HARVARD | BUSINESS | SCHOOL

Location Strategies for Agglomeration Economies

Juan Alcácer
Wilbur Chung

Working Paper

10-071

LOCATION STRATEGIES FOR AGGLOMERATION ECONOMIES

Juan Alcácer
Harvard Business School

Wilbur Chung
R.H. Smith School of Business
University of Maryland

November 2009

Abstract:

Geographically concentrated industry activity creates pools of skilled labor and specialized suppliers, and increases opportunities for knowledge spillovers. The strategic value of these agglomeration economies may vary by firm, depending upon the relative value of each economy, and upon firm and agglomeration economy traits. To better determine when a firm will be attracted to agglomeration economies, we develop a three-layer framework. The first layer assesses the relative importance of skilled labor, suppliers, and knowledge spillovers. The second layer considers whether firms can benefit from geographic concentration without co-locating. The final layer examines why some firms are more inclined to co-locate than others based upon firm and agglomeration economy traits. We test our framework on the U.S. location choices of new manufacturing entrants between 1985 and 1994 and find that firms are far more attracted to skilled labor and specialized suppliers than they are to potential knowledge spillovers, even in R&D intensive industries. We also find that leading firms will be more attracted to pools of labor, suppliers, and potential knowledge spillovers when their own contributions are less fungible, and cannot be easily leveraged for strategic advantage by proximate competitors.

JEL Classifications: R30, R12, L21

Keywords: agglomeration economies, location choice, firm strategy

---

We would like to thank audiences at Duke and the University of Illinois for helpful comments and discussion. Correspondence to jalcacer@hbs.edu or wchung@rhsmith.umd.edu, 3417 Van Munching Hall, College Park, MD 20742 301-405-8967.

1. Introduction

Marshall (1920) suggests that locations thick with similar activity generate valuable agglomeration economies for firms, namely better access to skilled labor (labor market pooling), specialized suppliers (shared inputs), and knowledge spillover from competing firms. As a result, firms' location choices may create competitive advantage by improving access to key resources. More recently, business research has begun exploring why these agglomeration economies are more valuable to some firms than to others. Notably, Shaver and Flyer (2000) argue that large firms may be less inclined to agglomerate because their presence would dramatically increase local economic activity, thereby reducing costs for neighboring competitors. Addressing potential knowledge spillovers specifically, Alcácer and Chung (2007) argue that the cost of knowledge lost to competitors depends upon whether competitors can absorb and use that knowledge. When competitors cannot leverage the knowledge gleaned from technically advanced firms, industry leaders are free to enjoy the benefits of agglomeration without the attendant risk.

In order to more fully understand the firm-level implications of agglomeration economies, and better predict when a particular firm will co-locate, we develop a three-layer framework of determinants. The first layer prioritizes the relative value of labor pools, specialized suppliers and knowledge inflows using the well-developed economics literature on production functions. The second layer assesses whether firms can take advantage of agglomeration economies *without* co-locating with firms that create those economies. The final layer examines differences between firms and among the three agglomeration economies to see if firm or agglomeration traits change the likelihood of aiding neighboring competitors. Larger firms or technical leader firms, for example, might be more likely to contribute resources to competitors through agglomeration economies, giving them less incentive to join a geographic concentration in which their firm's contribution would be particularly valuable to others. At the same time, factor inputs created by each agglomeration economy – the intermediate goods created by pools of specialized suppliers, for example – may be easier or harder for firms to access depending upon the extent of market imperfection. We argue that some factor inputs are more difficult to obtain through market transactions because of market imperfections, making it less likely that competing firms can leverage them for advantage.

We introduce measures for each agglomeration economy – share of industry employment, supplier-industry activity, and industry patent stock – which parallel recent work in economics.[1] For key dimensions of firm heterogeneity, we pair each agglomeration economy with a corresponding firm trait: the contribution to pools of skilled labor and specialized suppliers will change with firms' relative economic output; the contribution to knowledge spillovers will change with firms' technical capabilities. For agglomeration economy heterogeneity, we expect that factor markets for knowledge spillovers will be the most imperfect, followed by specialized suppliers and then skilled labor. Accounting for both firm and agglomeration economy heterogeneity allows us to predict when various firms might be more or less attracted to a particular agglomeration economy.

We test our hypotheses with a sample of first-time foreign entrants[2] to U.S. manufacturing industries between 1985 and 1994.[3] We find that industry employment and

[1] For example, Rosenthal and Strange (2001) examine differences in extent of agglomeration in U.S. manufacturing using levels of educational attainment, manufactured inputs as share of output value, and innovations per value of output. Ellison, Glaeser, and Kerr (2007) examine co-location of U.S. manufacturing industries using similarity in employment category arrays, extent of input-output links, and extent of patent citation links.
[2] We focus on first-time entrants because incumbent firms have prior investments that may affect subsequent location choices and create dependence among observations by the same firm.
[3] By looking at location choice, we evaluate how much value firms expect to gain by agglomerating (or not). The actual value might be evaluated after a move by examining performance measures, such as productivity. Examining

supplier activity are about ten times more attractive than industry patent stocks, suggesting that, on average, firms are relatively unconcerned about gaining knowledge from nearby competitors (knowledge inflows). Similarly, we find that large firms are more wary of co-locating in the presence of skilled labor pools than in the presence of specialized suppliers – presumably because competing firms can hire from pools of skilled labor more readily than they can leverage the supplier base developed by other firms.

In the next section we consider firms' strategic reactions to agglomeration economies in three stages. First, we establish the baseline attractiveness of each agglomeration economy. Second, we argue that some agglomeration economies are more location-specific, providing stronger incentive to co-locate. Third, we propose that firm and agglomeration economy heterogeneity will impact the propensity to co-locate. In later sections we explain our data, methods, and results. Finally, we highlight our findings and discuss implications for firm strategy.

2. Firm asymmetry for agglomeration economies

Marshall (1920) suggests three factors that reduce production costs for agglomerated firms: larger pools of skilled labor, more specialized suppliers, and knowledge inflows from competitors. The first two develop when increased local demand encourages specialization. Unskilled workers are more likely to seek specialized training when local jobs are concentrated in the same industry. In turn, firms save time and money they would otherwise spend on training. By the same logic, suppliers working alongside a concentration of same-industry firms are more likely to make industry-specific investments, reducing transportation and coordination costs for the firms they supply. Marshall's third agglomeration economy considers how firms combine

location choices is crucial because performance is endogenous, in that actual performance will be partially driven by initial location choices.

inputs. Firms with similar products may have very different production processes, varying by capital and labor intensity, by how capital is paired with labor, and by how labor is organized and trained, among other things. Some of these processes will be more productive than others, and over time these best practices may spread to neighboring firms, improving efficiency and lowering costs.

As a source of reduced costs, Marshall's agglomeration economies have important implications for firm strategy: firms that benefit from agglomeration economies will have an advantage over competitors. Whether a firm does, in fact, benefit from an agglomeration economy will depend upon three issues (see Figure 1). First, skilled labor, specialized suppliers, and knowledge spillover have different values to different firms. Second, the level of localization required to access a specific agglomeration economy will vary. Third, the value of a particular agglomeration economy will most likely vary depending upon both firm and agglomeration economy heterogeneity.

2.1 Are some agglomeration economies more intensively used?

Before exploring firms' strategic response to each of Marshall's agglomeration economies, we must determine which of them firms value most. Because trained labor, specialized suppliers, and knowledge inflows can each lower firms' costs, we assume that, on average, firms will be drawn to all three. We expect the attractiveness of each economy will vary by its corresponding importance to a firm's production processes: the more important the input for creation of the final output, the more valuable the corresponding agglomeration economy.

To assess labor, suppliers, and knowledge as inputs, we use the empirical research on firms' production functions as a guide. Fundamental work by Griliches and Mairesse (1983) and Hall (1993) link the set of factor inputs consumed by firms to the output produced. Production functions typically include measures of capital, labor, materials, and R&D. These authors have estimated the cost share, or weights, of these factor inputs, which reflect their relative contribution to output. While there is certainly variation by industry, on average the estimates of these weights are largest for materials, followed by capital and labor, with R&D last. For example, when looking across all U.S. manufacturing from 1974-1990, Nadiri and Kim (1996) report cost shares for capital, labor, materials, and R&D of 13.5%, 14.1%, 64.2%, and 8.2%, respectively. These weights suggest that materials would be most important for firms' productive processes, followed by capital and labor, and then by R&D.

These production function weights give us a baseline from which we can determine the relative attractiveness of various agglomeration economies. Assuming that the more important the input, the more important the corresponding agglomeration economy, we would initially expect firms to value specialized suppliers (the source of materials) most, followed by skilled labor, and finally, R&D activity.

2.2 Do firms need to co-locate to access agglomeration benefits?

The next step is considering how the relative importance of each agglomeration economy will affect firms' location choices. Initially, we would expect the relative draw of each economy to correspond to the value firms place on that economy's factor input – the corresponding production function weights. This baseline would be affected by the degree to which each economy is location specific. Where geographic proximity is necessary to leverage advantage from an economy's factor inputs, firms would have stronger incentive to co-locate.

Consider materials, for example. While materials have a large production weight, indicating their importance as a factor input, firms can source materials from a distance. The 1993 Commodity Flow Survey from the Bureau of Transportation Statistics shows that the distance travelled for the average of all commodities was 424 miles (with this distance growing in subsequent surveys). While commodity producers may be found only in certain locations, the market for commodities and their transportation is efficient. In contrast, costs for sourcing workers distantly are prohibitive. Workers are unlikely to commute more than a couple of hours between work and home. The U.S. Census County-to-County Worker Flow Files show that the average commute to work was 22.4 minutes in 1990 (and 25.5 minutes in 2000); and that less than three percent of workers travelled more than 90 minutes between work and home. Skilled labor for any given industry will likely be located in only a few locations, and labor markets across locations may be less efficient if people are hesitant to relocate. For example, contrary to common belief, geographic mobility in the U.S. is low. According to the Current Population Survey conducted by the Census Bureau since 1948, 17 percent of the population moved on average from 1985 to 1994 in the U.S. Of those movers, 36 percent moved to another county and 16 percent moved to another state. Similarly, there is growing recognition that knowledge is somewhat localized because it has tacit components. For example, Adams and Jaffe (1996) demonstrate that localized R&D spending (within 100 miles) improves proximate establishments' total-factor-productivity substantially more than distant R&D activity.

Because worker commuting patterns and R&D spillovers are more geographically bounded than the flow of commodities, the benefits emanating from pools of skilled labor and potential knowledge inflows will be more localized than the benefits from specialized suppliers. As a result, we expect that the production weights for labor and R&D activity understate firms' incentives to co-locate for skilled labor and knowledge inflows, while the weight for materials overstates the importance of specialized suppliers. But how large are these under- and over-statements?

To better benchmark the combined impact of production weights and location specificity on the draw of each agglomeration economy, we turn to prior empirical work. Rosenthal and Strange (2001) examine differences in the extent of agglomeration in U.S. manufacturing at the industry level and find that the explanatory power of Marshall's agglomeration economies differs when applied at three geographic levels. Labor-market pooling explains industry agglomeration at zip code, county, and state levels; access to intermediate inputs explains agglomeration only at the state level; and potential knowledge spillovers have effect at the zip code level. Applied to our setting, this study suggests a greater importance for suppliers and labor than for knowledge inflows, but the relative merits of skilled labor and specialized suppliers remain unclear. For that reason, we are agnostic about this particular bilateral comparison, and work with two bilateral rankings instead. More formally:

H1. Skilled labor *(share of same industry employment)* and specialized suppliers *(share of supplier-industry employment)* will be more attractive than potential knowledge spillover *(share of industry R&D activity)*.

2.3 Why are some firms more likely to pursue agglomeration economies?

The first and second layers of the framework assess the relative attraction of each agglomeration economy for the average firm. Our final step explores why and how individual firms will vary in their propensity to co-locate for each economy. With each economy offering reduced costs to proximate firms, we might expect geographic concentration to be universally appealing. To understand why this might not be the case, we examine how much proximate firms contribute to agglomeration economies, and whether their competitors can make use of those contributions.

Shaver and Flyer (2000) make two fundamental observations about firms' contributions to proximate competitors and the impact of those contributions on co-location. First, firms not only benefit from agglomeration economies, but they also contribute to them. It is the collective activity of proximate firms that forms, maintains, and grows these public goods. Second, firms differ in how much they gain from and contribute to the pool. Shaver and Flyer suggest the benefits of agglomerating are asymmetric – that larger firms contribute more to the pool, and gain less, than their smaller competitors.[4] As a result, the net benefit of agglomerating may be far less for some firms than for others. In some cases, the net benefit may actually be negative, creating incentive to avoid co-location entirely. Specifically, they show that while co-location improves survival for average-sized firms, larger entrants are much less likely to locate near competitors.

Because Shaver and Flyer address Marshall's agglomeration economies in aggregate, two issues must be resolved before applying their reasoning to *individual* agglomeration economies. First, they do not provide guidance on whether firms' behavior is similar for all types of agglomeration economies. Second, as a consequence, they cannot identify which firm traits may be more relevant for each agglomeration economy.

To address the first issue we start with Alcácer and Chung (2007) who argue that, in addition to how much a firm contributes to agglomeration economies, one must consider the usefulness of that contribution. Agglomeration economies offer potential benefits, but not all proximate firms will be able to benefit. Alcácer and Chung make this point in relation to knowledge spillovers. They argue that knowledge spills are not easy to pick up: firms may create a pool of potential knowledge spillovers, but all potential spillovers are not accessible or useful to all firms. Knowledge is difficult to transfer at arms length by virtue of its tacit nature, which

[4] A manifestation of this is that large firms can sometimes command an entourage of followers. The infrastructure and markets created by industry agglomeration would be less valuable to these firms, giving them more latitude to locate away from competitors. Martin, Mitchell and Swaminathan (1995) found this to be the case among some Japanese automotive manufacturers, who were able to locate away from Detroit because their suppliers followed them.

makes it less useful without repeated contact. The firms most able to overcome this difficulty and realize the benefits of potential knowledge spills are firms that have the capacity to absorb knowledge gleaned from more-sophisticated competitors. When competitors cannot materially benefit from knowledge outflows, the knowledge has limited usefulness; therefore the risk is reduced for the technically advanced firm that leaks that knowledge. As a result, advanced firms can readily locate among less-advanced competitors.

We generalize this limited-usefulness argument and apply it across the three agglomeration economies. These economies are potentially useful to firms because they create pools of factor inputs. However, accessing factor inputs involves exchange through the marketplace. If the market is imperfect, the factor input becomes difficult to access. A firm that contributes to an agglomeration economy that other firms cannot readily access through the marketplace is not risking its competitive advantage.

We expect labor markets to be the least imperfect – and therefore labor will be the easiest factor to access – followed by specialized suppliers and knowledge inflows from competitors. Labor markets are relatively easy to penetrate because they are thick with information. First, firms can clearly specify their labor needs: professions, job categories, and duties are fairly well defined.[5] Prices (wages) are readily set by supply and demand in the marketplace. Additionally, labor quality is somewhat observable from education, training, and relevant experience. Because of such market transparency, firms looking to hire new employees face few hurdles. Among professions normally employed, firms should be relatively well equipped to assess, hire, and integrate individuals into their organizations. Thus any contributions that firms make to enlarge and deepen labor pools should be useful for neighboring firms.

[5] The market is so well defined that the Bureau of Labor Statistics maintains the "Standard Occupation Classification" with 820 classifications covering all possible professions.

The market for suppliers' services can be more imperfect and thus less accessible for firms. While some services can be tightly specified, requiring only a one-time transaction; other services are more involved and difficult to completely specify ex-ante – for example when a supplier provides services over time for an ongoing project. In addition, supplier services received by one firm might not be suitable for another due to differences in business processes and routines. Due to this lack of standardization, pricing becomes more idiosyncratic. These challenges mean that the ability to identify good suppliers, engage them, and work successfully with them is not a given. As such, contributions made to pools of specialized suppliers may not be widely accessible.

The market for knowledge inflows from competitors is the most imperfect and hard to access because of knowledge's tacit nature, which makes its identification and transfer difficult. Identifying the potential value added from a piece of knowledge requires substantial interaction between firms, which is hard to do at arms-length. Transfer is likely to be even more problematic, since it may involve repeated interaction with an unwilling donor firm. With unwilling donors, prices are non-existent. As a result the accessibility of knowledge outflows to neighboring firms will vary, with some readily gleaned by all, and others being more opaque.

The ordering of agglomeration economies' relative usefulness suggests that firms would be most wary of contributing to skilled labor pools, because these pools are easily accessed by competitors. In contrast, with pools of specialized suppliers and knowledge spillovers, market imperfections would reduce the likelihood that competitors could glean competitive advantage, allowing contributing firms to co-locate without significant risk.

To address the second unresolved issue – which firm traits may be more relevant for each of the agglomeration economies – we draw upon prior research. Following Shaver and Flyer, we expect that larger firms will make larger contributions to pools of skilled labor and specialized suppliers because they employ more workers and engage in more supplier activity. Similarly, following Alcácer and Chung, we expect that firms with greater technical capabilities will engage in more R&D activity than other firms, making them more likely to contribute knowledge outflows.

The relative value of each agglomeration economy, firm heterogeneity (firm size and technical capability), and agglomeration economy heterogeneity (whether the market for the corresponding factor is more or less imperfect) will lead to different firm behavior. Larger firms will be more wary of thick labor markets because labor pools are easy for competitors to penetrate. In contrast, specialized suppliers and knowledge inflows are harder to leverage, making it less risky for larger or more technically advanced firms to co-locate with competitors. Stated more formally, we expect:

H2a. Economically larger firms will be less attracted to skilled labor *(share of same industry employment).*

H2b. Economically larger firms will be more attracted to specialized suppliers *(share of supplier industries employment).*

H2c. Technical-leader firms will be more attracted to potential knowledge spillovers *(share of industry R&D activity).*

3. Data

To assess the implications of our framework for firm behavior towards agglomeration economies, we examine how variation across locations in share of same-industry employment, supplier-industry employment, and industry R&D activity affects firms' location choices. To explain location choice as a function of both location-specific and firm-level traits, we draw upon several data sources.

Our dependent variable, the location within the U.S. of first-time foreign entrants, is drawn from *Foreign Direct Investment in the United States, Annual Transactions,* published by the International Trade Administration (ITA). We use only transactions in the continental U.S. with SIC codes in manufacturing (all four-digit SIC industries between 2000 and 3999) for 1985 through 1994.[6] We focus on manufacturing industries, rather than agriculture, mining or services, because we expect the impact of agglomeration economies on firms' location decisions to be the greatest.

We restrict our sample to transactions by firms that are first-time entrants because prior investments can influence subsequent location choices and create dependence among observations for the same firm.[7] We also restrict ourselves to firms making greenfield investments (excluding those that enter through acquisitions) so that we might observe firms making the least-constrained location choices.[8] We then match firms making these transactions to their accounting information from Worldscope. Restricting ourselves to first-time, greenfield entries by firms with publicly-available accounting information yields a sample of 657 investment transactions.[9] The five, 4-digit SIC industries that receive the most transactions are 3714 (automotive components), 2821 (plastics/materials), 3674 (semiconductors), and 2819 (industrial inorganic chemicals) with 49, 21, 17 and 14 transactions respectively.

[6] The data became available electronically in 1985; ITA stopped collecting it in 1994. The ITA provides investments' location data, four-digit SIC code, foreign investor's name and country of origin, transaction value and mode of entry. Kogut and Chang (1991) and Chung and Alcácer (2002) use this dataset to construct either their dependent variables or focal independent variables.

[7] To determine first-time entrant status, we review inward investment data back to 1975. We exclude transactions where a firm invests in the same industry as prior transactions, but include transactions when a firm invests in a different 4-digit industry from all prior investments.

[8] This is consistent with Shaver and Flyer (2000) and Alcácer and Chung (2007), both of whom use only greenfield investments.

[9] These transactions are made by 500 unique firms. We include in our definition of first-time entrants additional entries from the same parent firm if it made those investments in different 4-digit industries than all prior entries.

For each transaction, investment location can be defined broadly (the Pacific Northwest, New England), or more narrowly (the Bay Area, the Boston metropolitan area). In an effort to identify geographic areas that mimic economic activity rather than state or administrative boundaries, the Bureau of Economic Analysis (BEA) defined 170 economic areas spanning the continental U.S.[10] Each economic area consists of at least one node (a metropolitan or densely populated area that serves as center of economic activity) and the surrounding counties that are economically related to the node(s). Commuting patterns are the main factor used to determine economic relationships among counties. Each economic area includes, as far as possible, both the work site and residence of its labor force. Figure 2 shows the location of transactions in our sample.

For our first location-specific independent variable, we use *%_industry_employment*, the percentage of workers in a given industry present in each economic area, a variable that is widely used in recent papers looking at the geographic concentration of U.S. manufacturing industries. A location with a higher portion of existing-industry employment must have a greater amount of skilled labor with industry-specific attributes. We aggregate employment per industry-year from the county level to the economic area level using the definition of economic areas provided by the BEA. This, as well as all our focal variables, varies by economic area, industry, and year.

Second, we introduce *%_industry_suppliers*, the percentage of employment from supplying industries present in a given economic area. This variable, which captures the existence of specialized suppliers, is constructed using the Benchmark Input-Output Accounts of the U.S. for 1992 from the Department of Commerce. The input-output accounts provide the types and amounts of commodities made by all industries, as well as which commodities, in what

[10] The BEA introduced economic areas in 1977 and redefined them in 1983 and 1995 to adjust for changes in commuter patterns across time. We use the 1995 definition.

amounts, each industry uses. Using these commodities and quantities, we determine the fraction of inputs from all other industries used by each focal industry.[11] Then, to get our industry-economic area-year level measures, we multiply these fractions by the corresponding percents of industry employment by economic area and then sum across all industries to get percent of related supplying-industry employment. While a focal industry can potentially be fed by many supplying industries, we limit ourselves to each focal industry's top ten supplying industries in constructing this measure, while excluding the focal industry itself as one of the top input providers (since we already explicitly capture this contribution in *%_industry_employment*). These top ten supplying industries account for 63% of all inputs for the industries where firms in our sample invest.

For our third location-specific focal measure, we use *%_industry_patents*. We use patenting activity to proxy for potential knowledge spillovers similar to Ellison, Glaeser, and Kerr (2007) and Alcácer and Chung (2007). We calculate patent stocks by industry, economic area, source and year using patent data from the Micropatent dataset, which contains all information listed on the front page of every U.S. patent granted since 1975. We use four variables to characterize patents: inventor location (for economic area), technological class and subclass (for industry), and filing date (for year). The dataset provides 499,271 patents whose first inventors are located in the U.S. and whose application dates fall between 1985 and 1994.

The location of a patent is determined by the address of its first inventor. Most patents list one inventor; when multiple inventors are listed they tend to locate in the same economic area. Since we need information about cities and states to map locations to economic areas, we

[11] Due to the I-O accounts being commodity based, we are unable to match some commodities to SIC industries, resulting in the loss of some commodity flows in determining intra-industry supply linkages. Also when constructing the similar *%_industry_buyers* measure as a control variable, additional losses occur. Commodities can also be exported or consumed by federal, state, and local government. We exclude these "final use" categories from being considered as buying industries since an appropriate geographic location of such buyers is problematic.

exclude patents for which this information is not available, yielding a sample of 496,275 locations, representing 99% of the patents granted to assignees in the U.S. that were applied for between 1985 and 1994.

We map technological classes and subclasses onto industries using a concordance that links the International Patent Classification (IPC) system to the U.S. Standard Industrial Classification (SIC) system at the four-digit SIC level (see Silverman (1999) for more information). Alternatively, we use the *"Concordance Between the Standard Industrial Code (SIC) Classification System and the U.S. Patent Classification (USPC) System"* issued in January 2002 by the U.S. Patent and Trademark Office (USPTO).[12]

Finally, we identify the time dimension based on when the patent application was filed. The application date is the closest available to the generation of an innovation. We also calculate stock using the date the patent was granted, with very similar results. With economic area, industry, and year defined for each patent, we calculate patent stocks with different time lengths (three-, five-, and seven-year averages preceding the focal year) to smooth possible yearly fluctuations and to capture continuity in technological activity. Results are similar using any of the averages; subsequent results are based upon three-year averages. We then use these three-year averages to calculate percentages by economic area, industry, and year.

Among these three focal measures, *%_ind_employment* and *%_ind_patents* often take on a value of zero. This is because not every industry has direct employment or research in every part of the country; in contrast, a focal industry draws from many supplier industries that on average end up covering the entire country. At our economic area-industry-year level of analysis,

[12] Although these two methods of mapping patents to industries yield very similar stocks per industry, we use the IPC alternative because the USPTO concordance is only at a 2/3-digit SIC level versus the IPC concordance, which generates stocks at a 4-digit level.

we have 72,790 cells.[13] With *%_ind_employment* 20,907 cells (28.7%) take on a value of zero; while, with *%_ind_patents*, 56,124 cells (77.1%) take on a value of zero. Aware that such a distribution of values cannot be well represented by a single continuous measure, we introduce a dummy variable to indicate which economic area-industry-year cells take on zero values: *%_ind_employment_zero* and *%_ind_patents_zero*. These dummy variables should take on strongly negative values – the lack of any direct employment or research activity will make a location much less attractive for a potential entrant.

For our first firm-specific trait, firm technical capabilities, we use R&D intensity measures (R&D spending scaled by sales) obtained from Worldscope.[14] To classify firms as technical leaders or laggards, we use the year of entry and compare each firm's R&D intensity to the average R&D intensity for all U.S. firms in the same 4-digit industry.[15] Using industry-year R&D intensity data that was drawn from *Compustat*, we identify quartiles of R&D intensity. A firm whose R&D efforts would place it in the top quartile of R&D intensity in the U.S. is designated a *leader* firm; a firm whose R&D efforts are below the top quartile of U.S. firms is designated a *laggard* firm. We use the quartile definition instead of the mean or median because the R&D intensity of all foreign entrants to the U.S. is quite high and using a lower threshold would lead to most entrants being defined as leaders.[16]

---

[13] While there are 337,850 total cells possible (233 4-digit industries x 10 years x 145 economic areas), there are not entries into all industries in all years. On average across industries, there is only entry in 2.15 of the possible 10 years.

[14] Most first-time entrants are in the same industry as their parents. From our sample, 63% of all entries are in the same 4-digit code as their parent's primary 4-digit SIC and 93% are in the same 3-digit code as the parent.

[15] We use firms in the U.S. as our comparison group given our interest in location behavior among first-time entrants to strategically benefit from agglomeration economies. The bulk of creation and maintenance of agglomeration economies is going to come from the presence of proximate U.S. firms.

[16] Consistent with the internalization theory of foreign direct investment, we would expect foreign entrants to be large and R&D intense relative to incumbents encountered in the host country. These are the firms that were successful at home and decided to use their existing capabilities more widely.

For our second firm-specific trait, firm economic size, we use a measure of total firm assets, also obtained from Worldscope. To classify firms as economically large or small, we compare each firm's total assets to the total assets for all U.S. firms in the same 4-digit industry. A firm whose total assets are above the top quartile of U.S. firms is designated a *large* firm; a firm whose total assets are less than the top quartile of U.S. firms is designated a *small* firm.[17]

[Table 1 about here]

In addition to these focal variables, we need to control for other characteristics, such as access to consumers, market growth, and low-priced inputs such as land, etc. Collecting an exhaustive set of these variables at the economic-area level for all industries is practically impossible, so we use three variables to proxy for the host of attributes that attract entrants.

Our first and second control variables are *establishment_growth* and *employment_growth*. Both variables control for location-industry-time specific heterogeneity that can influence location decisions. Our establishment and employment data are drawn from *County Business Patterns* reports from the Bureau of Economic Analysis for 1985 to 1994.

Our third control variable is *%_industry_buyers*, which is constructed in a similar manner to the focal variable *%_industry_suppliers*, but uses the input-output tables to identify purchasing/buying industries of the focal industry. *%_industry_buyers* is an important control because proximity to buyers should be an important consideration when firms decide where to locate. Since an industry's buyers and suppliers are likely to be somewhat overlapped geographically, including suppliers only would create a classic omitted-variables problem: any

[17] Optimally we would use the entrant's establishment size, since it is the economic size of the investment that is contributing to and drawing upon agglomeration economies. However, entrant size is only intermittently reported (about 40% of the time) and is noisy itself since a firm might subsequently add to an initial investment. Instead of introducing this additional source of noise and losing many observations, we use parent-firm-level information for both entrants and U.S. comparators. While noisy, this measure performs well – consistent with our theoretic expectations – in our subsequent analysis.

significance of the omitted-buyer's proximity would inappropriately load onto the included measure for supplier proximity.

Finally, following Head, Ries, and Swenson (1995), we use alternative specific constants (ASCs) (dummy variables per economic area) to capture time-invariant attributes that may be attracting entrants. Examples of attributes captured by ASCs would be land area, population density, right-to-work laws, tax rates, and coastal location. To the extent that such attributes are relatively constant across time, they are reflected by the ASCs.

4. Method

We examine the attractiveness of same-industry employment, supplier-industry employment, and industry R&D activity, as well as how this attractiveness varies with firm heterogeneity.

We model the location-choice process, where firms choose one of 170 economic areas, through a conditional-logit model as described by McFadden (1974). Conditional logit has been used extensively for cases where choices are made from a large set of possible geographical locations (Head *et al*., 1995; Chung and Alcácer, 2002).

The conditional logistic regression is similar to the ordinary logistic regression models except that the data occur in groups. The idea is to fit a logistic model that explains why a given choice has a positive outcome in a group (choice set) conditional on the other existing alternatives in the choice set. The conditional-logit model is specified as follows. Let V_{ij} represent the value to firm i of choosing location j. V_{ij} depends on location characteristics that vary by industry, X_{ij}, according to:

$$V_{ij} = \beta' x_{ij} + e_{ij} \qquad (1)$$

Let Y_{ij} be our dependent variable equal to 1 if firm i chooses location j and 0 otherwise. Firm i will choose location j if $V_{ij} > V_{ik}$ for all $k \neq j$. Assuming that e_{ij} are independent and identically distributed with Type I Extreme Value Distribution, then the probability that location i is chosen is

$$\mathrm{P}(Y_{ij}=1) = \mathrm{P}(V_{ij} > V_{ik} \text{ for all } k \neq j) \quad = \frac{e^{\beta' x_{ij}}}{\sum_{k \neq j} e^{\beta' x_{ik}}} \qquad (2)$$

where k = 1 to m are all locations entered by at least one firm between 1985 and 1994. Since 25 economic areas are never chosen in our sample, our choice set is reduced from 171 to 145 potential locations. While dropping alternatives might seem to bias our results, a central feature of the conditional logit is its independence from irrelevant alternatives (IIA) property – the ratio of two alternatives' probabilities of being chosen is the same regardless of what other alternatives are available (or unavailable). For example, Train (2003) shows that estimated model parameters are consistent when using only a subset of alternatives in the decision making process.

Note that the conditional-logit model focuses on location traits instead of firm traits. Firm characteristics would be invariant within a choice set, appearing in both the numerator and denominator of equation 2 and dropping from the estimation. Chooser's traits can be modeled by interacting terms (multiplying economic area traits by firm traits) or by splitting the sample into sub-samples according to specific firm traits. We choose the second option for two reasons. First, it offers a more parsimonious presentation of the results because it uses fewer regressors than models with interaction effects; second, it avoids numerous cross-terms that are difficult to interpret with non-linear models. For example, Ai and Norton (2003) find that "the magnitude of the interaction effect in nonlinear models does not equal the marginal effect of the interaction term, can be of opposite sign, and its statistical significance is not calculated by standard software." Also, while interaction effects are by definition co-linear with the variables being interacted, splitting samples does not increase co-linearity levels. Finally, using sub-samples allows all variables in the model to vary for each sub-sample, rather than using interaction terms where only those variables explicitly interacted will vary.

5. Results

Table 2 provides an overview of each agglomeration economy's attractiveness. The first column is a benchmark specification of control variables. The subsequent columns introduce each independent variable separately.

[Table 2 about here]

Focusing on the measures for agglomeration economies, across all models, we see that *%_ind_employment* is consistently positive and significant. Including the zero dummy variable, *%_ind_employment_zero* drops the magnitude of *%_ind_employment* slightly, but otherwise remains very significant. Assessing the improvement from including *%_ind_employment_zero*, we look at the difference in log likelihood between columns 2 and 3: -2497.01 vs. -2476.21 for a difference of 20.80, which is significant at better than a 1% level. *%_ind_suppliers* is also consistently positive and significant across all models. For *%_ind_patents*, while consistently positive, it is only significant in column 5 before *%_ind_patents_zero* is included. Looking at the difference in log likelihood between columns 5 and 6, including *%_ind_patents_zero* does not significantly improve overall model fit.[18] We investigate the marginal significance of *%_ind_patents* further by separating out industries that are more R&D intensive later in the paper.

[18] In later specifications found in Table 3, we continue to include *%_ind_patents_zero* because it occasionally does have significant statistical effect, and we want to avoid an omitted variables problem in later specifications.

Our main interest in this table is to compare the economies' magnitudes of effect. To do so, we use the odds-ratios, which due to the conditional logit's form are exponentiated values of the coefficient estimates. Using odd-ratios' values from column 6, we see that *%_ind_employment* and *%_ind_suppliers* have similar magnitudes of effect; their odds-ratios are 1.090 and 1.089 vs. 1.007 for *%_ind_patents*. Since the corresponding independent variables are expressed as percentages, the 1.090 indicates that if an economic area had a 1% increase in *%_industry employment* this would lead to a 9.0% increase in likelihood of being chosen. The odds-ratios result and corresponding Welch's t-tests for differences indicate that skilled labor and specialized suppliers are of about the same attractiveness for firms, and that both are at least ten times more attractive than potential knowledge spillovers. This ordering is consistent with hypothesis H1.

Of note also is the odds-ratio for the zero employment dummy variable, *%_ind_employment_zero*. The odds-ratio from column 6 is 0.250, which means that an economic area that has no relevant employment is four times less likely of being chosen relative to if it had any industry employment.

Having established relative attractiveness across agglomeration economies, we turn to firms' strategic response. Table 3 splits the sample along the two key dimensions of firm heterogeneity: firm economic size and firm R&D intensity. Column 1 is again the baseline model of all firms – identical to that shown in Table 2, column 6. Columns 2 and 3 present the results for the split by firm economic size, while columns 4 and 5 present the results for the split by firm R&D intensity.

[Table 3 about here]

Looking at *%_ind_employment* in columns 2 and 3 where the firms are split into those that are economically large and economically small, the coefficient estimates are both positive and significant. Of note, the coefficient for the large-firm group is substantially smaller than for the small-firm group: 0.0757 versus 0.1284 (using Welch's t-test, this difference is significant at more than a 1% level). The corresponding odds-ratios (1.079 and 1.137) suggest that larger firms are roughly 40% less attracted than small firms to locations with skilled labor. This finding is consistent with hypothesis H2a – larger firms would be less attracted to locations with skilled labor. Another consistent explanation is that larger firms are better able to attract workers to wherever they locate and can be less concerned with existing pools of labor. Our difference between large and small firms is also consistent with Shaver and Flyer (2000), who find that large establishments are about 40% less attracted to locations thick with same-industry establishments.[19]

Also looking at *%_ind_suppliers* in columns 2 and 3, the coefficient estimates for large and small firms are starkly different: positive and significant for large firms, but slightly negative and non-significant for small firms. This suggests that only large firms are drawn to pools of specialized suppliers. This finding is consistent with hypothesis H2b – limited accessibility reduces concerns about aiding competitors because they may have difficulty leveraging other firms' supplier base. As a result, the greater contributors – large firms in this case – can enjoy agglomeration economies without concern over aiding smaller competitors.

The lack of significance for *%_ind_suppliers* with small firms is somewhat surprising – we might expect such firms to rely on the existing infrastructure of suppliers. Among this population at least, it appears that small firms' location choices are driven not by the presence of

[19] From Shaver and Flyer (2000): column 1 of Table 5 on page 1189, comparing the estimate for "PROPORTION" of 11.82 to the estimate for "PROPORTION x LARGE ESTABLISHMENT" of -4.79.

suppliers, but by the availability of skilled labor and potential knowledge inflows from competitors (the coefficient estimates for *%_ind_employment* and *%_ind_patents* are strongly significant). Both of these effects may indicate the importance of scale. Small firms do not account for enough economic activity to encourage workers to relocate, making them more dependent upon existing pools of labor. Similarly, small firms do not perform enough technical activity to keep pace with changing technologies and thus have to rely on the proximate activity of other firms as well.

Finally we look at *%_ind_patents* in columns 4 and 5, where firms are split into those that are more and less R&D intensive. Here again we expect to find that R&D-intensive firms are more attracted to R&D-intensive locations than firms less reliant upon R&D, but find little evidence of variation. Coefficient estimates for both leader and laggard firms are positive but not significant. This lack of significance is similar to prior research by Alcácer and Chung (2007), who find that R&D-intensive firms are more attracted to potential knowledge spillover. However, the 2007 study disaggregates by knowledge source (academic, industry, and government) and amount (high, low, and zero) to show several strong relationships. We do not pursue similar disaggregation, because our interest is in assessing firms' attraction to each of Marshall's three agglomeration economies instead of going into greater depth with a single one. These results for potential knowledge spillover, combined with our findings for the other economies, suggest that firms' considering new locations will place more value on labor and suppliers than they place on potential knowledge inflows from competitors.

In summary, our results show that *%_ind_employment* and *%_ind_suppliers* are several times more attractive to firms choosing locations than *%_ind_patents*. Firms vary in their response to these economies depending upon whether they are economically large or small and whether they can access and make use of the factor inputs available in a particular agglomeration. Consistent with concerns about aiding proximate competitors, large firms tend to be less attracted to locations with high *%_ind_employment*. In contrast, large firms appear more attracted to locations with high *%_ind_suppliers*, which is consistent with the limited usefulness of contributions to pools of supplier activity. Relative to these distinct location behaviors, firm response to potential knowledge spillovers is minor.

6. Robustness

Before settling on these results, we conduct additional tests to assess their robustness. In particular for H1 – skilled labor and specialized suppliers being more attractive than potential knowledge inflows – the strategic value of each agglomeration economy may vary based upon how intensively it is used in certain industries. For example, firms in labor-intensive industries might place a greater emphasis on geographic concentrations of skilled labor than on specialized suppliers or potential knowledge inflows. Similarly, firms in supplier-intensive industries might be more drawn to supplier activity; or firms in R&D-intensive industries might be more drawn to knowledge inflows. As such, we split our sample by labor intensity, extent of suppliers used, and R&D intensity.

To assign industries into high and low categories for labor intensity, we use the National Bureau of Economic Research's Manufacturing Industry Productivity Database, which provides industries' value-added as well as the inputs used in production. To determine labor intensity, we use an industry's ratio of total payroll divided by total value-added, averaged for the nine years of our investigation, 1985-1994. We designate industries above the median as labor intensive and those below as not. To assign industries to high supplier-use or low supplier-use, we use the Benchmark Input-Output Accounts of the U.S. for 1992 from the Department of Commerce, which reports for each industry the proportion of value coming from supplying industries. We also split industries into those that are more and less knowledge-intense. To assign industries to high-tech and low-tech categories, we follow the OECD classification from the report *OECD Science, Technology and Industry Scoreboard* (1999). Labor-intensive transactions make up 45.7% of all transactions, while 44.6% of transactions are classified as high supplier-use, and 60.9% of transactions are classified as high-tech. Results appear in Table 4 below.

[Table 4 about here]

In Table 4, column 1 is the baseline specification – identical to that shown in Table 2, column 6. Columns 2 and 3 present the results for the split by industry labor-intensity, columns 4 and 5 are for the split by industry- supplier-intensity, and columns 6 and 7 present results for the split by industry R&D-intensity.

Looking at *%_ind_employment* in columns 2 and 3, we see that both coefficient estimates are statistically significant, but they differ in magnitude. As we would expect, the attraction to pools of labor in labor-intensive industries is greater: 0.0979 vs. 0.0756. This difference in coefficient estimates is strongly statistically significant using Welch's t-test. Also of interest is how the coefficients for industry suppliers and patents differ for these two industry groups. With labor-intensive industries, only industry employment is a draw; suppliers and patents are strong draws in less labor-intensive industries. This is sensible, since by definition those two categories would be of greater importance in less labor-intensive industries. Looking at the odds-ratios even in this less labor-intensive group, the draw of employment and suppliers is still much greater than it is for patents: 1.079 and 1.207 versus 1.018.

Looking at *%_ind_suppliers* in columns 4 and 5, we see that the coefficient estimates are substantially different. For firms in high supplier-use industries, the estimate is positive and significant, while for firms in the low supplier-use group, the estimate is positive but not significant. The coefficient estimate for the high supplier-use industry group is also substantially larger than the baseline all-industry sample: 0.1504 vs. 0.0857. Analogously, the odds-ratio is also greater: 1.162 versus 1.089. Unsurprisingly, these differences suggest that the attractiveness of agglomerated supplier activity is more important for firms in industries that rely heavily on suppliers.

Looking at *%_ind_patents* in columns 6 and 7, the coefficient estimates are both positive but not significant. While we might expect a positive and significant estimate for the high R&D-intensity group, this basic split is perhaps still too aggregated to reveal an effect.

With these three industry splits, we see that even when examining industries that use certain inputs more intensively, hypothesis H1's expected relationship remains – skilled labor and specialized suppliers are more attractive to firms than potential knowledge inflow.

7. Conclusions

Locations thick with similar economic activity expose firms to pools of skilled labor, specialized suppliers, and potential inter-firm knowledge spillovers that can provide firms with opportunities for competitive advantage. While certainly attractive, the draw of these agglomeration economies will vary. Firms contribute differentially to the formation, maintenance, and growth of these agglomeration economies. Each agglomeration economy also differs in how readily competitors can leverage contributions made by others. As a result, some firms must be wary of aiding their competitors by co-locating with them.

To better understand how firms respond to agglomeration economies, we develop a framework of three interconnecting layers. The first layer uses the long-standing literature in economics on production functions to establish a baseline for each agglomeration economy's relative value; an economy should be more valuable if it provides factor inputs that firms use more intensively. The second layer modifies this baseline by examining whether particular economies are location-specific; firms will have more incentive to co-locate when an economy's factor inputs cannot be leveraged from afar. The final layer addresses why individual firms vary in their propensity to co-locate with competitors based upon both firm heterogeneity and agglomeration economy heterogeneity; firms that contribute greater amounts to agglomeration economies may be wary of co-locating, though this concern will be reduced when competitors cannot make use of those contributions due to more imperfect markets.

The framework generates two novel predictions. First, pools of skilled labor and specialized suppliers will be more attractive to firms than potential knowledge inflows. Second, the risk of aiding competitors might be reduced by an agglomeration economy's relative usefulness. Large firms might shy away from locations with skilled labor, because labor pools are easy for competitors to tap. The same large firms might be less concerned about enriching local supplier networks, because supplier networks are of more limited usefulness to competitors. Similarly, the limited usefulness of leaked knowledge would allow technical leaders to locate with less-advanced competitors.

We find empirical results consistent with the first prediction and somewhat consistent with the second. For first-time foreign entrants to the U.S. making greenfield investments in 1985-1994, we find that industry employment and supplier activity are about ten times more attractive than industry patent stocks, suggesting that firms, on average, place more value on pools of skilled labor and specialized suppliers than on potential knowledge inflows from competitors. The priority placed on labor and suppliers persists even for industries that are more R&D intensive. Introducing key dimensions of firms' heterogeneity reveals that economically larger firms are less attracted to industry employment, but are more attracted to supplier industry activity. This paired finding suggests that concern about aiding competitors is strong when competitors can leverage agglomeration economies for strategic advantage, and weak when they cannot.

For the strategy literature, our intent is to provide a comprehensive, consistent, and cohesive lens to view firms' location behavior in the presence of Marshall's three agglomeration economies. The framework separates each economy from the collective, allowing us to better understand the variety of firm responses composing strategic co-location. This structure both orients and integrates extant research, allowing us to resolve some apparent contradictions. For example, Shaver and Flyer (2000) find that new entrants might shy away from locating with competitors, while Alcácer and Chung (2007) find that new entrants might be drawn to competitors. Considering their studies in the context of our framework, there are two key differences: (1) Shaver and Flyer examine the set of three agglomeration economies collectively, while Alcácer and Chung only look at knowledge spillovers; and (2) concerns about aiding competitors predominate for the three agglomeration economies in aggregate, but wane when looking only at knowledge spillovers.

Our findings also contribute to the economics literature on agglomeration. Work by Rosenthal and Strange (2001) and Ellison, Glaeser, and Kerr (2007) make great strides in distinguishing and prioritizing Marshall's agglomeration economies by examining long-term incumbents. Both state that analysis of recent entrants' agglomeration behavior is critical for addressing alternate explanations not in line with their theoretic explanations. We specifically examine firm-level entry and find results prioritizing Marshall's externalities that are very consistent with these studies, namely, that access to skilled labor and specialized suppliers has a greater impact on location decisions than access to knowledge inflows from competitors.

A number of caveats for our results remain. Distinguishing the three agglomeration economies from one another depends upon our measures being strong enough to capture – and yet sharp enough to separate – the distinct effects. Undoubtedly there is some overlap among our measures. For example, from a theoretic standpoint, skilled labor is one of the main conduits through which knowledge would spill. While non-overlapped measures are desirable, we are hard-pressed to find sharper measures, especially when the economies are theoretically entangled. We are not unique in this area; the research on which we build faces the same measurement challenge. Another limitation is our particular context of new entrants into the U.S., which certainly reduces the generality of our results. The idiosyncrasies of entering firms' motives, which are driven by the prevailing economic topology, are likely specific to the U.S.

Overall, our framework suggests how firms would react to each agglomeration economy. We theoretically and empirically separate the set, and show that firms' location choices balance the perceived risk of aiding competitors with a recognition that some agglomeration economies will be of limited use to others. This integrative framework is important since understanding firms' behavior towards these agglomeration economies is a large component of what drives firms' location strategy, which can, in turn, be a source of competitive advantage.

References

Adams, J. D. and A. B. Jaffe. 1996. Bounding the effects of R&D: An investigation using matched establishment-firm data. *The Rand Journal of Economics* **27**(4): 700-721.

Ai, C. and E. C. Norton. 2003. Interaction terms in logit and probit models. *Economic Letters* **80**: 123-129.

Alcácer, Juan and Wilbur Chung. 2007. Location strategy and knowledge spillovers. *Management Science* **53**(5): 760-776.

Chung, Wilbur and Juan Alcácer. 2002. Knowledge seeking and location choice of Foreign Direct Investment in the United States. *Management Science* **48**(12): 1534-1554.

Ellison, G., E. L. Glaeser, and W. Kerr. 2007. *What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns*. U.S. Census Bureau - Center for Economic Studies.

Griliches, Zvi and Jacques Mairesse. 1983. Comparing Productivity Growth: An Exploration of French and US Industrial and Firm Data. *European Economic Review* **21**(1-2): 89-119.

Hall, Bronwyn H. 1993. Industrial research during the 1980s: Did the rate of return fall? *Brookings Papers on Economic Activity* **2**: 289-343.

Head, K., J. Ries and D. Swenson. 1995. Agglomeration benefits and location choice: Evidence from Japanese manufacturing investments in the United States. *Journal of International Economics* **38**(3,4): 223-247.

Jaffe, Adam B., Manuel Trajtenberg and Rebecca Henderson. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics* **108**(3): 577-598.

Kogut, B. and S. J. Chang. 1991. Technological Capabilities and Japanese Foreign Direct Investment in the United States. *The Review of Economics and Statistics* **73**(3): 401-413.

Marshall, Alfred. 1920. *Principles of Economics; An Introductory Volume*. Macmillan and Co.: London, U.K.

Martin, Xavier, Will Mitchell and Anand Swaminathan. 1995. Recreating and extending Japanese automobile buyer-supplier links in North America. *Strategic Management Journal* **16**(8): 589-619.

McFadden, D. 1974. *Conditional logit analysis of qualitative choice behavior*. Frontiers in Econometrics. P. Zarembka (ed). Academic Press**:** New York**:** 105-142.

Nadiri, M. I. and Seongjun Kim. 1996. R&D, production structure and productivity growth: A comparison of the U.S., Japanese and Korean manufacturing sectors. *NBER Working Paper* 5506.

Rosenthal, S. S. and W. C. Strange. 2001. The determinants of agglomeration. *Journal of Urban Economics* **50**(2): 191-229.

Shaver, J. Myles and Frederick Flyer. 2000. Agglomeration economies, firm heterogeneity, and Foreign Direct Investment in the United States. *Strategic Management Journal* **21**(12): 1175-1193.

Silverman, B. S. 1999. Technological resources and the direction of corporate diversification: Toward an integration of the resource-based view and transaction cost economics. *Management Science* **45**(8): 1109-1124.

Train, K. E. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press: Cambridge, U.K.

Figure 1: Framework to analyze agglomeration economies

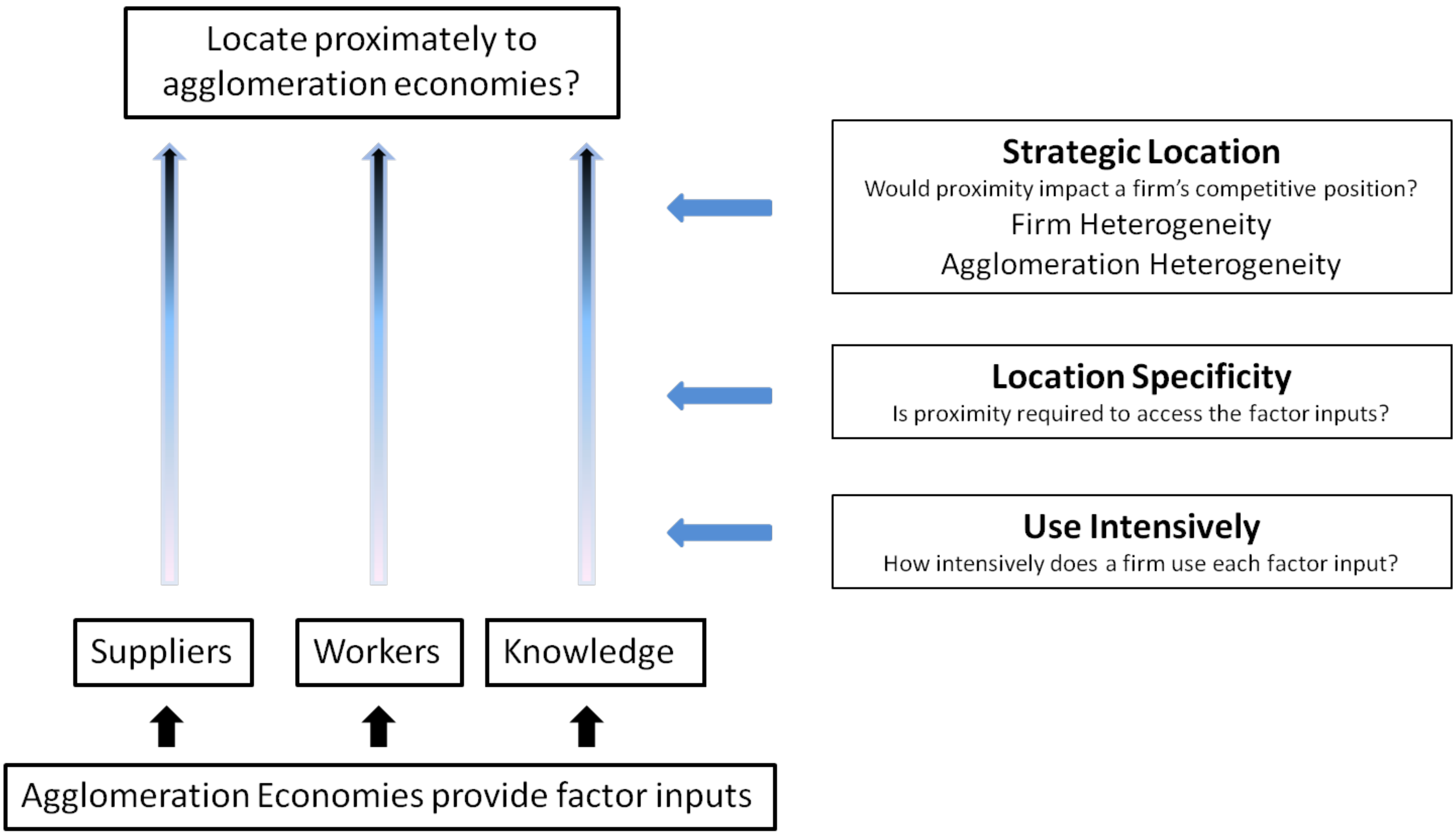

Figure 2: Location choice of manufacturing investments into the U.S., 1985-1994

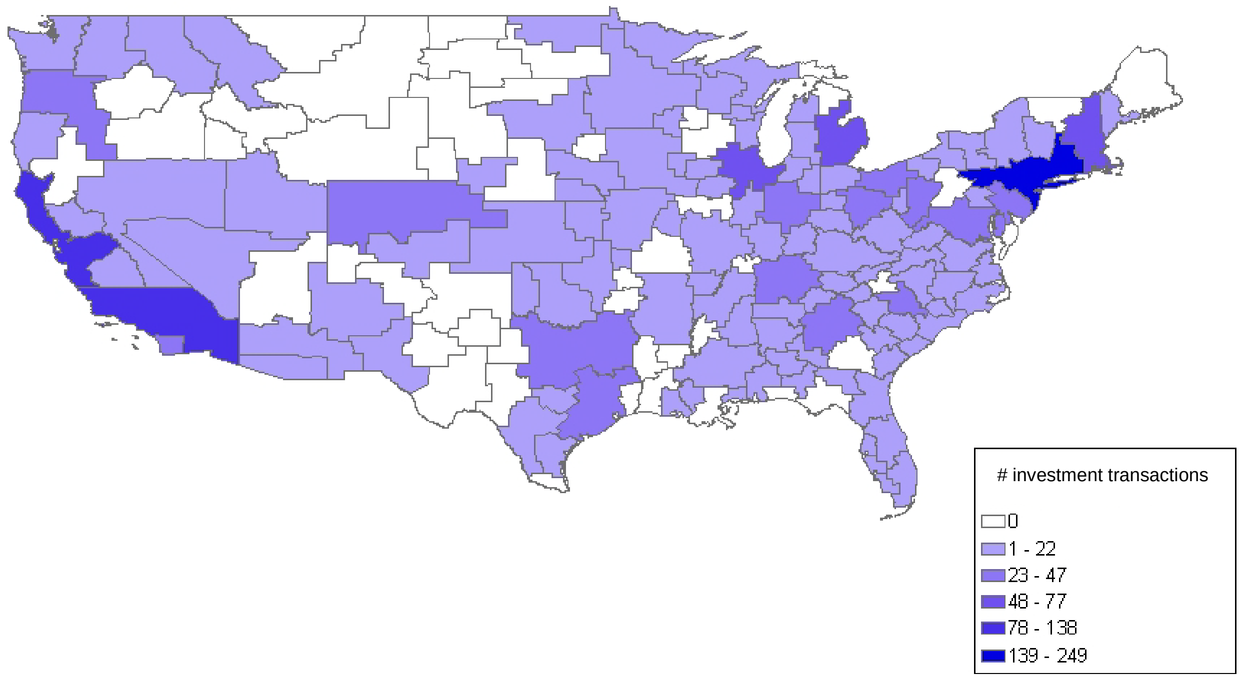

Table 1: Descriptive statistics

Descriptive statistics of economic areas characteristics (1985-1994 for 145 economic areas)*

	Variable	Source		N**	Mean	Std. Dev.	Min	Max
Location-specific variables	employment_growth	County Business Patterns (BEA)	percentage	72,790	0.361	23.155	-0.999	6094.0
	establishment_growth	County Business Patterns (BEA)	percentage	72,790	0.029	0.267	-0.833	26.0
	% buyers	County Business Patterns (BEA)	percentage	72,790	0.671	1.294	0.015	13.2
	% industry employment	County Business Patterns (BEA)	percentage	72,790	0.671	1.973	0.000	59.9
	% supplying industries	County Business Patterns (BEA)	percentage	72,790	0.672	1.261	0.000	19.3
	% industry patents (3-year stock)	Micropatent	percentage	72,790	0.635	3.192	0.000	100.0

* the continental US is composed of 170 economic areas; 25 economic areas do not receive any transactions
** 145 economic areas x 10 years x 233 industries = 337,850, but not all industries had entries in all years leaving 72,790

Descriptive statistics for foreign investment transactions (1985-1994)

	Variable	Source		N	Mean	Std. Dev.	Min	Max
Firm-specific variables	Economically Large	Worldscope, Compustat	dummy	657	0.732	0.443	0	1
	R&D Leader	Worldscope, Compustat	dummy	657	0.460	0.499	0	1

Table 2: Attractiveness of Agglomeration Economies
Conditional logit models of locations' attributes affecting likelihood of being chosen

Dependent variable: entry in 1 economic area	baseline	agglomeration economies				
	(1)	(2)	(3)	(4)	(5)	(6)
Econ. Area dummies	included	included	included	included	included	included
employment_growth	-0.1314 *	-0.1355 *	-0.1456 **	-0.1449 **	-0.1439 **	-0.1425 **
	(0.068)	(0.072)	(0.071)	(0.071)	(0.071)	(0.070)
	[0.877]	[0.873]	[0.865]	[0.865]	[0.866]	[0.867]
establishment_growth	0.1903	0.2364	0.2069	0.2029	0.2060	0.2027
	(0.166)	(0.171)	(0.167)	(0.167)	(0.167)	(0.167)
	[1.210]	[1.267]	[1.230]	[1.225]	[1.229]	[1.225]
%_industry_buyers	1.0884 ***	0.4979 ***	0.5050 ***	0.4590 ***	0.4517 ***	0.4507 ***
	(0.126)	(0.139)	(0.138)	(0.139)	(0.138)	(0.139)
	[2.970]	[1.645]	[1.657]	[1.582]	[1.571]	[1.569]
%_industry_employment		0.0958 ***	0.0914 ***	0.0877 ***	0.0867 ***	0.0862 ***
		(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
		[1.101]	[1.096]	[1.092]	[1.091]	[1.090]
%_ind_employment_zero			-1.4138 ***	-1.4050 ***	-1.4023 ***	-1.3849 ***
			(0.257)	(0.257)	(0.257)	(0.258)
			[0.243]	[0.245]	[0.246]	[0.250]
%_industry_suppliers				0.0886 ***	0.0876 ***	0.0857 ***
				(0.030)	(0.030)	(0.030)
				[1.093]	[1.092]	[1.089]
%_industry_patents					0.0091 *	0.0065
					(0.005)	(0.005)
					[1.009]	[1.007]
%_ind_patents_zero						-0.1633
						(0.125)
						[0.849]
Observations	95265	95265	95265	95265	95265	95265
no. of investments	657	657	657	657	657	657
no. of alternative choices	145	145	145	145	145	145
Log Likelihood	-2561.12	-2497.01	-2476.21	-2472.22	-2470.67	-2469.82
Improvement in Model Fit Test						
comparing		vs (1)	vs (2)	vs (3)	vs (4)	vs (5)
difference in Log Likelihood		64.11	20.80	3.99	1.55	0.86
additional d.o.f.		1	1	1	1	1
Chi-square test of ΔLogL		0.000 ***	0.000 ***	0.005 ***	0.079 *	0.191
Pseudo R-squared	0.217	0.236	0.243	0.244	0.244	0.245

standard error in parentheses, odds-ratio in square brackets
* significant at 10%; ** significant at 5%; *** significant at 1% for 2-tailed tests

Table 3: Strategic Response to Agglomeration Economies

Conditional logit models of locations' attributes affecting likelihood of being chosen

Dependent variable: entry in 1 economic area	baseline		sample split by firm economic size			sample split by firm R&D intensity	
	all firms		large firms		small firms		leaders		laggards	
	(1)		(2)		(3)		(4)		(5)	
Econ. Area dummies	included		included		included		included		included	
employment_growth	-0.1425	**	-0.1998	**	-0.0555		-0.2239	*	-0.0933	
	(0.070)		(0.096)		(0.094)		(0.124)		(0.079)	
	[0.867]		[0.819]		[0.946]		[0.799]		[0.911]	
establishment_growth	0.2027		0.2127		0.2309		0.1619		0.2563	
	(0.167)		(0.199)		(0.307)		(0.260)		(0.216)	
	[1.225]		[1.237]		[1.260]		[1.176]		[1.292]	
%_industry_buyers	0.4507	***	0.5229	***	0.2584		0.4148	**	0.4875	**
	(0.139)		(0.175)		(0.252)		(0.205)		(0.194)	
	[1.569]		[1.687]		[1.295]		[1.514]		[1.628]	
%_industry_employment	0.0862	***	0.0757	***	0.1284	***	0.0704	***	0.0971	***
	(0.008)		(0.009)		(0.018)		(0.013)		(0.011)	
	[1.090]		[1.079]		[1.137]		[1.073]		[1.102]	
%_ind_employment_zero	-1.3849	***	-1.3567	***	-1.4522	***	-1.3904	***	-1.3664	***
	(0.258)		(0.302)		(0.494)		(0.405)		(0.335)	
	[0.250]		[0.258]		[0.234]		[0.249]		[0.255]	
%_industry_suppliers	0.0857	***	0.1090	***	-0.0099		0.1084	**	0.0729	*
	(0.030)		(0.034)		(0.072)		(0.052)		(0.038)	
	[1.089]		[1.115]		[0.990]		[1.114]		[1.076]	
%_industry_patents	0.0065		-0.0027		0.0195	***	0.0023		0.0094	
	(0.005)		(0.008)		(0.008)		(0.009)		(0.007)	
	[1.007]		[0.997]		[1.020]		[1.002]		[1.009]	
%_ind_patents_zero	-0.1633		-0.2970	**	0.1833		-0.3011		-0.0610	
	(0.125)		(0.147)		(0.248)		(0.191)		(0.169)	
	[0.849]		[0.743]		[1.201]		[0.740]		[0.941]	
Observations	95265		69745		25520		43790		51475	
no. of investments	657		481		176		302		355	
no. of alternative choices	145		145		145		145		145	
Log Likelihood	-2469.82		-1793.06		-615.37		-1107.96		-1286.44	
Pseudo R-squared	0.245		0.251		0.297		0.263		0.272	

standard error in parentheses, odds-ratio in square brackets

* significant at 10%; ** significant at 5%; *** significant at 1% for 2-tailed tests

Table 4: Attractiveness of Agglomeration Economies - Split by Industry Intensity
Conditional logit models of locations' attributes affecting likelihood of being chosen

Dependent variable: entry in 1 economic area	baseline all industries	sample split by industry labor intensity		sample split by industry supplier intensity		sample split by industry R&D intensity	
		high use	low use	high use	low use	high R&D	low R&D
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Econ. Area dummies	included	included	included	included	included	included	included
employment_growth	-0.1425 **	-0.1700	-0.1194	-0.1184	-0.1624	-0.1674 *	-0.1127
	(0.070)	(0.115)	(0.089)	(0.099)	(0.100)	(0.100)	(0.097)
	[0.867]	[0.844]	[0.887]	[0.888]	[0.850]	[0.846]	[0.893]
establishment_growth	0.2027	0.1846	0.2423	0.0767	0.3063	0.1219	0.2858
	(0.167)	(0.261)	(0.224)	(0.259)	(0.217)	(0.227)	(0.244)
	[1.225]	[1.203]	[1.274]	[1.080]	[1.358]	[1.130]	[1.331]
%_industry_buyers	0.4507 ***	0.4763 *	0.3845 **	0.5658 ***	0.4477 *	0.4704 ***	0.0983
	(0.139)	(0.269)	(0.176)	(0.193)	(0.230)	(0.164)	(0.290)
	[1.569]	[1.610]	[1.469]	[1.761]	[1.565]	[1.601]	[1.103]
%_industry_employment	0.0862 ***	0.0979 ***	0.0756 ***	0.0650 ***	0.0972 ***	0.0979 ***	0.0673 ***
	(0.008)	(0.012)	(0.012)	(0.013)	(0.011)	(0.011)	(0.014)
	[1.090]	[1.103]	[1.079]	[1.067]	[1.102]	[1.103]	[1.070]
%_ind_employment_zero	-1.3849 ***	-1.1669 ***	-1.4973 ***	-1.8785 ***	-0.9978 ***	-1.4853 ***	-1.2530 ***
	(0.258)	(0.386)	(0.347)	(0.436)	(0.325)	(0.381)	(0.353)
	[0.250]	[0.311]	[0.224]	[0.153]	[0.369]	[0.226]	[0.286]
%_industry_suppliers	0.0857 ***	0.0354	0.1883 ***	0.1504 ***	0.0554	0.0423	0.1683 ***
	(0.030)	(0.040)	(0.056)	(0.056)	(0.038)	(0.046)	(0.047)
	[1.089]	[1.036]	[1.207]	[1.162]	[1.057]	[1.043]	[1.183]
%_industry_patents	0.0065	-0.0069	0.0174 **	0.0177 *	0.0006	0.0059	0.0154
	(0.005)	(0.010)	(0.007)	(0.010)	(0.007)	(0.006)	(0.014)
	[1.007]	[0.993]	[1.018]	[1.018]	[1.001]	[1.006]	[1.016]
%_ind_patents_zero	-0.1633	-0.2798	-0.0124	-0.0092	-0.2594	-0.1916	-0.1757
	(0.125)	(0.192)	(0.170)	(0.183)	(0.176)	(0.175)	(0.185)
	[0.849]	[0.756]	[0.988]	[0.991]	[0.772]	[0.826]	[0.839]
Observations	95265	43500	51765	42485	52780	58000	37265
no. of investments	657	300	357	293	364	400	257
no. of alternative choices	145	145	145	145	145	145	145
Log Likelihood	-2469.82	-1072.57	-1322.31	-1072.85	-1332.17	-1448.11	-920.72
Pseudo R-squared	0.245	0.282	0.256	0.264	0.265	0.273	0.280

standard error in parentheses, odds-ratio in square brackets
* significant at 10%; ** significant at 5%; *** significant at 1% for 2-tailed tests