What Causes Industry Agglomeration?
Evidence from Coagglomeration Patterns

By Glenn Ellison, Edward L. Glaeser, and William R. Kerr[*]

Why do firms cluster near one another? We test Marshall's theories of industrial agglomeration by examining which industries locate near one another, or coagglomerate. We construct pairwise coagglomeration indices for US manufacturing industries from the Economic Census. We then relate coagglomeration levels to the degree to which industry pairs share goods, labor, or ideas. To reduce reverse causality, where collocation drives input-output linkages or hiring patterns, we use data from UK industries and from US areas where the two industries are not collocated. All three of Marshall's theories of agglomeration are supported, with input-output linkages particularly important. (JEL L14, L60, O33, R23, R32)

Industries are geographically concentrated.[1] This concentration is too great to be explained by exogenous spatial differences in natural advantage.[2] Why does this concentration occur? There is no shortage of theories that can explain the agglomeration of industries.[3] But we have very little empirical work assessing the relative importance, or even general correctness, of these theories. This paper exploits patterns of industry coagglomeration to measure the relative importance of different theories of industry agglomeration.

The benefits of agglomeration ultimately reflect gains that occur when proximity reduces transport costs. Marshall (1920) emphasized three different types of transport costs—the costs of moving goods, people, and ideas—that can be reduced by industrial agglomeration. First, he argued that firms will locate near suppliers or customers to save shipping costs. Second, he developed a theory of labor market pooling to explain clustering. Finally, he began the theory of intellectual spillovers by arguing that in agglomerations, "the mysteries of the trade become no mystery, but are, as it were, in the air." Firms, such as those described by AnnaLee Saxenian (1996) in Silicon Valley, locate near one another to learn and to speed their rate of innovation.

There are certainly anecdotal examples of individual industries that have agglomerated to reduce one or more these transport costs. It is challenging, however, to assess their relative

[1] See P. Sargant Florence (1948), E. M. Hoover (1948), Victor R. Fuchs (1962), Paul R. Krugman (1991a), Ellison and Glaeser (1997), and Gilles Duranton and Henry G. Overman (2005, 2008).

[2] See Ellison and Glaeser (1999).

[3] See Johann Heinrich von Thünen (1826), Alfred Marshall (1920), Krugman (1991b), Thomas J. Holmes (1998), and Mohammad Arzaghi and J. Vernon Henderson (2008).

importance to agglomeration across industries as a whole. Each Marshallian theory predicts that the same thing will happen for similar reasons: plants will locate near other plants in the same industry because there is a benefit to locating near plants that share some characteristic. Our empirical approach exploits the information that can be found in coagglomeration patterns. Plants are similar to the other plants in their industry along many dimensions. But across industries, plants are similar in some dimensions and not in others. For example, some industry pairs exchange goods but employ very different workers. Other industries hire similar workers but never trade with each other. Hence, one can gain insight into which theories are more important by looking at which cross-industry similarities best predict which industries coagglomerate.[4]

Section I describes the data used to generate our coagglomeration indices. We use establishment level data from the Census of Manufacturing to calculate the discrete index of Ellison and Glaeser (1997) and an approximation of the continuous metric of Duranton and Overman (2005).

Section II reviews Marshall's three theories and discusses our empirical measure of the importance of each theory for each industry pair. For example, input-output linkages enable us to test whether different industries collocate to reduce the costs of shipping between customers and suppliers. Metrics for the extent to which industries share workers and ideas are similarly constructed. We also describe our calculations of the expected coagglomeration of each industry pair that would be expected to arise from the uneven spatial distribution of natural advantages, following Ellison and Glaeser (1999).

Section III presents our main empirical results. The ordinary least squares relationships support the importance of all three Marshallian theories and the importance of shared natural advantages. We estimate that shared natural advantages are more important than any single Marshallian factor, but not as important as the cumulative effect of the three Marshallian factors. Among the Marshallian factors, customer-supplier relationships have the strongest effect. These input-output linkages are closely followed by similar labor needs. Our proxies for intellectual spillovers are weaker than the other factors but still economically and statistically important. These relative rankings are subject to the caveat that we have imperfect proxies for the variables of interest and a finite number of measured natural advantages. Overall, each of the agglomeration theories plays a measurable role in agglomeration within manufacturing.

One concern with these results is that industrial relationships may be the result of collocation instead of the cause of collocation. Some industries may be flexible enough in their production processes that they adjust to nearby resources of labor and material inputs. If two industries locate near one another for random reasons, then those industries might both start using the same labor and raw materials that are readily available in their shared location. Section IV addresses this concern by developing two sets of instrumental variables. First, we use characteristics of UK industries. Coagglomeration patterns due to unobserved, shared natural advantages or purely random events may differ in the United Kingdom, in which case UK industry characteristics can help identify effects that are due to innate similarities between industry pairs.

Second, we measure the similarities of industries using only data from plants located in different parts of the United States. Even in highly coagglomerated industry pairs, there will typically be some plants in each industry that are not located near plants in the other industry. We can use these isolated plants to estimate input-output matrices and labor usage. Since these plants are not near the other industry, their inputs, outputs, and labor decisions are less likely to be driven by

---

[4] In a similar vein, Henderson (2003) examines how plant level productivity is related to the set of plants in the area. A second approach to this problem pioneered by David B. Audretsch and Maryann P. Feldman (1996) and Stuart S. Rosenthal and William C. Strange (2001) is to examine cross-industry variation in the degree of agglomeration, such as regressing the degree to which an industry is agglomerated on the importance of R&D to the industry. Michael Greenstone, Richard Hornbeck, and Enrico Moretti (2008), Carlo Menon (2008), and Mercedes Delgado, Michael E. Porter, and Scott Stern (2008) address related empirical issues.

common omitted factors or by the influence of proximity to the other industry. We use plant level detail from the Census of Manufacturing and individual level data from the Census Integrated Public Use Microdata Series (IPUMS) to develop measures of industry pair similarity based upon characteristics of the noncoagglomerated plants. Our IV regressions provide additional support for the view that input-output relationships and labor market pooling benefits are both significant drivers of industry agglomeration.

I. US Manufacturing Coagglomeration

We compute pairwise coagglomeration measures for manufacturing industries using the confidential plant level data from the US Census Bureau's Census of Manufacturing.[5] Each Census documents the operations of approximately 300,000 establishments employing about 17 million workers. We focus on the three-digit level of the 1987 Standard Industrial Classification (SIC3). The sample contains 7,381 industry pair observations per year: all distinct coagglomeration pairs from 122 industries.[6]

We quantify industry pair coagglomeration in two ways. First, we use the Ellison and Glaeser (1997, hereafter EG) metric of coagglomeration. We do this at the state, Primary Metropolitan Statistical Area (PMSA), and county levels. We also use the longitudinal nature of the Census Bureau data to analyze coagglomeration of startup firms.[7] The EG coagglomeration index takes a simple form when applied to industry pairs (as opposed to larger groups). The index for the coagglomeration of industries i and j is

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

where m indexes geographic areas. s_{mi} is the share of industry i's employment contained in area m. x_m measures the aggregate size of area m, which we model as the mean employment share in the region across manufacturing industries. The Mathematical Appendix demonstrates that this index can be regarded as a measure of the strength of agglomerative forces in a particular model of firm location.

Our second coagglomeration metric is a "lumpy" approximation to the continuous index developed by Duranton and Overman (2005, hereafter DO). DO criticize indices like EG that employ discrete spatial units. This discreteness in effect makes the distance from Detroit to Chicago equivalent to that of Detroit to Miami. DO instead propose analyzing coagglomeration through a continuous index

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

[5] See Timothy Dunne, Mark J. Roberts, and Larry Samuelson (1989a, 1989b) and Steven J. Davis, John C. Haltiwanger, and Scott Schuh (1996).

[6] We exclude Tobacco (210s), Apparel (230s), portions of Printing and Publishing (277–279), Secondary Nonferrous Metals (334), and Search and Navigation Equipment (381). These exclusions are primarily due to data constraints and are documented in the online Appendix.

[7] Relevant employments for each geographic unit are calculated by aggregating employments from individual establishments. Related work on entrepreneurship patterns includes Guy Dumais, Ellison, and Glaeser (2002), David H. Autor, William R. Kerr, and Adriana D. Kugler (2007), Glaeser and Kerr (2009), and Kerr and Ramana Nanda (2009).

TABLE 1—DESCRIPTIVE STATISTICS FOR PAIRWISE COAGGLOMERATION REGRESSIONS

	Mean	SD	Minimum	Maximum
Panel A. Pairwise EG coagglomeration measures				
EG state total employment coagglomeration	0.000	0.013	−0.065	0.207
EG PMSA total employment coagglomeration	0.000	0.006	−0.025	0.119
EG county total employment coagglomeration	0.000	0.003	−0.018	0.080
EG state firm birth employment coagglomeration	0.000	0.015	−0.082	0.259
EG expected coagglomeration due to natural advantages	0.000	0.001	−0.008	0.022

| | Industry count | Relevant industries (nonzero) | | |
		Mean	SD	Maximum
Panel B. Pairwise DO coagglomeration measures				
DO global localization coagglomeration, 1,000 mi.	7,371	0.133	0.073	0.454
DO global dispersion coagglomeration, 1,000 mi.	10	0.592	0.078	0.746
DO expected global localization coagglomeration, 1,000 mi.	7,381	0.181	0.027	0.256
DO global localization coagglomeration, 250 mi.	6,429	0.017	0.019	0.283
DO global dispersion coagglomeration, 250 mi.	952	0.042	0.029	0.307
DO expected global localization coagglomeration, 250 mi.	7,381	0.029	0.010	0.077

	Mean	SD	Minimum	Maximum
Panel C. Marshallian factors				
Labor correlation	0.470	0.226	−0.046	1.000
Input-output maximum	0.007	0.029	0.000	0.823
Input maximum	0.005	0.019	0.000	0.392
Output maximum	0.005	0.026	0.000	0.823
Scherer R&D technical maximum	0.005	0.026	0.000	0.625
Patent citation technical maximum	0.015	0.025	0.000	0.400

Notes: Descriptive statistics for coagglomeration estimations. All pairwise combinations of manufacturing SIC3 industries are included, except those listed in the text, for 7,381 observations. EG and DO coagglomeration metrics are calculated from the 1987 and 1997 Census of Manufacturers, respectively. The distance threshold for determining global localization or dispersion is adjusted across DO row groupings. Natural advantages coagglomeration is estimated through predicted state-industry shares developed from exogenous local cost variables (e.g., coastal access, energy prices) and industry cost dependencies. Labor correlation indices are calculated from the BLS National Industry-Occupation Employment Matrix for 1987. Input-output relationships are calculated from the BEA Benchmark Input-Output Matrix for 1987. Technology flows are calculated from the Scherer (1984) R&D tables for the 1970s and from the NBER Patent Citation Database for 1975–1997. Online Appendix Tables 1–5 provide additional descriptive statistics.

where $d_{r,s}$ is the Euclidean distance between plants r and s, f is a Gaussian kernel density function with bandwidth h, and n_i and n_j are the number of plants in industries i and j, respectively. The summations are over every bilateral distance between plants of industry i and industry j (i.e., $n_i n_j$ distances).

This observed coagglomeration density is then compared to an underlying distribution of manufacturing activity akin to the x_m of EG. An industry pair is said to exhibit global localization (dispersion) if the observed coagglomeration density is substantially higher (lower) than the underlying distribution of manufacturing activity. This comparison is done over a specified distance horizon. We vary this distance threshold below from 100 to 1,000 miles, with our primary results taken from a 250-mile distance horizon.[8]

Panel A of Table 1 presents descriptive statistics for the EG metric. The mean EG pairwise coagglomeration is approximately zero. This is largely by definition: our benchmark measure of

---

[8] Specified distance thresholds are required as densities sum to one over the support. Thus, industry pairs that are more localized at shorter distances will be more dispersed than aggregate manufacturing activity at longer distances. DO consider a threshold of 180 kilometers for the United Kingdom, which is motivated by the median plant to plant distance in their sample. This distance is equivalent to 112 miles. The median plant distance is much larger for the United States, falling between 900 and 1,000 miles depending upon the weighting. We chose our four thresholds to span the actual physical distance studied by DO and the median plant concept.

TABLE 2—HIGHEST PAIRWISE COAGGLOMERATIONS

Rank	Industry 1	Industry 2	Coagglomeration
Panel A. EG index using 1987 state total employments			
1	Broadwoven mills, cotton (221)	Yarn and thread mills (228)	0.207
2	Knitting mills (225)	Yarn and thread mills (228)	0.187
3	Broadwoven mills, fiber (222)	Textile finishing (226)	0.178
4	Broadwoven mills, cotton (221)	Broadwoven mills, fiber (222)	0.171
5	Broadwoven mills, fiber (222)	Yarn and thread mills (228)	0.164
6	Handbags (317)	Photographic equipment (386)	0.155
7	Broadwoven mills, wool (223)	Carpets and rugs (227)	0.149
8	Carpets and rugs (227)	Yarn and thread mills (228)	0.142
9	Photographic equipment (386)	Jewelry, silverware, plated ware (391)	0.139
10	Textile finishing (226)	Yarn and thread mills (228)	0.138
11	Broadwoven mills, cotton (221)	Textile finishing (226)	0.137
12	Broadwoven mills, cotton (221)	Carpets and rugs (227)	0.137
13	Broadwoven mills, cotton (221)	Knitting mills (225)	0.136
14	Carpets and rugs (227)	Pulp mills (261)	0.110
15	Jewelry, silverware, plated ware (391)	Costume jewelry and notions (396)	0.107
Panel B. DO index using 1997 firm employments, 250 mi. threshold			
1	Broadwoven mills, fiber (222)	Yarn and thread mills (228)	0.283
2	Carpets and rugs (227)	Yarn and thread mills (228)	0.262
3	Broadwoven mills, fiber (222)	Carpets and rugs (227)	0.226
4	Broadwoven mills, cotton (221)	Yarn and thread mills (228)	0.219
5	Broadwoven mills, cotton (221)	Carpets and rugs (227)	0.218
6	Footwear cut stock (313)	Costume jewelry and notions (396)	0.217
7	Jewelry, silverware, plated ware (391)	Costume jewelry and notions (396)	0.208
8	Knitting mills (225)	Yarn and thread mills (228)	0.200
9	Broadwoven mills, fiber (222)	Knitting mills (225)	0.190
10	Broadwoven mills, cotton (221)	Broadwoven mills, fiber (222)	0.175
11	Textile finishing (226)	Yarn and thread mills (228)	0.163
12	Footwear cut stock (313)	Jewelry, silverware, plated ware (391)	0.157
13	Handbags (317)	Costume jewelry and notions (396)	0.151
14	Broadwoven mills, cotton (221)	Knitting mills (225)	0.149
15	Women's and misses' outerwear (233)	Costume jewelry and notions (396)	0.149

Notes: See Table 1.

an area's "size" is its share of manufacturing employment, so each industry's deviations from the benchmark will be approximately uncorrelated with the average of the deviations of all other industries. The standard deviation of the coagglomeration index is more interesting because it reflects the extent to which industry pairs are positively and negatively coagglomerated. The standard deviation is 0.013 at the state level. This can be compared with the mean within industry agglomeration level of 0.051 in Ellison and Glaeser (1997). Panel B presents descriptive statistics for the DO metric. Eighty-seven percent of industry pairs exhibit some degree of global localization to the 250-mile threshold.

Table 2 lists the 15 most coagglomerated industry pairs for the EG and DO metrics. Textile and apparel industries rank very high on both scales. These industries are heavily concentrated in North Carolina, South Carolina, and Georgia. Despite this clustering, these coagglomerations are not as strong as the largest within industry agglomerations. Many industry pairs have approximately zero coagglomeration. Negative values of the EG index arise when pairs of industries are agglomerated in different areas. The lowest EG value of −0.065 obtains for Guided Missiles and Space Vehicles (376) and Railroad Equipment (374) industries. The most dispersed industry pair using the DO metric at 250 miles is Guided Missiles and Space Vehicles (376) and Pulp Mills (261). The correlation of EG and DO metrics across all industry pairs is 0.4.

The Data and Empirical Appendix provides additional information regarding the Census Bureau data, the construction of these two metrics, and their descriptive statistics. The

continuous DO methodology is computationally demanding, and the Appendix fully documents the approximate DO index that we computed to make using the index more tractable on 7,381 industry pairs. These approximations also respond to data limitations of the Census of Manufacturing.[9]

II. Why Do Firms Agglomerate? Empirical Methodology

The gains from concentration, whether in cities or geographic clusters, come from reducing some form of transport costs. Marshall emphasized that these transport costs could be for goods, people, or ideas. Our primary goal is to assess the relative importance of these Marshallian forces. We do so via cross-sectional regressions of pairwise coagglomeration on proxies for the importance of Marshall's agglomerative forces.

Agglomeration and coagglomeration can also appear empirically even if there are no gains from locational proximity. Natural advantages, such as the presence of natural inputs, differ spatially, and firms may choose locations to gain access to those inputs. We therefore also control for expected coagglomeration of industry pairs arising from common dependencies on certain natural advantages (e.g., coastal access, energy prices). Beyond just controlling for important omitted variables, we also view natural advantages as a benchmark for assessing the relative importance of goods, people, and ideas in the location decisions of manufacturing firms.

In the following subsections, we briefly discuss the Marshallian forces and the metrics we use to capture their relevance to each industry pair. We then describe how we model coagglomeration due to natural advantages and some of the limits of our approach. Our initial analysis will consist of OLS regressions of our concentration indices on these measures. Where possible, we focus our estimation and data construction on the 1987 cross section.[10] The online Appendix provides additional details and descriptive statistics.

A. *Proximity to Customers and Suppliers: Goods*

Firms locate near one another to reduce the costs of obtaining inputs or shipping goods to downstream customers. When inputs are far away from the eventual market, Marshall (1920) argued that firms will trade off the distance between customers and suppliers based on the costs of moving raw inputs and finished goods. For example, sugar refining was one of New York City's largest industries in the nineteenth century because of transport costs. Sugar was refined in New York, rather than on tropical plantations, because refined sugar crystals coalesce during a long sea voyage in a hot ship's hull. Sugar refining took place in New York, rather than in eventual small town markets, to exploit scale economies. Once Armour's refrigerated rail cars made it possible to ship cold beef, cattle were slaughtered in Chicago's vast stockyards to save the costs of shipping live beef east. The "new economic geography" of Masahisa Fujita, Krugman, and Anthony J. Venables (1999) views reducing the costs of transporting goods as the central driver behind agglomeration.

To assess the importance of this factor, we use the 1987 Benchmark Input-Output Accounts published by the Bureau of Economic Analysis (BEA) to measure the extent that industries buy

---

[9] Our main estimations employ random draws of plants from each industry. Distances between these plants are measured through county-to-county spatial distances, and the densities are weighted by plant employments, $e(r)$ and $e(s)$. The online Appendix reports similar results when using plant counts, $e = 1$, or when using aggregated firm counts and employment by county-industry. We discuss how the confidence intervals are adjusted under the approximations.

[10] Panel estimation techniques are limited in this setting due to the high persistence in pairwise coagglomeration (see online Appendix Table 2B). We also believe that industry pair connections do not change greatly over time, and data limitations prevent calculating several of our explanatory measures at higher frequency.

and sell from one another. The input-output tables provide commodity level flows which we aggregate to SIC3 industries. We define $Input_{i \leftarrow j}$ as the share of industry i's inputs that come from industry j. We also define $Output_{i \rightarrow j}$ as the share of industry i's outputs that are sold to industry j. These shares are calculated relative to all suppliers and customers, some of which may be non-manufacturing industries or final consumers, and range from zero to one.

The highest observed value of $Input_{i \leftarrow j}$ is 0.39, which represents the share of inputs that come to Leather Tanning and Finishing (SIC 311) from Meat Products (SIC 201). The highest relative value of $Output_{i \rightarrow j}$ is 0.82, which represents the importance of output sales from Public Building and Related Furniture (SIC 253) to Motor Vehicles and Equipment (SIC 371). For most industry pairs, of course, $Input_{i \leftarrow j}$ and $Output_{i \rightarrow j}$ are approximately zero— in fact, 70 percent are less than 0.0001. To construct a single proxy for the connection in goods between a pair of industries, we define undirectional versions of the input and output variables by $Input_{ij} = \max\{Input_{i \leftarrow j}, Input_{j \leftarrow i}\}$ and $Output_{ij} = \max\{Output_{i \rightarrow j}, Output_{j \rightarrow i}\}$. We also define a combined $InputOutput_{ij} = \max\{Input_{ij}, Output_{ij}\}$.

B. *Labor Market Pooling: People*

A second reason to agglomerate is to take advantage of scale economies associated with a large labor pool. Multiple theories have been proposed about the underlying benefits of these labor pools. Marshall emphasizes the risk sharing properties of a large labor market. As individual firms become more or less productive, workers can shift across employers, thereby maximizing productivity and reducing the variance of worker wages (e.g., Charles A. Diamond and Curtis J. Simon 1990; Krugman 1991a). A variant on this theory is that agglomerations facilitate better worker-firm matches (e.g., Robert W. Helsley and Strange 1990). Julio J. Rotemberg and Garth Saloner (2000) further model how workers are more likely to invest in human capital in clusters, knowing that they do not face ex post appropriation. Entrepreneurs may also locate in existing agglomerations due to the suitable labor force (e.g., Pierre-Philippe Combes and Duranton 2006; Michael S. Dahl and Steven Klepper 2007).

All of these models suggest that agglomeration occurs because workers are able to move across firms and industries. Labor movements across firms and industries, however, can occur only if the industries use the same type of workers. We will assess the importance of labor market pooling by looking at the extent to which industries that use the same type of workers coagglomerate with one another. We measure the extent to which industries use similar types of labor through the occupational employment patterns across industries catalogued in the 1987 National Industrial-Occupation Employment Matrix (NIOEM) published by the Bureau of Labor Statistics (BLS). The NIOEM matrix provides industry level employment in 277 occupations, and we define $Share_{io}$ as the fraction of industry i's employment in occupation o.

We measure the similarity of employments in industries i and j through the correlation of $Share_{io}$ and $Share_{jo}$ across occupations. Table 1 contains summary statistics for this $LaborCorrelation_{ij}$ variable. The mean value is 0.470. The measured correlations of one arise because the industry-occupation matrix reports data for NIOEM industries, which is a coarser division than SIC3 industries. Motor Vehicles (371) and Motorcycles, Bicycles and Parts (375) have the most similar employment patterns among industries with different NIOEM data at 0.984. Logging (241) and Aircrafts and Parts (372) have the least correlated labor needs at -0.046.

C. *Intellectual or Technology Spillovers: Ideas*

A final reason that firms collocate is to speed the flow of ideas. Marshall emphasizes that workers learn skills quickly from each other in an industrial cluster. Saxenian (1996) and others

focus on information exchanges among business leaders in industrial concentrations like Silicon Valley. Glaeser and Matthew E. Kahn (2001) argue that the urbanization of high human-capital industries, like finance, is evidence for the role that density plays in speeding the flow of ideas. Arzaghi and Henderson (2008) emphasize networking benefits among marketing firms in Manhattan. Unfortunately, our ability to capture the full range of these models is quite limited. We base our metrics of information flows on patents and research and development (R&D), which reflect only the highest level of information flows, rather than worker level spillovers.

Our first source of data on knowledge spillovers is Frederic M. Scherer's (1984) technology matrix that captures how R&D activity in one industry flows out to benefit another industry. This technology transfer occurs either through supplier-customer relationships between these two industries or through the likelihood that patented inventions obtained in one industry will find applications in the other industry. We develop two metrics, $TechIn_{i \leftarrow j}$ and $TechOut_{i \rightarrow j}$, for these technology flows that mirror $Input_{i \leftarrow j}$ and $Output_{i \rightarrow j}$ described above. The strongest relative technology flows are associated with Plastic Materials and Synthetics (282) and its relationships to Misc. Plastics Products (308), Tires and Inner Tubes (301), and Industrial Organic Chemicals (286).

Our second data source on information exchange is the NBER Patent Database. We measure the extent to which technologies associated with industry i cite technologies associated with industry j, and vice versa. The measures $PatentIn_{i \leftarrow j}$ and $PatentOut_{i \rightarrow j}$ are normalized by total citations for the industries.[11] For our regression analysis, we construct undirectional measures of the intellectual spillovers across an industry pair, $Tech_{ij}$ and $Patent_{ij}$, in a manner analogous to our construction of $InputOutput_{ij}$.

Intellectual spillovers are harder to identify than trade in goods and labor pooling. Many authors employ patent citations to assess intellectual spillovers, but they are only an imperfect measure of intellectual spillovers.[12] As Porter (1990) emphasizes, much knowledge sharing occurs between customers and suppliers, which may be captured more by input-output relationships than by these citations. Idea sharing through the exchange of workers may likewise be better captured by our occupation correlations. Our patent citation measure is a proxy for the importance of exchanging technology rather than a proxy for all forms of intellectual spillovers. Since our measures of idea sharing are weaker than our measures of input-output linkages, we anticipate their connection with coagglomeration to be weaker.

D. *Natural Advantages*

Some regions simply possess better natural environments for certain industries, and agglomeration can follow from these natural cost advantages. Desert areas are inadequate hosts to the logging industry. Areas with exogenously cheap electricity, due perhaps to hydroelectric power, attract aluminum producers. Coagglomeration may be observed if two industries are attracted to the same natural advantages, even if the industries would not otherwise have interacted through Marshallian forces. For example, the ship building and oil refining industries might be coagglomerated simply because both prefer coastal locations.

---

[11] The NBER Patent Data File was originally compiled by Bronwyn H. Hall, Adam B. Jaffe, and Manuel Trajtenberg (2001). It contains records for all patents granted by the United States Patent and Trademark Office (USPTO) from January 1975 to December 1999. The USPTO issues patents by technology categories rather than by industries. Combining the work of Daniel K.N. Johnson (1999), Brian S. Silverman (1999), and Kerr (2008), concordances are developed between the USPTO classification scheme and SIC3 industries (a probabilistic mapping).

[12] See Zvi Griliches (1990), Jaffe, Trajtenberg, and Rebecca M. Henderson (1993), Jaffe, Trajtenberg, and Michael S. Fogarty (2000), and Peter Thompson and Melanie Fox-Kean (2005).

To control for natural advantages based coagglomeration, we develop a predicted spatial distribution for each manufacturing industry based upon local cost advantages and industry traits. The core idea is to interact industry characteristics with costs that are relevant to those traits. This methodology follows Ellison and Glaeser (1999), who model 16 state level characteristics that afford natural advantages in terms of natural resources, transportation costs, and labor inputs. Combining these cost differences with each industry's intensity of factor use, Ellison and Glaeser (1999) estimate a spatial distribution of manufacturing activity by industry that would be expected due to cost differences alone. They find that 20 percent of the observed agglomeration of US manufacturing industries can be explained through these mostly exogenous local factors.[13]

We employ these expected spatial distributions of industries across states to calculate expected coagglomeration levels $Coagg_{ij}^{NA}$ for industry pairs. Separate expected coagglomerations due to natural advantages are constructed for the EG and DO metrics. These measures simply substitute the predicted spatial employments by industry into the EG and DO formulas outlined in Section I. Essentially, this approach measures how coagglomerated the two industries would be if their locations were determined entirely by the interactions of industry characteristics and local characteristics. The DO metric again requires some slight modifications, which we document in the online Appendix. The pairwise correlation between expected and actual coagglomeration using this technique is 0.2 and 0.4 for the EG and DO techniques, respectively.

While $Coagg_{ij}^{NA}$ offers an important control for our estimations, the metric is cruder than those possible in a more focused study of natural advantages (e.g., Holmes and Sanghoon Lee 2008) in a couple of ways. First, our 16 natural advantages will omit many traits that may be important in a subset of industries. Second, we constrain the effects of each natural advantage to be proportional to some factor, e.g., the industry's use of electricity, rather than estimating each effect freely. As with our Marshallian regressors, the resulting measurement error may downward bias our estimate of the importance of these natural conditions. It is also possible that some of the omitted natural advantages may be correlated, positively or negatively, with our Marshallian proxies. While mostly fixed, some of our natural advantages may be themselves endogenous, and that endogeneity could lead us to either over- or understate the importance of natural advantage. For example, if energy prices rise in areas where energy intensive firms locate for other reasons, then this will bias the coefficient on energy prices, complicating the interpretation of our results.

Recognizing these limitations, we believe that our measure of expected coagglomeration is both an important control variable and a natural baseline for comparing Marshallian agglomeration economies. As the imperfections in our natural advantages metric have the potential for biasing our Marshallian parameter estimates, we will test the sensitivity of our Marshallian findings to including or excluding expected coagglomeration measures.

III. Empirical Results: OLS Estimates

We now present our main empirical results of the forces contributing to manufacturing coagglomeration. Our core empirical specification is a simple OLS regression:

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

where $Coagg_{ij}$ is a measure of the pairwise coagglomeration between industries i and j. We separately test four variants of both the EG and DO metrics. We modify $Coagg_{ij}^{NA}$ to mirror the

---

[13] Ellison and Glaeser (1999) suggest that this 20 percent share likely underestimates the true portion of spatial agglomeration that can be explained through mostly fixed characteristics. Sukkoo Kim (1999) estimates natural regional advantages over a 100-year period.

TABLE 3—OLS Univariate Specifications for Pairwise Coagglomeration

Each entry reports separate estimation with single regressor	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

Notes: Each cell reports a separate regression of pairwise coagglomeration on a determinant of industrial co-location. Coagglomeration measures are calculated from the 1987 and 1997 Census of Manufacturers as listed in the column headers. All pairwise combinations of manufacturing SIC3 industries are included, except those listed in the text, for 7,381 observations. Natural advantages coagglomeration is estimated through predicted state-industry shares developed from exogenous local cost variables (e.g., coastal access, energy prices) and industry cost dependencies. Labor correlation indices are calculated from the BLS National Industry-Occupation Employment Matrix for 1987. Input-output relationships are calculated from the BEA Benchmark Input-Output Matrix for 1987. Technology flows are calculated from the Scherer (1984) R&D tables for the 1970s and from the NBER Patent Citation Database for 1975–1997. Maximum values for the pairwise combination are employed. Variables are transformed to have unit standard deviation for interpretation. Regressions are unweighted. Bootstrapped standard errors are reported in parentheses.

design of the dependent variable (EG or DO), while the Marshallian metrics remain the same. As $Coagg_{ij}^{NA}$ is a generated regressor (e.g., Adrian R. Pagan 1984), we report bootstrapped standard errors.

We normalize all variables to have a standard deviation of one. This normalization makes it easier to compare the coefficient estimates for the different variables and to assess the importance of each factor in explaining overall coagglomeration patterns. We will compare the individual contributions of the Marshallian factors, both amongst themselves and relative to natural advantages. We are also interested in comparing the total contribution of Marshallian agglomeration economies to the contribution of natural advantages. We evaluate this through a one standard deviation increase in all three Marshallian factors.

A. *Univariate Regressions*

Table 3 presents univariate regressions for each of our variables. Entries in the table are from 40 separate specifications, with columns reporting the coagglomeration index and rows reporting

explanatory variables. Column 1 finds fairly uniform coefficient magnitudes for the EG metric of state total employments. A one standard deviation increase in expected coagglomeration due to shared natural advantages is associated with a 0.21 standard deviation increase in actual coagglomeration. Input-output relationships also exhibit a 0.21 correlation. The other Marshallian factors are slightly weaker. The estimated coefficients are 0.18 for labor pooling and 0.08 to 0.18 for technology sharing. Columns 2 through 4 find comparable orderings when employing other variants of the EG metric, with some overall decline in the strength of all correlations also evident. On their own, each of the three variables can explain about the same share of the variation in coagglomeration across industry pairs.

Columns 5 through 8 consider four variants of the continuous DO metric where we adjust the threshold for identifying localization. Shared natural advantages are found to have greater explanatory power when using the continuous DO index than with the discrete EG metric, regardless of the distance threshold specified.[14] The Marshallian factors generally have similar coefficients in the EG and DO regressions. The EG state level results are particularly similar to the 250 mile DO results. These two measures are designed to reflect coagglomeration at similar scales, so this result provides added confidence that the effects we identify are robust to how coagglomeration is being measured.

The DO results do change substantially when we move to a 1,000-mile threshold: the patent citation measure appears to be uncorrelated with coagglomeration, and the labor pooling measure is negatively correlated with coagglomeration. We find these results to be reassuring. The 1,000-mile threshold is far beyond the distance at which one would expect labor to be highly mobile and ideas to be "in the air." Hence, we would not expect these regressions to identify strong effects of labor pooling and technological spillovers on coagglomeration.[15]

B. *Multivariate Regressions*

Table 4 presents our full multivariate specification. Each column reports coefficients from a single regression with a pairwise coagglomeration metric as the dependent variable. We concentrate on the EG metric that uses state employments and the DO metric with a 250-mile threshold, reporting four specifications for each.

The first column presents our base EG specification. The estimates show that each of our variables continues to be significant in multivariate frameworks. Natural advantages remains the strongest explanatory variable with a coefficient estimate of 0.16. The point estimates are largest for input-outputs (0.15), followed by labor pooling (0.12) and technology spillovers (0.10). But the differences between the coefficient estimates are not significant, which suggests that all three Marshallian forces are important and that their effects appear to be comparable in magnitude. Together these three variables explain more of the variation in coagglomeration than does natural advantage, which supports the view that agglomeration economies is a more important determinant of geographic location (as in Ellison and Glaeser 1999).

The second column excludes the natural advantages measure from the regression. While we believe that it makes sense generally to control for this measure, the potential endogeneity of the elements that drive the natural agglomeration measure make it reasonable to wonder whether our

---

[14] The higher correlations when we use the DO natural advantages measure extend from two sources. First, the more continuous horizon helps identify clustering along natural advantages across state borders (e.g., neighboring coastal states in New England). The online Appendix discusses a second, mechanical reason due to limitations in our procedure for constructing the DO natural advantages.

[15] Rosenthal and Strange (2003) and Arzaghi and Henderson (2008) emphasize even further the small spatial distances over which knowledge spillovers occur. The online Appendix further discusses the negative labor correlation with the 1,000-mile DO metric.

TABLE 4—OLS MULTIVARIATE SPECIFICATIONS FOR PAIRWISE COAGGLOMERATION

	EG coaggl. index with state total emp.				DO coaggl. index, 250 mi.			
	Base estimation	Exclude natural advantages	Separate input & output	Exclude pairs in same SIC2	Base estimation	Exclude natural advantages	Separate input & output	Exclude pairs in same SIC2
	(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

Notes: See Table 3. Regressions of pairwise coagglomeration on determinants of industrial co-location. Columns 4 and 8 exclude SIC3 pairwise combinations within the same SIC2. Online Appendix Table 6 provides additional robustness checks. Variables are transformed to have unit standard deviation for interpretation. Bootstrapped standard errors are reported in parentheses.

results change much when that measure is excluded. We find that Marshallian forces become slightly stronger when natural advantages are excluded. However, the coefficients in the two columns are sufficiently similar that it seems that the natural advantages and Marshallian factors are mostly orthogonal to one another. The third column disaggregates the input-output effect into separate input and output effects. The two effects are comparable in magnitude and both are quite significant.

The fourth column excludes all industry pairs in the same two-digit SIC industry (SIC2). There are both conceptual and methodological reasons for this exclusion. Conceptually, industries within the same SIC2 may be more likely to coagglomerate due to unobserved factors or due to geographic features that we have measured with error. Methodologically, some of our measures, like the technology flow measure, have variation that straddles the SIC2 and SIC3 divisions. The coefficient estimates in this regression are slightly lower, but similar in magnitude to the base regression in the first column. We will use this restricted sample in our instrumental variables analysis below.

Columns 5 through 8 present equivalent results for the DO index calculated with a distance threshold of 250 miles. The results are similar to those obtained with the state level EG index. All three Marshallian factors are important. Natural advantages are more important than any single Marshallian factor, but the three factors together are more important than natural advantage. The differences shown in Table 3 persist: natural advantages appear more important when we use the DO metrics for coagglomeration; and labor market pooling appears somewhat less important. Again, the broad similarity provides confidence that the coagglomeration metric design is not driving the basic conclusions of this paper.

Three general conclusions emerge from these regressions. First, all three of Marshall's (1920) theories regarding agglomeration find support in coagglomeration patterns. Second, the Marshallian factors appear to be relatively important in the sense that taken together they are more important than the natural advantages we have identified. Third, the input-output factor

comes through most consistently. Labor pooling follows closely on smaller spatial distances, but it has much less of an effect when we look at coagglomeration at a broader geographic scale.

The online Appendix documents the full set of outcomes for each variant of the coagglomeration metric. We also present robustness checks: using pairwise means rather than maximums for explanatory variables, including industry effects, weighting by the relative size of the industry pair, and substituting the patent-based technology measure for the Scherer metric. We further consider several variants on the EG and DO metrics. While minor differences emerge, the overall patterns presented in Tables 3 and 4 are quite stable.

IV. Instrumental Variables Analysis

A potential concern with the analysis presented above is that our measures of the potential for Marshallian spillovers between industries might endogenously reflect coagglomeration patterns. For example, the volume of trade between the shoemaking and leather industries may not only reflect inherent features of shoemaking technology. It could be that there would be less leather and more plastic in shoes if random events had led to the coagglomeration of the shoemaking and plastics industries. Similarly, the employment mix of an industry could be affected by where plants are located. Firms in some industries may be able to choose between a low tech production process that requires many unskilled laborers and a more automated process with a very different occupational mix. These choices could then be influenced by local labor market conditions.

To help with these concerns, our OLS regressions include controls for expected coagglomeration due to shared interests in natural advantages. Variance in coagglomeration due to unmodeled natural advantages, of course, will still bias our parameter estimates. Moreover, it will not help with the reverse causation problem noted above. In this section we present two sets of IV estimates designed to address these concerns.

A. *UK Instruments*

Our first set of instruments are constructed from data on characteristics of UK industries. If two industries are coagglomerated in the United States for purely random reasons or because they value different, unobserved natural advantages that are randomly correlated in the United States (e.g., if states with bauxite deposits are also close to sources of sugar cane), then one would expect that the industry pair would not be coagglomerated for these reasons in the United Kingdom. In this case, characteristics of UK industries provide measures of the Marshallian factors for the industry pair that are orthogonal to the endogenous variation in the United States.

Of course, this technique will work in only some situations. If two industries are coagglomerated in the United States because they have a greater need for a coastal location, then they will likely be coagglomerated in the United Kingdom as well. In this case, the UK characteristics of the industry pair could be affected by a correlated endogeneity. Our natural advantages metric should theoretically capture such situations, but unmodeled natural advantages that are not randomly distributed may again be present.

Our UK based instrument for input-output relationships builds from the 1989 Input-Output Balance for the United Kingdom published by the Central Statistical Office in 1992. The original table contained 102 sectors, and Keith E. Maskus, C. Sveikauskas, and Allan Webster (1994) aggregated those original categories into 80 sectors that can be matched with US industries. The construction of the UK instruments is otherwise comparable to that undertaken with BEA data. We will use these UK input-output measures as instruments for the US input-output relationships

under the identifying assumption that UK material flows are correlated with true Marshallian dependencies among US industries but uncorrelated with the reverse causation that may have arisen within the United States after industrial locations are determined.

Our UK based instrument for labor market similarities was constructed using data from the UK Labour Force Survey, which is roughly akin to the US Current Population Survey. The United Kingdom does not publish a detailed equivalent of the BLS NIOEM matrix, so we constructed our own matrix by pooling six years (2001–2006) of the UK Labour Force Survey. We mapped the British industry codes to the SIC3 system, but we did not map the British occupations to NIOEM equivalents. We instead calculated pairwise industry correlations on the British occupation vectors.

There is also a concern of endogeneity of intellectual exchanges, as industries may share technologies because of locational proximity. The online Appendix describes instruments developed through patent citations using instances where both the citing and cited USPTO patent were filed from the United Kingdom, but in practice we found it very difficult to instrument simultaneously for all three of Marshall's forces. We therefore focus on our IV specifications on the customer-supplier and labor pooling rationales, which are also more distinguishable intellectually and empirically.

B. *US Spatial Instruments*

Our second set of instruments are constructed using disaggregated data that allow us to examine industry characteristics in different parts of the United States. This approach measures input-output and labor patterns in one industry in places where the other industry is quite rare. By focusing on areas where the other industry is absent, we can ideally estimate input-output and hiring patterns that are correlated with innate industrial needs but not biased by geographic proximity to the other industry. Most industry pair coagglomerations are sufficiently weak so that one can find parts of the United States where industry i is present and industry j is not overrepresented. We measure the relatedness of each industry pair using data on the characteristics of industry i in areas where industry j is least present and data on the characteristics of industry j in areas where industry i is least present.

Measures of this form will be useful instruments if the endogenous variation in our Marshallian factors is due to a plant's input/labor choices being affected by the proximity of plants in the other industry. For example, they will be helpful if shoemaking plants choose to make leather shoes when located near leather manufacturers and choose to make plastic shoes otherwise. In such a situation, OLS estimates would overstate the importance of input/output relationships as an agglomeration force. By looking at the inputs used by shoemakers who are located far from leather makers, we may derive industry characteristics that are useful instruments.

There are other situations, of course, in which the spatial instruments will not help. One example is where there are economies of scale in the development of production technologies and technologies develop in light of the average distance between plants in industries i and j. In this scenario, firms in industry i still need to buy inputs from industry j even if no plants in industry j are nearby. Measuring characteristics from plants that are not collocated will not correct this situation.

Our spatial input-output instruments are developed using "material inputs trailers" of the 1987 Census of Manufacturing. This form asks plants to list their material inputs and associated expenditures. Our spatial instrument employs the microrecords to calculate industry i's input dependence on industry j in regions where industry j is least present. We specifically choose the 25 PMSAs where industry j is least present relative to all manufacturing to calculate industry i's dependency for j. The dependencies are relative to all plant inputs, including nonmanufacturing.

TABLE 5—IV MULTIVARIATE SPECIFICATIONS FOR PAIRWISE COAGGLOMERATION

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

Notes: See Table 3. OLS and IV regressions of pairwise coagglomeration on determinants of industrial co-location. All estimations exclude SIC3 pairwise combinations within the same SIC2. Online Appendix Tables 7 and 8 report first stages and additional robustness checks. Variables are transformed to have unit standard deviation for interpretation. Bootstrapped standard errors are reported in parentheses.

The Appendix describes the materials trailers data in greater detail and the variants of this instrument that we tested.[16]

Our spatial instruments for labor similarity are developed using the 1990 Census IPUMS. We again ordered PMSAs by the relative presence of each industry compared to all manufacturing activity. We chose the 25 PMSAs where industry i was least present to measure industry j's occupation needs, and vice versa. We then constructed the labor similarity correlation between industries i and j as described above. The online Appendix again describes these data in greater detail and the variants of this instrument that we tested.

We conduct our IV analysis on the restricted sample of 7,000 pairwise industry combinations that exclude SIC3 pairs within the same SIC2 sector. This restriction is for two reasons. First, some of the data for the instruments have limited variation across SIC3 pairs within an SIC2 sector. Second, our discussion of the instruments' conceptual liabilities has often centered on unobserved natural advantages missed by our expected coagglomeration metric. These confounding issues are most likely to exist among SIC3 industries within the same SIC2 category. As we saw in Table 4, the OLS relationships are stable including or excluding these closely-related industry pairs.

The Appendix documents the first-stage regression estimates for both sets of instruments. The t-statistics are over ten for the relevant instruments, and we satisfy relevant tests regarding weak instruments. The strength of these first-stage relationships does not change substantially when simultaneously instrumenting for both labor and input-output factors. Likewise, the inclusion or exclusion of our metric of expected coagglomeration due to natural advantages does not influence substantially the first-stage relationships for the Marshallian factors.

C. *IV Regression Results*

Table 5 presents our core instrumental variables results using UK and US spatial instruments. We instrument for the input-output and labor pooling factors using the instruments described

---

[16] For example, we have confirmed that using absolute thresholds of the bottom 25 cities delivers similar results to techniques using relative shares (e.g., the group of cities accounting for a small share of activity in an industry). We have also implemented a regional approach that includes rural areas.

above. The control of expected coagglomeration due to shared natural advantages is included and treated as exogenous. We do not include a technological spillover variable.[17] Columns 1 and 4 report OLS estimates of these specifications.

Columns 2 and 3 report IV regressions using the EG state level coagglomeration as the dependent variable and employing the UK instruments and US spatial instruments, respectively. Both instruments, despite their quite different construction and data sources, yield similar results. The role of labor is confirmed, and the instrumented elasticity is very similar to the OLS results. On the other hand, the input-output elasticity strengthens. Hausman tests do not reject the hypothesis that the OLS estimates are exogenous at the ten percent level.[18]

Our instrumental variables estimates employing the 250-mile DO metric also support the importance of both Marshallian factors. The input-output variable has a larger coefficient in the DO OLS regression than in the EG OLS regression, and the estimate remains significant and retains its magnitude in both IV estimates. The labor pooling variable had a much smaller effect in the OLS regression, but the estimate is much larger in the IV regressions. Hausman tests of equality for the OLS and IV specifications are rejected for the DO specifications with both instrument pairs. The online Appendix extends these IV estimations to other variants of EG and DO metrics.

V. Conclusions

At the broadest level, our paper provides strong support for Marshallian theories of agglomeration. We find consistent evidence for each of the three mechanisms—proximity to reduce the costs of moving goods, people, and ideas—in the US manufacturing sector. Taken together, the Marshallian factors appear to have a stronger effect on coagglomeration patterns than shared natural advantages, which Ellison and Glaeser (1999) found to drive a nontrivial fraction of within-industry agglomeration in the United States. We recognize, however, that we have modeled only a finite number of measured natural advantages and that our proxies are imperfect. This measurement error may lead us to understate the relative contribution of natural advantages versus Marshallian forces.

Which of Marshall's theories regarding industrial agglomeration are more important? Our basic conclusion from examining coagglomeration patterns is that all three forces are similar in magnitude, with input-output flows being the greater among equals. A one standard deviation growth in labor or input-output dependencies increases coagglomeration by around one seventh of a standard deviation. The importance of technology flows is weaker in some specifications, but of comparable magnitude in other estimations.

We do not know how our manufacturing results would generalize to other industries. Many services are more costly to transport since they involve face to face interaction, and therefore we might think that input-output relationships are particularly important in that sector (e.g., Jed Kolko 1997). The current excitement over service offshoring suggests, however, that segments within services like call centers may have rather low transport costs. Ideas and knowledge spillovers may be more important in very innovative sectors. We hope that future research defines Marshallian interactions in ways appropriate for industries outside of manufacturing.

It would likewise be interesting to understand better how these forces have changed over time. Transportation costs for physical goods have declined remarkably over the twentieth century (Glaeser and Janet E. Kohlhase 2004). These shipment costs have likely declined relative to the

---

[17] In general, our instruments for technology sharing do not adequately distinguish themselves from the input-output and labor pooling relationships. Results for the triple IV are reported in the online Appendix.

[18] The probability of rejecting the Chi-Squared test is 0.13 and 0.16 using the UK and US spatial IVs.

costs of moving labor, but it is not clear whether relative declines for physical goods are more or less than declines in transportation costs for ideas. Some types of information appear to flow very well over long distances, while others still require very close proximity. The impact of these relative changes in transportation costs are worthy of additional study. It would also be interesting to model the interaction between sunk investments in firm locations, made decades ago due to coagglomeration factors that are no longer as relevant, with changing relative transportation costs.

Although this paper is primarily about agglomeration and not about methodology, we hope that the approach it takes will be useful in future explorations of agglomerative forces. The coagglomeration patterns we explore could be examined in many different ways. And the UK and US spatial instruments we develop could be applied in many other areas in which the endogeneity of industry characteristics is a concern as well as in future studies of agglomeration and coagglomeration.

REFERENCES

Arzaghi, Mohammad, and J. Vernon Henderson. 2008. "Networking Off Madison Avenue." *Review of Economic Studies*, 75(4): 1011–38.

Audretsch, David B., and Maryann P. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review*, 86(3): 630–40.

Autor, David H., William R. Kerr, and Adriana D. Kugler. 2007. "Does Employment Protection Reduce Productivity? Evidence from US States." *Economic Journal*, 117(521): F189–217.

Combes, Pierre-Philippe, and Gilles Duranton. 2006. "Labour Pooling, Labour Poaching, and Spatial Clustering." *Regional Science and Urban Economics*, 36(1): 1–28.

Dahl, Michael S., and Steven Klepper. 2007. "Who Do New Firms Hire?" Unpublished.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction*. Cambridge, MA: MIT Press.

Delgado, Mercedes, Michael E. Porter, and Scott Stern. 2008. "Convergence, Clusters and Economic Performance." Unpublished.

Diamond, Charles A., and Curtis J. Simon. 1990. "Industrial Specialization and the Returns to Labor." *Journal of Labor Economics*, 8(2): 175–201.

Dumais, Guy, Glenn Ellison, and Edward L. Glaeser. 2002. "Geographic Concentration as a Dynamic Process." *Review of Economics and Statistics*, 84(2): 193–204.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson. 1989a. "Plant Turnover and Gross Employment Flows in the U.S. Manufacturing Sector." *Journal of Labor Economics*, 7(1): 48–71.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson. 1989b. "The Growth and Failure of U.S. Manufacturing Plants." *Quarterly Journal of Economics*, 104(4): 671–98.

Duranton, Gilles, and Henry G. Overman. 2005. "Testing for Localization Using Micro-Geographic Data." *Review of Economic Studies*, 72(4): 1077–106.

Duranton, Gilles, and Henry G. Overman. 2008. "Exploring the Detailed Location Patterns of U.K. Manufacturing Industries Using Microgeographic Data." *Journal of Regional Science*, 48(1): 213–43.

Ellison, Glenn, and Edward L. Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy*, 105(5): 889–927.

Ellison, Glenn, and Edward L. Glaeser. 1999. "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *American Economic Review*, 89(2): 311–16.

Florence, P. Sargant. 1948. *Investment, Location and Size of Plant*. Cambridge, UK: Cambridge University Press.

Fuchs, Victor R. 1962. *Changes in the Location of Manufacturing in the United States Since 1929*. New Haven, CT: Yale University Press.

Fujita, Masahisa, Paul Krugman, and Anthony J. Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press.

Glaeser, Edward L., and Matthew E. Kahn. 2001. "Decentralized Employment and the Transformation of the American City." National Bureau of Economic Research Working Paper 8117.

Glaeser, Edward L., and William R. Kerr. 2009. "Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain?" *Journal of Economics and Management Strategy*, 18(3): 623–63.

Glaeser, Edward L., and Janet E. Kohlhase. 2004. "Cities, Regions and the Decline of Transport Costs." *Papers in Regional Science*, 83(1): 197–228.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2008. "Identifying Agglomeration Spillovers: Evidence from Million Dollar Plants." National Bureau of Economic Research Working Paper 13833.

Griliches, Zvi. 1990. "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature*, 28(4): 1661–707.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. 2001. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." National Bureau of Economic Research Working Paper 8498.

Helsley, Robert W., and William C. Strange. 1990. "Matching and Agglomeration Economies in a System of Cities." *Regional Science and Urban Economics*, 20(2): 189–212.

Henderson, J. Vernon. 2003. "Marshall's Scale Economies." *Journal of Urban Economics*, 53(1): 1–28.

Holmes, Thomas J. 1998. "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders." *Journal of Political Economy*, 106(4): 667–705.

Holmes, Thomas J., and Sanghoon Lee. 2008. "Economies of Density Versus Natural Advantage: Crop Choice on the Back Forty." Unpublished.

Hoover, Edgar Malone. 1948. *The Location of Economic Activity*. New York: McGraw Hill.

Jaffe, Adam B., Manuel Trajtenberg, and Michael S. Fogarty. 2000. "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors." *American Economic Review*, 90(2): 215–18.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics*, 108(3): 577–98.

Johnson, Daniel K. N. 1999. "150 Years of American Invention: Methodology and a First Geographic Application." Wellesley College Department of Economics Working Paper 99-01.

Kerr, William R. 2008. "Ethnic Scientific Communities and International Technology Diffusion." *Review of Economics and Statistics*, 90(3): 518–37.

Kerr, William R., and Ramana Nanda. 2009. "Democratizing Entry: Banking Deregulations, Financing Constraints, and Entrepreneurship." *Journal of Financial Economics*, 94(1): 124–49.

Kim, Sukkoo. 1999. "Regions, Resources, and Economic Geography: Sources of U.S. Regional Comparative Advantage, 1880–1987." *Regional Science and Urban Economics*, 29(1): 1–32.

Kolko, Jed. 1999. "Can I Get Some Service Here? Information Technology, Service Industries, and the Future of Cities." Unpublished.

Krugman, Paul. 1991a. *Geography and Trade*. Cambridge, MA: MIT Press.

Krugman, Paul. 1991b. "Increasing Returns and Economic Geography." *Journal of Political Economy*, 99(3): 483–99.

Marshall, Alfred. 1920. *Principles of Economics*. London: MacMillan.

Maskus, Keith E., Catherine D. Sveikauskas, and Allan Webster. 1994. "The Composition of the Human Capital Stock and Its Relation to International Trade: Evidence from the U.S. And Britain." *Weltwirtschaftliches Archiv*, 130(1): 50–76.

Menon, Carlo. 2008. "The Bright Side of Gerrymandering: An Enquiry on the Determinants of Industrial Agglomeration in the United States." Unpublished.

Pagan, Adrian. 1984. "Econometric Issues in the Analysis of Regressions with Generated Regressors." *International Economic Review*, 25(1): 221–47.

Porter, Michael E. 1990. *The Competitive Advantage of Nations*. New York: The Free Press.

Rosenthal, Stuart S., and William C. Strange. 2001. "The Determinants of Agglomeration." *Journal of Urban Economics*, 50(2): 191–229.

Rosenthal, Stuart S., and William C. Strange. 2003. "Geography, Industrial Organization, and Agglomeration." *Review of Economics and Statistics*, 85(2): 377–93.

Rotemberg, Julio J., and Garth Saloner. 2000. "Competition and Human Capital Accumulation: A Theory of Interregional Specialization and Trade." *Regional Science and Urban Economics*, 30(4): 373–404.

Saxenian, AnnaLee. 1996. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Scherer, Frederic M. 1984. "Using Linked Patent and R&D Data to Measure Interindustry Technology Flows." In *R&D, Patents, and Productivity*, ed. Zvi Griliches, 417–61. Chicago: University of Chicago Press.

Silverman, Brian S. 1999. "Technological Resources and the Direction of Corporate Diversification: Toward an Integration of the Resource-Based View and Transaction Cost Economics." *Management Science*, 45(8): 1109–24.

Thompson, Peter, and Melanie Fox-Kean. 2005. "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment." *American Economic Review*, 95(1): 450–60.

von Thünen, Johann Heinrich. 1826. *Der Isolirte Staat in Beziehung auf Landwirthschaft und Nationalökonomie, oder Untersuchungen über den Einfluss, den die Getreidepreise, der Reichtum des Bodens und die Abgaben auf den Ackerbau Ausüben*. Reprinted in English as *Von Thünen's Isolated State*, Pergamon Press, 1966.