Journal of Extension

8-1-2017

Using Proprietary Databases to Overcome Data Suppression in Industry Cluster Analysis

Gilbert Michaud
Ohio University

G. Jason Jolley
Ohio University

Using Proprietary Databases to Overcome Data Suppression in Industry Cluster Analysis

Abstract

Extension agents are frequently tasked with determining industry clusters that exist in a region to support economic development. However, data suppression issues recurrently prohibit a comprehensive understanding of heavily concentrated firms in a region, particularly in rural areas. This article discusses the use of North American Industry Classification System codes within the LexisNexis Academic database as a technique for locating data about specific firms and analyzing regional industry clusters. This approach provides a practical and cost-efficient method for Extension agents and other researchers and practitioners to identify clusters and gather firm information in small geographies throughout the United States.

Gilbert Michaud
Adjunct Assistant
Professor
michaudg@ohio.edu

G. Jason Jolley
Assistant Professor
jolleyg1@ohio.edu

Voinovich School of
Leadership and Public
Affairs
Ohio University
Athens, Ohio

Introduction

Economic development researchers and Extension agents play an important role in helping communities understand their local economies, particularly when assessing regional strengths to develop jobs and community wealth (Blaine, Bowen-Ellzey, & Davis, 2011). For instance, business attraction efforts have often focused on the local competitive advantages created by large networks of interconnected organizations within a particular industry (Kuah, 2002). This strategy, known as cluster analysis, may help drive job creation, diversification, and overall economic growth (Boari, 2001; Feser, 2009; Matoon & Wang, 2014).

Nevertheless, there remain several limitations surrounding the identification of firms in a particular industry, especially in rural areas, where Extension agents are likely to work. In this article, we discuss the concept of industry clusters and then describe some of the key data issues, including suppression, in identifying these clusters. We describe the use of proprietary databases, specifically LexisNexis Academic, as a means of mitigating such suppression issues, using Ohio's plastics industry as an example.

The plastics industry in Ohio is a useful example due to the recent boom in natural gas extraction from the Marcellus and Utica shale plays in the southern and eastern parts of the state (Ohio Department of Natural Resources Division of Geological Survey, 2016). By-products of natural gas extraction are important inputs for supporting mid- and downstream processing and manufacturing opportunities in the plastics industry, thus representing an emerging and intriguing regional cluster.

Data Suppression in Industry Cluster Analysis

Industry clusters are "geographic concentrations of interconnected companies and institutions in a particular field" (Porter, 1998, p. 78) that are "linked by similar needs such as production inputs, specialized labor, and technology" (Hagadone & Grala, 2012, p. 16). Put another way, industry clusters represent a geographic concentration of interrelated businesses, vendors, service providers, academic institutions, and other affiliated organizations surrounding a specific industry (Feser & Bergman, 2000; Morgan, 2007; Slaper & Ortuzar, 2015). Industry clusters are often identified through the use of the North American Industry Classification System (NAICS), which is "the standard used by Federal statistical agencies in classifying business establishments" (U.S. Census Bureau, 2017, para. 1).

All U.S. businesses are associated with a NAICS code (Mulangu & Clark, 2012). This classification system operates in a hierarchical manner, ranging from 2- through 6-digit codes that represent progressively narrower categories. Higher level codes (i.e., 2-digit) are less likely to be associated with suppression issues as these codes represent the broadest level sector (e.g., NAICS 32: Manufacturing) and, thus, inherently capture the largest number of firms. The 3- and 4-digit NAICS codes represent subsectors and industry groups, respectively, and circumstantially suffer from suppression issues. Logically, 6-digit NAICS codes are the most likely to be associated with suppression as a larger number of establishments are included in this classification (e.g., NAICS 326160: Plastics Bottle Manufacturing).

Using this industry classification system involves trade-offs. At the state and regional levels, employment and establishment data are typically available at the 6-digit level. At the county level, 6-digit data are often suppressed due to the small number of establishments and/or lower levels of employment. Researchers particularly encounter such data suppression when investigating rural areas, areas with a single large employer, or industries with only a few employers (Porter, Ketels, Miller, & Bryden, 2004; U.S. Department of Labor Bureau of Labor Statistics, 2016). The Bureau of Labor Statistics purposely conceals its methods for protecting the privacy of companies to prevent researchers from disassembling the data to locate these hidden firms (U.S. Department of Labor Bureau of Labor Statistics, 2016).

The U.S. Cluster Mapping Project from Harvard University is a respected resource commonly used for cluster identification. This accessible, nonproprietary online tool offers strong surface-level data on employment and wages, as well as numerous data visualization tools, at the regional level. It groups over 1,000 6-digit NAICS codes into 51 traded clusters and 16 local clusters (U.S. Cluster Mapping Project, 2014). However, due to data suppression, this tool cannot provide county-level data on the number of establishments and employment at the industry level. This circumstance prohibits Extension agents from using the tool to conduct industry cluster analyses in smaller, rural geographies.

Using Proprietary Databases: Ohio Plastics Example

Most university libraries offer access to the LexisNexis Academic company dossier database. With this database, Extension agents and researchers can use the company list option to search for firms with specific NAICS codes in their regions and obtain company data such as name, contact information, number of employees, and annual revenue. Other business databases, such as Hoover's, Reference USA, or Dun & Bradstreet, also could be used, but for the example provided here, we used LexisNexis, which is a well-regarded and free business research platform.

We executed this methodology using 20 NAICS codes for plastics in Ohio. LexisNexis Academic displayed 1,300 relevant firms, but we dropped the 320 firms for which no employees were listed, although a researcher could

follow up with the applicable firms to obtain this information, if desired. Next, we deleted 18 firm duplicates, organized the firms by county, and then separated them by JobsOhio region. JobsOhio is an economic development nonprofit that arranges the state into six geographic regions. For each region, we determined the number of firms in the plastics industry cluster and the total and average numbers of employees, as displayed in Table 1.

Table 1.

Summary of Ohio's Plastics Industry Clusters

JobsOhio region	No. of firms	Total employment	Average employment per firm
TEAM NEO (northeast Ohio)	487	36,617	75
REGIONAL GROWTH PARTNERSHIP (northwest Ohio)	119	12,013	101
REGIONAL ECONOMIC DEVELOPMENT INITIATIVE CINCINNATI (Greater Cincinnati region)	108	6,737	79
DAYTON DEVELOPMENT COALITION (Dayton region/western Ohio)	108	6,561	61
COLUMBUS REGION (central Ohio)	106	5,170	49
APPALACHIAN PARTNERSHIP FOR ECONOMIC GROWTH (southeast Ohio/Appalachia)	34	1,055	31
Total/average	962	68,153	66

Currently, there are 962 firms in the plastics industry in Ohio. The largest cluster of plastics firms exists in northeast Ohio, with 487 firms and nearly 37,000 total employees.

Conclusion

This suggested technique passes the face validity test for use by economic development researchers seeking to locate specific firms within local industry clusters. Extension agents may find the identification of specific employers in certain sectors useful for business attraction efforts, job training programs, peer-to-peer information sharing, and business client education, among other purposes. One can easily determine which NAICS codes are relevant to an industry and then use LexisNexis Academic or a similar proprietary database to drill down to specific firm data by area code, county, city, or zip code. Such capability can be a boon to Extension researchers needing to perform cluster analyses for the communities they serve.

References

Blaine, T. W., Bowen-Ellzey, N., & Davis, G. A. (2011). Helping clientele understand elements of the local economy through input–output modeling. *Journal of Extension*, *49*(1), Article 1FEA5. Available at: https://joe.org/joe/2011february/a5.php

Boari, C. (2001). *Industrial clusters, focal firms, and economic dynamism: A perspective from Italy*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.195.5611&rep=rep1&type=pdf

Feser, E. (2009). Clusters and strategy in regional economic development. *Industry Clusters*, *3*, 26–38.

Feser, E. J., & Bergman, E. M. (2000). National industry cluster templates: A framework for applied regional cluster analysis. *Regional Studies*, *34*(1), 1–19.

Hagadone, T. A., & Grala, R. K. (2012). Business clusters in Mississippi's forest products industry. *Forest Policy and Economics*, *20*, 16–24.

Kuah, A. T. H. (2002). Cluster theory and practice: Advantages for the small business locating in a vibrant cluster. *Journal of Research in Marketing and Entrepreneurship*, *4*(3), 206–228.

Matoon, R., & Wang, N. (2014). Industry clusters and economic development in the Seventh District's largest cities. *Economic Perspectives*, *2Q*, 52–66.

Morgan, J. Q. (2007). Industry clusters and metropolitan economic growth and equality. *International Journal of Economic Development*, *9*(4), 307–375.

Mulangu, F., & Clark, J. (2012). Identifying and measuring food deserts in rural Ohio. *Journal of Extension*, *50*(3), Article 3FEA6. Available at: http://www.joe.org/joe/2012june/a6.php

Ohio Department of Natural Resources Division of Geological Survey. (2016). *Marcellus and Utica shales data*. Retrieved from http://geosurvey.ohiodnr.gov/energy-resources/marcellus-utica-shales

Porter, M. E. (1998). Clusters and the new economics of competition. *Harvard Business Review*, *76*(6), 77–90.

Porter, M. E., Ketels, C. H. M., Miller, K., & Bryden, R. T. (2004). *Competitiveness in rural U.S. regions: Learning and research agenda*. Retrieved from http://www.isc.hbs.edu/Documents/ced/EDA_RuralReport_20040621.pdf

Slaper, T., & Ortuzar, G. (2015). Industry clusters and economic development. *Indiana Business Review*, *Spring*, 7–9.

U.S. Census Bureau. (2017). *North American industry classification system*. Retrieved from http://www.census.gov/eos/www/naics/

U.S. Cluster Mapping Project. (2014). *About: Cluster mapping methodology*. Retrieved from http://clustermapping.us/content/cluster-mapping-methodology

U.S. Department of Labor Bureau of Labor Statistics. (2016). *Quarterly census of employment and wages: Frequently asked questions (FAQs)*. Retrieved from http://www.bls.gov/cew/cewfaq.htm

ed@joe.org.

If you have difficulties viewing or printing this page, please contact _JOE_ Technical Support