Spatial Economic Analysis

ISSN: 1742-1772 (Print) 1742-1780 (Online) Journal homepage: https://www.tandfonline.com/loi/rsea20

Measuring industry co-location across county borders

Zheng Tian, Paul D. Gottlieb & Stephan J. Goetz

To cite this article: Zheng Tian, Paul D. Gottlieb & Stephan J. Goetz (2019): Measuring industry co-location across county borders, Spatial Economic Analysis, DOI: 10.1080/17421772.2020.1673898

To link to this article: https://doi.org/10.1080/17421772.2020.1673898

SPATIAL ECONOMIC ANALYSIS
https://doi.org/10.1080/17421772.2020.1673898

Routledge
Taylor & Francis Group

Regional Studies Association

Check for updates

Measuring industry co-location across county borders

Zheng Tian[a], Paul D. Gottlieb[b] and Stephan J. Goetz[c]

ABSTRACT

The location quotient (LQ) measures regional industry concentration with the advantages of easy calculation and interpretation. However, it is a weak method for identifying industry clusters that consist of related industries geographically concentrated in contiguous counties. This paper proposes a new spatial input–output location quotient (SI-LQ) accounting for both the co-location of related industries and the spatial spillover of concentration into neighbouring counties. A bootstrap method is used to determine the cut-off values of the new measure. The practical advantages of the SI-LQ over the traditional LQ include attenuation of the extreme values of the LQ in less populous and remote counties and the identification of large substantive clusters. The SI-LQ outperforms the LQ in a regression analysis of the effect of industry concentration on total employment growth.

KEYWORDS

location quotients, spatial input–output location quotient (SI-LQ), industry agglomeration, input–output linkages, spatial correlation

JEL R12, R15

HISTORY Received 2 October 2018; in revised form 13 September 2019

INTRODUCTION

Industry cluster development is a widely applied strategy for regional economic development (Feser & Bergman, 2000; Nathan & Overman, 2013; Porter, 1998), and industry agglomeration is an enduring focus of regional and urban economics (Combes & Gobillon, 2015; Rosenthal & Strange, 2001). Both concepts emphasize the geographical concentration of interdependent industries, connected with supply–demand chains, taking advantage of input sharing, knowledge spillovers and labour market pooling (Cainelli & Iacobucci, 2016; Ellison, Glaeser, & Kerr, 2010; Helsley & Strange, 2014). Industry agglomeration in a region may also have spatial spillover effects on neighbouring regions (Greenstone, Hornbeck, & Moretti, 2008), forming a large area, such as Silicon Valley, which consists of several counties in the Bay Area of California. The objective of this paper is to construct a measure of industry agglomeration that takes into account co-location of related industries across adjacent regions, defined here as US counties.

CONTACT

[a] **(Corresponding author)** zzt36@psu.edu
Northeast Regional Center for Rural Development, Department of Agricultural Economics, Sociology and Education, Pennsylvania State University, University Park, PA, USA
[b] pdgott@sebs.rutgers.edu
Department of Agricultural, Food, and Resource Economics, Rutgers University, New Brunswick, NJ, USA.
[c] sjg16@psu.edu
Northeast Regional Center for Rural Development, Department of Agricultural Economics, Sociology and Education, Pennsylvania State University, University Park, PA, USA.

Supplemental data for this article can be accessed at https://doi.org/10.1080/17421772.2020.1673898

In the literature, various agglomeration measures have been proposed. Based on the type of spatial data, existing measures can be categorized into the discrete and continuous types. With areal unit data, such as states, counties or other administrative units, discrete measures include the Gini index, the Theil index and the Ellison–Glaeser index (Ellison & Glaeser, 1997). With spatial point data, and focusing on economic activities in continuous space, continuous measures include the Duranton–Overman index (Duranton & Overman, 2005) and variations of Ripley's *K*-function (Arbia, Espa, Giuliani, & Mazzitelli, 2010; Marcon & Puech, 2003).[1] However, both types of measures have limitations. The discrete measures lose the spatial dimension after summing up the shares of industries over all regions. They also fail to account for industry agglomeration across administrative borders, which is a manifestation of the modifiable areal unit problem (MAUP). The continuous measures require very detailed data on location (often the coordinates) of firms in a region. The high data requirement makes the continuous measures uncommon in practice for regional development practitioners.

This paper focuses on the location quotient (LQ), a commonly used measure of local industry concentration (Cromley & Hanink, 2012; Delgado, Porter, & Stern, 2016; Fracasso & Marzetti, 2018; Glaeser, Kallal, Scheinkman, & Shleifer, 1992). It is a discrete type measure but preserves both industry and spatial dimensions. In the literature, the LQ is used to denote several related concepts, including specialization (Mulligan & Schmidt, 2005), spatial concentration (or clustering) (Billings & Johnson, 2012), and industry agglomeration (O'Donoghue & Gleave, 2004). Although these concepts have often been used as synonyms, distinctions between them are important (Aiginger & Rossi-Hansberg, 2006; Brülhart, 1998). Specialization occurs when a region's industry structure is dominated by a single industry, whether or not the industry is large in an absolute sense. Spatial concentration occurs when one or a few regions have a large share of an industry's national employment. Unless an industry at the national level is very small, spatial concentration in a region normally implies an employment count that is large in absolute terms. Finally, agglomeration emphasizes a substantial presence of several related industries in spatial proximity that enjoy both scale and transactional economies.[2]

Consider now the extent to which the traditional LQ measures any or all of these concepts. Mathematically, the LQ has two alternative forms and related interpretations. It can be interpreted as measuring either relative specialization or relative spatial concentration, depending on how the terms in the numerator and denominator are arranged. A frequently-observed issue with the LQ's relative approach to industries is that small, highly specialized counties emerge with high LQs but very small employment counts (Carroll, Reid, & Smith, 2008). Such places would normally be rejected if the LQ were being used as an indicator of agglomeration. The LQ's bias toward specialization without regard to size also leads to a statistical problem: the distribution of LQs across counties will be highly skewed, with outliers that need to be dealt with, for example, in the context of regression.

A second issue with the LQ based on individual county data is that high-LQ counties may be physically isolated. This occurs because the measure does not account for industry shares in neighbouring spatial units. This is a manifestation of the MAUP for areal unit data: the regions are units of statistical convenience and not true economic regions.

Finally, the traditional LQ based on national industry statistics does not usually include key information on related industries. This is essential if one wishes to capture not only the concepts of specialization and spatial concentration but the broader concept of agglomeration as well.

Recent studies have endeavoured to improve the LQ by including new elements. Of special interest in the present context are those that calculate LQs using groups of industries connected via input–output (I-O) relationships or other characteristics such as specialized labour skills (Delgado et al., 2016). The main approach to embedding I-O relationships is to use national-level I-O tables to create a set of essentially aspatial industry clusters composed of North American Industry Classification System (NAICS)-coded industries. This technique was pioneered by Edward Feser

(Feser & Bergman, 2000; Feser, Renski, & Goldstein, 2008). Researchers who use such national cluster definitions for LQ analyses typically select counties, metropolitan statistical areas (MSAs) or other administrative units of convenience, rather than allowing groups of adjacent counties to be identified by the data (Slaper, Harmon, & Rubin, 2018).

In contrast, studies concerned with the 'county proximity problem' have used individual NAICS-coded industries to test their measures of spatial association. These measures are not, strictly speaking, always LQs. Gibbs and Bernat (1997) and Goetz, Shields, and Wang (2009) use a local Moran's I statistic to identify multi-county groupings of establishments in various NAICS-coded industries. (The second of these studies also uses LQs and I-O data, but never in combination.) Carroll et al. (2008) compare the Getis–Ord correlation statistic with LQs estimated on individual counties. Cromley and Hanink (2012) address the county proximity problem for LQs by proposing a 'focal LQ' or FLQ, which uses spatially weighted county-level data. It is similar to one component of our own composite measure. The FLQ was used in applications of public health data and employment data for two sectors, with counties in North Carolina as the study areas.

Leslie and Kronenfeld (2011) and Cromley, Hanink, and Bentley (2014) develop the co-location quotient (CLQ) for categorical spatial point data. The CLQ is defined as the ratio of the probability of one category locating near to another category in a given location relative to the probability that the two categories co-locate globally. While Leslie and Kronenfeld (2011) examine the co-location of industries in Phoenix, they do not use I-O relationships to select pairs of industries that they would expect to locate together.

In the context of a literature on an enhanced LQ index that remains young and somewhat disconnected, our motivation to modify the LQ is to make it a measure of regional agglomeration that incorporates both industry relatedness and spatial proximity. We show that modifying the LQ to include proximate counties–which is justified by our knowledge of the size of labour markets and economic regions, and modifying the LQ to include industries related through a common supply chain–which is justified by industry cluster theory (Porter, 1998), solve the problems of small isolated counties and extreme values that are associated with the traditional LQ. Finally, we use the new measure in a regression analysis of employment growth to test its validity in terms of common industry cluster predictions.

COMPUTING THE NEW MEASURE

The proposed new measure is based on the traditional LQ. For industry $i = 1, \ldots, M$ and county $j = 1, \ldots, N$, the LQ is defined as:

$$LQ_{ij} = \frac{x_{ij}}{x_j} / \frac{x_{in}}{x_n} \tag{1}$$

where x_{ij} is industry i's employment in county j; x_j is the total employment in county j; and x_{in} and x_n are the national counterparts.[3] LQ_{ij} reflects how industry i is specialized in county j relative to its national level. The new measure to be introduced extends the LQ with three variants to account for industry relatedness and spatial spillovers.

Measuring co-location of related industries in a county

The concentration of a single industry i may induce related industries to co-locate. The relatedness of industries can be represented by the coefficients in an I-O matrix. To measure the concentration of co-located industries, a weighted average of the LQs for industries related to industry i in county j can be used, with the weights being the I-O coefficients. Let $\mathbf{y}_j = (LQ_{1j}, LQ_{2j}, \ldots, LQ_{Mj})'$ be the $M \times 1$ vector of LQs in county j and $\mathbf{A}$ an $M \times M$ I-O coefficient matrix (i.e., the direct requirement matrix). Each row in $\mathbf{A}$, say $\mathbf{a}_i = (a_{i1}, a_{i2}, \ldots, a_{iM})'$, represents the demand for industry i's products as intermediate inputs to other industries. The inner product $\mathbf{a}'_i\mathbf{y}_j$ is a weighted average of the LQs for all industries related to i, denoted as ALQ_{ij}. The ALQ for all M industries in county j can be calculated simply as $\mathbf{A}\mathbf{y}_j$.

However, using $\mathbf{A}$ to calculate the measure of co-location has two problems. First, when the diagonal element of $\mathbf{A}$, a_{ii}, is non-zero, a high LQ_{ij} will produce a high ALQ_{ij} because of $a_{ii}LQ_{ij}$. Although an industry often has large intra-industry purchases, our purpose is to have a measure of co-location of related industries other than of its own industry. Also, it is not desirable to double count LQ_{ij} when combining ALQ_{ij} with LQ_{ij} in the final composite measure. Therefore, in calculating ALQ_{ij}, the diagonal elements in $\mathbf{A}$ are set to zero and each row is standardized by dividing all the elements by the row sum.

The second problem with $\mathbf{A}$ is zero elements. On the one hand, zeros may be a desired feature in computing ALQ_{ij} because the LQs of unrelated industries with industry i are excluded. On the other hand, the products of industry i may be inputs for industry k, the products of which are in turn inputs for industry l. Zero elements in $\mathbf{A}$ cannot reflect this second-round relatedness among industries, but the elements in $\mathbf{A}^2$ can (Miller & Blair, 2009). Moreover, artificially setting all diagonal elements of $\mathbf{A}$ to zero eliminates the intra-industry I-O relationship, which is important for some industries.[4] But the intra-industry relationship can be preserved in $\mathbf{A}^2$ that is computed before making its diagonal zero. Therefore, for the purpose of including the multi-stage I-O relationships and maintaining a sufficient number of zeros to exclude unrelated industries, a composite I-O matrix, $\mathbf{A} + \mathbf{A}^2$, is used in computing ALQ_{ij}. The composite matrix is then modified to have zero diagonal elements and standardized rows. As an example, the 2015 I-O matrix, except for the diagonal elements, has 460 zeros out of 4970 elements, and this number in the composite matrix becomes 94.

Measuring co-location across county borders

Industry clusters may spread across county borders, leading to spatial association of concentration. The strength of industry i's concentration in a county will be reinforced if the same industry and its related industries concentrate in surrounding counties. A straightforward way to account for such spatial concentration is to compute spatial lags of the LQ and ALQ measures, which are the weighted average of these variables in neighbouring counties, with the weights defined in a spatial weight matrix (LeSage & Pace, 2009). Suppose $\mathbf{W}$ is an $N \times N$ spatial weight matrix with zero diagonal elements and standardized rows. We use the queen-type contiguity to define neighbours. Let $\mathbf{y}_i$ and $\tilde{y}_i$ be the $N \times 1$ vectors of LQ_{ij} and ALQ_{ij} for industry i in all counties. The spatial lags of the LQ and ALQ are $\mathbf{W}\mathbf{y}_i$ and $\mathbf{W}\tilde{y}_i$, denoted as WLQ_{ij} and $WALQ_{ij}$, respectively. Spatial lags that reflect spatial spillover effects were also used in Cromley and Hanink (2012). Instead of computing the spatial lags of the LQ directly, they apply the spatial weight matrix to county-level employment, that is, x_{ij} and x_j in equation (1), and compute the FLQ with spatially weighted data.

Incorporating industry relatedness and spatial spillovers

We compute three variants of the LQ: (1) ALQ_{ij} for industry i's related industries in county j; (2) WLQ_{ij} for industry i's concentration in neighbouring counties to form a large spatial concentration; and (3) $WALQ_{ij}$ for industry i's related industry in neighbouring counties. While these variants can individually represent a facet of industry agglomeration, we need a composite measure that takes into account both industry relatedness and spatial association.

Based on the 10-step guide of computing composite indicators of the Organisation for Economic Co-operation and Development (OECD) (2008), we compute the spatial input–output location quotient (SI-LQ), as the geometric average of all LQ variants, that is:

$$SI - LQ_{ij} = [LQ_{ij}(ALQ_{ij} + \xi)(WLQ_{ij} + \xi)(WALQ_{ij} + \xi)]^{1/4} \quad (2)$$

where ξ is a small positive number to prevent the product from being zero if LQ_{ij} is non-zero. However, we allow $SI - LQ_{ij} = 0$ if $LQ_{ij} = 0$, keeping the two measures consistent in cases where industry i does not exist in county j.[5] Computing the SI-LQ as a geometric average effectively attenuates an extremely high LQ in a remote and small county with a simple industry structure. The ALQ, WLQ, and WALQ measures would all be relatively low in such a situation, resulting in a low SI-LQ. However, the SI-LQ would be high if all of its components were substantially high.

There are some caveats with the SI-LQ. First, the SI-LQ computed here defines industry relatedness in terms of the I-O linkages, omitting other important factors, such as worker skills or technology (Delgado et al., 2016), that connect industries. The I-O matrix in the ALQ can be replaced by any matrix quantifying other types of industry relatedness. That we use the I-O matrix to link various industries is based on the existing literature. The literature on the intermediate demand variables (IDVs) uses I-O matrices to compute IDVs for inter-industry dependence (Moghadam & Ballard, 1988; Rey & Jackson, 1999). The literature on integrated I-O and econometric models (Rey, 2000) advocates the use of both I-O matrices and spatial weight matrices to account for the inter-industry and spatial dependence in multiregional contexts. Second, we use the national I-O tables simply because they are readily obtained. At the regional level, local I-O tables can be used, but constructing such tables is beyond the scope of our work. Third, the choice of exponents in equation (2) raises the issue of parameterization, but it also provides flexibility. While the four components are given equal weights (exponents), users can assign each component a different weight to reflect their judgment of importance. (This option is discussed further in the sixth section.) Finally, while the construction of the SI-LQ lacks a theoretical foundation, such as the dartboard framework in Ellison and Glaeser (1997) and Guimarães, Figueiredo, and Woodward (2009), our main concern is practical applications of the SI-LQ for identifying industry agglomeration. We evaluate our new metric largely on its performance, and not on its adherence to statistical or behavioural theory.

DETERMINING CUT-OFF VALUES OF THE SI-LQ

While $LQ_{ij} > 1$ indicates industry i is concentrated in county j, various arbitrary cut-off values of the LQ have been used in previous studies, ranging from 1 to 5 (Crawley, Beynon, & Munday, 2013). To avoid arbitrary cut-offs, recent studies have attempted to construct confidence intervals (Moineddin, Beyene, & Boyle, 2003), derive test statistics based on the Ellison–Glaeser dartboard framework (Ellison & Glaeser, 1997; Guimarães et al., 2009), and model the LQ with the Poisson and binomial probabilistic processes (Billings & Johnson, 2012). Although it is desirable to apply these new methods in this paper, the non-linear computation of the SI-LQ makes a direct application difficult. Also, these new methods impose strong assumptions on the LQ, which may not hold for the SI-LQ.

To circumvent these difficulties, we employ a parametric bootstrap method to determine cut-off values for the SI-LQ. Assuming that the standardized LQ (SLQ, i.e., the z-scores of LQ) follows the standard normal distribution, O'Donoghue and Gleave (2004) use 1.63 and 1.96, the 90% and 95% critical values, as cut-off values for the SLQ. However, because the assumption of the normal distribution is not satisfied for many industries, the applicability of this method is limited. Tian (2013) follows the SLQ method but uses a simple non-parametric bootstrap method to simulate cut-off values of the SLQ from their empirical distributions. While the simple bootstrap method works well for independent samples, it becomes unreliable for temporally or spatially correlated data. The SI-LQ must be subject to spatial correlation because its computation involves spatial lags. Several bootstrap methods are available for spatially correlated data, including block bootstrap, wild bootstrap, pairs bootstrap, and parametric bootstrap (Klarl, 2014; Plant, 2012). We choose a parametric bootstrap method following Burridge and Fingleton (2010) and Plant (2012).

The parametric bootstrap method consists of the following procedure:

(1) Estimate a spatial error model (SEM). For each industry, let an $N \times 1$ vector $\mathbf{y}$ represent the SI-LQ of the industry; then the SEM is:

$$\mathbf{y} = \alpha \mathbf{1} + \mathbf{u}, \mathbf{u} = \rho \mathbf{W} \mathbf{u} + \epsilon, \epsilon \quad NID(0, \sigma^2 \mathbf{I})$$

where α is a constant; $\mathbf{1}$ is a vector of ones; $\mathbf{u}$ is a residual vector assumed to be spatially correlated, with a spatial autoregressive coefficient ρ and a spatial weight matrix $\mathbf{W}$; and ϵ is the disturbance vector assumed to be independent and identically distributed. It follows that $\mathbf{y}$ can be generated by:

$$\mathbf{y} = \alpha \mathbf{1} + (\mathbf{I} - \rho \mathbf{W})^{-1} \epsilon. \tag{3}$$

(2) On obtaining $\hat{\rho}$, $\hat{\alpha}$ and $\hat{\epsilon}$, resample the residuals $\hat{\epsilon}$ with replacement, that is, bootstrapping. We assume that the counties with non-zero SI-LQs and those with zero SI-LQs are heterogeneous in the error terms. Thus, the residuals belonging to the two groups are resampled separately.
(3) Compute the fitted values of the SI-LQ with the bootstrapped residuals according to equation (3). For each industry, 999 sets of bootstrapped residuals and fitted values are obtained.
(4) Compute the averages of the 90th and 95th percentiles from the bootstrapped SI-LQ as the estimates of the true 90% and 95% critical values of the population distribution, which are the cut-off values to detect industry agglomeration.

DATA SOURCE

We have two data sources for county-level industry employment. Except for Farm (NAICS 111 and 112) and Public administration (NAICS 92), we use the data from the County Business Patterns (CBP) provided by the W.E. Upjohn Institute to compute LQs. Based on the original CBP of the US Census Bureau, in which many county-level industry employment records are suppressed due to privacy protection, the data set from the institute uses the imputation method proposed by Isserman and Westervelt (2006) to estimate missing records. For Farm and Public administration, we use county-level employment data from the Regional Economic Accounts of the Bureau of Economic Analysis (BEA). The LQs for these two sectors are computed with the BEA data and then merged with the LQs computed from the CBP data for other industries.

Merging the CBP and BEA data also requires addressing the problem of independent cities in Virginia. Since the BEA uses the combined areas for independent cities and their counties, the county-level employment data in the corresponding spatial units from the CBP are aggregated according to the BEA's region definitions. Finally, we obtain the data set of LQs for 71 industries, of which most have three-digit NAICS, and four have two-digit NAICS (all NAICS and industry names are listed in Table 2).[6] The number of counties ranges from 3077 to 3080 due to the change in FIPS and missing records for a few counties from 1998 to 2015.

The aggregated I-O matrix is calculated with the use, make and final demand matrices from 1998 to 2015, obtained from the Employment Projections (EP) programme of the Bureau of

Labor Statistics. The original use and make matrices have 200 industries that are aggregated to the desired industry classification, and then the direct requirement matrix (i.e., the matrix **A**) is computed based on the instruction provided by the BEA (Horowitz & Planting, 2009), resulting in matrices with 71 industries in each year.[7]

The spatial weight matrix is generated with the generalized (1:500,000) counties shapefile from the TIGER products of the US Census Bureau, downloaded and converted to spatial polygon objects with the tigris package of R. The spatial polygons for independent cities are merged with counties. The spatial weight matrix is then computed as a queen-type contiguity-based matrix with row-standardization.[8]

RESULTS OF COMPUTING THE NEW MEASURE

Descriptive analyses

We computed the LQ, ALQ, WLQ, WALQ and SI-LQ measures for all the counties in the 48 continental states in the United States from 1998 to 2015. Table 1 shows all the descriptive statistics, computed by pooling all industries and using the observations that have non-zero SI-LQs in 2015. An improvement of the SI-LQ over the LQ is that the distribution of the former is more centred on the mean with a narrower range. While the minimums are close to zero for both the LQ and SI-LQ, the maximum SI-LQ is much smaller. This is the effect of geometric averaging over four measures so that the extreme values in LQs are controlled. As a result, the skewness of the SI-LQ is only one-third of the LQ. For each industry, we have similar attenuation in maximums and skewness, with the average ratio of the skewness between the SI-LQ and LQ being only 29%.[9]

A direct check of the outcome of the SI-LQ alleviating the extreme-value problem of the LQ is to compare the location of industries with maximum LQs and SI-LQs. The improvement in this aspect is noticeable, as presented in Table 2 and Figure 1, which show the locations and maximums of the LQ and SI-LQ that are averaged with the five-year data from 2011 to 2015. In the LQ column in Table 2, New York County, NY, does not show up for any industry, despite its prominence as a cluster of a variety of industries. With the SI-LQ, it has the highest concentration in Broadcasting (NAICS 515). Similar cases include Los Angeles County, CA, for Motion picture and sound recording (NAICS 512) and San Francisco County, CA, for Other information services (NAICS 519). Although some small counties still have the maximum SI-LQs, the average population size of counties with maximum SI-LQs is 212,374, much higher than 23,421 based on the maximum LQs. In Figure 1, many isolated counties that have maximum LQs do not show up on the map of maximum SI-LQs. For example, the map for the SI-LQ shows that most manufacturing industries still concentrate in the region spanning from the Rust Belt region down to Alabama, rather than being sporadically distributed in less populated counties in the northern Great Plains region, as indicated by the maximum LQs.

We can map each industry's spatial concentration using categorized SI-LQs arising from the bootstrapped percentiles. Based on the 25th–95th percentiles (see Table S5 in the supplemental data online), we categorize the SI-LQ into six groups, and define the 95–100% group to be the most concentrated regions and the 90–95% group to be highly concentrated regions. Consider Transportation equipment manufacturing (NAICS 336) as an example. For comparison, we also compute the bootstrapped percentiles for the LQ and categorize it. Panels (a) and (b) in Figure 2 compare the spatial distribution of the LQ and SI-LQ for NAICS 336 in 2010, respectively. An obvious difference is that with the SI-LQ, the counties in Michigan, Ohio and Indiana form a large area with a very high level of concentration, and this industry also concentrates in the southern states, including Kentucky, Tennessee, Alabama and South Carolina. In contrast, with the LQ, many counties identified as the most concentrated are relatively isolated, especially in the Midwest region, including the states from North Dakota and Minnesota down to Texas and Louisiana. Panel (c) shows the change of the concentration level from 2005 to 2015 based on

Table 1. Summary statistics of the spatial input–output location quotient (SI-LQ) and its components for all industries.

								Quantile				
	Total observations	**Non-zero observations**	**Mean**	**SD**	**Skewness**	**Minimum**	**Maximum**	**25%**	**50%**	**75%**	**90%**	**95%**
SI-LQ	218,609	151,170	1.09	1.26	8.86	0.00	55.57	0.58	0.86	1.16	1.77	2.69
LQ	218,609	151,170	2.19	9.63	26.99	0.00	810.64	0.33	0.79	1.48	3.37	6.84
ALQ	218,609	218,609	1.22	2.22	30.97	0.04	257.41	0.66	0.86	1.19	1.91	2.81
WLQ	218,609	202,167	1.62	4.60	14.41	0.00	257.44	0.33	0.77	1.39	2.84	5.24
WALQ	218,609	218,396	1.21	1.30	13.29	0.11	65.85	0.76	0.93	1.25	1.88	2.57

Note: The table is based on computation results in 2015. Statistics are computed across all industries. The number of counties is 3079; the number of industries is 71.

Table 2. Maximum LQs and SI-LQs averaged from 2011 to 2015.

NAICS	Industry	Counties with maximum LQs: Counties	Maximum LQs	Population	Counties with maximum SI-LQs: Counties	Maximum SI-LQs	Population
111–112	Crop and animal production	Sioux County, NE	38.73	1313	Sherman County, TX	11.06	3028
113	Forestry and logging	Echols County, GA	678.18	4030	Randolph County, GA	49.18	7671
114	Fishing, hunting and trapping	Patrick County, VA	737.59	18,479	St. Bernard Parish, LA	9.66	36,813
115	Support activities for agriculture and forestry	Owyhee County, ID	361.74	11,473	Motley County, TX	12.53	1206
211	Oil and gas extraction	Kent County, TX	309.13	809	Winkler County, TX	19.43	7079
212	Mining (except oil and gas)	Eureka County, NV	488.45	1993	Mercer County, ND	21.16	8426
213	Support activities for mining	Slope County, ND	158.41	728	Upton County, TX	17.34	3344
221	Utilities	Surry County, VA	124.92	7060	Mercer County, ND	20.33	8426
23	Construction	Clay County, GA	10.09	3161	Live Oak County, TX	3.58	11,547
311	Food manufacturing	Jerauld County, SD	51.84	2085	Parmer County, TX	10.21	10,274
312	Beverage and tobacco product manufacturing	Moore County, TN	347.74	6340	Anderson County, KY	6.52	21,448
313–314	Textile mills and textile product mills	Murray County, GA	255.77	39,541	Chattooga County, GA	16.34	25,956
315–316	Apparel, leather and allied product manufacturing	Shannon County, MO	146.77	8446	Cherokee County, AL	14.34	25,977
321	Wood product manufacturing	Webster County, GA	141.93	2777	Cleveland County, AR	31.91	8686
322	Paper manufacturing	Little River County, AR	129.19	13,137	Choctaw County, AL	12.98	13,840
323	Printing and related support activities	Linn County, MO	50.19	12,745	Crawford County, MO	4.00	24,617
324	Petroleum and coal products manufacturing	Crawford County, IL	126.67	19,810	Hutchinson County, TX	35.60	22,203
325	Chemical manufacturing	Esmeralda County, NV	60.61	782	St. James Parish, LA	7.41	22,006
326	Plastics and rubber products manufacturing	Bracken County, KY	64.84	8503	Pleasants County, WV	7.69	7583
327	Non-metallic mineral product manufacturing	Hancock County, GA	82.40	9419	Pershing County, NV	12.49	6744
331	Primary metal manufacturing	Hancock County, KY	157.46	8551	Monroe County, OH	10.28	14,579

(*Continued*)

Table 2. Continued.

		Counties with maximum LQs			Counties with maximum SI-LQs		
NAICS	Industry	Counties	Maximum LQs	Population	Counties	Maximum SI-LQs	Population
332	Fabricated metal product manufacturing	Cameron County, PA	39.12	5074	Hancock County, KY	6.29	8551
333	Machinery manufacturing	Sargent County, ND	78.71	3804	Shelby County, OH	5.52	49,311
334	Computer and electronic product manufacturing	Deuel County, SD	53.09	4349	Henry County, KY	4.65	15,380
335	Electrical equipment, appliance, and component manufacturing	Bland County, VA	116.41	6811	Fountain County, IN	7.03	17,272
336	Transportation equipment manufacturing	Edwards County, IL	46.03	6728	Perry County, IN	6.15	19,415
337	Furniture and related product manufacturing	Pontotoc County, MS	188.30	30,043	Chickasaw County, MS	14.51	17,413
339	Miscellaneous manufacturing	Jackson County, KY	62.73	13,497	Kosciusko County, IN	5.50	77,336
42	Wholesale trade	Brown County, IL	10.71	6913	St. Charles Parish, LA	1.66	52,844
441	Motor vehicle and parts dealers	Sully County, SD	8.92	1380	Pulaski County, IN	1.74	13,357
442–454	All other retail	McPherson County, NE	7.73	536	Dare County, NC	1.81	33,986
445	Food and beverage stores	Craig County, VA	9.12	5175	Petroleum County, MT	2.29	493
452	General merchandise stores	Hickman County, KY	6.41	4868	Chattooga County, GA	2.45	25,956
481	Air transportation	Rappahannock County, VA	36.98	7500	Cameron Parish, LA	4.74	6901
483	Water transportation	Lafourche Parish, LA	221.77	96,686	St. Bernard Parish, LA	22.08	36,813
484	Truck transportation	Benton County, MS	32.38	8696	Roberts County, TX	10.62	924
485	Transit and ground passenger transportation	Perry County, PA	37.38	45,999	Warren County, PA	3.93	41,755
486	Pipeline transportation	McMullen County, TX	355.72	713	Reagan County, TX	24.82	3348
487–488	Scenic and sightseeing transportation and support activities	Dale County, AL	55.23	50,360	St. James Parish, LA	6.85	22,006
492	Couriers and messengers	Jackson County, OK	24.83	26,467	St. Charles Parish, LA	2.29	52,844
493	Warehousing and storage	Lucas County, IA	56.03	8891	Union County, SC	3.01	28,905

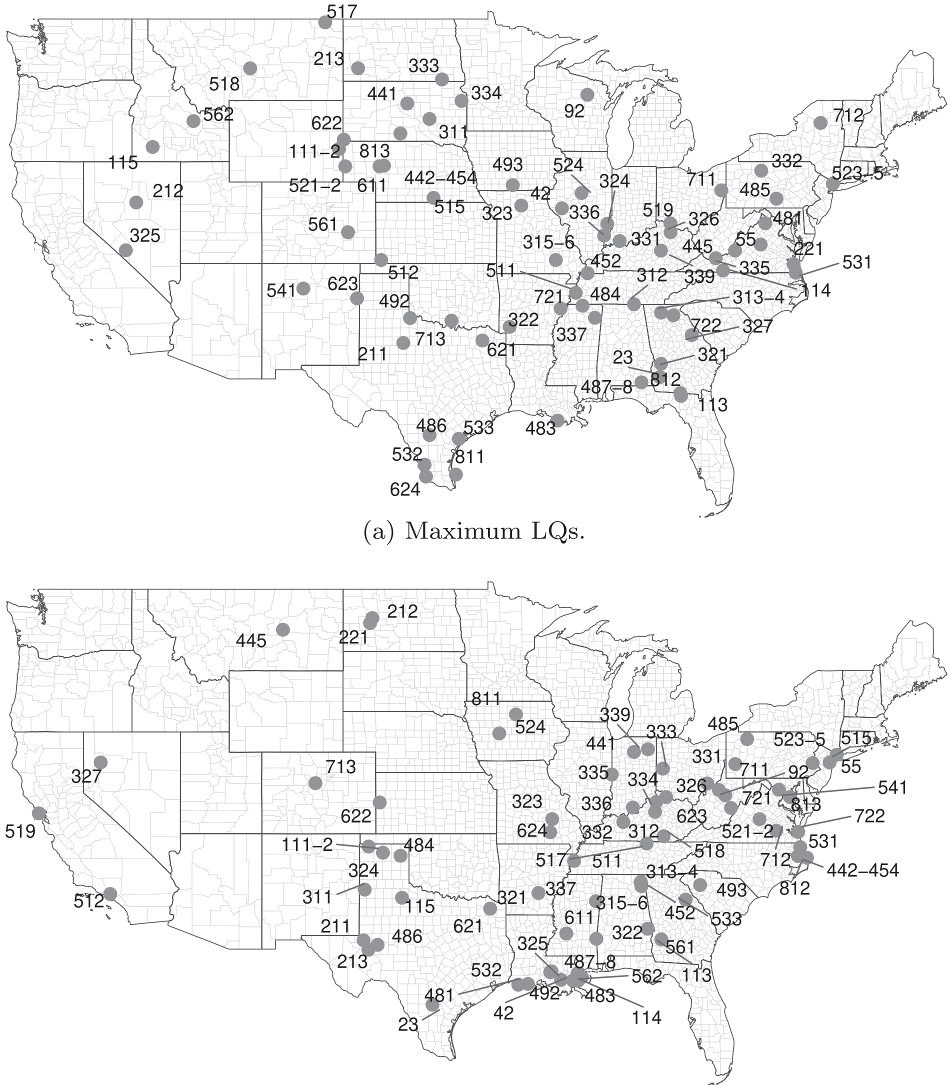
511	Publishing industries (except internet)	Mississippi County, AR	26.93	46,390	Lauderdale County, TN	2.61	27,742
512	Motion picture and sound recording industries	Morton County, KS	23.89	3237	Los Angeles County, CA	1.81	9,825,473
515	Broadcasting (except internet)	Kearney County, NE	14.87	6484	New York County, NY	1.82	1,588,530
517	Telecommunications	Daniels County, MT	33.37	1745	Jackson County, TN	1.78	11,596
518	Data processing, hosting, and related services	Wheatland County, MT	43.04	2160	Laurel County, KY	2.90	58,990
519	Other information services	Pendleton County, KY	16.53	14,902	San Francisco County, CA	4.03	805,766
521–522	Monetary authorities, credit intermediation, and related activities	Banner County, NE	18.97	696	Goochland County, VA	1.80	21,741
523–525	Securities, commodity contracts, fund, trusts and other related activities	Hudson County, NJ	12.25	636,254	Hudson County, NJ	2.71	636,254
524	Insurance carriers and related activities	McLean County, IL	13.59	169,857	Dallas County, IA	1.65	66,699
531	Real estate	Currituck County, NC	9.29	23,661	Currituck County, NC	2.26	23,661
532	Rental and leasing services	Jim Hogg County, TX	22.47	5286	Cameron Parish, LA	3.44	6901
533	Lessors of non-financial intangible assets (except copyrighted works)	Refugio County, TX	66.08	7360	Oconee County, GA	2.04	32,929
541	Professional, scientific, and technical services	Los Alamos County, NM	10.38	18,012	Arlington County, VA	1.94	209,449
55	Management of companies and enterprises	Goochland County, VA	8.50	21,741	Somerset County, NJ	1.84	324,158
561	Administrative and support services	Crowley County, CO	6.39	5852	Chattahoochee County, GA	1.45	11,178
562	Waste management and remediation services	Butte County, ID	159.84	2907	St. Bernard Parish, LA	4.28	36,813
611	Educational services	Arthur County, NE	20.95	464	Warren County, MS	2.02	48,816
621	Ambulatory health care services	Delta County, TX	10.94	5238	Franklin County, TX	1.51	10,597
622	Hospitals	Fall River County, SD	10.57	7111	Greeley County, KS	2.13	1259

(*Continued*)

Table 2. Continued.

		Counties with maximum LQs			Counties with maximum SI-LQs		
NAICS	**Industry**	**Counties**	**Maximum LQs**	**Population**	**Counties**	**Maximum SI-LQs**	**Population**
623	Nursing and residential care facilities	Oldham County, TX	21.25	2051	Robertson County, KY	2.66	2273
624	Social assistance	Starr County, TX	15.66	61,170	Reynolds County, MO	2.28	6679
711	Performing arts, spectator sports, and related industries	Hancock County, WV	20.13	30,659	Washington County, PA	1.72	207,900
712	Museums, historical sites, and similar institutions	Hamilton County, NY	48.91	4834	Charles City County, VA	3.49	7271
713	Amusement, gambling, and recreation industries	Love County, OK	45.89	9430	Clear Creek County, CO	4.55	9081
721	Accommodation	Tunica County, MS	47.12	10,748	Bath County, VA	3.56	4713
722	Food services and drinking places	Banks County, GA	3.19	18,408	Northampton County, VA	2.72	12,388
811	Repair and maintenance	Kenedy County, TX	52.49	415	Hancock County, IA	2.42	11,307
812	Personal and laundry services	Brooks County, GA	8.21	16,263	Dare County, NC	1.64	33,986
813	Religious, grant-making, civic, professional, and similar organizations	Todd County, SD	12.50	9644	Arlington County, VA	2.14	209,449
92	Public administration	Menominee County, WI	6.74	4265	Gilmer County, WV	3.26	8724

Note: LQs and SI-LQs are authors' calculation. Data on the county-level population in 2015 are from the US Census Bureau.



(a) Maximum LQs.

(b) Maximum SI-LQs.

Figure 1. Distribution of maximum location quotients (LQs) and spatial input–output location quotients (SI-LQs).
Note: Authors' calculation. Geographical data are downloaded with the tigris package in R, which are the same for the following maps. The log-transformed values are shown.

the SI-LQ. The counties in the darker colour are where the SI-LQ increases to a higher quantile group from 2005 to 2015, and vice versa. We can observe a sharp contrast between the Rust Belt region, such as Indiana, and the southern region, such as Alabama, which has been a preferred location for foreign direct investment in this industry.[10]

Regression analysis

We use a regression application to compare the performance of the SI-LQ and the LQ when they are used as regressors. We continue to use Transportation equipment manufacturing (NAICS 336) as an example to examine the effect of industry agglomeration on total employment

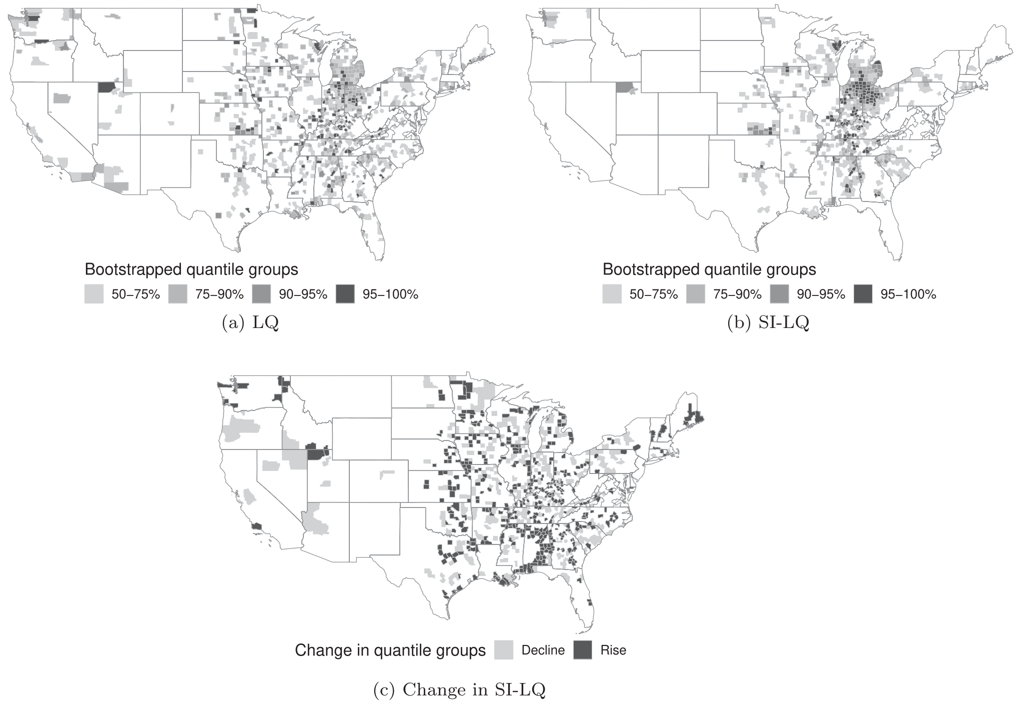


Figure 2. Spatial distribution of the concentration of transportation equipment manufacturing.
Note: Authors' calculation. The LQ and SI-LQ are grouped with the bootstrapped quantiles. The rise and decline in (c) represents the up and down in the quantile group by at least one group.

growth.[11] This industry depends heavily on I-O linkages, and large plants often locate in nearby counties to produce different parts and types of transportation equipment. Therefore, this industry is a case where it appears necessary to consider both inter-industry and inter-regional aspects of concentration when exploring outcomes related to economic vitality.

The regression model is a spatial panel model with county fixed effects.[12] The dependent variable is the growth rate of total wage employment for two time periods, 2005–10 and 2010–15, which are roughly before and after the trough of the Great Recession. Transportation equipment manufacturing in the United States was severely hit by the Great Recession. The areas where this industry was concentrated before the recession experienced substantial job losses during the downturn. In the recovery period, we would expect industry concentration to be an asset, with faster employment growth in high concentration counties. Therefore, we expect a positive coefficient on either the LQ or SI-LQ if they can correctly capture industry concentration. However, since the SI-LQ takes into account the concentration of related industries and spatial concentration in nearby counties, we hypothesize that the coefficient on SI-LQ will be higher than that of the LQ. Moreover, the problem of extreme LQs in remote areas without substantive concentration could further dampen the positive effect of the LQ on employment growth.

According to Elhorst (2014), a spatial panel model with county fixed effects can be specified as:

$$y_{it} = \lambda \sum_{j=1}^{n} w_{ij} y_{jt} + \mathbf{x}_{it}\beta + \mu_i + \varepsilon_{it},\ \varepsilon_{it} \sim N(0, \sigma^2) \tag{4}$$

where y_{it} represent the total wage employment growth rate of county i at time t, $t = \{2005 - 2010, 2010 - 2015\}$; λ is the spatial autoregressive coefficient; and w_{ij} is the (i, j)

element in a spatial weight matrix $\mathbf{W}$, which is the same queen-type contiguity matrix used in computing the WLQ and WALQ. To maintain model parsimony, we use μ_i to account for all time-invariant county-specific variables, and include two regressors in $\mathbf{x}_{it}$. The main regressor is the categorized LQ or SI-LQ that takes on integer values from 1 to 6, representing from the least to the most concentrated counties based on the bootstrapped percentiles. The main reason for using the categorical variables is to avoid the leverage effect from the extreme value of the LQ.[13] To mitigate potential endogeneity, the LQ and SI-LQ variables take the values at the beginning years of the two time periods, that is, 2005 and 2010. The other regressor in the model is the level of total wage employment in 2005 and 2010, to account for the initial size of local labour markets and to test for convergence across counties. The data on total wage employment are from the BEA.

The results are consistent with our expectation. Table 3 shows the estimated coefficients of the spatial panel models and the direct, indirect and total effects of regressors, which are defined according to LeSage and Pace (2009) to quantify the effect of a regressor on the dependent variable in local spatial units, neighbouring units and their sum, respectively. The contrast of the coefficients on the LQ and SI-LQ is straightforward. Using the SI-LQ, the effect of the concentration of Transportation equipment manufacturing on total wage employment growth is positive and statistically significant, but using the LQ, this effect is negative and insignificant.[14] The estimated models indicate a strong spatial spillover effect of total wage employment growth, with a positive and significant spatial autoregressive coefficient. They also suggest convergence in employment across the U.S. counties before and after the period of the recession, with a negative and significant coefficient on the initial level of total employment.

Table 3. Employment regressions with the LQ and SI-LQ for NAICS 336 as regressors.

	Regressors	Model with LQ	Model with SI-LQ
Coefficients	LQ quantile groups	−0.174 (0.272)	
	SI-LQ quantile groups		1.124 (0.325)***
	Initial level of wage employment	−0.257 (0.017)***	−0.257 (0.017)***
Direct effect	LQ quantile groups	−0.180 (0.270)	
	SI-LQ quantile groups		1.159 (0.345)***
	Initial level of wage employment	−0.265 (0.018)***	−0.265 (0.018)***
Indirect effect	LQ quantile groups	−0.106 (0.159)	
	SI-LQ quantile groups		0.674 (0.204)***
	Initial level of wage employment	−0.155 (0.015)***	−0.154 (0.014)***
Total effect	LQ quantile groups	−0.286 (0.428)	
	SI-LQ quantile groups		1.833 (0.546)***
	Initial level of wage employment	−0.420 (0.031)***	−0.419 (0.029)***
	λ	0.388 (0.017)***	0.387 (0.017)***
	Log likelihood	−20,666.887	−20,661.111
	Num. observations	3075 counties, two time periods	

Notes: The dependent variable in the two regressions is the growth rate of total wage employment in counties for two time periods, 2005–10 and 2010–15.
All the regressors take the value at the first year of the two periods, that is, 2005 and 2010, to mitigate the endogeneity problem.
The categorical variables, LQ and SI-LQ groups, are defined with the bootstrap quantile of the two variables.
The direct, indirect and total effects of regressors are defined according to LeSage and Pace (2009) to quantify the effect of a regressor on the dependent variable in the local spatial unit, neighbouring units and their sum, respectively.
Significance levels: *10%, **5%, ***1%.

In sum, both descriptive and regression analyses show that the SI-LQ outperforms the LQ as a measure of industry agglomeration. Next, we explore why and how the SI-LQ improves on the LQ, and give it an economic interpretation that highlights differences across industries in co-location patterns based on the supply chain.

ANALYSIS OF THE SI-LQ COMPONENTS

We examine the relationship among the components of the SI-LQ to understand the way in which the SI-LQ smooths the LQ. First, we compute the Spearman correlation coefficients of the LQ against the SI-LQ and the other components, which is presented in Figure 3. The Spearman correlation coefficient is used because it is the ranking, not the values of these variables, that

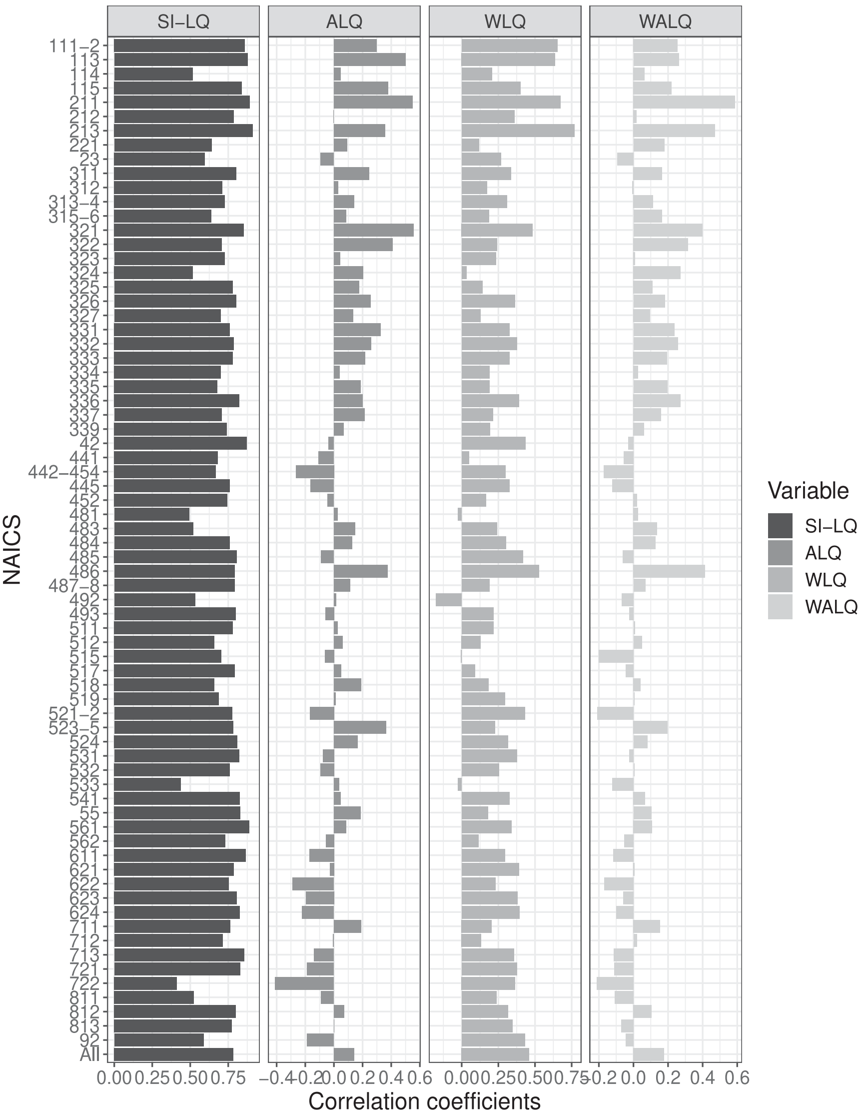


Figure 3. Spearman correlation of the components of SI-LQ with LQ.
Note: Authors' calculation.

matters for applied work.[15] The correlation coefficients between LQs and SI-LQs are higher than 0.5 for most industries (68 out of 71 industries), but only one industry (NAICS 213) has a Spearman coefficient higher than 0.9. This indicates that both variables provide the same general picture of industry concentration nationwide, but the SI-LQ changes the precise ordering of the counties due to its inclusion of more facets of agglomeration. The LQ and WLQ are positively correlated for all but three industries (NAICS 481, 492 and 533), indicating the existence of spatial spillover effects.

The key factor that distinguishes the SI-LQ from the LQ is the ALQ. The average of the Spearman correlation coefficients across all industries is only 0.07, the maximum is 0.56 and 22 industries have negative correlation coefficients. The correlation coefficients between the LQ and WALQ are very similar to those of the LQ-ALQ pair.

An interesting finding is that service, retail and wholesale industries tend to have negative correlation between the LQ and ALQ (similarly, WALQ), while manufacturing industries mostly have positive correlations. This is reasonable because manufacturing industries usually have more I-O linkages with local input suppliers than service industries. Indeed, when using the sum of all the elements in a row in the I-O matrix, **A**, to represent the share of intermediate demand in an industry's output, we find that with the I-O matrix in 2015, the average share of intermediate demand of manufacturing industries is 61%, 20% higher than that of service and other sectors. Therefore, we hypothesize that the higher the share of intermediate demand is, the more correlated are the LQ and ALQ. To test this hypothesis, we regress the Spearman (ranking) correlation coefficients between the LQ and the ALQ of all 71 industries in this study on the vector of the row sums of **A**. We run two types of regressions. The first is a linear regression of the Spearman correlation coefficients on the shares of intermediate demand, and the second is a probit regression with the dependent variable being an indicator variable that equals one if the correlation coefficient is positive and zero otherwise. Both models confirm the hypothesis with highly significant positive coefficients. (For the results, see in the upper panel in Table 4.)

To check the robustness of the effect of the share of intermediate demand on the LQ–ALQ relationship, we use quantile regression to evaluate the LQ–ALQ association along the 25th, 50th,

Table 4. Regression of Spearman correlation coefficients and quantile regression coefficients on the shares of intermediate demand.

Dependent variable: Spearman coefficient between LQ and ALQ by industry		
	Linear model	**Probit model**
	0.451 (0.159)	3.483 (1.218)
Dependent variable: quantile regression coefficients of regressing LQ on ALQ		
LQ percentiles		
25%	1.100 (0.397)	3.135 (1.162)
50%	1.201 (0.387)	3.596 (1.209)
75%	1.289 (0.379)	4.010 (1.253)
90%	1.283 (0.380)	4.005 (1.252)
95%	1.390 (0.383)	4.098 (1.235)

Notes: Dependent variables are the Spearman correlation coefficient in the upper panel and the quantile regression coefficients in regression models of the LQ on the ALQ in the lower panel. The independent variable is the share of intermediate demand of each industry.
Standard errors are parenthesized.
All coefficients are significant at the 5% level.
Number of observations is 71 industries.

75th, 90th and 95th percentiles of the LQ for each industry. Quantile regression is robust for a random variable with a heavy-tail distribution, like the LQ (Koenker & Hallock, 2001). (The results of the quantile regression are shown in the supplemental data online.) Upon obtaining the five quantile regression coefficients by industry, we regress them on the share of intermediate demand separately. Again, we use both linear and probit regressions. The results of these regressions are shown in the lower panel of Table 4. Not only do the coefficients remain positive but they also increase with the percentiles, implying that the LQ–ALQ association depends more on the I-O linkages when industries are more concentrated. This result is in line with the findings of Ellison et al. (2010) who use the maximum of the row and column elements of one industry in the I-O matrix to quantify the I-O relationship, and find a positive effect on the EG co-agglomeration index that measures co-location of two industries.

The follows are implications of the relationship between the LQ and the other components for constructing the SI-LQ. From a perspective of designing a composite indicator, the component measures should not be highly correlated. Otherwise, it would double count a particular aspect represented by more than one components (OECD, 2008). In the case of the SI-LQ, the WLQ is a double counting factor for most industries, but its mild correlation with the LQ to some extent offsets this drawback. On the other hand, the ALQ and WALQ are compensatory factors to the LQ for industries with negative correlation coefficients, which is useful to alleviate the extreme-value problem of the LQ. From a conceptual perspective of understanding regional industry concentration, the varying correlation coefficients reflect the complex location patterns of related industries. The idea of co-agglomeration may work well for industries with strong I-O linkages, such as manufacturing industries. As a solution, a varying weighting method may be needed to give each component a different weight when computing the SI-LQ as a geometric average, reflecting the expected variation across types of industries. However, the pros and cons of different weighting methods may further complicate discussion of the properties of the SI-LQ measures. For the present purpose, we use simple equal weights for each component, and leave the choice of a varying weighting method to future research.

CONCLUSIONS

In descriptive and applied regression analysis, the SI-LQ performs better than the LQ at capturing industry clusters and their hypothesized effects. Our analysis of the SI-LQ components reveals that the relationship between its components and the LQ varies across industries, suggesting that aggregating these components into a single index may require differential assignments of weights for each industry. But even with simple and intuitive equal weights, we have accomplished our goal of designing a measure of industry agglomeration that incorporates both industry relatedness and spatial correlation, and which attenuates the extreme-value problem in the LQ. We will post the SI-LQ and its component measures online so that users can download and try different weighting methods based on their interests and explore more applications.

ACKNOWLEDGEMENTS

The authors thank the W.E. Upjohn Institute for providing the unsuppressed County Business Patterns Data.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

This publication was prepared in part using Federal funds awarded to the Trustees of Indiana University and as a sub-component under award number ED17HDQ3120040 from the U.S. Economic Development Administration, U.S. Department of Commerce. The work was also supported in part by US Department of Agriculture National Institute of Food and Agriculture [grant number 2017-67023-26906] as well as the Multistate/Regional Research and/or Extension Appropriations [project number NE1749]. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the Economic Development Administration; the U.S. Department of Commerce, or the U.S. Department of Agriculture.

NOTES

1 See Nakamura and Paul (2009) for a survey.

2 In the following context, we consider these concepts as synonyms: regional industry concentration, industry concentration, spatial concentration, industry agglomeration, industry clusters and spatial clusters.

3 In the literature, the LQ is commonly computed with industry employment, but it is also computed with other variables for the size or importance of an industry, such as establishment counts (Billings & Johnson, 2012; Delgado et al., 2016) and value-added (Mulligan & Schmidt, 2005). Owing to data availability at the county level and industry aggregation as well as common practice in the literature, we chose to use employment data to compute the LQ.

4 Among all 71 industries, 22 industries have a diagonal element that is the maximum value of the column.

5 It is an interesting counterfactual attempt to add ξ to LQ_{ij}. This can be interpreted as the situation where even though industry i does not exist in county j, it has the potential to locate there because this region already possesses the localization and urbanization economies for industry i.

6 The main reason for choosing such an industry classification is for easy calculation. Mulligan and Schmidt (2012) also point out that it is undesirable to discuss statistical properties of LQs at a very granular level of industry classification and the level of NAICS with three to four digits is appropriate.

7 Rail transportation (NAICS 482) and private households (NAICS 814) are missing in both the CBP and BEA data sources. Therefore, the corresponding rows and columns in the direct requirement matrix of these two sectors are deleted. However, when calculating the direct requirement matrix from the use and make tables, these two sectors are preserved so that part of the IO information of these two sectors is saved in the direct requirement matrix.

8 A five-nearest-neighbour spatial weight is also used, yielding similar results.

9 For the tables comparing descriptive statistics of each industry, see the supplemental data online.

10 See http://www.madeinalabama.com/industries/industry/automotive/.

11 The literature on agglomeration generally predicts growth in both employment and productivity (Cohen, Coughlin, & Paul, 2019). In the short run, we can expect differences in agglomerative advantage to generate more firm births or, in a recessionary period, fewer plant closings. In contrast, the negative employment effect of enhanced labour productivity is likely to be a long-run phenomenon. Studies that explore the kind of correlations we hypothesize here, and over similar timespans, include Barkley, Henry, & Kim (1999), Rosenthal and Strange (2003) and Gabe (2003). To test the effect of agglomeration on productivity that is often represented by wage rates, we estimated a model with the growth rate of average wages being the dependent variable.

The result of this model also shows that the SI-LQ has a higher and more significant effect on the growth rate of average wages than the LQ. For the detailed result, see the supplemental data online.

[12] We also estimated spatial panel models with two-way (county and time) fixed effects. The results of the two-way fixed-effects model are similar to those of the county fixed-effects models, but the coefficients on both the LQ and SI-LQ are insignificant. Those results are available from the authors on request.

[13] Categorizing LQ and SI-LQ makes them comparable with the same magnitude scale. Using the original values does not appreciably change the regression results. The results are available from the authors on request.

[14] To compare the effects of both the LQ and SI-LQ directly, we estimated a model that includes both variables. Although not reported in Table 3 owing to the problem of multicollinearity between the LQ and SI-LQ variables, the nested model yields similar results and the negative coefficient on the LQ is statistically significant.

[15] We also compute the Pearson correlation coefficients with similar results.

REFERENCES

Aiginger, K., & Rossi-Hansberg, E. (2006). Specialization and concentration: A note on theory and evidence. *Empirica*, *33*(4), 255–266. doi:10.1007/s10663-006-9023-y

Arbia, G., Espa, G., Giuliani, D., & Mazzitelli, A. (2010). Detecting the existence of space–time clustering of firms. *Regional Science and Urban Economics*, *40*(5), 311–323. doi:10.1016/j.regsciurbeco.2009.10.004

Barkley, D. L., Henry, M. S., & Kim, Y. (1999). Industry agglomerations and employment change in non-metropolitan areas. *Review of Urban and Regional Development Studies*, *11*(3), 168–186. doi:10.1111/1467-940X.00014

Billings, S. B., & Johnson, E. B. (2012). The location quotient as an estimator of industrial concentration. *Regional Science and Urban Economics*, *42*(4), 642–647. doi:10.1016/j.regsciurbeco.2012.03.003

Brülhart, M. (1998). Economic geography, industry location and trade: The evidence. *The World Economy*, *21*(6), 775–801. doi:10.1111/1467-9701.00163

Burridge, P., & Fingleton, B. (2010). Bootstrap inference in spatial econometrics: The J-test. *Spatial Economic Analysis*, *5*(1), 93–119. doi:10.1080/17421770903511346

Cainelli, G., & Iacobucci, D. (2016). Local variety and firm diversification: An evolutionary economic geography perspective. *Journal of Economic Geography*, *16*(5), 1079–1100. doi:10.1093/jeg/lbv040

Carroll, M. C., Reid, N., & Smith, B. W. (2008). Location quotients versus spatial autocorrelation in identifying potential cluster regions. *The Annals of Regional Science*, *42*(2), 449–463. doi:10.1007/s00168-007-0163-1

Cohen, J. P., Coughlin, C. C., & Paul, C. J. M. (2019). Agglomeration, productivity and regional growth: Production theory approaches. In R. Capello & P. Nijkamp (Eds.), *Handbook of regional growth and development theories* (2nd ed., pp. 104–123). Cheltenham, UK: Edward Elgar Publishing.

Combes, P. P., & Gobillon, L. (2015). The empirics of agglomeration economies. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of regional and urban economics* (Vol. 5, pp. 247–348). Amsterdam: North-Holland.

Crawley, A., Beynon, M., & Munday, M. (2013). Making location quotients more relevant as a policy aid in regional spatial analysis. *Urban Studies*, *50*(9), 1854–1869. doi:10.1177/0042098012466601

Cromley, R. G., & Hanink, D. M. (2012). Focal location quotients: Specification and applications. *Geographical Analysis*, *44*(4), 398–410. doi:10.1111/j.1538-4632.2012.00852.x

Cromley, R. G., Hanink, D. M., & Bentley, G. C. (2014). Geographically weighted colocation quotients: Specification and application. *The Professional Geographer*, *66*(1), 138–148. doi:10.1080/00330124.2013.768130

Delgado, M., Porter, M. E., & Stern, S. (2016). Defining clusters of related industries. *Journal of Economic Geography*, *16*(1), 1–38. doi:10.1093/jeg/lbv017

Duranton, G., & Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, *72*(4), 1077–1106. doi:10.1111/0034-6527.00362

Elhorst, J. P. (2014). *Spatial econometrics: From cross-sectional data to spatial panels*. New York, NY: Springer.

Ellison, G., & Glaeser, E. L. (1997). Geographic concentration in U.S. manufacturing industries: A dartboard approach. *Journal of Political Economy*, *105*(5), 889–927. doi:10.1086/262098

Ellison, G., Glaeser, E. L., & Kerr, W. R. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, *100*(3), 1195–1213. doi:10.1257/aer.100.3.1195

Feser, E. J., & Bergman, E. M. (2000). National industry cluster templates: A framework for applied regional cluster analysis. *Regional Studies*, *34*(1), 1–19. doi:10.1080/00343400050005844

Feser, E., Renski, H., & Goldstein, H. (2008). Clusters and economic development outcomes. *Economic Development Quarterly*, *22*(4), 324–344. doi:10.1177/0891242408325419.

Fracasso, A., & Marzetti, G. V. (2018). Estimating dynamic localization economies: The inadvertent success of the specialization index and the location quotient. *Regional Studies*, *52*(1), 119–132. doi:10.1080/00343404.2017.1281388

Gabe, T. (2003). Local industry agglomeration and new business activity. *Growth and Change*, *34*(1), 17–39. doi:10.1111/1468-2257.00197

Gibbs, R. M., & Bernat, G. A. (1997). Rural industry clusters raise local earnings. *Rural Development Perspectives*, *12*(3), 18–25.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, *100*(6), 1126–1152. doi:10.1086/261856

Goetz, S. J., Shields, M., & Wang, Q. (2009). Identifying food industry clusters: A comparison of analytical tools. In S. J. Goetz, S. C. Deller, & T. R. Harris (Eds.), *Targeting regional economic development* (pp. 281–310). London: Routledge.

Greenstone, M., Hornbeck, R., & Moretti, E. (2008). *Identifying agglomeration spillovers: Evidence from million dollar plants* (NBER working paper: 13833) Cambridge, MA: National Bureau of Economic Research.

Guimarães, P., Figueiredo, O., & Woodward, D. (2009). Dartboard tests for the location quotient. *Regional Science and Urban Economics*, *39*(3), 360–364. doi:10.1016/j.regsciurbeco.2008.12.003

Helsley, R. W., & Strange, W. C. (2014). Coagglomeration, clusters, and the scale and composition of cities. *Journal of Political Economy*, *122*(5), 1064–1093. doi:10.1086/676557

Horowitz, K. J., & Planting, M. A. (2009). Concepts and methods of the U.S. input–output accounts (Technical report). The U.S. Bureau of Economic Analysis of the U.S. Department of Commerce. Retrieved from https://www.bea.gov/sites/default/files/methodologies/IOmanual_{0}92906.pdf.

Isserman, A. M., & Westervelt, J. (2006). 1.5 million missing numbers: Overcoming employment suppression in County Business Patterns data. *International Regional Science Review*, *29*(3), 311–335. doi:10.1177/0160017606290359

Klarl, T. (2014). Is spatial bootstrapping a panacea for valid inference? *Journal of Regional Science*, *54*(2), 304–312. doi:10.1111/jors.12075

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, *15*(4), 143–156. doi:10.1257/jep.15.4.143

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton, FL: Chapman and Hall/CRC.

Leslie, T., & Kronenfeld, B. (2011). The colocation quotient: A new measure of spatial association between categorical subsets of points. *Geographical Analysis*, *43*(3), 306–326. doi:10.1111/j.1538-4632.2011.00821.x

Marcon, E., & Puech, F. (2003). Evaluating the geographic concentration of industries using distance-based methods. *Journal of Economic Geography*, *3*(4), 409–428. doi:10.1093/jeg/lbg016

Miller, R. E., & Blair, P. D. (2009). *Input–output analysis: Foundations and extensions* (2nd ed). Cambridge, MA: Cambridge University Press.

Moghadam, K., & Ballard, K. P. (1988). Small area modeling of the industrial sector (SAMIS): An integrated econometric—interindustry approach. *Environment and Planning A: Economy and Space*, *20*(5), 655–668. doi:10.1068/a200655

Moineddin, R., Beyene, J., & Boyle, E. (2003). On the location quotient confidence interval. *Geographical Analysis*, *35*(3), 249–256. doi:10.1111/j.1538-4632.2003.tb01112.x

Mulligan, G. F., & Schmidt, C. (2005). A note on localization and specialization. *Growth and Change*, *36*(4), 565–576. doi:10.1111/j.1468-2257.2005.00295.x

Nakamura, R., & Paul, C. J. M. (2009). Measuring agglomeration. In R. Capello & P. Nijkamp (Eds.), *Handbook of regional growth and development theories* (pp. 305–328). Northampton, MA: Edward Elgar Publishing.

Nathan, M., & Overman, H. G. (2013). Agglomeration, clusters, and industrial policy. *Oxford Review of Economic Policy*, *29*(2), 383–404. doi:10.1093/oxrep/grt019

O'Donoghue, D., & Gleave, B. (2004). A note on methods for measuring industrial agglomeration. *Regional Studies*, *38*(4), 419–427. doi:10.1080/03434002000213932

Organisation for Economic Co-operation and Development (OECD). (2008). *Handbook on constructing composite indicators: Methodology and user guide*. Paris: OECD Publ. Retrieved from https://www.oecd.org/sdd/42495745.pdf

Plant, R. E. (2012). *Spatial data analysis in ecology and agriculture using R*. Boca Raton, FL: CRC Press.

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

Rey, S. J. (2000). Integrated regional econometric + input − output modeling: Issues and opportunities. *Papers in Regional Science*, *79*(3), 271–292. doi:10.1007/PL00013613

Rey, S. J., & Jackson, R. W. (1999). Labor-productivity changes in regional econometric +input−output models. *Environment and Planning A: Economy and Space*, *31*(9), 1583–1599. doi:10.1068/a311583

Rosenthal, S. S., & Strange, W. C. (2001). The determinants of agglomeration. *Journal of Urban Economics*, *50*(2), 191–229. doi:10.1006/juec.2001.2230

Rosenthal, S. S., & Strange, W. C. (2003). Geography, industrial organization, and agglomeration. *Review of Economics and Statistics*, *85*(2), 377–393. doi:10.1162/003465303765299882

Slaper, T. F., Harmon, K. M., & Rubin, B. M. (2018). Industry clusters and regional economic performance: A study across U.S. metropolitan statistical areas. *Economic Development Quarterly*, *32*(1), 44–59. doi:10.1177/0891242417752248

Tian, Z. (2013). Measuring agglomeration using the standardized location quotient with a bootstrap method. *The Journal of Regional Analysis & Policy*, *43*(1), 186–197.