Evolution of Industrial Agglomeration and Its Causal Relation with Road Networks in the U.S.

Yena Song* · William P. Anderson** · T.R. Lakshmanan***

미국의 산업집적 추이와 도로교통망의 인과관계 분석

송예나* · William P. Anderson** · T.R. Lakshmanan***

Abstract : Industrial agglomeration is an old theme in economic geography and many studies have been devoted to this topic. But only few have empirically looked at the time trend of industrial agglomeration. This study measured agglomeration of U.S. industries over last 29 years and measurement results indicated that industrial clustering has occurred during the study period in all study industries without a common time trend shared amongst the study industries. The agglomeration levels then were plugged in to investigate causalities, i.e. causal relations, around industrial agglomeration. Three variables were selected to see causal relations with agglomeration levels based on literatures, and our focus was given to the causality between transport network and agglomeration. Causal relation from transport to agglomeration was found in various industries and this supports the argument that the development of transportation influences industrial agglomeration. At the same time inverse and bi-directional causalities were also revealed implying more complex relationship between these two.

Key Words : industrial agglomeration, geographical concentration, transportation networks, Granger causality

요약 : 산업집적은 경제지리학에서 오랫동안 연구되어온 주제 중의 하나이며 지금까지 다양한 방법을 통해 산업집적 현상을 설명하고 이의 영향을 평가해왔다. 하지만 시계열 데이터를 이용해 집적의 추이를 살펴본 연구는 아직 활발히 이루어지지 않고 있다. 본 연구는 지난 29년 간의 데이터를 이용해 사례로 선정된 산업들의 집적 정도를 평가하고 이러한 시계열 패턴과 도로 네트워크의 확장이 어떠한 인과관계를 맺고 있는 지 살펴보고 있다. 집적 정도를 측정한 결과 사례로 선정된 산업의 종사자들은 지리적으로 균등하게 분포되지 않고 있었다. 또한 사례 산업들 간에 공통된 시계열적 변화 특성은 나타나지 않았으나 각 산업의 발전 단계 및 비즈니스 환경 변화가 개별 산업의 집적 정도에 영향을 주고 있는 것으로 보이고 있다. 집적 정도와 도로교통망 사이의 관계를 살펴보기 위해 각 산업 별로 Granger causality test를 실시하였고 도로 교통망이 산업 집적에 영향을 주고 있음을 몇몇 사례를 통해 확인할 수 있었다. 그러나 이와 반대의 경우 혹은 상호 간에 영향을 주는 사례도 나타나 교통망과 산업집적이 보다 복잡한 관계를 맺고 있음을 보여준다.

주요어 : 산업집적, 지리적 집중, 교통 네트워크, Granger causality

* Transportation Research Group, University of Southampton, Southampton, SO17 1BJ, U.K., Y.Song@soton.ac.uk

** Cross-border Transportation Centre, University of Windsor, 401 Sunset Ave., Windsor, Ontario N9B 3P4, Canada, bander@uwindsor.ca

*** Department of Geography and Environment, Boston University, 675 Commonwealth Ave., Boston, MA 02215, U.S., laksh@bu.edu

1. Introduction

Ever since Marshall's pioneering work (1890), industrial agglomeration has been one of favorite themes of economic geography and also has been extensively studied. Moreover, recent studies (Piore and Sabel, 1984; Saxenian, 1994; Scott, 1998; Cooke and Morgan, 1998; Lee *et al.*, 2000) showed that industrial agglomeration encourages endogenous growth, which boosted the interest in this traditional subject.

Industrial agglomeration can be divided into two categories: agglomeration of firms and economic agents in the same or related industries and more diversified agglomeration of various types of entities. The first type of agglomeration provides localization economies, which is in the same line of Marshall's external economies, to firms in the agglomerated area. Localized agglomeration is what usually referred to industrial agglomeration and have been studied more extensively than the second type. Diversified agglomeration has its roots in Jacobs's study (1969). In Jacob's work, diversity is described as a fertilizer of regional economic growth via information and knowledge spillovers among various industries, which in turn, leads innovation. In this study, we focus on Marshallian agglomeration and hereinafter, the term, agglomeration refers only to localized agglomeration.

There are many studies dealing with agglomeration mechanisms. New economic geographers offer theoretically solid ground explaining how industrial clusters occur. They demonstrate how the interaction of scale economies on production side and a preference for variety on consumption side leads to the spatial concentration of economic activities in the form of general equilibrium models (Fujita *et al.*, 2001). In this process, the level of transportation cost, regional wage gaps and strength of industrial linkages affect the speed and location of industrial agglomeration (Puga and Venables, 1996; Hanson, 1998; Venables and Gasiorek, 1999; Fujita *et al.*, 2001). Mobile labor tends to move to the regions that compensate them more. So the larger the wage variation between regions is, the more laborers move to where the wage is higher. At the same time, companies in agglomerated areas should pay higher wages to compensate congestion costs which include the living cost such as rent. This side of story appears to be a self-reinforcing circle: higher income attracts mobile labors causing congestion and thus workers demand higher wage for the increased living cost. Industrial specific characteristics and linkage among various industries determine the order and the speed of agglomeration. Puga and Venables (1996) showed how and what industries move out of cluster to other places by simulations. According to their simulation results, labor intensive and weakly linked industries tend to leave agglomerated areas first to save production cost. Transportation is a key to moving labor and/or products. Low transport costs make it cheaper to produce more than locally needed and trade to other places, which results in industrial agglomeration. On the other hand, it allows diffusion of production activities to new places and reduces price gap caused by transport cost. Thus transport cost approaches close to zero, stability of agglomeration could be broken due to weakened home market effects, and production activities spread out. Along with economic theory, Lakshmanan and Anderson (2007) argued that well-developed transportation system contributes to economic growth in various ways, such as market expansion, regional specialization, industrial clustering, and innovation process. Their study addresses the positive relationship between transportation networks and industrial agglomeration. On the other hand, Boarnet (1998) and Chandra and Thompson (2000) claimed that extending transportation networks does not always mean economic benefits. They empirically showed that developed transport services could make some regions worse off and also dissolve the existing clusters.

This study seeks to find further evidence that would prove or disapprove suggested from the literature with very basic questions: is there industrial agglomeration in reality?; if there is agglomeration in a certain industry, how intensive is it, and how has it changed over time?; does transport infrastructure have causal relation with the industrial agglomeration? To answer these questions it is essential to track down temporal evolution of industrial agglomeration. Only small number of agglomeration studies has actually used time series data set (Stephanedes and Eagles, 1987; Kim, 1995; Jiwattanakulpaisarn, 2009) whilst many have dealt with cross sectional data (Ellison and Glaeser, 1997; Barkley et. al., 2001; Duranton and Overman2002; Spieza, 2003; Sohn, 2004). We investigate how the intensity of industrial agglomeration has changed over time and the role of transportation networks behind such temporal trend will be explored. Though there are many explanations for the underlying reasons of agglomeration, transport's role will be emphasized here because without transportation system, agglomeration is not feasible at all.

2. Case study

1) Study Area, Period, and Data

This study is on the subject of industrial agglomeration, in other words, industrial clustering in the U.S. A cluster refers to regions showing the same or similar property to adjacent regions. Therefore it would not be proper to include islands or any regions apart from mainland. For that reason, we focus on the continental U.S., which excludes states of Hawaii and Alaska and thus 48 states and District of Columbia constitute the study area.

Many national level analyses have been conducted at the state level and agglomeration studies have been no exception. However, several studies have shown that we may find different result by adopting more detailed spatial level dataset. For example, Ellison and Gleaser (1997) pointed out that their localization measure fell when they used county data instead of state data, which is due to Modifiable Areal Unit Problem (MAUP). There are two sources of MAUP, aggregation and zoning (Reynolds, 1998), and zoning effect cannot be avoided or reduced when discrete spatial data set is used. But finer data may diminish the problem caused by aggregation. County is the finest geographic unit that has aggregate employment and establishment data. The Census Bureau publishes zip-code level data from 1998 but it does not include industrial employment information. So the spatial unit used here for measuring concentration level is the county and there are 3079 counties in our data set.

This study covers a period from 1977 to 2005. County level industrial data have been annually published by Census Bureau since 1964 and at irregular intervals back to 1946. However, earlier data can be obtained in printed form, which makes hard to use them in analysis. Digitized data exists only from 1977 so the study period begins in 1977 covering around 30 years, which would be a sufficient length of time for a time series analysis.

An agglomeration study requires data on economic activities for each regional unit. This study uses county level employment and wage data published by the Census Bureau each year in the form of County Business Patterns (CBP). CBP is a geographically detailed data set, and it also provides even zip code level information for the recent years. However, such finesse could unintentionally reveal the personal/ firm or establishment level information. So Census Bureau withholds data in cases when only 1 or 2 establishments exist in a certain county. Missing values that were undisclosed due to the purpose of information protection account for approximately 32% of the whole data set. When the employment data of a county is not disclosed, the wage data is also left missing. Since missing cells form around 1/3 of employment and wage data, these should not be ignored. To estimate the suppressed employment data, we use establishment size data given for each county. CBP reports the number of establishments and the total employment in establishments belonging to 9 employment size categories[1)] for each industry. Even though total number of employment is not revealed, Census Bureau reports the number of establishments for each category. Thus we assign the mean employment values of U.S. for 9 categories and make estimation of the employment size. To check how well the estimation works, we compare the estimated employment size and employment ranges[2)] given by the Census Bureau, and over 90% of our estimations fit into their given employment ranges. One of the biggest problems of estimating wage data is that wage level varies not only by establishment size but by region, e.g. metropolitan area's average wage is normally higher than that of a rural area. To alleviate this issue, missing wage cells are firstly estimated following the same procedure as employment estimation, but one more step is added, weighting by geography. A weight matrix was developed to reflect each county's annual mean relative wage and it is, then, multiplied by estimated wages in the previous step.

Finally transportation data come from Highway Statistics published by Federal Highway Administration. It contains several components useful in time series analysis, such as highway mileage, local road and street mileage, and annual disbursement in highway system.

2) Selection of Industries

At the county level, the finest industrial category is 4 digit in SIC or 6 digit in NAICS systems. However, the combination of fine area and fine industry results in poor data quality as, already noted, Census Bureau withholds data if there is a danger of identity disclosure of individual firms. Thereby we consider SIC-2-digit or NAICS-3-digit to be the finest industrial classification level with reliable data.

To select study industries, first, general growth pattern in terms of industry GDP, employment, and productivity levels are examined. This works as a prescreening step and informs us which industries show more dynamic growth patterns. Second, industries that produce transferable goods or services are only considered. If only local customers consumed goods or services produced in a certain region, such industries could not be good candidates as agglomeration as clustering indirectly means a higher level of production activity at certain region than is locally needed. For instance, 'Eating and drinking places (SIC 58)' is not a good choice while 'Business Services (SIC 73)' has possibility to be selected. Then, information from cluster theories is used to determine the final candidates of the study. New economic geography theory considers differentiated goods, increasing returns and transport costs (Fujita *et al.*, 2001) to be crucial factors in agglomeration process whilst innovation literatures value knowledge intensive sectors (Stuart and Sorenson, 2003; Storper, 1997; Saxenian, 1994; Scott, 1993; Piore and Sabel, 1984).

Based on these criteria, three sectors from manufacturing and four sectors from service industry are selected: 'Apparel and other textile products (SIC 23)', 'Chemicals and allied products (SIC 28)', 'Electrical and electronic equipment (SIC 36)', 'Security, commodity brokers, and services (SIC 62)', 'Insurance carriers, agents, brokers, and services (SIC 63, 64)', 'Business services (SIC 73)', and 'Legal services (SIC 81)'. Also the products or services produced by those industries easily can be transferred to other areas.

3. Concentration Measurement

There are various measures that calculate geographic concentration level. Concentration measures compare the observed distribution pattern against a hypothetic distribution representing the absence of spatial concentration. In this study localization index developed by Ellison and Gleaser (1997) is used to assess the geographical concentration level. This index has been employed in recent agglomeration studies (Duranton and Overman, 2002; Holmes and Stevens, 2002; and Santa Marĭa *et al.*, 2005) due to its sound grounding as well as its different definition of concentration. It distinguishes any potential agglomeration from random distribution. So it is different from the traditional geographical concentration measures whose no-concentration condition refers to uniform distribution. Ellison and Glaeser (1997) illustrated this measure's superiority with an example of vacuum industry: four plants employed 75% of the workforce of this industry, and then four locations must have accounted for at least 75% of the employment even if they were located separately. Traditional concentration measures would find this industry was heavily concentrated but their new index can distinguish unevenness from localization and may find it random.

The localization index is expressed as below equation (1).

$$\gamma \equiv \frac{\sum_{i=1}^{M}(s_i - x_i)^2 - (1-\sum_{i=1}^{M} x_i^2)^2 - \sum_{j=1}^{N} z_j^2}{(1-\sum_{i=1}^{M} x_i^2)(1-\sum_{j=1}^{N} z_j^2)} \qquad (1)$$

where s_i is the share of an industry's employment in each of M geographic areas, the share x_i of total employment in each of area and means and $\sum_{j=1}^{N} z_j^2$ is the Hirschman-Herfindahl index of the industry plant size distribution[3].

The higher γ means more severe geographical concentration in an industry. This index is very good comparison tool across industries and over time, but in absolute term there is no critical point above which we can say that there exists geographical concentration. But if the index is over 0, we can at least say there is agglomeration at some level.

Estimation results are shown in Figures 1 and 2, which cover 29 years of geographical agglomeration levels of 7 study industries. General findings can be summarized as following two points. First, we can conclude that 7 study industries have been geographically agglomerated during study period with some variations by industry and measurement method. In other words, study industries have not been geographically distributed as general employment pattern. Second, there was no unanimous time trend of agglomeration, which indicates that industry-specific factors or reactions to common factors might have affected agglomeration trend rather than any universal power.

Although we notice that concentration measures suggest different time trend of industrial agglomeration in detail, apparel and finance sectors had very distinctive and impressive time trends that appeared common in all three measures: increase in apparel's agglomeration level and decline of agglomeration intensity in finance sector. Such an opposite concentration pattern could be explained by industrial expansion. During the study period, finance industry has grown and geographically diffused very rapidly and this fast falling concentration level reflects its industrial development cycle. To add more specifics, the financial

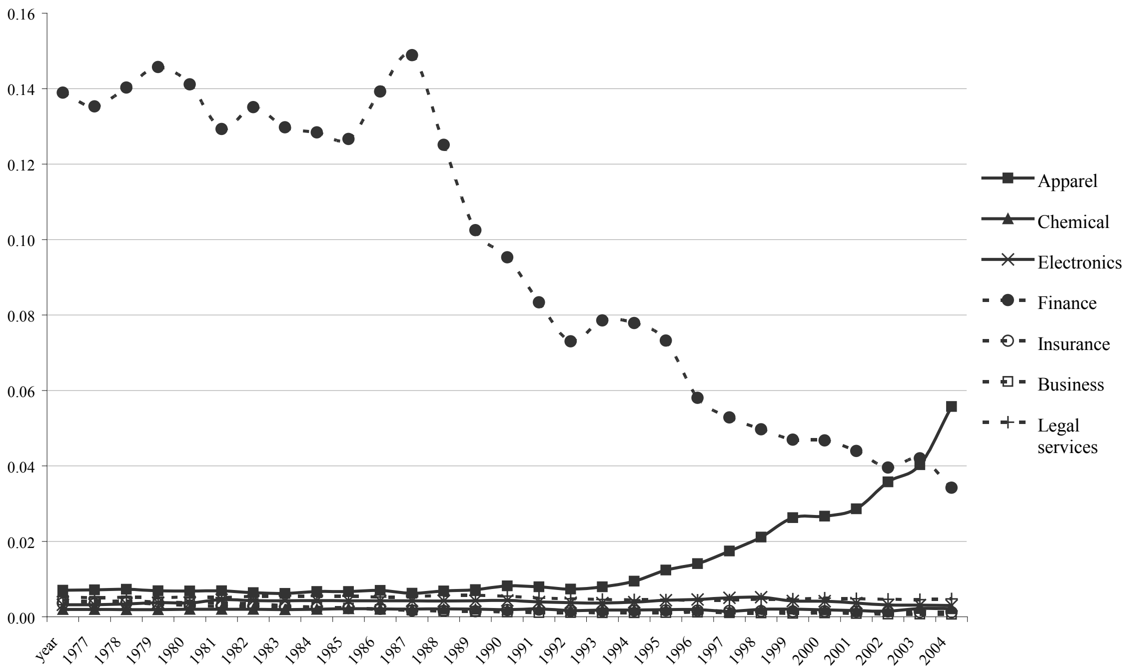

Figure 1. Estimated index of localization

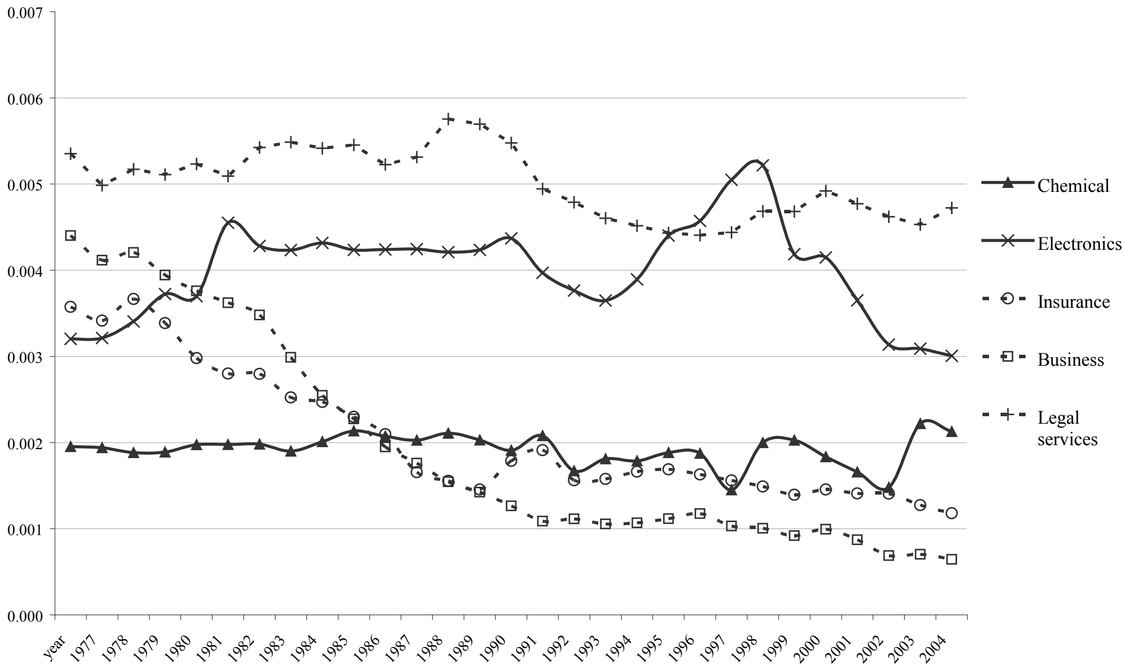

Figure 2. Estimated index of localization except apparel and finance

industry began to geographically diffuse from a few metropolitan areas to regional centers. In 1977, only 1005 among 3079 counties had at least one employee in financial industry within their boundaries and most of them were in or near big cities like New York, Boston, and Chicago. At the end of study period, the number of counties with financial workers has doubled: 2339 counties had employment in this sector. On the other hand, apparel has reduced its size in terms of employment, establishment, and production outsourcing many of its functions, especially mass production part. This results in loss of workforce and establishments, and finally, appears as a higher level of agglomeration as employment remains in only a few regions. Agglomeration trends of these two industries imply that industrial development cycle or structural change can have serious impacts on industrial clustering.

4. Causality between Transportation and Agglomeration

1) Methodology

Whether transportation networks have really mattered in industrial clustering is the issue being tested here. As transportation factor, one of two measures, interstate highway mileage or public road mileage, is applied in empirical models[4). These two are all indicative of transport networks properties but they can be interpreted in slightly different manners. Interstate highway mileage stands for national transport connectivity whilst public road for the smaller scale. Though the focus is the causal relation between transportation and agglomeration variables, two more factors, industrial wage variation and industrial GDP, are considered to minimize the bias from omitted variables[5). Literature indicates that the former could facilitate mobile workers to move and at the same time could affect location decision of industries. Industrial GDP is of interest for two reasons: first, it was found that concentration levels can be significantly associated with industrial expansion or shrinkage in previous section; and agglomeration brings scale economy and could lead to better productivity.

Our presumed causal relationships among mentioned variables and agglomeration are depicted in Figure 3.

To investigate the causal relations between the industrial agglomeration and transportation networks, Granger causality (Granger, 1969) is tested for each study industry. This test begins from an intuitive idea that a cause cannot come after the effect (Sturn, 1998). So if a variable improved the prediction of the other, then it implies that the former caused the latter.

Granger causality test is performed in the form of Vector Auto Regression (VAR) systems. There are two

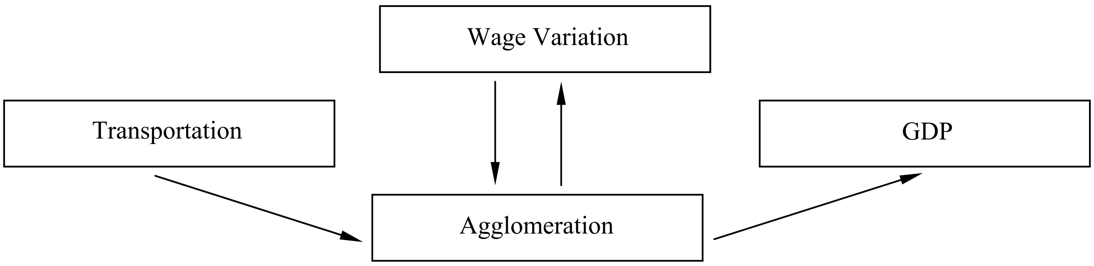

Figure 3. Presumed causal relations

advantages of introducing VAR model in Granger causality test: the simultaneity effects can be avoided and at the same time, no a priori causality directions imposed (Strum, 1998). In other words, unlike other statistical test, Granger causality test allows the system to figure out statistically significant causalities from all possibilities without a priori assumption posed by researchers. A k-th order VAR model for the variable vector y is

$$y_t = c + \sum_{p=1}^{k} \prod\nolimits_p y_{t-p} + \varepsilon_t \tag{2}$$

where c is an m×1 vector of unknown deterministic terms, $\prod_p$ is m×m matrix of unknown parameters.

Granger causality is determined by testing the assumption that the coefficients of lagged variables are equal to zero. Suppose a s-th order VAR system only with two variable, x and z. Then this system has two equations as below.

$$x_t = a_{10} + \sum_{p=1}^{s} \o_p x_{t-p} + \sum_{p=1}^{s} \pi_p z_{t-p} + \varepsilon_{1t}$$

$$z_t = a_{20} + \sum_{p=1}^{s} \eta_p x_{t-p} + \sum_{p=1}^{s} \mu_p z_{t-p} + \varepsilon_{2t} \tag{3}$$

In this imaginary system, x does not Granger cause z if all η_p are not significantly different from zero. Similarly, z does not Granger cause x only when all coefficients of lagged z are zero. As such Granger causality test is very intuitively attractive and easy to perform. However, this statistical measure can lead to the ecological fallacy, in which causal relation is given simply due to the sequential realization of two events.

VAR assumes that included variables are stationary in its system. Since time series data are in use, the stationary of selected variables cannot be guaranteed. Augmented Dickey-Fuller (ADF) test is employed to examine the presence of unit roots in the level of variables as well as in their first differences. The variables are transformed to natural logarithmic form and the optimal lag lengths[6)] are determined by the Akaike Information Criterion (AIC).

Table 1 shows that most of series are not stationary as they are, but they are stationary in first differences in 95% significance level, i.e. integrated in order 1. Since the variables are identified to be integrated in order 1, we difference VAR equation and move to Vector Error Correction Model (VECM) that use the variables in the first-differenced form and include additional Error Correction (EC) term as shown in equation (4).

$$\Delta y_t = c + \sum_{p=1}^{k-1} \Gamma_p \Delta y_{t-p} + \Phi y_{t-1} + \varepsilon_t \tag{4}$$

where Γ_p and y_t are coefficient matrices (Enders, 2004).

If all elements of y_t are stationary, Φ is a full rank of m×m matrix. If Φ matrix's rank is 0, then the standard VAR model is appropriate in first differences. If the rank of Φ is not zero, the system will have cointegrating relationships as many as the rank of Φ and ignoring the cointegration relationships causes mis-specified VAR model (Johansen, 1988).

The lag length in VECM is chosen on the basis of likelihood ratio proposed by Sims (1980), and in this estimation 95% significance level is applied. To establish the cointegration relationship among variables, Johansen's (1988) reduced rank method is used and the critical values are calculated by the procedure developed by Osterwald-Lenum (1992). In this test, a model with constant and no trend is applied. The rank of long-run coefficients matrix determines the number of cointegrating relationships, and the results from these tests on each set of VECM are below Table 2.

Table 2 shows the lags and number of cointegrating relations determined by maximal eigenvalue and trace statistics respectively. We find the same number

Table 1. Unit root test results

				First difference	
		ADF t-value	Lags	ADF t-value	Lags
Apparel	Localization index	-0.1695	3	-4.5045***	3
	Wage variance	-2.8138	2	-4.3379**	3
	GDP	-1.1493	2	-6.6921***	1
Chemical	Localization index	-3.8238**	1	-5.1867***	1
	Wage variance	-2.7214	1	-6.1547***	1
	GDP	-3.3542*	2	-6.2054***	2
Electronics	Localization index	-2.1529	3	-4.7127***	3
	Wage variance	-2.9856	1	-5.5830***	1
	GDP	-1.0528	1	-4.5683***	3
Finance	Localization index	-2.3352	1	-3.9525**	1
	Wage variance	-1.5140	3	-5.1963***	1
	GDP	-1.1796	1	-5.0960***	2
Insurance	Localization index	-2.1053	1	-4.5304***	1
	Wage variance	-2.0662	1	-4.0208**	1
	GDP	-1.5786	2	-6.1863***	3
Business services	Localization index	-2.0581	3	-6.6199***	1
	Wage variance	-2.8501	2	-5.3059***	1
	GDP	-3.3219*	1	-4.5260***	2
Legal services	Localization index	-2.4404	2	-4.2201**	2
	Wage variance	-1.6336	1	-4.9974***	2
	GDP	-2.3835	2	-4.2715***	2
Transportation	Interstate highway mileage	-1.6226	1	-5.0585***	1
	Public road mileage	-2.5851	2	-6.9924***	1

*, **, and *** indicate that the null hypothesis of a unit root is rejected at 10%, 5%, and 1% significance level respectively.

Table 2. Johansen's reduced rank method results*

Industries	Interstate highway mileage			Public road mileage		
	lags	$\lambda_{\max}$	λ_{trace}	lags	$\lambda_{\max}$	λ_{trace}
Apparel	3	2	1	3	2	1
Chemical	3	1	0	2	0	0
Electronics	3	1	1	3	0	0
Finance	3	1	0	1	0	0
Insurance	3	2	2	3	1	1
Business services	3	1	1	3	2	1
Legal services	3	2	2	3	2	2

* 95% confidence level is applied in determining the number of cointegrating relations.

of cointegrating vectors by two different statistics in most of cases but sometimes different inference is possible. In such case, the minimal number is applied to VECM.

The results from Sims test and Johansen's rank method indicate that it would be better to develop a customized model for each industry-transportation variable pair. The numbers of lags and cointegrating relations are different by industry-transportation pair and some of them have no cointegrating relations at all indicating no long-term relationship. If cointegrating relationship was not found, VAR using first difference is estimated instead of VECM.

As VECM is in use, we can identify three sources of causation from the right-hand side variables to the dependent variable. Testing the significance of the lagged variables in equation (3) reveals short-run causality as the dependent variable responds to the short-term shocks (Asafu-Adjaye, 2000). Long-run causality can be found when the coefficient of the EC terms, Φ in equation (4) is significant, showing how quickly the deviations from the long-run equilibrium are eliminated. Finally, there is a joint test of these two sources of causation and it is often referred to as a strong Granger causality test. This test tells which variables bear burden of short-run adjustment to catch up the long-run equilibrium (Oh and Lee, 2004). All these three sources of causation between transportation variable and agglomeration are tested in this study and results are reported in following section.

2) Results

Table 3 presents the Granger causality test results between industrial agglomeration and transportation network variables, i.e. interstate highway mileage and public road mileage. Particular caution is required in interpreting the causal relations provided in Table 3. As noted, Granger causality is based on the time sequence of events, not on actual causality. Therefore without reasonable or scientifically solid explanation, provided Granger causality does not imply an actual causal relationship.

In general different industries have different causal relations between transportation networks and ag-

Table 3. Granger causality test results

		Interstate Highway		Public road	
Industry	Dependent V.	Aggl.	Trans.	Aggl.	Trans.
Apparel	*Short-run causality (Wald test)*				
	Aggl.		18.75***		1.56
	Trans.	1.32		3.83	
	Long-run causality (z-score)				
	ECTs	-1.95*	1.69*	-2.8***	0.51
	Granger endogeneity (joint Wald test)				
	Aggl.		19.3***		1.64
	Trans.	4.33		12.84**	
	R^2	0.8185	0.9052	0.8833	0.5389
Chemical	*Short-run causality (Wald test)*				
	Aggl.		5.565		0.645
	Trans.	3.998		3.503*	
	R^2	0.6883	0.5442	0.3306	0.0682

		Interstate Highway		Public road	
Industry	Dependent V.	Aggl.	Trans.	Aggl.	Trans.
Electronics	*Short-run causality* (*Wald test*)				
	Aggl.		27.78***		5.915
	Trans.	8.67**		2.40	
	Long-run causality (*z-score*)				
	ECTs	-1.75*	-9***		
	Granger endogeneity (*joint Wald test*)				
	Aggl.		88.40***		
	Trans.	13.39***			
	R^2	0.7717	0.9754	0.5352	0.4630
Finance	*Short-run causality* (*Wald test*)				
	Aggl.		0.292		0.073
	Trans.	0.543		0.078	
	R^2	0.1961	0.1173	0.1823	0.0955
Insurance	*Short-run causality* (*Wald test*)				
	Aggl.		8.99**		1.61
	Trans.	17.21***		0.94	
	Long-run causality (*coefficient & z-score*)				
	ECTs (no wage)	-3.92***	-0.72	-1.85*	-1.43
	(no agglomeration)	-2.92***	-1.72*		
	Granger endogeneity (*joint Wald test*)				
	Aggl.		94.5***		3.46
	Trans.	69.93***		4.97	
	R^2	0.9680	0.9841	0.8072	0.5501
Business Services	*Short-run causality* (*Wald test*)				
	Aggl.		4.25		3.71
	Trans.	2.02		16.18***	
	Long-run causality (*z-score*)				
	ECTs	0.82	-4.7***	-4.33***	-0.09
	Granger endogeneity (*joint Wald test*)				
	Aggl.		57.07***		3.87
	Trans.	2.11		19.81***	
	R^2	0.8059	0.9460	0.9174	0.5410
Legal Services	*Short-run causality* (*Wald test*)				
	Aggl.		17.74***		2.31
	Trans.	5.47		3.89	
	Long-run causality (*z-score*)				
	ECTs (no wage)	-2.7***	-3.42***	-1.52	0.68
	(no agglomeration)	-2.48**	-3.9***	1.08	0.44
	Granger endogeneity (*joint Wald test*)				
	Aggl.		23.73***		12.91**
	Trans.	11.13**		11.80**	
	R^2	0.7535	0.9545	0.7614	0.7257

*, **, and *** indicate that the null hypothesis of no-causality is rejected at 10%, 5%, and 1% significance level respectively.

glomeration levels and different transport infrastructures lead to different causal relationship even in the same industry. Also short-run and long-run causalities do not necessarily correspond to each other. Two transportation networks, interstate highway and public road show different causalities between two key variables and the interstate highway system appears to have more causal association with industrial agglomeration than the local road networks in either ways.

The test results provide support for the argument that the development of transportation has impact on industrial agglomeration. Such a relationship is found between Apparel, Chemical, Insurance, and Business service industries' agglomeration levels and the public road network. Interestingly when the interstate highway is used in the model, opposite causality is found in the Business Service industry. This clearly indicates that the coverage and nature of the transportation networks could lead a different association with the industrial agglomeration. The test results of the Business Services can be interpreted as following: enhanced local transport network would lead same number of industries has opposite direction of causality. This might mean that causal relation between transportation infrastructure and agglomeration is more complex and agglomeration could also Granger-cause the change of transport infrastructure. Possible explanation for this inverse relationship is that industrial agglomeration caused congestion and therefore pressure for the expansion of road networks or it is simply due to realization timing. Bidirectional causality is also found in a few industries such as Apparel, Electronics, Insurance and Legal Services. Two-way Granger causality between these two key variables indicates possible self-reinforcing relations in certain industries, i.e. better transport networks attract more workers and more business, and at the same time, active economic activities require better transportation networks. Lastly it also needs to be noted that Finance industry showed no causal association with either transportation network. This may be because this sector began to develop most recently and has developed rapidly in a short period, which could have caused significantly different concentration trend from other industries and thus different relationships with the other key variable, i.e. transportation networks, in the system.

5. Conclusion

In this study, geographical concentration levels over last 29 years were calculated and displayed using a localization index. This index revealed that industries in this case study have been agglomerated at some levels with variation over time and across industries. Among seven sectors, apparel and finance presented opposite concentration pattern that could be explained by industrial development cycles and the change in business environment. Using one of concentration measures, the causal relations between agglomeration and transport were investigated by industry. Under VAR framework, we employed four variables, i.e. industrial agglomeration level, transportation networks, wage variation and industrial GDP, that are identified in the literature as having causal relationships, and found support for the theories in spite of variation across industries.

This study deals with a traditional topic in economic geography and used popular measures in its analysis. But this study can differentiate itself from others in a few points. We measured concentration levels of about 30 years using county data set. As noted earlier, there has been no effort to measure geographical concentration in time series manner since Kim's study and his study was based on state/census division level data set. Such lack of agglomeration studies using time series data is mainly because of data limitation and hardship in data processing. We showed also causal relationships among transport, agglomeration, and industrial variables. New economic geographers demonstrate theoretically how transport and other factors affect agglomeration, but there has been little analysis of those relationships using real data. There are a couple of studies dealing with causality between infrastructure investment and economic developments (Sturm, 1998; Peterson and Jessup, 2008), but neither of them explicitly used agglomeration in their variables.

Although many meaningful results were found through our analysis, there is significant scope for further research. First of all, most agglomeration studies start from the discrete space, such as county, state, or country, and this study is no exception. However, discrete space studies are vulnerable to the modifiable areal unit problem (MAUP) as many geographers have pointed out. Duranton and Overman (2002) tried to solve this problem by applying K-density function on point data of U.K. With point data accessible through Census Research Data Center program, the same or advanced methodology can be applied to U.S. industries and the result from it would be a good comparison with our result.Second, we have studied industrial agglomeration and its relationship to transportation infrastructure from 1977. The interstate highway system had embarked mid-1950 and by 1977 it was almost in its full shape, as was the public road network. In fact, interstate highway system was extended only 11% and public road only 3.3% during the study period. Since transport infrastructure has diminishing returns over time (Nadiri and Marmeneus, 1996), this study may underestimate the impact of transportation. This, in turn, calls for a future study with extended time span preferably before 1950.

Acknowledgement

The study data was initially collected and cleaned as a part of the lead author's Ph.D. dissertation. However, authors carried out time series analysis again employing more sophisticated methodology in preparation of this manuscript and this paper is based upon new analysis results.

Notes

1) The categories are classified according to the employment size of each establishment. 9 groups are 1-4, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500-999, 1000 or more. For each category, total number of establishment and total employment are reported in a national level data set.

2) There are 12 employment size classes in CBP reports and details are as following. A: 0-19, B: 20-99, C: 100-249, E: 250-499, F: 500-999, G: 1000-2499, H: 2500-4999, I: 5000-9999, J: 10000-2499, K: 25000-49999, L: 50000-99999, M: 100000 or more.

3) This index is HHI in original context measuring market competition. To calculate this index, we need information of each establishment size, which was not accessible. To get a proxy for each industry, categorized establishment and employment data from Census Bureau are used and it is assumed that the sizes of the establishments within each class are uniformly distributed on a range centered on the mean with its boundary as the rule of thumb proven by Schmalensee (Schmalensee, 1977). Then each establishment was assigned by a number of employees produced randomly but according to the assumption. This procedure may not produce the exact HHI, but it would estimate the plausible HHI given the limited data set.

4) We focus on the road networks in this study although other transportation related factors such as containerization, adoption of multi-modal transport system, and decline of railway networks could be closed related to the industrial agglomeration. These factors were excluded from the analysis because existing data cover short time span or are sporadic and it was better to keep the models simple if we could construct valid models. Still the potential role of other

transport factors should not be under-emphasized and may need to be explored in other context.

5) Omitted variables in the VAR system hinder finding cointegrating relation among the variables within which cointegrating relationship indeed exists and thus long-run relationship should be found. We initially tried to develop bi-variable models using only transportation and agglomeration variables and rarely found cointegrating relationship in our VECM framework. Thus we tested various possibilities and concluded two more variables needed to be included in the system to make valid models.

6) We have 29 observations in this time-series data and the common rule of determining the maximum number of lag length is $T^{1/3}$, where T is the number of observations.

Reference

Asafu-Adjaye, J., 2000, The relationship between energy consumption, energy prices and economic growth: time series evidence from Asian developing countries, *Energy Economics*, 22(6), 615-625.

Aschauer, D.A., 1989, Is public expenditure productive?, *Journal of Monetary Economics*, 23(2), 177-200.

Boarnet, M.G., 1998, Spillovers and the locational effects of public infrastructure, *Journal of Regional Science*, 38(3), 381-400.

Chandra, A., and Thompson, E., 2000, Does public infrastructure affect economic activity? Evidence from the rural interstate highway system, *Regional Science and Urban Economics*, 30(4), 457-490.

Cooke, P., and Morgan, K., 1998, *The associational economy*, Oxford University Press, Oxford.

Duranton, G., and Overman, H.G., 2002, Testing for localization using micro-geographic data, *Review of Economic Studies*, 72(4), 1077-1106.

Eagle, D., and Stephanedes, Y.J., 1987, Dynamic highway impacts on regional employment, *Transportation Research Record*, 1116, 56-62.

Ellison, G., and Glaeser, E.L., 1997, Geographic concentration in U.S. manufacturing industries: A dartboard approach, *Journal of Political Economy*, 105(5), 889-927.

Enders, W., 2004, *Applied econometric time series,* Wiley, Hoboken, NJ.

Fujita, M., Krugman, P., and Venables, A.J., 2001, *The spatial economy*, MIT Press, Cambridge, MA.

Granger, C., 1969, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, 37(3), 424-438.

Granovetter, M., 1985, Economic action and social structure: the problem of embeddedness, *American Journal of Sociology*, 91(3), 481-510.

Hanson, H., 1998, Market potential, increasing returns, and geographic concentration, *NBER Working Paper*, 6429.

Holmes, T.J., and Stevens, J.J., 2002, Geographic concentration and establishment scale, *The Review of Economics and Statistics*, 84(2), 682-690.

Jacobs, J., 1969, *The economy of cities*, Random House, New York, NY.

Jiwattanakulpaisarn, P., Noland, R.b., Graham, D.J., and Polak, J.W., 2009, Highway infrastructure and state-level employment: A causal spatial analysis, *Papers in Regional Science*, 88(1), 133-160.

Johansen, S., 1988, Statistical analysis of cointegration vectors, *Journal of economic dynamics and control*, 12(2-3), 231-254.

Kim, S., 1995, Expansion of markets and the geographic distribution of economic activities: The trends in U.S. regional manufacturing structure, 1860-1987, *The Quarterly Journal of Economics*, 110(4), 881-908.

Lakshmanan, T.R., and Anderson, W.P., 2007, *Transport's role in regional integration processes*, Paper presented at the OECD Regional Round Table 134, OECD-ECMT, Paris.

Lee, J., Miller, W.F., Hancock, M.G., and Rowen, H.S., 2000, *Silicon Valley Edge*, Stanford University Press, Stanford, CA.

Marshall, A., 1890. *Principles of economics*, Macmillan, London.

Nadiri, I.M., and Manuneas, T.P., 1996, *Constitution of highway capital to industry and national productiv-*

ity groups, Report prepared for FHWA.

Oh, W., Lee, K., 2004, Causal relationship between energy consumption and GDP revisited: the case of Korea 1970-1990, *Energy Economics*, 26(1), 51-59.

Osterwald-Lenum, M., 1992, A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics, *Oxford Bulletin of Economics and Statistics*, 54(3), 461-472.

Peterson, S.K., and Jessup, E.L., 2008, Evaluating the relationship between transportation infrastructure and economic activity: evidence from Washington state, *Journal of the Transportation Research Forum*, 47(2), 21-39.

Piore, M., and Sabel, C., 1984, *The Second Industrial Divide: Possibilities for Prosperity*, Basic Books, New York, NY.

Puga, D., and Venables, A.J., 1996, The spread of industry: spatial agglomeration and economic development, *Journal of the Japanese and International Economies*, 10(4), 440-464.

Reynolds, H.D., 1998, *The modifiable area unit problem: Empirical analysis by statistical simulation.* Ph.D. thesis, University of Toronto.

Santa Marĭa, M.J., Pĕrez, J.M.G., and Olivares, A.F., 2005, *The industrial location in Spain: New methods for measuring industrial agglomeration*, Paper presented at 45th European Congress of the European Regional Science Association, Amsterdam, Netherlands.

Schmalensee, R., 1977, Using the H-index of concentration with published data, *Review of Economics and Statistics*, 59(2), 186-193.

Scott, A.J., 1993, *Technopolis: High-technology industry and regional development in Southern California*, University of California Press, Berkeley, CA.

Scott, A.J., 1998, *Regions and the world economy*, Oxford University Press, Oxford.

Sims, C.A., 1980, Macroeconomics and reality, *Econometrica*, 48(1), 1-48.

Sohn, J., 2004, Do birds of a feather flock together?: Economic linkage and geographic proximity, *The Annals of Regional Science*, 38(1), 47-73.

Spiezia, V., 2003, *Geographic concentration of production and unemployment in OECD countries*, The Journal of Cities and Regions (available online: http://www.oecd.org/dataoecd/43/0/15179780.doc).

Strum, J.E., 1998, *Public Capital Expenditure in OECD Countries: the Causes and Impact of the Decline in Public Capital Spending*, Edward Elgar Publishing, Cheltenham.

Stuart, T., and Sorenson, O., 2003, The geography of opportunity: Spatial heterogeneity in founding rates and the performance of biotechnology firms, *Research Policy*, 32(2), 229-253.

Tatom, J.A., 1991, Public capital and private sector performance, *Federal Reserve Bank of St. Louis Review*, 73(3), 3-15.

Venables, A., and Gasiorek, M., 1999, *The welfare implications of transport improvements in the presence of market failure: The incidence of imperfect competition in UK sectors and regions*, SACTRA Report.

Corresponding Address: Yena Song, Transportation Research Group, University of Southampton, Southampton, SO17 1BJ, U.K., E-mail: Y.Song@soton.ac.uk, phone: +44 (0)23 8059 2834, Fax: +44-(0)23 8059 3152

교신: 송예나, 사우스햄튼 대학교, 교통연구그룹, 사우스햄튼, SO17 1BJ, 영국, e-mail: Y.Song@soton.ac.uk, phone: +44 (0)23 8059 2834, Fax: +44 (0)23 8059 3152

Recieved January 29, 2013
Revised February 15, 2013
Accepted February 21, 2013