The Journal of
Regional
Analysis
& Policy

JRAP 43(2): 186-197.

Measuring Agglomeration Using the Standardized Location Quotient with a Bootstrap Method

Zheng Tian
West Virginia University – USA

Abstract. The location quotient is a widely used index to measure agglomeration. However, a problem concerning the usefulness of the location quotient is how to obtain an objective cut-off value to identify the existence of agglomeration for an industry in a region. This paper extends the idea of O'Donoghue and Gleave (2004) and proposes a bootstrap method to determine the cut-off value based on the standardized location quotient. The advantage of our method is that the bootstrap method does not rely on any assumption regarding the statistical distribution of the location quotient, which is a major limitation of O'Donoghue and Gleave's (2004) approach. Then we apply the method to measure agglomeration of manufacturing industries at the county level in the United States.

1. Introduction

The phenomenon of industrial agglomeration[1] has drawn interest from both researchers and policy makers during the last few decades. In the academic field voluminous theoretical and empirical studies on industrial agglomeration have been emerging, motivated by the New Economic Geography, since the 1990's (Krugman, 1991). Policymakers, inspired by the idea of industrial clustering (Porter, 1990), have adopted the cluster-based economic development strategy (Carroll et al., 2008) as a policy tool to promote the local economy. For both researchers and practitioners, it is critical to first have some stylized facts about industrial agglomeration. One of the questions of the stylized facts is to understand the extent to which industrial agglomeration occurs, that is, how to measure agglomeration.

Constructing an index to measure industrial agglomeration is an important aspect of empirical studies in regional economics. Economists have long been seeking to develop an index that can accurately reflect the degree of agglomeration across industries, time, and space. Among many indices that have been discovered, we focus on the location quotient (LQ), which is used widely in regional science due to its computational simplicity and low data requirements. The LQ measures the ratio between the local and national share of productive activities of a particular industry in a region, usually using employment to represent productive activities. $LQ > 1$ can be interpreted as indicating that the industry under study is more concentrated in the region than the national average. Apart from using unity as the cut-off value, some researchers use other values, such as $LQ > 1.25$ or 2, to delimit agglomeration of an industry in a region. However, people can still question how large the value of the LQ should be to ensure the existence of agglomeration.

To find an objective cut-off value of the LQ for identifying agglomeration, O'Donoghue and Gleave (2004) propose an approach of computing the standardized location quotient (SLQ), which is simply the z-statistic of the LQ, and using the 5% critical value

[1] In this paper, the terms of "agglomeration", "concentration", and "clustering" are used interchangeably, although these terms have their specific implications. See Brülhart (1998) and Franceschi et al. (2009).

of the standard normal distribution as the cut-off value of the SLQ. However, the limitation of their approach is that if the LQ does not follow the normal distribution, then the cut-off value determined by this approach is not reliable. Therefore, following the idea of the SLQ approach, we suggest an alternative way, a bootstrap method, to obtain an objective cut-off value of the LQ without any assumption about the statistical distribution. Then we apply this method to measure agglomeration of manufacturing industries in the U.S. counties.

The remainder of the paper is organized as follows. Section 2 briefly surveys the existing agglomeration indices and justifies the choice of the specific method used in this study. Section 3 introduces the method of computing the SLQ and using the bootstrap method to delimit agglomeration. Sections 4 and 5 present the data used and the results of computation. Finally, section 6 concludes the paper and proposes some future research for the use of the SLQ method.

2. Literature review

Combes et al. (2008, Chapter 10) and Nakamura and Paul (2009) have provided a comprehensive literature review that covers most of the existing agglomeration indices. Thus, we will not try to explain all existing indices in detail. Instead, we will briefly survey some important aspects about the indices and explain why we use the SLQ approach to measuring industrial agglomeration.

The existing indices can be categorized into two types: discrete and continuous. The discrete indices apply to areal data that are discrete spatial units, like counties, states, countries, etc. The majority of agglomeration indices belong to the discrete type, including the LQ, the Gini index, the Theil index, the Isard index, the Herfindahl–Hirschman (HH) index, the Ellison-Glaeser (EG) index (Ellison and Glaeser, 1997), and the Maurel-Sedillot (MS) index (Maurel and Sedillot, 1999). Among the discrete indices, the EG index has been widely adopted by researchers in measuring industrial agglomeration (Rosenthal, 2001; Holmes and Stevens, 2004; Bertinelli and Decrop, 2005). The dartboard framework, under which the EG index is derived, also initiates a way to discover other indices and to introduce statistical tests on indices (Maurel and Sedillot, 1999; Guimarães et al., 2009). However, the EG, MS, and Gini indices suffer a common problem in their application. They are unable to evaluate the degree of agglomeration of an industry in a particular region because the regional dimension is integrated out in the process of computation. Another problem that all discrete indices fail to address is the modifiable areal unit problem (MAUP). The MAUP makes an agglomeration index biased when industries actually agglomerate across the administrative boundaries of spatial units for which the data are collected. The continuous agglomeration indices are designed to address the MAUP.

With the assumption of continuous space, the continuous agglomeration indices are applied to spatial point objects represented by geographical coordinates. A typical continuous index measures the density of economic activities along the link between pairs of points. The studies exploring the continuous-type indices include Duranton and Overman (2005), which analyzes Duranton-Overman K-density, and Marcon and Puech (2003, 2009) and Arbia et al. (2010), which investigate Ripley's K-function. However, a practical problem of the continuous indices is that data requirements are so high that ordinary researchers without access to a dataset with very detailed information, especially the coordinates of each establishment, cannot compute the continuous index. Moreover, the continuous indices usually measure agglomeration without reference to any administrative entities, so their implications for local economic policy makers are not readily applicable.

Establishing a guideline for new agglomeration indices, Duranton and Overman (2005) advance five properties that an index should satisfy. Combes et al. (2008, Chapter 10) and Kominers (2007) provide three additional properties. Together, these properties require that an agglomeration index should (1) be comparable across industries, (2) be comparable across spatial scales, (3) be unbiased with respect to arbitrary changes to spatial classification, (4) be unbiased with respect to arbitrary changes to industrial classification, (5) control for the overall distribution of economic activities, (6) allow for a statistical significance test, (7) be computable in the closed form from accessible data, and (8) be justified by a suitable model.

Although it is desirable to find an agglomeration index that can meet most of these properties, the index that a researcher actually chooses is often constrained by data availability and the purpose of the study. This paper aims to identify industrial agglomeration in an administrative spatial unit. Data on such units, like states and counties, are readily available. It implies that neither the Gini and EG index nor the continuous indices are suitable for the purpose as the former provides no information regarding agglomeration in a specific spatial unit and the latter are applied only to spatial point objects. In contrast, taking advantage of its flexibility in application at any level of industrial and spatial classification, the LQ-type index can serve the purpose well.

However, a problem in using the LQ relates to how one objectively determines the cut-off value for defining agglomeration. Choosing any arbitrary value of the LQ as the cut-off value can always call the validity into question. Besides the SLQ approach of O'Donoghue and Gleave (2004), three studies attempt to address this problem by building some statistical tests to determine the cut-off value. Moineddin et al. (2003) derive an expression of the standard deviation of the LQ and construct the confidence interval by assuming the normal distribution of the LQ. Following the dartboard framework of Ellison and Glaeser (1997), Guimarães et al. (2009) provide a theoretical foundation of the LQ and derive two test statistics that are asymptotically chi-squared distributed. Billings and Johnson (2012) examine the accuracy of statistical tests about the LQ, which is assumed to follow a scaled Binomial or Poisson distribution. One common drawback of these previous studies is that the assumptions on the statistical distribution of the LQ and test statistics may not hold, so their results are not reliable. In this paper, we propose a bootstrap method that does not depend on any assumption about the statistical distribution by extending O'Donoghue and Gleave's (2004) SLQ approach.

3. The Standardized Location Quotient

Let $i = 1,2,\ldots,I$ denote industries and $j = 1,2,\ldots,J$ denote regions. Then the LQ of industry i in region j is defined as

$$LQ_{ij} = \frac{s_{ij}}{s_{*j}} = \frac{x_{ij}/x_{i*}}{x_{*j}/x_{**}} \tag{1}$$

where x_{ij} represents employment of industry i in region j, x_{i*} is total employment of industry i in all regions, x_{*j} is total employment of all industries in region j, and x_{**} is total employment of the overall economy. Thus, s_{ij} is the share of industry i's employment in region j relative to total employment of industry i, and s_{*j} is the share of region j's employment relative to total employment in the overall economy. If LQ_{ij} is above one, then industry i is said to be concentrated in region j. Arguing against using unity or other arbitrary values of the LQ to delimit agglomeration, O'Donoghue and Gleave (2004) propose using the standardized location quotient (SLQ), simply the z-statistic of the original LQ, for which the equation of industry i in region j is,

$$SLQ_{ij} = \frac{LQ_{ij}-\overline{LQ}_i}{std(LQ_i)} \tag{2}$$

where $\overline{LQ}_i$ and $std(LQ_i)$ are the mean and standard deviation of the LQ of industry i.

Before being converted to the z-statistic, the LQ is tested for whether it is normally distributed using the Kolmogorov-Smirnov normality test. If the test fails to confirm the normal distribution, the logarithmic function is used to transform the LQ, followed by another test for normality of log(LQ). Passing the normality test implies that the SLQ, or the standardized log(LQ), should conform to the standard normal distribution. The cut-off level for confirming the existence of agglomeration in a region is then determined by the critical value of the standard normal distribution at the 5% level, i.e., 1.96 for a two-tailed test or 1.64 for a one-tailed test.

O'Donoghue and Gleave's (2004) approach hinges on the assumption that the LQ is normally distributed. Hence, the most serious limitation of their approach is that the critical value of the standard normal distribution may not be reliable if the normality assumption is invalid. Moreover, an implicit assumption in the approach is that the statistical distributions of the LQ indices of all industries are the same. This assumption is questionable because the actual data generating process of the LQ indices is very likely to be different among industries, determined by industry-specific characteristics. A simple regression model can illustrate the problem more clearly. Consider the following model

$$\boldsymbol{LQ}_i = \alpha\boldsymbol{\iota} + \boldsymbol{u_i} \tag{3}$$

where $\boldsymbol{LQ}_i$ is a $J \times 1$ vector of the LQ of industry i, $\boldsymbol{\iota}$ is a $J \times 1$ vector of 1s, and $\boldsymbol{u_i}$ is a vector of random errors following some statistical distribution. It follows that the SLQ is simply the residuals from the ordinary least square estimation of (3) divided by the standard deviation of the residuals. If, in the true data generating process, $\boldsymbol{u_i}$ is not normally distributed, neither is the distribution of the residuals, i.e., the SLQ. Further, it is not necessarily true that $\boldsymbol{u_i}$ for all $i = 1,2,\ldots,I$ conforms to the same distribution. Therefore, there is no well-founded justification for using the critical value of the standard normal distribution at the 5% level to determine the cut-off level of the SLQ.

O'Donoghue and Gleave (2004) acknowledge this limitation and suggest not using the SLQ approach if the normality test on the LQ fails, but the authors provide no alternatives to solve the problem. To circumvent the obstacle imposed by the normal distribution assumption, we propose to use the bootstrap method to get the estimated critical value of the actual distribution of the SLQ at the 5% level. Essentially, the bootstrap method is based on the Fundamental Theorem of Statistics, asserting that the empirical distribution function of a random variable X, which can be obtained by bootstrapping, consistently estimates the true cumulative distribution function of X. It follows that test statistics constructed from the empirical distribution are also the consistent estimates of the exact statistics from the true distribution (Davidson and MacKinnon, 2004).

The steps of using the bootstrap method to determine the cut-off level of the SLQ are as follows:

1) Computing the SLQ for each industry in all regions.
2) Bootstrap resampling the SLQ for each industry. A bootstrap resampling is a process of randomly drawing samples from the whole original sample with replacement to get a bootstrap sample with the same length as the original sample.
3) Obtaining the 95^{th} percentile for each bootstrap sample and performing bootstrap resampling N (set N = 999) times to get a set of N 95^{th} percentiles. The purpose of this step is to draw the 95^{th} percentile from its empirical distribution.
4) Using the sample mean of the N 95^{th} percentile as the estimate of the critical value at the 5% level of the true distribution.

By using the bootstrap method, the determination of the cut-off value of the SLQ does not rely on any assumption of the statistical distribution. Moreover, the cut-off value of the SLQ for each industry is unique because we apply the bootstrap method to each industry individually. Another advantage of using the bootstrap method is the simplicity in implementation. Our method enhances the applicability of the original SLQ approach by relaxing the normality assumption.

In this study, we do not intend to address other issues regarding an agglomeration index that should satisfy the properties proposed by Duranton and Overman (2005) and other researchers. Specifically, our method can be construed as partly solving the sixth property, which requires a test of the statistical significance of an index. To meet other properties, we need a more theoretically-based method to derive an agglomeration index, which is definitely not an easy task but a promising direction for future studies.

4. Data

We use the data from the County Business Patterns (CBP) for 2002 imputed by Isserman and Westervelt (2006). The CBP, published by the U.S. Bureau of the Census, contains a comprehensive annual compilation of information about the location of establishments with employment in the United States. The CBP data set is constructed with a hierarchical structure, in which industries are categorized by the two to six digit NAICS codes and spatial units cover four levels of spatial aggregation: the nation, states, counties, and zip-code areas. However, the nondisclosure problem of the CBP impairs the usefulness of the data set. For the purpose of protecting private business confidentiality, some data on employees are suppressed by the Bureau of the Census. Instead, employment flags are used to indicate the range of the missing value of an industry/county pair with undisclosed data.

Isserman and Westervelt (2006) propose a two-stage method to complete the CBP data set. The first stage identifies the smallest possible range for each withheld data entry, given the information provided by employment flags. Taking advantage of the hierarchical structure of the CBP data set, the second stage estimates the missing values, and iteratively adjusts each estimate to ensure that the estimated number of employees can add up correctly to the total number of employees in the higher levels of aggregation along both the industrial and spatial hierarchies. We downloaded their complete data set of the 2002 CBP from www.wholedata.com.

In this paper we use the county-level employment of manufacturing industries with 3-digit NAICS codes to compute the SLQ. Table 1 shows the names of all manufacturing industries with 3-digit NAICS codes. The total sample size is 35,328, with each observation being a unique industry/county pair from the 21 manufacturing industries and 3,076 counties.

Table 1. The 3-digit-NAICS manufacturing industries

NAICS	Manufacturing Industries
311	Food mfg
312	Beverage & tobacco product mfg
313	Textile mills
314	Textile product mills
315	Apparel manufacturing
316	Leather & allied product mfg
321	Wood product mfg
322	Paper mfg
323	Printing & related activities
324	Petroleum & coal products mfg
325	Chemical mfg
326	Plastics & rubber products mfg
327	Nonmetallic mineral product mfg
331	Primary metal mfg
332	Fabricated metal product mfg
333	Machinery mfg
334	Computer & electronic product mfg
335	Electrical equip, appliance mfg
336	Transportation equipment mfg
337	Furniture & related product mfg
339	Miscellaneous mfg

5. Results

We compute four forms of the LQ indices: LQ, log(LQ), SLQ, and standardized log(LQ) (referred to as SLLQ hereafter). In this section we first report the effects of standardization and the logarithmic transformation of the original LQ on the skewness problem and the statistical distribution of the LQ. Then we use the bootstrap method to obtain the cut-off values of the SLQ and SLLQ for all manufacturing industries.

Table 2 shows that the LQ indices for all manufacturing industries exhibit severe skewness. The gap between the median and mean is considerable, and the ranges of the LQ indices differ widely among industries. The minimum values of the LQ indices for all manufacturing industries are approximately zero, but the maximum values can be as high as 338.20 (Beverage and Tobacco Product Manufacturing) and as low as 27.05 (Computer and Electronic Product Manufacturing). Also, only a few of the mean values of the LQ indices are close to unity (Printing & Related Support Activities and Computer & Electronic Product manufacturing), and the variation of the mean values of the industries is also substantial. This invalidates the practice of using unity or any other arbitrary value as the cut-off value for identifying agglomeration.

Table 2. LQ summary statistics for manufacturing industries.

	Min	Median	Mean	Max
Food mfg	0.00	0.63	2.50	65.54
Beverage & tobacco product mfg	0.00	0.57	2.82	338.20
Textile mills	0.00	0.43	7.40	170.90
Textile product mills	0.01	0.45	4.16	250.10
Apparel manufacturing	0.00	0.43	3.76	134.10
Leather & allied product mfg	0.01	0.88	8.30	298.70
Wood product mfg	0.00	1.45	4.78	81.22
Paper mfg	0.00	1.23	4.28	85.68
Printing & related activities	0.01	0.39	1.09	47.93
Petroleum & coal products mfg	0.00	0.44	3.96	193.20
Chemical mfg	0.00	0.53	1.89	71.64
Plastics & rubber products mfg	0.00	1.01	2.41	55.25
Nonmetallic mineral product mfg	0.01	0.82	2.11	82.63
Primary metal mfg	0.00	0.79	3.43	106.70
Fabricated metal product mfg	0.00	0.79	1.46	34.06
Machinery mfg	0.00	0.79	1.86	65.28
Computer & electronic product mfg	0.00	0.33	1.13	27.05
Electrical equip, appliance mfg	0.00	0.86	3.36	87.68
Transportation equipment mfg	0.00	0.63	1.93	47.27
Furniture & related product mfg	0.01	0.50	2.39	113.50
Miscellaneous mfg	0.01	0.45	1.44	52.42

Using the logarithmic transformation and standardization of the LQ can effectively alleviate the skewness problem. As shown in Table 3, the gap between the mean and median of log(LQ) is remarkably shortened. For example, the mean and median of log(LQ) in Food Manufacturing are almost equal. Moreover, the differences among the maximum values of log(LQ) between industries are not as large as the original LQ indices, and the range between the minimum and maximum values of log(LQ) is, to some extent, balanced around the means for most industries. Standardization can also reduce the gap between the mean and median, but the range of the SLQ values is not as balanced as log(LQ) (see Table 4). Finally, the SLLQ combines the effects of both the logarithmic transformation and standardization, resulting in a balanced range around the mean and median that is close to zero (see Table 5).

The normality test of the SLQ and SLLQ provides little support for the normal distribution assumption, which is the key in O'Donoghue and Gleave's (2004) approach to determining the cut-off value. The SLQ indices for all industries fail to pass the Kolmogorov-Smirnov normality test. Table 6 shows that p-values for all SLQ indices are zero, which means the null hypothesis of the normal distribution is rejected. Although the test result for the SLLQ is better than for the SLQ, the SLLQ indices for only three industries (Food Manufacturing, Beverage & Tobacco Product Manufacturing, and Miscellaneous Manufacturing) have a p-value greater than 0.05, which would support the normality assumption. Figure 1 displays histograms of the SLQ and SLLQ for Food Manufacturing, from which we can observe that the histogram of the SLLQ looks more like a normal distribution than does that of the SLQ.

Table 3. Summary statistics for log(LQ) of manufacturing industries

	Min	Median	Mean	Max
Food mfg	-5.57	-0.46	-0.46	4.18
Beverage & tobacco product mfg	-5.81	-0.56	-0.61	5.82
Textile mills	-6.15	-0.85	-0.63	5.14
Textile product mills	-4.91	-0.79	-0.56	5.52
Apparel manufacturing	-5.99	-0.84	-0.66	4.90
Leather & allied product mfg	-5.13	-0.13	0.00	5.70
Wood product mfg	-5.94	0.37	0.29	4.40
Paper mfg	-6.00	0.20	0.11	4.45
Printing & related activities	-4.48	-0.95	-0.88	3.87
Petroleum & coal products mfg	-5.53	-0.83	-0.57	5.26
Chemical mfg	-5.84	-0.63	-0.72	4.27
Plastics & rubber products mfg	-6.77	0.01	-0.18	4.01
Nonmetallic mineral product mfg	-5.01	-0.20	-0.14	4.41
Primary metal mfg	-6.57	-0.23	-0.32	4.67
Fabricated metal product mfg	-6.68	-0.23	-0.40	3.53
Machinery mfg	-5.50	-0.23	-0.38	4.18
Computer & electronic product mfg	-6.45	-1.10	-1.23	3.30
Electrical equip, appliance mfg	-5.99	-0.15	-0.32	4.47
Transportation equipment mfg	-6.92	-0.46	-0.74	3.86
Furniture & related product mfg	-4.90	-0.70	-0.59	4.73
Miscellaneous mfg	-4.90	-0.79	-0.79	3.96

Table 4. Summary statistics for SLQ of manufacturing industries.

	Min	Median	Mean	Max
Food mfg	-0.47	-0.35	-0.00	11.77
Beverage & tobacco product mfg	-0.23	-0.18	-0.00	26.96
Textile mills	-0.42	-0.40	0.00	9.30
Textile product mills	-0.28	-0.25	0.00	16.48
Apparel manufacturing	-0.38	-0.34	-0.00	13.24
Leather & allied product mfg	-0.32	-0.29	0.00	11.24
Wood product mfg	-0.56	-0.39	-0.00	8.95
Paper mfg	-0.45	-0.32	-0.00	8.55
Printing & related activities	-0.39	-0.26	0.00	17.05
Petroleum & coal products mfg	-0.28	-0.25	0.00	13.52
Chemical mfg	-0.39	-0.28	-0.00	14.42
Plastics & rubber products mfg	-0.60	-0.35	0.00	13.18
Nonmetallic mineral product mfg	-0.47	-0.29	0.00	17.92
Primary metal mfg	-0.42	-0.33	0.00	12.78
Fabricated metal product mfg	-0.67	-0.31	-0.00	14.93
Machinery mfg	-0.57	-0.33	-0.00	19.31
Computer & electronic product mfg	-0.50	-0.35	0.00	11.50
Electrical equip, appliance mfg	-0.46	-0.34	-0.00	11.55
Transportation equipment mfg	-0.54	-0.36	0.00	12.70
Furniture & related product mfg	-0.32	-0.25	-0.00	14.89
Miscellaneous mfg	-0.44	-0.30	-0.00	15.73

Table 5. Summary statistics for SLLQ of manufacturing industries.

	Min	Median	Mean	Max
Food mfg	-2.90	-0.00	-0.00	2.63
Beverage & tobacco product mfg	-2.90	0.03	-0.00	3.59
Textile mills	-2.12	-0.09	-0.00	2.21
Textile product mills	-2.33	-0.12	0.00	3.25
Apparel manufacturing	-2.46	-0.08	0.00	2.57
Leather & allied product mfg	-2.42	-0.06	-0.00	2.68
Wood product mfg	-3.51	0.05	0.00	2.32
Paper mfg	-3.40	0.05	0.00	2.42
Printing & related activities	-2.70	-0.05	-0.00	3.57
Petroleum & coal products mfg	-2.68	-0.14	-0.00	3.15
Chemical mfg	-2.96	-0.05	0.00	2.89
Plastics & rubber products mfg	-3.90	0.11	-0.00	2.48
Nonmetallic mineral product mfg	-3.76	-0.05	0.00	3.51
Primary metal mfg	-3.12	0.04	-0.00	2.49
Fabricated metal product mfg	-4.59	0.12	0.00	2.87
Machinery mfg	-3.24	0.09	0.00	2.88
Computer & electronic product mfg	-2.79	0.07	0.00	2.42
Electrical equip, appliance mfg	-2.83	0.09	0.00	2.40
Transportation equipment mfg	-3.15	0.14	0.00	2.34
Furniture & related product mfg	-2.70	-0.07	-0.00	3.34
Miscellaneous mfg	-2.68	-0.00	-0.00	3.10

Table 6. The Kolmogorov-Smirnov Normality Test of Standardized LQs.

	SLQ		SLLQ	
	statistic	p value	statistic	p value
Food mfg	0.32	0.00	0.01	0.41
Beverage & tobacco product mfg	0.41	0.00	0.02	0.24
Textile mills	0.34	0.00	0.07	0.00
Textile product mills	0.39	0.00	0.06	0.00
Apparel manufacturing	0.35	0.00	0.04	0.00
Leather & allied product mfg	0.37	0.00	0.04	0.01
Wood product mfg	0.29	0.00	0.03	0.00
Paper mfg	0.33	0.00	0.06	0.00
Printing & related support activities	0.35	0.00	0.02	0.04
Petroleum & coal products mfg	0.39	0.00	0.06	0.00
Chemical mfg	0.35	0.00	0.02	0.01
Plastics & rubber products mfg	0.27	0.00	0.05	0.00
Nonmetallic mineral product mfg	0.32	0.00	0.03	0.00
Primary metal mfg	0.34	0.00	0.04	0.00
Fabricated metal product mfg	0.25	0.00	0.05	0.00
Machinery mfg	0.29	0.00	0.04	0.00
Computer & electronic product mfg	0.31	0.00	0.05	0.00
Electrical equip, appliance mfg	0.32	0.00	0.04	0.00
Transportation equipment mfg	0.29	0.00	0.06	0.00
Furniture & related product mfg	0.37	0.00	0.05	0.00
Miscellaneous mfg	0.33	0.00	0.02	0.18

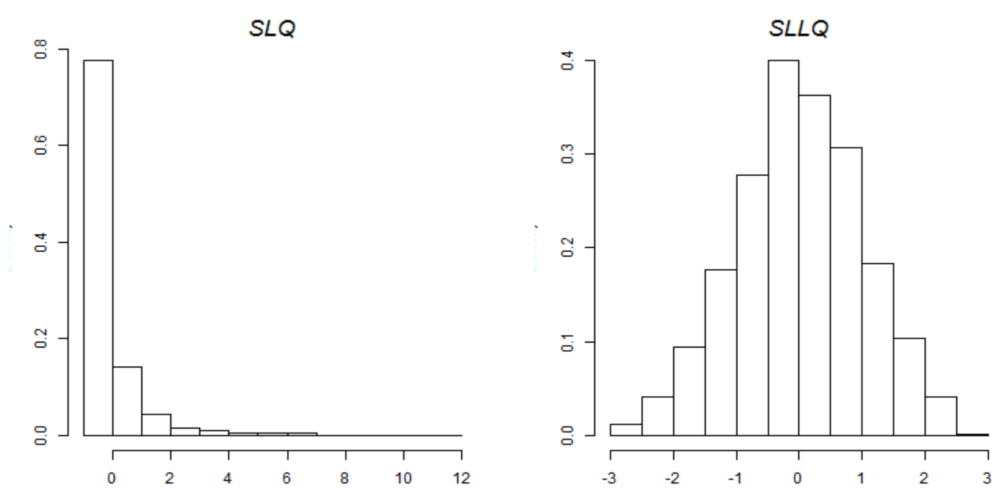

Figure 1. Comparison of histograms between SLQ and SLLQ of Food Manufacturing.

The failure of the normality test for the LQ indices of most industries suggests using the bootstrap method to obtain the cut-off value to delimit agglomeration. Tables 7 and 8 present the cut-off values of manufacturing industries based on the SLQ and SLLQ, respectively, using the bootstrap method. As shown in the first column in both tables, the cut-off values for most industries are different from 1.64, the 5% critical value of the standard normal distribution for a one-tailed test. Six industries have cut-off values greater than 1.64 using the SLQ, while nine industries have cut-off values greater than 1.64 using the SLLQ. That means that if 1.64 is used as the cut-off value, as O'Donoghue and Gleave (2004) suggest, for more than half of 3-digit NAICS manufacturing industries we would identify fewer counties as having agglomeration of a particular industry than if we use the bootstrap method to define the cut-off value. Further, the cut-off values vary across industries, which is expected as we use the bootstrap method for each industry independently.

Among the SLQ indices, the lowest cut-off level, 0.78 is for Beverage & Tobacco Product Manufacturing, and the highest level, 2.01, is for Wood Product Manufacturing. The cut-off values of the SLLQ indices change in an even narrower range than the SLQ indices. This is because the logarithmic transformation diminishes the leverage effect from the extreme values. Interestingly, the average of the cut-off values of the SLLQ is 1.63, close to the 5% critical value of the standardized normal distribution, which means that the cut-off value defined by O'Donoghue and Gleave (2004) is reasonable in an average sense but not for each individual industry.

The second and third columns in Tables 7 and 8 show that, for counties that are identified as having agglomeration of a particular industry, the shares of employment that these counties account for are nearly identical for using both the SLQ and SLLQ. Also, there is only marginal discrepancy in the number of these counties between the two tables. We cannot get the same kind of results if we rely on the standard normal distribution to get the cut-off values because the SLQ and SLLQ have different statistical distributions. In contrast, the bootstrap method estimates the cut-off values, the means of the bootstrap samples of the 95th percentiles, from the empirical distribution that is the consistent estimate of the true data generating process, regardless of statistical distributions of the SLQ and SLLQ. Therefore, we can identify the same set of counties as having agglomeration of some industry using either the SLQ or the SLLQ even though they have different distributions.

Table 7. The SLQ cut-off value for identifying agglomeration.

	The Cutoff Value	Share of employment	Number of Counties
Food mfg	1.63	0.14	125
Beverage & tobacco product mfg	0.78	0.27	51
Textile mills	1.98	0.19	44
Textile product mills	1.17	0.30	73
Apparel manufacturing	1.50	0.11	74
Leather & allied product mfg	1.14	0.21	31
Wood product mfg	2.01	0.13	121
Paper mfg	1.56	0.12	58
Printing & related activities	0.94	0.12	108
Petroleum & coal products mfg	1.05	0.27	43
Chemical mfg	1.15	0.16	85
Plastics & rubber products mfg	1.70	0.10	91
Nonmetallic mineral product mfg	1.31	0.12	125
Primary metal mfg	1.36	0.20	69
Fabricated metal product mfg	1.68	0.08	131
Machinery mfg	1.53	0.09	112
Computer & electronic product mfg	1.64	0.26	68
Electrical equip, appliance mfg	1.66	0.12	60
Transportation equipment mfg	1.86	0.16	93
Furniture & related product mfg	0.94	0.27	107
Miscellaneous mfg	1.49	0.12	104

Table 8. The SLLQ cut-off level for identifying agglomeration.

	The Cutoff Value	Share of employment	Number of Counties
Food mfg	1.63	0.14	125
Beverage & tobacco product mfg	1.75	0.27	51
Textile mills	1.67	0.19	44
Textile product mills	1.95	0.30	73
Apparel manufacturing	1.65	0.11	74
Leather & allied product mfg	1.70	0.21	31
Wood product mfg	1.58	0.14	123
Paper mfg	1.58	0.11	57
Printing & related activities	1.64	0.12	108
Petroleum & coal products mfg	1.88	0.27	44
Chemical mfg	1.58	0.16	86
Plastics & rubber products mfg	1.43	0.10	91
Nonmetallic mineral product mfg	1.70	0.13	126
Primary metal mfg	1.49	0.20	70
Fabricated metal product mfg	1.48	0.08	131
Machinery mfg	1.46	0.09	112
Computer & electronic product mfg	1.50	0.26	68
Electrical equip, appliance mfg	1.53	0.12	60
Transportation equipment mfg	1.47	0.16	93
Furniture & related product mfg	1.77	0.27	108
Miscellaneous mfg	1.71	0.12	104

Tables 7 and 8 also illustrate that the degree of agglomeration is disparate across manufacturing industries. For example, 8% of employment in Fabricated Metal Manufacturing is found in 131 counties. In contrast, in Leather & Allied Product Manufacturing only 31 counties account for over 20% of the industry's employment. Maps make such comparison more clear. Figures 2 and 3 show that Fabricated Metal Manufacturing is more widely spread across counties than Leather & Allied Product Manufacturing, which is concentrated in only a few counties.

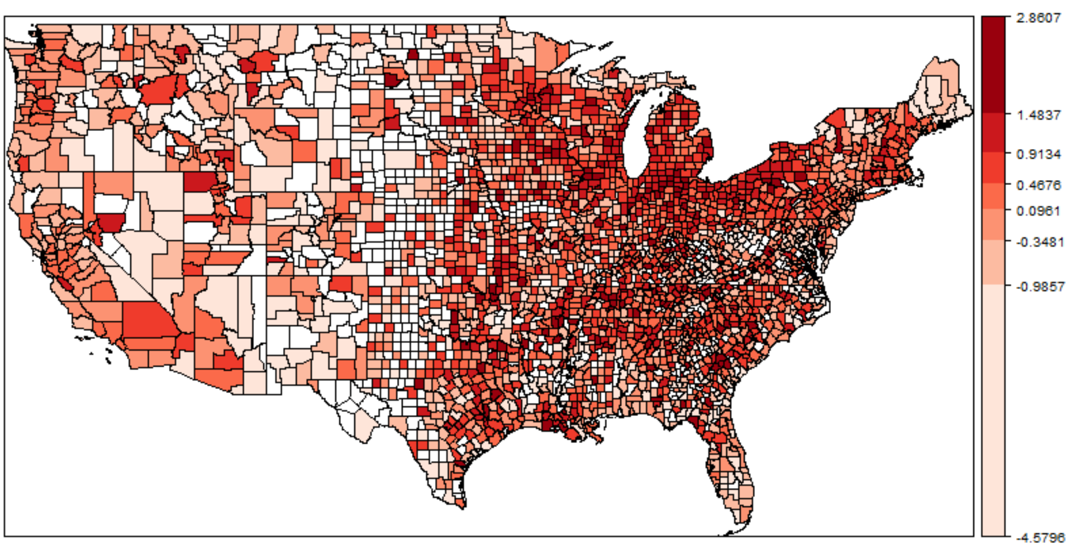

Figure 2. The distribution of Standardized log(LQ) of Fabricated Metal Manufacturing.

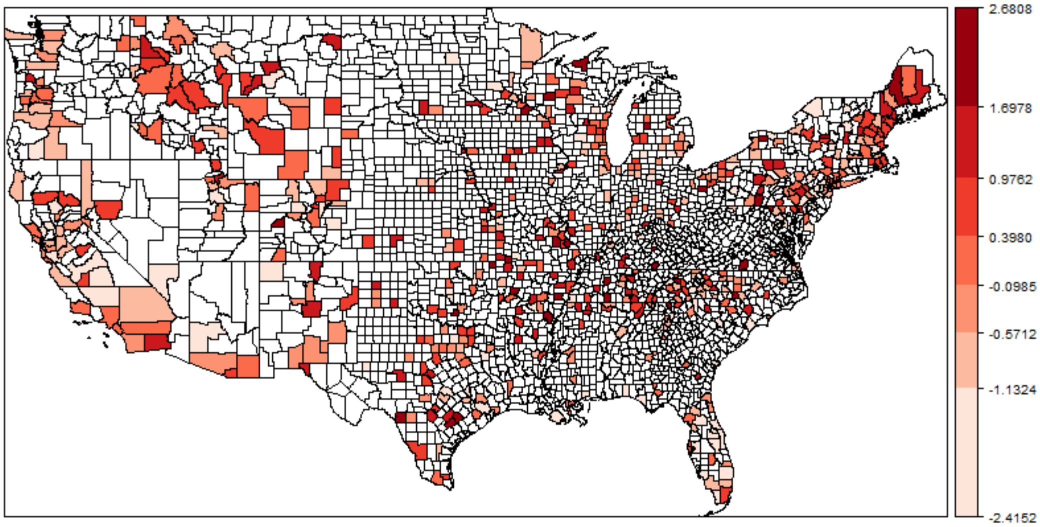

Figure 3. The distribution of Standardized log(LQ) of Leather & Allied Product Manufacturing.

6. Conclusion

This paper provides a simple bootstrap method to obtain the cut-off value of the SLQ for identifying the existence of industrial agglomeration in a region. The bootstrap method contributes to the SLQ approach of O'Donoghue and Gleave (2004) by relaxing the normality assumption. The analysis of agglomeration using 2002 data for U.S manufacturing industries demonstrates the usefulness of the bootstrap method in identifying agglomeration. However, this method does not address other issues concerning measuring agglomeration. Searching for an agglomeration index satisfying the set of properties advocated by Duranton and Overman (2005) is still a task for future studies.

The bootstrap method for identifying agglomeration can serve as the starting point for other related studies. The SLQ indices can be readily fed into a spatial econometric regression model in which the spatial weight matrix requires the dependent variable to be region-specific. The spatial econometric model can study the spatial spillover effect of the SLQ from the surrounding regions which may result from estimation errors due to the modifiable areal unit problem. Also, using relatively accurate cut-off values obtained from the bootstrap method, we can divide regions into two contrasting groups, one with agglomeration and another without agglomeration, and then examine the distinctive industrial and local characteristics within each group which can determine the formation of agglomeration.

Acknowledgements

I would like to thank Professor Brian Cushing for his extremely careful correcting the draft. I am also grateful to Professor Randall Jackson, Professor Peter Schaeffer, Dr. Mulugeta Kahsai, and Dr. Christa D. Court for their feedback, encouragement, and help.

References

Arbia, G., G. Espa, D. Giuliani, and A. Mazzitelli. 2010. Detecting the Existence of Space-time Clustering of Firms. *Regional Science and Urban Economics* 40 (5): 311-323.

Bertinelli, L., and J. Decrop. 2005. Geographical Agglomeration: Ellison and Glaeser's Index Applied to the Case of Belgian Manufacturing Industry. *Regional Studies* 39 (5): 567-583.

Billings, S. B., and E. B. Johnson. (2012) The Location Quotient as an Estimator of Industrial Concentration. *Regional Science and Urban Economics* 42: 642–647

Brülhart, M. 1998. Economic Geography, Industry Location and Trade: The Evidence. *The World Economy* 21 (6): 775-801.

Carroll, M., N. Reid, and B. Smith. 2008. Location Quotients Versus Spatial Autocorrelation in Identifying Potential Cluster Regions. *The Annals of Regional Science* 42: 449–463.

Combes, P., T. Mayer, and J. F. Thisse. 2008. *Economic geography: the integration of regions and nations*. Princeton: Princeton University Press.

Davidson, R., and J. G. MacKinnon. 2004. *Econometric Theory and Methods*. Oxford University Press.

Duranton, G., and H. G. Overman. 2005. Testing for Localization Using Micro- geographic Data. *The Review of Economic Studies* 72(4): 1077–1106.

Ellison, G., and E. L. Glaeser. 1997. Geographic Concentration in U.S. Manufacturing Industries: a Dartboard Approach. *Journal of Political Economy*, 105(5): 889.

Franceschi, F., M. Mussoni, and G. Pelloni. 2009. A Note on New Measures of Agglomeration and Specialization. Unpublished, University of Bologna. amsacta.cib.unibo.it/2683/.

Guimarães, P., O. Figueiredo, and D. Woodward. 2009. Dartboard Tests for the Location Quotient. *Regional Science and Urban Economics*, 39(3): 360–364.

Holmes, T. J., and J. J. Stevens. 2004. Spatial Distribution of Economic Activities in North America. In J. V. Henderson and J. F. Thisse(ed.), *Handbook of Regional and Urban Economics*, volume 4 of *Handbook of Regional and Urban Economics*. Amsterdam: Elsevier.

Isserman, A. M., and J. Westervelt. 2006. 1.5 Million Missing Numbers: Overcoming Employment Suppression in County Business Patterns Data. *International Regional Science Review*, 29(3): 311–335.

Kominers, S. D. Measuring Agglomeration. 2007. Unpublished expository paper. www.scottkom.com/articles/measure_agglomeration.pdf.

Krugman, P. 1991. Increasing Returns and Economic Geography. *Journal of Political Economy* 99(3): 483-499.

Marcon, E., and F. Puech. 2003. Evaluating the Geographic Concentration of Industries Using Distance-based Methods. *Journal of Economic Geography* 3(4): 409–428.

Marcon, E., and F. Puech. 2009. Measures of the Geographic Concentration of Industries: Improving Distance-based Methods. *Journal of Economic Geography* 10(5):745–762.

Maurel, F., and B. Sedillot. 1999. A Measure of the Geographic Concentration in French Manufacturing Industries. *Regional Science and Urban Economics* 29(5):575–604.

Moineddin, R., J. Beyene, and E. Boyle. 2003. On the Location Quotient Confidence Interval. *Geographical Analysis* 35(3): 249–256.

Nakamura, R., and C. J. M. Paul. 2009. Measuring agglomeration. In R. Capello and P. Nijkamp (ed.), *Handbook of Regional Growth and Development Theories*, Chapter 16, pages 305–328. Northampton, MA : Edward Elgar Publishing.

O'Donoghue, D., and B. Gleave. 2004. A Note on Methods for Measuring Industrial Agglomeration. *Regional Studies*, 38(4): 419–427.

M. E. Porter. 1990. *Competitive Advantage of Nations*. New York: Free Press.

Rosenthal, S. 2001. The Determinants of Agglomeration. *Journal of Urban Economics* 50(2):191–229.