

Testing the Role of Input-Output Linkages in Industry Agglomeration with the Spatial Econometric and Input-Output Model

Zheng Tian
Ph.D. Candidate
Department of Economics
Regional Research Institute
West Virginia University
886 Chestnut Ridge Road
Morgantown, WV 26506-6825
Phone: 304-685-9429
zheng.tian@mail.wvu.edu

Abstract

This paper examines the relationship between input-output linkages and industry agglomeration by combining the strength of spatial econometrics and the input-output model. This paper uses the location quotient as a measure of industrial agglomeration and regresses it on the intermediate demand variable, which is an important variable in the integrated econometric and input-output models. The nature of the dataset and the construction of key variables are consistent with the use of a spatial Durbin model, estimated by “borrowing” spatial panel model techniques. With the model labeled as the spatial pseudo-panel industry fixed-effects Durbin model, this paper estimates it with the maximum likelihood method using manufacturing industries with 3-digit NAICS codes at the county level in the U.S. The results confirm the importance of input-output linkages in promoting the formation of industry agglomeration in a region.

Keywords: industry agglomeration, input-output linkages, the integrated spatial econometrics and input-output model

JEL Classification: C21, C33, R12, R15

1 Introduction

Industry agglomeration¹ is an economic issue with two facets. First, it is a phenomenon of economic geography. Industry activities are spread unevenly in geographic space, with some regions attracting a large variety of industries but some regions losing industries. Second, industry agglomeration is a phenomenon of industrial organization. The interactions between industries through input-output linkages encourage interrelated industries to cluster in some regions and promote regional economic growth (Porter, 1990, 1998). These two facets call for research that can incorporate both spatial and industry dimensions to study industry agglomeration.

This paper uses an integrated spatial econometric and input-output model to meet this demand. The integrated spatial econometric and input-output model combines the strengths of spatial econometrics and the input-output model. Spatial econometrics is capable of accounting for spatial spillover effects between regions, while the input-output model is useful to describe in detail the input-output linkages between industries. This paper combines spatial econometrics and the input-output model using the embedding strategy (Rey, 2000) through the spatial intermediate demand variable (Rey and Dev, 1997; Rey, 2000).

The nature of the dataset and the construction of the spatial intermediate demand variable suggest using a spatial Durbin model, estimated by “borrowing” the spatial panel time fixed-effects model techniques. The term “borrowing” means that the model is not a panel model in the normal sense, with observations on individual spatial units over several years. However, it preserves the essential form of a panel since each spatial unit has observations across several industries. Replacing time fixed effects with industry fixed effects, I label the model as the spatial pseudo-panel industry fixed-effects Durbin model. This innovation in methodology facilitates making full use of information in the dataset, taking into account variations in variables along both spatial and industry dimensions, controlling for industry

¹Hereafter, industry agglomeration is also referred to as agglomeration or economic agglomeration. Also, the terms of “agglomeration”, “concentration”, and “clustering” are used interchangeably in this paper, although these terms have their specific implications (Brühlhart, 1998).

heterogeneity and enhancing the efficiency of estimation.

Understanding the role of input-output linkages on industry agglomeration is of practical importance. Regional development policies, backed by the growth-pole theory (Perroux, 1950) and the cluster theory (Porter, 1998), rely on the notion of cumulative causation stressing the self-reinforcing process of industry agglomeration. Regional policy makers often selectively attract some industries to locate in their regions, hoping that other industries that have strong input-output linkages with the chosen ones can follow suit and the process can be fed on by itself, leading to a higher degree of agglomeration. The effectiveness of such regional development policies hinges on the extent that input-output linkages work towards industry agglomeration, requiring empirical studies to evaluate this causal relationship.

The rest of the paper is structured as follows. Section 2 briefly reviews the literature on industry agglomeration and input-output linkages. Section 3 introduces the approach to constructing key variables through which the integrated spatial econometric and input-output model is implemented. Section 4 presents the estimation results, with a discussion of some estimation issues. Section 5 draws a conclusion.

2 Literature Review²

Economic theories have emphasized the role of input-output linkages in the formation of industry agglomeration for a long time. von Thünen (1826) observes that it is efficient for factories to locate closely because “machines to produce machines are themselves the product of many different factories”(Fujita, 2012). Marshall (1890) advances the concept of input sharing of industries as an ingredient for agglomeration in his famous trinity of the Marshallian external economies (input sharing, labor market pooling, and knowledge spillovers). The vertical linkage models (Krugman and Venables, 1995; Venables, 1996; Robert-Nicoud, 2006) in New Economic Geography (NEG) formalize the analysis of input-

²See Fujita et al. (1999), Fujita and Thisse (2002), and Combes et al. (2008) for comprehensive surveys of theoretical studies and Head and Mayer (2004), Combes et al. (2008) and Redding (2010) for comprehensive surveys of empirical studies.

output linkages and agglomeration under the general equilibrium framework using the Dixit-Stiglitz monopolistic competition model.

In empirical studies, a straightforward approach to examining the role of input-output linkages on industry agglomeration is to regress some agglomeration indices on a set of explanatory variables according to relevant theories. Head and Mayer (2004) summarize this approach as the concentration regressions, which can be considered as the reduced-form approach to testing the NEG theories. The general form of the concentration regressions is as follows,

$$CONC_s = a + bTRCOSTS_s + cIRS_s + dLINKAGES_s + \dots + e_s \quad (1)$$

The dependent variable, $CONC_s$, is a particular agglomeration index for industry s . $TRCOSTS_s$ and IRS_s are proxies for trade costs and the degree of increasing returns, respectively. $LINKAGES_s$ measures the industry's reliance on intermediate inputs. Additional variables can be included to control other plausible factors for agglomeration. Table 1 lists the existing empirical studies pertaining to the concentration regression category, which aim at testing the effect of input-output linkages on agglomeration. I review these papers by asking the following questions: What are the purposes of these studies? What is their estimation method? What are the measures for agglomeration and input-output linkages?

All papers, except Mion (2004), set the goal of testing the sources of industry agglomeration based on Marshallian external economies. Rosenthal and Strange (2001), Glaeser and Kerr (2009), and Ellison et al. (2010) attempt to distinguish three sources of Marshallian external economies by controlling other factors such as natural endowments for U.S. manufacturing industries. Dumais et al. (1997) introduce a dynamic process to examine the effect of Marshallian external economies on the change of manufacturing employment in the U.S. states. Jofre-Monseny et al. (2011) develop a discrete choice model to examine the three sources for manufacturing industries in Spain. Overman and Puga (2010) focus

Table 1: The Summary of Literature on the Effect of Input-Output Linkages on Agglomeration

Authors, Year	Main Purpose	Estimation Method	Measures for Agglomeration	Measures for I-O Linkages
Dumais et al. (1997)	The authors regress the changes in employment share on Marshallian agglomeration forces	linear regressions using OLS with state and industry fixed effects; Non-spatial	$\Delta E_{ist}^j / E_{it}$ and $\log(1 + \Delta E_{ist}^j)$	$input_{ist} \equiv \sum_{j \neq i} I_{ji} E_{jst} / E_{jt}$ and $output_{ist} \equiv \sum_{j \neq i} O_{ji} E_{jst} / E_{jt}$
Rosenthal and Strange (2001)	Examine the determinants of agglomeration economies for U.S. manufacturing industries, including Marshallian agglomeration economies	linear regressions using OLS with industry fixed effects; Non-spatial	the EG index	energy, natural resources, water, manufactured inputs, non-manufactured inputs as shares of shipment
Mion (2004)	Examines the role of agglomeration externalities stemming from input-output linkages in the location of U.S. manufacturing plants based on NEG theories	the Tobit model using MLE; Spatially lagged independent variables	$Spec_{ci} \equiv \ln(1 + \frac{emp_{ci}/emp_c}{emp_i/emp})$	$IntD_{ci} = \ln(1 + \sum_k \mu^{c,k} r_k emp_{ki})$, $IntS_{ci} = \ln(1 + \sum_k \mu^{k,c} r_k emp_{ki})$
Glaeser and Kerr (2009)	Examine the impact of local conditions on entrepreneurial rates, including Marshallian agglomeration economies	linear regressions using OLS with city and industry fixed effects; Non-spatial	Log entry employment of new firms for city-industries	$Input_{ci} = - \sum_k Input_{i \leftarrow k} - \frac{E_{ikc}}{E_c} $, $Output_{ci} = \frac{[\sum_k Output_{i \rightarrow k} \frac{E_{kic}}{E_c}]}{[\sum_k Output_{\rightarrow k} \frac{E_{kic}}{E_c}]}$
Ellison et al. (2010)	Examine the sources of coagglomeration of industries in U.S. manufacturing based on Marshallian agglomeration economies	linear regression using OLS; Non-spatial	EG coagglomeration index and the Duranton and Overmans index	$Input_{ij} = \max\{inputs_{i \rightarrow j}, inputs_{j \rightarrow i}\}$, $Output_{ij} = \max\{Outputs_{i \rightarrow j}, Outputs_{j \rightarrow i}\}$, $InputOutput_{ij} = \max\{inputs_{ij}, Outputs_{ij}\}$
Overman and Puga (2010)	Examine the importance of labor pooling for agglomeration based on Marshallian agglomeration economies	Panel model with industry fixed effects; Non-spatial	the EG index	A set of production factors as share of inputs; IO weighted EG index
Jofre-Monseny et al. (2011)	Examine three Marshallian agglomeration forces in determining location of new firms in Manufacturing in Spain	Poisson regression using MLE; Non-spatial	The number of new firms	$input_{ic} = \sum_{j-i} (W_{ij}^I L_{cj})$ where $W_{ij}^I = \frac{inputs_{i \rightarrow j}}{totalinputs_i}$ $output_{ic} = \sum_{j-i} (W_{ij}^O L_{cj})$ where $W_{ij}^O = \frac{outputs_{i \rightarrow j}}{totaloutputs_i}$

on the labor market pooling effect and include input-output linkages as a control variable for manufacturing industries in the U.K. Different from these studies, Mion (2004) directly bases his regression on the NEG theory and tests the role of final demand and input-output linkages in agglomeration.

As for the estimation methods, the majority estimate linear regression models using ordinary least square (OLS). Some studies include fixed effects for industries and/or spatial units (Dumais et al., 1997; Rosenthal and Strange, 2001; Glaeser and Kerr, 2009; Ellison et al., 2010; Overman and Puga, 2010). Mion (2004) estimates a Tobit model with the maximum likelihood (ML) method to account for possible bias caused by censored data due to zeros in the dependent variable for many observations. Jofre-Monseny et al. (2011) employ an innovative method in which a Poisson regression model is derived from the discrete choice model and estimated using the ML method. However, none of the existing studies use spatial econometric models by including the spatial lag of the dependent variable.

One prerequisite condition for the concentration regressions is to find an agglomeration index.³ Three studies in Table 1 use the Ellison-Glaeser (EG)⁴ index to measure agglomeration (Rosenthal and Strange, 2001; Ellison et al., 2010; Overman and Puga, 2010). The EG index is derived by Ellison and Glaeser (1997) based on the discrete choice model, designed to correct the bias caused by industry structure on the agglomeration measurement. However, the computation process of the EG index involves the summation over regions, resulting in losing the spatial dimension, which is indispensable in spatial econometric models. In contrast, other studies in Table 1 compute agglomeration measures that possess both spatial and industry dimensions. Mion (2004) uses the logarithmically transformed location quotient, which is another widely used agglomeration index (Kim, 1995; Holmes and Stevens,

³Combes et al. (2008, Chapter 10) and Nakamura and Paul (2009) provide a survey for agglomeration indices.

⁴The EG index for industry i takes the form as follows: $\gamma_i = \frac{G_i / (1 - \sum_r s_{ir}^2) - H_i}{1 - H_i}$, where s_{ir} is the share of industry i 's employment in region r , $G_i \equiv \sum_r (s_{ir} - s_r)^2$ is a raw concentration index, s_r is the share of total employment of region r , $H_i \equiv \sum_k \frac{e_{ik}^2}{(\sum_k e_{ik})}$ is a Herfindahl index of the plant-level concentration of employment and e_{ik} is the level of employment in the k th plant in industry i .

2004; Billings and Johnson, 2012) and, importantly, has the two dimensions.

Another key variable of interest is the measure of input-output linkages. The simplest measures take the form as the shares of particular production factors in total inputs or outputs. Rosenthal and Strange (2001) use the shares of energy, natural resources, water, manufactured inputs, and non-manufactured inputs in total shipment to measure the dependency of an industry on natural endowment and use input-output linkages to distinguish their individual effects on agglomeration. Other studies use the information in the input-output tables to compute linkage measures. Ellison et al. (2010) obtain summarized coefficients directly from the 1987 benchmark input-output accounts. They define $Input_{i \leftarrow j}$ as the share of industry i 's inputs that come from industry j , and similarly, $Output_{i \rightarrow j}$ as the share of industry i 's outputs that are sold to industry j . The shortcoming of these measures is that they only have the industry dimension. A more sophisticated approach to including the input-output table information is to combine the input-output coefficients with other variables, typically employment, to obtain measures for input-output linkages, which have both spatial and industry dimensions as in Dumais et al. (1997), Mion (2004), Glaeser and Kerr (2009) and Jofre-Monseny et al. (2011). The spatial intermediate demand variable belongs to this type of measure for input-output linkages.

3 Methodology

The methodology of this paper is to estimate a spatial Durbin model by using the spatial panel time fixed-effects estimation method. Different from the normal panel model that includes individual and/or time fixed effects, the model in this paper replaces time fixed effects with industry fixed effects, thereby labeled as the spatial pseudo-panel industry fixed-effects Durbin model. The analysis uses this methodology based on three considerations: the nature of data, the approach to constructing key variables, and, most importantly, the necessity of studying industry agglomeration with respect to both spatial and industry

dimensions.

3.1 Data

I obtain industry employment data in counties from the 2002 County Business Patterns (CBP) imputed by using the method of Isserman and Westervelt (2006). The CBP of the U.S. Bureau of Census provides comprehensive annual records on employment, payroll, and the number of establishments by detailed industry for all counties in the U.S. However, the nondisclosure problem of the CBP limits its usefulness. To construct variables used in the model, a complete CBP dataset is necessary. To deal with the nondisclosure problem, Isserman and Westervelt (2006) propose a method to impute the withheld data, taking advantage of the hierarchical structure of the CBP dataset. I download the complete imputed 2002 CBP dataset⁵ and use employment of manufacturing industries with 3-digit NAICS codes⁶ at the county level to compute location quotients and the intermediate demand variable.

The data format of the CBP dataset is analogous to a panel even though it is for only one year. The county and industry attributes of the data make it resemble a panel. With the dummy variables for industries replacing those for time periods, the model can be estimated as a spatial panel time fixed-effects model. Moreover, I make the structure of the panel balanced in that each county has the same number of industries, imposing zeros for employment of industries that do not exist in the county. Balancing the panel makes the spatial weight matrix remain the same for each industry. The purpose of constructing the pseudo-panel data is to model industry heterogeneity and make full use of information in the data, improving accuracy and efficiency of estimation.

Other data sources include the Bureau of Economic Analysis (BEA) for the 2002 input-output table at the summary level, USA CountiesTM of the Census Bureau for demographic variables, and the Economic Research Service of the United States Department of Agricul-

⁵The dataset is downloaded from www.wholedata.com. Unfortunately, this website has been shut down recently.

⁶To compute the intermediate demand variable, I use nearly all 3-digit NAICS industries, except for several 2-digit NAICS code industries in BEA's 2002 input-output table at the summary level.

ture (USDA) for the natural amenities scale and the 2003 rural-urban continuum codes. The spatial weight matrix is created using the cartographic boundary files from the Census Bureau.

3.2 Variable Construction

The Dependent Variable

The dependent variable in the model is the logarithmically transformed location quotient (LQ). Let $i = 1, 2, \dots, S$ denotes industries and $r = 1, 2, \dots, N$ denotes regions, then the LQ of industry i in region j is defined as

$$LQ_{ir} = \frac{E_{ir}/E_{i*}}{E_{*r}/E_{**}}$$

where E_{ir} is the employment of industry i in region r , E_{i*} is the total employment of industry i in all regions, E_{*r} is the total employment of all industries in region r , and E_{**} is the total employment of the overall economy. The problems of influential observations and skewness in the LQ suggest using logarithmic transformation on the LQ. For observations for which LQs are zero, adding a small positive number can avoid the logarithmic transformation yielding negative infinity. Thus, the dependent variable is $\log(LQ + \epsilon)$, where $\epsilon = \frac{1}{10} \min \{LQ : LQ > 0\}$.⁷

The spatial lag of the dependent variable is necessary to deal with spatial spillover effects and measurement errors. Anselin et al. (2008) consider a spatial econometric model with the spatially lagged dependent variable as “the formal specification for the equilibrium outcome of a spatial or social interaction process, in which the value of the dependent variable for one agent is jointly determined with that of the neighboring agents”. Moreover, the spatial lag

⁷Mion (2004) uses $\log(LQ + 1)$ as the dependent variable. However, if the LQ is zero for an industry in a county, then $LQ + 1$ will bring up the level of the LQ near to the national average, which is implausible. In fact, I experiment with different values for ϵ by dividing the minimum nonzero LQ by 10 to the power of 1 to 10. The power of 10 gives $\log(LQ)$ the smallest standard deviation. Also, the signs of estimated coefficients are not sensitive for ϵ , including for $\epsilon = 1$.

of the dependent variable is of methodological importance for using a discrete agglomeration index, like the LQ, which suffers from the modifiable area unit problem (MAUP) (Openshaw, 1981). The MAUP makes an agglomeration index biased when industries actually agglomerate across the administrative boundaries of spatial units from which the data are collected. Addressing measurement errors across spatial units requires the spatial lag of the dependent variable (LeSage and Pace, 2009).

Independent Variables

This paper uses the intermediate demand variable (IDV) to represent input-output linkages. The IDV is the key variable through which the embedding strategy of the integrated econometrics and input-output model is implemented. The IDV is first advanced by Moghadam and Ballard (1988) as⁸

$$IDV_{ir} = \sum_{j=1}^S a_{ij} E_{jr}, \text{ for } r = 1, \dots, N. \quad (2)$$

The IDV of industry i in region r is the weighted sum of employment of all other industries in region r . The weight a_{ij} represents the share of inputs from industry i in the output of industry j at the national level, i.e. the coefficient in the input-output table.

Rey and Dev (1997) and Rey (2000) extend the IDV for addressing spatial linkages. The spatial intermediate demand variable (SIDV) is defined as

$$SIDV_{ir} = \sum_{s \neq r} \sum_j \phi_{rs} \gamma_{ij}^s a_{ij} E_{js}, \text{ for } r = 1, \dots, N \quad (3)$$

where ϕ_{rs} is the freeness of trade between regions r and s , which can be estimated as the decaying function of distance between r and s . In Rey and Dev (1997) it is $\phi_{rs} = 1/d_{rs}$ and

⁸The IDV in Moghadam and Ballard (1988), Rey and Dev (1997), and Rey (2000) has a time dimension to introduce dynamics into the input-output model. However, I only have data for one year, thereby I omit the subscript t of the IDV without loss of clarity. Also, the summation should be over $j \neq i$ in a rigorous sense. Since intra-industry trade is possible, especially for industry aggregation into 3-digit NAICS codes in the I-O table, the summation over all industries is reasonable.

d_{rs} is the distance. γ_{ij}^s is the import propensity of industry j in region s for intermediate inputs produced by industry i elsewhere. For simplicity, I assume $\gamma_{ij}^s = 1$ for all regions and industries. Also, I include the labor productivity adjustment discussed in Rey and Jackson (1999) to overcome the dimensional inconsistency problem in the original IDV. Thus, $SIDV_{ir}$ can be re-written as

$$SIDV_{ir} = \sum_{s \neq r} \sum_j \phi_{rs} \varphi_{ij} a_{ij} E_{js}, \text{ for } r = 1, \dots, N \quad (4)$$

where $\varphi_{ij} \equiv l_j/l_i$ is the adjustment factor of labor productivity, l , of industry i and j .

Using the spatial weight matrix makes the computation of $SIDV_{ir}$ easier. An inverse-distance spatial weight matrix, \mathbf{W} , is an $N \times N$ matrix that has off-diagonal elements $w_{rs} = 1/d_{rs}$ for $r \neq s$ and diagonal elements of zero, i.e. $w_{rr} = 0$. Define the local IDV_{ir} for industry i in region r as $\sum_j \varphi_{ij} a_{ij} E_{jr}$. Then replacing ϕ_{rs} with w_{rs} in Equation 4, yields the spatial lag of the local IDV. That is, $SIDV_{ir} = \sum_{s \neq r} w_{rs} \sum_j \varphi_{ij} a_{ij} E_{js} = \sum_{s \neq r} w_{rs} IDV_{is}$. In matrix notation, let \mathbf{IDV}_i be an $N \times 1$ vector of the IDV for industry i in all regions. Then the SIDV for industry i is equal to $\mathbf{W} \cdot \mathbf{IDV}_i$. As such, \mathbf{IDV}_i and $\mathbf{W} \cdot \mathbf{IDV}_i$ in the spatial model represent intermediate demands from the local and neighboring areas.

Besides input-output linkages, another important factor to explain industry agglomeration is market potential for final demand markets. I use the nominal market potential (NMP) variable (Harris, 1954; Head and Mayer, 2004) to describe the strength of final demand markets. The NMP for industry i in region s is defined as $NMP_{ir} = \sum_s \phi_{rs} \mu_{is} Y_s$. Similar to the IDV, the NMP can also be constructed using the inverse-distance spatial weight matrix, \mathbf{W} , if ϕ_{rs} is approximated as $1/d_{rs}$. Define the local market potential for industry i in region r as $MP_{ir} = \mu_{ir} Y_r$, then $NMP_{ir} = \sum_{s \neq r} w_{rs} \mu_{is} Y_s = \sum_{s \neq r} w_{rs} MP_{is}$. In matrix notation, \mathbf{MP}_i is the vector of local market potentials for industry i and $\mathbf{W} \cdot \mathbf{MP}_i$ is its spatial lag.

Additionally, I include two groups of control variables. The first group represents the exogenous natural endowment of regions, consisting of three variables: a natural amenities

scale, a dummy variable for coastal counties, and a dummy variable for metro counties. The natural amenities scale is a measure of the physical characteristics of a county area. The scale was constructed by combining six measures of climate, topography, and water area that reflect environmental qualities most people prefer. The definition of coastal counties is provided by the Strategic Environmental Assessments Division of the National Oceanic and Atmospheric Administration. The coastal dummy variable takes the value of one for coastal counties and zero otherwise. The dummy variable for metro counties takes the value of one if the 2003 rural-urban continuum code computed by USDA is less than 4, otherwise being zero for nonmetro counties. The second group consists of demographic variables for counties, including the growth rate of population from 1980 to 2000, the percentage of persons who are 25 years and over with a Bachelor’s or higher degree in 2000, and the number of violent crimes relative to the national number in 2000. Table 2 shows the summary statistics of all variables in the model and the inverse-distance spatial weight matrix.

Table 2: Descriptive Statistics of Variables

Variable	Unit	Min	Max	Mean	Std
LQ	national level=1	0.00	259.95	1.37	5.57
IDV	thousand persons	0.00	46.24	0.15	0.66
MP	million dollars	0.00	8728.40	18.77	120.26
Amenity	z-scores	-6.40	11.17	0.05	2.29
PopGrowth	percentage	-1.86	25.12	0.72	1.53
Bachelor	percentage	0.05	0.60	0.16	0.08
Crime	percentage	0.00	6.67	0.03	0.19

	Number of Nonzero Links	Percentage of Nonzero Links	Min. Number of Neighbors	Max. Number of Neighbors
W	247,730	2.65	3	169

(1) Number of observations: 64,239 (3,059 counties in 48 states and 21 manufacturing industries with 3-digit NAICS code). Independent cities in VA are included in the locating counties. The number of coastal counties is 621, and the number of metro counties is 1,052.

(2) Data Sources: The CBP, USA CountiesTM and the cartographic boundary files from the Census Bureau, the ERS of the USDA.

(3) W is the inverse-distance spatial weight matrix, non-standardized, being constructed using the cut-off distance that is 1.5 times as long as the distance ensuring each county to have at least one neighbor.

3.3 Model Specification and the Estimation Method

The approach to constructing the SIDV and NMP as spatial lags of IDV and MP dictates the model specification as a spatial Durbin model, estimated as a spatial panel industry fixed-effects model. For each industry $i = 1, 2, \dots, S$, the model is

$$\mathbf{y}_i = \rho \mathbf{W} \mathbf{y}_i + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{W} \mathbf{X}_i \boldsymbol{\theta} + \boldsymbol{\iota}_N \eta_i + \boldsymbol{\varepsilon}_i, \text{ and } \boldsymbol{\varepsilon}_i \sim IID(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_N) \quad (5)$$

where \mathbf{y}_i is an $N \times 1$ vector of the dependent variable, \mathbf{W} is the $N \times N$ inverse-distance spatial weight matrix, ρ is the spatial autoregressive coefficient. \mathbf{X}_i is an $N \times K$ matrix of explanatory variables (i.e. IDV, MP, and control variables) excluding the constant term, and $\boldsymbol{\beta}$ is the $K \times 1$ coefficient vector. $\boldsymbol{\iota}_N \eta_i$ is the industry fixed-effects dummy variable, where $\boldsymbol{\iota}_N$ is an $N \times 1$ vector of ones, and $\boldsymbol{\varepsilon}_i$ denotes disturbances assumed as being independently and identically distributed with the mean of zero and the variance of σ_ε^2 .

Stacking Equation (5) over all industries, the model can be expressed in the panel format,

$$\mathbf{y} = \rho \widetilde{\mathbf{W}} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \widetilde{\mathbf{W}} \mathbf{X} \boldsymbol{\theta} + \mathbf{D} \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \text{ and } \boldsymbol{\varepsilon} \sim IID(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_{NS}) \quad (6)$$

where $\widetilde{\mathbf{W}}$ is the spatial weight matrix for the panel data, i.e. $\widetilde{\mathbf{W}} = \mathbf{I}_S \otimes \mathbf{W}$, the Kronecker product of the $S \times S$ identity matrix and the spatial weight matrix \mathbf{W} for one industry equation. \mathbf{y} , $\boldsymbol{\varepsilon}$, and $\boldsymbol{\eta}$ are $NS \times 1$ vectors, \mathbf{X} is an $NS \times K$ matrix for explanatory variables, $\mathbf{D} = \mathbf{I}_S \otimes \boldsymbol{\iota}_N$ is an $NS \times T$ matrix and $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_S)'$ is an $S \times 1$ vector for industry fixed effects.

I estimate Equation 6 using the maximum likelihood method.⁹ According to Anselin et al. (2008) and Elhorst (2009), the log-likelihood function of the spatial panel industry

⁹Lee and Yu (2010) derive the asymptotic properties of the ML estimation of a spatial dynamic panel data model with both time and individual fixed effects. The MATLAB[®] routines for the spatial panel model are explained in Elhorst (2012) and downloaded from <http://www.regroningen.nl/elhorst/software.shtml>.

fixed-effects Durbin model is given by

$$\log L = -\frac{NS}{2} \log \sigma_\varepsilon^2 + S \log |\mathbf{I}_N - \rho \mathbf{W}| - \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^S \varepsilon_i' \varepsilon_i \quad (7)$$

where $\varepsilon_i = \mathbf{y}_i - \rho \mathbf{W} \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{W} \mathbf{X}_i \boldsymbol{\theta} - \boldsymbol{\iota}_N \eta_i$. Maximizing Equation 7 with respect to coefficients yields the estimation results.

4 Estimation

Table 3 shows the estimation results of three groups of models. The first two groups of models (Models 1-3 and 4-6) pertain to the spatial Durbin model (SDM). In Models 1 to 3, the IDV and MP are in linear form. In Models 4 to 6, they are in logarithmic form. The first model in each of these two groups only contains the IDV and MP along with their spatial lags. The second models in each group (Model 2 and 5) add control variables for natural endowment. The third models in each group (Model 3 and 6) add control variables for demographic factors. Model 7 is a spatial autoregressive model (SAR) in which the IDV and W*IDV are combined to form the SIDV, and MP and W*MP are combined to form the NMP. LeSage and Pace (2009) argue that it is inappropriate to directly interpret coefficient estimates in spatial econometric models that include a spatial lag of the dependent variable. Table 4 reports the direct, indirect, and total effects of explanatory variables in Models 3, 6, and 7.

The estimation results confirm the positive role of input-output linkages in agglomeration. The coefficient estimates on the IDV are significantly positive at the 1% level in all models. Despite the negative coefficients on W*IDV in the second group of models, the direct, indirect, and total effects of the IDV are significantly positive in all models. The direct effect of the IDV on $\log(\text{LQ})$ is 0.4657. Given that the unit of the IDV is thousands of employees, this means that for the manufacturing industry, say M, in a county, an increase of one thousand employees in related industries at the local market leads to a 46.57% in-

Table 3: Estimation Results of Coefficients in Various Models

Model Types	SDM			SDM			SAR
Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
IDV	0.5505*** (19.8103)	0.4627*** (17.0148)	0.4527*** (14.2691)	—	—	—	—
MP	0.0008*** (5.3512)	0.0005*** (3.5171)	0.0004*** (2.6184)	—	—	—	—
log(IDV)	—	—	—	2.1049*** (118.3016)	2.1157*** (115.7535)	2.1270*** (113.7215)	—
log(MP)	—	—	—	-1.0193*** (-46.7863)	-1.1119*** (-45.2803)	-1.0902*** (-40.6999)	—
SIDV	—	—	—	—	—	—	0.0638*** (32.7371)
NMP	—	—	—	—	—	—	0.0000 (0.7215)
Amenity	—	-0.0230** (-2.2233)	-0.0579*** (-5.2761)	—	-0.0266*** (-2.9832)	-0.0245*** (-2.6454)	-0.1522*** (-20.2846)
Coastal	—	0.7387*** (12.7222)	0.9177*** (15.4122)	—	-0.0519 (-1.0381)	0.1022** (2.0121)	0.1471*** (3.6321)
Metro	—	1.2676*** (35.7200)	0.9647*** (24.9733)	—	0.3489*** (10.1435)	0.3162*** (9.0491)	1.3246*** (35.1372)
PopGrowth	—	—	0.1685*** (12.8274)	—	—	0.1136*** (10.1519)	0.2331*** (19.5532)
Bachelor	—	—	3.3131*** (14.2217)	—	—	-1.2336*** (-5.7266)	3.5269*** (15.6629)
Crime	—	—	-0.0353 (-0.3158)	—	—	-0.7535*** (-10.4653)	0.9134*** (10.8782)
W*IDV	0.0811*** (38.5143)	0.0149*** (8.3199)	0.0082*** (4.1861)	—	—	—	—
W*MP	0.0001*** (6.6731)	-0.0000*** (-4.7324)	-0.0001*** (-5.7150)	—	—	—	—
W*log(IDV)	—	—	—	-0.0072*** (-63.5440)	-0.0075*** (-52.7933)	-0.0052*** (-29.1417)	—
W*log(MP)	—	—	—	0.0015*** (8.0864)	0.0009*** (4.6091)	0.0009*** (4.5630)	—
W*Amenity	—	-0.0021*** (-7.9404)	-0.0014*** (-4.1921)	—	0.0004** (2.0422)	0.0011*** (3.7493)	—
W*Coastal	—	-0.0212*** (-14.6998)	-0.0244*** (-16.5399)	—	-0.0037*** (-2.9846)	-0.0081*** (-6.4522)	—
W*Metro	—	0.0620*** (40.4278)	0.0431*** (16.7092)	—	0.0075*** (5.8364)	-0.0089*** (-3.9920)	—
W*PopGrowth	—	—	-0.0006 (-1.0313)	—	—	-0.0016*** (-3.4504)	—
W*Bachelor	—	—	0.1272*** (15.3790)	—	—	0.1090*** (12.3670)	—
W*Crime	—	—	-0.0356*** (-3.1053)	—	—	-0.0085 (-0.9480)	—
ρ	0.0030*** (14.3604)	0.0060*** (91.1849)	0.0070*** (201.8096)	0.0070*** (298.3578)	0.0080*** (1914.6862)	0.0070*** (261.7192)	0.0030*** (15.1920)
R^2	0.2287	0.2881	0.2974	0.4980	0.5001	0.5027	0.2656
log(likelihood)	-179121	-178501	-178694	-170047	-171377	-169997	-177630

Notes: (1) Significant levels: * for 10%, ** for 5%, and *** for 1%.

(2) t-statistics are in parentheses.

Table 4: Estimation of Direct, Indirect and Total Effects

	Effects	Direct Effects	Indirect Effects	Total Effects
Model 3	IDV	0.4657*** (15.0622)	2.2286*** (5.8286)	2.6943*** (7.0696)
	MP	0.0004** (2.5445)	-0.0095*** (-5.6359)	-0.0091*** (-5.4338)
	Amenity	-0.0607*** (-5.5241)	-0.3504*** (-6.4582)	-0.4111*** (-7.7385)
	Coastal	0.8951*** (15.5893)	-3.6250*** (-16.6098)	-2.7299*** (-12.9652)
	Metro	1.0156*** (28.0351)	9.8627*** (20.1825)	10.8783*** (22.3223)
	PopGrowth	0.1691*** (12.1889)	0.1073 (1.1262)	0.2765*** (2.9321)
	Bachelor	3.4691*** (16.8323)	29.6834*** (17.1703)	33.1525*** (19.3147)
	Crime	-0.0707 (-0.6295)	-7.0966*** (-3.2703)	-7.1673*** (-3.3073)
Model 6	log(IDV)	2.1378*** (118.8789)	1.8187*** (28.5479)	3.9565*** (64.7359)
	log(MP)	-1.0978*** (-40.9269)	-1.2786*** (-16.2743)	-2.3765*** (-32.1809)
	Amenity	-0.0238*** (-2.5811)	0.1776*** (3.6230)	0.1538*** (3.1944)
	Coastal	0.0942* (1.8829)	-1.5101*** (-7.4542)	-1.4159*** (-7.2127)
	Metro	0.3110*** (8.5539)	-1.3329*** (-2.9648)	-1.0219** (-2.2804)
	PopGrowth	0.1123*** (9.8912)	-0.1623* (-1.9168)	-0.0499 (-0.5952)
	Bachelor	-1.1269*** (-5.3634)	20.1907*** (11.2355)	19.0638*** (10.6817)
	Crime	-0.7734*** (-10.6066)	-2.7295 (-1.5151)	-3.5029* (-1.9460)
Model 7	SIDV	0.0638*** (32.7447)	0.0176*** (6.3151)	0.0814*** (40.8088)
	NMP	0.0000 (0.6669)	0.0000 (0.1120)	0.0000 (0.6664)
	Amenity	-0.1524*** (-20.0542)	-0.0421*** (-3.2824)	-0.1946*** (-18.8064)
	Coastal	0.1477*** (3.8615)	0.0408 (0.6582)	0.1885*** (3.8689)
	Metro	1.3266*** (36.1197)	0.3668*** (5.4260)	1.6934*** (29.8432)
	PopGrowth	0.2338*** (19.3484)	0.0646*** (3.1905)	0.2984*** (18.3567)
	Bachelor	3.5247*** (15.8283)	0.9746*** (2.6133)	4.4993*** (15.0407)
	Crime	0.9158*** (11.6871)	0.2533* (1.9493)	1.1691*** (11.2787)

Notes: (1) Significant levels: * for 10%, ** for 5%, and *** for 1%.

(2) t-statistics are in parentheses.

crease in the LQ for industry M. Take the chemical manufacturing industry (NAICS 325) in Monongalia County, WV as a concrete example. The LQ for chemical manufacturing in the county in 2002 is 4.51, implying that the industry is concentrated in the county. Given that the number of employees in the county's chemical industry in 2002 is 1,347, with other things being unchanged, the LQ increasing by 46.57% implies an increase of 626 employees in this industry as a result of an additional thousand employees in other industries that have input-output linkages with the chemical industry in the county. The indirect effect of the IDV is 2.2286, implying that the LQ will rise more than two times when the intermediate demands in neighboring counties increase by one thousand employees. For Monongalia County, which has 114 neighbors (the farthest one is Amherst, VA) in the spatial weight matrix, an increase of one thousand employees in related industries in all its neighboring counties will contribute to 2,999 more employees in the chemical industry in Monongalia County. Combining the direct and indirect effects, the total effect of the IDV is 2.6943, further strengthening the impacts of the IDV on $\log(\text{LQ})$.

The estimation results for market potential are counterintuitive. While the signs on the MP and its spatial lag in the first group of models and the second group are opposite, it does not change the negative sign of the total effects of the MP. In Model 3, the direct effect of the MP is 0.0004, which means that the LQ for an industry can increase by 0.04% if the local final demand for the industry increases by \$1 million. Continuing with the example of Monongalia County, the \$1 million increase in final demand in the county only results in the chemical industry employment increasing by less than one person. However, the small gain will be overwhelmed by the indirect effect. An increase in market potential in neighboring counties will pull some manufacturing production out of Monongalia county. In Model 7, which combines the MP and its spatial lag as the NMP, the direct, indirect, and total effects are all insignificantly different from zero, further illustrating the counteraction of the local and neighboring final demand markets. Mion (2004) also gets negative coefficients on market potential. He explains that manufacturing activities are pushed aside from the center

by sectors characterized by higher transportation costs. Facing higher wages and rents in the center, manufacturing industries, which do not depend on face-to-face interactions, tend to locate closer to intermediate demand markets than to final demand markets. Even though this explanation sounds plausible, it cannot exclude the possibility that the variable used to measure market potential is misspecified. Head and Mayer (2004) cite several studies that use the real market potential (RMP)¹⁰ to explain the regional variation in wages and they find positive results. They argue that studies getting negative signs on market potential may be using inappropriate measures, such as the nominal market potential.

As for the control variables used in each model, some have coefficients with expected signs, while others do not. The most unexpected result is for natural amenities, as measured by the natural amenities scale from USDA. The coefficients on the natural amenities scale and its spatial lags are negative in seven out of nine models. Its total effect in Model 6 is positive but is negative in Models 3 and 7. While the negative sign might be attributed to measurement problem, it may also be true that natural amenities do not affect location decisions of manufacturing industries. The direct effects of coastal and metro dummies are positive, which is expected as coastal and metro counties tend to have more manufacturing industries than other counties. For demographic variables, the population growth from 1980 to 2000 has positive effects in Model 3 and 7, the share of population with at least a Bachelor's degree has strong positive effects in most effect estimates, while the crime rate has negative effects in Models 3 and 6.

Issues with the Estimation Method

While the estimation results validate the role of input-output linkages on agglomeration, some issues with the estimation method suggest caution. The first issue is the endogeneity problem. The IDV, MP, and the demographic control variables may be endogenous. They

¹⁰The real market potential is defined as $RMP_{ir} \equiv \sum_s \phi_{rs} \mu_{is} Y_s P_s^{\sigma-1}$, where ϕ_{rs} is the freeness of trade, μ_{is} is the share of total income Y_s in region s in the consumption of goods produced by industry i , P_s is the price index in region s and σ is the elasticity of substitution.

might be determined by other exogenous variables that can also explain industry agglomeration. Even worse, the causal relationship between agglomeration and these variables may be reverse. The consequence of the endogeneity problem is that the ML estimates may be inconsistent because the true data generating process is not fully specified merely by Equation 5 or 6. A solution to this problem is to use the IV/GMM estimation method, which is proposed by Kelejian and Prucha (1998, 1999) to deal with the existence of $\rho W y$ and simplify the computation process in the ML estimation. Mutl and Pfaffermayr (2011) and Millo and Piras (2012) explain the IV/GMM method for spatial panel models. The advantage of the IV/GMM method is that the instrumental variables can be used to handle the endogeneity problem of both the spatially lagged dependent variable and suspicious explanatory variables.

The second issue concerns balancing the panel by imposing zeros in an industry's employment in counties where the industry does not exist. This may cause estimation problems resulting from censored data or sample selection issues. The Tobit model is often used to overcome the censored data problem. LeSage and Pace (2009) estimate the spatial Tobit model using a Bayesian MCMC framework. The sample selectivity problem is often handled with Heckman's two step method (Davidson and MacKinnon, 2004). Flores-lagunes and Schnier (2012) consider the estimation of a sample selection model with spatial autoregressive errors. However, there is no application of Heckman's method in spatial panel models. Alternatively, the model may also be estimated by the unbalanced panel model method (Egger et al., 2005). However, the appearance of zeros in observations is likely to be meaningful so that removing these observations may throw away useful information.

5 Conclusion

This paper examines the role of input-output linkages between industries in the formation of industry agglomeration. The spatial and industry dimensions of industry agglomeration

suggest building the model under the framework of the integrated spatial econometric and input-output model through the spatial intermediate demand variable. Taking advantage of the data format and the construction of the spatial intermediate demand variable and the nominal market potential variable, this paper estimates the model as the spatial pseudo-panel industry fixed-effects Durbin model. Using manufacturing industries in the U.S. counties, the role of input-output linkages is supported. However, some potential estimation issues, such as endogeneity, censored data, and sample selectivity, need further study. Future research can also seek to refine the variable construction for the spatial intermediate demand variable and the real market potential in order to get more accurate estimates.

Bibliography

- Anselin, L., Gallo, J. L., Jayet, H., 2008. Spatial Panel Econometrics. In: Mátyás, L., Sevestre, P. (Eds.), *The Econometrics of Panel Data*. Springer Berlin Heidelberg, Ch. 19, pp. 625–660.
- 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.
- Brülhart, M., 1998. Economic geography, Industry, Location and Trade. *The World Economy* 21 (6), 775–801.
- Combes, P.-P., Mayer, T., Thisse, J.-F., 2008. *Economic geography: the integration of regions and nations*. Princeton University Press.
- Davidson, R., MacKinnon, J. G., 2004. *Econometric Theory and Methods*. Oxford University Press, New York Oxford.
- Dumais, G., Ellison, G., Glaeser, E. L., 1997. Geographic Concentration as a Dynamic Process. Working Paper Series No.6270, National Bureau of Economic Research.

- Egger, P., Pfaffermayr, M., Winner, H., 2005. An unbalanced spatial panel data approach to US state tax competition. *Economics Letters* 88 (3), 329–335.
- Elhorst, J. P., 2009. Spatial Panel Data Models. In: Fisher, M. M., Getis, A. (Eds.), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Springer Berlin Heidelberg and New York, Ch. C.2, pp. 377–405.
- Elhorst, J. P., 2012. Matlab Software for Spatial Panels. *International Regional Science Review*, 1–22.
- Ellison, G., Glaeser, E. L., 1997. Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy* 105 (5), 889.
- Ellison, G., Glaeser, E. L., Kerr, W. R., 2010. What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review* 100 (June), 1195–1213.
- Flores-lagunes, A., Schnier, K. E., 2012. Estimation of sample selection models with spatial dependence. *Journal of Applied Econometrics* 27 (2), 173–204.
- Fujita, M., 2012. Thünen and the New Economic Geography. *Regional Science and Urban Economics* 42 (6), 907–912.
- Fujita, M., Krugman, P. R., Venables, A. J., 1999. *The Spatial Economy: Cities, Regions and International trade*. MIT Press, Cambridge, MA.
- Fujita, M., Thisse, J.-F., 2002. *Economics of agglomeration: cities, industrial location, and regional growth*. Cambridge University Press, Cambridge, UK.
- Glaeser, E. L., Kerr, W. R., 2009. Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain? *Journal of Economics Management Strategy* 18 (3), 623–663.
- Harris, C. D., 1954. The Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers* 44 (4), 315–348.

- Head, K., Mayer, T., 2004. The empirics of agglomeration and trade. *Handbook of Regional and Urban Economics* 4 (04), 2609–2669.
- Holmes, T. J., Stevens, J. J., 2004. Spatial distribution of economic activities in North America. In: Henderson, J. V., Thisse, J.-F. (Eds.), *Handbook of Regional and Urban Economics*. Vol. 4. Elsevier, Ch. 63, pp. 2797–2843.
- 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.
- Jofre-Monseny, J., Marín-López, R., Viladecans-Marsal, E., 2011. The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms. *Journal of Urban Economics* 70 (2-3), 61–74.
- Kelejian, H. H., Prucha, I. R., 1998. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics* 17 (1), 99–121.
- Kelejian, H. H., Prucha, I. R., 1999. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* 40 (2), 509–533.
- 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.
- Krugman, P. R., Venables, A. J., 1995. Globalization and the Inequality of Nations. *Quarterly Journal of Economics* 110 (4), 857–880.
- Lee, L.-f., Yu, J., 2010. A Spatial Dynamic Panel Data Model With Both Time And Individual Fixed Effects. *Econometric Theory* 26 (02), 564–597.

- LeSage, J. P., Pace, R. K., 2009. *Introduction to Spatial Econometrics*. Chapman and Hall/CRC.
- Marshall, A., 1890. *Principle of economics*. Macmillan, London, UK.
- Millo, G., Piras, G., 2012. splm : Spatial Panel Data Models in R. *Journal Of Statistical Software* 47 (1).
- Mion, G., 2004. Input-Output Linkages, Proximity to Final Demand and the Location of Manufacturing Industries. CORE Discussion Paper No. 2004/53.
- Moghadam, K., Ballard, K. P., 1988. Small area modeling of the industrial sector (SAMIS): an integrated econometric - interindustry approach. *Environment and Planning A* 20 (5), 655–668.
- Mutl, J., Pfaffermayr, M., 2011. The Hausman test in a Cliff and Ord panel model. *The Econometrics Journal* 14 (1), 48–76.
- Nakamura, R., Paul, C. J. M., 2009. Measuring agglomeration. In: Capello, R., Nijkamp, P. (Eds.), *Handbook of regional growth and development theories*. Edward Elgar Publishing, Northampton, MA, Ch. 16, pp. 305–328.
- Openshaw, S., 1981. The modifiable areal unit problem. In: Wrigley, N., Bennett, R. J. (Eds.), *Quantitative geography a British view*. Vol. 38 of *Concepts and techniques in modern geography*,. Geo Books, Ch. 5, pp. 60–70.
- Overman, H. G., Puga, D., 2010. Labor Pooling as a Source of Agglomeration: An Empirical Investigation. In: Glaeser, E. L. (Ed.), *Agglomeration Economics*. University of Chicago Press, Ch. 4, pp. 133–150.
- Perroux, F., 1950. *Economic Space : Theory and Applications*. *The Quarterly Journal of Economics* 64 (1), 89–104.

- Porter, M. E., 1990. *Competitive Advantage of Nations*. Free Press.
- Porter, M. E., 1998. Clusters and competition. *On competition* 76 (6), 197–287.
- Redding, S. J., 2010. The Empirics Of New Economic Geography. *Journal of Regional Science* 50 (1), 297–311.
- Rey, S. J., 2000. Integrated regional econometric+input-output modeling: Issues and opportunities. *Papers in Regional Science* 79 (3), 271–292.
- Rey, S. J., Dev, B., 1997. Integrating Econometric and Input-Output Models in a Multiregional Context. *Growth and Change* 28 (2), 222–243.
- Rey, S. J., Jackson, R. W., 1999. Labor-productivity changes in regional econometric+input-output models. *Environment and Planning A* 31, 1583–1599.
- Robert-Nicoud, F., 2006. Agglomeration and Trade with InputOutput Linkages and Capital Mobility. *Spatial Economic Analysis* 1 (1), 101–126.
- Rosenthal, S. S., Strange, W. C., 2001. The Determinants of Agglomeration. *Journal of Urban Economics* 50 (2), 191–229.
- Venables, A. J., 1996. Equilibrium locations of vertically linked industries. *International Economic Review* 37 (2), 341–359.
- von Thünen, J. H., 1826. *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, Hamburg. English translation by C.M. Wartenberg: *von Thünen's Isolated State*, Oxford, Pergammon Press (1966), Hamburg, Perthes.