{1}{I}\sum_{ik} 2\Delta z_{ikt}\tilde{z}_{ikt} + \Delta z^2_{ikt} \\ &= \frac{1}{I}\sum_{ik} 2(\hat{\beta}\tilde{z}_{ikt} + \hat{\epsilon}_{ikt})\tilde{z}_{ikt} + (\hat{\beta}\tilde{z}_{ikt} + \hat{\epsilon}_{ikt})^2 \\ &= \frac{1}{I}(2\hat{\beta} + \hat{\beta}^2)\sum_{ik}\tilde{z}^2_{ikt} + \frac{1}{I}\sum_{ik}\hat{\epsilon}^2_{ikt} \\ &= (2\hat{\beta} + \hat{\beta}^2)\frac{1}{I}\sum_{ik}(H_{it} - \frac{1}{N_i}) + \frac{1}{I}\sum_{ik}\hat{\epsilon}^2_{ikt} \end{aligned}$$

Let

$$\Delta z^j_{ikt} = \frac{\Delta e^j_{ikt} - z_{ikt}\Delta E^j_{it}}{E_{it+1}}.$$

If we estimate separate regressions for each of the J components of changes in the plant-industry employment shares,

$$\Delta z^j_{ikt} = \hat{\alpha}_j + \hat{\beta}_j\tilde{z}_{ikt} + \hat{\epsilon}^j_{ikt},$$

the estimates will be related to the estimates obtained from the regression (A1) of overall share changes by $\hat{\beta} = \sum_j \hat{\beta}_j$ and $\hat{\epsilon}_{ikt} = \sum_j \hat{\epsilon}^j_{ikt}$. The fact that the ΔH^j_t defined in (4) in the text satisfy $H_{t+1} - H_t = \sum_j \Delta H^j_t$ then follows immediately from the expression for $H_{t+1} - H_t$ given above.

Again, the definition can be motivated by regarding the formula as attributing to events in the jth category both the change in the plant Herfindahl which would have resulted if the those events were the only employment changes and a portion of the additional change in the Herfindahl which results from the correlation between events of the jth and other types. In thinking about correlation here (and if one wants to think about mean reversion and randomness) it is important to keep in mind that the relevant correlations here are only those at the level of the individual plant.

Appendix B

This appendix provides additional information on the data used in the paper.

Industries

The industry codes provided in the LRD are available in two forms, based on the SIC prevailing when the census was taken and based on the current (1987) SIC. We use the latter, so that industry definitions are consistent. Industries are aggregated to the 3-digit level, which includes 140 industry groups. Six of these were dropped: the four industries covering tobacco products (211, 212, 213 and 214), as well as the search and navigation equipment and fur goods industries (381 and 237). The tobacco industries were excluded because of the importance of industry switches among them, which probably do not reflect fundamental changes in the operations of the plants from which the state-industry employment numbers are derived. Similarly, there were large reported switches between the fur goods and women's outerwear industry which would have made the fur goods industry a large outlier in its concentration change. The search and navigation equipment industry was excluded because of a major discontinuity in employments over time, which might be due to recoding in the years prior to 1987 to make them consistent with the 1987 SIC.

$Input$ **and** $Output$

The I_{ji} and O_{ij} coefficients are calculated using the "Use of Commodities by Industries" table from the 1987 Benchmark Input-Output Accounts. This table provides numbers on the value of each group of commodities used as input in each industry at the national level. While some commodities can partly be produced by other industries than the one associated with these commodities, we ignore this distinction and therefore interpret the numbers from the table as providing an estimate of how much of an industry's production is used as an input to other industries. Since the industry groups used in producing the input- output tables differ from the SIC ones, some of the numbers had to be reallocated to make them consistent with our 3-digit SIC industries. In general, I-O industries consist of one or a group of 4-digit SIC industries. In most cases, these I- O industries don't overlap 3-digit SIC industries. In this case, the rows (columns) corresponding to the I-O industries are collapsed into a single row (column). If there is some overlap, the numbers for that I-O industry are allocated to the 3-digit SIC industries based on the total value of shipments of the 4-digit SIC industries that make up the I-O industry. A column of the resulting matrix, say i, says how much of the corresponding 3-digit industry's inputs come from the set of all 3-digit industries. The input coefficient, I_{ji}, is then equal to the (j,i) entry of the matrix divided by the sum of the entries in the ith column excluding the (i,i) entry, i.e. it is the fraction of industry i's inputs (excluding the inputs coming from within the industry) coming from industry j. To obtain the O_{ij} coefficient, we divide the (i,j) entry by industry i's total output (as calculated as the sum of all uses for the commodities comprised in industry i in the original I-O table, including final uses such as consumption, investment, etc.).

The $Input$ variable for a given industry-state, say (i, j) is calculated by summing over all other industries the products of the I_{ji} coefficients by the corresponding fractions E_{jst}/E_{jt}, which are equal to the shares of employment in all other industries that is located in state s. The O_{ij} variable is obtained in a similar fashion. For $Input\_SU$ and $Output\_SU$, we calculate the shares of employment E^S_{jst}/E^S_{jt} based on single-unit firms only. Otherwise, the calculations are the same.

$LaborMix$ **and** $College$

The L_{io} variable comes from the National Industry-Occupation Employment Matrix (NIOEM) for 1987, which presents employment numbers (at the national level) for a group of 277 occupations and 185 broadly-defined industries, including 51 manufacturing industries. Each 3-digit SIC industry is assumed to possess the same composition of employment (by occupation) as that of the NIOEM industry to which it belongs. L_{io} is simply the fraction of industry i's employment that is in occupation o. $College$ comes from the same source. Each of the 277 occupations was classified (based on the authors' best guess) as either requiring or not requiring a college degree. $College_i$ equals the fraction of employment in industry i that is part of these occupations.

$Techflow$

The T_{ji} numbers are derived from Table 20.1 of Scherer (1984). Each entry in that table is a dollar amount of 1974 R&D spending in a given industry that is assumed to flow out to benefit another industry. The conversion from the 38 broadly-defined groups of manufacturing industries reported there to our 3-digit industries was achieved by apportioning the number for a given entry in the table to the corresponding 3-digit industries based on the latter industries' total value of shipments (obtained from the 1987 Census of Manufactures). For instance, if T^*_{mn} is the entry in Scherer's table corresponding to the dollar flow of benefits from industry m to industry n, and j (resp., i) is a 3-digit industry that is part of industry group m (resp., n) and accounts for a fraction w_j (resp, w_i of all shipments in that industry group, then T_{ji} = $w_i w_j$ T^*_{mn}.

$Integration$

The employment numbers used to calculate the W_{ij} variable come from the LRD.

$Weight$

This measure of transportation costs for the industry is equal to the total weight of shipments (in tons) divided by their total value, as obtained from the 1977 Census of Transportation (the last year for which these numbers were available). The distance measure mentioned in the text is equal to ton-miles divided by total weight (which is a ton-weighted average of distance shipped).

Table 1: Mean levels of geographic concentration 1972 - 1992

	1972	1977	1982	1987	1992
Ellison-Glaeser index (γ)	.039	.039	.038	.036	.034
Raw concentration (G)	.049	.049	.049	.046	.045
Plant Herfindahl (H)	.013	.012	.012	.012	.013
Employment weighted mean γ	.038	.038	.037	.035	.034

Table 2: Correlation of Ellison-Glaeser index over time (1972-92)

	1972	1977	1982	1987
1977	.973			
1982	.967	.973		
1987	.918	.924	.969	
1992	.917	.925	.962	.975

Table 3: Pattern of raw concentration changes across industries

Set of Industries [number of industries in brackets]	Mean γ (1972)	Average correlation between 1972 and 1992 state shares	Estimates		Average five-year percent change in raw concentration		
			β	σ	Total	Mean Reversion	Disper-sion
Full sample [134]	.039	.86	-.062	.010	-2.4	-12.0	9.6
Geog. concentrated [45]	.088	.88	-.043	.010	-2.5	-8.4	5.9
Geog. unconcentrated [45]	.006	.86	-.116	.008	-0.1	-21.7	-21.5
Conc. high technology [6]	.103	.82	-.065	.013	-4.8	-11.9	7.0
Conc. natural resource [11]	.052	.90	-.059	.007	-5.7	-11.1	5.4
Conc. textile & apparel [14]	.111	.88	-.015	.010	2.1	-2.8	4.8
Conc. crafts [6]	.048	.79	-.064	.011	-1.6	-12.2	10.7

Table 4: Raw concentration changes and industry movement over time

Time Period	Percentage change in raw concentration		
	Total	Mean Reversion	Dispersion
1972-77	-1.3	-9.6	8.3
1977-82	-1.4	-10.3	8.9
1982-87	-4.3	-16.6	12.3
1987-92	-2.7	-11.7	9.0

Table 5: Employment changes at various life cycle stages

Independent variable	Dependent variables: components of employment share changes, Δs^j_{ist}					
	Total change	New firm births	Old firm births	Closures	Expansions/ Contractions	Switches
$s_{ist} - s_{st}$	-.062	-.023	-.018	.024	-.031	-.014
	(.002)	(.0004)	(.0006)	(.001)	(.001)	(.001)
$s_{st+1} - s_{st}$	1.000	.174	.198	.138	.488	.002
	(.023)	(.005)	(.007)	(.012)	(.014)	(.016)
R^2	.05	.11	.05	.02	.07	.00
$\sigma(\times 100)$	.95	.22	.31	.51	.56	.65

Standard errors in parentheses.

Table 6: Life cycle decomposition of changes in geographic concentration

Time period	Percent change in $\tilde{\gamma}_t$	Percent change in $\tilde{\gamma}_t$ attributed to:				
		New firm births	Old firm births	Closures	Expansions/ Contractions	Switches
1972-77	0.4	-2.1	-1.8	2.9	-0.7	2.2
1977-82	-2.7	-2.8	-0.8	2.6	-2.8	1.2
1982-87	-5.9	-2.3	-1.2	-0.1	-2.4	0.2
1987-92	-4.7	-2.7	0.2	0.6	-2.9	0.4
Average of estimates	-3.2	-2.5	-0.9	1.5	-2.2	1.0

Table 7: Life cycle decompositions for various subsets of industries

Set of industries	Average percent change in $\tilde{\gamma}_t$	Percent change in $\tilde{\gamma}_t$ attributed to:				
		New firm births	Old firm births	Closures	Expansions/ Contractions	Switches
Full sample	-3.2	-2.5	-0.9	1.5	-2.2	1.0
Geographically concentrated	-2.2	-3.1	-1.3	2.5	-2.6	1.2
Geographically unconcentrated	4.9	-0.0	1.6	1.3	0.4	2.4
Conc. high technology	-4.1	-1.8	-2.7	5.1	-5.6	0.8
Conc. natural resource	-8.3	-3.2	-0.7	0.4	-3.9	0.8
Conc. textile & apparel	0.9	-1.5	0.3	0.4	0.5	1.9
Conc. crafts	-0.1	-4.3	-0.7	2.5	-1.6	3.0

Table 8: Life cycle employment changes with plant age and size effects

Independent variables	Dependent variables: $\Delta E^j_{ist}/E_{it}$ New firm births	Old firm births	Closures	Expansions/ Contractions
$s_{ist}(\Delta E^j_{it}/E_{it})$	.63	.72	.88	1.16
	(8.6)	(4.0)	(18.4)	(16.9)
$s_{st}(\Delta E^j_{it}/E_{it})$	.39	-.01	.24	-.32
	(5.9)	(0.0)	(4.5)	(-5.2)
log(avg. plant size)			.07	-.02
			(4.2)	(-2.0)
Share 0-4 yrs old			-.02	.02
			(-8.4)	(3.7)
Share 5-9 yrs old			-.01	.02
			(-5.8)	(3.8)
Adjusted R^2	.51	.23	.57	.35
Number of obs.	164,552	164,552	85,598	85,598

Regressions at MSA level include MSA, industry, and year fixed effects. Estimated t-statistics in parentheses.

Table 9: Employment changes and initial state-industry employment by 2-digit industry

Industry	Dependent variables: $\Delta E^j_{ist}/E_{it}$							
	New firm births		Old firm births		Closures		Expansions/ Contractions	
	s_{ist}	s_{st}	s_{ist}	s_{st}	s_{ist}	s_{st}	s_{ist}	s_{st}
20. Food	.46	.49	.67	.03	.76	.24	1.11	-.14
22. Textiles	.57	.63	.99	.08	.84	.35	1.06	-.36
23. Apparel	1.11	.18	.71	.01	1.15	02	.80	.01
24. Lumber & wood	.55	.20	.81	-.13	.84	.11	1.11	-.10
25. Furniture & fixtures	.38	.52	.37	.53	.56	.56	1.14	-.00
26. Paper	.27	.66	.84	-.12	.84	.37	1.10	-.38
27. Printing & Publishing	.41	.55	.38	.63	.56	.72	1.16	-.50
28. Chemicals	.41	.51	.54	.16	.67	.37	1.07	-.18
29. Petroleum & coal	.44	.34	.61	.29	.88	-.03	1.08	-.07
30. Rubber & misc. plastics	.39	.61	.32	.37	.95	.34	1.32	-.58
31. Leather	.47	.87	.94	-.21	1.00	.30	.99	-.42
32. Stone, clay, & glass	.20	.64	.25	.35	.65	.35	1.32	-.38
33. Primary metals	.50	.47	.38	.21	.98	.13	1.10	-.17
34. Fabricated metal products	.58	.29	.56	.03	.80	.22	1.19	-.25
35. Industrial machinery & equip.	.69	.17	.85	-.12	.85	.27	1.12	-.30
36. Electronic & electric equip.	.44	.59	.36	.54	.60	.67	1.17	-.31
37. Transportation equipment	.21	.64	.94	-.58	.73	.34	1.06	-.25
38. Instruments	.16	.81	.79	-.15	.73	.19	1.42	-.64
39. Miscellaneous	.68	.23	.85	-.12	.90	.24	1.93	-1.26

Regressions at MSA level include MSA, industry and year fixed effects. Regressions for closures and expansions/contractions include age and size controls.

Table 10: Employment growth due to births, closures, and expansions: linear specification, MSA data

Independent variables	Dependent variables: $\Delta E^j_{ist}/E_{it}$			
	New firm births	Old firm births	Closures	Expansions/ Contractions
Input	.07	-.01	-.01	-.01
	(1.5)	(0.8)	(-0.8)	(-0.5)
Output	.00	.06	-.04	.01
	(0.1)	(2.9)	(-2.2)	(-0.4)
Labor mix	.06	-.01	-.08	-.01
	(5.2)	(-1.0)	(-2.4)	(-1.4)
Integration	.05	-.03	-.04	-.01
	(1.1)	(-0.8)	(-2.2)	(-1.4)
Tech. flows	-.06	.03	.01	-.00
	(-2.4)	(1.6)	(1.8)	(-0.3)
$s_{ist}(\Delta E^j_{it}/E_{it})$	.60	.73	.86	1.16
	(9.9)	(3.9)	(17.7)	(16.8)
$s_{st}(\Delta E^j_{it}/E_{it})$	.36	-.03	.24	-.32
	(4.8)	(-0.2)	(4.6)	(-5.0)
log(avg. plant size)			.05	-.02
			(3.4)	(-1.9)
Share 0-4 yrs old			-.04	.02
			(-9.1)	(3.1)
Share 5-9 yrs old			-.02	.01
			(-6.0)	(3.4)

Regressions include MSA, industry, and year fixed effects.
t-statistics in parentheses.

Table 11: Employment growth due to births and closures: log specification, MSA data

Independent variables	Dependent variable: log(1 + ΔE_{ist}^{j})				
	New firm births		Old firm births		Closures
Input	-.00	-.00	.03	.03	-.01
	(0.0)	(0.0)	(2.8)	(3.1)	(-1.9)
Output	.02	.02	.01	.01	.00
	(3.2)	(3.4)	(0.6)	(1.2)	(-0.8)
Labor Mix	.18	.17	.13	.13	-.07
	(13.4)	(12.6)	(9.7)	(9.4)	(-5.5)
Labor Mix*Closure rate		.02		.01	.02
		(6.1)		(3.2)	(4.4)
Integration	.08	.08	.10	.11	.00
	(5.0)	(5.2)	(4.9)	(5.6)	(-0.2)
Integration*College		.01		.06	-.01
		(1.2)		(5.0)	(-2.0)
Technological Flows	.00	.00	.04	.04	-.01
	(0.3)	(0.3)	(4.9)	(4.4)	(-4.0)
log($(1 + E_{ist})\|\Delta E_{it}^{j}\|/E_{it}$)	.16	.16	.14	.14	1.24
	(68.4)	(68.4)	(54.5)	(54.5)	(173.1)
log($s_{st}\|\Delta E_{it}^{j}\|$)	.04	.04	-.04	-.04	-.34
	(6.9)	(6.9)	(-8.0)	(-8.0)	(-20.6)
log(avg. plant size)					.75
					(114.3)
Share 0-4 yrs old					-.08
					(-27.6)
Share 5-9 yrs old					-.02
					(-6.4)
Adjusted R^2	0.56	0.56	0.32	0.32	0.57
Number of obs.	163,938	163,938	163,938	163,938	85,588

Regressions include MSA, industry, and year fixed effects. Dependent variable for closure regression is $-\log(1 - \Delta E_{ist}^{\text{closure}})$. t-statistics in parentheses.

Table 12: Industry specific coefficients on $LaborMix$ and $Integration$

Industry	Dependent variable: log(1+ΔE_{ist}^{j})					
	New firm births		Old firm births		Closures	
	Labor	Integ.	Labor	Integ.	Labor	Integ.
20. Food	.09	.44	.26	.45	-.15	-.11
22. Textiles	.15	.06	.10	.02	-.34	-.04
23. Apparel	.18	.18	.08	.12	.07	.01
24. Lumber & wood	.37	.24	.39	.03	.01	.03
25. Furniture & fixtures	.32	.21	.09	.15	.04	-.01
26. Paper	-.07	.03	.06	.03	.08	-.01
27. Printing & publishing	.25	.06	.29	.10	.01	.01
28. Chemicals	-.04	.23	.25	.37	.05	-.03
29. Petroleum & coal	-.05	-.21	.20	-.29	.10	.07
30. Rubber & misc. plastics	.17	.09	.23	.10	-.14	.01
31. Leather	-.23	.07	-.08	-.03	-.25	-.05
32. Stone, clay, & glass	.07	.19	.10	.21	-.05	.01
33. Primary metals	.10	.04	.11	.04	-.29	-.09
34. Fabricated metal products	.33	.16	.22	.18	.12	.01
35. Industrial machinery & equip.	.47	.24	.33	.30	.03	.01
36. Electronic & electric equip.	.27	.17	.26	.19	-.02	-.04
37. Transportation	.08	.10	.08	.08	-.19	-.03
38. Instruments	.30	.22	.34	.31	-.16	-.02
39. Miscellaneous	.16	.26	-.04	.22	.01	.02

Regressions include MSA, industry, and year fixed effects. Regression for closures has dependent variable $-log(1-\Delta E_{ist}^{\text{closure}})$ and includes plant age and size controls.

Table 13: Employment changes due to birth and closures: log specification, state data

Independent variables	Dependent variable: $\log(1+\Delta E_{ist}^{j})$				
	New firm births		Old firm births		Closures
Input	.04	.04	.08	.07	-.01
	(5.5)	(5.6)	(7.8)	(7.5)	(-2.5)
Output	.03	.03	.01	.02	-.02
	(5.1)	(5.1)	(1.6)	(2.6)	(-3.8)
Labor mix	.43	.43	.25	.25	-.06
	(11.0)	(10.6)	(5.4)	(5.3)	(-2.0)
Labor mix*Closure rate		.00		.02	
		(0.0)		(2.4)	
Integration	.06	.06	.09	.04	.01
	(5.9)	(5.9)	(6.5)	(2.4)	(4.4)
Integration*College		-.05		.50	
		(-1.5)		(5.2)	
Technological Flows	-.01	-.01	.03	.02	-.01
	(-1.5)	(-1.5)	(3.3)	(3.0)	(-2.1)
$\log(1+E_{ist}\|\Delta E_{it}^{j}\|/E_{it})$	.20	.20	.25	.25	1.28
	(29.5)	(29.5)	(30.0)	(30.0)	(92.5)
$\log(s_{st}\|\Delta E_{it}^{j}\|)$	.34	.34	.14	.14	-.36
	(13.5)	(13.5)	(6.3)	(6.2)	(-11.0)
log(avg. plant size)					.72
					(57.7)
Share 0-4 yrs old					-.05
					(-11.2)
Share 5-9 yrs old					-.01
					(-3.1)
Adjusted R^2	.73	.73	.54	.54	.74
Number of observations	27,234	27,234	27,336	27,336	23,473

Regressions include state, industry, and year fixed effects. Regression for closures has dependent variable $-log(1-\Delta E_{ist}^{\text{closure}})$ and includes plant age and size controls.

Table 14: Initial employment, MSA data

Independent variables	Dependent Variable: log(1+E_{ist})
Input	.01
	(.08)
Output	.06
	(6.9)
Labor mix	.41
	(21.7)
Integration	.04
	(3.8)
Technological Flows	.03
	(6.5)
log($s_{st}E_{it}$)	.58
	(45.8)
Adjusted R^2	.54
Number of observations	164,552

Regressions include MSA, industry, and year fixed effects.
t-statistics in parentheses.