

Urbanization, Agglomeration, and Co-Agglomeration of Service Industries

Jed Kolko*

Public Policy Institute of California

Original version: April 2007

This version: January 2008

ABSTRACT

Services industries are more urbanized though less agglomerated than manufacturing industries are. Furthermore, nearly every agglomerated services industry is highly urbanized, whereas many agglomerated manufacturing industries exist outside big cities. Services are distinguished from manufacturing by a lower reliance on natural inputs and, more importantly, different costs of transporting output: some services output can be transported electronically while some must be delivered in-person. While traditional analyses of the micro-foundations of agglomeration have not focused on these differences between manufacturing and services and do not easily incorporate transport costs, analyzing why pairs of industries co-agglomerate reveals how the nature of transport costs contributes to the location patterns we observe among manufacturing and services.

The results show that co-agglomeration between pairs of services industries is driven by 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 and help explain why services are more urbanized yet less agglomerated than manufacturing industries are.

* Thanks to Davin Reed for excellent research help. Thanks for useful comments to Ed Glaeser, Mark Partridge, and to participants at PPIC seminars, the April 2007 Kiel Institute Workshop on Agglomeration and Growth in Knowledge-Based Societies, the November 2007 Regional Science meetings, and the December 2007 NBER workshop on agglomeration.

Services now dominate the US and other advanced economies, most of all in large cities. Employment in most services industries is strikingly urbanized, while manufacturing is the least urbanized sector of the economy, aside from natural-resource-dependent sectors of agricultural support and mining. The economic future of cities depends on services, and understanding why services are in cities is essential to understanding the function of cities in modern economies.

Although services are highly urbanized, they are less concentrated than manufacturing industries are. While there are examples of highly agglomerated services industries – such as investment banking and motion picture production – services are, on average, less agglomerated than manufacturing at the county level and much less so at the state level. Furthermore, nearly all agglomerated services industries are highly urbanized, whereas many manufacturing industries agglomerate in smaller cities or rural areas. The traditional study of agglomeration, which focused on manufacturing, could safely divorce the question of urbanization from agglomeration. To analyze why some services industries are agglomerated, it is important to ask at the same time why services are urbanized.

To assess why services are less agglomerated yet highly urbanized, this chapter looks at co-agglomeration between pairs of industries, which helps reveal the micro-foundations for the observed location patterns in services and manufacturing and the differences between them. Services differ from manufacturing in that services rely much less on physical inputs, freeing services from having to locate near natural resources, and service transactions often happen in-person, making it important for some services industries to be near their customers. Both of these characteristics of services contribute

to their tendency to be in larger cities and to be less agglomerated than manufacturing industries.

The main empirical finding is that services industries that trade with each other are more likely to co-locate in the same zip code, though not in the same county or the same state; in contrast, manufacturing industries that trade with each other are more likely to co-locate in the same county or state but not at the zip-code level. This finding is consistent with the in-person delivery of many services, including consumer services like haircuts and professional services like management consulting. Furthermore, services industries that rely more on information technology are even less likely to co-locate at the state level, which suggests that the Internet substitutes for phone, mail, and travel in services industries though not for in-person interactions. The importance of proximity to trading partners underscores the advantages of an urban location for services industries, especially for services whose customers are other businesses.

Urbanization and Agglomeration Patterns

Services are strikingly urbanized.¹ Ranking sectors by the average metropolitan population where jobs in that sector are located, the most urbanized sectors are all services, including professional, scientific, and technical services; information; and management of companies and enterprises (see Table 2). Services that serve households rather than primarily other businesses, like health care and accommodation and food

¹ In this paper, “services” refers to industries involved in the production and distribution of intangible goods. These include the information sector; finance, insurance, and real estate; professional services; education, health care, leisure, and hospitality; and other personal and business services. Together these are covered by North American Industry Classification System, or NAICS, codes 51-81 and accounted for 59% of private, non-farm employment in 2004. Manufacturing, covered by NAICS codes 31-33, accounted for 12%. The remainder of private, non-farm employment is in the natural resources industries including mining, utilities, construction, wholesale and retail trade, and transportation. See Table 1.

services, are less urbanized, but still more so than the manufacturing sector is.

Aggregating all services sectors, 55% of services jobs are in metro areas with at least two million population, compared with 42% of manufacturing jobs. The urbanization of services is even more apparent when considering the county density of the average job because services are disproportionately clustered in downtowns of metropolitan areas. The county population density of the average services job is 3583; for manufacturing jobs, the average county density is only 1400 (see Table 3).²

While services are more urbanized than manufacturing and other sectors on average, the most urbanized industries are spread across multiple sectors. Among the ten most extremely urbanized industries are four services industries: investment banking, motion picture production, teleproduction & postproduction services, and agents & managers (see Table 4). The rest of the top ten are apparel manufacturers and wholesalers and related industries. In these most urbanized industries, the metro area population of the average job is over 12 million, implying that these industries are all highly concentrated in New York and Los Angeles.

Services are less agglomerated than manufacturing at the county level and much less so at the state level.³ The average Ellison-Glaeser county-level agglomeration index for services industries at the 6-digit NAICS level, weighted by national employment, is .0068, compared with the average for manufacturing of .0132. The difference is even larger for state-level agglomeration: the average index for services is .0135, and for

² Employment data are from County Business Patterns (CBP) and the National Establishment Time-Series (NETS) database. The NETS was available for California only and is used for calculating zip-code-level agglomeration and co-agglomeration; CBP is available nationally and is used for all other analyses in the paper. Detailed on the datasets are given in the appendix.

³ Following earlier work on agglomeration, this paper uses the Ellison-Glaeser (1997) index of agglomeration, which 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 formula is given in the appendix.

manufacturing is .0406. 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.⁴

Still, some services industries are highly agglomerated, and services account for five of the ten most agglomerated industries at the county level (see Table 5). 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 6). 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 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).

A pronounced difference between manufacturing and services is that agglomeration and urbanization are more correlated for services than for manufacturing.

⁴ 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).

The correlation between average metro size and the county-level agglomeration index, weighted by industry employment, is .48 for services and .26 for manufacturing; using the state-level agglomeration index, the correlations are .29 for services and -.08 for manufacturing (see Table 3). Nearly all of the agglomerated services industries are highly urbanized, the only exception being casino hotels, where employment is concentrated in Las Vegas rather than in the largest metropolitan areas (see Figure). In contrast, many manufacturing industries are highly agglomerated, even at the county level, without being highly urbanized, such as the photographic equipment manufacturing industry, based in mid-sized Rochester, NY.

Micro-foundations of Agglomeration and Urbanization

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. 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), 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), and a study, similar to this paper, that uses co-agglomeration to assess the micro-foundations of agglomeration (Ellison, Glaeser, and Kerr 2007). 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).⁵

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 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

⁵ 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).

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.⁶

⁶ 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)

To explain the variation in agglomeration levels across industries, the analysis uses four measures at the industry level: occupational specialization, natural resource inputs, workers with graduate degrees, and share of output going to consumers. This section follows Rosenthal and Strange (2001) and others in using industry-level measures for agglomeration forces as explanations for observed agglomeration. These four measures are also used to explain urbanization. The intuitive meaning of these measures is outlined in the following paragraphs. The formulas and data sources for these measures are summarized in Table 7 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' (BLS's) National Industry-Occupation Employment Matrix (NIOEM).⁷

show that, theoretically, electronic and face-to-face communications can instead be complements rather than substitutes. Kolko (2000b) and Sinai and Waldfoegel (2004) offer empirical evidence that the Internet both substitutes and complements for non-electronic communications, depending on the nature of the communications.

⁷ The summary statistics in Table 7 reveal no difference between services industries and manufacturing industries in their levels of occupational specialization.

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.⁸ The 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).⁹

The fourth measure is the share of industry output going directly to consumers, as opposed to other businesses or government. Although the agglomeration literature has not suggested that the share of output going to consumers should affect industry agglomeration, the finding that business services are generally more urbanized than consumer services raises the possibility that the nature of industry output could affect industry location decisions. This measure comes from the BEA's IO accounts.

⁸ 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 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.

⁹ 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.

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.¹⁰

The index of agglomeration is regressed on these four 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 inputs, workers with graduate degrees, and share of output going to consumers, along with national industry employment as a control; and k refers to zip codes, counties, or states, depending on the specification.¹¹ Agglomeration is measured at the zip-code level in California and at the county and state levels for the US. The four measures that proxy for agglomeration factors – occupational specialization, natural resource inputs, graduate degrees, and share of output going to consumers – 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 specification is then repeated with urbanization in place

¹⁰ See the appendix for more detail on the industry classification.

¹¹ This set-up follows Rosenthal and Strange (2001).

of agglomeration, with the average metropolitan population where jobs in that industry are located as the dependent variable.

The purpose of this part of the analysis is two-fold: first, to see what, if anything, explains agglomeration and urbanization 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 8 and 9, 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.¹² 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.¹³ That the relationship between occupational specialization and agglomeration is strongest at the state level is consistent with interpreting occupational 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.

¹² 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.

¹³ 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>.

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 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.

Turning to urbanization, for manufacturing industries, the reliance on natural resource inputs is negatively correlated with urbanization. While the coefficient on natural resource inputs is statistically insignificant for urbanization in services industries, this is probably due less to a qualitatively different relationship between natural resources and urbanization in the services sector and due more to the very low mean and variance of natural resource inputs within services (the standard error on the coefficient is quite high). Share of workers with graduate degrees is positively correlated with urbanization for both manufacturing and services, though not quite at the 5% level for services. The important difference in explaining urbanization across the sectors, however, is the share of output going to consumers, which is positively correlated with urbanization for manufacturing industries and negatively correlated with urbanization for services industries (and statistically significant for both sectors). This difference between manufacturing and services suggests that the cost of transporting output to customers could affect location decisions differently for manufacturing and for services, which will be explored in greater depth in the co-agglomeration analysis.

Transport Costs in Manufacturing and Services

Comparing the factors affecting location patterns in manufacturing and services, two differences between the sectors stand out: (1) occupational specialization contributes to agglomeration only for manufacturing industries, not services, and (2) the fraction of output going to consumers encourages urbanization for manufacturing industries and discourages urbanization for services industries. This section outlines a theory why these differences might arise.

First, why does occupational specialization not contribute to agglomeration in services? To the extent that occupational specialization is a proxy for the benefits of labor pooling, industries that are urbanized have less need to agglomerate to take advantage of labor pooling. The theory of labor pooling posits that workers will require higher wages if there are fewer local employment opportunities outside their firm, whereas agglomeration from firm-specific shocks (though not industry-wide shocks). If being in a large labor market also protects workers from shocks, then the benefits of labor pooling could be weakened for urbanized industries, which tend to be services.

Second, why should the type of customer – consumers or businesses – affect manufacturing and service location decisions differently? This striking difference between the sectors warrants a closer examination of the costs each sector faces in transporting output, and how the nature of output and customers affects location decisions. In fact, the question of transport costs goes to the essence of how manufacturing and services differ, since manufacturing by definition produces tangible output and services produce intangible output. What follows is a simple yet plausible

model of transport costs for tangible and intangible goods and – by implication -- for manufacturing and services industries.

For tangible goods, transport costs rise 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.¹⁴ 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.¹⁵ 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 likely to be sensitive to proximity to trading partners, and therefore would exhibit agglomeration at larger geographic levels.¹⁶

For intangibles, transport costs are quite different. Consider a service that must be consumed in person, like a haircut or a face-to-face legal discussion, where what is

¹⁴ 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.

¹⁵ 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.

¹⁶ 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.

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.¹⁷ 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.¹⁸ 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 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

¹⁷ 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 its value.

¹⁸ 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, and 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.

being 2000 miles away from customer is relatively small – at least relative to manufacturing.¹⁹ 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 cannot.²⁰ 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 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

¹⁹ 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.

²⁰ 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.

industries of very close proximity to customers as much as they reduce the benefits of longer-distance proximity to customers.²¹

What Co-Agglomeration Reveals about Transport Costs

To assess how important trading relationships are for industry location decisions, this section analyzes whether industries that trade with each other also agglomerate together, and whether trading affects co-agglomeration differently for services than for manufacturing. In addition to examining the effect of trading relationships between industries, 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.²² One can characterize the degree of similarity of multiple variables between firms in different industries on a continuous scale, so co-agglomeration allows for more refined testing of the earlier finding that occupational specialization affects manufacturing location decisions but not those of services.

To measure co-agglomeration, this paper uses the extension of the Ellison-Glaeser (1997) index to co-agglomeration.²³ Their co-agglomeration index measures the extent to

²¹ 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.

²² 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).

²³ There has been very little research on industries locating near each other. Three 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 increasingly important, rather than industrial specialization, which implies greater linkages between, rather than within, industries. Most recently, Ellison, Glaeser, and Kerr (2007) analyze the reasons for agglomeration, including direct trading relationships, through the lens of co-agglomeration using a framework closely related to the one in this paper. Their study, however, only includes manufacturing

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.²⁴

The empirical strategy for measuring co-agglomeration is:

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

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

industries and therefore does not address the question of whether the different nature of transport costs in service industries affects services' location patterns.

²⁴ 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.

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 are discussed below.

The main variable of interest is direct trade, measured using the volume of direct trade between industries i and j as a share of overall inputs and outputs of both industries. The expectation is that trading intensity should contribute to services co-agglomeration for smaller geographies and to manufacturing co-agglomeration only for larger geographies. 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 with the direct trade measure. It is further 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.²⁵

²⁵ 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.

The other variable of interest is the occupational similarity between the industry pair. Industries with similar workers could co-agglomerate to for labor pooling if they have workers in occupations facing thin labor markets; industries with similar workers could co-agglomerate also because they benefit from knowledge spillovers between similarly skilled workers. To attempt to separate out these factors, and to isolate whether labor pooling specifically indeed applies uniquely to manufacturing, occupational similarity enters the model twice: once interacted with the average of the occupational specialization of each industry (to reflect labor pooling) and once interacted with the average of the share of workers with graduate degrees in each industry (to reflect knowledge spillovers).²⁶

In addition, three further measures are included as controls. 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-

²⁶ 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. Ellison, Glaeser, and Kerr (2007) use patent citation data and Scherer technology flows to proxy for knowledge spillovers between pairs of industries, rather than any workforce measures. Both patent data and technology flows are available in greater detail for manufacturing than for services industries: for instance, the U.S. Patent and Trademark Office's correspondence between patent codes and industry codes combines all non-manufacturing SIC's into a single category. Analyzing both manufacturing and services, therefore, restricts the set of usable data sources to those that are meaningful and disaggregated for both sectors, such as the industry-occupation matrix.

agglomerate because their workers consider the same locations to be high-amenity.²⁷ The other two controls, the similarity of the industries' inputs and the similarity of the industries' outputs, capture whether the two industries in the pair have similar suppliers and customers. Table 10 summarizes all of these measures and their interpretations, and the appendix defines them in detail.

The results of the co-agglomeration analysis for manufacturing are presented in Table 11 and for services in Table 12. The regressions include all of the measures described above.²⁸ 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 most notable difference between manufacturing and services is the effect of trading intensity. For manufacturing, 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

²⁷ 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. Furthermore, omitting the demographic similarity variable has essentially no effect on the coefficient estimates for the labor pooling variable.

²⁸ 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.

and not significant for zip code co-agglomeration. For services, however, direct trading relationship contributes to the co-agglomeration of services at the zip code level, and is negatively and significantly related to co-agglomeration at the state level. This difference between services and manufacturing is consistent with the simple models of transport costs sketched above, which predicted that direct trade would lead services to co-agglomerate only at small geographies and manufacturing only over larger geographies.²⁹

The interaction between trading intensity and information-technology usage is more complicated. For manufacturing, the interaction between trade and information 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.

For services, the interaction between direct trade and information technology intensity is also 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 less likely to locate in the same state if the industries rely more on information technology. The fact that the interaction coefficient is

²⁹ 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 suggest that it could be negative; the negative sign on this coefficient is surprising and remains unexplained.

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.

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.³⁰

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

³⁰ 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 11, column 5) remains positive but is no longer statistically significant.

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.³¹

Turning to the other variable of interest: occupational similarity interacted with specialization contributes to manufacturing co-agglomeration at all levels of geography (Table 11) and services co-agglomeration at none (Table 12). There is, however, little difference between manufacturing and services in the effect on co-agglomeration of occupational similarity interacted with worker graduate degrees. For both sectors, this interaction term is positive and statistically significant at the 5% level only for zip-code co-agglomeration, and the standardized betas are similar for the two sectors (Tables 13 and 14). These findings reinforce the initial finding that the labor pooling explanation applies specifically to manufacturing, though there are reasons other than labor pooling that encourage industries with similar occupational mixes to co-agglomerate.

Conclusions

These findings on urbanization, agglomeration, and co-agglomeration reveal why services are more urbanized yet less agglomerated than manufacturing. First, transport costs for services output encourage services to locate near their customers. This acts as a force against industry-level agglomeration. Further, because services industries often have business customers across diverse industries, it is optimal for many services to

³¹ These conclusions about services industries are based on regressions that include all services industries. It is possible that many consumer-facing services simply locate where consumers are without regard to agglomerative forces; Table 15 therefore repeats the analysis in Table 12, excluding industries in which 95% or more of output goes to consumers. The results in Table 15 are nearly identical to those in Table 12, with the exception that similarity of inputs and similarity of outputs are no longer significant for any level of geography in Table 15 (columns 1-3), knowledge spillovers become statistically significant at the county level, and direct trade become statistically insignificant at the state level. Also, the interaction between IT and direct trade becomes statistically significant only at the 10% level at the zip code level.

locate within a dense, diverse set of businesses, which explains not only the tendency for services to urbanize but also for services to be in the denser portions of urban areas. Second, services industries rely less on natural resources than manufacturing industries do, which allows services to urbanize to a greater extent than manufacturing. Third, although the level of occupational specialization is as high in the average services industry as in the average manufacturing industry, occupational specialization does not lead to agglomeration for services as it does for manufacturing, suggesting that labor pooling does not affect services industries' location decisions as it does manufacturing. In fact, the tendency for services to urbanize due to transport costs and non-reliance on natural resources may itself make labor pooling less important since urbanized services industries are already in thick labor markets even in the absence of industry agglomeration.

These results also suggest that the increasing reliance on information technology could continue to change business location decisions, but in a way that could favor cities. 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. As information technology becomes further integrated into business processes, the benefit for services of proximity in dense areas increases even though IT enables some types of output to be transported more cheaply – good news for the future of cities.

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.
- Ellison, Glenn, and Edward Glaeser, "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," Journal of Political Economy, Vol. 105, No. 5, pp. 889-927, 1997.
- Ellison, Glenn, Edward Glaeser, and William Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," NBER Working Paper #13068, 2007.
- 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 Waldfoegel, "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):

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

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

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

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.

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

$H = \sum_j w_j^2 H_j$ (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:

$$outputsim_{ij} = \frac{\left(2 - \sum_k \left| \frac{b_{i \rightarrow k}}{output_i} - \frac{b_{j \rightarrow k}}{output_j} \right| \right)}{2}, \text{ which equals 1 if industries i and j}$$

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:

$$inputsim_{ij} = \frac{\left(2 - \sum_k \left| \frac{b_{k \rightarrow i}}{input_i} - \frac{b_{k \rightarrow j}}{input_j} \right| \right)}{2}, \text{ which equals 1 if industries i and j have}$$

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 :

$$occsim_{ij} = \frac{\left(2 - \sum_k |occ_{ik} - occ_{jk}|\right)}{2}, \text{ which equals 1 if industries } i \text{ and } j$$

have perfectly overlapping distributions of occupations and 0 if they have non-overlapping distributions of occupations.

The **occupational specialization** variable, $occspec_i$, 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 :

$$occspec_i = \frac{\left(2 - \sum_k |occ_{ik} - occ_k|\right)}{2}, \text{ which equals 1 if industry } i \text{ has a}$$

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.

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 :

$$demosim_{ij} = \frac{\left(2 - \sum_k |demo_{ik} - demo_{jk}|\right)}{2}, \text{ which equals 1 if industries } i \text{ and } j \text{ have perfectly overlapping distributions of age-education cells and 0 if they have non-overlapping distributions of age-education cells.}$$

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: Urbanization, All Sectors (NAICS 2-Digit)

SECTOR	MEAN METRO SIZE
Professional, Scientific, and Technical Services	6674642
Information	6445303
Management of Companies and Enterprises	6212818
Educational Services	6195770
Wholesale Trade	6038354
Real Estate and Rental and Leasing	6005790
Finance and Insurance	5955742
Administrative and Support and Waste Management and Remediation Services	5700440
Arts, Entertainment, and Recreation	5456450
Transportation and Warehousing	5176784
Other Services (except Public Administration)	5019998
Health Care and Social Assistance	4889552
Construction	4704008
Retail Trade	4572272
Accommodation and Food Services	4359364
Utilities	4278660
Manufacturing	4049816
Forestry, Fishing, Hunting, and Agriculture Support	1697923
Mining	1591765

Note: Average metro size refers to the mean population of the Combined Statistical Area, Core Based Statistical Area (if not in a CSA), or county (if not in a CBSA), averaged over all jobs in the sector.

Table 3: Urbanization and Agglomeration (NAICS 6-digit Industries)

	All	Manufacturing	Services
Average metro size	5,113,050	4,046,245	5,450,653
% in metros over 2 million	.52	.42	.55
Average county population density	2902	1400	3583
County-level agglomeration	.0066	.0132	.0068
State-level agglomeration	.0154	.0406	.0135
Correlation between average metro size and county-level agglomeration	.3487	.2581	.4830
Correlation between average metro size and state-level agglomeration	.0776	-.0773	.2938
N	1084	471	334

Note: weighted by 6-digit NAICS industry employment.

Table 4: Most Urbanized Industries (NAICS 6-Digit)

SECTOR	MEAN METRO SIZE
Investment Banking and Securities Dealing	16545537
Women's and Girls' Cut and Sew Dress Manufacturing	15500985
Women's, Children's, and Infants' Clothing and Accessories Merchant Wholesalers	14950006
Women's and Girls' Cut and Sew Blouse and Shirt Manufacturing	14907902
Women's, Girls', and Infants' Cut and Sew Apparel Contractors	14607557
Motion Picture and Video Production	14303119
Teleproduction and Other Postproduction Services	13652323
Agents and Managers for Artists, Athletes, Entertainers, and Other Public Figures	13503575
Jewelry, Watch, Precious Stone, and Precious Metal Merchant Wholesalers	12336946
Piece Goods, Notions, and Other Dry Goods Merchant Wholesalers	12304385

Note: Average metro size refers to the mean population of the Combined Statistical Area, Core Based Statistical Area (if not in a CSA), or county (if not in a CBSA), averaged over all jobs in the industry. Only industries with at least 10,000 national employment included.

Table 5: Most agglomerated industries: County

Industry	EG Index
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 6: Most agglomerated industries: State

Industry	EG Index
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)

Figure

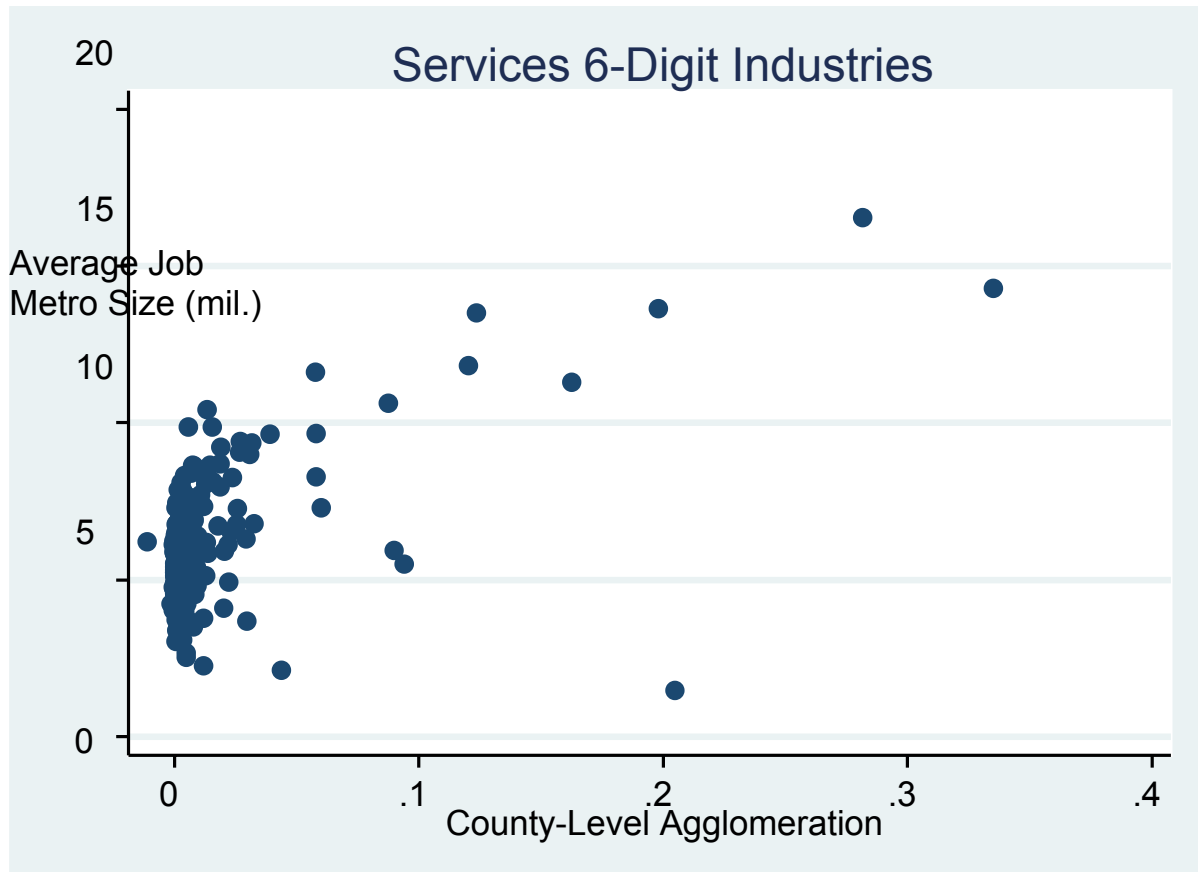


Table 7: Measures, and Summary Statistics of Agglomeration

Measure	Mean (mfg)	S.D. (mfg)	Mean (services)	S.D. (services)
Zipcode agglomeration (EG index)	.009	.014	.003	.004
County agglomeration (EG index)	.008	.012	.005	.013
State agglomeration (EG index)	.031	.039	.008	.014
Occupational specialization	.625	.045	.619	.103
% natural resources inputs (agriculture, forestry, fishing, logging, and mining)	.088	.175	.005	.018
Share of workers with graduate degrees	.040	.038	.106	.099
Share of output going to consumers	.354	.336	.518	.370
N	64	64	54	54

Table 8: Agglomeration and Urbanization in Manufacturing Industries

	Zip code agglomeration	County agglomeration	State agglomeration	Average metro size
Occupational specialization	0.06535	0.07310**	0.41916**	-5.53399
	(0.04265)	(0.03543)	(0.10965)	(4.93500)
Share natural resource inputs	0.00762	-0.00868	0.00059	-2.97760**
	(0.01009)	(0.00838)	(0.02595)	(1.16769)
Share of workers with grad degrees	0.03481	0.05172	0.13130	15.25900**
	(0.05042)	(0.04189)	(0.12965)	(5.83486)
Fraction of output to consumers	0.00764	0.00450	0.00206	1.31775**
	(0.00520)	(0.00432)	(0.01337)	(0.60169)
R-squared	0.09	0.13	0.24	0.33
N	64	64	64	64

National industry employment included as control but not reported

Standard errors in parentheses

** significance at 5% level; * significance at 10% level

Table 9: Agglomeration and Urbanization in Services Industries

	Zip code agglomeration	County agglomeration	State agglomeration	Average metro size
Occupational specialization	0.00039	0.00639	0.00423	1.10561
	(0.00608)	(0.01965)	(0.02179)	(1.81134)
Share natural resource inputs	0.00538	-0.04528	-0.07347	-4.40207
	(0.03498)	(0.11314)	(0.12548)	(10.42858)
Share of workers with grad degrees	0.01374**	0.00716	0.00718	3.38075*
	(0.00575)	(0.01859)	(0.02061)	(1.71309)
Fraction of output to consumers	0.00129	-0.00298	-0.00348	-1.27118**
	(0.00165)	(0.00532)	(0.00590)	(0.49056)
R-squared	0.15	0.04	0.03	0.24
N	54	54	54	54

National industry employment included as control but not reported

Standard errors in parentheses

** significance at 5% level; * significance at 10% level

Table 10: Forces, Measures, and Summary Statistics of Co-agglomeration

Measure	Notation (see appendix for definitions)	Mean (mfg)	S.D. (mfg)	Mean (servs)	S.D. (servs)
Zipcode co-agglomeration (EG index)		.0012	.0026	-.00002	.0007
County co-agglomeration (EG index)		.0006	.0018	.0002	.0020
State co-agglomeration (EG index)		.0033	.0114	.0004	.0036
Occupational similarity of industry pair, un-interacted	$Occsim_{ij}$.522	.144	.263	.131
Occupational similarity of industry pair interacted with occupational specialization of <i>each</i> industry	$Occsim_{ij}*(occspec_i+occspec_j)/2$.326	.089	.158	.073
Occupational similarity of industry pair interacted with worker graduate degrees of <i>each</i> industry	$Occsim_{ij}*(grad_i+grad_j)/2$.021	.014	.029	.026
Demographic similarity of industry pair	$Demosim_{ij}$.805	.100	.673	.127
Direct trade between industry pair	$Trade_{ij}$.006	.015	.008	.014
Input similarity of industry pair	$Inputs_{ij}$.455	.122	.535	.101
Output similarity of industry pair	$Outputs_{ij}$.302	.222	.449	.266
Direct trade between industry pair interacted with information technology intensity in <i>each</i> industry	$Trade_{ij}*(tech_i+tech_j)/2$.0003	.0011	.0010	.0017
N		2016	2016	1431	1431

Table 11: Co-agglomeration in Manufacturing Industries

	Zip code	County	State	Zip code	County	State
Direct trade	0.00831 (0.00571)	0.01055* (0.00573)	0.10587** (0.03545)	-0.00000 (0.00473)	0.00570 (0.00625)	0.10255** (0.04165)
Occupational similarity * specialization	0.06974** (0.01832)	0.05853** (0.02233)	0.38056** (0.12590)	0.06629** (0.01832)	0.05652** (0.02228)	0.37918** (0.12812)
Occupational similarity * graduate degrees	0.09157** (0.02072)	0.03214* (0.01759)	0.01191 (0.08766)	0.06618** (0.02025)	0.01732 (0.01800)	0.00177 (0.09420)
Demographic similarity	0.00507** (0.00091)	0.00299** (0.00076)	0.01431** (0.00406)	0.00496** (0.00091)	0.00292** (0.00076)	0.01426** (0.00408)
Similarity of outputs	-0.00034 (0.00027)	-0.00003 (0.00025)	0.00028 (0.00152)	-0.00031 (0.00027)	-0.00001 (0.00025)	0.00030 (0.00152)
Similarity of inputs	0.00228** (0.00078)	0.00253** (0.00073)	0.02809** (0.00585)	0.00244** (0.00075)	0.00262** (0.00074)	0.02815** (0.00586)
IT * direct trade				0.51463** (0.12484)	0.30023** (0.10774)	0.20550 (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 10 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 12: Co-agglomeration in Services Industries

	Zip code	County	State	Zip code	County	State
Direct trade	0.00604** (0.00233)	0.00142 (0.00558)	-0.02568** (0.00953)	-0.00003 (0.00277)	0.00274 (0.00668)	-0.01131 (0.01030)
Occupational similarity * specialization	0.00174 (0.00239)	-0.00454 (0.00564)	-0.00047 (0.00787)	0.00154 (0.00240)	-0.00450 (0.00560)	0.00000 (0.00780)
Occupational similarity * graduate degrees	0.01234** (0.00246)	0.00746 (0.00471)	-0.00276 (0.00899)	0.01147** (0.00240)	0.00765 (0.00477)	-0.00069 (0.00890)
Demographic similarity	0.00018 (0.00019)	0.00196** (0.00054)	0.00437** (0.00109)	0.00020 (0.00019)	0.00196** (0.00054)	0.00433** (0.00108)
Similarity of outputs	0.00037** (0.00009)	0.00023 (0.00018)	0.00122** (0.00039)	0.00033** (0.00009)	0.00025 (0.00018)	0.00133** (0.00040)
Similarity of inputs	0.00090** (0.00043)	0.00187** (0.00087)	0.00684** (0.00268)	0.00083* (0.00043)	0.00188** (0.00087)	0.00701** (0.00270)
IT * direct trade				0.15834** (0.06041)	-0.03438 (0.05444)	-0.37499** (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 10 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 13: Co-agglomeration in Manufacturing Industries – Standardized Betas

	Zip code	County	State	Zip code	County	State
Direct trade	0.05	0.09	0.14	0.00	0.05	0.14
Occupational similarity * specialization	2.43	2.99	3.00	2.31	2.89	2.99
Occupational similarity * graduate degrees	0.52	0.27	0.02	0.37	0.14	0.00
Demographic similarity	0.20	0.17	0.13	0.19	0.17	0.13
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 11.

Table 14: Co-agglomeration in Services Industries – Standardized Betas

	Zip code	County	State	Zip code	County	State
Direct trade	0.12	0.01	-0.10	0.00	0.02	-0.04
Occupational similarity * specialization	0.18	-0.17	-0.01	0.16	-0.17	0.00
Occupational similarity * graduate degrees	0.44	0.10	-0.02	0.41	0.10	0.00
Demographic similarity	0.03	0.13	0.15	0.04	0.13	0.15
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 12.

Table 15: Co-agglomeration in Services Industries, excluding Consumer Services

	Zip code	County	State	Zip code	County	State
Direct trade	0.00739** (0.00266)	0.00337 (0.00629)	-0.01389 (0.01243)	0.00257 (0.00345)	0.00514 (0.00794)	0.01107 (0.01340)
Occupational similarity * specialization	0.00599* (0.00347)	-0.00653 (0.00850)	-0.01918* (0.01063)	0.00566 (0.00350)	-0.00641 (0.00844)	-0.01747* (0.01056)
Occupational similarity * graduate degrees	0.01081** (0.00320)	0.01510** (0.00723)	0.01484 (0.01212)	0.01064** (0.00320)	0.01517** (0.00724)	0.01573 (0.01202)
Demographic similarity	0.00031 (0.00027)	0.00197** (0.00092)	0.00410** (0.00155)	0.00032 (0.00027)	0.00197** (0.00092)	0.00407** (0.00155)
Similarity of outputs	0.00025* (0.00014)	-0.00034 (0.00047)	-0.00018 (0.00087)	0.00024* (0.00014)	-0.00034 (0.00047)	-0.00013 (0.00087)
Similarity of inputs	0.00047 (0.00059)	0.00073 (0.00117)	0.00392 (0.00340)	0.00044 (0.00059)	0.00074 (0.00117)	0.00407 (0.00340)
IT * direct trade				0.10656* (0.06419)	-0.03919 (0.07443)	-0.55236** (0.18933)
Observations	820	820	820	820	820	820
R-squared	0.28	0.23	0.28	0.29	0.23	0.29

See Table 10 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

Excludes industries in which 95% or more of output goes to consumers: elementary and secondary schools, health care, museums and historical sites, personal care services, death care services, and religious organizations

** denotes significance at 5% level; * denotes significance at 10% level