

What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns (AER, 2010)

By G. Ellison, E. L. Glaeser, and W. R. Kerr

Palermo Penano

April 12, 2017

Research Question

- Why do firms cluster near one another?

Research Question

- **Why do firms cluster near one another?**
 - ① Reduce shipping cost of inputs and finished goods (input-output linkage)
 - ② Take advantage of gains from a larger labour pool
 - ③ Take advantage of gains from intellectual or technological spillovers
 - ④ Natural advantages
- Marshall (1920)

Method

- Regression analysis to test the contribution of each of the three Marshallian theories of industrial agglomeration + natural advantages to industry coagglomeration
- Addresses reverse causality where coagglomeration patterns \implies industrial relationships

Main Results

- Individually, each Marshallian force are important for explaining variation in industrial coagglomeration
- Taken together, Marshallian forces explain more of the variation in coagglomeration than natural advantages

Measuring Extent of Industry Coagglomeration

Ellison-Glaeser Index (JPE, 1997)

$$\gamma_{ij}^c = \frac{\sum_{m=1}^M (s_{mi} - x_m)(s_{mj} - x_m)}{1 - \sum_{m=1}^M x_m^2}$$

- s_{mi} = share of industry i 's employment
- x_m =
mean employment share in region m across all manufacturing industries
- Two industry version of index developed in Ellison and Glaeser, JPE 1997
- Derived from a model where local industry-specific spillovers, natural advantages, and random chance all contribute to geographic concentration
- Level: state, PMSA, county

Ellison-Glaeser Index

$$\gamma_{ij}^c = \frac{\sum_{m=1}^M (s_{mi} - x_m)(s_{mj} - x_m)}{1 - \sum_{m=1}^M x_m^2}$$

- + Comparable across industries
- + Accounts for region “size”
- + Localization vs industrial concentration
- – Establishments arbitrarily allocated to spatial units
 - distance between establishments not explicitly accounted for
 - different size spatial units
 - Agglomerated industries cut by an administrative border?
- – Statistical significance (departure from agglomeration arising from randomness)

Duranton-Overman Index (ReStud, 2005)

$$\hat{K}_{ij}^{Emp}(d) = \frac{1}{h \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} e(r)e(s)} \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} e(r)e(s) f\left(\frac{d - d_{r,s}}{h}\right)$$

- $e(r)$: employment in establishment r
- $d_{r,s}$: distance between establishments r and s
- h : bandwidth
- n_i : number of firms in industry i
- $f(.)$: Gaussian kernel (bell curve)
- d : scale at which localization is measured
- Estimates the probability density function of the the distribution of bilateral distances between every establishment in industry i and every establishment in industry j

Duranton-Overman Index

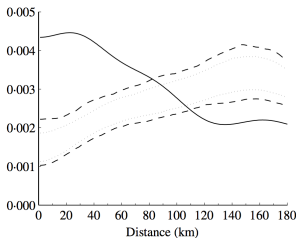
$$\hat{K}_{ij}^{Emp}(d) = \frac{1}{h \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} e(r)e(s)} \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} e(r)e(s) f\left(\frac{d - d_{r,s}}{h}\right)$$

- Explicitly accounts for distance between establishment (missing in EG index)
- Index of localization:

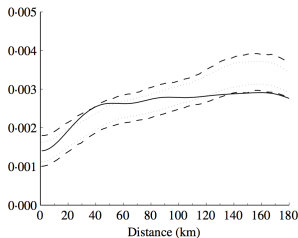
$$\gamma_{ij}(d) \equiv \max(\hat{K}_{ij}^{Emp}(d) - \bar{K}_{ij}^{Emp}(d), 0)$$

- $\bar{K}_{ij}^{Emp}(d)$ is the upper 5% confidence interval in randomly generated counterfactual distribution
- Estimate for each industry pair ij and each distance $d \in \{100mi, 250mi, 500mi, 1000mi\}$

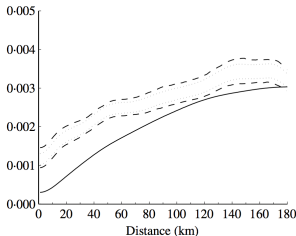
Duranton-Overman Index



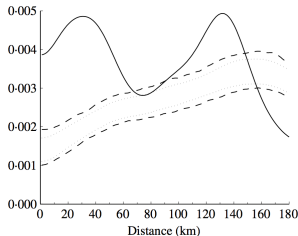
(a) Basic Pharmaceuticals
(SIC2441)



(b) Pharmaceutical Preparations
(SIC2442)



(c) Other Agricultural and Forestry
Machinery (SIC2932)



(d) Machinery for Textile, Apparel and
Leather Production (SIC2954)

Data used to estimate indices

- US Census Bureau's Census of Manufacturing
- 300,000 establishments employing 17 million workers
- 3-digit level SIC3
- 7381 industry pairs per year (1987, 1997). 122 industries

Why do Firms Agglomerate?

$$Coagg_{ij} = \alpha + \beta_{NA} Coagg_{ij}^{NA} + \beta_L LaborCorrelation_{ij} + \beta_{IO} InputOutput_{ij} + \beta_T Tech_{ij} + \varepsilon_{ij}$$

A. Proximity to Customers: Goods

- Firms cluster to reduce shipping costs
- 1987 Input-Output account as measure of extent to which industries buy and sell from one another
- $Input_{i \rightarrow j} = 0.39$: Leather Tanning and Finishing input from Meat Products
- $Output_{i \rightarrow j} = 0.82$: Public building and Related Furniture sales to Motor Vehicles and Equipment
- Most values for an industry pair is zero
- $Input_{ij} = \max\{Input_{i \rightarrow j}, Input_{i \leftarrow j}\},$
 $Output_{ij} = \max\{Output_{i \rightarrow j}, Output_{i \leftarrow j}\}$
- $InputOutput_{ij} = \max\{Input_{ij}, Output_{ij}\}$

B. Labour Market Pooling: People

- Larger labour pool \implies risk sharing, worker-firm matches, higher investment in human capital by workers (lower ex post appropriation in clusters)
- Need measure of extent to which industries use same type of worker
- $Share_{io}$: share of industry i 's employment in occupation o
- $LabourCorrelation_{ij} = corr(Share_{io}, Share_{jo})$
- Most correlated: Motor vehicles and Motorcycles, Bicycles, and Parts (0.984)
- Least correlated: Logging and Aircraft and Parts (-0.046)

C. Intellectual or Tech. Spillovers: Ideas

- Firms cluster to speed the flow of ideas
- Scherer (1984)'s measure of extent to which R&D activity from one industry flow out to benefit another
- NBER Patent data: measure extent to which tech. in industry i cite tech. in industry j (normalized by total citations for the industries)
- $Tech_{ij}$ and $Patent_{ij}$ (construction similar to $InputOutput_{ij}$)

D. Natural Advantages

- Firms cluster in a region because the region has resources useful for their production
- Forests, areas with exogenously cheap electricity, coastal regions, etc.
- Coagglomeration may exist if two industries are attracted to the same natural advantage
 - even if industries would not otherwise have interacted through Marshallian forces
 - e.g. ship building and oil refining
- Controls for coagglomeration between i and j due to natural advantages
 - $Coagg_{ij}^{NA}$
 - Ellison and Glaeser (AER PP, 1999)

D. Natural Advantages: Ellison-Glaeser (AER PP, 1999)

TABLE 1—EFFECT OF “NATURAL ADVANTAGES”
ON STATE-INDUSTRY EMPLOYMENT

A. State variable × industry variable	Coefficient (t statistic)
(a) Electricity price × electricity use	0.170 (17.62)
(b) Natural gas price × natural gas use	0.117 (6.91)
(c) Coal price × coal use	0.119 (4.55)
(d) Percentage farmland × agricultural inputs	0.026 (2.58)
(e) Per capita cattle × livestock inputs	0.053 (5.08)
(f) Percentage timberland × lumber inputs	0.152 (11.98)
(g) Average mfg wage × wages/value added	0.059 (4.11)
(h) Average mfg wage × exports/output	−0.014 (−1.28)
(i) Average mfg wage × import competition	0.036 (3.10)
(j) Percentage without HS degree × percentage unskilled	0.157 (7.38)
(k) Unionization percentage × percentage precision products	0.100 (12.17)
(l) Percentage with B.A. or more × percentage executive/professional	0.170 (12.70)
(m) Coast dummy × heavy exports	−0.031 (−2.20)
(n) Coast dummy × heavy imports	0.017 (0.92)
(o) Population density × percentage to consumers	0.043 (3.68)
(p) (Income share − mfg share) × percentage to consumers	0.025 (4.49)

Empirical Results

OLS

$$Coagg_{ij} = \alpha + \beta_{NA} Coagg_{ij}^{NA} + \beta_L LaborCorrelation_{ij} + \beta_{IO} InputOutput_{ij} + \beta_T Tech_{ij} + \varepsilon_{ij}$$

OLS-Univariate

TABLE 3—OLS UNIVARIATE SPECIFICATIONS FOR PAIRWISE COAGGLOMERATION

Each entry reports separate estimation with <u>single regressor</u>	EG coagglomeration index, 1987				DO coagglomeration index, 1997			
	State total empl.	PMSA total empl.	County total empl.	State entry empl.	Bilateral firm employments with localization threshold			
					1,000 mi.	500 mi.	250 mi.	100 mi.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Natural advantages [DV Specific]	0.210 (0.020)	0.188 (0.017)	0.222 (0.014)	0.120 (0.016)	0.442 (0.013)	0.406 (0.012)	0.253 (0.013)	0.531 (0.019)
R^2	0.044	0.036	0.049	0.014	0.200	0.165	0.064	0.282
Labor correlation	0.180 (0.014)	0.106 (0.016)	0.082 (0.013)	0.077 (0.015)	−0.155 (0.012)	0.008 (0.012)	0.127 (0.015)	0.103 (0.013)
R^2	0.032	0.011	0.007	0.006	0.024	0.000	0.016	0.011
Input-output	0.205 (0.037)	0.167 (0.028)	0.130 (0.022)	0.112 (0.022)	0.100 (0.019)	0.162 (0.029)	0.188 (0.036)	0.112 (0.029)
R^2	0.042	0.028	0.017	0.012	0.010	0.026	0.035	0.013
Technology flows Scherer R&D	0.180 (0.037)	0.148 (0.031)	0.107 (0.019)	0.089 (0.024)	0.046 (0.019)	0.107 (0.029)	0.136 (0.038)	0.094 (0.029)
R^2	0.032	0.022	0.012	0.008	0.002	0.011	0.019	0.009
Technology flows patent citations	0.081 (0.012)	0.100 (0.016)	0.085 (0.013)	0.068 (0.013)	−0.001 (0.012)	0.056 (0.012)	0.103 (0.012)	0.092 (0.013)
R^2	0.007	0.010	0.007	0.005	0.000	0.031	0.011	0.008

OLS-Multivariate

TABLE 4—OLS MULTIVARIATE SPECIFICATIONS FOR PAIRWISE COAGGLOMERATION

	EG coaggl. index with state total emp.				DO coaggl. index, 250 mi.			
	Base	Exclude	Separate	Exclude	Base	Exclude	Separate	Exclude
	estimation	natural	input &	pairs in	estimation	natural	input &	pairs in
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Natural advantages [DV specific]	0.163 (0.017)		0.162 (0.017)	0.172 (0.016)	0.251 (0.012)		0.252 (0.012)	0.253 (0.013)
Labor correlation	0.118 (0.011)	0.146 (0.012)	0.114 (0.011)	0.085 (0.012)	0.069 (0.012)	0.098 (0.013)	0.066 (0.012)	0.029 (0.012)
Input-output	0.146 (0.032)	0.149 (0.032)		0.110 (0.022)	0.162 (0.035)	0.150 (0.035)		0.177 (0.032)
Input			0.106 (0.029)				0.097 (0.029)	
Output			0.093 (0.039)				0.107 (0.038)	
Technology flows Scherer R&D	0.096 (0.035)	0.112 (0.035)	0.079 (0.035)	0.046 (0.019)	0.076 (0.033)	0.075 (0.034)	0.065 (0.032)	0.033 (0.020)
R^2	0.103	0.077	0.110	0.059	0.113	0.051	0.117	0.102
Observations	7,381	7,381	7,381	7,000	7,381	7,381	7,381	7,000

IV

- Industries may be collocated not because of Marshallian factors.
- Reverse causality: collocation could be driving input-output linkages or hiring patterns
 - Industries may be collocated due to some random event, but their flexible production technologies allow them to adjust to nearby labour and material inputs
- Unmodeled natural advantages may bias estimates

IV

- UK IV
 - Measures of input-out linkage, labour market, and technology for U.K. industries (two industries randomly coagglomerated in U.S. will not be coagglomerated in the U.K.)
- US Spatial IV
 - Measures input-output and labour patterns in one industry in places where the other industry is quite rare
 - Addresses issue where input / labour choices is affected by proximity of plants in the other industry (biasing measures of the Marshallian forces variables)

IV-Results

TABLE 5—IV MULTIVARIATE SPECIFICATIONS FOR PAIRWISE COAGGLOMERATION

	EG coaggl. index with state total emp.			DO coaggl. index, 250 mi.		
	Base	UK	US spatial	Base	UK	US spatial
	OLS	IV	IV	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Natural advantages	0.173	0.173	0.171	0.254	0.210	0.233
[DV specific]	(0.016)	(0.019)	(0.016)	(0.013)	(0.016)	(0.012)
Labor correlation	0.083	0.079	0.091	0.027	0.501	0.248
	(0.012)	(0.060)	(0.023)	(0.012)	(0.060)	(0.023)
Input-output	0.122	0.191	0.185	0.186	0.164	0.213
	(0.023)	(0.048)	(0.036)	(0.031)	(0.054)	(0.049)
Observations	7,000	7,000	7,000	7,000	7,000	7,000

End