Economic Development Quarterly
Volume 22 Number 4
November 2008 324-344

10.1177/0891242408325419
http://edq.sagepub.com
hosted at
http://online.sagepub.com

Clusters and Economic Development Outcomes

An Analysis of the Link Between Clustering and Industry Growth

Edward Feser
University of Illinois at Urbana-Champaign
Henry Renski
University of Massachusetts Amherst
Harvey Goldstein
MODUL University-Vienna

Much of the existing empirical research on industry clusters focuses on the detection of clusters for economic development purposes. There are comparatively few studies that relate identified clusters to business and industry growth or that trace changes in designated clusters over time. This article seeks to better understand the link between industry clustering and regional economic outcomes. In a comprehensive study sponsored by the Appalachian Regional Commission and released in 2002, the authors identified technology-based clusters within and on the border of the Appalachian region. The Appalachian technology clusters constituted subregional concentrations of related industrial, research and development, and university-based strengths as of the middle to late 1990s. In this article, the authors investigate how the industries in the identified clusters fared over the subsequent several years in terms of employment and new business formation. They find evidence that clustering is associated with new business formation for selected technology industries but not with employment growth.

Keywords: *industry cluster; Appalachia; regional growth; entrepreneurship*

In late 2001 and early 2002, we conducted an extensive analysis of technology-oriented economic activity in the 406-county region served by the Appalachian Regional Commission (ARC). We examined technology-related assets within and bordering Appalachia from two perspectives: the industrial base (technology-related goods and services production and industry innovation activity) and the knowledge base (knowledge-creating institutions and programs). Our objective was to identify functional and spatial clusters of activity that demonstrated legitimate existing or potential strength in the region vis-à-vis the broader U.S. economy. We defined the areas of overlap between the industrial and the knowledge/innovation strengths as Appalachia's unique technology clusters, places where a sophisticated knowledge infrastructure was joined with a substantial related economic base. The investigation found some 100 high-technology clusters in 8 different industrial categories. Although we completed the work in 2002, the majority of the data defining the clusters were vintage 1998 because of lags in the release of various data series.

The clusters helped characterize broader intraregional differences in the location of technology-oriented activity in Appalachia, such as between the region's core and periphery, its main subregions (north, central, and south), and its metro and nonmetro areas. However, of greater interest to the funding agency (the ARC) was the utility of the clusters as economic development targets. Although we did not attempt to formally forecast the growth of the clusters or the regions in which they were situated, we did claim that they represented logical candidates for technology-based economic development strategies given their demonstrated competitive success. Our analysis was not unlike the applied industry cluster studies that are now in vogue in economic development practice around the world, where the aim is to sift through large volumes of data to identify the existing and emerging economic strengths that appear poised to drive a region's growth.

In this article, we take advantage of the passage of time since our 2002 study to explore the relationship between clustering as we ourselves defined it and subsequent economic performance. Specifically, we ask two empirical questions. The first is whether there is an association between the level of clustering and employment growth for various technology industries. Because our study produced several continuous measures of clustering, we test for the association between growth and clustering by first separating Appalachian counties into growth and no-growth samples for each technology industry and then testing for mean differences in clustering levels between the two samples. The analysis asks, "Are the counties in which we observe employment growth between 1998 and 2002 the same counties in which we observe clustering in 1998?" Our second empirical question asks whether clustering is associated with more business start-ups. We examine the link between clustering and start-ups using a zero-inflated negative binomial model of new business entrants.

The second section of our article summarizes the data and methods used to detect significant regional concentrations of high-technology activity in Appalachia, followed by sections that investigate the link between clustering and performance by testing for an association between employment growth and clustering and between business start-up rates and clustering. We find very little evidence of a relationship between job growth and clustering in the region. In effect, spatial clustering as defined in the late 1990s is not a good predictor of job growth over the subsequent several years. We do find, however, higher new business formation rates for clustered technology industries, although the results vary by the measure used to specify the clustering and by the type of business establishment. We also find that new business formation may be stronger in clusters of relatively recent vintage, indicative of cycles in cluster growth. The article concludes with a brief summary and discussion of the research and policy implications of our findings.

Measuring Clustering: The ARC Study

In our 2002 study (Feser, Goldstein, Renski, & Renault, 2002), our methodology for identifying Appalachia's technology clusters was based on an explicit strategy of triangulation.[1] Given the myriad plausible ways in which high-technology activity might be defined and measured, we opted to use multiple data sources, classification schemes, competitiveness criteria, and indicators. The logic was that we could be more confident of the strength and depth of the science and technology base of a given subregion within Appalachia if it stood out along multiple science and technology dimensions rather than simply along one or two.

Any quantitative exercise of this kind is analytically and empirically challenging. We do not claim that the classification schemes we adopted were without any weaknesses or that the measures or criteria we used were beyond dispute. Moreover, we do not claim to have evaluated exhaustively all salient kinds of technology-based activity in the region. The findings we generated should be used cautiously given potential aggregation biases related to industry, geography, and discipline or technology area, as well as the absence of information on the presence of informal networks and the extent of real knowledge spillovers between businesses. Nevertheless, despite the attendant limitations, we believe that our approach of systematically quantifying clustering across the region is useful for highlighting trends, weaknesses, and strengths in Appalachia's economy that warrant further investigation.

Study Area

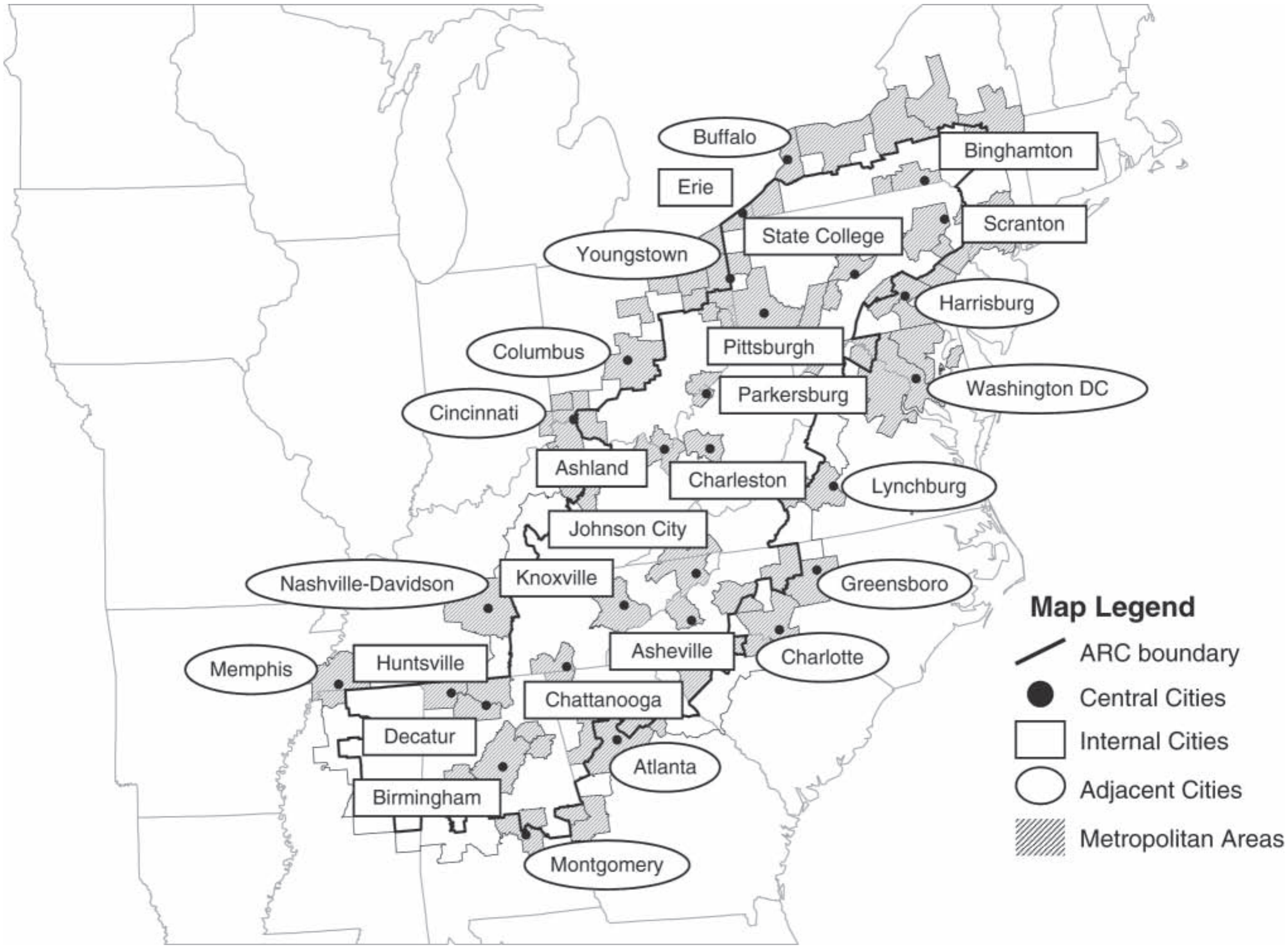
The study area was the 406 counties under the policy jurisdiction of the ARC. To accommodate spatial spillovers from neighboring areas, we also considered activity within a border territory or "buffer" of counties and metropolitan areas adjacent to the ARC region. Figure 1 depicts the ARC region, the buffer area, and metropolitan areas within and bordering the region.

The High Technology Industrial Base

Our first step in characterizing geographic clustering in Appalachia's high-technology industrial base was to identify the set of technology industries that would be the focus of study. We began with a classification created by Feser and Koo (2000) that organizes all industries identified by the U.S. Bureau of Labor Statistics as technology-intensive into a total of eight value chains based on a detailed analysis of national input-output patterns. The value chains represent the core technology-intensive buyer-supplier chains in the U.S. economy as of the mid-1990s. Each of the eight chains listed comprised between 8 and 30 diverse 4-digit Standard Industrial Classification (SIC) codes. The chains were not mutually exclusive because some sectors are linked to multiple industries. The chains are a good starting point for assessing unique industrial specializations in Appalachia because they are groups of industries that share similar competitive pressures and, in many cases, utilize similar production technologies. They served as the common reference classification of functional high-technology areas for the remainder of the study. We established concordances between the value-chain classification and all other variable classifications (e.g., university disciplines, patents) to derive a single consistent set of functional technology overlays.

Our decision to use a classification of eight technology-based value chains as the study baseline met our practical need for a methodology that was manageable for a large

Figure 1
Appalachian (ARC) Study Region, Major Internal and Adjacent Cities and Metropolitan Areas



and diverse region of hundreds of counties. We recognized that the approach might obscure some important underlying industrial strengths if such strengths were subspecializations in one or more of the eight value chains or if they spanned various chains in unique ways. A common classification was necessary because project resources did not permit conducting the equivalent of 406 individual county cluster studies. Moreover, we believe the value chains best captured the notion of linked and related industries underlying most industry cluster theories.

We used location quotients and the Getis/Ord *G* statistic (Getis, 1984; Ord & Getis, 1995) to identify localized concentrations of the value chains. Location quotients evaluate the level of specialization or relative concentration of a given industry for individual counties, ignoring activity in neighboring counties. The *G* statistic, which is very similar to the local Moran's *I,* helps reveal broader multicounty areas where technology-related activity is especially pronounced. The *G* works by analyzing the full multicounty spatial distribution of values of a given indicator to detect where high and low values of the indicator are proximate. Data were from the confidential unsuppressed Unemployment Insurance Data Base (UDB) of the U.S. Bureau of Labor Statistics. The UDB data, which contain employment and wage figures by establishment for all 50 states, permitted a fine-grained look at employment patterns even in very small counties.

The *G* measure for areal unit *i* for a given industry cluster is calculated as

$$G_i^* = \frac{\sum_j w_{ij}x_j - W_i\bar{x}}{s\sqrt{(nS_{1i} - W_i^2)/(n-1)}},\ all\ j \tag{1}$$

where x is the variable of interest (e.g., employment or patents), $\{w_{ij}\}$ is a spatial weights matrix that defines neighboring areas j to areal unit i, W_i is the sum of weights in $\{w_{ij}\}$, and

$$\bar{x} = \sum_j x_j/(n-1)$$
$$S_{1i} = \sum_j w_{ij}^2$$
$$s^2 = (\sum_j x_j^2/n-1) - (\bar{x})^2.$$

Although the normality of G depends partially on the number of neighbors, we made 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% significance level from a two-tailed normal distribution.

To implement G, we developed a spatial weights matrix $\{w_{ij}\}$ with adjacency defined by immediate neighbor counties inclusive of the county itself. Nonneighboring areal units were given a weight of 0. The value x of neighboring county j to county i was weighted by the degree of expected interaction between counties j and i

$$w_j = \frac{X_i X_j}{\sum_j X_i X_j} \tag{2}$$

where X is total exportable (or basic sector) 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 means that larger counties exert a heavier influence on neighboring counties than do smaller counties. There is also an implicit assumption of no interaction between nonneighboring counties.

To detect unique specializations of activity in and around Appalachia, we needed a means of controlling for the general tendency of industry to concentrate. To the extent that much of the commerce in a given area serves a local population or industry base, we should expect more employment in urban centers simply because those centers boast larger populations and industry employment (Feser & Sweeney, 2000; Sweeney & Feser, 1998). We sought to find local concentrations of activity beyond those that might be expected by the general distribution of employment and population. To do that, we regressed sectoral employment on total export-base employment using counties as the units of analysis. The residual from the regression—the difference between the predicted and actual employment—is an estimate of local activity other than that expected by the overall size of the place. The G is then calculated using the regression residuals.

The *current* level of industrial concentration may be a misleading indicator of the strength of a local cluster. Existing concentrations are sometimes the historical residue of an eroding economic advantage or outdated technology. From a development perspective, emerging clusters often are of greater policy interest than existing ones. Small but fast-growing clusters may offer greater opportunities for near-term employment growth in new technology areas or specialized market niches. We defined emerging clusters in Appalachia as regions with a concentration of value-chain employment growth. Specifically, we regressed each county's value-chain employment growth on its total employment growth to create the residuals for use with the G statistic, identifying regional growth hot spots.

Knowledge Infrastructure

The county-level location quotients, G statistic on the level of employment, and G statistic on employment growth provide three different measures of industrial clustering for the value-chain industries. Our next step was to construct indicators of the region's knowledge infrastructure, which is based in 18 research universities and a limited number of other research institutions (e.g., federal government laboratories), nonprofit research and development (R&D) organizations, state-sponsored technology agencies, and private sector businesses engaged in innovation. The region's higher education network consists of over 250 universities, colleges, and community colleges offering degree programs and specialized training in 15 science- and engineering-related fields. In the academic year 1997 to 1998, 4-year institutions in the region and bordering areas conferred over 23,600 science and engineering degrees and 2-year colleges and institutes granted an additional 12,200 degrees. Many more students obtained relevant training at Appalachian community colleges without earning degrees. We limited our attention to the influence of research universities and industrial innovation, the latter as measured by patenting activity. Programs of state science and technology initiatives and nonprofits cannot be reliably associated with a specific substate area of service.

There are 11 research universities located in ARC territory: Carnegie Mellon University, Clemson University, Cornell University, Mississippi State University, Ohio University (consolidated in the available data but dominated by Ohio State University), Pennsylvania State University, the University of Alabama at Birmingham, the University of Pittsburgh, the University of Tennessee, Virginia Polytechnic Institute, and West Virginia

University. There are an additional six universities situated adjacent to or very near the ARC boundary: Auburn University, Georgia Institute of Technology, Emory University, the University of Georgia, the University of Mississippi, and the University of Alabama at Huntsville. We included the six adjacent schools in the analysis on the assumption that their close spatial proximity might yield spillovers into the ARC region. Unlike all of the other campuses, the University of Alabama at Huntsville is not classified as a doctoral university (extensive) by the Carnegie Foundation. We included it because of its very strong technology focus.

We developed three measures of university competitiveness or strength by discipline: (a) perceived faculty quality as judged by peers in 1995, (b) external research funding receipts in 1991 and 1999, and (c) the number of full-time graduate students enrolled in 1991 and 1999. To establish a common scale for combining the disparate dimensions of research strength, we converted the measures into national rankings and established a concordance between the eight value-chain categories and the disciplines. We then averaged the rankings across the disciplines within each technology area. For example, Cornell University's rank for sponsored research relevant to the chemicals and plastics industry is the arithmetic average of its ranks for the chemical engineering, materials engineering, and chemistry disciplines.

Given the rankings on the three indicators, we identified Tier 1 universities as those with an average rank in the U.S. top 20 for at least two out of the three measures. Tier 2 schools are those with (a) an average rank in the U.S. top 20 for research funding or faculty quality, (b) an average rank in the U.S. top 40 for all three measures, or (c) an average rank in the U.S. top 20 for number of graduate students and a rank in the U.S. top 40 for either (or both) faculty quality or research funding (see Table 1). Our criteria effectively considered sponsored research and faculty quality as the leading barometers of a university's research capacity and output.

Appalachia's science and innovation base also includes many private sector businesses that actively engaged in research, applied innovation, and development. Unfortunately, data on private sector R&D activity are very limited. The National Science Foundation's industry surveys are based on very small samples and cannot be disaggregated to the substate level. Although counts of patents cannot be regarded as direct proxies of private sector R&D generally, they can provide a partial picture of the geographical distribution of private sector science and innovation in the region. A patent is an attempt by an inventor to appropriate fully and exclusively any returns derived from an innovation, at least for a limited period. Patent grants by sector are thus a partial indicator of applied innovative activity. Although some patents are granted to universities and nonprofit R&D performers, the vast majority are secured by private industry.

We used county-level utility patent data from the U.S. Patent and Trademark Office (USPTO, 2000) for the period from 1990 to 1999 to calculate patent-based G statistics for the study region. The USPTO assigns patents to counties based on residence of the inventor. Patents are initially classified by invention or product, which the USPTO then reclassifies into industries using the 1972 SIC definitions. Using the USPTO SICs, we organized patents into 10 technology categories that corresponded as closely as possible to the 8 value-chain categories. The USPTO commonly assigns a single patent to multiple SICs and therefore a patent could be included in more than one technology category, in turn making it relevant to more than one value chain.

Identifying Technology Clusters

Table 2 summarizes our various measures of spatial clustering. For each technology value chain and for each Appalachian county, we have 1998 value-chain employment (an absolute, single-county measure), a location quotient on 1998 employment (a relative, single-county measure), a G statistic on 1998 employment concentration (a multicounty spatial clustering measure of the *level* of value-chain employment), a G statistic on 1989–1998 employment growth concentration (a multicounty spatial clustering measure of the *change* in value-chain employment), a G statistic-based measure of 1990–1999 patent concentration (a multicounty spatial clustering measure of innovation activity) and the distance (in miles) to the nearest university ranked in the first or second tier in one or more relevant scientific disciplines. Using a geographic information system (GIS), we overlaid electronic maps of these variables and, from the overlays, identified subregions of joint industrial and knowledge infrastructure strength for each of the eight value chains.

Figure 2 is an example of one of the overlays. Although the specific geography of the clusters is necessarily inexact, several general findings emerged. First, Appalachia's major localized technology-related strengths as of the late 1990s were in three major areas: chemicals/plastics, industrial machinery, and motor vehicles and related industries. There were relatively few places within the region with overlapping industrial and knowledge infrastructure in information technology (IT), communication services and software, and pharmaceuticals.

Table 1
University Research and Development Strengths by Technology Area: First- and Second-Tier Strengths Based on U.S. Rank

University	Chemicals and Plastics	Information Technology and Instruments	Industrial Machinery	Motor Vehicles	Aerospace	Household Appliances	Communication Software and Services	Pharmaceuticals and Medical Technologies
University of Alabama at Birmingham								
University of Alabama at Huntsville			2	2		2		
Auburn University								
Carnegie Mellon University		2		1	2	1	2	2
Clemson University						2		
Cornell University	1	1	1	1	1	1	1	1
Emory University								2
Georgia Institute of Technology	1	1	1	1	2	1	2	
University of Georgia								
University of Kentucky								
Mississippi State University								
University of Mississippi								
Ohio University								
Pennsylvania State University	1	1	1	1	2	1	2	
University of Pittsburgh								2
University of Tennessee at Knoxville			2					
Virginia Polytechnic Institute		2		1		1	2	
West Virginia University								

Source: Feser, Goldstein, Renski, and Renault (2002).

Table 2
Measures of Technology Clusters in Appalachia

Concept	Classification	Variable	Measure	Data Source
Value chain absolute size in county	Standard Industrial Classification (SIC) industries in technology value chains identified via an input-output analysis of buyer-supplier patterns	Value-chain employment, 1998	Total (or level)	Confidential Covered Employment and Wages (CEW) files, U.S. Bureau of Labor Statistics (BLS)
Value chain relative size in county	SIC industries in technology value chains identified via an input-output (IO) analysis of buyer-supplier patterns	Value-chain employment, 1998	Location quotient	Confidential CEW files, U.S. BLS
Existing multicounty technology value chain clusters	Value chains developed via an IO analysis of buyer-supplier patterns among technology-intensive industries	Value-chain employment, 1998	*G* statistic	Confidential CEW files, U.S. BLS
Emerging multicounty technology value chain clusters	Value chains developed via an IO analysis of buyer-supplier patterns among technology-intensive industries	Change in value-chain employment, 1989 to 1998	*G* statistic	Confidential CEW files, U.S. BLS
Clusters of patent activity	In fields relevant to value chains	Utility patents granted during the period from 1990 to 1999	*G* statistic	U.S. Patent and Trademark Office (USPTO, 2000)
University research infrastructure	In disciplines related to value chains	University nationally ranked in related disciplines	Distance (in miles) to ranked university	National Science Foundation Web CASPAR; National Research Council (1995)
Technology cluster county	A dummy variable, by value chain	= 1 if (either the 1998 employment *G* or employment change *G* is statistically significant, or the employment location quotient > 1) AND (either the patent *G* is statistically significant, a university ranked in relevant field is within or adjacent to the county, or the patent location quotient > 1), set equal to 0 otherwise		

Source: Feser, Goldstein, Renski, and Renault (2002).

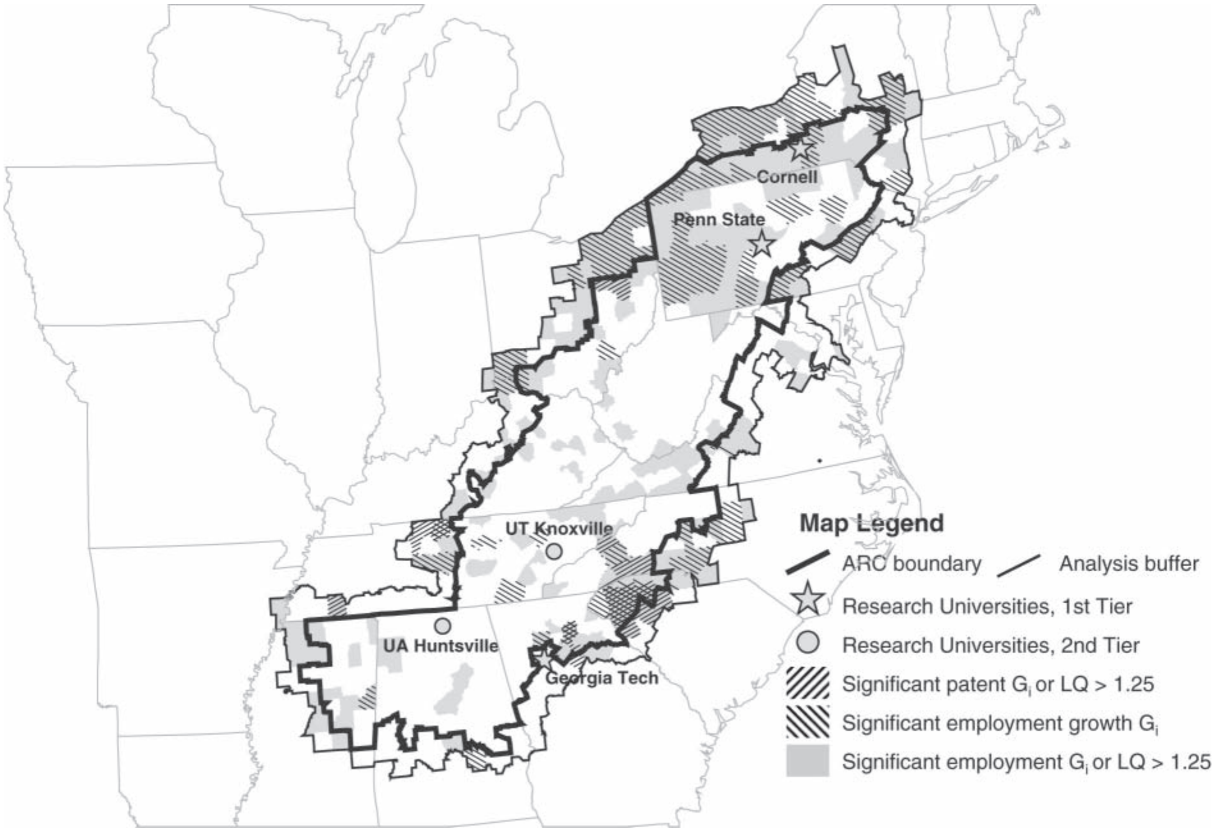
Second, the distribution of clusters in the region was highly uneven, with most found in the northern (New York, Pennsylvania, and northern Ohio) and southern (North and South Carolina, Tennessee, Georgia, and Alabama) thirds of the region, and relatively few in central Appalachia (southern Ohio, West Virginia, Virginia, and Kentucky). Third, the uneven geography of the clusters varied substantially by technology area. The chemicals/plastics and IT/instruments clusters are relatively evenly distributed among the northern, central, and southern thirds of Appalachia, whereas industrial machinery is nearly exclusively a northern and southern strength and clusters in communications services and software, aerospace, and pharmaceuticals and medical technologies were most common in the north. Finally, many of the region's technology clusters are located on the periphery, often anchored in metropolitan areas centered just outside the region, such as Cincinnati, Atlanta, and Washington, D.C.

Our 2002 study produced a rich data set. With the passage of time, it can be used to explore the relationship between clustering and subsequent economic growth, the focus of the next two sections.

Clustering and Employment Growth in Appalachia

In this section, we investigate the association between several different measures of localized clustering as of 1998 and 1999 and employment growth during the period from 1998 to 2002. Clusters are particularly alluring to economic developers because of the prevailing belief that they produce measurably higher income and employment gains than do businesses located in relative isolation (Porter, 2003). Clustering presumably confers substantial economic advantages on member firms and industries.

Figure 2
Technology Clusters in Appalachia: Industrial Machinery



A major challenge facing studies of clustering and economic performance is the substantial change in industrial classification schemes over time, making tracing the growth of identified spatial clusters very difficult. In our case, our portrait of clustering in Appalachia was developed with data reported under the SIC system. Most federal data series in the United States switched to the North American Industrial Classification System (NAICS) in 1998. Therefore, our *ex post* industry performance data are reported under a substantially different classification scheme from those of our *ex ante* analysis. With NAICS data, we are unable to construct industry value chains after 1998 that are directly comparable to those developed for preceding years using the SIC system.

We work around the problem by using industries rather than aggregate clusters as the units of analysis. Although the concordance between SIC and NAICS codes is imperfect, making consistent measurement of the growth of an SIC-based *cluster as a whole* after 1998 impossible, we can still investigate whether higher performing NAICS technology industries that are closely aligned with various SIC-based value chains tended to be situated in Appalachian regions with identified technology clusters at the beginning of the period. We exclude one value chain—household appliances—from our analysis in this section because it is present in too few counties in the region to yield sufficient sample sizes for our statistical tests.[2] Our employment data for technology-intensive NAICS industries are from 1998 to 2002 *County Business Patterns* files that have been adjusted for confidentiality suppressions (Isserman & Westervelt, 2006).

We use a series of differences of means tests to explore the relationship between technology clustering and subsequent industry employment growth. Specifically, we separate Appalachian counties into two samples: those that posted employment gains in a given NAICS technology industry *i* between 1998 and 2002 (*growth counties*) and those that did not (*no-growth counties*). We then test whether the mean level of beginning-of-period clustering for the sample of growth counties significantly differs from the mean level for the sample of no-growth counties. In the case of employment concentration,

Table 3
Employment Trends, Technology-Intensive Industries in Appalachia

NAICS Industries in Value Chains	1998	2002	Percentage Change 1998–2001	Percentage Change 2001–2002	Percentage Change 1998–2002
Chemicals and plastics	86,067	69,189	–13.6	–7.0	–19.6
Information technology and instruments	136,253	104,636	–6.9	–17.5	–23.2
Industrial machinery	65,445	56,377	–0.1	–13.8	–13.9
Motor vehicles	79,871	77,858	5.3	–7.4	–2.5
Aerospace	20,676	15,796	–23.4	–0.3	–23.6
Communications services and software	178,619	210,053	16.6	0.8	17.6
Pharmaceuticals and medical technologies	60,354	65,109	5.3	2.5	7.9

Source: Suppression-adjusted *County Business Patterns* (Isserman & Westervelt, 2006).
Note: NAICS = North American Industrial Classification System.

employment growth concentration, and patenting concentration, we hypothesize that the mean clustering levels for growth counties will exceed the mean levels for no-growth counties. In the case of distance to ranked universities, we expect the mean level for growth counties to be lower than the mean level for no-growth counties, because we are hypothesizing that NAICS technology industries benefit from being closer to universities with strengths in related scientific disciplines.

We constructed samples of growth/no-growth counties two different ways as a robustness check. In the first set of difference of means tests, growth counties are those posting employment growth in NAICS technology industry *i* of at least one worker whereas no-growth counties are those posting an employment decline in the industry of at least one worker. In the second set of tests, growth counties are those registering employment gains in NAICS industry *i* of least 50 workers during the period, and no-growth counties are those that suffered declines of at least 50 workers. The second set of tests uses growth/no-growth samples in which the employment change is more significant to reduce the influence of small, comparatively negligible employment changes. Both tests exclude counties in which employment did not change at all, the overwhelming majority of which are counties with no activity in the given industry in either period.

It is possible the relationship between clustering and subsequent performance is different prior to the 2001 U.S. recession from that during the recession. Therefore, we also test for differences in mean clustering for two subperiods—1998 to 2001 and 2001 to 2002—in addition to the 4-year period from 1998 to 2002. With six different measures of clustering, two sets of samples, seven value chains, and three periods, we tested a total of 252 differences. Essentially, we are looking for broad patterns across the 252 tests that accord with our hypothesis that technology industries located in spatial technology clusters in 1998 subsequently produced more net new jobs than did industries located in noncluster Appalachian locations.

We should note that employment in many technology-intensive industries in Appalachia contracted during the period from 1998 to 2002. Worst hit were industries in IT and aerospace; both groups posted job declines of more than 23% (see Table 3). Overall job growth occurred in communications services/software and pharmaceuticals/medical technologies. These overall trends mask subregional trends, however. Job dynamics during the period vary widely across Appalachia, with some counties adding jobs and others losing them. Indeed, the subregional variation in employment growth across technology value chains provides the basis for our statistical tests.

The results of the tests for the period from 1998 to 2002 are reported in Table 4, and results for the periods from 1998 to 2001 and from 2001 to 2002 are available upon request. Table 5 summarizes all of the results qualitatively. Overall, of 252 differences, we find only 47 that are statistically significant using a relatively liberal p value of .10. Of those 47, just 14 significant differences are in the hypothesized direction. Put differently, we find a significant difference in the predicted direction for roughly 6% of the 252 cases we tested. The strongest results in terms of significant differences are for the chemicals and plastics value chain (although only 4 out of 14 are in the predicted direction); almost no mean differences are significant for pharmaceuticals/medical technologies and aerospace. These broad results are our first hint that technology-based industries in Appalachia did not generally grow faster where they were located nearby or within identified technology clusters in the region.

Table 4
Differences of Means, 1998 to 2002

	Growth/Decline > 0 Workers, Chain *i*						Growth/Decline > 49 Workers, Chain *i*					
	Growing Counties		Declining Counties				Growing Counties		Declining Counties			
	N	*Mean*	*N*	*Mean*	Difference	*p-val*	*N*	*Mean*	*N*	*Mean*	Difference	*p-val*
Chemicals and plastics												
Location quotient 1998	97	0.93	135	4.79	–3.86	**.000**	19	1.81	54	9.22	–7.41	**.016**
Employment, 1998	97	138.00	135	535.00	–397.00	**.001**	19	343.00	54	1136.00	–792.00	**.049**
G, employment, 1998	97	0.44	135	0.94	–0.5	.126	19	0.91	54	1.54	–0.63	.389
G, employment change, 1989–1998	97	0.27	135	0.06	0.21	.295	19	0.42	54	–0.04	0.46	.258
G, patents, 1990–1998	97	–0.15	135	–0.08	–0.07	.287	19	–0.20	54	0.02	–0.22	.109
Distance to university ranked in related fields	97	153.00	135	156.00	–3.00	.783	19	160.00	54	168.00	–8.00	.721
Information technology and electronics												
Location quotient 1998	90	0.87	155	1.63	–0.76	**.009**	31	1.23	81	2.39	–1.16	**.043**
Employment, 1998	90	251.00	155	732.00	–481.00	**.008**	31	534.00	81	1280.00	–746.00	**.061**
G, employment, 1998	90	–0.44	155	–0.34	–0.10	.409	31	–0.58	81	–0.30	–0.28	.254
G, employment change, 1989–1998	90	–0.04	155	–0.07	0.03	.805	31	0.32	81	–0.08	0.40	**.050**
G, patents, 1990–1998	90	–0.14	155	–0.08	–0.07	.354	31	–0.11	81	0.03	–0.13	.334
Distance to university ranked in related fields	90	112.00	155	105.00	7.00	.359	31	101.00	81	97.00	3.00	.799
Industrial machinery												
Location quotient 1998	86	1.19	143	3.52	–2.33	**.012**	24	2.50	59	6.97	–4.46	**.076**
Employment, 1998	86	201.00	143	334.00	–133.00	**.049**	24	548.00	59	688.00	–140.00	.390
G, employment, 1998	86	–0.02	143	0.56	–0.59	**.017**	24	0.75	59	1.03	–0.28	.527
G, employment change, 1989–1998	86	–0.82	143	–0.52	–0.3	.414	24	–1.12	59	–0.76	–0.36	.685
G, patents, 1990–1998	86	0.03	143	0.07	–0.04	.768	24	0.42	59	0.30	0.11	.703
Distance to university ranked in related fields	86	92.00	143	92.00	0.00	.99	24	87.00	59	78.00	9.00	.433
Motor vehicles												
Location quotient 1998	91	2.25	98	4.31	–2.07	**.054**	40	3.57	47	7.39	–3.82	**.071**
Employment, 1998	91	330.00	98	502.00	–172.00	.123	40	648.00	47	936.00	–288.00	.175
G, employment, 1998	91	0.06	98	0.07	–0.01	.889	40	0.22	47	0.18	0.05	.774
G, employment change, 1989–1998	91	0.20	98	0.25	–0.05	.82	40	0.42	47	0.17	0.25	.527
G, patents, 1990–1998	91	0.09	98	–0.09	0.18	**.037**	40	0.25	47	–0.04	0.3	**.063**
Distance to university ranked in related fields	91	94.00	98	97.00	–3.00	.7.00	40	101.00	47	97.00	4.00	.755
Aerospace												
Location quotient 1998	34	3.05	40	2.24	0.81	.645	12	2.38	11	7.19	–4.82	.270
Employment, 1998	34	259.00	40	296.00	–38.00	.753	12	504.00	11	875.00	–371.00	.244
G, employment, 1998	34	–0.38	40	–0.34	–0.04	.865	12	–0.58	11	–0.39	–0.20	.719
G, employment change, 1989–1998	34	0.10	40	–0.34	0.45	**.074**	12	–0.07	11	–0.96	0.89	.204
G, patents, 1990–1998	34	0.11	40	0.42	–0.31	.337	12	–0.01	11	0.73	–0.74	.315
Distance to university ranked in related fields	34	139.00	40	125.00	14.00	.452	12	111.00	11	116.00	–5.00	.879
Communications services and software												
Location quotient 1998	258	0.37	132	0.49	–0.11	**.076**	92	0.60	26	1.17	–0.57	**.001**
Employment, 1998	258	525.00	132	325.00	200.00	.382	92	1336.00	26	1199.00	137.00	.870
G, employment, 1998	258	–0.40	132	–0.21	–0.19	**.070**	92	–0.22	26	–0.07	–0.14	.646
G, employment change, 1989–1998	258	–0.02	132	–0.07	0.05	.556	92	0.27	26	0.30	–0.03	.927
G, patents, 1990–1998	258	–0.11	132	–0.14	0.02	.598	92	–0.07	26	–0.19	0.12	.356
Distance to university ranked in related fields	258	113.00	132	125.00	–12.00	**.070**	92	97.00	26	111.00	–14.00	.278

(continued)

Table 4 (continued)

	Growth/Decline > 0 Workers, Chain *i*						Growth/Decline > 49 Workers, Chain *i*					
	Growing Counties		Declining Counties				Growing Counties		Declining Counties			
	N	*Mean*	*N*	*Mean*	Difference	*p-val*	*N*	*Mean*	*N*	*Mean*	Difference	*p-val*
Pharmaceuticals and medical technologies												
Location quotient 1998	160	0.65	154	0.79	–0.15	.487	38	1.58	26	3.13	–1.55	**.077**
Employment, 1998	160	204.00	154	180.00	24.00	.744	38	744.00	26	879.00	–136.00	.677
G, employment, 1998	160	–0.20	154	–0.35	0.15	.165	38	–0.13	26	–0.59	0.46	.142
G, employment change, 1989–1998	160	–0.11	154	–0.16	**0.05**	.550	38	–0.24	26	–0.15	–0.09	.720
G, patents, 1990–1998	160	–0.16	154	–0.17	**0.01**	.798	38	–0.16	26	–0.20	0.04	.648
Distance to university	160	144.00	154	140.00	4.00	.602	38	118.00	26	130.00	–12.00	.493

Note: Boldface type indicates $p \leq .10$.

Consider the results for chemicals and plastics industries more closely. Employment in chemicals and plastics expanded in 97 counties and declined in 135 counties in the region between 1998 and 2002 (see Table 4). The mean 1998 chemicals and plastics value chain location quotient for the set of 97 growth counties was 0.93 compared to 4.79 for the set of declining counties. Moreover, 1998 employment in the value chain in the 97 counties posting growth in chemicals and plastics averaged 138, compared to 535 for the set of declining counties. Clearly, chemicals and plastics industries in counties with *less* chemicals/plastics activity in 1998 generally fared *better* during the period from 1998 to 2002 in terms of employment growth. Nineteen counties added at least 50 net jobs in the chemicals and plastics whereas 54 counties lost at least 50 jobs, but the mean differences are similar in direction: The employment gains tended to occur in noncluster counties. The broader spatial cluster measures, including the three *G* statistics and distance to ranked universities, yielded no significant mean differences for the value chain. Across all 7 chains, of the 23 significant mean differences during the period from 1998 to 2002 reported in Table 4, 18 suggest that industries fared better in *noncluster* locations.

Overall, for this period and region, there would appear to be little evidence to support the notion that establishments in technology-oriented regional clusters outperformed establishments in noncluster locations, at least in terms of net new job creation. Indeed, if anything, the results indicate the opposite. Of course, employment is only one—arguably imperfect—measure of performance. It is possible that output and/or productivity grew substantially in clustered technology industries in the region as employment declined. Although we do not have the output data necessary to examine that possibility, we are able to investigate another possible outcome measure: new business formation.

Start-Up Activity in Appalachia

Entrepreneurship is a central facet of contemporary theories of regional growth and development. New businesses are an important source of job creation. Roughly 26% of the jobs added to the U.S. economy between 1991 and 1996 came from establishment births, compared with approximately 17% attributed to the expansion of existing firms (Acs & Armington, 2004). Porter (1998) identifies new business creation as one of the three primary benefits to industrial clustering. His claims are supported by consistent evidence of a positive association between the location of new and existing plants (Armington & Acs, 2002; Gabe & Kraybill, 2002; Rosenthal & Strange, 2003). Many new firms are spin-offs from existing businesses. Spin-offs may be direct, as when large firms downsize by out-sourcing formerly internal functions (Harrison, 1994), or they may be informal, such as when entrepreneurs start their own businesses based on experience gained in previous employment. In either case, the new business is likely to be in a similar industry as the parent and, because of "home-bias" preferences, it is also likely to start up in the same region (Figueiredo, Guimarães, & Woodward, 2002; Meester, 2004).

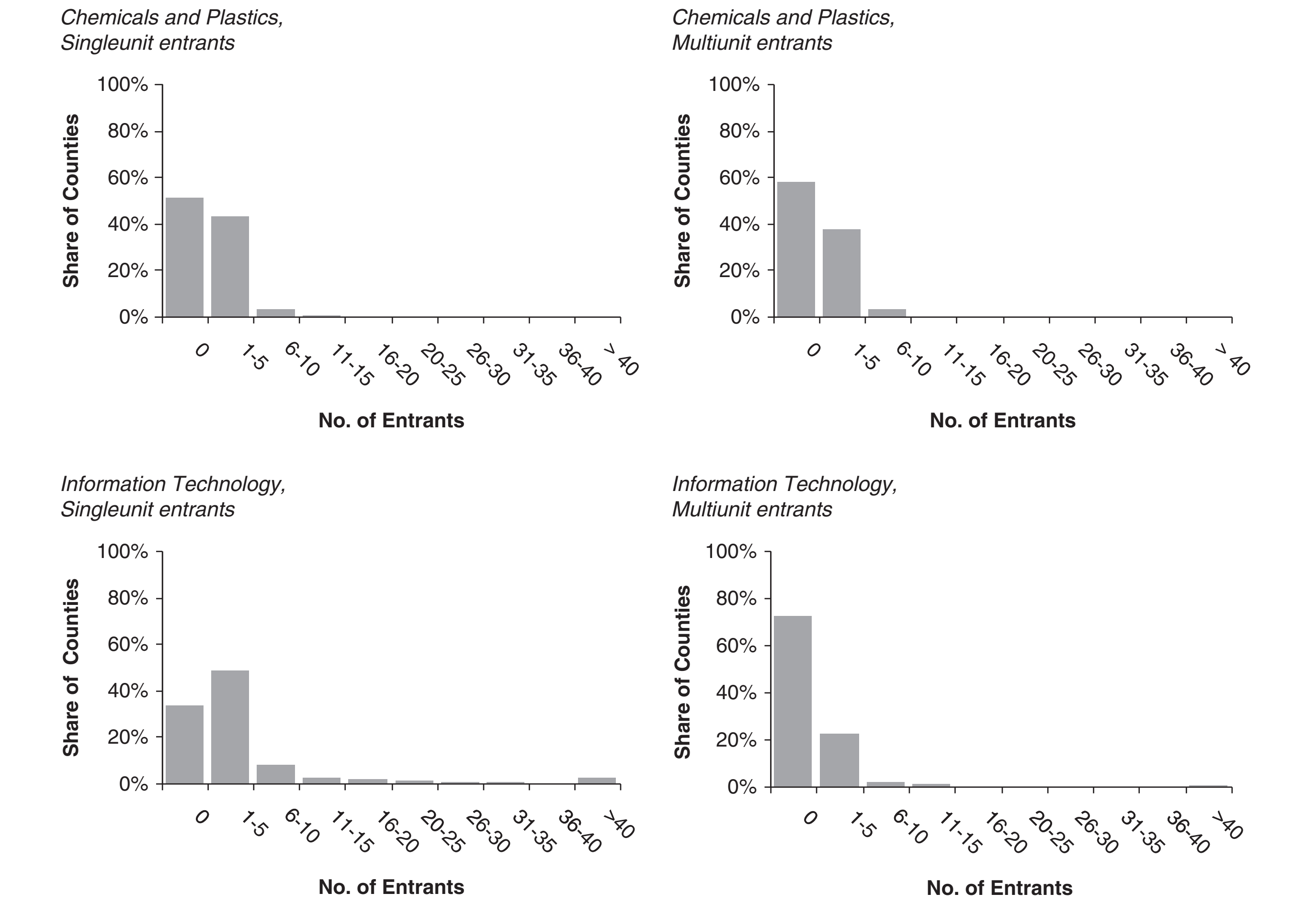
It is less clear whether entry rates are proportionate to the size of existing industry in a place or if external

Table 5
Summary of Difference of Means Tests, 1998 to 2002

	Growth/Decline > 0 Workers, Chain *i*			Growth/Decline > 49 Workers, Chain *i*		
NAICS industries in technology value chain	1998–2002	1998–2001	2001–2002	1998–2002	1998–2001	2001–2002
Chemicals and plastics						
Location quotient, 1998 employment	–	–	–	–		–
1998 employment	–	–		–	–	
G statistic, 1998 employment						
G statistics, 1989–1998 employment change		+				+
G statistic, patents granted over period 1990–1999			–			
Distance to university ranked in related fields			–			–
Information technology and electronics						
Location quotient, 1998 employment	–		–	–		
1998 employment	–		–	–		–
G statistic, 1998 employment						
G statistics, 1989–1998 employment change				+		
G statistic, patents granted over period 1990–1999						
Distance to university ranked in related fields						
Industrial machinery						
Location quotient, 1998 employment	–			–		
1998 employment	–		–			–
G statistic, 1998 employment	–				–	
G statistics, 1989–1998 employment change						
G statistic, patents granted over period 1990–1999						
Distance to university ranked in related fields					+	
Motor vehicles						
Location quotient, 1998 employment	–	–		–		
1998 employment						
G statistic, 1998 employment						
G statistics, 1989–1998 employment change						
G statistic, patents granted over period 1990–1999	+	+		+	+	
Distance to university ranked in related fields						
Aerospace						
Location quotient, 1998 employment						
1998 employment						
G statistic, 1998 employment						
G statistics, 1989–1998 employment change	+	+			+	
G statistic, patents granted over period 1990–1999						
Distance to university ranked in related fields						
Communication services and software						
Location quotient, 1998 employment	–			–	–	
1998 employment						
G statistic, 1998 employment	–					
G statistics, 1989–1998 employment change						
G statistic, patents granted over period 1990–1999						
Distance to university ranked in related fields	–					
Pharmaceuticals and medical technologies						
Location quotient, 1998 employment				–		
1998 employment						
G statistic, 1998 employment					+	
G statistics, 1989–1998 employment change						
G statistic, patents granted over period 1990–1999						
Distance to university ranked in related fields						

Note: NAICS = North American Industrial Classification System; positive (+) and negative (-) differences indicate $p \leq .10$.

Figure 3
Distributions of Single- and Multiunit Business Entrants in Appalachia



economies and other benefits that underlie localization actually induce the creation of new business beyond the level expected from existing industry. Areas in which industry is concentrated are also likely to be places where external economies and business networks are particularly well developed. Industrial concentration may provide a favorable entrepreneurial climate that induces increasing returns in new business formation. Dumais, Ellison, and Glaeser (2002) provide some contrary evidence, finding that the employment from births has had a slight deconcentrating effect on regional industrial specialization over time. Modeling the industry base with a measure of relative specialization, as opposed to levels, can help isolate the influence of external economies from proportionate spin-offs.

We investigate these issues in Appalachia using counts of county entrants by industry from a special tabulation of the U.S. *County Business Patterns.* We began by reclassifying single and multiunit entrants by three-digit SIC industry into the eight technology areas using the SIC concordance from the 2002 study. We then excluded the aerospace and appliances value chains from further analysis because the numbers of entrants in those chains were too low during the period to produce substantial regional variation. As shown in Figure 3 and Table 6, the total number of entrants and their distribution across counties varies substantially depending on the value chain in question. By far, the largest number of entrants is in the communication services/software and IT value chains, coinciding with 1990's boom years for those industries.

We analyzed single-unit and multiunit entrants separately. New single (i.e., independent) plants may react differently to their regional environment than do multiunit

Table 6
Descriptive Statistics: Establishment Entry Levels and Rates by County in Appalachia

	Chemicals and Plastics		Information Technology and Instruments		Industrial Machinery		Motor Vehicles		Communications Services and Software		Pharmaceuticals and Medical Technologies	
	Single	Multi	Single	Multi	Single	Multi	Single	Multi	Single	Multi	Single	Multi
Total entrants	644	474	2677	475	280	41	327	59	3935	669	295	103
Entry per county												
M	1.59	1.17	6.59	1.17	0.69	0.10	0.81	0.15	9.69	1.65	0.73	0.25
SD	3.61	2.69	28.73	5.35	1.72	0.39	1.83	0.45	38.19	7.55	3.12	1.25
Maximum	38.00	32.00	463.00	66.00	18.00	4.00	17.00	4.00	567.00	91.00	46.00	19.00
Moran (first order)	0.11	0.13	0.09	0.05	0.28	0.05	0.23	0.09	0.09	0.03	0.02	0.01
County entry rate (per 1,000 persons)												
M	0.02	0.02	0.06	0.01	0.01	0.00	0.01	0.00	0.10	0.01	0.01	0.00
SD	0.03	0.03	0.08	0.02	0.02	0.01	0.02	0.01	0.11	0.03	0.02	0.01
Maximum	0.24	0.19	0.86	0.18	0.2	0.08	0.19	0.09	1.05	0.27	0.12	0.13
Moran (first order)	0.08	0.14	0.3	0.07	0.08	0.03	0.05	0.04	0.27	0.06	–0.03	0.01

establishments. Consequently, the two are best modeled as samples drawn from separate populations. In brief, the process governing the spatial distribution of multiunit establishments can be viewed as a location or expansion decision situated within the overall organization strategy of the parent firm (Caves & Porter, 1977). Multiunit plants are typically larger and have access to the financial and nonmonetary resources of the parent firm, making them less dependent on local sources of knowledge and capital. By contrast, the formation of an independent establishment is typically based on the decision of a single person or small number of people. Such businesses are usually smaller, are often financed through personal savings or equity capital, and have limited access to external information networks, making them more dependent on locational assets than are their multiunit counterparts.

Our dependent variable is the number of entrants per county measured during the combined years of 1998, 1999, and 2000. Guimarães, Fiqueiredo, and Woodward (2003) have proven a statistical equivalence between Poisson count processes and the conditional logit specification preferred for modeling firm location choices. Several of the value chains have an exceptionally large number of zero-entrant counties, particularly for multiunit entrants, and Poisson count models are not well suited to such data. As a consequence, we use a zero-inflated negative binomial (ZINB) model to measure the influence of clustering on entry. The ZINB model accounts for excess zeros through a two-stage process, similar to Heckman's (1979) sample selection model. The first stage distinguishes counties with no realistic potential for entry in a value chain from those where positive entry is possible even if a zero count is observed. The probability of selection into the no-entry potential group is captured by the estimated parameter ψ_i, determined by a binary logit model:

$$\psi_i = F\,(z_i\gamma) \tag{3}$$

where F is the logistic cumulative density function. For simplicity, we assume that z_i comprises each county's population in 1998 and a constant. Explaining potential entry as a function of population corresponds to the "labor market" view of entry.

The second stage adjusts the count model by the probability that a county might fall into the no-entry potential group. The probability function is:

$$\begin{aligned} &\Pr(y_i = 0 \mid x_i) = \psi_i + (1-\psi_i)\left(\frac{\theta}{\theta+\mu_i}\right) \\ &\Pr(y_i = x_i) = (1-\psi_i)\frac{\Gamma(\theta+y_i)}{\Gamma(\theta)\Gamma(y_i+1)}\left(\frac{\theta}{\theta+\mu_i}\right)^{\theta}\left(\frac{\theta}{\theta+\mu_i}\right)^{y_i} \text{ for } y = 1, 2, \ldots \end{aligned} \tag{4}$$

where $\mu_i = \exp(x_i\beta)$ and Γ is a gamma function with parameter θ. Measures of clustering are those analyzed in the previous section, with one addition: We also test a technology cluster dummy variable that takes a value of 1 for a given county and value chain *i* if the following are true: Either the county's 1998 employment G_i or employment change G_i is statistically significant, or the county's employment location quotient > 1, *and* either the county's

Table 7
Zero-Inflated Negative Binomial Model: Chemicals and Plastics Value Chain

	Single-Unit Establishments						Multiunit Establishments					
Number of observations		406			406			406			406	
Nonzero observations		198			198			170			170	
Likelihood ratio (LR) χ^2		177.8			190			112.8			114.3	
Probability > χ^2		0			0			0			0	
Log likelihood		–520.1415			–514.0539			–474.2			–473.419	
	Model 1			Model 2			Model 1			Model 2		
	B	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|
Negative binomial (NegBin)												
Constant	0.42	1.53	0.011	0.32	1.38	0.054	–0.06	0.19	0.728	–0.09	0.91	0.276
Employment (G_i)	0.05	1.05	0.142	0.01	1.01	0.697	0.08	0.04	0.039	0.07	1.07	0.142
Employment growth (G_i)	–0.01	0.99	0.807	0.02	1.02	0.635	–0.01	0.05	0.922	0.00	1.00	0.105
Patents (G_i)	0.02	1.03	0.869	–0.02	0.98	0.900	–0.08	0.18	0.670	–0.11	0.89	0.240
Population (thousands)	0.01	1.01	0.000	0.01	1.01	0.000	0.00	0.00	0.000	0.00	1.00	0.000
University (distance)	0.00	1.00	0.029	0.00	1.00	0.098	0.00	0.00	0.534	0.00	1.00	0.003
Technology cluster dummy				0.74		0.000				0.32		0.829
Logit												
Constant	2.25	9.44	0.000	2.21	9.12	0.000	1.99	0.37	0.000	1.97	7.20	2.694
Population (thousands)	–0.09	0.92	0.000	–0.08	0.92	0.000	–0.07	0.01	0.000	–0.06	0.94	0.000
LR test (overdispersion)		133.10	0.000		108.60	0.000		47.50	0.000		41.90	0.000
Vuong test (versus NegBin)		3.96	0.000		4.16	0.000		3.17	0.001		3.21	0.001

patent G_i is statistically significant, a university ranked in a relevant field is within or adjacent to the county, or the patent location quotient > 1. The technology cluster dummy is a categorical indicator of whether a county is home to both industry and knowledge-based activity in a given value chain.

The results of our modeling are presented in tables 7 to 12. The tables report separate estimates for single and multiunit samples, as well as models exclusive and inclusive of the technology cluster dummy variable (models 1 and 2, respectively). Each table also includes estimates for both the first stage binary logit and the second stage negative binomial count model. Under the assumption that external economies in the industry or knowledge base favor entry, the parameter estimates for the employment concentration measure, employment growth concentration measure, and patent concentration measure are all expected to be positive in the count model. The distance-to-universities variable is expected to be negative, assuming that proximity to knowledge infrastructure favors entry. The logit model estimates the likelihood of being in the zero-outcome group. Assuming that population size favors entry, we expect this variable to be negative in the first stage model and positive in the count model.

Exponentiated coefficients from both models are readily interpreted as odds ratios. Because the *G* variables (employment, employment growth, and patents) follow an approximate normal distribution, one might interpret their coefficients in terms of changes in standard deviation. We include the results of the likelihood ratio test for overdispersion and the Vuong statistic as a test of whether the zero-inflated model is significantly different from the standard negative binomial model. The Vuong test is distributed approximately normal with large positive values favoring the zero-inflated model and large negative values favoring the standard negative binomial. In nearly all cases, significant likelihood ratio tests suggest overdispersion, confirming our use of the negative binomial specification over the Poisson. Positive and significant Vuong statistics also suggest a superior fit for the zero-inflated specification.

Population is the only universally significant predictor of entry across all model specifications. Population size significantly explains the difference between counties with entry potential and those without entry potential and in the count model explains why some counties have more entrants. That is not surprising, given that counties with a larger number of residents have a larger pool of candidate entrepreneurs. Larger counties also have larger labor pools and greater access to local markets, which may help explain the attraction for new branch and subsidiary plants. Furthermore, some of the smallest

Table 8
Zero-Inflated Negative Binomial Model: Information Technology and Electronics Value Chain

	Single-Unit Establishments						Multiunit Establishments					
Number of observations	406			406			406			406		
Nonzero observations	270			270			110			110		
Likelihood ratio (LR) c^2	370.9			374.2			138.2			141.1		
Probability> c^2	0			0			0			0		
Log likelihood	–813.2954			–811.6542			–350.3331			–348.8893		
	Model 1			Model 2			Model 1			Model 2		
	b	Exp *b*	*P* > |*z*|	*B*	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|
Negative binomial (NegBin)												
Constant	1.03	2.79	0.000	0.56	1.75	0.076	–0.80	0.45	0.007	–0.78	0.46	–0.209
Employment (G_i)	–0.11	0.90	0.117	–0.15	0.86	0.035	0.19	1.21	0.056	0.09	1.10	0.311
Employment growth (G_i)	0.22	1.24	0.001	0.22	1.25	0.000	0.19	1.21	0.122	0.14	1.16	0.383
Patents (G_i)	0.27	1.31	0.026	0.23	1.26	0.058	–0.23	0.79	0.310	–0.38	0.68	0.096
Population (thousands)	0.01	1.01	0.000	0.01	1.01	0.000	0.01	1.01	0.000	0.01	1.01	0.000
University (distance)	0.00	1.00	0.001	0.00	1.00	0.001	0.00	1.00	0.465	0.00	1.00	0.005
Technology cluster dummy				1.01		0.000				0.96		2.062
Logit												
Constant	3.14	23.16	0.000	3.13	22.94	0.000	2.25	9.48	0.000	2.27	9.72	3.196
Population (thousands)	–0.19	0.83	0.000	–0.18	0.83	0.000	–0.06	0.94	0.000	–0.06	0.95	0.000
LR test (overdispersion)		1707.00	0.000		1224.72	0.000		160.20	0.000		86.07	0.000
Vuong test (versus NegBin)		4.59	0.004		4.62	0.000		2.42	0.008		2.43	0.008

Table 9
Zero-inflated Negative Binomial Model: Industrial Machinery Value Chain

	Single-Unit Establishments						Multiunit Establishments					
Number of observations	406			406			406			406		
Nonzero observations	130			130			33			33		
Likelihood ratio c^2	77.4			78.3			24.0			24.2		
Probability > c^2	0.0			0.0			0.0			0.0		
Log likelihood	–368.7			–368.2			–108.2			–108.2		
	Model 1			Model 2			Model 1			Model 2		
	b	Exp *b*	P > |*z*|	*b*	Exp *b*	P > |*z*|	*b*	Exp *b*	P > |*z*|	*b*	Exp *b*	P > |*z*|
Negative binomial												
Constant	–0.71	0.49	0.005	0.32	1.38	0.316	–2.30	0.10	0.000	–2.27	0.10	–1.389
Employment G_i	0.14	1.16	0.018	0.14	1.15	0.026	0.05	1.05	0.711	0.05	1.06	0.323
Employment Growth G_i	–0.06	0.94	0.089	–0.06	0.94	0.083	–0.01	0.99	0.931	–0.01	0.99	0.173
Patents G_i	–0.06	0.94	0.680	–0.13	0.88	0.427	–0.42	0.66	0.080	–0.39	0.68	0.118
Population (000's)	0.01	1.01	0.000	0.00	1.00	0.000	0.00	1.00	0.000	0.00	1.00	0.000
University (distance)	0.00	1.00	0.652	0.00	1.00	0.694	0.00	1.00	0.638	0.00	1.00	0.009
Technology cluster dummy				–0.69		0.006				–0.24		1.108
Logit												
Constant	2.27	9.66	0.000	2.25	9.49	0.000	3.72	41.18	0.004	3.73	41.68	6.210
Population (000's)	–0.09	0.91	0.006	–0.09	0.92	0.006	–0.10	0.90	0.028	–0.10	0.90	0.000
Likelihood ratio test (overdispersion)		52.17	0.000		39.53	0.000		0.00	1.000		0.00	1.000
Vuong test (vs NegBin)		2.72	0.003		2.65	0.004			.	.		

Table 10
Zero-Inflated Negative Binomial Model: Motor Vehicles Value Chain

	Single-Unit Establishments		Multiunit Establishments
Number of observations	406	406	406
Nonzero observations	149	149	47
Likelihood ratio (LR) c^2	89.2	89.3	21.2
Probability > c^2	0	0	0
Log likelihood	–399.3279	–399.2955	–146.9232

	Model 1			Model 2			Model 1			Model 2		
	b	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|
Negative binomial (NegBin)												
Constant	–0.16	0.85	0.437	–0.16	0.85	0.430	–2.12	0.12	0.000			
Employment (G_i)	0.28	1.33	0.007	0.28	1.33	0.007	–0.06	0.94	0.795			
Employment growth (G_i)	0.01	1.01	0.886	0.01	1.01	0.917	–0.09	0.92	0.429			
Patents (G_i)	0.22	1.24	0.082	0.21	1.23	0.115	–0.52	0.60	0.008			
Population (thousands)	0.01	1.01	0.000	0.00	1.00	0.000	0.00	1.00	0.000			
University (distance)	0.00	1.00	0.158	0.00	1.00	0.171	0.01	1.01	0.028			
Technology cluster dummy				0.08		0.799						
Logit												
Constant	2.09	8.09	0.000	2.09	8.12	0.000	2.89	18.07	0.000			
Population (thousands)	–0.08	0.92	0.001	0.00	1.00	0.000	–0.06	0.94	0.003			
LR test of alpha = 0: Probability > c^2		50.67	0.000		46.26	0.000		0.00	1.000			
Vuong test: Probability > z		2.69	0.004		2.60	0.005		0.46	0.332			

Note: Model 2 for multiunit establishments failed to converge.

Table 11
Zero-Inflated Negative Binomial Model: Communications Services and Software Value Chain

	Single-Unit Establishments		Multiunit Establishments	
Number of observations	406	406	406	406
Nonzero observations	319	319	129	129
Likelihood ratio (LR) c^2	473.5	484.8	156.5	156.5
Probability> c^2	0	0	0	0
Log likelihood	–975.746	–970.1018	–413.9784	–413.9717

	Model 1			Model 2			Model 1			Model 2		
	b	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|
Negative binomial (NegBin)												
Constant	1.19	3.30	0.000	1.11	3.05	0.000	–0.73	0.48	0.004	–0.74	0.48	–0.240
Employment (G_i)	–0.06	0.95	0.338	–0.10	0.91	0.082	0.24	1.28	0.015	0.24	1.27	0.458
Employment Growth (G_i)	0.30	1.35	0.000	0.25	1.28	0.001	–0.19	0.83	0.138	–0.19	0.83	0.060
Patents (G_i)	–0.08	0.93	0.463	–0.11	0.90	0.267	–0.07	0.94	0.738	–0.07	0.93	0.317
Population (thousands)	0.01	1.01	0.000	0.01	1.01	0.000	0.01	1.01	0.000	0.01	1.01	0.000
University (distance)	0.00	1.00	0.000	0.00	1.00	0.004	0.00	1.00	0.662	0.00	1.00	0.004
Technology cluster dummy				0.84		0.001				0.06		1.101
Logit												
Constant	1.47	4.35	0.008	1.48	4.38	0.006	2.82	16.70	0.000	2.82	16.74	4.098
Population (thousands)	–0.16	0.85	0.000	–0.15	0.86	0.000	–0.10	0.90	0.000	–0.10	0.90	0.000
LR test (overdispersion)		2290.00	0.000		2301.00	0.000		366.10	0.000		354.64	0.000
Vuong test (versus NegBin)		2.94	0.002		3.13	0.001		2.72	0.003		2.72	0.003

Table 12
Zero-Inflated Negative Binomial Model: Pharmaceuticals and Medical Technologies Value Chain

	Single-Unit Establishments						Multiunit Establishments					
Number of observations	406			406			406			406		
Nonzero observations	103			103			43			43		
Likelihood ratio (LR) c^2	113.6			114			32.1			32.1		
Probability> c^2	0			0			0			0		
Log likelihood	–301.2322			–301.0581			–155.7452			–155.7281		
	Model 1			Model 2			Model 1			Model 2		
	b	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|	*b*	Exp *b*	*P* > |*z*|
Negative binomial (NegBin)												
Constant	–0.469	0.63	0.119	–0.370	0.69	0.281	–1.019	0.36	0.067	–0.965	0.38	0.267
Employment (G_i)	0.010	1.01	0.902	0.010	1.01	0.905	0.297	1.35	0.040	0.301	1.35	0.588
Employment growth (G_i)	0.123	1.13	0.292	0.131	1.14	0.259	0.114	1.12	0.561	0.109	1.12	0.496
Patents (G_i)	0.940	2.56	0.027	1.044	2.84	0.022	1.144	3.14	0.123	1.177	3.24	2.659
Population (thousands)	0.007	1.01	0.000	0.007	1.01	0.000	0.007	1.01	0.000	0.007	1.01	0.000
University (distance)	–0.001	1.00	0.659	–0.001	1.00	0.544	–0.001	1.00	0.630	–0.001	1.00	0.004
Technology cluster dummy				–0.380	0.68	0.557				–0.169	0.84	1.618
Logit												
Constant	2.507	12.27	0.000	2.535	12.61	0.000	2.978	19.65	0.000	2.993	19.95	4.406
Population (thousands)	–0.061	0.94	0.000	–0.060	0.94	0.000	–0.048	0.95	0.002	–0.048	0.95	0.000
LR test (overdispersion)		73.24	0.000		30.78	0.000		51.52	0.000		35.09	0.000
Vuong test (versus NegBin)		2.71	0.003		2.56	0.005		2.06	0.002		1.90	0.029

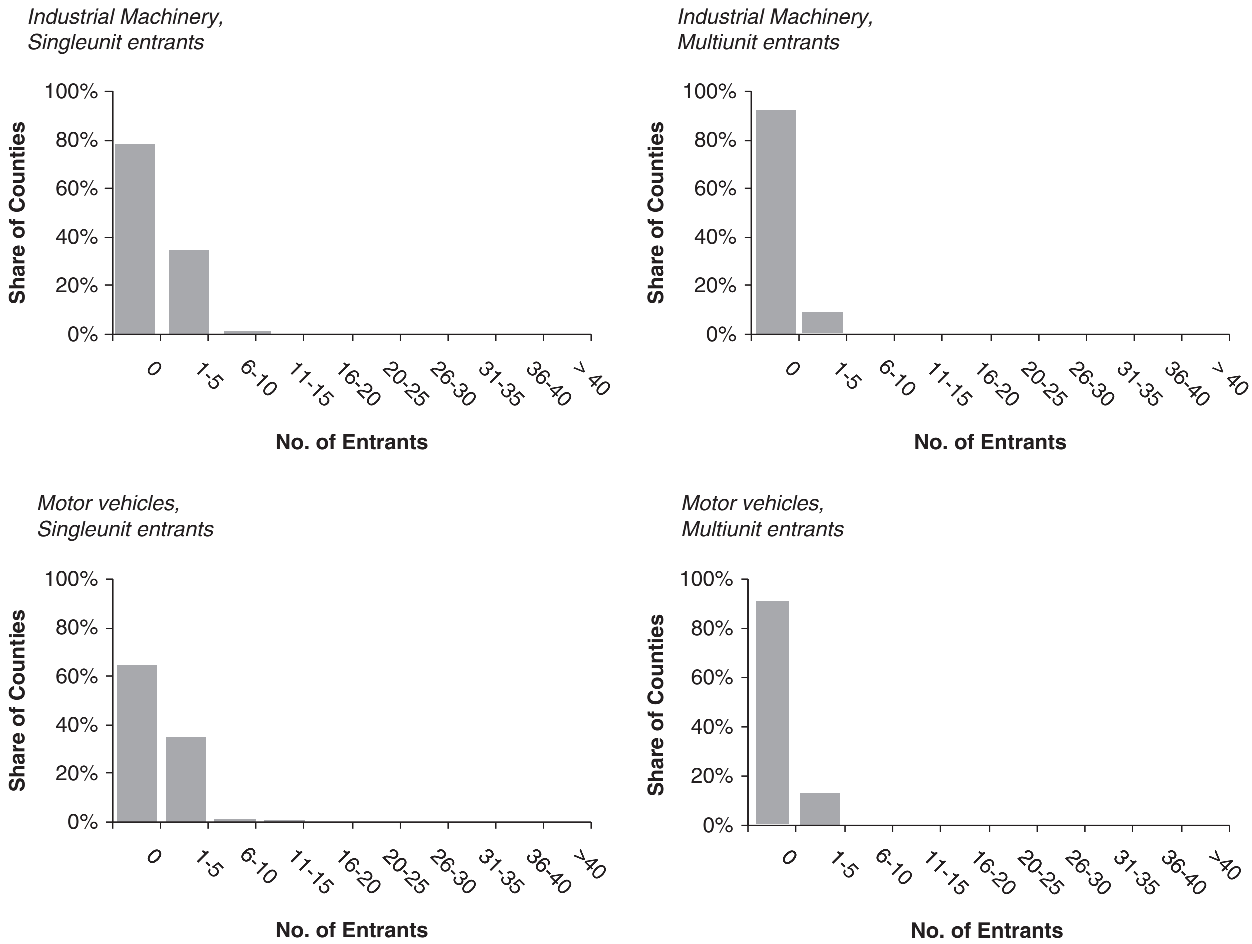
counties in Appalachia still lack some infrastructure or have topographical barriers that restrain development.

Consider first the models of single-unit entry that do not include the technology cluster dummy. Single-unit entrants generally favor proximity to the sources of knowledge. Distance from a university in relevant disciplines has a negative and significant effect on entry in the chemicals and plastics, IT, and communications services/software value chains. A concentration of industrial patenting is associated with higher entry in IT, pharmaceuticals, and at a lesser level of statistical significance, motor vehicles. The greatest influence is for pharmaceuticals and medical technologies, where a standard deviation increase in the relative concentration of patents increases the odds of new firm entry by 153%.

Not surprisingly, new firms in emerging technologies such as IT and communications services and software seem to favor counties with emerging concentrations over larger, established ones. A standard deviation increase in a county's concentration of value-chain employment growth is associated with 25% (in IT) and 35% (in communications services and software) higher single-unit entry. New firms in more traditional manufacturing industries (i.e., industrial machinery and motor vehicles) are predominantly located where employment is already concentrated. In motor vehicles, for instance, a standard deviation increase in the employment G increases the odds of new firm entry by 33%. The effect is more modest in industrial machinery, where a standard deviation change raises the odds of entry by 16%.

The technology cluster dummy variable seeks to capture the spatial coincidence between evidence of clustering in a value chain and a concentration of knowledge infrastructure. The results with the dummy suggest that joint clustering does favor single-plant entry over and above the effects of the individual clustering measures in half of the cases. New single-plant entry is significantly higher in those counties where we see joint industry and knowledge infrastructure clustering for the chemicals and plastics, IT, and communications services and software value chains. The most dramatic impact is for the IT value chain, in which the odds of entry in joint cluster counties is nearly three times that of entry in other counties. Single-unit entry in industrial machinery is about 50% lower in joint cluster counties. There is no obvious explanation for this counterintuitive finding. In most cases, the addition of the joint cluster dummy variable

Figure 4



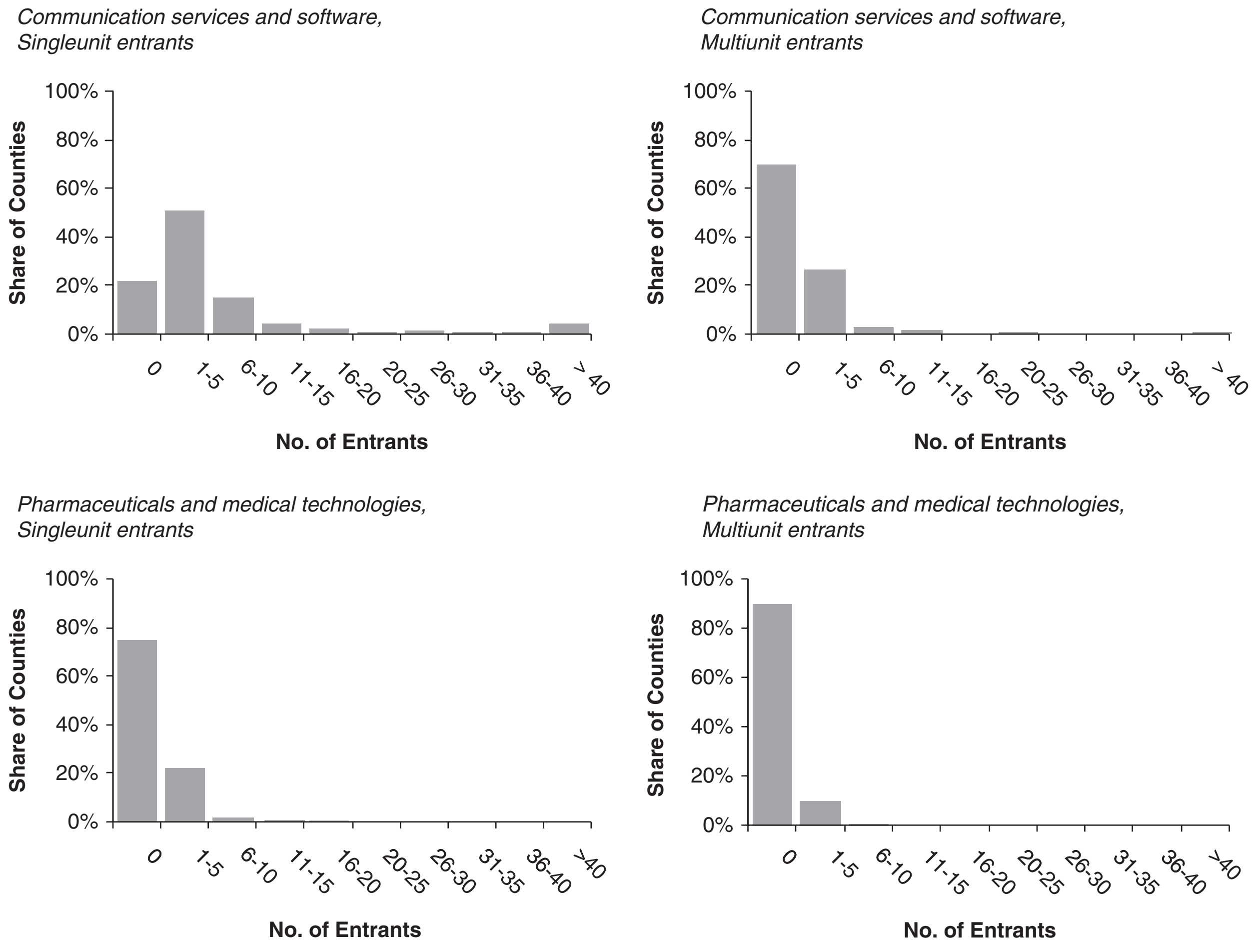
reduces the magnitude and significance of the other independent variables only by a relatively small amount.

There are clear differences between single-unit and multiunit establishments, even among those in the same value chain. In general, multiunit establishments are less influenced by proximity to universities or innovation activity. Indeed, proximity to universities or patenting hot spots typically has an adverse effect on multiunit entry, as seen for IT, industrial machinery, motor vehicles, and communications services and software. The sole exception is for multiunit pharmaceuticals entrants. They are more likely to locate near a top-tier university but only when we add the joint cluster dummy. The location decisions of multiunit pharmaceutical companies are likely driven by greater needs for highly educated scientists and engineers and the desire to engage in collaborative research and the need for clinical trials. Multiunit establishments in communications services and software and pharmaceuticals and IT show some preference for location in existing industry concentrations. This contrasts with the single-unit entrants in these sectors, for whom existing employment concentration did not matter. The joint cluster dummy variable does not have a significant impact on multiunit entry, although its introduction does eliminate the significance of the employment concentration variable in the three cases where it was previously significant.

Discussion

Our results suggest that the answer to the question of whether related industries generate more jobs where those industries are clustered is a conditional “no,” at least for this region of the United States and for the period from 1998 to 2002. We found little evidence that technology industries in spatial clusters in Appalachia created more jobs than the same industries in noncluster locations. For the most part, there were comparatively few differences in the level of clustering in growth and

Figure 5



nongrowth counties in the region. Where we did find significant differences, they more often suggested that technology industries in noncluster locations tended to fare better in terms of employment growth.

The relationship between new business formation and clustering is more complex. We did find some evidence that clustering is associated with higher levels of new business entry for some value chains and for both single-unit and multiunit establishments. The relationship is most pronounced for recent-vintage chains like IT and communications services and software. In addition, we find that the links between start-up activity and proximity to innovation activity (as measured by patents) and ranked universities tend to be stronger than the links between entry and existing value-chain employment. In one case—the industrial machinery value chain—entry rates tended to be lower in cluster versus noncluster locations. Overall, clustering does not guarantee employment growth in Appalachia, but it does appear to be associated with higher rates of new business formation in some of the more technology-intensive of the value chains (IT and communications services and software).

Our findings imply that economic development policy cannot be premised simply on the notion that spatial industry clusters observed in a given point in time are likely to be the sites of substantial subsequent employment growth. Put differently, clusters may represent useful targets for economic development but not because they are assured to produce strong job gains. The appropriate economic development strategies for an identified cluster in a given region may actually be job retention or workforce development programs aimed at helping redundant workers in the cluster obtain the necessary training and skills to assume employment in newer, emerging industries. We find that in Appalachia, most job growth occurred—albeit over a relatively slow-growth period regionally and nationally—among companies that were not substantially clustered with other businesses in their own value chain. At the same time, entrepreneurship policies may have more success if focused where target industries are already concentrated. It is important to realize, however, that we have not investigated the causal factors behind an association between agglomeration and new firm formation. Although it may be tempting to assume that spatial externalities are the explanation, careful longitudinal analysis by others (e.g., Buenstorf & Klepper, 2005) posits that such an observed relationship can also be explained by the normal process of new firm formation from existing firms in a region, not externalities per se. Clearly, much more work needs to be done on the link between business performance and clustering. Meanwhile, local economic development professionals should understand that clusters are not necessarily growth centers.

Notes

1. See Feser, Goldstein, Renski, and Renault (2002) for extensive documentation of all data sources and methods used in the original study.

2. A list of the North American Industrial Classification System industries associated with each value chain from the 2002 study is available from the authors on request.

References

Acs, Z. J., & Armington, C. (2004). Employment growth and entrepreneurial activity in cities. *Regional Studies, 38*(9), 911-927.

Armington, C., & Acs, Z. (2002). The determinants of regional variation in new firm formation. *Regional Studies, 36*(1), 33-45.

Buenstorf, G., & Klepper, S. (2005). *Heritage and agglomeration: The Akron tire cluster revisited* (Papers on Economics and Evolution, 2005-08). Jena, Germany: Max Planck Institute of Economics, Evolutionary Economics Group.

Caves, R. E., & Porter, M. E. (1977). From entry barriers to mobility barriers. *Quarterly Journal of Economics, 91,* 241-261.

Dumais, G., Ellison, G., & Glaeser, E. L. (2002). Geographic concentration as a dynamic process. *Review of Economics and Statistics, 84*(2), 193-204.

Feser, E., Goldstein, H., Renski, H., & Renault, C. (2002). *Regional technology assets and opportunities: The geographic clustering of high-tech industry, science and innovation in Appalachia.* Washington, DC: Appalachian Regional Commission.

Feser, E., & Koo, J. (2000). *High-tech clusters in North Carolina.* Raleigh: North Carolina Board of Science and Technology.

Feser, E. J., & Sweeney, S. H. (2000). A test for the coincident economic and spatial clustering of business enterprises. *Journal of Geographical Systems, 2,* 349-373.

Figueiredo, O., Guimarães, P., & Woodward, D. (2002). Home-field advantage: Location decisions of Portuguese entrepreneurs. *Journal of Urban Economics, 52,* 341-361.

Gabe, T. M., & Kraybill, D. S. (2002). The effect of state economic development incentives on employment growth of establishments. *Journal of Regional Science, 42,* 703-730.

Getis, A. (1984). Interaction modeling using second-order analysis. *Environment and Planning A, 16,* 173-183.

Guimarães, P., Figueiredo, O., & Woodward, D. (2003). A tractable approach to the firm location decision problem. *Review of Economics and Statistics, 85*(1), 201-204.

Harrison, B. (1994). *Lean and mean: The changing landscape of corporate power in the age of flexibility.* New York: Basic Books.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47,* 161-196.

Isserman, A., & Westervelt, J. (2006). 1.5 million missing numbers: Overcoming employment suppression in *County Business Patterns* data. *International Regional Science Review, 29,* 311-339.

Meester, W. (2004). *Locational preferences of entrepreneurs: Stated preference in the Netherlands and Germany.* Heidelberg and New York: Physica.

National Research Council (1995). *Research-doctorate programs in the United States: Continuity and change.* Washington, DC: Author, National Academy of Sciences.

Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis, 27*(4), 286-306.

Porter, M. E. (1998, November/December). Clusters and the new economics of competition. *Harvard Business Review,* 77-90.

Porter, M. E. (2003). The economic performance of regions. *Regional Studies, 37*(6&7), 549-578.

Rosenthal, S. S., & Strange, W. C. (2003). Geography, industrial organization, and agglomeration. *Review of Economics and Statistics, 85*(2), 377-393.

Sweeney, S. H., & Feser, E. J. (1998). Plant size and clustering of manufacturing activity. *Geographical Analysis, 30*(1), 45-64.

U. S. Patent & Trademark Office. (2000). *United States patent grants by state, county, and metropolitan area (utility patents 1990-1999).* Washington, DC: Author, Office for Patent and Trademark Information.

Acknowledgment

Special thanks to Andrew Isserman for sharing his suppression-adjusted County Business Patterns employment data for 1998 and 2002.

Edward Feser is a professor and interim head in the Department of Urban and Regional Planning at the University of Illinois at Urbana-Champaign.

Henry Renski is an assistant professor in the Department of Landscape Architecture and Regional Planning at the University of Massachusetts at Amherst.

Harvey Goldstein is a professor and director of the Public Governance and Management Program at MODUL University in Vienna, Austria.