Excerpted and adapted from "The Economic Performance of Regions" by Michael E. Porter, *Regional Studies* Volume 37, Numbers 6-7 / August-October 2003

A. The Composition Of Regional Economies: Local, Resource Dependent and Traded Industries

The distribution of economic activity by industry over geography reveals three different broad types of industries, with very different patterns of spatial competition and locational drivers.

The first type of industry in regional economies is ***local*** industries. In these industries, employment is evenly distributed across all regions — that is, employment is roughly proportional to regional population. Local industries provide goods and services primarily to the local market, or the region in which the employment is located. Such industries compete in only a limited way with other regions. Most are services including local health services, most utilities, retailing, and many types of construction. A few goods producing industries are revealed as local, including bottled and canned soft drinks, newspapers, concrete products, and ready-mixed concrete.

A second type of industry is ***resource dependent*** industries. Employment in these industries is located primarily where the needed natural resources are found, but these industries compete with other domestic and international locations. Examples of such industries include uranium ore, logging, beet sugar, and freight transportation on the Great Lakes.

The third type of industries in regional economies is ***traded*** industries that are not resource dependent. These industries sell products and services across regions and often to other countries. They locate in a particular region based not on resources but on broader competitive considerations, and employment concentration varies markedly by region. Examples of traded industries include aircraft engines and engine parts, motion picture and videotape production, and automobile assembly.

We utilize the actual distribution of employment by industry to separate industries into these three groups, using data for 1996.[1] The U.S. Census County Business Patterns (CBP) data *understates* the true geographic concentration of traded industries by region because the local sales, service, distribution, and other support activities of traded industries based elsewhere are counted in the region in which the employment appears, even though the primary and headquarters activities are based elsewhere. This might be termed the local portion of traded industries. We utilize three measures of the variation of industry employment across geography to separate industries: the share of national employment for all states with LQ≥1; the mean location quotient (LQ) for the top five states ranked by LQ; and the employment GINI coefficient.

After examining the pattern of employment across geography in many industries, cutoffs were established for each variable: employment in states with LQ≥1 of ≥ 50% of total employment; mean LQ of the top five states ≥2; and employment GINI of 0.3. The vast majority of the 879 industries in the SIC system were clearly traded or local based on all three criteria. For the industries that met two but not all three criteria, we examined the actual distribution of employment as well as the industry definitions. Of those 62 industries, 18 were categorized as traded and the rest as local. We also identified a number of industries that were traded based on all three criteria but were local based on the industry definition (mostly retailers). We classified all of those as local after examining the employment distribution.

This process resulted in 241 local industries out of 879. Of the 638 traded industries, 48 had locational distributions and industry definitions tied heavily to the location of resource endowments. Our designation of resource-dependent industries was conservative, and only industries clearly dominated by resource endowments were included. This left 590 non-resource dependent traded industries. While the cutoff points used in developing the classifications were arbitrary, modifying the cutoffs led to only minor changes in the results.

[1] The categorization is highly stable from year to year.

B. Clusters of Traded Industries

A major constraint to the analysis of clusters has been the lack of a systematic approach to defining the industries that should be included in each cluster and the absence of consistent empirical data on cluster composition across a large sample of regional economies. Lack of large sample empirical data is understandable, since knowledge spillovers and other positive externalities are difficult if not impossible to measure directly.

We proceed indirectly, using the locational correlation of employment across traded industries to reveal externalities and define cluster boundaries. For example, if computer hardware employment is nearly always associated geographically with software employment, this provides a strong indication of locational linkages. Such a methodology exploits the unique characteristics of the U.S. economy which is by far the largest economy in the world, in which virtually every industry and cluster in any economy is present, and which consists of a large number of distinct but interdependent regions. This approach is not feasible in most if not all other countries.

We utilized states as the base unit of geography for computing locational correlations for two reasons. First, states involve less data suppression in the CBP data than EAs. Second, starting with larger geographic regions mitigates the problem of artificially high locational employment correlation coefficients when employment in a given traded industry is small or zero in many regions. The use of small regions, then,

can cause locational correlation across many industries to appear very high. The relevant geographic unit for a cluster varies by cluster and region. Clusters are often concentrated within a state and, conversely, clusters sometimes cross state lines. However, states are large enough and sufficiently diverse in economic landscape to reveal clusters. After defining clusters using states, we repeated the analysis using EAs. While the correlations were generally higher for EAs, the patterns were nearly identical to clusters defined using states.

Using CBP data for 1996, we identified pairs and then groups of tightly linked industries based on statistically significant locational correlations.[i] Standard clustering algorithms proved inadequate to revealing the multiple patterns of linkages across industries. To build up clusters, then, we proceeded pragmatically, beginning with small groups of obviously related industries and then tracing correlation patterns to others.

The major complexity arises because of spurious correlation, which can occur for several reasons. First, SIC industry definitions tend to be overly broad, hence two industries may be correlated overall though only a small portion of one industry involves the linked products or services. Second, the CBP data do not distinguish between employment in headquarters activities and that employment dispersed to serve local markets. This overstates true traded industry employment in many locations. Third, industries with a major presence in large employment states like California and New York can appear highly correlated with each other even though there is no economic relationship. Fourth, small industries can register small or zero employment in many

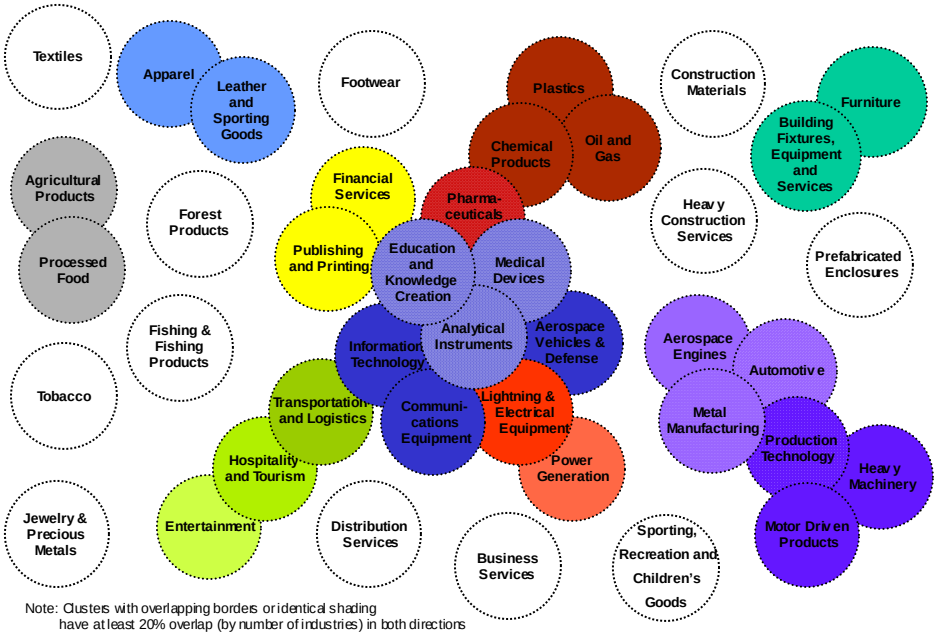
locations, making them appear correlated. Finally, industries can register high locational correlation if they are part of different clusters that appear in some of the same larger states, either by chance or for historical reasons related to natural resources. The strong position in Michigan of both automotive industries and industries related to office and commercial furniture, for example, creates a statistical correlation between the two groups of industries even though they are located in different parts of the state and have little or no economic relationship with each other.

We employed a sequence of steps to eliminate spurious correlation. First, we used detailed four-digit SIC industry definitions and lists of products included in each industry, together with industry knowledge, to reveal the likely presence of logical externalities. Focused case studies were conducted in unfamiliar industries to better understand the possible externalities present. Second, where there were no apparent externalities, we utilized the National 1992 Input-Output (I-O) Accounts from the Bureau of Economic Analysis to look for meaningful cross-industry flows.[ii] Note that input output links are just one of many forms of externalities or linkages between industries within a cluster, but have the advantage that systematic data is available even though industry definitions in the I-O tables are more aggregated than the 4-digit SIC codes we employ. Where there was no logical externality *and* the I-O data revealed no meaningful product flows, a correlation pair was excluded as spurious. Through this sequence of steps, we eliminated those pairs of correlated industries where there was no apparent basis for linkages.

This process resulted in 41 traded clusters in the U.S. economy, with an average of about 29 industries each.[iii] Each cluster has a different geographic pattern of employment. Clusters often contain both manufacturing and service industries as well as industries from various parts of the SIC system. Clusters, then, represent a different way of dividing the economy than is embodied in conventional industrial classification systems which are based primarily on product type and similarities in production.

We expected overlap of industries across clusters, and such overlap was indeed present empirically. Total cluster employment including overlap is 204% of total traded employment in 2000. So that, on average, each industry is part of about two clusters. Figure 18 provides a schematic representation of those clusters with substantial overlap. Some clusters are linked with several others, such as education and knowledge creation (significant overlap with eight other clusters) and analytic instruments (significant overlap with seven other clusters). Other clusters (e.g., textiles, forest products, distribution services) are relatively independent.

Figure 18: Schematic Diagram of Cluster Overlap in the United States Economy



Note: Clusters with overlapping borders or identical shading have at least 20% overlap (by number of industries) in both directions

The presence of overlapping industries across clusters leads to double counting of employment. In order to eliminate double counting for some analyses, we designated *broad* and *narrow* cluster definitions. Broad cluster definitions include all the industries included in a cluster. Narrow cluster definitions involve assigning each industry to the single cluster with which it has the strongest locational correlation. Here clusters are mutually exclusive.

We also subdivided each cluster into *subclusters*. Subclusters are subgroups of industries *within* the cluster whose locational correlations with each other were higher than with remaining industries. Subclusters are important because they can differ in sophistication, wage, and patenting rates. Different regions often have differing concentrations in some subclusters relative to others.

Separate subclusters were defined for the set of industries included in the narrow cluster definition and those in the remaining industries. In most cases, subclusters were quite sharply delineated. In other cases, judgments based on detailed industry definitions were made or subclusters were designated with only one constituent industry. In all, there were 264 subclusters for narrowly defined clusters, or an average of 6.4 subclusters per cluster. There were a total of 550 subclusters for broadly defined clusters, or an average of 13.4.

---

[i] This analysis was based on 1996, the most recent year then available. Replicating the analysis using more recent years revealed no material differences.

[ii] Application of the input-output method can be found in Tiebout (1957); Miernyk (1965); Nevin, Roe and Round (1966); Yan (1969); Richardson (1972); Lewis and McNicoll (1978); and Pullen and Proops (1983) and others. See Armstrong and Taylor (1985) for the summary on the input-output approach.

[iii] Amended clusters have been defined using NAICS data which prove to be very similar in composition but include a moderate number of additional industries. We do not utilize NAICS clusters here due to lack of historical data.