JOURNAL OF REGIONAL SCIENCE, VOL. 45, NO. 2, 2005, pp. 395–419

A DESCRIPTIVE ANALYSIS OF DISCRETE U.S. INDUSTRIAL COMPLEXES*

Edward Feser

Department of Urban and Regional Planning & Regional Economic Applications Laboratory (REAL), University of Illinois at Urbana-Champaign, Champaign, IL 61820. E-mail: feser@uiuc.edu

Stuart Sweeney

Geography Department, University of California, Santa Barbara, Santa Barbara, CA 93106-4060. E-mail: sweeney@geog.ucsb.edu

Henry Renski

Department of Urban and Regional Planning, University of North Carolina at Chapel Hill, CB 3140, Chapel Hill, NC 27599-3140. E-mail: hrenski@email.unc.edu

ABSTRACT. We use the Getis/Ord local *G* statistic and detailed county-level industry employment data from the U.S. Bureau of Labor Statistics to isolate discrete industrial complexes—or groups of nominally linked industries clustered in particular locations—for two recent years: 1989 and 1997. We describe the characteristics of the complexes in terms of their number, spatial extent, broad regional distribution, and other factors. Data from the two periods help illustrate key shifts in industrial locations, including the continuing concentration of the apparel industry in the Southeast and the ongoing southern shift in U.S. vehicle production.

1. INTRODUCTION

Today there is a vigorous debate underway regarding the most appropriate methods for identifying what Harvard strategic management theorist Michael Porter has popularized as industry clusters: "geographic concentrations of interconnected companies and institutions in a particular field" (Porter, 1998, p. 78). Roughly 25 years ago, the regional science literature was full of studies concerned with the "identification of clusters and complexes, or of groups of industries linked by flows of goods and services, or showing significant mutual locational attraction" (Czamanski and Ablas,

*This research was supported by a grant from the National Science Foundation (BCS-9986541). Data are used with special permission from the U.S. Bureau of Labor Statistics (BLS). Research results and conclusions expressed are the authors' and do not necessarily indicate concurrence by the BLS.

Received April 2001; Revised May 2004; Accepted July 2004

Blackwell Publishing, Inc., 350 Main Street, Malden, MA 02148, USA and 9600 Garsington Road, Oxford, OX4 2DQ, UK.

1979, p. 61).[1] At that time, regional scientists defined industry clusters as groups of sectors linked through input–output relationships, regardless of their geography, and industry complexes as industry clusters whose members (i.e., sectors) exhibited common location patterns. Both the concepts of complexes and modern industry clusters are premised on a hypothesized association between industrial interdependence and spatial proximity.

In this paper, we return to the early distinction between clusters and complexes to develop a systematic means of identifying and examining the characteristics of geographically localized groups of nominally linked industries in the U.S. economy. By systematic, we mean a methodology that does not rely on *a priori* information as to the location of such agglomerations. Most of the current research on industry clusters adopts the case study method, focusing on documenting ties among co-located businesses in well-known clusters such as computers and software in Silicon Valley, biotechnology in Boston's Route 128, aerospace and software in Seattle, and automobiles in Detroit. The research presented here is consistent with the earlier industrial complex literature in that we use exogenous information on interdependence (input–output) and geographically disaggregated regional data to search for patterns of concentration among linked industries. However, unlike that early literature, which focuses only on detecting overall inter-industry location patterns, we use a local indicator of spatial association (LISA; see Anselin, 1995) to pinpoint the actual locations of various complexes in the United States. The approach, which we view as a type of exploratory analysis, produces descriptive information about the changing geography of U.S. production that can be used for designing formal tests of location theory.

Using definitions of extended buyer–supplier value chains derived from an analysis of U.S. input–output data, we use a G statistic (Getis and Ord, 1992; Ord and Getis, 1995) to isolate regions in the United States where such chains are localized. We term such agglomerations "discrete industrial complexes." We then calculate basic measures of the average size, spatial extent, metropolitan orientation, and broad regional distribution of the complexes for each value chain. While analysis of the descriptive statistics must be undertaken cautiously given unavoidable modifiable areal unit issues, reliable trends can be discerned in combination with the careful inspection of maps of each set of complexes. We focus not only on findings generated with 1997 employment data but also briefly summarize trends over the 1989–1997 period for two value chains: apparel and vehicle manufacturing. We argue that specifying a form of inter-industry independence (of which input–output is only one among many that might be envisioned), and then examining location patterns of identified related sectors, is extremely useful for studying clustering and agglomeration.

[1]See, for example, Isard, Schooler, and Vietorisz (1959); Richter (1969); Streit (1969); Bergsman, Greenston, and Healy (1972); Campbell (1972, 1974); Czamanski (1974, 1976, 1977); Roepke et al. (1974); Bergsman, Greenston, and Healy (1975); Latham (1976, 1977); ÓhUallacháin (1984, 1985); Howe (1991).

2. METHODOLOGY

We require two things to derive a set of discrete industry complexes: a means of identifying businesses linked via buyer–supplier relationships and a measure of geographic concentration that reveals local agglomerations or “hot spots” in the spatial pattern of U.S. industrial activity. To isolate linked businesses, we use a set of “benchmark” value chains derived with a methodology described briefly here and outlined in detail in Feser and Bergman (2000). Input–output data provide a useful characterization of trading patterns and general technological similarities between industries, with an emphasis on manufacturing. The methodology utilizes a statistical factor analysis that produces summary measures of inter-industry linkage between three- and four-digit Standard Industrial Classification (SIC) system industries. The analysis excludes local-serving sectors such as personal services, construction (though not construction equipment), retail, wholesale, government, and education. Farming is also excluded, primarily because of inappropriate data. In general, the included sectors are those with the greatest potential to export goods and services outside sub-state regions.

Very briefly, the method starts with a 491×491 inter-industry transactions matrix, $\mathbf{A}$, assembled from the 1992 detailed industry by commodity benchmark input–output accounts of the United States. Following Czamanski (1974), two matrices, $\mathbf{X}$ and $\mathbf{Y}$, are derived with elements:

$$x_{ij} = \frac{a_{ij}}{a_{+j}}, y_{ij} = \frac{a_{ij}}{a_{i+}}$$

where a_{ij} is the dollar value of goods and services sold by industry i in some period to industry j, and a_{+j} and a_{i+} are total intermediate good purchases and sales, respectively, of industries i and j over the same period. The term x_{ij} is intermediate good purchases by sector j from i as a proportion of j’s total intermediate good purchases. A large value for x_{ij} implies that j depends on industry i as a source for a significant share of its total intermediate purchases. The term y_{ij} captures intermediate good sales from i to j as a proportion of i’s total intermediate good sales, so that a large value for y_{ij} implies that i depends on industry j as a market for a large proportion of its total intermediate good sales. The columns of $\mathbf{X}$ are the intermediate input-purchasing pattern of each industry j, and the rows of $\mathbf{Y}$ are the intermediate output sales pattern of each industry i.

For any two industries (A and B), with the column vectors of $\mathbf{X}$ defined as $\mathbf{x}_\mathrm{A}$ and $\mathbf{x}_\mathrm{B}$ and the row vectors of $\mathbf{Y}$ defined as $\mathbf{y}_\mathrm{A}$ and $\mathbf{y}_\mathrm{B}$, four correlations on the sales and purchasing vectors of any two industries are derived: (1) $\mathrm{r}(\mathbf{x}_\mathrm{A} \cdot \mathbf{x}_\mathrm{B})$ measures the similarity in input purchasing patterns of industries A and B; (2) $\mathrm{r}(\mathbf{y}_\mathrm{A} \cdot \mathbf{y}_\mathrm{B})$ measures the degree to which A and B possess similar output selling patterns, that is, the degree to which they sell goods to a similar mix of intermediate input buyers; (3) $\mathrm{r}(\mathbf{x}_\mathrm{A} \cdot \mathbf{y}_\mathrm{B})$ measures the degree to which the buying pattern of industry A is similar to the selling pattern of industry B, that is, the degree to which industry A purchases inputs from industries in which B supplies (a second-tier linkage); and (4) $r(\mathbf{x}_B \cdot \mathbf{y}_A)$ measures the degree to which the buying pattern of industry B is similar to the selling pattern of industry A, that is, the degree to which industry B purchases inputs from industries in which A supplies. A linkage matrix, **L**, comprising the largest of the four correlations for each pair of sectors, summarizes the degree of linkage between and among all 491 sectors. Before performing the factor analysis, **L** is reduced to a 367×367 matrix by eliminating natural resource-based and primarily locally-serving sectors, as noted above.

The factor analysis of **L** identified 26 value chains, of which we selected five for examination in this study: information technology (IT) and instruments, motor vehicle manufacturing, apparel, transportation and shipping, and pharmaceuticals. The industry components of each value chain are provided in the Appendix. The chains are not mutually exclusive and while some are dominated by industries within the same broad sector, most include a range of component sectors that span major industry categories. For example, the motor vehicles chain includes industries in SICs 22, 23, 25, 28, 30, 31, 32, 34, 35, 36, and 37. In contrast, because the majority of inter-industry trade in pharmaceuticals occurs within the industry itself, the pharmaceuticals chain includes only sectors in SIC 283 (drugs).

We selected the five chains to represent varying technologies, historical vintages, and locational imperatives. IT and pharmaceuticals are both comparatively technology-intensive chains with strong urban orientations. The motor vehicles chain is a heavy, technology-intensive industry that has undergone a pronounced shift in its regional orientation from the industrial upper Midwest to southern states such as Kentucky, Tennessee, and South Carolina. The mature apparel industry is notoriously location cost-sensitive with a heavy concentration in the Southeast but continuing presence in major metropolitan areas such as Los Angeles and New York. The transportation and shipping industry has historically concentrated in population centers and at highway nodes, rail nodes, and ports, but is also believed to be attaining an increasing degree of locational flexibility with the growth of Internet and catalog shopping, increases in time-sensitive shipping (e.g., medical testers, computer assemblers, etc.), and the development of air cargo hubs. Leading states in each chain—according to total 1997 employment—are reported in Table 1.

Note that our approach to identifying linked industries produces a set of average value chains for the United States as a whole. The definitions of the chains ignore potential regional variations in trading patterns, whether determined by differences in price, transportation costs, or technology. They are best viewed as an alternative industrial classification scheme determined by aggregate trading ties as opposed to the similarity-of-product rationale underlying the SIC system. We seek to understand the changing geography of the most likely set of trading sectors while emphasizing that we have no evidence of actual trading linkages at the sub-national scale. The chains are designed to

TABLE 1: Leading States, Five Value Chains (Based on Total Employment, 1997)

IT	Apparel	Vehicles	Transportation	Pharmaceuticals
California	North Carolina	Michigan	California	California
Texas	California	Ohio	Texas	Texas
New York	Georgia	Indiana	New York	New York
Massachusetts	South Carolina	California	Illinois	Illinois
Illinois	New York	Illinois	Florida	Florida
Florida	Alabama	Tennessee	Pennsylvania	Pennsylvania
Pennsylvania	Virginia	Texas	New Jersey	New Jersey
Virginia	Pennsylvania	Pennsylvania	Ohio	Ohio
Minnesota	Tennessee	Georgia	Georgia	Georgia
New Jersey	Texas	New York	Tennessee	Tennessee

Source: U.S. Bureau of Labor Statistics.

help reveal patterns in the U.S. industrial geography that would not be evident if we simply utilized default sectoral categories. The objective is to use the benchmark chains, in combination with a spatial association measure, to scan the landscape to highlight localized complexes that can then be studied in detail. The scanning can also reveal shifts in overall levels of concentration when applied for different time periods, as we suggest below.

We identified discrete industrial complexes for value chain i in time t—or distinct regions where i is especially concentrated—in four steps. First, we calculated total value chain employment for counties in all contiguous 48 states except Wyoming and Massachusetts, anomalies we discuss in more detail below. Second, we regressed value chain i employment on total export-oriented employment using counties as the units of analysis. Third, we calculated a local G statistic for all counties using the regression residuals from step 2. Fourth, we used maps to visually identify discrete complexes as spatially contiguous groups of counties posting high G values. We then assembled data on the basic characteristics of the set of complexes for each chain. The following sections discuss the methodology in detail.

Units of Analysis and Data

The basic data source for the study is the U.S. Bureau of Labor Statistics' comprehensive ES-202 file, which reports employment and wages for all businesses subject to federal and state employment security law. At the time of study, 1997 was the most recent year available to us, with reliable data stretching back to 1989. We selected counties as the primary unit of analysis because they are the smallest meaningful areal unit to which we could reliably aggregate the micro-level data. In ongoing research, we are using the business addresses contained in the ES-202 file to study industrial concentration patterns using point process models and techniques. Point process models have the advantage of avoiding inevitable modifiable areal unit problems that can often only be acknowledged rather than solved. There is still substantial value, however, in continuing to refine areal methods since area-based economic data are more commonly available than point data (e.g., via the Economic Census, County Business Patterns, or proprietary sources).

The unsuppressed BLS ES-202 file does not include sole proprietorships, owner-operated businesses with a single employee (the owner). While over 90 percent of business establishments in the United States are represented overall, it is possible that the findings for any value chain with a significant share of sole proprietorships (such as consultants) could be biased. Four out of five of the chains examined here are unlikely to be affected because they are comprised exclusively of sectors where sole proprietorships are relatively few in number (e.g., manufacturing sectors or industries dominated by large companies). However, the IT value chain includes an important segment—information services and software—that is likely to include many sole proprietorships. As a result, the findings for IT should be interpreted as reflective of the manufacturing core of that chain (IT hardware) along with larger software and information services companies. The exclusion of sole proprietorships in the software industry may account for our finding of a comparatively even regional distribution of IT complexes across U.S. Census divisions.

Our base data include unsuppressed employment figures at the four-digit SIC level for every county in every state in the continental United States except Wyoming and Massachusetts. Wyoming denied us access to its confidential data altogether while Massachusetts allowed the use of their data absent all sub-state geographic identifiers. Because Wyoming has a small and predominantly natural resource-based economy, its omission should not create major distortions in our analysis of the five value chains included here. Large and diverse Massachusetts, however, is a key location of several of the chains, most notably IT and pharmaceuticals. We therefore elected to include the Massachusetts data in the spatial analysis, essentially treating the state as one large county. Aggregating Massachusetts' counties into one unit is preferred to omission because national employment totals are used in calculating the local G. But while the estimated G values for non-adjacent regions are unaffected, aggregation is likely to generate localized distortions in and around Massachusetts itself. The individual counties of the state are likely to be more specialized in particular industries than the state as a whole and thus sub-state pockets of concentrated activity will likely be masked. Because the G is also calculated across neighboring county boundaries, statistics for the counties that abut that state must be interpreted cautiously.

Measure of Concentration

We are primarily interested in identifying regions where counties posting comparatively high levels of employment for a given value chain are clustered together. Such spatial "hot spots" in the geographic distribution of value chain activity can be detected effectively using the local G statistic outlined in Ord and Getis (1995; see also Getis and Ord, 1992), which has the advantage of distinguishing between the clustering of high versus low values of the variable under study. The value of G for a given county is based on value chain employment of both the county itself as well as neighboring counties. In this way, G detects concentrations of cluster activity across county boundaries, rather than just within them, as is the case with simple measures such as the location quotient, the local Gini coefficient (Krugman, 1991), or the localization coefficient of Ellison and Glaeser (1997).

The original G statistic described by Getis and Ord (1992) was developed as a share-based measure of spatial concentration, that is, the amount of activity (e.g., employment) in a multi-unit region divided the total activity across all units. In this paper, we use the newer variant of the statistic outlined in Ord and Getis (1995). Because the new statistic is expressed in standard deviations from the mean, the results can be interpreted roughly as z-scores along the normal curve. The new measure also permits the use of non-binary spatial weights and numeric values that do not have a natural origin. The measure for county i for a given industry cluster is calculated as

$$G_i^* = \frac{\sum_j w_{ij}x_j - W_i\bar{x}}{s\sqrt{\frac{nS_{1i}-W_i^2}{n-1}}}$$

for all j, where x is a measure of value chain size (e.g., employment), w_{ij} is a spatial weight that defines neighboring counties j to county i, W_i is the sum of weights w_{ij}, $\bar{x} = \sum_j x_j \big/ (n-1)$, $S_{1i} = \sum_j w_{ij}^2$, and $s^2 = \left(\sum_j x_j^2 \big/ n-1\right) - (\bar{x})^2$. Although the normality of G_i* depends partially on the number of neighbors (Getis and Ord, 1992), we make the common simplifying assumption that it follows a normal distribution for each county. Significant counties are identified as those posting values of 1.96 or greater, the 95 percent significance level from a two-tailed normal distribution.

The first step in calculating G_i* is to develop the spatial weights w_{ij}. We define adjacency by immediate neighbor counties inclusive of the county itself. Non-neighboring counties are given a weight of zero. The value x of neighboring county j to county i is weighted by the degree of expected interaction between counties j and i:

$$w_j = \frac{X_iX_j}{\sum_j X_iX_j}$$

where upper-case X is total export-oriented employment and the denominator is the sum of interactions between county i and all its neighboring counties j. Dividing by the sum of interactions row-standardizes the matrix, turning each cell's weight into a percentage of the total interactions between adjacent counties. The weighting scheme adopts a gravity logic: that larger employment centers exert a heavier influence on neighboring counties than do smaller areas. There is also an implicit assumption that non-neighboring counties do not interact.[2]

Controlling for Urban Orientation

One of the most persistent conceptual and empirical challenges in studying geographic concentration is distinguishing first-order and second-order spatial processes. First-order effects relate to variation in the mean value of the process in space, a global or large-scale trend. Second-order effects result from the spatial correlation structure, or deviations of local value from the large-scale trend (Bailey and Gatrell, 1995). The existence of a strong first-order trend is a clear violation of the randomization assumption implicit in most statistical hypothesis tests. In industrial geography, a first-order trend is generated through the uneven distribution of population and employment into a small number of metropolitan areas. To the extent that much commerce serves the local population and that infrastructure is more developed in urban areas, we should expect more employment in urban centers because there is more activity there to begin with.

Although changes in first-order employment trends such as the rise of the Sunbelt and decline of the industrial Midwest are of interest to many geographers, our investigation is primarily concerned with identifying significant second-order spatial trends. In the present context, those are the local concentrations of industrial activity beyond those expected to follow the general distribution of population and employment. In practice, it is often difficult to distinguish first- and second-order spatial trends. In an earlier study (Feser and Sweeney, 2000), we investigated the global clustering of value chain establishments using a case-control framework developed by Diggle and Chetwynd (1991). However, that approach is inappropriate for detecting specific local "hot spots" of industrial activity.

The most straightforward implementation of $G_i{}^*$ is to set x equal to total value chain employment. The practical consequence of that approach, however, is that the largest counties (typically those that include New York, Los Angeles, Chicago, and other large cities) are identified as complexes for nearly every chain. Such counties post high levels of value chain employment, regardless of the whether the chain is large in relative terms. Because $G_i{}^*$ is a relative measure, the dominance of the largest counties also means that it misses smaller places with large concentrations of value chain employment. In other words, first-order trends dominate the outcome.

One way to offset the large-size effects is to transform value chain employment such that the skewed place size distribution is dampened (e.g., $x = ln\ x$). Such an approach is successful in identifying small and medium-sized places

[2]We introduced the gravity specification of the weights after initial tests with binary weights failed to distinguish between discrete complexes in heavily industrialized areas, mainly in the Northeast and Midwest.

washed out by the dominance of the largest metropolitan counties. Although a transformation can be usefully applied where the objective is to identify absolute concentrations of activity, it does not isolate second-order spatial processes that are of primary interest to researchers investigating spatial externalities, spillovers, and other advantages of proximity between businesses. The G_i^* statistic might also be calculated using share-based measures of value chain employment (percentages or location quotients), reducing the influence of first-order effects. However, while biasing against places that are more diverse, the use of relative size measures exhibits a noted preference for smaller, highly specialized local economies. The smallest counties with small absolute levels of cluster employment appear significant whenever they also have a very small overall economic base.

In general, it is clear that there are two principal dimensions of the importance of a given industry in a region: absolute size (magnitude) and relative size (share). The ideal measure would capture both dimensions by reflecting local concentrations of activity while also accounting for overall magnitudes in such a way that the concentrations, and not just a lack of underlying industrial diversity, are detected.

To help isolate second-order from first-order spatial processes and strike a balance between the magnitude and share trade-off, this study uses the residuals from regressions of county value chain employment on total county export employment. The method assumes that first-order spatial processes are captured by the spatial distribution of the total export base employment. The difference between predicted and actual value chain employment is essentially value chain employment beyond that expected by the overall size of the county's export base (a measure of overall size of the place). The residuals then become x in the calculation of the local statistic G_i*. Examining all the value chains developed via the factor analysis, we found that the residual method is reasonably successful in excluding very small, non-diverse places with high relative concentrations of value chain activity (e.g., branch plant-dominated rural economies). Other advantages of the residual method are that it is intuitive, easily implemented, and allows for alternative indicators to control for first-order trends, such as population, total employment, or other measures of economic size. In conceptual terms, the residual approach is akin to original attempts by early studies of industry complexes to control for the urban orientation of value chains by regressing county value chain employment on county population (Czamanski and Ablas 1979).

3. RESULTS

The results for each of the five value chains are summarized most effectively visually. Figures 1–5 plot values of G_i* for 1997, with the darkest shaded counties indicating statistical significance. We visually inspected each map and identified individual clusters of significant counties, or what we are calling discrete complexes. Tables 2 and 3 characterize fundamental

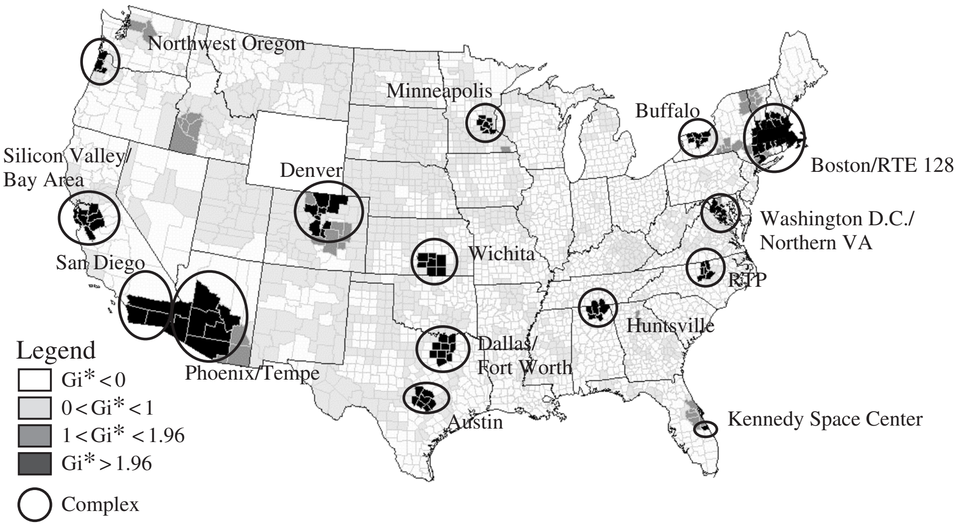

FIGURE 1: Discrete Complexes, Information Technology, 1997.

differences among the five value chains in terms of the number, size, metropolitan orientation, and regional distribution of their respective complexes.

However, first examining the set of maps, the success of the residual approach in combination with the *G* statistic is evidenced simply by the plausibility of the results. In the case of IT (Figure 1), well-known places of significant activity are identified, including the Silicon Valley/Bay Area, Los Angeles/San Diego, Phoenix, Austin, Boston, northern Virginia, and North

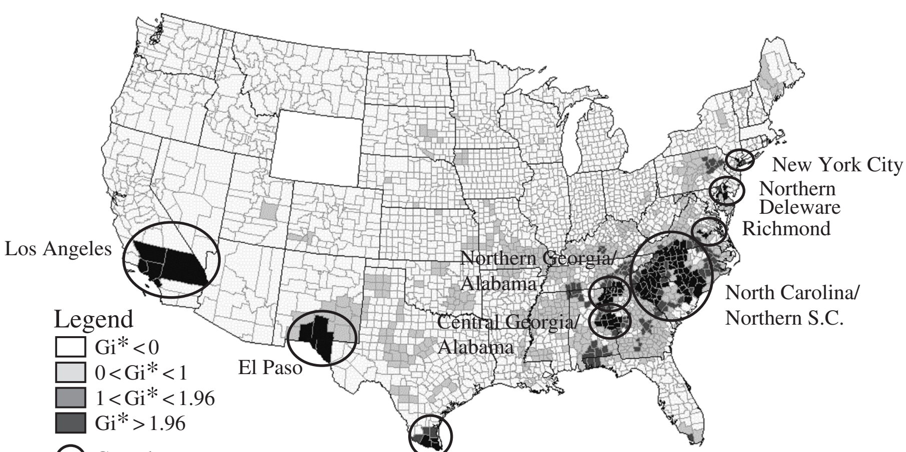

FIGURE 2: Discrete Complexes, Apparel, 1997.

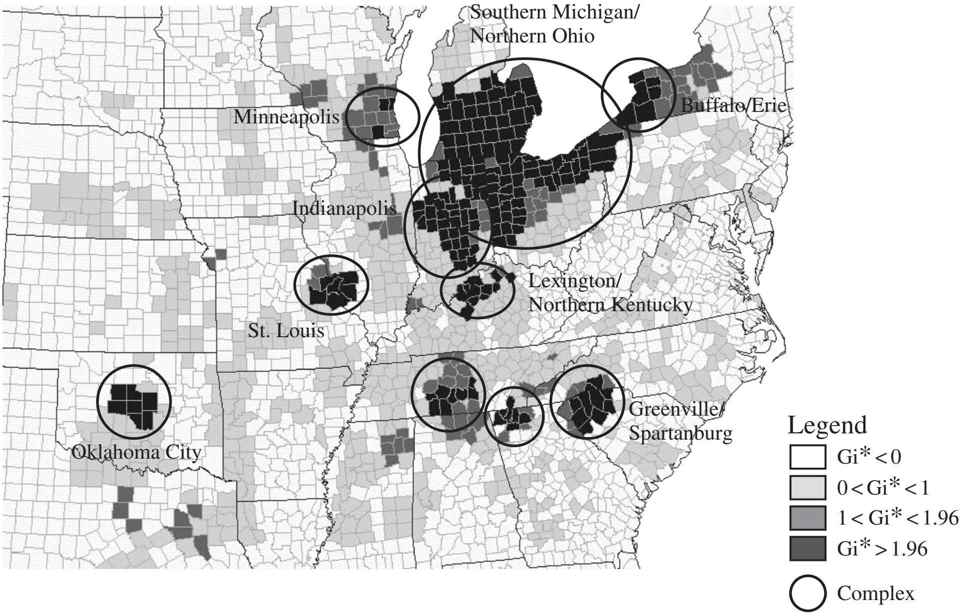

FIGURE 3: Discrete Complexes, Motor Vehicles, 1997.

Carolina's Research Triangle. Less well-known concentrations include Dallas/ Forth Worth, Wichita, Denver/Front Range, Minneapolis, and Buffalo. Indeed, the latter illustrates the value of the technique, namely, its ability

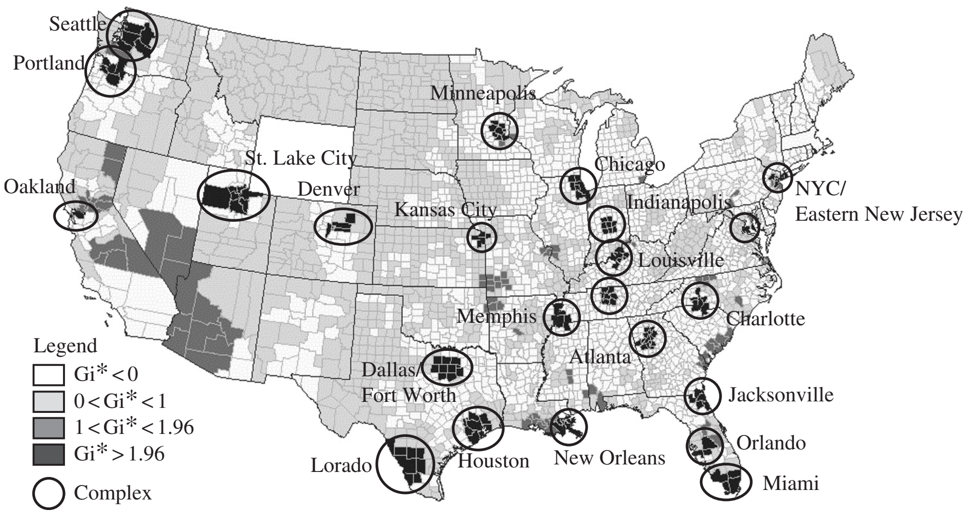

FIGURE 4: Discrete Complexes, Transportation and Shipping, 1997.

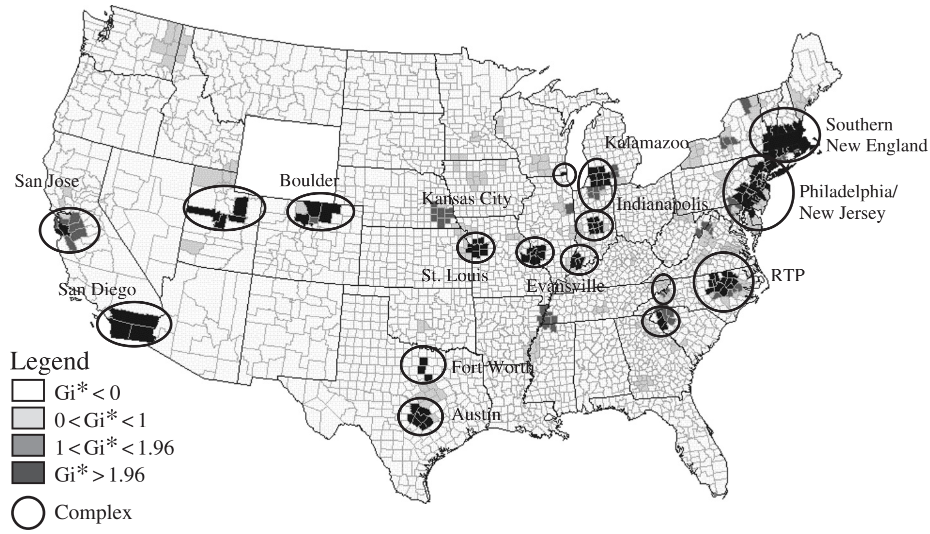

FIGURE 5: Discrete Complexes, Pharmaceuticals, 1997.

to systematically detect notable but less commonly cited concentrations of activity in related industries. Such places are excellent candidates for further analysis, including case studies. The fact that the results are ultimately dependent on specific measures of value chain size, or, for that matter, the measure of interdependence itself, is not problematic when the goal of the research is ultimately exploratory in nature. Indeed, varying the analysis to include shares, location quotients, or logged values, experimenting with different quantity measures, such as output, value added, or wages, or altering the means of evaluating industrial interdependence could generate valuable complementary information for both research and policy purposes.

Figure 2 illustrates the heavy dominance of the Southeast in apparel, while also highlighting apparel centers in New York and Los Angeles and cost-sensitive production along the U.S.–Mexico border. The core motor vehicle complexes (Figure 3) in Michigan, Ohio, and Indiana are clearly evident, along with relatively newer centers in Kentucky, Greenville/Spartanburg, and central Tennessee. Many of the transportation and shipping complexes are also predictable (port and rail centers like Seattle, Oakland, New York, Miami, Chicago, along with passenger and/or freight air hubs in Atlanta, Indianapolis, Louisville, Memphis, Denver, and Salt Lake City; see Figure 4), as are the pharmaceutical complexes (San Jose/Silicon Valley, Boulder, Indianapolis, Research Triangle Park, and Boston; see Figure 5).

In some cases, the lack of complexes in certain areas is informative. For example, a concentration of transportation and shipping activity is evident in St. Louis but is not particularly strong relative to other complexes. Indeed, along the Mississippi, we find shipping complexes in Minneapolis, Memphis,

TABLE 2: Basic Characteristics, Identified Complexes, 1997

Value Chain	IT	Apparel	Vehicles	Transportation	Pharmaceuticals
R-squared, value chain employment on total employment	65.1	41.5	41.7	85.5	22.1
Number of discrete complexes	15	9	10	28	17
Number of states with significant counties	20	13	14	29	23
Number of significant counties	115	126	186	163	129
Mean square miles, significant counties	1,297	846	510	1,043	822
Percent share of U.S. counties that are part of complexes	3.7	4.1	6.0	5.3	4.2
Percent of U.S. county area in complexes	5.1	3.7	3.3	5.8	3.7
Percent U.S. value chain employment in complexes	37.7	46.1	35.7	30.5	52.6
Percent significant counties in metropolitan statistical areas (MSAs)	69.6	44.4	55.4	78.5	76.0
Percent complex employment in MSA counties	99.0	67.2	81.9	94.4	99.4
Range, number of counties in discrete complexes (smallest/largest)	1/20	3/70	3/96	1/13	1/27
Percent significant counties in largest complex	17.4	55.6	51.6	8.0	20.9
Smallest areal complex (square miles)	489	532	1,257	418	226
Largest areal complex (square miles)	46,778	96,964	50,933	26,016	21,313
Mean area of discrete complexes	9,946	11,850	9,489	6,069	6,393
Percent total complex area in largest complex	31.4	34.7	53.7	15.3	20.1
Percent value chain employment in largest areal complex	22.3	41.5	72.1	14.3	50.2

Source: U.S. Bureau of Labor Statistics and authors' calculations.

and New Orleans. An interesting question is whether the localization of shipping activity in places like Minneapolis and Memphis is driven more by the airborne rather than waterborne segments of the shipping value chain. We would expect that an analysis conducted with much earlier data, say from the early twentieth century when shipping along the Mississippi was much more significant relative to other modes, might find complexes in St. Louis as well as other smaller river port communities.

Table 2 summarizes that the transportation and shipping value chain exhibited the most complexes (28), an unsurprising finding given the nature of the industry. Next were pharmaceuticals (17 complexes) and IT

TABLE 3: Regional Distribution, Identified Complexes, 1997

Census Division	IT	Apparel	Vehicles	Transportation	Pharmaceuticals
Percent U.S. complex counties					
New England	13.9	0.0	0.0	0.0	11.6
Middle Atlantic	8.7	5.6	4.8	8.0	24.8
East North Central	0.0	0.0	70.4	12.9	18.6
West North Central	13.0	0.0	2.7	3.1	10.1
South Atlantic	17.4	75.4	8.6	22.1	16.3
East South	6.1	8.7	9.7	9.8	0.8
West South	14.8	4.0	3.8	25.2	7.8
Mountain	13.9	1.6	0.0	9.2	6.2
Pacific	12.2	4.8	0.0	9.8	3.9
Percent U.S. complex employment					
New England	21.2	0.0	0.0	0.0	7.2
Middle Atlantic	5.8	15.4	2.8	14.7	56.5
East North Central	0.0	0.0	81.8	17.8	14.3
West North Central	8.2	0.0	3.3	0.8	5.4
South Atlantic	12.6	51.9	6.5	20.2	9.3
East South	1.9	4.0	4.1	7.4	0.0
West South	14.9	4.6	1.6	22.3	1.7
Mountain	9.0	0.1	0.0	7.3	0.2
Pacific	26.4	24.1	0.0	9.5	5.4

Source: U.S. Bureau of Labor Statistics and authors' calculations.

(15 complexes). Apparel and motor vehicles revealed the fewest complexes (9 and 10, respectively), in part because of their extended geographic presence in certain regions (Michigan/Ohio in the case of motor vehicles, and North and South Carolina in the case of apparel). The largest apparel complex is roughly 97,000 square miles in size, compared to 21,000 for the largest pharmaceuticals complex. Of course, those values very much depend on how the complexes are delineated as well as differing county sizes (a modifiable areal unit problem). However, general trends in Table 2 appear to accord with mapped results as well as intuition. Tightly self-contained complexes are more common for IT, pharmaceuticals, and transportation and shipping than apparel and motor vehicles, both of which have been important sources of rural employment. This is confirmed by looking at the distribution of complex activity in metro versus non-metro counties. Seventy-nine percent of transportation and shipping, 76 percent of pharmaceuticals, and 70 percent of IT complex counties are designated as metro counties, compared to 55 percent and 44 percent for motor vehicles and apparel, respectively.

Assigning complexes to Census Divisions helps illustrate fairly strong regional trends (Table 3). The Pacific and New England Divisions account for nearly 50 percent of IT complex employment while the South Atlantic is home to over half of workers in the country's apparel complexes. The East North Central Division accounts for 82 percent of workers in U.S. motor vehicles complexes. When interpreting these results, it is important to note that the share of total value chain employment located in complexes themselves varies widely. For example, over 53 percent of pharmaceuticals value chain employment in the contiguous United States (excluding Wyoming) is located in multi-county complexes, for example, compared to 31 percent for transportation and shipping. Considerable shares of value chain employment (and counties containing value chain employment) are not part of discrete multi-county complexes at all.

Trends 1989–1997

One useful feature of the industrial complex analysis as outlined here is its ability to examine trends over time. Unfortunately, our exploratory analysis with data from 1989 to 1997 indicates that an eight-year time horizon is often too short to detect meaningful patterns for many value chains. Clear trends are only discernable for those value chains that have been undergoing fairly dramatic regional shifts.

Two such examples are apparel and motor vehicles. Even in the relatively mature apparel industry, its continued southward shift and overall national contraction is evident. For example, Figure 6 shows that comparatively large apparel complexes are detected in Pennsylvania in 1989 compared to 1997. Furthermore, the apparel complexes in the Carolinas and Georgia/Tennessee actually shrink in spatial extent over the period while complexes along the

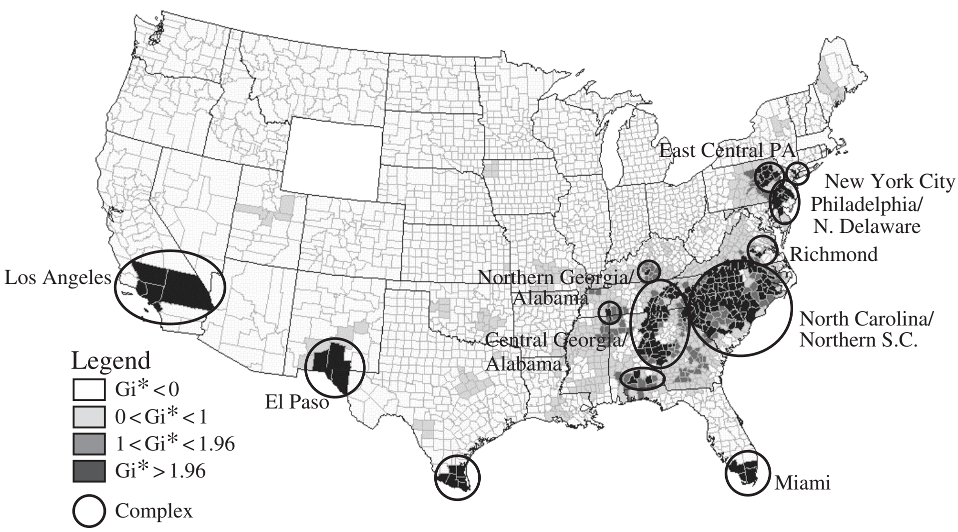

FIGURE 6: Discrete Complexes, Apparel, 1989.

U.S.–Mexico border expand. In the case of motor vehicles (Figure 7), complexes in central Tennessee and Kentucky that appear in 1997 (undoubtedly due to the recent location of major assemblers such as General Motors, Ford, and Toyota and their core suppliers in those states) are not detected in 1989.

Geographical changes over the period in the apparel and motor vehicles value chains can be further evaluated by comparing the descriptive measures in Tables 4 and 5 to those in Tables 2 and 3. In the case of apparel, the number of discrete complexes fell from 13 to 9 over the period while the number of component counties declined from 213 to 126. In 1989, the Middle Atlantic region accounted for about 19 percent of apparel complex employment, a figure that had fallen to 15 percent by 1997. By contrast, while the number of vehicle production complexes also fell (from 13 to 10), the number of significant complex counties increased (from 151 to 186). The three southern Census division accounted for 12.2 percent of vehicle production complex employment by 1997, compared to just 6.4 percent in 1989.

4. SUMMARY

In this paper, we develop and illustrate an exploratory method for detecting discrete industrial complexes. With data for the period 1989–1997, our application focuses on five U.S. value chains: IT, apparel, motor vehicles, transportation and shipping, and pharmaceuticals. The value of the approach

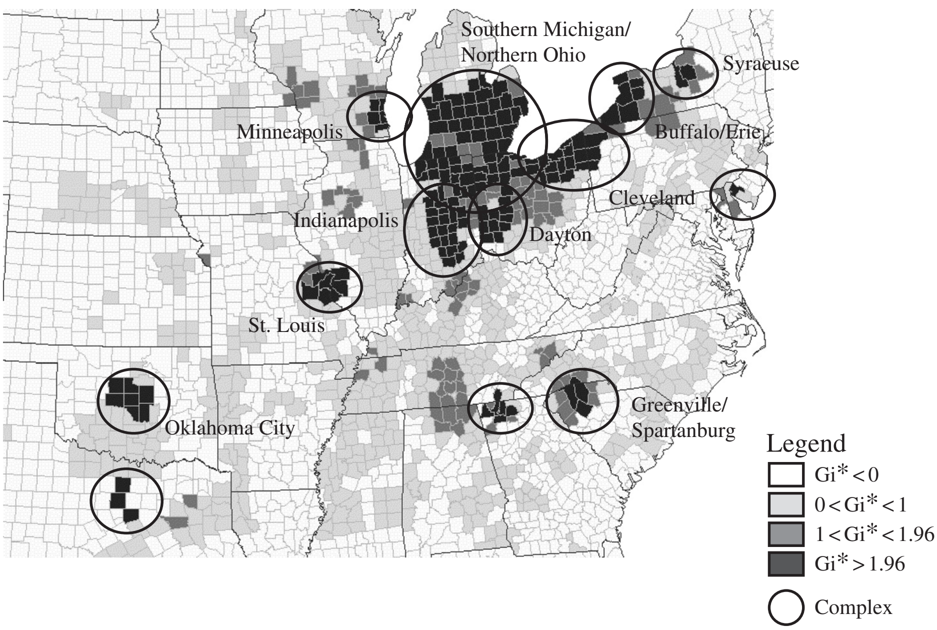

FIGURE 7: Discrete Complexes, Motor Vehicles, 1989.

TABLE 4: Basic Characteristics, Identified Complexes, 1989

Value Chain	Apparel	Vehicles
R-squared, value chain employment on total employment	47.5	45.1
Number of discrete complexes	13	13
Number of states with significant counties	16	15
Number of significant counties	213	151
Mean square miles, significant counties	737	537
Percent share of U.S. counties that are part of complexes	6.9	4.9
Percent of U.S. county area in complexes	5.4	2.8
Percent U.S. value chain employment in complexes	52.5	35.1
Percent significant counties in MSAs	40.8	57.0
Percent value chain employment in significant MSA counties	63.7	86.0
Range, number of counties in discrete complexes (smallest/largest)	1/104	1/57
Percent significant counties in largest complex	51.2	37.7
Smallest areal complex (square miles)	403	351
Largest areal complex (square miles)	55,308	31,793
Mean area of discrete complexes	12,636	6,242
Percent total complex area in largest complex	36.5	56.6
Percent value chain employment in largest areal complex	50.1	56.6

Source: U.S. Bureau of Labor Statistics and authors' calculations.

TABLE 5: Regional Distribution, Identified Complexes, 1989

Census Division	Apparel	Vehicles
Percent U.S. complex counties		
New England	0.0	0.0
Middle Atlantic	11.3	7.9
East North Central	0.0	75.5
West North Central	0.0	3.3
South Atlantic	65.3	6.0
East South	16.4	0.7
West South	3.8	6.6
Mountain	0.9	0.0
Pacific	2.3	0.0
Percent U.S. complex employment		
New England	0.0	0.0
Middle Atlantic	18.7	5.2
East North Central	0.0	84.7
West North Central	0.0	3.6
South Atlantic	59.8	4.8
East South	8.0	0.0
West South	2.8	1.6
Mountain	0.1	0.0
Pacific	10.6	0.0

Source: U.S. Bureau of Labor Statistics and authors' calculations.

resides in its capacity to systematically identify significant spatial concentrations of linked and related industries. While the measure of interdependence between sectors is narrowly confined to input–output linkages (thus, our use of the term "complexes"), the basic approach has the potential to inform current research on industry clusters, which tends to emphasize a broader array of interrelationships between co-located businesses. On the one hand, case studies could be conducted on specific complexes to document formal and informal relationships between firms in the given region. On the other hand, given suitable data, other dimensions of interdependence (e.g., shared innovations, exchange of information, and shared labor pools) could be used to initially determine groups of related industries that would then be subjected to spatial analysis of the kind conducted here.

We have also explored in some depth the utilization of one particular local spatial autocorrelation measure for the analysis of industrial geography (the G statistic). It is clear that there remain significant challenges in the application of exploratory spatial analysis techniques to industrial location questions, given both data limitations and the complexity of underlying spatial processes. We suggest, after testing of a variety of possible approaches, a simple strategy for isolating concentrations of value chain activity above what would be expected given the general distribution of economic activity. Further applications, such as that presented here, are essential for the continued refinement and eventual widespread adoption of the many new techniques emerging in the exploratory spatial data analysis literature.

It is important to emphasize that it is the conceptual separation between the notions of industrial interdependence and spatial proximity that forms the basis of our strategy for scanning the industrial landscape for locations that may be home to business clusters in the sense described by Porter (Porter, 1990, 1998, 2000, 2002) and his adherents. Much of the current empirical work on industry clusters focuses on documenting and analyzing different kinds of business ties in already identified, usually well-known, concentrations (e.g., the Silicon Valley computer industry, probably the archetype industry cluster). Thus, while the number of case studies of unique places has grown rapidly, the development of methods for finding clusters that we do not already know about (or do not find out about through some informal and idiosyncratic means) has lagged (Steiner, 1998; Roelandt and Hertog, 1999; Gordon and McCann, 2000; Hertog, Bergman, and Charles, 2001a, b). The industrial complex approach documented here offers a means of locating potential clusters that is flexible enough to admit different forms of interdependence while being feasible to apply given current data and spatial analysis methods.

Our methodology has important limitations that must be acknowledged when interpreting the findings. First, no measure of spatial concentration can account for all theoretically relevant dimensions of localization in which researchers might be interested. One could make the case that our use of value chain employment ignores the concentration of smaller enterprises that would be detected if establishments were studied instead. Or, it could be argued that the use of residuals from regressions of value chain on total (exportable) employment excessively dampens the significance of value chain activity in the largest metropolitan areas. Our defense is that our objective is to illustrate how to use a combination of methods—analysis of sectoral interdependence with exploratory spatial techniques—to develop richer information about industrial geography. Other measures of value chain activity are not necessarily any less relevant, depending on the specific question at hand.

Second, the varying size of counties across the United States admittedly biases our descriptive areal size measures, particularly for those complexes represented in both the eastern and western halves of the country. Although this problem is unavoidable, inaccurate conclusions can be minimized by comparing the univariate descriptive measures with the maps of complexes. In addition, we hope to shed further light on the problem by utilizing zip code-based data derived from the confidential ES-202 files.

Third, the use of ES-202 data means that the very smallest businesses, namely sole proprietorships, are ignored. That places the focus on larger companies. Above, we argue that this problem mainly applies to the IT value chain, which therefore cannot be interpreted to include many of the most dynamic elements of the software and information services industries.

Finally, the spatial results are not without occasional anomalies that are difficult to explain without more detailed research. For example, the finding of an IT complex that is comprised solely of Indian River County, Florida, without the likely influential counties immediately to the north (particularly Brevard, the location of Kennedy Space Center) may be a function of the underlying specification of the interaction among these counties in the spatial weights matrix of the *G* statistic. Altering the significance level bounds used to derive the complex maps is one way to explore the sensitivity of the results and identify complexes that require more intensive study. For example, the IT value chain around Indian River County would include Brevard County and Kennedy Space Center if we considered counties significant at the 90 percent level. More generally, this issue underscores the flexibility of the approach as well as its exploratory nature. Maps generated by altering assumptions regarding statistical significance, the variables, the weights matrix, and other features of the analysis can be compared and evaluated against different hypotheses of industrial location.

REFERENCES

Anselin, Luc. 1995. “Local Indicators of Spatial Autocorrelation – Lisa,” *Geographical Analysis*, 27, 93–115.

Bailey, Trevor C., and Anthony C. Gatrell. 1995. *Interactive Spatial Data Analysis*. Essex, England: Longman.

Bergsman, Joel, Peter Greenston, andRobert Healy. 1972. “The Agglomeration Process in Urban Growth,” *Urban Studies*, 9, 263–288.

———. 1975. “A Classification of Economic Activities Based on Location Patterns,” *Journal of Urban Economics*, 2, 1–28.

Campbell, John. 1972. “Growth Pole Theory, Digraph Analysis and Interindustry Relationships,” *Tijdschrift Voor Economische En Sociale Geographie*, 63, 79–87.

———. 1974. “Selected Aspects of the Interindustry Structure of the State of Washington, 1967,” *Economic Geography*, 50, 35–46.

Czamanski, Stan. 1974. *Study of Clustering of Industries*. Halifax, Canada: Institute of Public Affairs, Dalhousie University.

———. 1976. *Study of Spatial Industrial Complexes*. Halifax, Canada: Institute of Public Affairs, Dalhousie University.

———. 1977. “Needless Complexity in the Identification of Industrial Complexes: A Comment,” *Journal of Regional Science*, 17, 455–457.

Czamanski, Stan, and Luiz Augusto Ablas. 1979. “Identification of Industrial Clusters and Complexes: A Comparison of Methods and Findings,” *Urban Studies*, 16, 61–80.

Diggle, Peter J., and Amanda D. Chetwynd. 1991. “Second-Order Analysis of Spatial Clustering for Inhomogenous Populations,” *Biometrics*, 47, 1155–1163.

Ellison, Glenn, and Edward L. Glaeser. 1997. “Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach,” *Journal of Political Economy*, 105, 889–927.

Feser, Edward J., and Edward M. Bergman. 2000. “National Industry Cluster Templates: A Framework for Applied Regional Cluster Analysis,” *Regional Studies*, 34, 1–19.

Feser, Edward J., and Stuart H. Sweeney. 2000. “A Test for the Coincident Economic and Spatial Clustering of Business Enterprises,” *Journal of Geographical Systems*, 2, 349–373.

Getis, Arthur, and J. Keith Ord. 1992. “The Analysis of Spatial Association by Use of Distance Statistics,” *Geographical Analysis*, 24, 189–206.

Gordon, Ian R., and Philip McCann. 2000. “Industrial Clusters: Complexes, Agglomeration and/or Social Networks?” *Urban Studies*, 37, 513–532.

Hertog, Pim den, Edward M. Bergman, and David Charles. 2001a. “Creating and Sustaining Innovative Clusters: Towards a Synthesis,” in P. d. Hertog, E. M. Bergman, and D. Charles (eds.), *Innovative Clusters: Drivers of National Innovation Systems*. Paris: OECD, pp. 405–419.

Hertog, Pim den, Edward M. Bergman, and David Charles, eds. 2001b. *Innovative Clusters: Drivers of National Innovation Systems.* Paris: Organisation for Economic Cooperation and Development.

Howe, Eric. 1991. “Simple Industrial Complexes,” *Papers in Regional Science*, 70, 71–80.

Isard, Walter, Eugene W. Schooler, and Thomas Vietorisz. 1959. *Industrial Complex Analysis and Regional Development*. Cambridge, MA: MIT Press.

Krugman, Paul. 1991. *Geography and Trade*. Cambridge, MA: MIT Press.

Latham, William R. 1976. “Needless Complexity in the Identification of Industrial Complexes,” *Journal of Regional Science*, 16, 45–55.

———. 1977. “Needless Complexity in the Identification of Industrial Complexes: A Reply,” *Journal of Regional Science*, 17, 459–461.

Ó hUallacháin, Breandán. 1984. “The Identification of Industrial Complexes,” *Annals of the Association of American Geographers*, 74, 420–436.

———. 1985. “Complementary Linkages and the Hierarchical Structure of Regional Economies,” *Geographical Analysis*, 17, 130–142.

Ord, J. Keith, and Arthur Getis. 1995. “Local Spatial Autocorrelation Statistics: Distributional Issues and an Application,” *Geographical Analysis*, 27, 286–306.

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

———. 2002. *Clusters of Innovation: Regional Foundations of U.S. Competitiveness*. Washington, DC, Council on Competitiveness.

———. 1998. “Clusters and the New Economics of Competition,” *Harvard Business Review*, 77–90.

———. 2000. “Location, Competition, and Economic Development: Local Clusters in a Global Economy,” *Economic Development Quarterly*, 14, 15–34.

Richter, Charles. 1969. "The Impact of Industrial Linkages on Geographic Association," *Journal of Regional Science*, 9, 119–128.

Roelandt, Theo J A and Pim den Hertog, eds. 1999. *Boosting Innovation: The Cluster Approach*. Paris: Organisation for Economic Co-operation and Development.

Roepke, Howard, David Adams, and Robert Wiseman. 1974. "A New Approach to the Identification of Industrial Complexes Using Input-Output Data," *Journal of Regional Science*, 14, 15–29.

Steiner, Michael, ed. 1998. *Clusters and Regional Specialisation. European Research in Regional Science*. London: Pion.

Streit, Manfred E. 1969. "Spatial Associations and Economic Linkages between Industries," *Journal of Regional Science*, 9, 177–188.

APPENDIX

COMPONENT INDUSTRIES, BENCHMARK VALUE CHAINS

Information Technology and Instruments

SIC	Description
3471	Plating and polishing
3571	Electronic computers
3572	Computer storage devices
3575	Computer terminals
3577	Computer peripheral equipment, nec
3578	Calculating and accounting equipment
3579	Office machines, nec
3596	Scales and balances, except laboratory
3625	Relays and industrial controls
3629	Electrical industrial apparatus, nec
3651	Household audio and video equipment
3661	Telephone and telegraph apparatus
3663	Radio and TV communications equipment
3669	Communications equipment, nec
3672	Printed circuit boards
3674	Semiconductors and related devices
3675	Electronic capacitors
3676	Electronic resistors
3677	Electronic coils and transformers
3678	Electronic connectors
3679	Electronic components, nec
3694	Engine electrical equipment
3699	Electrical equipment and supplies, nec
3728	Aircraft parts and equipment, nec
3761	Guided missiles and space vehicles
3769	Space vehicle equipment, nec
3812	Search and navigation equipment
3821	Laboratory apparatus and furniture
3822	Environmental controls
3823	Process control instruments
3824	Fluid meters and counting devices
3825	Instruments to measure electricity

3826	Analytical instruments
3827	Optical instruments and lenses
3829	Measuring and controlling devices, nec
3841	Surgical and medical instruments
3844	X-ray apparatus and tubes
3845	Electromedical equipment
3861	Photographic equipment and supplies
3873	Watches, clocks, watchcases and parts
3931	Musical instruments
7371	Computer programming services
7372	Prepackaged software
7373	Computer integrated systems design
7374	Data processing and preparation
7375	Information retrieval services
7376	Computer facilities management
7377	Computer rental and leasing
7378	Computer maintenance and repair
7379	Computer related services, nec

Apparel

SIC	**Description**
2211	Broadwoven fabric mills, cotton
2221	Broadwoven fabric mills, manmade
2231	Broadwoven fabric mills, wool
2241	Narrow fabric and other smallwares mills
2251	Women's hosiery, except socks
2252	Hosiery, nec
2253	Knit outerwear mills
2254	Knit underwear mills
2257	Weft knit fabric mills
2258	Lace and warp knit fabric mills
2259	Knitting mills, nec
2261	Finishing plants, cotton
2262	Finishing plants, manmade
2269	Finishing plants, nec
2273	Carpets and rugs
2281	Yarn spinning mills
2282	Throwing and winding mills
2284	Thread mills
2296	Tire cord and fabrics
2297	Nonwoven fabrics
2298	Cordage and twine
2299	Textile goods, nec
2311	Men's and boys' suits, coats, and overcoats
2321	Men's and boys' shirts
2322	Men's and boys' underwear + nightwear
2323	Men's and boys' neckwear
2325	Men's and boys' trousers and slacks

2326	Men's and boys' work clothing
2329	Men's and boys' clothing, nec
2331	Women's and misses' blouses and shirts
2335	Women's, junior's, and misses' dresses
2337	Women's and misses' suits and coats
2339	Women's and misses' outerwear, nec
2341	Women's and children's underwear
2342	Bras, girdles, and allied garments
2353	Hats, caps, and millinery
2361	Girls' and children's dresses, blouses
2369	Girls' and children's outerwear, nec
2371	Fur goods
2381	Fabric dress and work gloves
2384	Robes and dressing gowns
2385	Waterproof outerwear
2386	Leather and sheep-lined clothing
2387	Apparel belts
2389	Apparel and accessories, nec
2395	Pleating and stitching
2397	Schiffli machine embroideries
2824	Organic fibers, noncellulosic
3965	Fasteners, buttons, needles, and pins

Motor Vehicle Manufacturing

SIC	**Description**
2273	Carpets and rugs
2299	Textile goods, nec
2396	Automotive and apparel trimmings
2399	Fabricated textile products, nec
2531	Public building and related furniture
2599	Furniture and fixtures, nec
2851	Paints, varnishes, lacquers, enamels, etc.
2891	Adhesives and sealants
3011	Tires and inner tubes
3052	Rubber and plastics hose and belting
3061	Mechanical rubber goods
3069	Fabricated rubber products, nec
3081	Unsupported plastics film and sheet
3082	Unsupported plastics profile shapes
3083	Laminated plastics plate and sheet
3084	Plastics pipe
3085	Plastics bottles
3086	Plastics foam products
3087	Custom compound purchased resins
3088	Plastics plumbing fixtures
3089	Plastics products, nec
3142	House slippers
3211	Flat glass

3229	Pressed and blown glass, nec
3231	Glass products, made of purchased glass
3465	Automotive stampings
3493	Steel springs, except wire
3519	Internal combustion engines, nec
3524	Lawn and garden equipment
3585	Refrigeration and heating equipment
3592	Carburetors, pistons, rings, valves
3641	Electric lamps
3651	Household audio and video equipment
3694	Engine electrical equipment
3711	Motor vehicles and car bodies
3713	Truck and bus bodies
3714	Motor vehicle parts and accessories
3715	Truck trailers
3716	Motor homes

Transportation and Shipping

SIC	**Description**
4212	Local Trucking without Storage
4213	Trucking, except local
4214	Local Trucking with Storage
4215	Courier Services, except Air
4221	Farm product warehousing and storage
4222	Refrigerated warehousing and storage
4225	General warehousing and storage
4226	Special warehousing and storage, nec
4231	Trucking terminal facilities
4311	U.S. Postal Service
4412	Deep sea foreign transportation of freight
4424	Deep sea domestic transportation of freight
4432	Freight transportation on Great Lakes-St. Lawrence
4449	Water transport of freight, nec
4481	Deep sea passenger transportation, except ferry
4482	Ferries
4489	Water passenger transportation, nec
4491	Marine cargo handling
4492	Towing and tugboat service
4493	Marinas
4499	Water transportation services, nec
4512	Air transportation, scheduled
4513	Air courier services
4522	Air transportation, nonscheduled
4612	Crude petroleum pipelines
4613	Refined petroleum pipelines
4619	Pipelines, nec
4731	Freight transportation arrangement
4741	Rental of railroad cars

4783	Packing and crating
4785	Vehicle inspection and weighing services
4789	Transportation services, nec

Pharmaceuticals

SIC	Description
2833	Medicinals and botanicals
2834	Pharmaceutical preparations
2835	Diagnostic substances
2836	Biological products except diagnostic