Economic Development Quarterly
2018, Vol. 32(1) 44–59

DOI: 10.1177/0891242417752248
journals.sagepub.com/home/edq
SAGE

Industry Clusters and Regional Economic Performance: A Study Across U.S. Metropolitan Statistical Areas

Timothy F. Slaper[1], Karter M. Harmon[1], and Barry M. Rubin[1]

Abstract

Cluster theory provides a framework for analyzing regional industry dynamics. Definitions and categorizations of clusters vary, however, hindering the development of econometric models for cluster analysis. The authors propose empirical measures relevant to researchers and practitioners for cluster strength/specialization, cluster diversity/diversification, and identifying regional growth clusters. The latter measure uses location quotients, shift-share analysis, and other criteria to identify robust clusters that are important for thriving regions. These measures are calculated for local and traded clusters using employment data for 366 U.S. metropolitan statistical areas. Additionally, the authors estimated the relationship of their cluster performance measures to four traditional measures of economic performance: growth in gross domestic product, productivity per employee, compensation per employee, and personal income. The authors find traded cluster strength is positively related to compensation per employee growth and positively related to productivity growth, the latter being consistent with expected Marshall–Arrow–Romer externalities. Traded growth clusters are positively related to gross domestic product growth.

Keywords

cluster growth, regional economic performance, regional analysis, econometrics, agglomeration economies

For over two decades, policy makers and economic development professionals have stressed the importance of encouraging and supporting industry clusters to promote job creation and regional competitiveness. Several researchers—Michael Porter and Christian Ketels, among others—have developed the study of cluster-based economic development and touted the employment and competitive benefits of cluster-based development strategies (Ketels, 2013; Ketels & Memedovic, 2008; Porter, 2008). Several studies have quantified the benefits of clusters in terms of employment growth and competitiveness as it relates to patent rates, with Delgado, Porter, and Stern (2014) being the most comprehensive. Yet other economic benefits along the dimensions of productivity, wage, and income growth have been largely overlooked. In this study, we assess the benefits of clusters across several traditional measures of economic performance.

This study is organized as follows: We first discuss key concepts related to clusters in a brief literature review, including a discussion on the role of related and unrelated variety as they are connected to cluster definitions. Second, we present the data, analytical methods, and the rationale and construction for three measures of regional cluster characteristics, namely strength, diversity, and growth. These three measures are applied to both traded and local/nontraded clusters. The third section reports the empirical results, with the fourth section discussing the findings and conclusions about the usefulness of our measures.

Review of the Literature

Industry clusters are agglomerations of closely related industries (Delgado, Porter, & Stern, 2010). One might say that clusters are a network of economic relationships that create a competitive advantage for the related firms in a particular region. This advantage then becomes an enticement for similar industries to develop or relocate to a region.

Developing industry clusters has become a key goal for regional economic development, as clusters have been shown to strengthen competitiveness by increasing productivity, stimulating innovative new partnerships (even among competitors), and presenting opportunities for new businesses.

[1]Indiana University, Bloomington, IN, USA

Corresponding Author:
Karter M. Harmon, Indiana Business Research Center, Kelley School of Business, Indiana University, 1315 East 10th Street, Bloomington, IN 47405, USA.
Email: kartharm@indiana.edu

Michael Porter and others have identified the industries that tend to cluster together (Porter, 1998), and with colleagues Delgado and Stern, they have more recently created the U.S. Benchmark Cluster Definitions (Delgado, Porter, & Stern, 2016).[1] This serves as the analytical foundation for cluster-based economic development strategies that may target certain types of industries to locate in a region to strengthen a cluster or strategies that may target regional resources to help bolster a developing cluster.

Porter (2000) maintained that the colocation of companies, customers, suppliers, and other institutions create an environment of increased rivalry that leads to higher pressure to innovate. Colocation also facilitates building collaborative relationships and social networks, sharing information, and the diffusion of technology. Industry clusters are also important for innovation because they facilitate the growth of start-up firms, and they are considered key agents of innovation (Delgado et al., 2010). In comparison with more mature firms, start-ups may be more likely to identify new technologies and new market opportunities.

That said, clusters may decrease incentives for new business formation due to increased competition and crowding-out effects (or congestion costs) that result in diminishing marginal returns to new firms entering the region. On the other hand, some studies suggest clusters might lower the cost of starting a business by providing specialized suppliers, a local customer base, and producers of complementary products and services (Feldman, Francis, & Bercovitz, 2005; Glaeser & Kerr, 2009; Porter, 1998; Saxenian, 1994). Delgado et al. (2010) conducted an empirical analysis to inform this debate and concluded that, while there is evidence of negative crowding-out effects on entrepreneurship, the cluster environment that surrounds an industry expands the pool of available resources and reduces the barriers to starting a new business.

Because the forces of agglomeration as expressed in the formation of regional clusters is partially based on the diversification of industries, it may be helpful to contrast the Jacobian urbanization diversity benefits (Jacobs, 1969) of the unrelated variety of economic sectors with how cluster diversification is operationalized for our analysis. The findings of Delgado et al. (2014) pointed to cluster variety—multiple clusters in a region that are related—as having a positive effect on innovation as measured by patenting rates. Frenken, Van Oort, and Verburg (2007) hypothesized that Jacobian, or unrelated variety, externalities are positively related to employment, but they also discussed an additional dimension to unrelated variety that we note below. In our case, a diversity of clusters is akin to the diversification of stocks in a portfolio. Cluster diversity used here is not a measure of how, and in what ways, unrelated clusters are different from each other; rather, diversity is more synonymous with balance. Following Frenken et al. (2007), we use an entropy index to measure cluster diversity.

In contrast to diversification, strength is viewed as the relative concentration of a cluster, without regard to the balance or concentration of industries within that cluster, in the same way that industry strength, or specialization, is viewed as the relative concentration of an industry as measured by a location quotient (LQ). In other words, how concentrated is this cluster relative to other clusters in the region? This aligns with the notion of related variety discussed by Frenken et al. (2007), who categorized industries based on their technological and material requirements. In a similar way, Delgado et al. (2016) used industry input–output relationships to categorize industries into clusters. The agglomerative benefits of such related variety are often called "localization economies" (Frenken et al., 2007) and were first conceptualized by Marshall over a century ago (1890/1966), and since refined by Arrow (1962) and Romer (1986). The agglomeration byproduct of related variety is often referred to as Marshall–Arrow–Romer (MAR) externalities. MAR externalities are, within industries, usually broadly defined sectors. In our case, however, MAR externalities would be in evidence within more narrowly defined clusters of related industries.

Delgado et al. (2014) used the term "cluster specialization" rather than cluster strength (as used here) for their measure of the relative concentration of a cluster in a region. They note the correlation between industry concentration—abandoning both the terms specialization and strength for the moment to reduce confusion—and cluster concentration may be small. That is, although a single industry may be salient in a region, aggregating it into a cluster with other industries dilutes its concentration and the cluster as a whole may or may not be salient. Delgado et al. (2014) went on to state that industry versus cluster concentration in a region "seem to capture different agglomeration forces" (p. 1791). It may well be the case that both MAR (related) and Jacobian (unrelated) forces are at play within a cluster.

Having established how clusters align with related and unrelated variety, our focus here is to assess the degree to which economic performance of regions is concomitant with growing clusters. In other words, we accept the economic theory of agglomeration externalities on that cluster growth theory stands, and apply the analysis and definitions of clusters as reflected in the work of Delgado, Porter, and Stern (2014, 2016). Neffke, Henning, and Boschma (2011), Frenken et al. (2007), and Henderson (1997). Our task is to broaden the empirical findings.

We do so by measuring the degree to which a region's profile of cluster diversity, strength, and growth relate to broad economic performance. This study develops three measures for cluster diversity/diversification, strength/specialization, and a measure to identify and quantify a region's growth clusters. To address the question of how cluster performance relates to regional competitiveness and economic growth, we regress a set of six cluster measures—each of the three measures noted above is applied to both traded and nontraded industries—on rates of GDP growth, GDP per employee growth, compensation per employee growth, and per capita personal income (PCPI) growth for the years 2002-2013, using data for 366 metropolitan statistical areas (MSAs). The regressions include control variables to avoid model misspecification.

Data and Method

Data Sources

We investigate these relationships using a new employment data set developed by the Indiana Business Research Center. This county-based data set originates as the Bureau of Labor Statistics Quarterly Census of Employment and Wages (QCEW), for where there are a significant number of null or suppressed data, especially as the level of industry and geographic detail increase. The Indiana Business Research Center has created a "QCEW-complete" data set by estimating the suppressed and undisclosed cells for each county-industry pair.[2]

The U.S. CMP uses County Business Pattern (CBP) data, along with data on input–output relationships, to identify 67 industry clusters for the nation (Delgado et al., 2016). This study uses the CMP definitions to group 6-digit North American Industry Classification System (NAICS) industries into 67 traded and local industry clusters.[3] Traded industries are those that agglomerate in particular regions and sell products or services to other regions or countries; for example, a computer chip manufacturer. Local industries, in contrast, primarily serve the local population and, as a result, local industries are not typically overrepresented or underrepresented in any one region (e.g., primary care physicians or grocery stores). Thus, we aggregated 6-digit QCEW-complete data using CMP cluster categories.

We consider our QCEW-complete data superior to the raw CBP data for three reasons. First, QCEW data are more timely. CBP data are not released for more than a year after collection. Second, the CBP data are based on employment in 1 week in March and are therefore subject to seasonality distortions (e.g., in March, accountants and tax preparers are overrepresented, while amusement park workers are underrepresented). Third, the QCEW data are more comprehensive. It covers more industries; for example, government employees and agricultural workers. As a result of the latter, we expanded the CMP cluster list to include four additional industry clusters, bringing the cluster count to 71. All four of the cluster performance measures were derived from the QCEW-complete data. Industry-to-cluster assignments from the CMP can be found in Appendix A.

The Bureau of Economic Analysis is the source for regional GDP and wage data. The QCEW-complete is the source for employment estimates used to derive "per worker" dependent variables. The Bureau of the Census American Community Survey is the source of the population, population density, and educational attainment (both bachelors and high school) figures. Natural resource dependency is measured by the share of QCEW-complete employment in resource-based CMP-defined clusters. The Bureau of Labor Statistics is the source for unemployment data. The Department of Education is the source for data on research universities.

All variables are presented at the MSA level. MSA data for Alaska, Hawaii, Virginia, and Washington, District of Columbia are not included due to missing values and geographic reporting differences. Our final data set contains observations for 366 MSAs for the years 2002-2013.

As aforementioned, we use employment data at a high level of industry granularity—6-digit NAICS—to construct the cluster employment aggregations. Many data are estimates for suppressed cells, not official data. As a result, there may be some noise in the estimates. To reduce the potential of a particularly large margin of error for one data point in one particular year for one particular industry, we took the average of the first 2 years and last 2 years of the industry employment time series.

Cluster Diversity

"Diversity" is a concept that features prominently in a variety of disparate disciplines—primarily ecology—but also in the physical, life, and information sciences, as well as in economics. Industrial diversification has long been of interest to regional economists, especially as it relates to a region's economic stability. For example, Bahl, Firestine, and Phares (1971) presented Flint, Michigan as an example of a region that enjoyed a relatively high level of income, but was subject to great cyclical instability because of the dominance of durable goods manufacturing. Hackbart and Anderson (1975) applied the "Shannon entropy function" to sectoral employment shares to measure diversity, to which Wasylenko and Erickson (1978) countered that absolute industry diversity may not result in greater economic stability than regions that have one dominant sector, citing examples like large universities or a federal government installation.

Frenken et al. (2007) expected the employment effects of related and unrelated variety to differ. That is, unrelated variety would better protect a region against economic shocks and, therefore, rising unemployment. This led them to hypothesize that unrelated variety would militate against unemployment growth and construct an entropy measure to discern diversity effects on their dependent variables. They found that regions with higher unrelated (industry) variety experience lower rates of unemployment growth. More recently, Feser, Mix, White, and Poole (2014) used an entropy index to measure industrial diversity, following Malizia and Ke (1993).

Dependence on a particular industry cluster implies vulnerability to the economic gains and losses of that cluster. To limit the vulnerability to the "ups and downs" of a key regional cluster, policy makers often advocate a strategy of regional industrial diversification.[4] If the key sectors of a region see their competitive position threatened, that region will become vulnerable to the struggles of these industries (Feser et al., 2014). Thus, in the sense of economic resiliency or vulnerability, the notion of diversification (or diversity) is one of unrelated variety.

The concern of Feser et al. (2014) findings of a region's vulnerability to a particular cluster raises the question of whether clusters—aggregations of related firms—can themselves be related to other clusters, thus implying that a region could be vulnerable to a set of clusters. These clusters may be components of a specialized supply chain or they share similar labor force skill sets. Delgado et al. (2014) found that clusters can overlap—that is, be related to each other—like the automotive and metal manufacturing clusters. The degree to which a set of clusters are related can be subject to the same demand shocks that would be relevant information to regional officials drafting a development strategy. But, that is a different question from simply addressing whether cluster diversification—putting the economic eggs in several baskets—influences economic performance broadly.

Which is the best measure of diversity? Stirling (1998) developed a set of criteria for selecting a measure or index: balance, variety, and disparity. Balance is a function of the pattern of apportionment of elements across categories. It is analogous to statistical variance. All else being equal, the more even the balance, the greater the diversity. Variety refers to the number of categories into which the quantity in question can be partitioned. The greater the variety of a system, the greater the diversity. Disparity is related to how different each category is from other categories. Stirling suggested that for his "dual concept" of diversity—that is, balance and variety—the Shannon function is preferred. (See also Stirling, 2007, for an in-depth discussion.)

What defines one cluster in contrast to another cluster is based on differences in their production functions; that is, differences in their supply chains and necessary inputs, and differences in employment concentration. We posit, therefore, that the disparity element discussed by Stirling is already accounted for within a cluster as related variety. Outside a cluster, on the other hand, is unrelated variety, based on the cluster definition methodology.[5] In short, industries are different outside a cluster, although not absolutely as already noted (Delgado et al., 2014). As a result, we measure the diversity of our CMP-based cluster aggregations using the Shannon Evenness Index (SEI). For each cluster (*cl)* in the region of interest (*g)*, we calculate the proportion of employment in that cluster $(p^{cl}_{g,te})$ of the past 2 years available (*te)*. For each region with n number of clusters with an average employment of five or more, the SEI, SEI_g is as follows:

$$SEI_g = \frac{\sum_{cl=1}^{n}[p^{cl}_{g,te} * ln\left(p^{cl}_{g,te}\right)]}{ln\left(n\right)} \tag{1}$$

An SEI score of one indicates a completely even distribution of employment across industry clusters within a region, while a score close to zero indicates great concentration in one or two clusters. Finally, the SEI for the region of interest is divided by the national SEI to index it relative to the nation. The national SEI for 2012 and 2013 (averaged) is 0.78.

The profiles for traded and nontraded, or local, clusters are expected to be different. The relative size of local clusters—industries that serve the region's population—should be consistent across regions, while the relative size of traded clusters can vary dramatically due to the presence or absence of natural resources or the legacy of specialization in certain industries. For this reason, the SEI was calculated for traded and nontraded clusters separately.

Cluster Strength

Our cluster strength, or specialization, measure applies LQs to determine which clusters may dominate employment in a given region. Cluster strength in a particular year is a modified LQ analysis that uses the number of clusters in a region to scale the sum of all LQs in the region. Like an LQ, the proportion of employment in a cluster in a region is compared with the proportion of the nation then summed across all clusters in region *g*.

The sum of LQs has little analytical power, but if divided by the number of clusters/industries/options, the cluster strength measure, $clstrsstr_g$ approaches one as the region becomes more diverse and similar to the nation as a whole, and increases above 1 as clusters and their relative size become more dominant. Like LQs, the relative value for a particular cluster provides an indication of concentration for a cluster (or industry) relative to clusters in the region. The sum of LQs for the United States is equal to the number of clusters (71).

The first step in calculating a region's cluster strength is to calculate the cluster LQs; that is, to divide the cluster proportion component for a cluster in region $(p^{cl}_{g,te})$ by the corresponding national value for that cluster $(p^{cl}_{US,te})$. Then, cluster strength, $clstrsstr_g$, in region g is the sum of all the individual cluster LQs in the region divided by the n number of clusters present in the region. The sum of LQs for the United States is equal to the number of clusters, namely 71.

$$clstrsstr_g = \frac{\sum_{cl=1}^{n}\left[LQ^{cl}_{g,te}\right]}{n} = \frac{\sum_{cl=1}^{n}\left[\frac{p^{cl}_{g,te}}{p^{cl}_{US,te}}\right]}{n} \tag{2}$$

The concentration of employment for traded and local industries is expected to be much different. The relative concentration of local clusters is not expected to vary across regions. The concentration of employment for traded clusters can vary dramatically as regions typically specialize in different industries. For this reason, the measure for cluster strength was calculated for traded and nontraded clusters separately.

Cluster Growth

Our third measure, the cluster growth factor, measures the percentage of employment growth in a region—even in a region that is losing jobs in the aggregate—that can be attributed to the relative concentration of clusters.[6]

This technique identifies regional growth clusters (RGC); that is, regional clusters of industries that are growing more rapidly than the national average growth rate for those groupings of industries. An RGC is subject to multiple criteria that are frequently applied to regional analysis of industry concentration and growth: it is a nontrivial portion of a region's economy, it has experienced employment growth, and it has increased its share of the regional economy. Shift-share analysis and the use of LQs are standard-issue tools that regional economists and development practitioners use to understand their region's economy (Stimson, Stough, & Roberts, 2006). In this case, we apply these tools to identify the percentage of total regional employment change that can be attributed to RGC. The greater the percentage, the greater the role that RGCs had in a region's job growth.

Delgado et al. (2014) found that initial cluster strength facilitates the emergence of industries in a regional cluster and that those emerging firms could be start-ups or incumbent firms from outside the region entering through new or existing plants. They also noted that an analysis of the types of firms driving the emergence-related growth is an important area for additional research. Indeed, we merely have different terms to describe the types of firms: firms generated from the region's economic metabolism are akin to start-ups and firms that are magnetically attracted to the region due to the benefits of agglomeration are akin to incumbent firms entering the region.

There is, however, a cautionary note regarding those incumbent firms entering the region with new investment in plant and equipment. Without conditions or limits, a rate of growth value for a region does not discriminate between metabolic cluster growth—growth attributed to investment and resources internal to the region but open to technologies, human capital, and knowledge developed from outside the region; magnetic growth—new incoming investments and establishments owned by firms from outside the region—and what may be called "parachute growth." In the case of parachute growth, an industry or cluster can simply appear as highly concentrated in the employment statistics after an incoming firm hires many workers for a large new plant in a region that heretofore, in that industry (or cluster), had little or no employment. An example may include Amazon building fulfillment and distribution centers in selected cities across the country.

Given that the underlying drivers of clusters are economies of agglomeration and regional characteristics that have sometimes been described as "something in the air," metabolic and magnetic growth are in closer alignment with cluster development than parachute growth. One could argue that firms parachuting into a region and immediately dominating the regional economic and employment landscape do so for reasons unrelated to the presence of a well-developed cluster.[7] Rather, those firms are likely seeking lower cost labor, cheaper electricity, transportation corridors, looser environmental regulations, and the like. To remove from the analysis those instances that a cluster may appear without any apparent externalities associated with agglomeration, there is a criterion—a threshold to identify false positives—to remove clusters that are likely attributed to parachute jobs.

The goal is to unambiguously identify the clusters that drive regional growth, or even counterbalance the effects of regional employment decline, while excluding parachute employment gains.

First, we calculate a modified shift-share analysis for region g for the past 10 years of available data. The shift-share ratio for cluster cl in region g, $SS^{cl}{}_{g}$, is the LQs for the region at the end of the time series, te, divided by the LQ for the region at the beginning of the period, tb. In addition, we use the 2-year average of LQs at the beginning and end of the period to moderate the influences of economic cycles or noise in the data. We also remove missing or zero-employment clusters.[8]

$$SS^{cl}{}_{g} = {LQ^{cl}_{g,te}} \Big/ {LQ^{cl}_{g,tb}} \tag{3}$$

There are four criteria for selecting a RGC:

1. The cluster has been growing (i.e., the change in employment has been positive).
2. $SS^{cl}{}_{g}$ is greater than 1 (i.e., the cluster grew in relative importance in the region).
3. $SS^{cl}{}_{g}$ is less than 1 plus 2 standard deviations of $SS^{cl}{}_{g}$.
4. The proportion of employment in cluster cl during the period te is greater than 0.005.

As stated above, the goal is to unambiguously identify clusters that drive economic growth. While a cluster of relatively high regional concentration may well benefit from the economies of agglomeration even as its overall employment in the region is shrinking, employment growth would unassailably indicate the region's competitive advantage in that cluster. This would also be the case for the cluster's share of employment in the region to grow over time as stated in criterion number 2. To be considered a RGC, the rate of employment growth needs to be greater than the national average growth rate for that cluster to indicate the region having a relative competitive advantage in that cluster.

To remove the parachute growth phenomenon for any particular cluster, the upper limit to $SS^{cl}{}_{g}$ is the employment "false positive flag" to eliminate falsely identified RGC. The parachute growth false positive flag is operationalized as one plus twice the standard deviation of the array of $SS^{cl}{}_{g}$ (criteria 3 above).[9] This keeps high-growth industries in the RGC category, while removing the most likely parachute establishments. The threshold is loose enough to allow for magnetic growth, but sufficiently tight to remove the most egregious instances of parachute growth.

The purpose of criterion 4—the proportion threshold—is to remove clusters that have little or no material or economic influence on a region; however, because they have started very small and grew to small over the period, they satisfied criteria 1 through 3. For example, the objective here is to exclude the artisanal leather handbag maker who may have started with 3 workers and grew to 10 over the period. Such a firm (or clusters) would have a high-growth rate, but would be of little economic significance. The aforementioned set of criteria was developed to be flexible enough to be applied to any regional definition, urban or rural, a single county, or a regional collection of counties.

The growth of dominant clusters—*clstrgrw*—is a percentage that represents the employment growth of RGCs compared with the total employment for the entire region ($ttlemp^{QCEW}{}_{g,tb}$):

$$clstrgrw_{g} = \frac{\sum_{rgc=1}^{Nrgc}\left(emp^{rgc}_{g,te}/2 - emp^{rgc}_{g,tb}/2 \right)}{\left(\frac{\left(ttlemp^{QCEW}_{g,te} \right)}{2} \right)} \tag{4}$$

Nrgc is the number of RGC in region *g* and $emp^{rgc}{}_{g,te}$ is employment in RGCs in region *g* and during period *te*. $clstrgrw_{g}$ is interpreted as the percentage of total employment attributable to RGC. The greater the percentage, the greater the role that RGCs had in job growth. There is no obvious national value for $clstrgrw_{g}$ because at the national level there is no regional cluster specialization. The percentage is for interregional comparison.

Finally, we further disaggregate cluster growth into local and traded cluster growth. As noted earlier, we used cluster designations as defined by the most recent iteration of the CMP (Delgado et al., 2016), an update of work by Porter (2003). The CMP distinguished between industries based on their competitive environment both within the region and external to the region: local (or nontraded) and traded. Local industries primarily serve local residents and markets, and, for this reason, the proportion of local cluster employment to total population does not greatly vary across regions. For example, the total square footage and total employment for all grocery stores in a region is a function of the population size. For example, a region does not specialize—as evidenced by high employment concentrations—in grocery stores. On the other hand, regions do specialize in the production of traded industries (e.g., engineering services or medical devices).

The thrust of cluster-based development strategies is to encourage the competitive advantages that accrue to industrial specialization and growth. For this reason, Delgado et al. (2014) analyzed traded industries to explore the strength of clusters on employment and innovation. Also for this reason, we divide the RGC factor into local and traded as well as to tease out the potential effects of robust local clusters that are growing because of a rapidly growing regional population in contrast to growth from traded industries.

Analytical Methods

We apply a set of multivariate ordinary least squares (OLS) regression models to estimate the relationship of our cluster performance measures to economic development indicators. Dependent variables are GDP growth, GDP per worker growth, compensation per worker growth, and per capita income growth. GDP growth is a commonly used measure of economic productivity and GDP-per-worker growth represents the change in labor productivity. Compensation per worker growth can be interpreted as the change in wages, plus benefits and supplements. This represents the change in the total cost of labor to employers. PCPI is another commonly used measure of economic development and well-being, but the measure is not as directly tied to current returns to labor as compensation because it includes transfer payments, proprietors' income, rental income, and personal dividend income, among other sources of income. All dependent variables are presented as a rate of change from the 2002-2003 average to the 2012-2013 average.

We include the proportion of the population with a high school diploma, the proportion of the population with a bachelor's degree, and the number of accredited research universities as independent variables. This allows us to control for the effects of human capital on regional economic performance. High school graduation rates are calculated for aged 18 to 25 years. College completion rates are calculated for aged 25 years and above. Education variables are calculated from measurements taken at the end of the study period (ACS data for educational attainment were not collected in 2002 and 2003).

The change in the unemployment rate from the 2002-2003 average to the 2012-2013 average is included to control for business cycles, as well as for tight labor markets. The population change over the study period and 2013 population density are included to address endogenous growth due to larger regional markets, availability of labor, and agglomeration economies that accompany greater density and size.

Table 1. Descriptive Statistics for Variables Used in Regression Analysis.

Variable	*N*	*M*	*SD*	Sum	Median	Min	Max
GDP growth	366	0.03764	0.01519	13.77689	0.0353771	−0.00549	0.14802
GDP-per-worker growth	366	0.03337	0.00952	12.2116	0.0324	0.0098	0.1051
Per capita income growth	366	0.32727	0.07843	119.7821	0.31755	0.147	0.9183
Compensation per worker growth	366	0.02512	0.00872	9.19382	0.0259749	−0.05698	0.05038
Cluster diversity*	366	0.9665	0.04219	353.74	0.97	0.83	1.1
Cluster strength*	366	1.24104	0.56835	454.22	1.08	0.61	4.87
Local cluster growth	366	0.0799	0.04438	29.24445	0.0736064	0.000516	0.43243
Traded cluster growth	365	0.04498	0.03376	16.46173	0.0379332	0	0.24243
Resource dependency	366	0.00759	0.02271	2.778	0.0009	0	0.2727
Population growth	366	0.01112	0.01418	4.07	0.01	−0.06	0.12
Bachelor's	366	0.16288	0.04419	59.6151	0.1596	0.0747	0.3188
High school	366	0.86469	0.04683	316.477	0.865	0.697	0.985
University count	366	0.49454	1.14136	181	0	0	13
Population density	366	274.2235	310.3354	100.366	187.3021464	7.23969	2670
Unemployment	366	0.28124	0.19249	102.933	0.26915	−0.3968	0.9077

Note. GDP = gross domestic product.

The natural resource dependency of an MSA is used to control for the relative windfall that regions endowed with natural resources or their processing and transportation may enjoy when resource prices are high. Slaper, Hart, Hall, and Thompson (2011) found the presence of fossil fuel extraction to have a positive effect on GDP-per-worker growth from 1997 to 2006. In this case, we broaden the definition to include all resources[10] and their associated processing and distribution as prescribed by the CMP (see Appendix B for the four clusters and their NAICS industry components). This variable represents the proportion of employment in resource-based clusters during the boom years of 2012-2013, and helps control for situations for which an extractive industry distorts overall trends in regional development (e.g., rapid expansion of oil industry employment in regions where petroleum is discovered).

After aggregating all data to the MSA level, we estimated a set of four OLS regression models of the following general specification:

$$\begin{aligned} Y = \beta_0 &+ \beta_1 LocalClusterDiversity \\ &+ \beta_2 TradedClusterDiversity \\ &+ \beta_3 LocalClusterStrength \\ &+ \beta_4 TradedClusterStrength \\ &+ \beta_5 LocalClusterGrowth \\ &+ \beta_6 TradedClusterGrowth \\ &+ \beta_7 PercentHS \\ &+ \beta_8 PercentBachelors + \beta_9 UniCount \\ &+ \beta_{10} \Delta PopTotal \\ &+ \beta_{11} PopDensity + \beta_{12} \Delta Unemployment \\ &+ \beta_{13} \Delta ResourceDependency + \varepsilon \end{aligned} \tag{5}$$

where Y is the dependent variable of interest and ε is a normally distributed error term. All modeling and calculations were completed using Microsoft Excel and SAS 9.4.

Results

Descriptive Statistics

Descriptive statistics for all variables are presented in Table 1.

The correlation matrix is presented in Table 2. Three of the four dependent variables—change in GDP, change in GDP per employee, and change in PCPI—are moderately to strongly positively related. Change in average employee compensation per worker, $\Delta Comp/Emp$, is negatively related to GDP growth and PCPI growth, and only weakly positively related to our productivity growth measure, GDP per employee. In terms of the explanatory variables and $\Delta Comp/Emp$, one sees a negative relationship with unemployment growth—as unemployment goes up, compensation goes down—and that population and local cluster growth also have a negative effect on compensation growth.

The other correlation of particular note is the strong positive correlation between local (or nontraded) cluster diversity and cluster strength. This relationship makes sense as the relative concentration and diversification of local industries and clusters should be similar across all regions, given that the scale of these regional industries is primarily driven by a region's population size and then secondarily by a region's demographic make-up. Also as expected, we do not see positive relationships between the traded and local components of cluster strength or diversity.

Table 2. Correlation Matrix.

Pearson correlation coefficients, $N = 366$

Probability > |r| under H_0: Rho = 0

	GDP growth	GDP/worker growth	Compensation/ worker growth	PCI growth	Traded CD	Local CD	Traded CS	Local CS	Traded CG	Local CG	Resource dependency	Population growth	Percentage of bachelor's	Percentage of HS	R1 universities	Population density	Unemployment growth
GDP growth	1																
GDP/worker growth	0.78	1															
Compensation/worker growth	−0.39	0.20	1														
PCI growth	0.70	0.60	−0.13	1													
Traded CD	0.04	0.14	0.09	−0.06	1												
Local CD	−0.19	−0.18	0.00	−0.02	−0.66	1											
Traded CS	0.03	0.16	0.17	0.00	0.51	−0.42	1										
Local CS	−0.17	−0.15	−0.02	0.03	−0.48	0.86	−0.26	1									
Traded CG	0.45	0.29	−0.20	0.29	0.29	−0.45	0.21	−0.35	1								
Local CG	0.28	−0.07	−0.49	0.16	−0.15	0.11	0.03	0.06	−0.08	1							
Resource dependency	0.60	0.64	−0.01	0.59	0.02	−0.10	0.15	0.01	0.33	−0.04	1						
Population growth	0.37	0.06	−0.46	0.11	−0.09	−0.03	−0.08	−0.08	0.19	0.31	0.12	1					
Percentage of bachelor's	0.10	−0.05	−0.24	−0.07	0.04	−0.06	−0.43	−0.05	−0.11	−0.01	−0.13	0.13	1				
Percentage of HS	−0.02	−0.01	0.02	−0.02	0.08	−0.03	−0.23	−0.04	−0.11	−0.01	−0.19	−0.06	0.53	1			
R1 universities	0.01	0.03	0.03	−0.09	−0.08	0.10	−0.05	0.07	−0.03	0.00	−0.02	0.01	−0.01	−0.09	1		
Population density	−0.16	−0.12	0.06	−0.16	−0.01	0.06	−0.22	0.09	−0.27	−0.16	−0.13	−0.08	0.29	0.05	−0.04	1	
Unemployment growth	−0.38	−0.34	0.16	−0.40	−0.15	0.04	−0.02	−0.04	−0.14	0.02	−0.36	0.01	−0.12	−0.02	0.01	0.12	1

Note. GDP = gross domestic product; PCI = per capita income; CD = cluster diversity; CS = cluster strength; HS = high school; CG = cluster growth.

Table 3. Results of OLS Regressions (All Variables Included).

	Dependent variables			
	ΔGDP	ΔGDP/employee	ΔCompensation/employee	ΔPCI
Adjusted R^2	0.6173	0.4605	0.4777	0.4454
$\hat{y}$ ± MSE	0.03764 ± 0.0094	0.03337 ± 0.00699	0.02512 ± 0.0063	0.32727 ± 0.05841
$p(F)$	<.0001*	<.0001*	<.0001*	<.0001*
Intercept	0.0112	−0.0086	−0.0007	0.2523
Local cluster diversity	0.1133 (.1827)	0.2723 (.0072)*	0.2089 (.0358)*	−0.0159 (.8764)
Traded cluster diversity	−0.0313 (.5320)	0.0582 (.3279)	0.0530 (.3652)	−0.0927 (.1244)
Local cluster strength	−0.2020 (.0036)*	−0.3285 (<.0001)*	−0.2145 (.0081)*	0.0014 (.9866)
Traded cluster strength	−0.0184 (.6926)	0.0804 (.1466)	0.1456 (.0078)*	0.0898 (.0898)
Local cluster growth	0.2801 (<.0001)*	−0.0351 (.4081)	−0.4394 (<.0001)*	0.1927 (<.0001)*
Traded cluster growth	0.2761 (<.0001)*	0.0665 (.1713)	−0.2454 (<.0001)*	0.1566 (.0016)*
HS graduation	0.0415 (.3019)	0.1367 (.0044)*	0.1517 (.0013)*	0.1187 (.0145)*
Bachelor's degree	0.1174 (.0116)*	−0.0193 (.7254)	−0.2275 (<.0001)*	−0.1052 (.0597)
R1 university count	0.0312 (.3421)	0.0590 (.1305)	0.0385 (.3159)	−0.0013 (.0755)
Population growth	0.1506 (<.0001)*	−0.0102 (.8153)	−0.2487 (<.0001)*	−0.0360 (.4138)
Population density	0.0295 (.4273)	0.0239 (.5882)	0.0102 (.8143)	−0.0013 (.9762)
Unemployment growth	−0.0108 (<.0001)*	−0.1254 (.0043)*	0.1536 (.0004)*	−0.2346 (<.0001)*
Resource dependency	0.4809 (<.0001)*	0.6189 (<.0001)*	0.1406 (.0019)*	0.4894 (<.0001)*

Note. OLS = ordinary least squares; GDP = gross domestic product; PCI = per capita income; HS = high school; MSE = mean squared error. Standardized beta coefficients reported for all independent variables. Asterisks on *p* values indicate statistical significance.

Regression Analysis

All models were statistically significant at the 0.01 level. In each of the four generally specified models, several of the independent variables were not statistically significant. Population density and the number of research universities were not significant in any model. The control variables for resource dependence and unemployment were significant across the board.

As these models are exploratory, we tested a second procedure wherein insignificant variables are removed in an iterative fashion to avoid bias due to overspecification. All independent variables included in the final models are significant at the 0.05 level. Table 3 shows the full model, while Table 4 shows the final model.

Our final model for GDP growth is as follows:

$$\begin{aligned}\Delta GDP = \beta_0 &+ \beta_1 LocalClusterDiversity \\ &+ \beta_2 LocalClusterStrength \\ &+ \beta_3 LocalClusterGrowth \\ &+ \beta_4 TradedClusterGrowth \\ &+ \beta_5 PercentBachelors \\ &+ \beta_6 \Delta PopTotal - \beta_7 \Delta Unemployment \\ &+ \beta_8 ResourceDependency + \varepsilon\end{aligned} \tag{6}$$

This model has the highest practical significance of the four, with the adjusted R^2 indicating that 61.9% of the variation in regional GDP is accounted for by our model. The standardized parameter estimates indicate that natural resource dependency has by far the strongest positive effect on GDP growth. Traded cluster growth is next most significant, followed by local cluster growth. Local cluster diversity and strength are also both significant, but present opposite signs for their standardized betas, indicating that a local cluster profile resembling the national profile for local clusters is slightly positively related to overall GDP growth, while regions with a relatively high concentration of local clusters tend to be negatively related to GDP growth. This may be explained by the fact that wages in many local service industries are lower than average, and that regions having a super abundance of retail or hospitality enterprises do not experience higher rates of GDP growth. Population growth and college education are shown to affect GDP growth as well. These variables are all significant at the $p < .05$ level or better, with results that are consistent with regional development theory.

Our "productivity" model for GDP per employee growth is as follows:

$$\begin{aligned}\frac{\Delta GDP}{Emp} = \beta_0 &+ \beta_1 LocalClusterDiversity \\ &+ \beta_2 LocalClusterStrength \\ &+ \beta_3 TradedClusterStrength \\ &+ \beta_4 PercentHS \\ &- \beta_5 \Delta Unemployment \\ &+ \beta_6 ResourceDependency + \varepsilon\end{aligned} \tag{7}$$

Table 4. Results of OLS Regressions (Insignificant Variables Iteratively Removed).

	Dependent variables			
	ΔGDP	ΔGDP/employee	ΔCompensation/employee	ΔPCI
Adjusted R^2	.6195	.4605	.4724	.4449
$\hat{y} \pm$ MSE	0.03764 ± 0.00937	0.03337 ± 0.00699	0.02512 ± 0.00634	0.32727 ± 0.05843
$p(F)$	<.0001*	<.0001*	<.0001*	<.0001*
Intercept	0.0082	0.0081	0.0147	0.1782
Local cluster diversity	0.1579 (.0215)*	0.1885 (.0259)*		
Traded cluster diversity				
Local cluster strength	−0.2170 (.0010)*	−0.2993 (.0002)*		
Traded cluster strength		0.0922 (.0415)*	0.1399 (.0013)*	−0.1371 (.0019)*
Local cluster growth	0.2757 (<.0001)*		−0.4423 (<.0001)*	0.1942 (<.0001)*
Traded cluster growth	0.2710 (<.0001)*		−0.2461 (<.0001)*	0.1361 (.0013)*
HS graduation		0.1234 (.0027)*	0.1525 (.0010)*	0.1243 (.0079)*
Bachelor's degree	0.1572 (<.0001)*		−0.2308 (<.0001)*	−0.1329 (.0084)*
R1 university count				
Population growth	0.1479 (<.0001)*		−0.2398(<.0001)*	
Population density				
Unemployment growth	−0.1710 (<.0001)*	−0.13258 (.0015)*	0.1543 (0.0002)*	−0.2255 (<.0001)*
Resource dependency	0.4808 (<.0001)*	0.6245 (<.0001)*	0.1152 (.0090)*	0.4949 (<.0001)*

Note. OLS = ordinary least squares; GDP = gross domestic product; PCI = per capita income; HS = high school; MSE = mean squared error. Standardized beta coefficients reported for all independent variables. Asterisks on p values indicate statistical significance.

This model has moderate practical significance, with the adjusted R^2 indicating that 46.1% of the variation in GDP per employee growth is accounted for by the independent variables. As with GDP growth, natural resource dependency has the strongest effect on GDP per worker, with a standardized parameter estimate of .624. Also, following the GDP growth model, local cluster diversity and strength present opposite signs and also hint that a region's productivity growth can be affected positivly or negatively depending how closely its cluster profile reflects the national average. Traded cluster strength surfaces as having a small, positive effect on productivity (standardized beta of .09), in line with the expected benefits of MAR externalities.[11] High school attainment is significant and positive, whereas the level of bachelor's degree attainment was not.

Our model for compensation per worker growth is as follows:

$$\begin{aligned}\frac{\Delta Comp}{Emp} = {} & \beta_0 + \beta_1 TradedClusterStrength \\ & - \beta_2 LocalClusterGrowth \\ & - \beta_3 TradedClusterGrowth \\ & + \beta_4 PercentHS - \beta_5 PercentBachelors \\ & - \beta_6 \Delta PopTotal + \beta_7 \Delta Unemployment \\ & + \beta_8 ResourceDependency + \varepsilon\end{aligned} \tag{8}$$

This model has moderate practical significance, with the adjusted R^2 indicating that 47.2% of the variation in compensation per worker growth is accounted for by our model. In contrast to the other models where natural resource dependency has the greatest effect on the dependent variable, traded cluster strength has the strongest positive impact on compensation per worker growth, followed by high school attainment and unemployment growth. Several variables have negative standardized betas that exceeded the magnitude of the positive betas mentioned above, namely both traded and local cluster growth, together with population growth and bachelor's degree attainment. These results may be pointing to the fact that average workers have not experienced an increase in their paychecks over the period, irrespective of working in a traded or local cluster. Moreover, the fact that only resource dependent areas have seen their average paychecks increase is because of the boom in the oil and gas industry—an industry that typically pays well—and legacy unionized industries that can bargain for increased benefits such as health care, even as their take-home pay remains constant.

Our model for per capita income growth is as follows:

$$\begin{aligned}\Delta PCI = {} & \beta_0 + \beta_1 TradedClusterStrength \\ & + \beta_2 LocalClusterGrowth \\ & + \beta_3 TradedClusterGrowth + \beta_4 PercentHS \\ & + \beta_5 PercentBachelors - \beta_6 \Delta Unemployment \\ & + \beta_7 ResourceDependency + \varepsilon\end{aligned} \tag{9}$$

This model has moderate practical significance as well, with an adjusted R^2 indicating that 44.4% of the variation in PCPI is explained by our model. Standardized estimates indicate that natural resource dependency again has the strongest positive effect on per capita income, followed by both local and traded cluster growth. The unemployment rate and PCPI have the strongest negative relationship—which is expected—followed by traded cluster strength—which was not expected given the mildly positive effect on productivity and average compensation. What may explain the negative relationship is both the components of PCPI and how those components have changed since 2002. PCPI includes not only compensation (67.1% of PCPI in 2002) but also proprietary income (9.5%), transfer payments (14.0%), and income from assets such as stock or rental housing (2.4%). (There are others.) The proportion of compensation fell from 2002 to 2012 by 7.8%, while the proportion of income related to personal dividend income and transfer payments both increased 37.8% and 21.2%, respectively.[12] In other words, personal income can be increasing for a small subgroup, while most average employees are not enjoying any increase in compensation. Thus, one should not be surprised to see strong traded clusters register positively in terms of productivity measures and compensation growth measures, but not in PCPI measures.

All models were tested for violations of the general linear model. Ubiquitous low-variance inflation factors indicate that near multicollinearity is not an issue in our models. We tested for heteroskedasticity using the SPEC procedure in SAS, which compares White's (1980) heteroskedasticity-consistent covariance matrices against the standard OLS covariance matrices. We did not find evidence for heteroskedasticity in any of the four models. Our Durbin–Watson statistics, which can reveal potential misspecification errors through patterns in the residuals, show no evidence of this problem affecting the models. Because the data are cross-sectional and no spatial variables are included in the models, autocorrelation should not be a problem.

Discussion

Out of the six measures of cluster performance developed for this study, five were statistically significant in terms of their relationship to at least one of the four traditional economic development indicators. Traded cluster growth is shown to have an effect on three of four of our dependent variables.

Traded cluster growth and local cluster growth are generally statistically significant at or near the .01 level in three of the four regressions, with each measure explaining a substantial proportion of the total explained variation in the dependent variables. The signs on these two cluster growth indicators are positive with respect to GDP growth and per capita income growth, and negative for growth in compensation per employee. The positive signs on traded cluster growth and local cluster growth are consistent with the predictions of regional economic development theory, particularly for the regressions on GDP growth and PCPI growth. The reversal of sign for the cluster growth indicators, as well as their insignificance for the GDP per employee model, is consistent with the stagnation of wage growth across the U.S. economy throughout the sample period of 2002-2013 that is addressed previously. As noted by Mishel and Shierholz (2012), wage growth during the study period grew only slightly or stagnated, particularly during the Great Recession. Thus, productivity dramatically outperformed compensation growth, leading to our hypothesis that the negative coefficients for change in compensation per employee can be attributed to the nature of population growth and labor market dynamics. Assuming this hypothesis is accurate, our measures of cluster growth have a theoretically valid and statistically significant relationship to three of our four dependent variables. Since these are among the generally accepted indicators of regional economic development, this analysis supports the efficacy of the traded cluster growth and local cluster growth measures as indicators of cluster development.

The measure of local cluster diversification is significant with a positive sign in the regressions on GDP growth and GDP per employee growth, whereas traded cluster diversification does not appear in any of the final models as a significant variable. Given that these two measures relate more to economic resilience to decline or cyclical economic shock, rather than economic growth, they may only be relevant for regions that are experiencing slow growth or decline. Large-scale regional economic stagnation that is the result of national economic trends and decaying regional export linkages—the United States was losing millions of manufacturing jobs well before the Great Recession—would likely negate any positive effects of traded cluster diversification. On the other hand, the greater the effect of such poor national economic performance, the more important local diversification becomes as an insulator against these trends beyond the region. Additional research is needed to explore the efficacy of these measures in situations beyond those explored in this study.

Traded cluster strength is statistically significant with positive signs in the per employee regressions and a negative sign in the per capita income change model, whereas local cluster strength is statistically significant with negative signs in the two GDP regressions. These results indicate that as the concentration of export-oriented traded clusters increase in a region, the change in GDP per employee and compensation per employee also increase. This is consistent with regional economic theory concerning the benefits of agglomeration economies. The negative sign for the per capita income regression, as noted above, is again likely due to changes in trends for per capita income over the sample period. On the other hand, as indicated previously, the negative signs for local cluster strength in the GDP regressions likely reflect lower than average wages for retail and service industries, and that regions with high concentrations of retail or hospitality enterprises experience lower rates of GDP growth.

Out of all the control variables included in the analysis, natural resource dependency has the strongest influence on GDP growth, GDP per employee growth, and per capita income growth. This can be attributed to the rise of hydraulic fracking and exploiting unconventional oil and gas resources. These factors have resulted in phenomenal investments in infrastructure and employment in regions that even 10 years ago could have been considered economic backwaters. All but one of the top 10 MSAs in terms of GDP per employee growth was resource based. When times are good for natural resource industries, they are very good.

Our disaggregation of traded and local clusters, based on CMP definitions, may reflect poor assignments of what have been defined as "local" industries. Many medical services are no longer local and many regional hospitals serve clients who live well beyond the boundaries of the MSA. Clinics are replacing hospitals in many rural areas, ceding even relatively minor procedures to larger facilities in the city. Portland, Maine ranked 88th in employment growth among the 366 MSAs, but driven by a large uptick in local health services—a cluster that pays relatively well—it ranked ninth in GDP-per-worker growth. These results suggest that adjustments to the CMP local-versus-traded industry cluster assignments may be warranted to correctly attribute a region's trade outside the official boundaries of an MSA.

Our model of compensation per worker growth raises questions about what exactly is driving compensation per worker change at the MSA level. More specifically, it calls into question whether population growth is overrepresented by college graduates moving to more amenity-rich locations even though they may be underemployed in personal services. There may be an oversupply of bachelor's degrees in some regions, putting downward pressure on wages and compensation. Moreover, amenity-driven migration would not likely be affected by regional unemployment rates, in contrast to regional labor market dynamism attracting skilled or semiskilled workers into goods-producing industries.

An alternative interpretation is that the negative signs on our cluster growth, college education, and population growth variables, as well as the positive sign on unemployment, could be due to the substitution of labor for capital during the study period. This period was characterized by shifts to advanced manufacturing, increased use of information technology, and higher capital-to-labor ratios. Moreover, wage growth during the study period grew only slightly or stagnated, particularly during the Great Recession. Thus, productivity dramatically outperformed compensation growth (Mishel & Shierholz, 2012). The tension in this interpretation, however, is that during the Great Recession and the years following, both productivity growth and business investment rates were subpar with respect to historical levels.[13] This may explain low compensation gains.

Given the nature of the results of this analysis, it is important to point out a potential limitation to their interpretation. It is possible, due to reverse causality in the relationships, that some of the cluster development measures have the potential for nonzero correlations with the error terms of the regression models. For example, although traded cluster growth has a highly significant positive association with GDP growth (i.e., likely to also be causal), changes in GDP growth may in turn affect the demand for employment for both traded and local industries within an MSA. Since employment is the primary measure of economic activity that comprises the formula for traded cluster growth, changes in GDP growth may also affect the level of traded cluster employment growth. This potential bidirectional causality would result in endogeneity bias in the parameter estimate for traded cluster growth in this regression.

Although such endogeneity bias may change the parameter estimates somewhat, it is unlikely to negate the nature or direction of the identified associations in this exploratory research context. Moreover, the measures for strength/specialization and diversity/diversification describe regional characteristics and only weakly covary with the dependent variables, and negatively more often than not. The corrective mechanisms for endogeneity problems, either in the form of simultaneous econometric models or the use of an appropriate (yet difficult to identify) instrumental variable, would be beyond the scope of the current research, which is to identify if there are additional economic benefits concomitant with cluster employment growth.

Conclusion

This research categorizes U.S. regional industrial cluster development using six measures: traded cluster growth, local cluster growth, traded cluster diversity, local cluster diversity, traded cluster strength, and local cluster strength. These measures are developed from industry employment data from the Bureau of Labor Statistics Quarterly Census of Employment and Wages and other economic data from the Bureau of Economic Analysis. Data are from the period 2002-2003 to 2012-2013, and are aggregated to the level of the MSA. The final data set contains observations for 366 MSAs.

The two measures of cluster diversity, one for traded clusters and one for local clusters, utilize the SEI derived initially from the natural sciences. The SEI is a measure of diversity that allows comparison of cluster diversification across MSAs with that of the United States. The two measures of cluster strength use a LQ approach to identify the degree to which clusters dominate employment in an MSA. The two measures of cluster growth indicate the percentage of employment growth in a region that can be attributed to cluster employment growth.

To provide a mechanism for validating and testing these cluster development measures, this research estimates four regressions of regional economic development indicators on the cluster development measures, along with a set of control variables. The dependent variables are GDP growth, GDP per employee growth, compensation per employee growth, and per capita income growth. These are some of the most commonly utilized indicators of regional economic development and economic performance.

The results from these regressions provide preliminary evidence of the validity and efficacy of five of the six cluster growth and development measures: local cluster diversity, local and traded cluster strength, and local and traded cluster growth. These results indicate that 40% to 60% of the variation in the dependent variables is explained by the cluster development measures, along with the control variables.

It is also important to recognize that all four of these measures are based on employment data. As such, they are subject to the usual assumptions concerning labor productivity, industry convergence, capital/labor substitution, regional consumption patterns, spatial dependencies, the accuracy of reporting, and so forth. Further research may be able to address these issues through the use of value-added measures during appropriate time periods and other sociodemographic characteristics that help describe the differences among regions. Moreover, to conduct detailed interregional analysis on cluster development region by region, additional exploration may be needed to determine if the threshold criteria for identifying RGC need to be adjusted. Economic development practitioners considering devoting resources to develop a particular cluster in their region would be interested in knowing what other regions are strong and growing in that particular cluster as well as the socioeconomic characteristics of those regions.

Given the research results, these measures of traded and local cluster development provide a tool for disaggregating and measuring the concept of industrial clusters. Considering the existing lack of quantitative indicators of cluster growth and development, these measures represent a significant contribution to the research on this topic. However, these measures should be explored in other contexts and with other types of data to further assess their validity.

Even with the limitations of this analysis and the preliminary nature of the results, and pending further verification of the identified relationships, this research has direct implications for both public policy and regional economic development strategies. The positive impact of the measures of traded and local cluster growth on the change in regional GDP and per capita income empirically confirms the importance of industrial clusters to public policy. Given that these two dependent variables are some of the most common indicators of regional economic development, it would behoove economic developers and local governments to continue to emphasize strategic economic development policies that focus on the expansion and development of both traded and local clusters. Whereas the role of traded clusters has long been identified in theory, these results also demonstrate the importance of local clusters to economic development policy.

This research has also affirmed the impact of local cluster diversity with respect to GDP-related growth. Traded clusters, with their strong linkages to the external economy, may have greater sensitivity to business cycles or economic decline. An economic development strategy that identifies potential gaps in the diversity of local clusters (or industries) compared with national averages may signal that local clusters need building to enhance resiliency as well as to make visible unexploited market opportunities for local businesses.

In contrast to local cluster strength, our results also emphasize the importance of traded cluster strength or specialization to enhance per employee growth in either GDP or compensation. Given the relevance of these indicators to employment and regional welfare, focusing on traded clusters with high degrees of regional strength or specialization may maximize wages and salaries for regional employees.

Together, these results can help map a blueprint to enhance regional economic development using a targeted strategy based on industrial cluster development. However, it is critical to emphasize that, as is the case with any other development tool or policy, such a focus on targeted industrial cluster development must be part of an overall strategic regional development plan that incorporates the other critical elements of economic development policy and practice.

Appendix A

Cluster Definitions

For a complete presentation of the cluster definitions used here, please see,

> http://www.statsamerica.org/ii2/reports/PorterClusters_expanded.xlsx.

See also, the Cluster Mapping Project at,

> http://clustermapping.us/content/cluster-mapping-methodology

Appendix B

Natural Resource-Based Industries and Clusters.

NAICS code	NAICS title	Cluster title
212111	Bituminous Coal and Lignite Surface Mining	Coal Mining
212112	Bituminous Coal Underground Mining	Coal Mining
212113	Anthracite Mining	Coal Mining
213113	Support Activities for Coal Mining	Coal Mining
113110	Timber Tract Operations	Forestry
113210	Forest Nurseries and Gathering of Forest Products	Forestry
113310	Logging	Forestry
115310	Support Activities for Forestry	Forestry
212210	Iron Ore Mining	Metal Mining
212221	Gold Ore Mining	Metal Mining
212222	Silver Ore Mining	Metal Mining
212231	Lead Ore and Zinc Ore Mining	Metal Mining
212234	Copper Ore and Nickel Ore Mining	Metal Mining
212291	Uranium-Radium-Vanadium Ore Mining	Metal Mining
212299	All Other Metal Ore Mining	Metal Mining
213114	Support Activities for Metal Mining	Metal Mining
211111	Crude Petroleum and Natural Gas Extraction	Oil and Gas Production and Transportation
211112	Natural Gas Liquid Extraction	Oil and Gas Production and Transportation
213111	Drilling Oil and Gas Wells	Oil and Gas Production and Transportation
213112	Support Activities for Oil and Gas Operations	Oil and Gas Production and Transportation
324110	Petroleum Refineries	Oil and Gas Production and Transportation
324199	All Other Petroleum and Coal Products Manufacturing	Oil and Gas Production and Transportation
333132	Oil and Gas Field Machinery and Equipment Manufacturing	Oil and Gas Production and Transportation
486110	Pipeline Transportation of Crude Oil	Oil and Gas Production and Transportation
486210	Pipeline Transportation of Natural Gas	Oil and Gas Production and Transportation
486910	Pipeline Transportation of Refined Petroleum Products	Oil and Gas Production and Transportation
486990	All Other Pipeline Transportation	Oil and Gas Production and Transportation
541360	Geophysical Surveying and Mapping Services	Oil and Gas Production and Transportation

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Notes

1. See also the U.S. Cluster Mapping Project (CMP) at http://www.clustermapping.us.
2. The approach is not unlike Isserman and Westervelt (2006).
3. These are not the clusters defined by Porter (2003) or used by Delgado et al. (2014).
4. It is important to note that diversification does not necessarily imply higher incomes or faster growth. An economy might be considered diverse because, for example, a large manufacturing plant recently closed, thereby decreasing the regional level of specialization. Successful diversification will promote several areas of specialization in such a way that the decline in one sector can be offset by the growth in another sector (Feser et al., 2014).
5. Of course, the converse is true and is the more common expression of a cluster definition; that is, similar location, supply chain, and labor requirements.
6. A cluster-based approach starts with the industries and assets that are already present in the region and regional stakeholders pursue initiatives to make them better. An approach for entirely new clusters is a strategy to improve overall business environment conditions by upgrading skills, by providing access to finance and infrastructure, by streamlining government rules and regulations, by supporting sophisticated local demand, and by being open to foreign investment and competition. Clusters of industries present in a region do not necessarily need public sector strategies to exist—they exist regardless—but the right public strategies can help the businesses within a cluster become more successful and competitive. A cluster-based strategy is not, in other words, primarily organized around attracting large entities from elsewhere.
7. Zheng and Slaper (2017) found that mixed results on the question of whether counties with strong clusters—that is, high relative employment concentration—are more likely to attract greenfield foreign direct investment.
8. Keeping the zero/missing clusters distorts the standard deviation calculation.

9. It is acknowledged that the threshold of 2 standard deviations is not based on theory. Using a false positive flag to remove clusters that are likely a result of parachute jobs is motivated by cluster-driven growth from economies of agglomeration versus employment growth based on other business decisions (like tax incentives) or regional characteristics (like low wages). On empirical grounds, it is common practice to remove outliers that overly influence regression/empirical results. Moreover, the fact that regression by quartiles is common practice in regional econometric work points to the fact that regions in the highest and lowest quartiles/quintiles can have drastically different coefficients among the explanatory variables.
10. Those four natural resource clusters include coal mining, forestry, metal mining, and oil and gas production and transportation. Nonmetal mining was not included as it is composed of many regional industries like sand and gravel mining. Even if one had the view that nonmetal mining should be considered a traded industry, one could argue that this cluster of industries is not as subject to the boom and bust as those industries vulnerable to global price shocks.
11. To double check the positive relationship between productivity growth and traded cluster strength, we removed the local cluster diversity and strength variables. Traded cluster strength remained positively related to GDP-per-worker growth.
12. Personal income data from the Bureau of Economic Analysis, Table 2.1. Personal Income and Its Disposition.
13. The following article highlights the tension in interpretation (http://www.ttnews.com/articles/basetemplate.aspx?storyid=39977&page=1) "Productivity has struggled to develop a sustained pickup since the United States emerged from a recession in 2009, in part because of *sluggish business spending and subdued returns on investment* in new technology such as computers" [Italics added].

References

Arrow, J. (1962). The economic implications of learning by doing. *Review of Economic Studies*, *29*, 155-173.

Bahl, R. W., Firestine, R., & Phares, D. (1971). Industrial diversity in urban areas: Alternative measures and intermetropolitan comparisons. *Economic Geography*, *47*, 414-425.

Delgado, M., Porter, M. E., & Stern, S. (2010). Clusters and entrepreneurship. *Journal of Economic Geography*, *10*, 495-518.

Delgado, M., Porter, M. E., & Stern, S. (2014). Clusters, convergence, and economic performance. *Research Policy*, *43*, 1785-1799.

Delgado, M., Porter, M. E., & Stern, S. (2016). Defining clusters of related industries. *Journal of Economic Geography*, *16*, 1-38.

Feldman, M. P., Francis, J., & Bercovitz, J. (2005). Creating a cluster while building a firm: Entrepreneurs and the formation of industrial clusters. *Regional Studies*, *39*, 129-141.

Feser, E. J., Mix, T. D., White, M., & Poole, K. (2014). *Economic diversity in Appalachia: Statistics, strategies, and guides for action*. Washington, DC: Appalachian Regional Commission.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, *41*, 685-697.

Glaeser, E. L., & Kerr, W. R. (2009). Local industrial conditions and entrepreneurship: How much of the spatial distribution can we explain? *Journal of Economics & Management Strategy*, *18*, 623-663.

Hackbart, M. M., & Anderson, D. A. (1975). On measuring economic diversification. *Land Economics*, *51*, 374-378.

Henderson, V. (1997). Externalities and industrial development. *Journal of Urban Economics*, *42*, 449-470.

Isserman, A. M., & Westervelt, J. (2006). 1.5 Million missing numbers: Overcoming employment suppression in County Business Patterns data. *International Regional Science Review*, *29*, 311-335.

Jacobs, J. (1969). *The economy of cities*. New York, NY: Vintage Books.

Ketels, C. (2013). Cluster policy: A guide to the state of the debate. In P. Meusburger, J. Glückler & M. el Meskioui (Eds.), *Knowledge and the economy* (pp. 249-269). Dordrecht, Netherlands: Springer.

Ketels, C. H., & Memedovic, O. (2008). From clusters to cluster-based economic development. *International Journal of Technological Learning, Innovation and Development*, *1*, 375-392.

Marshall, A. (1966). *Principles of economics*. London, England: Macmillan. (Original work published 1890)

Malizia, E. E., & Ke, S. (1993). The influence of economic diversity on unemployment and stability. *Journal of Regional Science*, *33*, 221-235.

Mishel, L., & Shierholz, H. (2012). *A decade of flat wages: The key barrier to shared prosperity and a rising middle class* (Briefing Paper No. 365). Washington, DC: Economic Policy Institute.

Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography*, *87*, 237-265.

Porter, M. E. (1998). Clusters and competition: New agendas for companies, governments, and institutions. In M. E. Porter (Ed.), *On competition* (pp. 197-299). Boston, MA: Harvard Business School Press.

Porter, M. E. (2000). Location, competition, and economic development: Local clusters in a global economy. *Economic Development Quarterly*, *14*, 15-34.

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

Porter, M. E. (2008). *On competition* (Updated and expanded ed.). Boston, MA: Harvard Business School Press.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, *94*, 1002-1037.

Saxenian, A. (1994). *Regional advantage: Culture and competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Slaper, T. F., Hart, N. R., Hall, T. J., & Thompson, M. F. (2011). The index of innovation: A new tool for regional analysis. *Economic Development Quarterly*, *25*, 36-53.

Stimson, R. J., Stough, R. R., & Roberts, B. H. (2006). *Regional economic development: Analysis and planning strategy*. Berlin, Germany: Springer Science + Business Media.

Stirling, A. (1998). *On the economics and analysis of diversity* (Electronic Working Papers Series No. 28). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.144.8865&rep=rep1&type=pdf

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, *4*, 707-719.

Wasylenko, M. J., & Erickson, R. A. (1978). "On measuring economic diversification": Comment. *Land Economics*, *54*, 106-109.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, *48*, 817-838.

Zheng, P., & Slaper, T. F. (2017). *A shining externality on a hill: Do the competitive advantages of industry clusters attract investment in new plant and equipment from outside the region?* Bloomington: Indiana University Press.

Author Biographies

Timothy F. Slaper leads a research team engaged in industry and workforce analysis, economic impact studies, regional economic analyses, demographic estimates and projections, trade and foreign investment analysis, measuring innovation and economic performance, and researching economic growth.

Karter M. Harmon is a researcher working under Dr. Timothy Slaper. His work focuses on topics such as economic development, industry clusters, and social capital.

Barry M. Rubin is a professor emeritus at the Indiana University School of Public and Environmental Affairs. His research focuses on public and environmental policy issues at urban and regional levels.