Measuring Industry Concentration: Discussion of Ellison and Glaeser, JPE 1997

Nathan Schiff
Shanghai University of Finance and Economics

Graduate Urban Economics, Lecture 9
April 26th, 2023

Midterm Research Outline: Due May 10

Each student should write a “research outline” using the guidelines I send you. The outline should be 2-3 pages and is due two weeks from today.

The purpose is just to help you make progress on your final research proposal, which is due at the end of class

The most important part of the outline is a clear discussion of your intended research question, as well as discussion of the existing literature (see guideline document)

Write as much as you can, including discussion of any potential problems with your project. I will provide detailed individual feedback to each student, so the more information I have the more helpful I can (hopefully) be.

Industrial clusters in Guangdong

Selected specialised clusters in southern China

Town, *city*	Product	Town's claimed contribution
Ouhai, *Wenzhou*	Eyewear	90% of global-brand products (2016)
Jinjiang, *Quanzhou*	Zips	10% of global production (2011)
Chenghai, *Shantou*	Toys	30% of global production (2014)
Shiling, *Guangzhou*	Bags, suitcases	70% of European and US mass market (2013)
Gurao, *Shantou*	Bras	12% of China's production (2014)
Tangxia, *Dongguan*	Golf products	40% of global production (2011)

*Places indicated may include multiple clusters

Economist.com

Measuring Concentration

Many theory papers suggest large productivity advantages through clustering of firms in same geographic location (ex Duranton and Puga 2004 review)

Famous examples of Silicon Valley and Detroit, or Dalton, GA carpet cluster (Krugman)

This paper: how do we know when an industry is clustered?

Extremely influential paper: 4000+ citations

Many Follow-up Papers

Papers offering new methods of measurement:

1. Ellison and Glaeser, AER PP, 1999
2. Dumais, Ellison, Glaeser, ReStat 2002
3. Duranton and Overman, ReStud, 2005
4. Mori, Nishikimi, Smith, ReStat, 2005
5. Guimaraes, Figueiredo, Woodward, Journal Regional Science, 2007
6. Ellison, Glaeser, Kerr, AER 2010

Countless papers use the methods of Ellison and Glaeser

Why do we care about clusters?

Urban economists: cluster externalities could be one explanation for the productivity advantages of cities. Large cities have many firms in the same industry, may lead to localization economies.

IO economists: understanding clustering helps to understand firm production

Note: EG looks at clusters of firms selling products nationally or internationally. There is a separate literature on clusters of retail firms selling locally; this generates a trade-off between demand/search externalities and competition (Wolinksy 1983, Konishi 2005)

Policy-makers: if there are clustering externalities, then coordination failures may result in too few clusters or clusters that are too small. This suggests there may be a role for policy in creating clusters.

Cluster Policy

Productivity advantages of clusters popularized by work of Michael Porter (see Porter, 1990, "The Competitive Advantage of Nations"; also nice free article in Harvard Business Review, 1998)

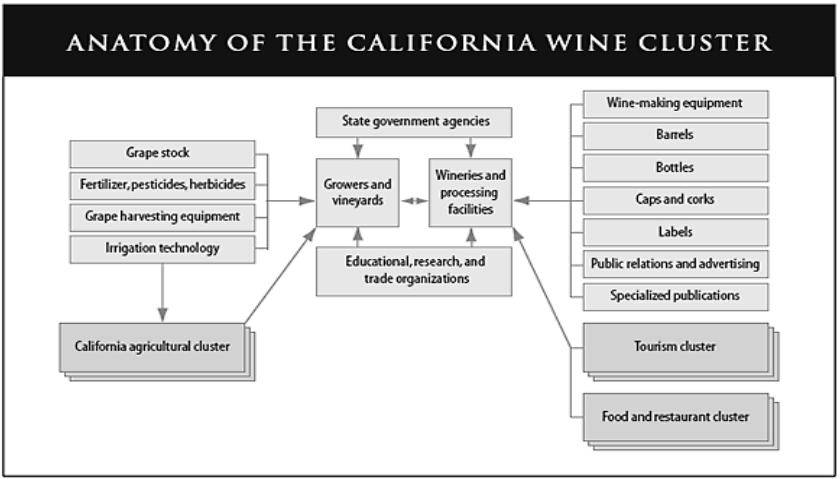
Many governments have tried to create clusters

Evidence suggests that it is quite difficult to create a cluster, see Duranton, "California Dreamin': The Feeble Case for Cluster Policies," (2011)

Related work studies China's research park policies. Zheng, Sun, Wu, and Kahn (JUE 2017, and follow-up papers) find some evidence of spillovers, and also growth in local housing and consumption

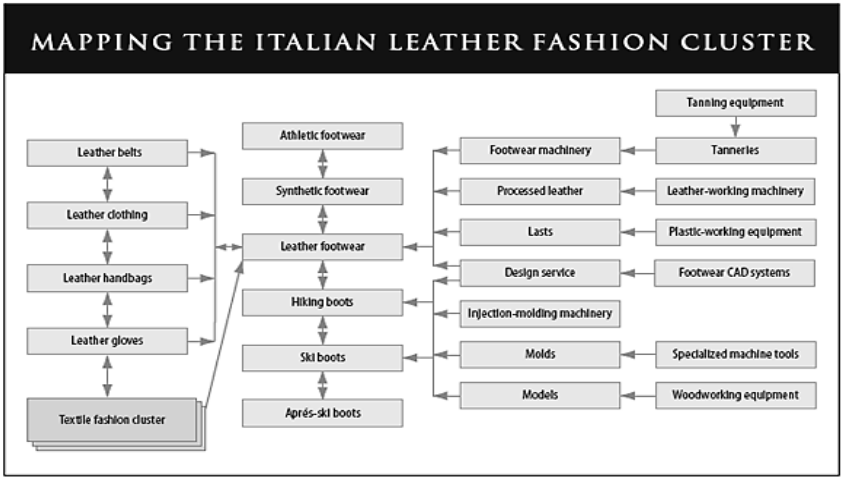
Note: Zheng et. al. would be a good paper for a student presentation

California Wine Cluster from Porter, HBR 1998



Anatomy Of The California Wine Cluster

Italian Leather Cluster from Porter, HBR 1998



Mapping the Italian Leather Fashion Cluster

Choice Model of Location

Profit of a location

Within an industry, the profit to business (plant) k of location i is:

$$\log \pi_{ki} = \log \bar{\pi}_i + g_i(\nu_i, ..., \nu_{k-1}) + \epsilon_{ki} \quad (1)$$

The variable $\bar{\pi}_i$ captures fixed location characteristics; it does not depend on number of firms choosing location i

These are commonly referred to as “natural advantages”; EG cite wine regions and coastal ship-building areas as examples

The function $g()$ captures spillover or agglomeration effects of previous $k-1$ firms (ν) choosing location i

The error term ϵ_{ki} is an idiosyncratic term, often thought of as a match between firm k and location i

Choice model with no spillovers: Assumption 1

If we assume that ϵ_{ki} are i.i.d. Extreme Value Type 1 then we have the logit model:

$$prob\{\nu_k = i|\bar{\pi}_1, ..., \bar{\pi}_M\} = \frac{\bar{\pi}_i}{\sum_j \bar{\pi}_j} \tag{1a}$$

First assumption: assume expected probability of firm k in some industry j choosing location i is equal to overall manufacturing employment of that location: x_i (x_i is all manufacturing, not just in industry j):

$$E_{\bar{\pi}_1,...,\bar{\pi}_M} \frac{\bar{\pi}_i}{\sum_j \bar{\pi}_j} = x_i \tag{2}$$

Choice Model: Assumption 2

Second assumption: assume variance of joint distribution of natural advantages (*na*) is governed by single parameter γ^{na}:

$$var\left(\frac{\bar{\pi}_i}{\sum_j \bar{\pi}_j}\right) = \gamma^{na} x_i(1 - x_i), \text{ where } \gamma^{na} \in [0, 1] \tag{3}$$

If $\gamma^{na} = 0$ there is no variance and plants choice probabilities perfectly match overall manufacturing distribution x_i

If $\gamma^{na} = 1$ then variance is maximized, which requires one location has all firms

If one location gets all firms then share of firms in any location is either one or zero but expected share is x_i; this is like $\frac{\bar{\pi}_i}{\sum_j \bar{\pi}_j}$ as the probability p from a Bernoulli distribution (0 or 1)

Implementing Distributional Assumptions

Authors note that one way to allow 2) and 3) is:

- Assume $\{\bar{\pi}\}$ are i.i.d. where $2[(1-\gamma^{na})/\gamma^{na}]\bar{\pi}_i \sim \chi^2$ with df=$2[(1-\gamma^{na})/\gamma^{na}]x_i$
- Then $E[\bar{\pi}_i] = x_i$ and $var[\bar{\pi}_i] = [\gamma^{na}/(1-\gamma^{na})]x_i$
- Note that $\chi^2(k)$ has mean=k and variance=$2k$, where k is d.f.
- Therefore if $E[\bar{\pi}_i * k/x_i] = k$ then $E[\bar{\pi}_i] = k * x_i/k$. Similarly if $Var[\bar{\pi}_i * k/x_i] = 2k$ then $Var[\bar{\pi}_i] = 2k * (x_i^2/k^2)$

Guimaraes et. al. show an easier way to implement this using a Dirichlet distribution

Allowing for Spillovers

$$\log \pi_{ki} = \log \bar{\pi}_i + g_i(\nu_i, ..., \nu_{k-1}) + \epsilon_{ki} \tag{1}$$

In order to implement $g()$ authors assume that if spillovers exist between two plants then the plants must locate in the same location

$$\log \pi_{ki} = \log \bar{\pi}_i + \sum_{l \neq k} \ell_{kl}(1 - u_{li})(-\infty) + \epsilon_{ki} \tag{4}$$

The variable u_{li} is equal to 1 if plant l is in location i

The $\{\ell_{kl}\}$ are Bernoulli variables equal to one with probability γ^s, indicating whether a spillover exists between plants k and l

Authors note that firm k only considers previous $k-1$ firms and that this is consistent with forward looking plants in rational expectations model

Defining Geographic Concentration

Let s_i be share of industry's employment in area i

Industry geographic concentration can be specified as G:

$$G \equiv \sum_i (s_i - x_i)^2 \qquad \text{(p895.1)}$$

Where share s_i is determined endogenously as:

$$s_i = \sum_k z_k u_{ki} \qquad \text{(p895.2)}$$

The z_k is kth plant's share of industry employment, u_{ki} indicates whether plant k located in site i

Expected Value of Geographic Concentration

$$E(G) = (1 - \sum x_i^2)[\gamma + (1 - \gamma)H]$$
$$H \equiv \sum_k z_k^2 \qquad \text{(Prop1)}$$
$$\gamma = \gamma^{na} + \gamma^{s} - \gamma^{na}\gamma^{s}$$

This expression for $E(G)$ comes from application of law of iterated expectations, see proof p896

Observational equivalence of natural advantage and spillovers: any $\gamma \in [0, 1]$ is compatible with spillovers, natural advantage, or both

Important point: we cannot distinguish spillovers from location specific features using concentration data alone!

So how can we measure spillovers (agglomeration effects)?

Co-Agglomeration

Co-agglomeration is defined as pairs or groups of industries locating together

The idea is that there may be spillovers *across* industries (urbanization), rather than solely *within* industries (localization)

Authors use more “reduced-form” approach to define expected concentration of *r* industries in a *group* in terms of correlation of location choices

The group is a collection of *r* industries that might have reason to co-locate, either due to shared natural advantages (ex: multiple industries may rely on access to the coast) or spillovers

Co-Agglomeration of Industries in Group

First authors specify parameter of co-location:

$$corr(u_{ki}, u_{li}) = \begin{cases} \gamma_j, & \text{if plants } k \text{ and } l \text{ both belong to industry } j \\ \gamma_0, & \text{otherwise} \end{cases}$$

Then expected concentration of the *group* of r industries is:

$$E(G) = \left(1 - \sum_i x_i^2\right)\left[H + \gamma_0\left(1 - \sum_{j=1}^{r} \omega_j^2\right) + \sum_{j=1}^{r} \gamma_j \omega_j^2 (1 - H_j)\right] \qquad \text{(p898)}$$

In above, ω is industry j's share of total employment in r industries

If $\gamma_0 = 0$ then *no* co-agglomeration (typo on p899); if $\gamma_0 = \gamma_1 = \gamma_r$ then agglomeration benefit within industry same as across industries

Measuring Industry Concentration and Coagglomeration with Industry-level Data on Employment

Measurement: Industry Concentration Index

Solve $E(G)$ equation for γ:

$$\gamma \equiv \frac{G - \left(1 - \sum_i x_i^2\right) H}{\left(1 - \sum_i x_i^2\right)(1 - H)} \tag{5}$$

or

$$\gamma \equiv \frac{\sum_{i=1}^{M}(s_i - x_i)^2 - \left(1 - \sum_{i=1}^{M}(x_i)^2\right)\sum_{j=1}^{N} z_j^2}{\left(1 - \sum_{i=1}^{M}(x_i)^2\right)\left(1 - \sum_{j=1}^{N} z_j^2\right)} \tag{5}$$

This is unbiased estimator of γ (we inserted G for $E[G]$)

Requires data on distribution of: 1) overall manufacturing employment 2) industry employment 3) plant employment

Properties of EG Index

According to EG:

1. Easy to compute because only requires limited data (NS: plant size data often unavailable, can assume uniform distribution so $H = 1/N$
2. Scale allows comparison with null hypothesis of no concentration beyond overall manufacturing: $E[\gamma] = 0$. Footnote 13 shows how to calculate variance for G; for γ authors assume Dirichlet (see Guimaraes et.al. 2007)
3. Comparable across industries where size distribution of firms differs: $E[G]$ is independent of number of plants and distribution of sizes (NS: Mori, Nishikimi, Smith argue differently)
4. Index is scale invariant: value of G should be the same no matter how data is aggregated, *if spillovers only exist at identical locations–infinite spatial decay*. This is an important issue in urban/spatial work (modifiable areal unit problem).

Understanding Magnitude

If $\gamma = 0$ no concentration beyond overall manufacturing; $\gamma = 1$ maximum expected concentration, but when is γ "big?"

Exercise 1: Compare to estimates of elasticity η of location wrt costs

If elasticity $\eta = 25$, then a 3% decrease in costs results in a 75% increase in likelihood of locating in a place

Following EG model, one standard deviation increase in probability p_i of locating in i is $sd(p_i) = \sqrt{\gamma * x_i * (1 - x_i)} = \sqrt{\gamma * (1 - x_i)/x_i} * x_i$

There are 50 states, so if $x_i = 0.02$ and $\gamma = .01$, then $sd(p_i) = 0.7 * x_i$, or about a 70% increase (close to effect of $\eta = 25$ and 3% cost shock)

A 9% cost shock with $\eta = 25$ is similar to $\gamma = .1$ (with $x_i = 0.02$)

EG note that sd of wages across states is 8-10%, thus $\gamma = 0.01$ is small, $\gamma = 0.1$ is big

Natural Advantage and State Size

Exercise 2: Use model and assume a χ^2 distribution of $\bar{\pi}_i$ to compare effect of state size and natural advantages on location:

TABLE 1

EFFECT OF γ NATURAL ADVANTAGE RELATIVE TO STATE SIZE

γ^{na}	prob$\{\bar{\pi}_{IA} > \bar{\pi}_{GA}\}$	prob$\{\bar{\pi}_{IA} > \bar{\pi}_{MI}\}$	prob$\{\bar{\pi}_{IA} > \bar{\pi}_{CA}\}$
.005	.07	.006	.00
.01	.14	.03	.00
.02	.20	.08	.006
.05	.25	.14	.04
.10	.26	.15	.07
1.00	.27	.17	.09

Measurement of Coagglomeration

Can derive a similar index of coagglomeration:

$$\gamma^c \equiv \frac{\left[G/\left(1-\sum_i x_i^2\right)\right] - H - \sum_{j=1}^{r} \hat{\gamma}_j \omega_j^2 (1 - H_j)}{1 - \sum_{j=1}^{r} \omega_j^2} \tag{6}$$

This γ^c is a measure of the γ_0–coagglomeration effect

Finally, define a measure of what proportion of group concentration is due to industry-specific concentration:

$$\lambda \equiv \frac{\gamma^c}{\sum_j \omega_j \hat{\gamma}_j} \tag{7}$$

Results: Estimated Values of Concentration and Co-Agglomeration Indices using US Data

Results for US

Use 459 manufacturing industries (four digit level) from 1987

Use 50 US states plus Washington D.C. as geographic regions

Employment data from Census of Manufactures

Look at:

1. When is G statistically different from null value with no spillovers or natural advantages? Null value of G with no sp or na is $(1 - \sum x_i^2) * H$. Use formula for $Var(G)$ from footnote 13 to calculate statistical significance of $G - (1 - \sum x_i^2) * H$.
2. What is overall distribution of γ across industries?
3. How do measures compare across different spatial units?

Findings on concentration

1. Nearly all 459 industries show statistically significant concentration (*G* larger than null)
2. Most show only "slight" concentration: 43% of industries have $\gamma < .02$
3. However, there is a thick right tail with 25% having $\gamma > .05$ (concentrated) and 14% very concentrated ($\gamma > .1$)
4. Authors find accounting for randomness or overall manufacturing distribution is important: in $1/3$ of industries randomness accounts for same amount of concentration as spillovers + natural advantage
5. Within county spillovers are stronger than across county (based on difference in γ using counties vs states)
6. However, does appear that there are spillovers across counties (at state level)

Histogram of γ, 4-digit industries

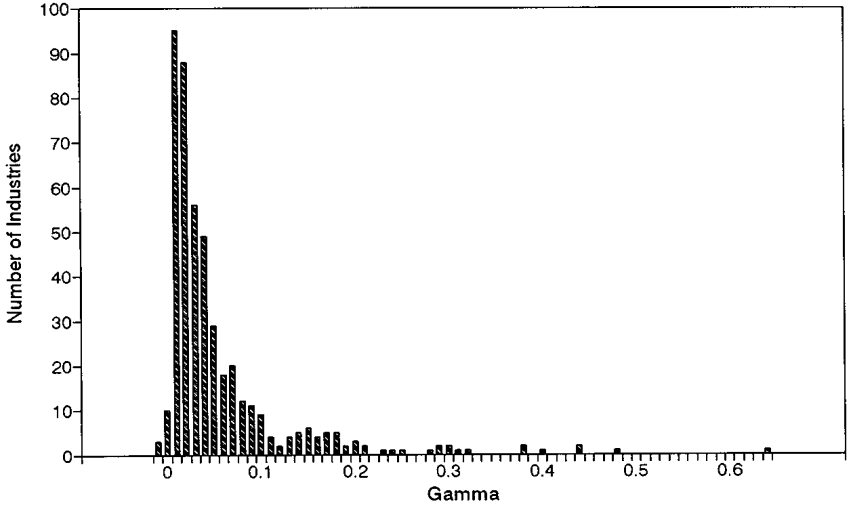


FIG. 1.—Histogram of γ (four-digit industries)

Concentration of Industries

TABLE 3

CONCENTRATION BY TWO-DIGIT CATEGORY

TWO-DIGIT INDUSTRY	NUMBER OF FOUR-DIGIT SUBINDUSTRIES	PERCENTAGE OF FOUR-DIGIT INDUSTRIES WITH		
		$\gamma < .02$	$\gamma \in [.02, .05]$	$\gamma > .05$
20 Food and kindred products	49	47	18	35
21 Tobacco products	4	0	0	100
22 Textile mill products	23	9	13	78
23 Apparel and other textile products	31	13	42	45
24 Lumber and wood products	17	29	47	24
25 Furniture and fixtures	13	69	8	23
26 Paper and allied products	17	53	47	0
27 Printing and publishing	14	71	14	14
28 Chemicals and allied products	31	38	24	38
29 Petroleum and coal products	5	60	0	40
30 Rubber and miscellaneous plastics	15	73	27	0
31 Leather and leather products	11	0	36	64
32 Stone, clay, and glass products	26	58	27	15
33 Primary metal industries	26	39	35	27
34 Fabricated metal products	38	61	32	8
35 Industrial machinery and equipment	51	49	26	26
36 Electronic and other electric equipment	37	41	46	14
37 Transportation equipment	18	28	33	39
38 Instruments and related products	17	47	41	11
39 Miscellaneous manufacturing industries	18	44	22	33

Most and Least Concentrated

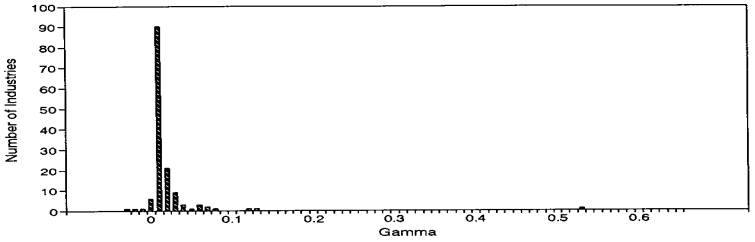
TABLE 4

MOST AND LEAST LOCALIZED INDUSTRIES

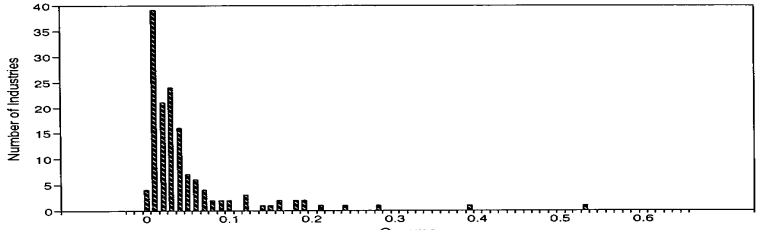
Four-Digit Industry	H	G	γ
	15 Most Localized Industries		
2371 Fur goods	.007	.60	.63
2084 Wines, brandy, brandy spirits	.041	.48	.48
2252 Hosiery not elsewhere classified	.008	.42	.44
3533 Oil and gas field machinery	.015	.42	.43
2251 Women's hosiery	.028	.40	.40
2273 Carpets and rugs	.013	.37	.38
2429 Special product sawmills not elsewhere classified	.009	.36	.37
3961 Costume jewelry	.017	.32	.32
2895 Carbon black	.054	.32	.30
3915 Jewelers' materials, lapidary	.025	.30	.30
2874 Phosphatic fertilizers	.066	.32	.29
2061 Raw cane sugar	.038	.30	.29
2281 Yarn mills, except wool	.005	.27	.28
2034 Dehydrated fruits, vegetables, soups	.030	.29	.28
3761 Guided missiles, space vehicles	.046	.27	.25
	15 Least Localized Industries		
3021 Rubber and plastics footwear	.06	.05	−.013
2032 Canned specialties	.03	.02	−.012
2082 Malt beverages	.04	.03	−.010
3635 Household vacuum cleaners	.18	.17	−.009
3652 Prerecorded records and tapes	.04	.03	−.008
3482 Small-arms ammunition	.18	.17	−.004
3324 Steel investment foundries	.04	.04	−.003
3534 Elevators and moving stairways	.03	.03	−.001
2052 Cookies and crackers	.03	.03	−.0009
2098 Macaroni and spaghetti	.03	.03	−.0008
3262 Vitreous china table, kitchenware	.13	.12	−.0006
2035 Pickles, sauces, salad dressings	.01	.01	−.0003
3821 Laboratory apparatus and furniture	.02	.02	−.0002
2062 Cane sugar refining	.11	.10	.0002
3433 Heating equipment except electric	.01	.01	.0002

Comparing γ at county and state levels

County Level Gammas



State Level Gammas



Findings on coagglomeration

To look at coagglomeration the authors use the classification codes to create industry groups, such as all industries in same 3 digit class or same 2 digit class

Look at measure of λ, find:

1. Value of λ evenly spread between 0 and 0.8 (fig 3)
2. Substantial heterogeneity for both 3 and 2 digit classes
3. Also try and look at colocation of upstream-downstream industries (upstream provides inputs to downstream); find highly concentrated (not surprising)

Histogram of coagglomeration estimates (λ)

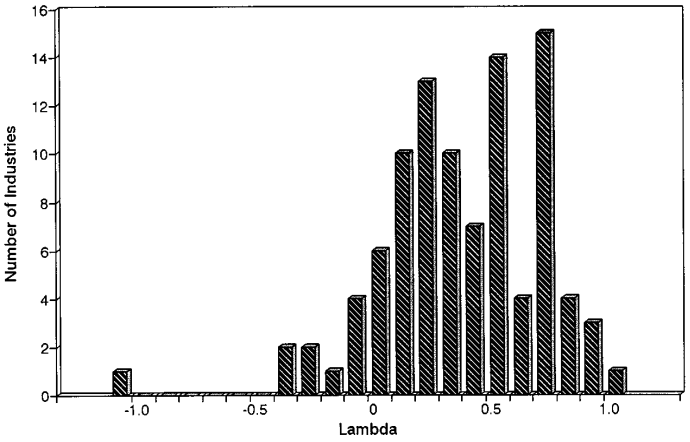


FIG. 3.—Histogram of λ: extent of spillovers between four-digit subindustries of three-digit industries.

Coagglomeration: spillovers across industries

TABLE 6

EXTENT OF SPILLOVERS BETWEEN THREE-DIGIT INDUSTRIES

Two-Digit Industry	γ^c	λ
Food and kindred products	.002	.14
Tobacco products	.151	.88
Textile mill products	.115	.61
Apparel and other textiles	.010	.29
Lumber and wood products	.016	.63
Furniture and fixtures	.001	.02
Paper and allied products	.005	.31
Printing and publishing	.005	.48
Chemicals and allied products	.007	.25
Petroleum and coal products	.007	.12
Rubber and miscellaneous plastics	.003	.38
Leather and leather products	.017	.31
Stone, clay, and glass products	.002	.20
Primary metal industries	.012	.41
Fabricated metal products	.003	.22
Industrial machinery and equipment	.000	.00
Electronic and other electric equipment	.000	.02
Transportation equipment	−.001	−.08
Instruments and related products	.013	.36
Miscellaneous manufacturing	.011	.34

Very Impressive Paper, Still Some Critics

1. Method requires plant-level employment data, which is often unavailable
2. Further, Figueiredo, Guimaraes, Woodward (FGW) write that use of employment leads to strange results; argue using plant locations alone is more intuitive measure of localization
3. Mori, Nishikimi, Smith (MNS) show that ordering of industries by concentration is mostly unaffected by simply assuming equal distribution of employment across plants
4. Duranton and Overman (DO) (and EG) note γ does not allow for spatial effects: all locations are assumed i.i.d. when in fact distances between plants vary tremendously
5. MNS show that comparison against overall manufacturing does bias index because larger industries are naturally a larger part of overall manufacturing

Extensions

Duranton and Overman (ReStud 2005): use point data (lat,long) on businesses in England to look at concentration. Show how using inter-firm distances can allow for concentration at varying distances (allows for spatial decay of spillovers); somewhat similar to Ripley's K

Mori, Nishikimi, Smith (ReStat 2005): use null spatial distribution of complete spatial randomness (CSR); show that this allows for more robust comparisons across industries of different sizes

Figueiredo, Guimaraes, Woodward (JRS 2007): redo EG model using Dirichlet distribution and plant count. Provides a simpler measure with roughly same intuition and smaller variance

FGW (JRS 2011): extend original EG measure to incorporate spatial correlation; (for geographers: basically add Moran's I to original EG measure)