What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns[1]

Glenn D. Ellison
MIT and NBER

Edward L. Glaeser
Harvard University and NBER

William R. Kerr
Harvard Business School

August 2008

[1]email: gellison@mit.edu, glaeser@fas.harvard.edu, wkerr@hbs.edu. We are grateful to Mohammad Arzaghi, Alex Bryson, Jim Davis, Keith Maskus, Pete Schott, and Debbie Smeaton for data assistance. We thank Gilles Duranton for assistance in calculating our approximation of the Duranton and Overman (2005) metric. The research in this paper was conducted while the authors were Special Sworn Status researchers of the U.S. Census Bureau at the Boston Census Research Data Center (BRDC). Support for this research from NSF grants ITR-0427889 and SES-0550897 is gratefully acknowledged. Research results and conclusions expressed are our own and do not necessarily reflect the views of the Census Bureau or NSF. This paper has been screened to insure that no confidential data are revealed.

Abstract

Why do people and firms cluster near one another? This paper tests Marshall's theories of industrial agglomeration by examining which industries locate near one another, or coagglomerate. We use Census Bureau's Longitudinal Research Database to measure pairwise coagglomeration indices for U.S. manufacturing industries. We then relate the degree of coagglomeration to attributes of each industry pair, such as the tendency of the two industries to buy from one another or to hire similar types of workers. To reduce reverse causality, where co-location drives input-output linkages or hiring patterns, we use data from U.K. industries and from U.S. areas where the two industries are not co-located. Our evidence supports all three of Alfred Marshall's theories of agglomeration, but input-output linkages seem particularly important for these manufacturing industries.

1 Introduction

We know that industries are geographically concentrated.[1] We know that this concentration is too great to be explained by exogenous spatial differences in natural advantage.[2] There are an abundance of theories for this concentration.[3] But we do not know which of these theories are important or even right. This paper exploits patterns of industry coagglomeration to measure the relative importance of different theories of industry agglomeration.

Marshall (1920) emphasized three different types of transport costs – the costs of moving goods, people, and ideas – that could 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 (1994) in Silicon Valley, locate near one another to learn and to speed their rate of innovation.

Although each of these determinants certainly contributes to agglomeration in some industries, it is challenging to assess their relative importance using data on which industries are agglomerated. 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.[4] The motivation for our empirical approach is that there is also information in coagglomeration patterns.[5] Plants are similar to the other plants in their industry in many dimensions. But when we look 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

[1] See P. Sargant Florence (1948), E. M. Hoover (1948), Victor Fuchs (1990), Paul Krugman (1991a), Glenn Ellison and Edward L. Glaeser (1997), and Gilles Duranton and Henry Overman (2005, 2008).

[2] See Ellison and Glaeser (1999).

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

[4] One 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.

[5] In a similar vein, Henderson (2003) examines how plant-level productivity is related to the set of plants in the area.

hire similar workers but never trade with each other. Hence, one can gain insight into which theories are more important by looking at which similarities are most predictive of whether industry pairs are coagglomerated.

Section II 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 III 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 co-locate to reduce the costs of shipping between customers and suppliers. 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 IV presents our main empirical results. The ordinary least squares relationships support the importance of all three Marshallian theories. Coagglomeration arising through shared natural advantages is estimated to be more important than any single Marshallian factor, but not as important as the cumulative effect of the three Marshallian factors. Input/output relationships have the largest effect of the Marshallian factors we consider, which is striking given the remarkable decline of transportation costs over the 20th century (Glaeser and Janet E. Kohlhase, 2004). Customer-supplier relationships are closely followed by similar labor needs. Our proxies for intellectual spillovers are weaker, but still economically and statistically important.

One concern with these results is that industrial relationships may be the result of co-location instead of the cause of co-location. Some industries may be flexible enough in their production processes that they adjust to nearby resources of labor and material inputs. Section V addresses this concern by developing two sets of instrumental variables. First, we use characteristics of U.K. industries. Coagglomeration patterns due to unobserved, shared natural advantages (or purely random events) may differ in the U.K., in which case U.K. industry characteristics can help identify effects that are due to innate similarities between industry pairs. Second, we use data on characteristics of plants in different parts of the U.S.

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. These more isolated plants will not have a geographic reason for changing their inputs to accommodate the neighbors of more typical plants in their industry. We use plant-level detail from the Census of Manufacturing and individual-level data from the Census IPUMS to develop measures of industry-pair similarity based upon characteristics of the non-coagglomerated 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.

2 U.S. Manufacturing Coagglomeration

We compute pairwise coagglomeration measures for manufacturing industries using the confidential plant-level data from the U.S. Census Bureau's *Census of Manufacturing*.[6] 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 7381 industry-pair observations per year: all distinct coagglomeration pairs from 122 industries.[7]

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, PMSA, and county levels. We also use the longitudinal nature of the Census Bureau data to analyze coagglomeration of start-up firms.[8] 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

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

[7] We exclude Tobacco (210s), Apparel industries (230s), portions of Printing and Publishing (277-279), Secondary Non-Ferrous Metals (334), and Search and Navigation Equipment (381). These exclusions are primarily due to data constraints and are documented in the appendix.

[8] 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 Autor, William Kerr, and Adriana Kugler (2007), Glaeser and Kerr (2008), and Kerr and Ramana Nanda (2008).

in area m, and 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

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

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 1000 miles, with our primary results taken from a 250 mile distance horizon.[9]

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 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

[9] 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 U.K., 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 U.S., falling between 900 and 1000 miles depending upon the weighting. We chose our four thresholds to span the actual physical distance studied by DO and the median plant concept.

Our main results focus on densities that are weighted by plant employments, $e(r)$ and $e(s)$, and the appendix reports similar findings when using plant counts, $e = 1$.

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 reported in Ellison and Glaeser (1997). Panel B presents descriptive statistics for DO coagglomeration indices. Eighty-seven percent of the industry-pairs exhibit some degree of global localization to the 250 mile threshold.

Table 2 lists the fifteen 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 which we computed to make using the index more tractable on 7381 industry pairs (and to respond to data limitations).

3 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. We also control for expected coagglomeration of industry pairs arising from common dependencies on certain natural advantages (e.g., coastal access, energy prices). Our goal is not only to learn about coagglomeration, but to learn more generally about 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. Our initial analysis will consist of OLS regressions of our concentration indices on these measures. Where possible, we focus our estimations and data construction on the 1987 cross section.[10] The appendix provides additional details and descriptive statistics.

3.1 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. The "new economic geography" of Masahisa Fujita, Paul Krugman, and Anthony Venables (1999) views reducing the costs of transporting goods as the driver behind agglomeration. To assess the importance of this factor, we use the 1987 Benchmark Input-Output Accounts of the Bureau of Economic Analysis (BEA) to measure the extent that industries buy and sell from one another. The input-output tables provide commodity-level flows which we aggregate to the SIC3 level.

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 whom 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%

[10] Panel estimation techniques are limited in this setting due to the high persistence in pairwise coagglomeration (see App. 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.

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 $Output_{ij} = \max\{Output_{i \to j}, Output_{j \to i}\}$. We also define a combined $InputOutput_{ij} = \max\{Input_{ij}, Output_{ij}\}$.

3.2 Labor market pooling: People

A second reason to agglomerate is to take advantage of scale economies associated with a large labor pool.[11] Labor movements across firms and industries are only likely if the industries use the same type of workers, which we model through occupations. We start with 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 (0.984). Logging (241) and Aircrafts and Parts (372) have the least correlated labor needs (-0.046).

3.3 Intellectual or technology spillovers: Ideas

A final reason that firms co-locate is to speed the flow of ideas. Marshall emphasizes how workers learn skills quickly from each other in an industrial cluster, while Saxenian (1994)

[11] 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., Diamond and Simon, 1990, Krugman, 1991a). A variant on this theory is that agglomerations facilitate better worker-firm matches through a wider range of alternatives. Rotemberg and 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. Finally, Combes and Duranton (2006) model how start-ups may locate near older firms to hire away their workers. All of these models suggest that agglomeration occurs because workers are able to move across firms and industries, and our empirical approach does not distinguish among them.

and others focus on information exchanges among business leaders in industrial concentrations like Silicon Valley. Glaeser and Matthew 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.

We construct proxies using data from two different sources. The first is Frederic 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 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.[12] 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 obviously only an imperfect measure of intellectual spillovers.[13] As Michael Porter (1990) emphasizes, much knowledge sharing occurs between consumers and suppliers, which may be captured more by input-output relationships than by these citations. Idea sharing through the ex-

[12] The NBER Patent Data File was originally compiled by Bronwyn Hall, Adam 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 Johnson (1999), Brian Silverman (1999), and Kerr (2008), concordances are developed between the USPTO classification scheme and SIC3 industries (a probabilistic mapping).

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

change of workers may likewise be better captured by our occupation correlations. As such, we see our patent citation measure as a proxy for the importance of exchanging technology rather than as 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.

3.4 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, and areas with cheap electricity 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.

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. This work follows Ellison and Glaeser (1999), who model sixteen 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 that would be expected due to cost differences alone (plus population distributions). They find that 20% of the observed agglomeration of U.S. manufacturing industries can be explained through these mostly exogenous local factors.[14]

We employ these state-industry expected spatial distributions 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. The pairwise correlation between expected and actual coagglomeration using this technique is 0.2 and 0.4 for the EG and DO techniques, respectively.

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

4 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 design of the dependent variable (EG or DO), while the Marshallian metrics remain the same. We normalize all variables to have a standard deviation of one. Coefficient estimates on the variables can thereby be more easily compared and thought of as measures of the importance of each factor in explaining overall coagglomeration patterns.

4.1 Univariate regressions

Before discussing multivariate analyses, Table 3 presents univariate regressions for each of our variables. Entries in the table are from forty 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 at 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 index than with EG, regardless of the threshold specified.[15] The Marshallian factors generally have similar coefficients in

[15] The higher correlation vis-a-vis the EG natural advantages extends from two sources. First, the more-

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 1000 mile threshold: the patent citation measure appears to be uncorrelated with coagglomeration, and the labor pooling measure is negatively correlated. We also find these results to be reassuring. The 1000 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.[16]

4.2 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 employment 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 continue 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, so the most important takeaway is that all three Marshallian forces appear to be 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).

continuous horizon does help identify clustering along natural advantages across state borders (e.g., neighboring coastal states in New England). The appendix discusses a second, mechanical reason due to limitations in our procedure for constructing the DO natural advantages.

[16] Rosenthal and Strange (2003) and Arzaghi and Henderson (2008) emphasize even further the small spatial distances over which knowledge spillovers occur. The appendix further discusses the negative labor correlation with the 1000 mile DO metric.

The second column finds 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. 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 we saw 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 appendix documents the full set of outcomes for each variant of the coagglomeration metric. We also present robustness checks: using pairwise means rather 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.

5 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. Our OLS regressions include controls for expected coagglomeration due to natural advantages, but any variance in coagglomeration due to omitted natural advantages and random events would create an endogeneity problem in such situations. In this section we present two sets of IV estimates designed to address this concern.

5.1 U.K. instruments

Our first set of instruments are constructed from data on characteristics of U.K. industries. If two industries are coagglomerated in the U.S. for purely random reasons or because they value different (unobserved) natural advantages that are randomly correlated in the U.S. (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 U.K. In this case, characteristics of U.K. industries provide measures of the Marshallian factors for the industry pair orthogonal to the endogenous variation in the U.S. Of course, this will only work in some situations. If two industries are coagglomerated in the U.S. because they have a greater need for a coastal location, then they will likely be coagglomerated in the U.K. as well. In this case, the U.K. characteristics of the industry pair could be affected by a correlated endogeneity.

Our U.K.-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 Maskus, C. Sveikauskas, and Allan Webster (1994) and Maskus and Webster (1995) aggregated those into 80 sectors that could be matched with U.S. industries. The construction of the U.K. instruments is otherwise comparable to that undertaken with BEA data. We will use these U.K. input-output measures as instruments for the U.S. input-output relationships under the identifying assumption that U.K. material flows are correlated with true Marshallian dependencies among U.S. industries but uncorrelated with the reverse causation that may have arisen within the U.S. after industrial locations are determined.

Our U.K.-based instrument for labor market similarities was constructed using data from the U.K. Labour Force Survey (which is roughly akin to the U.S. Current Population Survey). The U.K. does not publish a detailed equivalent of the BLS NIOEM matrix, so we constructed our own by pooling six years (2001-2006) of the U.K. Labour Force Survey. We mapped the British industry codes to the SIC3 system, but did not attempt to map the British occupations to NIOEM equivalents. We just 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 appendix describes instruments developed through patent citations using instances where both the citing and cited U.S. patent were filed from the U.K., 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.

5.2 U.S. spatial instruments

Our second set of instruments are constructed using disaggregate data that allow us to examine industry characteristics in different parts of the U.S. Most industry-pair coagglomerations are sufficiently weak so that one can find parts of the U.S. where industry i is present and industry j is not overrepresented. We construct measures of 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. Of course, there are other situations in which the 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.

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 micro-records 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 non-manufacturing. The appendix describes the materials trailers data in greater detail and the variants of this instrument that we tested.

Our spatial instruments for labor similarity are developed from 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 appendix again describes these data in greater detail and the variants of this instrument that we tested.

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 do not change substantially when simultaneously instrumenting for both labor and input-output factors.

5.3 IV regression results

Table 5 presents our core instrumental variables results using U.K. and U.S. spatial instruments. We instrument for the input-output and labor pooling factors using the instruments described above. The control of coagglomeration due to 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. The estimations are carried out on the restricted sample of 7000 pairwise industry combinations that exclude SIC3 pairs within the same SIC2 sector.

Columns (2) and (3) report IV regressions using the EG state-level coagglomeration as the dependent variable and employ the U.K. instruments and U.S. spatial instruments, respectively. Both instruments, despite their quite different construction, 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 10% 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 appendix extends these IV estimations to other variants of EG and DO metrics.

[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 appendix.

[18] The probability of rejecting the Chi-Squared test is 0.13 and 0.16 using the U.K. and U.S. spatial IVs, respectively.

6 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. Taken together, the Marshallian factors appear to have a stronger effect on coagglomeration patterns than the natural advantages which Ellison and Glaeser (1999) found to drive a nontrivial fraction of within-industry agglomeration in the U.S.

Which of Marshall's theories regarding 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. But at least in manufacturing, transport costs for goods, people, and ideas all still seem to matter, and all three of Marshall's theories find vindication in the data.

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 U.K. and U.S. 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 Henderson, J. Vernon (2008): "Networking off Madison Avenue," *Review of Economic Studies*, forthcoming.

Audretsch, David B. and Feldman, Maryann P. (1996): "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 86, 630–640.

Autor, David, Kerr, William and Kugler, Adriana (2007): "Does Employment Protection Reduce Productivity? Evidence from U.S. States," *The Economic Journal*, 117, 189–217.

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

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

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

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

Dunne, Timothy, Roberts, Mark and Samuelson, Larry (1989a): "The Growth and Failure of Manufacturing Plants in the U.S.," *Quarterly Journal of Economics*, 104, 671–698.

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

Duranton, Gilles and Overman, Henry (2005): "Testing for Localization Using Micro-Geographic Data," *Review of Economic Studies*, 72, 1077–1106.

Duranton, Gilles and Overman, Henry (2008): "Exploring the Detailed Location Patterns of UK Manufacturing Industries Using Microgeographic Data," *Journal of Regional Science*, 48, 313–343.

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

Ellison, Glenn and Glaeser, Edward L. (1999): "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?," *Papers and Proceedings, American Economic Review*, 89, 311–316.

Florence, P. Sargant (1948): *Investment, Location and Size of Plant.* London, U.K.: Cambridge University Press.

Fuchs, Victor (1962): *Changes in the Location of Manufacturing in the US Since 1929.* New Haven, CT: Yale University Press.

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

Glaeser, Edward L. and Kerr, William (2008): "Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain?," Harvard University Working Paper.

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

Glaeser, Edward L. and Khan, Matthew (2001): "Decentralized Employment and the Transformation of the American City," NBER Working Paper 8117.

Griliches, Zvi (1990): "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature,* 28, 1661–1707.

Hall, Bronwyn, Jaffe, Adam and Trajtenberg, Manuel (2001): "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," NBER Working Paper 8498.

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

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

Hoover, E.M. (1948): *The Location of Economic Activity.* New York, NY: McGraw Hill.

Jaffe, Adam, Trajtenberg, Manuel and Fogarty, Michael (2000): "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors," *Papers and Proceedings, American Economic Review*, 90, 215–218.

Jaffe, Adam, Trajtenberg, Manuel and Henderson, Rebecca (1993): "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108, 577–598.

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

Kerr, William (2008): "Ethnic Scientific Communities and International Technology Diffusion," *The Review of Economics and Statistics*, 90, 518–537.

Kerr, William and Nanda, Ramana (2008): "Democratizing Entry: Banking Deregulations, Financing Constraints, and Entrepreneurship," HBS Working Paper 07-033.

Kim, Sukkoo (1999): "Regions, Resources, and Economic Geography: Sources of U.S. Regional Comparative Advantage, 1880-1987," *Regional Science and Urban Economics*, 29, 123–137.

Kolko, Jed (1999): "Can I Get Some Service Here? Information Technology, Service Industries, and the Future of Cities," Harvard University Working Paper.

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

Krugman, Paul (1991b): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483–499.

Marshall, Alfred (1920): *Principles of Economics*. London, U.K.: MacMillan and Co.

Maskus, Keith, Sveikauskas, C. and Webster, Allan (1994): "The Composition of the Human Capital Stock and Its Relation to International Trade: Evidence from the U.S. and Britain," *Weltwirtschaftliches Archiv*, 1994, Band 130, Heft 1.

Maskus, Keith and Webster, Allan (1995): "Factor Specialization in U.S. and U.K. Trade: Simple Departures from the Factor-Content Theory," *Swiss Journal of Economics and Statistics*, 1, 419–440.

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

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

Rosenthal, Stuart S. and Strange, William C. (2003): "Geography, Industrial Organization, and Agglomeration," *The Review of Economics and Statistics*, 85, 377–393.

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

Saxenian, AnnaLee (1994): *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Scherer, Frederic M. (1984): "Using Linked Patent Data and R&D Data to Measure Technology Flows," in Griliches, Zvi (ed.) *R & D, Patents and Productivity*. Chicago, IL: The University of Chicago Press.

Silverman, Brian (1999): "Technological Resources and the Direction of Corporate Diversification: Toward an Integration of the Resource-Based View and Transaction Cost Economics," *Management Science*, 45, 1109–1124.

Thompson, Peter and Fox-Kean, Melanie (2005): "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment," *American Economic Review,* 95, 450–460.

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.

Table 1: Descriptive Statistics for Pairwise Coagglomeration Regressions

A. Pairwise EG Coaggl. Measures	Mean	Stand. Dev.	Minimum	Maximum
EG State Total Empl. Coaggl.	0.000	0.013	-0.065	0.207
EG PMSA Total Empl. Coaggl.	0.000	0.006	-0.025	0.119
EG County Total Empl. Coaggl.	0.000	0.003	-0.018	0.080
EG State Firm Birth Empl. Coaggl.	0.000	0.015	-0.082	0.259
EG Expected Coaggl. Due to Natural Advantages	0.000	0.001	-0.008	0.022

B. Pairwise DO Coaggl. Measures	Industry	Relevant Industries (non-zero)		
	Count	Mean	Stand. Dev.	Maximum
DO Global Localization Coaggl., 1000 mi.	7371	0.133	0.073	0.454
DO Global Dispersion Coaggl., 1000 mi.	10	0.592	0.078	0.746
DO Expected Global Localization Coaggl., 1000 mi.	7381	0.181	0.027	0.256
DO Global Localization Coaggl., 250 mi.	6429	0.017	0.019	0.283
DO Global Dispersion Coaggl., 250 mi.	952	0.042	0.029	0.307
DO Expected Global Localization Coaggl., 250 mi.	7381	0.029	0.010	0.077

C. Marshallian Factors	Mean	Stand. Dev.	Minimum	Maximum
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 Tech Maximum	0.005	0.026	0.000	0.625
Patent Citation Tech 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 7381 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. App. Tables 1-5 provide additional descriptive statistics.

Table 2: Highest Pairwise Coagglomerations

Rank	Industry 1	Industry 2	Coaggl.
		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
		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.

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 Employment	PMSA Total Employment	County Total Employment	State Entry Employment	Bilateral Firm Employments with Localization Threshold 1000 mi.	500 mi.	250 mi.	100 mi.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Natural Advantages [DV Specific]	0.210	0.188	0.222	0.120	0.442	0.406	0.253	0.531
	(0.018)	(0.015)	(0.013)	(0.017)	(0.013)	(0.012)	(0.012)	(0.019)
R^2	0.044	0.036	0.049	0.014	0.200	0.165	0.064	0.282
Labor Correlation	0.180	0.106	0.082	0.077	-0.155	0.008	0.127	0.103
	(0.016)	(0.016)	(0.014)	(0.013)	(0.012)	(0.013)	(0.014)	(0.014)
R^2	0.032	0.011	0.007	0.006	0.024	0.000	0.016	0.011
Input-Output	0.205	0.167	0.130	0.112	0.100	0.162	0.188	0.112
	(0.037)	(0.027)	(0.020)	(0.028)	(0.020)	(0.030)	(0.037)	(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	0.148	0.107	0.089	0.046	0.107	0.136	0.094
	(0.034)	(0.029)	(0.019)	(0.021)	(0.017)	(0.026)	(0.034)	(0.027)
R^2	0.032	0.022	0.012	0.008	0.002	0.011	0.019	0.009
Technology Flows Patent Citations	0.081	0.100	0.085	0.068	-0.001	0.056	0.103	0.092
	(0.011)	(0.015)	(0.014)	(0.011)	(0.012)	(0.012)	(0.013)	(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 7381 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. Robust standard errors are reported in parentheses.

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		0.162	0.172	0.251		0.252	0.253
	(0.015)		(0.015)	(0.015)	(0.012)		(0.012)	(0.013)
Labor Correlation	0.118	0.146	0.114	0.085	0.069	0.098	0.066	0.029
	(0.013)	(0.014)	(0.013)	(0.012)	(0.012)	(0.013)	(0.012)	(0.011)
Input-Output	0.146	0.149		0.110	0.162	0.150		0.177
	(0.032)	(0.032)		(0.024)	(0.035)	(0.035)		(0.034)
Input			0.106				0.097	
			(0.026)				(0.029)	
Output			0.093				0.107	
			(0.035)				(0.038)	
Technology Flows Scherer R&D	0.096	0.112	0.079	0.046	0.076	0.075	0.065	0.033
	(0.031)	(0.031)	(0.030)	(0.022)	(0.033)	(0.034)	(0.032)	(0.022)
R^2	0.103	0.077	0.110	0.059	0.113	0.051	0.117	0.102
Observations	7381	7381	7381	7000	7381	7381	7381	7000

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. App. Table 6 provides additional robustness checks. Variables are transformed to have unit standard deviation for interpretation. Robust standard errors are reported in parentheses.

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 Estimation	UK IV	US Spatial IV
	(1)	(2)	(3)	(4)	(5)	(6)
Natural Advantages	0.173	0.173	0.171	0.254	0.210	0.233
[DV Specific]	(0.015)	(0.018)	(0.016)	(0.012)	(0.016)	(0.013)
Labor Correlation	0.083	0.079	0.091	0.027	0.501	0.248
	(0.012)	(0.059)	(0.025)	(0.011)	(0.070)	(0.024)
Input-Output	0.122	0.191	0.185	0.186	0.164	0.213
	(0.025)	(0.049)	(0.041)	(0.033)	(0.059)	(0.050)
Observations	7000	7000	7000	7000	7000	7000

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. App. Tables 7 and 8 report first stages and additional robustness checks. Variables are transformed to have unit standard deviation for interpretation. Robust standard errors are reported in parentheses.