Agglomeration and Co-Agglomeration of Services Industries

Kolko, Jed
Public Policy Institute of California

April 2007

Agglomeration and Co-Agglomeration of Services Industries

Jed Kolko*
Public Policy Institute of California
April 2007

ABSTRACT

Economic research on industry location and agglomeration has focused nearly exclusively on manufacturing. This paper shows that services are prominent among the most agglomerated industries, especially at the county level. Because traditional measures of knowledge spillovers, natural resource inputs, and labor pooling explain little of agglomeration in services industries, this paper takes an alternative approach and looks at co-agglomeration to assess why industries cluster together. By considering the location patterns of pairs of industries instead of individual industries, the traditional agglomeration explanations can be measured more richly, and additional measures – like the need to locate near suppliers or customers – can be incorporated.

The results show that co-agglomeration between pairs of services industries is driven by knowledge spillovers and the direct trading relationship between the industries, especially at the zip code level. Information technology weakens the need for services industries to co-agglomerate at the state level, perhaps because electronic transport of services outputs lowers the value of longer-distance proximity. These results are in sharp contrast to results for manufacturing, for which labor pooling contributes most to co-agglomeration, and the direct-trading relationship contributes more to state-level co-agglomeration. These differences between services and manufacturing are consistent with simple models of transport costs.

* Thanks to Davin Reed for excellent research help and to seminar participants at PPIC and the Kiel Institute Workshop on Agglomeration and Growth in Knowledge-Based Societies for comments.

Services now dominate the US and other advanced economies, but little is known about the location decisions that services firms make. Empirical work in economic geography has focused almost exclusively on manufacturing even though there are many reasons to expect that findings from research on manufacturing industries cannot be generalized to services industries.

This paper extends research on agglomeration in manufacturing to services industries. Services are prominent among the most agglomerated industries, especially at the county level. After showing that the traditional explanations for manufacturing agglomeration explain little of the agglomeration in services industries, the paper proposes that co-agglomeration provides a richer way to explain location patterns in both manufacturing and services. The analysis reveals that several forces, including the need to be near suppliers and customers, contribute to co-agglomeration in services. There are significant differences between services and manufacturing, some of which can be explained by how transport costs in the two sectors differ. Advances in information technology appear to lower transport costs and discourage co-agglomeration for services over longer distances while reinforcing the benefits of geographic proximity at very short distances.

Background

Theoretical work explaining the micro-foundations of agglomeration has drawn primarily on manufacturing industries as examples and for inspiration. Marshall (1892) wrote down his explanations for why firms in an industry cluster in an era when manufacturing dominated the economy. Recent theoretical work, however, has pointed to examples from services, at least in passing. Krugman (1991) draws from manufacturing-sector examples of carpet-making, high-tech equipment, and automobiles, but he does highlight tradable services like insurance and entertainment as highly agglomerated and notes that the "most spectacular examples of localization in today's world are, in fact, based on services rather than manufacturing" (p. 66). Fujita, Krugman, and Venables (1999) mention financial services before focusing on manufacturing industries in their formal model of industrial clustering.

Empirical work on economic geography has been more consistent in excluding services industries, even though this research relies on a variety of methodologies and data sources. Studies focusing only on manufacturing include research that explains differences in industrial concentration level with measures that proxy for micro-foundations of agglomeration (Rosenthal and Strange 2001), research that assesses geographic patterns of innovative activity (Audretsch and Feldman 1996), and research that attributes differences in local industry growth rates or business creation to initial differences in local industrial diversity or specialization (Dumais, Ellison, and Glaeser, 2002; Henderson 1997). Other fields like urban planning and sociology have long studied the geography of services industries, particularly the tendency of business services functions to be highly agglomerated and to locate in major urban areas (Sassen 1991).[1]

There are some exceptions to the exclusive focus on manufacturing in economic geography. Jaffe, Trajtenberg, and Henderson (1993) do not restrict their study of the localization of patent citations to manufacturing, though they do point out that patents refer to commercial uses of "devices," not ideas or algorithms, so the implications of their

[1] Jane Jacobs (1969), the inspiration for much recent research on the economic advantages of industrial diversity, points to examples from both manufacturing (brassieres, p. 51; adhesives, p. 52) and services (health care, p. 54; restaurants, p. 56).

work apply primarily to sectors involved in the production and distribution of tangible outputs. Glaeser et al. (1992) look at growth in the largest local industries, regardless of sector, and therefore include several services industries. However, because they select the largest industries by metropolitan area ranked by 1956 employment, manufacturing and goods-distribution industries dominate their sample. Finally, Kolko (2000a) looks at differences between services and manufacturing industries in how their growth depends on the proximity of suppliers, customers, and labor; this paper also demonstrates the increasing concentration of business services in large cities.

Including services industries in the study of agglomeration comes at some cost. Many of the measures used to explain or illustrate agglomeration in manufacturing industries – such as R&D spending, patents, or the importance of natural resource inputs – are harder to interpret in the context of services industries, if they can even be constructed from available data. Bringing services into the same empirical framework as manufacturing, therefore, requires using datasets that cover all sectors of the economy and developing measures for agglomerative forces that are available and plausible across sectors.

However, the benefits of extending agglomeration research to services could be considerable. Services constitute a large and increasing share of the economy, and any public policy or business decision that is based on evidence from manufacturing could be misleading. Furthermore, with their lower reliance on natural resources inputs, the location decisions of services industries are potentially less determined by nature than those of manufacturing are. Thus, services may be in a better position to reveal the micro-foundations of why firms locate together, and, from a policy perspective, could be particularly responsive to public-sector attempts to lure industries. Finally, if recent improvements in information technology (IT) affect the location decisions of firms, services should be affected most. The output of many services industries can be transmitted electronically, so advances in IT represent a dramatic reduction in transport costs for many services industries.[2]

What Services Are

Studying services industries begins by defining them. Private-sector economic activity divides into the production and distribution of tangible goods and the production and distribution of intangible goods. The production and distribution of tangibles includes natural resource and extractive industries, construction, manufacturing, trade, and transportation (North American Industry Classification System, or NAICS, codes 11-49). Together, these industries accounted for 41% of private, non-farm employment in 2004 according to County Business Patterns. The manufacturing sector alone accounted for 12% of employment (see Table 1).

The production and distribution of intangible goods accounted for the remaining 59% of private, non-farm employment. These include professional and business services like law firms and management consultants; personal services, like health care and education; and the entertainment and hospitality industries (NAICS 54-81). The information industries (NAICS 51) and the financial, insurance, and real estate industries

[2] The effect of reduced communication costs on location decisions is ambiguous, and, despite predictions in the 1990's to the contrary, the Internet did not cause cities to become obsolete. Gaspar and Glaeser (1996) show that, theoretically, electronic and face-to-face communications can instead be complements rather than substitutes. Kolko (2000b) and Sinai and Waldfogel (2004) offer empirical evidence that the Internet both substitutes and complements for non-electronic communications, depending on the nature of the communications.

(NAICS 52-53) also involve the production and distribution of primarily intangible goods, even though the NAICS and earlier SIC classification do not call all of these industries "services."[3]

In this paper, "services" includes industries involved in the production and distribution of intangibles, including the information, finance, insurance, and real estate sectors, corresponding to NAICS 51-81. "Manufacturing" refers to goods-producing industries, corresponding to NAICS 31-33, which is consistent with the definition of manufacturing in other research on economic geography and agglomeration. The characteristics of services industries, outlined below, apply to the information and financial, insurance, and real estate industries, justifying their inclusion in "services" here.

Although services, as defined in this paper, is a broad category covering a diverse set of industries, services share several features that distinguish them from manufacturing, all arising from the intangibility of their output, and these distinguishing features could affect their location decisions. First, services rely less on natural resource inputs than manufacturing does.[4] If clustering around location-dependent natural resources, like coal, contributes to agglomeration, the relative unimportance of natural resources for services removes one reason for agglomeration. Second, services consume less land than manufacturing does, so services might be able to pack together more

---

[3] The Standard Industrial Classification (SIC) separated finance, insurance, and real estate (SIC 6) from services (SIC 7 and 8). The SIC did not categorize information as its own sector: some information industries, like computer programming and data processing (SIC 737) and motion pictures (SIC 78) were part of services; others, like publishing (SIC 27), were considered manufacturing industries; still others, like mobile phone providers (SIC 4812), were considered communications industries. See the NAICS-SIC correspondence site at http://www.census.gov/epcd/www/naicstab.htm.
[4] See summary statistics in Table 4.

closely in space than manufacturing industries can.[5] Third, services and manufacturing face different transport costs. This third point merits elaboration.

In a simple but plausible model, the transport cost of manufacturing output rises linearly with distance and includes a fixed cost that reflects the loading and unloading of goods at both ends. Over short distances, the fixed costs are large relative to the portion of costs that vary with distance.[6] Over long distances, the fixed costs diminish relative to the variable cost, so shipping goods coast to coast costs close to twice as much as shipping goods halfway across the country.[7] Firms facing these shipping costs that trade with each other benefit little from being in the same zip code, since the transport cost savings of being in the same zip code is minimal relative to the total transport cost; over longer-distances, though, the fixed costs shrink in importance relative to distance costs, and firms that trade manufactured goods can reduce their over transport costs by a larger percentage by locating, say, 250 miles apart rather than 500 or 1000 miles apart. Therefore, we might expect manufacturing firms to be indifferent to the distance from their trading partners within a certain radius, and therefore find little advantage in agglomerating at a small geographic level; beyond that radius, firms would be more

[5] Casual observation will have to suffice: while many services occupy high-rise buildings in dense downtowns where land is costly, manufacturing production is typically in low or single-story buildings where land is less expensive.

[6] Residential moves, for instance, are priced nearly identically for a one-mile move or a two-mile move: the only difference would be the marginal cost of the time needed to drive the truck the second mile.

[7] According to www.upsfreight.com, shipping 1000 pounds by truck from San Francisco (zip=94111) costs \$368 for 15 miles (to Oakland, zip=94601), \$517 for 56 miles (to Santa Rosa, zip=95401), \$627 for 388 miles (to Los Angeles, zip=90001), \$1167 for 2132 miles (to Chicago, zip=60601), and \$1543 for 2809 miles (to Washington, zip=20009). This suggests a fixed cost of shipping this weight of over \$300, and a per-mile cost of 30-40 cents.

likely to be sensitive to proximity to trading partners, and therefore would exhibit agglomeration at larger geographic levels.[8]

A simple model of transport costs for services looks different. Consider a service that must be consumed in person, like a haircut or a face-to-face legal discussion, where what is transported is a person (the customer to the barber shop, or the lawyer to her client). Over very short distances, the transport cost equals the opportunity cost of the traveler's time: it costs essentially twice as much to walk four blocks as two blocks, or to drive 10 miles as 5 miles.[9] Beyond the distance at which flying becomes the preferred mode, transport cost varies relatively little by distance: for instance, for a San Francisco management consultant to attend a client meeting in-person, it matters little in cost or time whether that client is in Chicago or New York.[10] If the service output lends itself to being transported by phone or mail – such as a document for signature – over a very short distance it may still be optimal to deliver it face-to-face, but beyond that short distance the cost of the phone call or of using priority mail may be invariant to distance. In these examples, the cost of transporting services rises over short distances when face-to-face is possible, and beyond the face-to-face distance, transport costs are relatively flat with respect to distance. For services that can be delivered electronically, such as data

[8] Here, a firm's trading partners are not necessarily other firms in the same industry. Trading partners could be firms in other industries. The section on co-agglomeration, below, will discuss this in more detail.

[9] For some services that must be delivered in-person, like management consulting engagements, the value may be sufficiently high to warrant paying the travel and time cost to bring in consultants based in another city; for lower-value in-person services like haircuts, almost no one travels any significant distance for a haircut and the cost of transporting the output of haircut services is so high relative to the its value.

[10] To attend a 10 am meeting in Chicago, the San Franciscan might fly the day before at 3 pm, arrive in Chicago at 9 pm, depart Chicago on a 1 pm to arrive at SFO at 3 pm, a 24-hour trip. To attend a 10 am meeting in New York, the San Franciscan would leave home at 1 pm the day before to arrive New York 9 pm, and depart New York on a 1 pm to arrive at SFO at 4 pm, a 27-hour trip. Traveling 50% farther raises the time cost by three hours – a 1/8 increase. The cost of the ticket, booked in advance, would be in the \$300-\$500 range, and even if the New York ticket were 50% more expensive the difference in ticket cost is very small relative to the opportunity cost of the management consultant's time, who might be billed at several hundred dollars per hour.

processing services, the cost of transport is effectively zero regardless of distance. Generalizing across services industries, the absence of fixed costs over short distances suggests that being in the same building or immediate neighborhood as customers could lower transport costs for services industries considerably relative to being across town from customers, though the advantage of being 500 miles away from a customer over being 2000 miles away from customer is relatively small – at least relative to manufacturing.[11] Face-to-face, low-value services, like laundry or haircuts, must be near customers and should exhibit low industry-level agglomeration, but face-to-face and low-value characterize only a subset of the broad category of services. Therefore, we might expect services firms to benefit from proximity to trading partners within a certain radius, and therefore find it advantageous to agglomerate at a small geographic level; beyond that radius, firms would be less sensitive to proximity to trading partners, and therefore would exhibit less agglomeration at larger geographic levels – the opposite of the logic that applies to manufacturing.

These simple models of transport costs imply that information technology usage could affect the location decisions of services and manufacturing differently. The direct effect on information technology is to lower the transport cost of intangibles only: a spreadsheet can be emailed, but a motor can't.[12] Advances in IT might be expected to affect services industry location decisions more than manufacturing location decisions. Electronic communication is, however, a closer substitute for mail and telephone

---

[11] Theoretically, services industries that could rely entirely on phone, mail, or electronic communication with customers would be indifferent to how far away from customers they are, but in practice it is hard to come up with services industries that never use face-to-face communication.

[12] Improvements in information technology can lower the transport costs for tangibles indirectly if it is less costly to arrange for shipping online than by phone; improvements in information technology can also lower transport costs for the entire distribution system by improving tracking, coordination, and other logistics.

communication than it is for face-to-face communication – many interactions, like education or complex negotiations, still are largely face-to-face even though the output is intangible. Information technology, therefore, may not reduce the benefits to services industries of very close proximity to customers as much as they reduce the benefits of longer-distance proximity to customers.[13]

Agglomeration of Services Industries

Following earlier work on agglomeration, this paper uses the Ellison-Glaeser (1997) index of agglomeration. Their index equals the sum across regions of squared deviations between (1) the region's share of an industry's national employment and (2) the region's share of total national employment. If industry employment is distributed across regions identically to the distribution of aggregate employment across regions, that industry exhibits no agglomeration. The index adjusts for both the distribution of region sizes and the level of establishment-level concentration, allowing comparisons of agglomeration at different levels of geography and of industrial aggregation. The index is given in the appendix.

Data on industry location come from two sources. County Business Patterns (CBP) provides employment counts by industry and county. To protect business confidentiality, exact employment counts are suppressed in many cases and instead given as a range, so imputation is sometimes necessary, especially when looking at more disaggregated industries.[14] The descriptive section of this paper considers the finest level

[13] Kolko (2000b) finds that the geographic distribution of commercial Internet domains was highest in isolated larger cities, suggesting that the Internet is a complement for face-to-face interactions (that are primarily within-city) and a substitute for longer-distance communication like phone and postal mail.
[14] The imputation procedure is described in the appendix.

of disaggregation – 6-digits NAICS – while the analytical section relies on 3- or 4-digit NAICS classifications. The other data source, the National Establishment Time-Series (NETS) database, is a longitudinal micro-data set of establishments and includes employment, industry at the 6-digit NAICS level, and street address, including zip code and county. The NETS is constructed by Walls & Associates from the annual Dun & Bradstreet data, which captures nearly all establishments operating in the U.S. Because these data are not confidential, there is no suppression of data, so the employment counts are exact.[15] However, the NETS is expensive, and only a subset containing all California establishments was available for this research. Therefore, this paper will rely on the NETS for constructing zip-code level agglomeration measures, but only within California. For county- and state-level analyses, agglomeration will be measured using CBP data.

Despite the focus of agglomeration research on manufacturing, service industries are among the most agglomerated industries. Using the Ellison-Glaeser index with 2004 CBP data and 6-digit NAICS industries, services account for five of the ten most agglomerated industries at the county level (see Table 2).[16] Motion picture and video production, teleproduction and post-production services, and payroll services are all highly agglomerated and concentrated in Los Angeles County; investment banking is agglomerated in Manhattan, and casino hotels in Clark County, Nevada (Las Vegas). At the state level, services account for only two of the top ten most agglomerated industries (see Table 3). Motion picture and video production is sufficiently concentrated in Los Angeles that its concentration when averaged with the rest of California still causes it to

[15] For a detailed assessment of the NETS database, see Neumark, Zhang, and Wall (forthcoming).
[16] These are the top ten most agglomerated industries among those with at least 10,000 workers in the U.S.

be agglomerated at the state level. Casino hotels have disproportionately high employment in other parts of Nevada, like Reno, not just in Las Vegas, contributing to its high agglomeration at the state level. Several manufacturing industries, like wineries, carpet and rug mills, and cigarette manufacturing, are much more agglomerated at the state level than at the county level: these industries tend to be agglomerated over a larger geographic area that covers multiple counties within a state (California, Georgia, and North Carolina, respectively).

Averaging across all industries in each sector, services are not as agglomerated as manufacturing industries at the county level, and the gap is even larger at the state level. The mean agglomeration index value is .011 for services and .014 for manufacturing at the county level, but the median level of agglomeration at the county level is twice as high for manufacturing as it is for services, suggesting that the average for services is raised more by a few highly agglomerated service industries. At the state level, the mean agglomeration index is .017 for services and .042 for manufacturing, and the medians are .0056 for services and .0207 for manufacturing. Services account for most of the non-agglomerated industries: five of the seven industries with agglomeration indices at or below zero are services, including newspaper publishers, monetary authorities, consumer electronics repair & maintenance, blood and organ banks, and sports teams and clubs.[17]

To explain the variation in agglomeration levels across industries, the analysis uses three measures at the industry level: occupational specialization, natural resource inputs, and workers with graduate degrees. This section follows Rosenthal and Strange

[17] The Ellison-Glaeser index can be negative if, by design or agreement, establishments are located far from each other to prevent competition (which could explain the negative index for sports teams and clubs) or to provide more uniform geographic coverage than the population (which could explain monetary authorities and blood and organ banks).

(2001) and others in using industry-level measures for agglomeration forces as explanations for observed agglomeration. The intuitive meaning of these measures is outlined in the following paragraphs. The formulas and data sources for these measures are summarized in Table 4 and detailed in the appendix.

The occupational specialization measure is intended to capture the importance of labor pooling. Intuitively, if an occupation is concentrated in an industry, then the employment opportunities for workers in that occupation are concentrated in that industry, and those workers should be willing to accept a lower wage if that industry is geographically concentrated so workers could switch employers in the event of a firm-specific shock. In contrast, an industry that hires workers in occupations common to many industries would have less advantage in agglomerating since workers in that occupation would have opportunities outside that industry. The occupational specialization index captures how much an industry's occupational mix diverges from the national occupational mix and is generated from the Bureau of Labor Statistics's (BLS's) National Industry-Occupation Employment Matrix (NIOEM).[18]

The second measure, natural resource inputs, is designed to capture whether an industry depends on a location-specific input like coal or lumber and therefore agglomerates to be near that input. The literature on agglomeration, having been developed to explain the geography of manufacturing industries, has focused on these natural resource inputs, even though they are presumably less important for service industries, which are more labor-intensive and less materials-intensive.[19] The natural-

[18] The summary statistics in Table 4 reveal no difference between services industries and manufacturing industries in their levels of occupational specialization.
[19] The "natural resource" inputs that could help explain why services agglomerate – and where – might include location-specific determinants of the supply of specialized labor. These could include natural

resources measure is the share of an industry's inputs that come from agricultural, forestry, fishing, hunting, and mining industries and is generated from the Bureau of Economic Analysis's (BEA's) Input-Output (IO) accounts. The third measure is the share of workers with graduate degrees, as a proxy for knowledge spillovers; this measure comes from the 2000 Public Use Microdata Sample (PUMS).[20]

Because the data sources use different industry classifications, the 6-digit NAICS industries were aggregated so that each resulting industry had unique information from each of the data sources. This aggregation process resulted in 64 manufacturing industries and 54 services industries. Despite the much larger share of employment in services industries, there are more manufacturing industries in the analysis because the data sources provide information at a finer level of disaggregation for manufacturing than for services industries. In the analysis, manufacturing industries typically correspond to the 4-digit NAICS level, whereas services industries usually correspond to the 3-digit NAICS level.[21]

The index of agglomeration is regressed on these three measures separately for these manufacturing and services industries by estimating the following:

$$agglom_{i,k} = \beta X_i + \varepsilon_{i,k}$$

where $agglom_{i,k}$ is the agglomeration index for industry i at the level of geography k; X_i is the set of industry measures, including occupational specialization, natural resource

amenities that raise quality-of-life, like good weather and proximity to a coast (see Glaeser, Kolko, and Saiz 2001), as well as institutions like universities.

[20] While these measures are typical proxies for the reasons for agglomeration in the literature, there is no consensus on which theoretical explanations for agglomeration are represented by each measure. Rosenthal and Strange (2001) point out the difficulty of choosing a proxy measure for labor pooling; they use three alternatives, one of which is the percentage of workers with high levels of education, which is arguably as suitable a proxy for knowledge spillovers as it is for labor pooling.

[21] See the appendix for more detail on the industry classification.

inputs, and workers with graduate degrees, along with national industry employment as a control; and k refers to zip codes, counties, or states, depending on the specification.[22] Agglomeration is measured at the zip-code level in California and at the county and state levels for the US. The three measures that proxy for agglomeration factors – occupational specialization, natural resource inputs, and graduate degrees – are calculated for each industry at the national level, so their values do not vary with the level of geography at which agglomeration is calculated.

The purpose of this part of the analysis is two-fold: first, to see what, if anything, explains agglomeration in services industries, and second, to see if similar forces explain agglomeration in both manufacturing and services. The results for manufacturing and services are in Tables 5 and 6, respectively. For manufacturing, the relationship between occupational specialization and agglomeration is positive and significant at the 5% level for both county and state agglomeration; the relationship is positive but significant only at the 10% level for zip code agglomeration. The theory of labor pooling suggests that the workers and therefore firms benefit from agglomerating in the same labor market.[23] Labor markets are larger than either zip codes or counties, so an industry should benefit from labor pooling so long as it is agglomerated within a state, even if spread over multiple counties.[24] That the relationship between occupational specialization and agglomeration is strongest at the state level is consistent with interpreting occupational

[22] This set-up follows Rosenthal and Strange (2001).

[23] Labor pooling mitigates the cost to workers of firm-specific shocks only if other firms in the industry are within the same labor market, so workers can switch firms within the same industry without incurring moving costs to a new labor market.

[24] In defining a metropolitan area, the U.S. Office of Management and Budget includes territory with a "high degree of social and economic integration with the core as measured by commuting ties," so metropolitan areas are a reasonable approximation for a local labor market. Metropolitan areas consist of one or, typically, multiple counties, so it is natural to think of labor markets as somewhat larger than a county though not as large as a state. See the OMB standards for defining metropolitan areas at http://www.census.gov/population/www/estimates/00-32997.pdf.

specialization as a proxy for labor pooling. There is no statistically significant relationship between either natural resource inputs or knowledge spillovers, as measured by workers' graduate degrees, and agglomeration at any geographic level for manufacturing industries.

For services, the only measure that contributes to agglomeration is the percent of workers with graduate degrees, interpreted as knowledge spillovers, which is positive and statistically significant only at the zip code level. Neither occupational specialization, interpreted as labor pooling, nor natural resource inputs help explain agglomeration at any level of geography. These findings shed little light on why services agglomerate, although they do suggest that services and manufacturing may agglomerate for different reasons since none of the three factors was significant for both manufacturing and services agglomeration at any level of geography.

Co-Agglomeration of Services Industries

The innovation in this paper is to consider the forces that cause different industries to cluster together, or co-agglomerate. Whereas agglomeration reflects the extent to which firms in the same industry locate near each other, co-agglomeration reflects the extent to which firms in different industries locate near each other. Looking at co-agglomeration offers two advantages over agglomeration, and these advantages are especially relevant when applied to service industries. This section explains the advantages of co-agglomeration and then demonstrates that co-agglomeration analysis highlights significant differences in the factors affecting services and manufacturing industries' location decisions.

The first advantage is that co-agglomeration allows for the possibility that agglomerative forces, like labor pooling, knowledge spillovers, or input sharing, exist between firms in different industries as well as between firms in the same industry.[25] Furthermore, one can characterize the degree of similarity of multiple variables between firms in different industries on a continuous scale: for instance, the opportunity for labor pooling is presumably much higher between the legal services industry and management consulting industry since both employ lawyers than it is between the legal services industry and the bakery products manufacturing industry. In contrast, industry-level agglomeration implicitly characterizes the degree of similarity between firms as discrete: either firms are in the same industry and therefore could benefit from labor pooling, knowledge spillovers, natural resource inputs, and so on, or they are in different industries and therefore can not.

The second advantage of analyzing co-agglomeration is that co-agglomeration makes it possible to consider an additional reason why firms in different industries could locate near each other: industries trade with each other, and transport costs rise with distance. Incorporating trading relationships especially useful for thinking about services because of the different nature of transport costs for services, as outlined above. If service and manufacturing outputs involve different transport costs, then this might be reflected in differences between services and manufacturing in whether industries that trade with each other therefore co-agglomerate..[26]

---

[25] Jacobs (1969) argues that the innovative activity arises in interactions between industries, not within an industry: "when new work is added to older work, the addition often cuts ruthlessly across categories of work, no matter how one may analyze the categories" (p. 62).

[26] There has been very little research on industries locating near each other. Two examples: Ellison and Glaeser (1997) find co-agglomeration to be higher between pairs of manufacturing industries where one is a significant input to the other. Also, Duranton and Puga (2002) show that functional specialization is

To measure co-agglomeration, this paper uses the extension of the Ellison-Glaeser (1997) index to co-agglomeration. Their co-agglomeration index measures the extent to which multiple industries are clustered together geographically in excess of the agglomeration of each of the industries. Like the agglomeration index, their co-agglomeration index adjusts for both the distribution of region sizes and the level of establishment-level concentration. The formula for the co-agglomeration index is provided in the appendix.

As with agglomeration, co-agglomeration can be measured at different levels of geography. For example, (1) tobacco manufacturing and (2) fiber, yarn, and thread mills are highly co-agglomerated at the state level but not at the county or zip code level: these two industries are both concentrated in North Carolina, but each is concentrated in different counties and zipcodes within North Carolina. The same is true for (1) audio and video equipment manufacturing and (2) motion picture, video, and sound recording: both concentrated in California, but the former is in the Bay Area and the latter is in Los Angeles. At the zip-code level, (1) museums, historical sites, and similar institutions and (2) accommodations are highly co-agglomerated, though not at either the county or state level: most counties and states have both of these industries, but within a county the two types of industries tend to concentrate in the same immediate neighborhoods.[27]

The empirical strategy for measuring co-agglomeration is:

$$coagglom_{i,j,k} = \beta X_{i,j} + \Lambda + \varepsilon_{i,j,k}$$

increasingly important, rather than industrial specialization, which implies greater linkages between, rather than within, industries.

[27] Since zip code data were available only in California, these industries are co-agglomerated at the zip-code level within California; their county and state level co-agglomerations were measured using CBP, which is available for the nation.

Whereas the agglomeration analysis uses the industry as the unit of observation, the co-agglomeration analysis uses the pair of industries (i,j) as the unit of observation. Co-agglomeration is measured at level of geography k, which refers to zip codes, counties, or states, depending on the specification. The vector Λ captures industry fixed-effects, and the elements of the vector equal 1 for industries i and j and 0 for all other industries. The vector $X_{i,j}$ is a set of variables capturing the reasons for co-agglomeration between industries i and j, which are independent of geography, and these include:

1. Occupational similarity: how similar are the occupation compositions of industry i and industry j, which will be interacted with other variables.
2. Demographic similarity: how similar are the worker age-education distributions of industry i and industry j.
3. Input similarity: how similar are the inputs for industry i and industry j.
4. Output similarity: how similar are the customers for the output of industry i and industry j.
5. Direct trade: how much of industry j's output is an input for industry i, and how much of industry i's output is an input for industry j, which will be interacted with other variables.

Table 7 summarizes all of these measures and their interpretations, and the appendix defines them in detail.

These measures, plus interactions, map to several forces that could contribute to co-agglomeration. This paper describes the relationship between these measures and the forces of co-agglomeration intuitively with the understanding that alternative interpretations of these measures may be appropriate. These forces and measures are summarized in Table 7.

The importance of labor pooling is represented by the occupational similarity between industries i and j interacted with the occupational specialization of industry i and the occupational specialization of industry j. Intuitively, firms in different industries might benefit from labor pooling if (1) they hire similar labor to each other (occupational similarity) but (2) they hire labor that is different from other industries (occupational specialization). In short, workers (and therefore firms) benefit from labor pooling between industry i and j if workers in industry i have opportunities in industry j and few opportunities in industries other than i or j.

The importance of knowledge spillovers is also represented by occupational similarity between industries i and j, but interacted with the percent of workers with graduate degrees in each industry. Intuitively, knowledge spillovers should be more common in industries with workers in similar occupations whose knowledge could therefore overlap, but more so if workers are highly skilled or specially trained, as proxied by graduate degrees.[28]

The importance of direct trade is measured using the volume of direct trade between industries i and j as a share of overall inputs and outputs of both industries. Because one of the key differences between services and manufacturing is the nature of transport costs, and because information technology is hypothesized to affect transport costs, the levels of information technology intensity in industries i and j are interacted

[28] Although Rosenthal and Strange (2001) use education level as a measure of the potential for labor pooling, it seems more plausible that labor pooling could arise from specialized labor at any skill level, whereas the knowledge spillovers that contribute to innovative activity arise from highly skilled labor, regardless of whether that skilled labor is uniquely employed by a given industry.

with the direct trade measure. It is hypothesized that information technology affects the location decisions of services more than of manufacturing because information technology lowers the cost of transporting many intangible outputs. Thus, for services, the relationship of direct trade between two industries on co-agglomeration is expected to be weaker for more information-technology intensive industries; for manufacturing, the relationship between direct trade and co-agglomeration should not be affected by how information-technology intensive the industries are. The level of information technology intensity is proxied using the share of employees in computing-specialty occupations.[29]

The measure of demographic similarity of workers is designed to capture the possibility that firms follow workers: namely, that industries locate where their workers want to live, and that local amenities serve as a compensating differential that enables firms to pay less for labor than they would in lower-amenity locations. Rather than attempt to identify high-amenity places, this paper assumes that different workers put a different amenity value on different places, and age and education help predict which amenities workers demand. Industries with workers that are demographically similar are hypothesized to co-agglomerate because their workers consider the same locations to be high-amenity.[30]

The two final measures, the similarity of inputs and the similarity of outputs, capture whether the two industries in the pair have similar suppliers and customers.

---

[29] An alternative measure would be the percentage of workers using a computer, or the Internet, or email at work. While the Current Population Survey (CPS) does ask these questions sporadically, the number of responses is very low for many industries, so using CPS data would require aggregating an already-small number of industries further.

[30] Implicit in this interpretation is the assumption that demographics, not occupation, influence tastes for location amenities, and occupation, not demographics, contributes to labor pooling. However, occupational categories do not fully describe how skilled or specialized a worker is, and demographic characteristics are probably correlated with the portion of skills and specialization not fully captured by occupational categories. Nonetheless, the inclusion of this demographic similarity measure is an improvement on past research in the field that did not consider an amenity-driven explanation for firm location decision.

The results of the co-agglomeration analysis are presented in Tables 8-11. The regressions include all of the measures described above.[31] With 64 manufacturing industries, the number of unique manufacturing pairs is (64*63)/2 = 2016, and with 54 services industries, the number of unique services pairs is 1431. In each table, columns 1, 2, and 3 show the results for zip code, county, and state co-agglomeration with all variables except the interaction between direct trade and information technology intensity, and columns 4, 5, and 6 repeat the analysis with the interaction between direct trade and information technology intensity.

The results for manufacturing in Table 8 show that labor pooling, as measured by the interaction between occupational similarity and occupational specialization, is positive and significant at all three levels of geography, with the largest magnitude at the state level, which is consistent with the finding in Table 5 that labor pooling contributes to manufacturing agglomeration. Knowledge spillovers are also positive and significant at the 5% level for zip code co-agglomeration and at the 10% for county co-agglomeration, but not at the state level. The similarity in worker-demanded amenities, proxied by demographics, is also positive and significant at all three levels of geography. The similarity of customers has no relation to co-agglomeration, while the similarity of suppliers is positively and significantly related to co-agglomeration at all three geographic levels. The coefficient on the direct trading relationship between the industries is positive and significant at the 5% level only for state co-agglomeration; for county co-agglomeration it is positive and significant at the 10% level and not significant for zip code co-agglomeration. However, the interaction between trade and information

[31] The industry-level values for occupational specialization, graduate degrees, and information-technology intensity are absorbed in the industry fixed effects variables. The un-interacted occupational similarity measure for the pair of industries is included in every specification but not shown.

technology intensity is positive and significant for both zip code and county co-agglomeration. This means that manufacturing industries that trade with each other are more likely to locate in the same zip codes and counties if the industries rely more on information technology. In other words, high-tech manufacturing industries that trade with each other are more likely to be neighbors than low-tech manufacturing industries that trade with each other.

Because neither the co-agglomeration variable nor the explanatory variables are measured in units that have inherent meaning (like dollars or years), it is useful to look at standardized coefficients that show the effect of a one standard-deviation change in the independent variable on the dependent variable, also measured in standard deviations. Table 9 repeats the results from Table 8 with only the standardized coefficients. Although many forces have a statistically significant positive effect, labor pooling has a much larger effect on co-agglomeration in manufacturing than any other, at all geographic levels. Knowledge spillovers are the second-most important, but the magnitude of their effects are one-fifth and one-tenth of the effect of labor pooling at the zip code and county levels, respectively.

The reasons for co-agglomeration between services industries are rather different. The results are shown in Table 10, with standardized coefficients in Table 11.

Knowledge spillovers are positively related to co-agglomeration at the zip code level only, not at either the county or state level. This is similar to the role of knowledge spillovers in manufacturing co-agglomeration, in that the effect is strongest for zip code co-agglomeration and weakens at larger geographies. Also as with manufacturing, the similarity of demographics is positively related to co-agglomeration at the county and state level, though not at the zip-code level.

The coefficient on the labor pooling variable is not significant at any level of geography, which is a striking difference from manufacturing. A possible reason is that services tend to be more urbanized than manufacturing is (Kolko 2000a). The theory of labor pooling posits that workers will require higher wages if there are fewer local employment opportunities outside their firm, whereas agglomeration or co-agglomeration protects workers from firm-specific shocks. If being in a large labor market also protects workers from shocks, then the benefits of labor pooling within an industry or between a specific pair of industries could be weakened. This could explain why labor pooling contributes less to co-agglomeration of services than to co-agglomeration of manufacturing.

The effect of input-output relationships (direct trade, similarity of customers, and similarity of inputs) is quite different for services than for manufacturing. The similarity of customers is positively and significantly related to co-agglomeration at the zip code and state levels for services, and not for manufacturing. Even more striking is that the strength of the direct trading relationship contributes to the co-agglomeration of services at the zip code level, and is <u>negatively</u> and significantly related to co-agglomeration at the state level. This contrasts the role of direct trade for manufacturing co-agglomeration, where the relationship is positive and significant only at the state level. This difference between services and manufacturing is consistent with the simple models of transport costs sketched above.[32] For services, the model of rapidly increasing transport costs at

[32] The model of transport costs in services industries suggests that the effect of the direct trading relationship on co-agglomeration for services at the state level would be small or zero. The model did not

short distances is consistent with the finding that services industries that trade with each other would benefit from very close proximity; the idea that transport costs are relatively invariant to distance at longer distances is consistent with the finding that, for services, the extent of direct trade contributes more to co-agglomeration at the zip code level, followed by the county level, and least of all at the state level. For manufacturing, the finding that direct trade has the strongest effect on state-level agglomeration is consistent with the theory that transporting tangible goods has a fixed cost, regardless of distance, which is large relative to the variable costs of transport at short distances.

Furthermore, the interaction between direct trade and information technology intensity is positively related to co-agglomeration at the zip code level but negatively at the state level. This means that, as in manufacturing, services industries that trade with each other are more likely to locate in the same zip codes and counties if the industries rely more on information technology. However, unlike manufacturing industries, service industries that trade with each other are <u>less</u> likely to locate in the same state if the industries rely more on information technology. The fact that the interaction coefficient is smaller for services than manufacturing at the county and state levels is consistent with the hypothesis that information technology should lower the transport cost for services output and not for manufacturing output: it is less important for firms that trade to be near each other if they can trade electronically. The positive coefficient on the interaction term for services at the zip code level suggests that information technology might not be a good substitute for the face-to-face interactions that cause services firms that trade to cluster in the zip code, block, or building.

---

suggest that it could be negative; the negative sign on this coefficient is surprising and remains unexplained.

However, the simple model outlined above would imply that the coefficient on the interaction term between direct trade and information technology should be zero (not positive) when looking at manufacturing industries; if IT does not affect the cost of face-to-face communication, then the interaction term should be zero (not positive) for services industries at the zip code level as well. The positive coefficient on the interaction term for manufacturing at the zip code and county levels and for services at the zip code level is unexplained by the simple model of transport costs. This suggests that information-technology intensity could affect location decisions for reasons other than its effect on transport costs. Trade between information-technology intensive industries may require more coordination between the supplier and the customer if the output is more abstract or complex than in non-information-technology intensive industries.[33] Furthermore, the information technology itself could add complexity if the supplier and customer need to agree on electronic formats or application standards. If some of this coordination happens face-to-face, this could explain why coefficient on the interaction between information-technology intensity and direct trade could be positive and larger in magnitude for co-agglomeration at smaller levels of geography.

Conclusions

These findings tell us, first, that some of the micro-foundations assessed in the manufacturing literature help explain the location patterns of services industries. Looking

[33] This appears to be the effect of information technology *per se* and not complexity or technical detail in a general sense. When an interaction term between the direct trading relationship and the percent of workers in the industries with graduate degrees is included, the signs and significance on the interaction between direct trade and IT-intensity do not change for services; for manufacturing, the coefficient in the county-level co-agglomeration regression (table 7, column 5) remains positive but is no longer statistically significant.

at services through the lens of co-agglomeration reveals these relationships more clearly than a simpler framework of industry agglomeration. Analyzing co-agglomeration rather than agglomeration makes it possible to consider more factors that explain location decisions, like proximity to suppliers and customers. Co-agglomeration also allows for the possibility that traditional agglomerative forces, like knowledge spillovers and labor pooling, apply between firms in different industries, not only between firm in the same industry. Although this paper uses co-agglomeration in order to highlight location patterns in services industries, using co-agglomeration also deepens our understanding about why manufacturing industries locate where they do.

A second conclusion is that the factors that explain manufacturing industry location differ from those that explain services industry location. In the co-agglomeration framework, the three micro-foundations of agglomeration emphasized in previous research – labor pooling, knowledge spillovers, and input-sharing – are all significant for manufacturing at multiple levels of geography. For services, however, labor pooling has no significant effect, perhaps due to the prevalence of services in larger labor markets, yet other forces explain co-agglomeration better for services than for manufacturing, such as the direct trading relationship, which affects zip code level co-agglomeration for services only, and sharing the same customers, which affects zip code and state level co-agglomeration for services yet has no affect on any geographic level of co-agglomeration for manufacturing.

A third conclusion is that proximity between suppliers and customers does not affect only the geography of services industries. The magnitude of the direct trade coefficient is larger for manufacturing (at the state level: .14) than it is for services (at the zip code level: .12) While it is easier to imagine extreme examples within services, like haircuts, of industries whose need to be near customers would trump all other location decisions, in fact manufacturing industries also locate together if they trade with each other. The key difference is not that proximity help trade for services and not for manufacturing; rather, services benefit more from proximity with trading partners only at short distances, while manufacturing benefits from proximity with trading partners even at longer distances.

Finally, these results suggest that information technology can either encourage or discourage co-agglomeration between industries that trade with each other. Information technology encourages co-agglomeration for services that trade with each other at the zip code level and discourages it at the state level, while encouraging co-agglomeration for manufacturing at both the zip code and county levels, with no effect at the state level. This paper argues that the differential effect of information technology on manufacturing and services is because electronic communication dramatically lowers the cost of transporting intangibles, especially over longer distances, but not the cost of transporting tangible goods. However, because information technology encourages co-agglomeration, information technology appears to have other effects on firms that trade with each other. While information technology lowers transport costs, high-IT industries appear to benefit more from face-to-face coordination than low-IT industries do.

REFERENCES

Audretsch, David, and Maryann Feldman, "R&D Spillovers and the Geography of Innovation and Production," American Economic Review, Vol. 86, No. 3, pp. 630-640, 1996.

Dumais, Guy, Glenn Ellison, and Edward Glaeser, "Geographic Concentration as a Dynamic Process," The Review of Economics and Statistics, Vol. 84, No. 2, pp. 193-204, 2002.

Duranton, Gilles, and Diego Puga, "From Sectoral to Functional Urban Specialization," NBER Working Paper #9112, 2002.

Fujita, Masahisa, Paul Krugman, and Anthony Venables, The Spatial Economy, Cambridge, MA: MIT Press, 1999.

Gaspar, Jess, and Edward Glaeser, "Information Technology and the Future of Cities," Journal of Urban Economics, Vol. 43, pp.136-156, 1998.

Glaeser, Edward, Hedi Kallal, Jose Scheinkman, and Andrei Shleifer, "Growth in Cities," Journal of Political Economy, Vol. 100, No. 6, pp. 1126-1152, 1992.

Glaeser, Edward, Jed Kolko, and Albert Saiz, "The Consumer City," Journal of Economic Geography, Vol. 1, No. 1, 2001.

Henderson, Vernon, "Externalities and Industrial Development," Journal of Urban Economics, Vol. 42, No. 3, pp. 449-470, 1997.

Jacobs, Jane, The Economy of Cities, New York: Vintage, 1969.

Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," Quarterly Journal of Economics, Vol. 108, No. 3, pp. 577-598., 1993.

Kolko, Jed, "Can I Get Some Service Here? Information Technology, Service Industries, and the Future of Cities," unpublished manuscript, 2000a.

Kolko, Jed, "The Death of Distance? The Death of Cities? Evidence from the Geography of Commercial Internet Usage," in The Internet Upheaval, eds. Ingo Vogelsang and Benjamin Compaine, MIT Press: 2000b.

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

Marshall, Alfred. 1892. *Elements of the Economics of Industry*. London: Macmillan.

Neumark, David, Junfu Zhang, and Brandon Wall, "Employment Dynamics and Business Relocation: New Evidence from the National Establishment Time-Series," Research in Labor Economics, forthcoming.

Rosenthal, Stuart, and William Strange, "The Determinants of Agglomeration," Journal of Urban Economics, Vol. 50, pp. 191-229, 2001.

Sassen, Saskia, The Global City: New York, London, Tokyo, Princeton: Princeton University Press: 1991.

Sinai, Todd, and Joel Waldfogel, "Geography and the Internet: is the Internet a Substitute or Complement for Cities?", Journal of Urban Economics, Vol. 56, pp. 1-24, 2004.

APPENDIX: DATA SOURCES AND VARIABLE DEFINITIONS

County Business Patterns

County Business Patterns (CBP) is the source for employment counts at the county and state levels. CBP is an annual tabulation of the Census Bureau's register of all business establishments, which is generated from the quinquennial Economic Censuses, the annual Company Organization Survey, the Annual Survey of Manufactures, and administrative records. CBP covers all private-sector non-farm employment in establishments with at least one paid employee. The total employment covered by CBP was around 115 million employees in 2004.

A record in CBP is a county-industry cell, where industries are reported down to the 4-digit SIC level. For each industry-county cell, an employment figure is given, which is either an exact figure or a range (1-4, 5-9, 10-19, etc.). A range, rather than an exact figure, is given when the number of establishments is sufficiently small that an exact figure would disclose information about a particular establishment. Also reported for each industry-county cell is the number of establishments and the number of establishments in each of several establishment-size ranges (1-4, 5-9, etc.). These establishment counts are always exact, never ranges. To impute industry-county employment figures when only a range is given, a second range is constructed using the establishments-by-establishment-size count. Thus, the exact employment count lies, with certainty, in the intersection of the two ranges. For each industry, a point in the intersection of the ranges was chosen such that the resulting estimates, when added to the exact figures for other cells, added up to the industry's national employment total. That point was a uniform distance between the lower and upper bound of each cell's range (say, 40% from the lower bound) for each industry; for each industry a separate distance was calculated.

The actual (or, where necessary, estimated) employment count for industry i in county x is $emp_{i,x}$ in the variable definitions, below. Total employment across industries in county x is emp_x, and total employment across counties in industry i is emp_i. Total national employment, in all industries and in all counties, is *emp*.

Documentation for the CBP is available on-line at http://www.census.gov/epcd/cbp/view/cbpview.html.

The National Establishment Time-Series (NETS) Database

The NETS is the source for employment counts at the zip code level. The NETS is a longitudinal file, created by Walls & Associates, from the register of business establishments tracked by Dun & Bradstreet. For this research only a subset of California data were available. The NETS provides uncensored employment counts and addresses at the establishment level, so no imputation is necessary in creating employment counts at the zip code – industry level. Detailed information about the NETS and an assessment of its quality is available in Neumark, Zhang, and Wall (forthcoming).

CBP is the basis for calculating agglomeration and co-agglomeration at the county and state levels. The NETS is the basis for calculating agglomeration and co-agglomeration at the zip code level. The agglomeration and co-agglomeration measures follow Ellison and Glaeser (1997).

Ellison-Glaeser measure of agglomeration (following their notation):

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

$$G = \sum_i (s_i - x_i)^2$$

$H = \sum_j z_j^2$ (industry Herfindahl index)

s_i=share of industry employment in geographic area i

x_i=share of national employment in geographic area i.

z_j=share of industry employment in establishment j

The index is the sum of squared differences between industry and national employment shares across geographic areas, adjusted for (1) the size distribution of geographic areas and (2) the Herfindahl index of the industry establishment size distribution.

Ellison-Glaeser measure of co-agglomeration (following their notation) across J industries, j = 1 to J, which constitute an industry group.

$$\gamma^c = \frac{\left[\dfrac{G}{1-\sum_i x_i^2}\right] - H - \sum_j \hat{\gamma}_j w_j^2 \left(1 - H_j\right)}{1 - \sum_j w_j^2}$$

$H = \sum_j w_j^2 H_j$ (weighted Herfindahls of industry establishment size distributions)

G is the raw concentration (as defined above) for industry group employment

H_j is the Herfindahl index of industry j's establishment size distribution

W_j is industry j's share of industry group employment

γ_j is the agglomeration index for industry j (as defined above)

Input-Output Accounts

The 2004 Input-Output (IO) accounts are the source for information on customer-supplier relationships among industries and consumption by final users (consumers and government). The IO accounts estimate the value of commodity flows between pairs of industries. The IO accounts are developed by the Bureau of Economic Analysis, based on the quinquennial Economic Censuses conducted by the Census Bureau and numerous other sources. Both physical (i.e., manufacturing) and non-physical (i.e., services) goods are included. Additional input sources and output destinations are included: namely, labor is included as an input source, and households and government are included as output destinations.

Documentation for the IO accounts is available on-line at http://www.bea.gov/bea/papers/IOmanual_092906.pdf.

In the IO accounts, industries can use their own output as an input in the production process. These "circular flows" are excluded. The key variables generated from the IO accounts are the direct trade variable, the similarity of inputs, and the similarity of outputs for the co-agglomeration analysis, as well as the natural resource inputs variable for the agglomeration analysis.

Between any pair of industries i and j, there are four possible measures of the strength of direct trade between them. Let input_k and output_k represent the total inputs from other industries consumed in industry k's production process and the total outputs generated by industry k's production process, excluding output from industry k that is also an input for industry k. If $b_{i \rightarrow j}$ equals the value of industry i's output used as an input by industry j and $b_{j \rightarrow i}$ equals the value of industry j's output used as an input by industry i, then the four measures of direct trade are:

1. $b_{i \rightarrow j}/input_j$
2. $b_{i \rightarrow j}/output_i$
3. $b_{j \rightarrow i}/input_i$
4. $b_{j \rightarrow i}/output_j$

These four measures reflect the fact that industry i and j might be of different size, and the amount of trade $b_{j \rightarrow i}$, for instance, could reflect a very different share of industry i's overall inputs than it does of industry j's overall outputs.

The **direct trade** ($trade_{ij}$) variable is calculated as the average of the four underlying measures, and the results of the analysis are not changed when using only the maximum of the four measures.

The **output similarity** variable, $outputsim_{ij}$, is equal to the sum of absolute differences between the shares of industries' i and j's outputs going to each customer k, where k=all other industries, consumers, and government:

$$outputsim_{ij} = \frac{\left(2 - \sum_k \left| \frac{b_{i \to k}}{output_i} - \frac{b_{j \to k}}{output_j} \right| \right)}{2}$$

, which equals 1 if industries i and j have perfectly overlapping distributions of customers and 0 if they have non-overlapping distributions of customers.

The **input similarity** variable, $inputsim_{ij}$, is equal to the sum of absolute differences between the shares of industries' i and j's inputs coming from each supplier k, where k=all other industries:

$$inputsim_{ij} = \frac{\left(2 - \sum_k \left| \frac{b_{k \to i}}{input_i} - \frac{b_{k \to j}}{input_j} \right| \right)}{2}$$

, which equals 1 if industries i and j have perfectly overlapping distributions of suppliers and 0 if they have non-overlapping distributions of suppliers.

The **natural resource inputs** ($nature_i$) measure is the share of inputs to industry i that come from crop or animal production, forestry, logging, fishing, or mining (NAICS 11 and 21, with the exception of support activities within those categories).

National Industry-Occupation Employment Matrix

The National Industry-Occupation Employment Matrix 2004 (NIOEM) is the source for occupation data. The Bureau of Labor Statistics produces the NIOEM from Occupational Employment Statistics, Current Employment Statistics, and the Current Population Survey.

The NIOEM presents employment counts in industry-occupation cells for around 300 industries and around 700 occupations. This paper uses the summary occupation codes, which aggregate the 700 occupations into 93 occupational groups.

The **occupational similarity** variable, $occsim_{ij}$, is equal to the sum of absolute differences between the shares of industries' i and j's workforces in occupation k, where occ_{ik}= share of industry i's workforce in occupation k:

$$occsim_{ij} = \frac{\left(2 - \sum_k \left| occ_{ik} - occ_{jk} \right| \right)}{2}$$

, which equals 1 if industries i and j have perfectly overlapping distributions of occupations and 0 if they have non-overlapping distributions of occupations.

The **occupational specialization** variable, occ_i, is equal to the sum of absolute differences between the share of occupation k in the economy (occ_k) and the share of occupation k of employment in industry i:

$$occspec_i = \frac{\left(2 - \sum_k \left| occ_{ik} - occ_k \right| \right)}{2}$$

, which equals 1 if industry i has a distributions of occupations identical to the economy in aggregate.

The NIOEM also provides the **share of workers with in computer specialist occupations** ($tech_i$) used in the co-agglomeration analysis interacted with the direct trade measure measure.

Documentation for the NIOEM is available on-line at http://www.bls.gov/emp/nioem/empioan.htm.

Public Use Microdata Sample

The 2000 Public Use Microdata Sample (PUMS) of the U.S. Census provides individual-level data on age and education level of workers, by industry. Using six age groups and eight education categories, the distribution of workers across 48 age-education cells was calculated by industry.

The **demographics similarity** variable, demosim$_{ij}$, is equal to the sum of absolute differences between the shares of industries' i and j's workforces in each age-education cell k, where demo$_{ik}$= share of industry i's workforce in age-education cell k:

$$demosim_{ij} = \frac{\left(2 - \sum_k \left| demo_{ik} - demo_{jk} \right| \right)}{2}$$

, which equals 1 if industries i and j have perfectly overlapping distributions of age-education cells and 0 if they have non-overlapping distributions of age-education cells.

The PUMS also provides the **share of workers with graduate degrees** (grad$_i$) used in the agglomeration analysis as well as in the co-agglomeration analysis interacted with the occupational similarity measure.

Industry definitions

Data on employment in the CBP and the NETS are available at the 6-digit NAICS level. The other data sources – the IO accounts, NIOEM, and the PUMS – are available at the 4-digit NAICS level or, for many industries, only at the 3- or 2-digit level. In creating the industry classification used in this paper, the classifications from all four data sources were aggregated so that each industry has a unique value from each data set.

For instance, one industry used in this paper is NAICS 722, Food Services and Drinking Places, rather than using the underlying 4-digit industries: NAICS 7221 (Full Service Restaurants), 7222 (Limited Service Restaurants), 7223 (Special Food Services, like caterers), and 7224 (Drinking Places). The CBP, NETS, and NIOEM provide separate data for NAICS 7221, 7222, 7223, and 7224. However, the Census industry code 868, used in the PUMS, combines NAICS 7221, 7222, and 7223, and Census code 869 corresponds to NAICS 7224. The IO accounts use BLS industry code 168, which correspond to NAICS 722 in aggregate. Thus, in order to avoid measurement error from assigning values from Census code 868 or BLS code 168 to all the component 4-digit NAICS codes, the industry classification in this paper uses NAICS 722, for which data is available for every source. The greater precision in the CBP, NETS, and NIOEM is lost, of course, by not using their data at the finest level of disaggregation available.

This table shows the number of industries that each data source uses within the manufacturing and services sector:

	Manufacturing	Services
County Business Patterns & NETS (NAICS-based)	86	109
NIOEM (NAICS-based)	84	100

IO accounts (BLS sectors)	86	66
PUMS (Census-based)	77	83
Classification in this paper	**64**	**54**

For manufacturing, the Census classification, used in the PUMS, provides the least detailed breakdown; for services, the BLS sector classification, used in the IO accounts, is the least detailed. Aggregating across all four sources results in 64 manufacturing industries and 54 services industries, which is the maximum number of codes such that none is a subset of any industry code in any of the data sources.

TABLES

Table 1: Share of US private, non-farm employment by sector, 2004

Sector	Share	NAICS
Forestry, fishing, hunting, mining, utilities, construction	7%	11, 21-23
Manufacturing	12%	31-33
Trade and transportation	22%	42-49
Information, finance, insurance, and real estate	10%	51-53
Business and personal services	49%	54-81

Source: County Business Patterns

Table 2: Most agglomerated industries: County

Deep sea passenger transportation	.454
Motion picture and video production*	.335
Investment banking and securities dealing*	.282
Women’s cut & sew blouse & shirt mfg	.265
Photographic film, paper, plate, and chem. mfg	.236
Casino hotels*	.205
Teleproduction and other post production svc*	.198
Women’s cut & sew apparel contractors	.194
Payroll services*	.163
Oil & gas field equipment/machinery mfg	.152

Highest Ellison-Glaeser agglomeration values, 6-digit NAICS industries, national employment >=10000

* denotes services industries (NAICS 51-81)

Table 3: Most agglomerated industries: State

Wineries	.448
Deep sea passenger transportation	.437
Oil & gas field equipment/machinery mfg	.403
Carpet and rug mills	.381
Other (non-sheer) hosiery and sock mills	.370
Cigarette manufacturing	.333
Motion picture and video production*	.327
Casino hotels*	.322
Women’s cut & sew blouse & shirt mfg	.300
Yarn spinning mills	.270

Highest Ellison-Glaeser agglomeration values, 6-digit NAICS industries, national employment >=10000

* denotes services industries (NAICS 51-81)

Table 4: Forces, Measures, and Summary Statistics of Agglomeration

Force	**Measure**	**Notation (see appendix for definitions)**	**Mean (mfg)**	**S.D. (mfg)**	**Mean (services)**	**S.D. (services)**
Agglomeration (zip code)	Ellison-Glaeser agglomeration index		.009	.014	.003	.004
Agglomeration (county)	Ellison-Glaeser agglomeration index		.008	.012	.005	.013
Agglomeration (state)	Ellison-Glaeser agglomeration index		.031	.039	.008	.014
Labor pooling	Occupational specialization	$Occspec_i$	.625	.045	.619	.103
Natural resource inputs	Share of inputs from agriculture, forestry, fishing, logging, and mining	$Nature_i$	.088	.175	.005	.018
Knowledge spillovers	Share of workers with graduate degrees	$Grad_i$	.040	.038	.106	.099
N			64	64	54	54

Table 5: Agglomeration in Manufacturing Industries

	Zip code	County	State
Labor pooling	.06937	.07547*	.42025*
	(.04298)	(.03539)	(.10852)
Natural resource inputs	.00750	-.00875	.00056
	(.01019)	(.00839)	(.02573)
Knowledge spillovers	.04041	.05501	.13281
	(.05077)	(.04180)	(.12820)
R-squared	.06	.11	.24
N	64	64	64

National industry employment included as control but not reported
Standard errors in parentheses
* denotes significance at 5% level

Table 6: Agglomeration in Services Industries

	Zip code	County	State
Labor pooling	.00218	.00224	-.00062
	(.00560)	(.01807)	(.02005)
Natural resource inputs	-.00056	-.03152	-.05740
	(.03401)	(.10967)	(.12167)
Knowledge spillovers	.01364*	.00739	.00744
	(.00572)	(.01845)	(.02047)
R-squared	.14	.03	.03
N	54	54	54

National industry employment included as control but not reported
Standard errors in parentheses
* denotes significance at 5% level

Table 7: Forces, Measures, and Summary Statistics of Co-agglomeration

Force	**Measure**	**Notation (see appendix for definitions)**	**Mean (mfg)**	**S.D. (mfg)**	**Mean (servs)**	**S.D. (servs)**
Co-agglomeration (zip code)	Ellison-Glaeser co-agglomeration index		.0012	.0026	-.00002	.0007
Co-agglomeration (county)	Ellison-Glaeser co-agglomeration index		.0006	.0018	.0002	.0020
Co-agglomeration (state)	Ellison-Glaeser co-agglomeration index		.0033	.0114	.0004	.0036
Occupational similarity (un-interacted)	Occupational similarity of industry pair	$Occsim_{ij}$	.522	.144	.263	.131
Labor pooling	Occupational similarity of industry pair **interacted with** occupational specialization of *each* industry	$Occsim_{ij}*(occspec_i+occspec_j)/2$	.326	.089	.158	.073
Knowledge spillovers	Occupational similarity of industry pair **interacted with** worker graduate degrees of *each* industry	$Occsim_{ij}*(grad_i+grad_j)/2$	.021	.014	.029	.026
Amenity demand	Demographic similarity of industry pair	$Demosim_{ij}$	.805	.100	.673	.127
Direct trade	Direct trade between industry pair	$Trade_{ij}$	.006	.015	.008	.014
Similarity of inputs	Input similarity of industry pair	$Inputsim_{ij}$	.455	.122	.535	.101
Similarity of outputs	Output similarity of industry pair	$Outputsim_{ij}$	.302	.222	.449	.266
Effect of information technology on transport costs	Direct trade between industry pair **interacted with** information technology intensity in *each* industry	$Trade_{ij}*(tech_i+tech_j)/2$	.0003	.0011	.0010	.0017
N			2016	2016	1431	1431

Table 8: Co-agglomeration in Manufacturing Industries

	Zip code	County	State	Zip code	County	State
Labor pooling	0.06974**	0.05853**	0.38056**	0.06629**	0.05652**	0.37918**
	(0.01832)	(0.02233)	(0.12590)	(0.01832)	(0.02228)	(0.12812)
Knowledge spillovers	0.09157**	0.03214*	0.01191	0.06618**	0.01732	0.00177
	(0.02072)	(0.01759)	(0.08766)	(0.02025)	(0.01800)	(0.09420)
Amenity demand	0.00507**	0.00299**	0.01431**	0.00496**	0.00292**	0.01426**
	(0.00091)	(0.00076)	(0.00406)	(0.00091)	(0.00076)	(0.00408)
Direct trade	0.00831	0.01055*	0.10587**	-0.00000	0.00570	0.10255**
	(0.00571)	(0.00573)	(0.03545)	(0.00473)	(0.00625)	(0.04165)
Similarity of outputs	-0.00034	-0.00003	0.00028	-0.00031	-0.00001	0.00030
	(0.00027)	(0.00025)	(0.00152)	(0.00027)	(0.00025)	(0.00152)
Similarity of inputs	0.00228**	0.00253**	0.02809**	0.00244**	0.00262**	0.02815**
	(0.00078)	(0.00073)	(0.00585)	(0.00075)	(0.00074)	(0.00586)
IT * direct trade				0.51463**	0.30023**	0.20550
				(0.12484)	(0.10774)	(0.49637)
Observations	2016	2016	2016	2016	2016	2016
R-squared	0.32	0.21	0.28	0.33	0.21	0.28

See Table 7 for variable definitions
Robust standard errors in parentheses
Industry fixed effects for industries i and j included in all specifications
Occupational similarity (uninteracted) also included but not shown in all specifications
** denotes significance at 5% level; * denotes significance at 10% level

Table 9: Co-agglomeration in Manufacturing Industries – Standardized Betas

	Zip code	County	State	Zip code	County	State
Labor pooling	2.43	2.99	3.00	2.31	2.89	2.99
Knowledge spillovers	0.52	0.27	0.02	0.37	0.14	0.00
Amenity demand	0.20	0.17	0.13	0.19	0.17	0.13
Direct trade	0.05	0.09	0.14	0.00	0.05	0.14
Similarity of outputs	-0.03	0.00	0.01	-0.03	0.00	0.01
Similarity of inputs	0.11	0.18	0.30	0.12	0.18	0.30
IT * direct trade				0.11	0.10	0.01

Standardized Betas correspond to results from regressions in Table 8.

Table 10: Co-agglomeration in Services Industries

	Zip code	County	State	Zip code	County	State
Labor pooling	0.00174	-0.00454	-0.00047	0.00154	-0.00450	0.00000
	(0.00239)	(0.00564)	(0.00787)	(0.00240)	(0.00560)	(0.00780)
Knowledge spillovers	0.01234**	0.00746	-0.00276	0.01147**	0.00765	-0.00069
	(0.00246)	(0.00471)	(0.00899)	(0.00240)	(0.00477)	(0.00890)
Amenity demand	0.00018	0.00196**	0.00437**	0.00020	0.00196**	0.00433**
	(0.00019)	(0.00054)	(0.00109)	(0.00019)	(0.00054)	(0.00108)
Direct trade	0.00604**	0.00142	-0.02568**	-0.00003	0.00274	-0.01131
	(0.00233)	(0.00558)	(0.00953)	(0.00277)	(0.00668)	(0.01030)
Similarity of outputs	0.00037**	0.00023	0.00122**	0.00033**	0.00025	0.00133**
	(0.00009)	(0.00018)	(0.00039)	(0.00009)	(0.00018)	(0.00040)
Similarity of inputs	0.00090**	0.00187**	0.00684**	0.00083*	0.00188**	0.00701**
	(0.00043)	(0.00087)	(0.00268)	(0.00043)	(0.00087)	(0.00270)
IT * direct trade				0.15834**	-0.03438	-0.37499**
				(0.06041)	(0.05444)	(0.15011)
Observations	1431	1431	1431	1431	1431	1431
R-squared	0.30	0.18	0.17	0.32	0.18	0.17

See Table 7 for variable definitions
Robust standard errors in parentheses
Industry fixed effects for industries i and j included in all specifications
Occupational similarity (uninteracted) also included but not shown in all specifications
** denotes significance at 5% level; * denotes significance at 10% level

Table 11: Co-agglomeration in Services Industries – Standardized Betas

	Zip code	County	State	Zip code	County	State
Labor pooling	0.18	-0.17	-0.01	0.16	-0.17	0.00
Knowledge spillovers	0.44	0.10	-0.02	0.41	0.10	0.00
Amenity demand	0.03	0.13	0.15	0.04	0.13	0.15
Direct trade	0.12	0.01	-0.10	0.00	0.02	-0.04
Similarity of outputs	0.14	0.03	0.09	0.12	0.03	0.10
Similarity of inputs	0.13	0.10	0.19	0.12	0.10	0.20
IT * direct trade				0.21	-0.02	-0.10

Standardized Betas correspond to results from regressions in Table 10.