NBER WORKING PAPER SERIES

THE GLOBAL AGGLOMERATION OF MULTINATIONAL FIRMS

Laura Alfaro
Maggie Chen

Working Paper 15576
http://www.nber.org/papers/w15576

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2009

We thank Jim Anderson, Bruce Blonigen, Jim Harrigan, Keith Head, Tarun Khanna, Keith Maskus, Mike Moore, John Ries, Roberto Samaneigo, Stephen Yeaple, Tony Yezer, and seminar and conference participants at Harvard Business School, the University of Virginia Darden School of Business, George Washington Univeristy, the AEA meeting, the ETSG plenary session, the EIIT meeting, and the LACEA Trade, Integration and Growth Meeting for helpful comments and suggestions. We also thank William Kerr for kindly providing us the patent concordance data, Bill Simpson for excellent advice in calculating the agglomeration indexes, and Francisco Pino for help with the GIS software. We are grateful to Dun & Bradstreet and Dennis Jacques for helping us with the D&B data set and HBS and GW for financial support. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

The Global Agglomeration of Multinational Firms
Laura Alfaro and Maggie Chen
NBER Working Paper No. 15576
December 2009 Revised September 2010
JEL No. D2,F2,R1,R3

ABSTRACT

The proliferation of multinational activities has led to the emergence of new industrial clusters around the world. In this paper, we examine how "first nature" location fundamentals and "second nature" agglomeration economies jointly determine the global landscape of multinational firms. Using a unique worldwide plant dataset that reports detailed location, ownership, and operation information for plants in more than 100 countries, we construct a spatially continuous index of global agglomeration and investigate the patterns and determinants of clustering between multinational firms. Our analysis indicates that multinationals' agglomeration goes above and beyond first-nature driven geographic concentration. Second-nature forces including knowledge spillovers, capital-market externalities, and vertical production linkages play a significant role. In comparison to domestic plants, knowledge spillovers and capital market externalities exert a stronger e¤ect on the clustering of multinational firms while labor market pooling has a weaker impact. These findings remain robust when we examine entry decisions and explore the process of agglomeration.

Laura Alfaro
Harvard Business School
Morgan Hall 263
Soldiers Field
Boston, MA 02163
and NBER
lalfaro@hbs.edu

Maggie Chen
Dept. of Economics
George Washington University
2115 G ST, NW, #367
Washington, DC 20052
xchen@gwu.edu

1 Introduction

The explosion of multinational corporation (MNC) activities in recent decades has precipitated the emergence of new industrial clusters around the world. Firms that agglomerated in, for example, Silicon Valley and Detroit now have subsidiary plants clustering in Bangalore and Slovakia (termed, respectively, the Silicon Valley of India and Detroit of the East). But are these clusters the rule or the exception? Do multinationals, known for their global resource and market seeking activities, agglomerate and, if so, what motivates their agglomeration?

The extensive empirical literature on foreign direct investment (FDI) led by, for example, Brainard (1997), Carr, Markusen, and Maskus (2001), Yeaple (2003), and Head and Mayer (2004a), stresses the importance of market access and comparative advantage to multinationals' foreign location decisions. Their evidence suggests that multinationals' foreign activities are motivated by "first nature" incentives to access markets and reduce production costs. A separate strand of literature in economic geography, focused on domestic industrial landscape in countries such as the U.S. and the U.K., finds that agglomeration economies, also referred to as "second nature" including proximity to customers and suppliers, labor market pooling, and knowledge spillovers, play a significant role in domestic firms' location choices within a given country (see, e.g., Ellison and Glaeser, 1999; Rosenthal and Strange, 2001; Duranton and Overman, 2005, Ellison, Glaeser and Kerr, 2009).[1] But how the "first nature" fundamentals in FDI and "second nature" agglomeration forces in domestic economic geography jointly explain the worldwide location decisions of multinational firms remains unknown.

In this paper we examine the significance and determinants of the global agglomeration of multinational firms. We investigate: first, do multinationals agglomerate with one another around the world? Second, if they do agglomerate, is the agglomeration driven by shared, first-nature incentives such as market access and comparative advantages or do second-nature forces play a role as well? Finally, are multinationals' overseas clusters a simple projection of their headquarters clusters at home, and are the effects of agglomeration economies similar between multinational and domestic manufacturing firms?

Answers to these questions are central to academic and policy debates centered on foreign direct investment (FDI). Growing evidence suggests that multinationals play a significant role in the performance of local economies. Compared to domestic firms in the same country and industry, multinationals tend to have greater sales, employment and capital and higher productivity (see, Helpman, Melitz and Yeaple, 2004). Given these characteristics, they are found

[1] A number of firm-level studies in international trade including, for example, Head, Ries and Swenson (1995), Head and Mayer (2004a), Crozet, Mayer and Mucchielli (2004), Blonigen, Ellis and Fausten (2005), Bobonis and Shatz (2007), Amiti and Javorcik (2008), Debaere, Lee and Park (2010), have examined the role of vertical production linkages in MNCs' agglomeration decisions in host countries. These studies offer importance evidence on the incentives of vertically linked MNCs to cluster within a host country. Our analysis builds on these studies. Specifically, it evaluates the importance of a variety of agglomeration forces in explaining the geographic distribution of MNCs. Section 2 provides a detailed discussion of related literature.

to raise local wages (see, e.g., Aitken, Harrison and Lipsey, 1996) and generate productivity spillovers (see, e.g., Javorcik, 2004). On the other hand, multinationals are by nature prone to adjusting location choices in response to changing local market conditions. Recognizing these benefits and risks, many countries including both host and home countries have long offered lucrative incentives to MNCs in the hope of building and sustaining FDI clusters. Understanding the interdependencies of multinational firms and how the movement of one MNC influences the movements of others is critical to such policy makings.

The first goal of this paper is to establish the global agglomeration patterns of multinational firms. We focus on patterns of agglomeration between industries, also referred to as coagglomeration in the urban economics literature (see, e.g., Ellison, Glaeser and Kerr, 2009). This enables us to disentangle the effects of first-nature location fundamentals and different second-nature agglomeration forces, a goal that is difficult to achieve with within-industry agglomeration information as plants in the same industry often exhibit all dimensions of agglomeration incentives.

We use a unique worldwide manufacturing plant dataset, WorldBase, that provides detailed location, ownership, and activity information for establishments in more than 100 countries. We obtain precise latitude and longitude information for each establishment in the dataset and compute the actual distance (and trade cost) between each establishment pair. Our data indicates that a significant fraction of MNC agglomeration occurs across borders. For example, among multinationals located within 200 km (approximately the distance between New York and Washington, DC), more than 20 percent of establishment pairs are in two different countries. For multinationals within 400 km, the percentage increases to 45 percent.[2] This observation suggests that it is important to examine agglomeration using cross-country datasets and, moreover, not restrict the definition of agglomeration by administrative borders. To achieve this goal, we construct a spatially continuous index of agglomeration based on pairwise distances between plants. We adopt a new empirical methodology from urban economics introduced by Duranton and Overman (2005) (henceforth, DO) in which we estimate actual geographic distributions of MNC plants and compare them to distributions of counterfactuals. The constructed index enables us to identify the significance and extent of MNC agglomeration between industries.

Having identified global agglomeration patterns of MNCs, we investigate the role of first-nature incentives and second-nature forces in multinationals' clustering decisions and, further, how the importance of these factors varies between MNCs and domestic firms. We evaluate the effect of first-nature location fundamentals by obtaining expected geographic distributions of multinationals predicted exclusively by factors such as market size, comparative advantage, and trade costs. Controlling for the first-nature predicted distributions, we then examine the degree to which second-nature agglomeration forces formalized in Marshall's theories of industrial agglomeration, including not only vertical production linkages, the force most documented, but also labor- and capital-market externalities and knowledge spillovers, explain the extent of

[2] See Table A.1.

agglomeration among multinational firms.

Our analysis indicates that first-nature economic characteristics, although they play a significant role in explaining the agglomeration of multinational firms, are not the only driving forces. Multinationals' agglomeration goes above and beyond first-nature driven geographic concentration. In addition to their common motivation to gain market access and comparative advantage, multinationals exhibit significant incentives to cluster with one another around the world. Second-nature forces including not only vertical production linkages but also knowledge spillovers and externalities in capital markets, forces not previously considered for multinationals, all exert an important effect. Labor market pooling has little effect on the clustering of individual subsidiaries, but matters for the agglomeration of MNC subsidiary employment. These results point out that despite the sharp declines in trade and telecommunication costs in recent decades, even the most active and mobile participants of globalization exhibit significant localization incentives.

When comparing their relative importances, we find, for the agglomeration of individual subsidiaries, the effect of first-nature incentives to be pronounced mostly at the aggregate geographic level and second-nature agglomeration economies to be the driving force at disaggregated geographic levels. For the agglomeration of MNC subsidiary employment, both natures have a significant effect at all geographic levels, with the impact of first nature exceeding the cumulative impact of second-nature agglomeration forces. A one-standard-deviation increase in the former is associated with a 0.31 standard-deviation increase in employment agglomeration at 200 km, whereas the cumulative effect of second-nature agglomeration economies is around 0.17. Across agglomeration economies, we find, in accordance with the physical- and knowledge-capital intensity of multinational firms, knowledge spillovers and capital-market externalities to have a particularly important effect on the clustering of MNCs. The finding persists when we compare the agglomeration of multinational firms with that of domestic counterparts in the same pairwise industries. Multinationals exhibit sharply different agglomeration motives compared to domestic firms: the former are relatively more influenced by knowledge- and physical-capital considerations and less by labor market pooling.

The importance of agglomeration forces also varies between multinational headquarters and subsidiaries. Our results suggest that as headquarters and subsidiaries become increasingly specialized at different tasks within the boundary of firms, they follow distinct agglomeration patterns. All agglomeration forces except vertical production linkages exert a significant effect on the clustering of MNC headquarters, whereas all factors including vertical production linkages play a role in the clustering of MNC subsidiaries. These results are consistent with the growing tendency of MNC headquarters to specialize in human- and physical-capital intensive activities such as management, marketing, and R&D services and the tendency of MNC subsidiaries to concentrate on production and sourcing activities.

The role of second-nature forces remains robust when we explore the dynamics of multinational-firm agglomeration. Examining the entry decisions of multinational firms enables us to look at not just the pattern, but also the process, of agglomeration and identify the causal effect of agglomeration forces. It permits us to address two potential econometric concerns that can arise in the analysis, (i) potential reverse causality between economic fundamentals and agglomeration patterns and (ii) omitted variable bias. Our results suggest that relative to domestic firms multinational entrants display stronger propensities to cluster with incumbent multinationals. This, again, is especially true when there are relatively strong knowledge spillover benefits, capital-market externalities, and vertical production linkages.

Our paper offers, to our knowledge, the first empirical analysis on the global patterns and determinants of multinational-firm agglomeration. Using a worldwide dataset of domestic and MNC establishments enables us to analyze the world landscape of multinational production unconstrained by geographic region and expand the definition of spatial agglomeration beyond political borders. Activities separated by administrative boundaries but proximate in actual distance (and trade cost) are taken into account. As described earlier, this type of activities constitute a significant fraction of MNC agglomeration. Limiting analysis to a particular country or region can omit these activities and mis-measure agglomeration.

The distinct feature of our dataset is complemented by the spatially continuous index adopted in our study. Previous indices have tended to treat space as discrete areas and define agglomeration as activities (the number of firms or the size of production) located within a given region. As a result, the accuracy of these indices is critically dependent on the definition and scale of geographic units. Firms can be sparsely located within a large geographic unit but considered agglomerative; conversely, firms in two different geographic units can be proximate in actual distance but considered dispersed. In this study, we use the actual distance between each pair of establishments to construct an agglomeration index that is independent of the level of geographic disaggregation.

Finally, by accounting for both first-nature location fundamentals and second-nature agglomeration forces, our paper is able to systematically evaluate their relative importances in MNCs' global location decisions. Our results suggest that when analyzing the location patterns of multinational firms, it is critical to take into account both natures including agglomeration economies such as knowledge spillovers and capital-market externalities. Policies that affect the degree of these agglomeration economies can have a potentially important impact on multinationals' location decisions.[3]

The rest of the paper is structured as follows. Section 2 reviews the related literature. Section 3 describes the data. Section 4 explains the methodology used to construct pairwise-industry agglomeration indices and Section 5 the construction of control variables. Section 6 reports main

[3] Note, however, that the process of Marshallian externalities is complex and difficult to replicate. See Harrison and Rodríguez-Clare (2009) for a critical evaluation of industrial policy. More broadly, our work also relates to the macro-level evidence on the role of complementarities in maximizing the benefits of FDI (see Alfaro et al., 2004).

empirical evidence. Section 7 presents further econometric analysis. The last section concludes.

2 Related Literature

This study is related to three broad strands of research analyzing the location decisions of firms. Two main categories of factors have been stressed by these literatures in explaining the structures of production, trade, and multinational activities. The first category is "first nature" factors (Krugman, 1993), also known as "natural advantage" (Ellison and Glaeser, 1999) and "location fundamentals" (Davis and Weinstein, 2002). In the economic geography literature, it relates to natural resource endowments and physical geography; in the factor-proportion trade theory and knowledge-capital FDI literature, it relates to factor endowments and trade costs. The second category is "second nature" factors, also known as "agglomeration economies". It refers to benefits of geographic proximity between economic agents including proximity between customers and suppliers, firms sharing labor-market externalities, and firms with knowledge spillovers. A thorough review of these large and diverse literatures is beyond the scope of this paper. We limit discussion to studies of particular relevance to this paper.

First, we build on the broader literature that examines the first-nature motives of FDI. Theoretical studies in this area, led by Markusen (1984), Helpman (1984), and Markusen and Venables (1998), have established two primary motives for multinationals' investment abroad: market access and comparative advantage. These motives are synthesized in the knowledge-capital model (see, e.g., Carr, Markusen and Maskus, 2001) and tested in a number of empirical studies. While Brainard (1997) finds evidence in favor of the market access motive, empirical analyses by Carr, Markusen and Maskus (2001), Yeaple (2003), Head and Mayer (2004a), and Alfaro and Charlton (2009) indicate the existence of both types of investments.

Second, a number of studies in international trade including Head, Ries and Swenson (1995), Head and Mayer (2004a), Crozet, Mayer and Mucchielli (2004), Blonigen, Ellis and Fausten (2005), Bobonis and Shatz (2007), Amiti and Javorcik (2008), and Debaere, Lee and Park (2010) have examined the role of distance and linkages in individual multinationals' location decisions. The results of these studies, which suggest that MNCs with vertical linkages tend to agglomerate within a host country/region, shed light on the role of vertical production relationship in multinationals' location choices.

Head, Ries and Swenson (1995) and Blonigen, Ellis and Fausten (2005) exploit large Japanese industrial groupings (keiretsu) and examine the location interdependence of vertically and horizontally linked Japanese plants. Their evidence suggests that members of the same keiretsu tend to choose the same states in the United States. Head and Mayer's (2004a) study of the location choices of Japanese firms in Europe finds regions with larger numbers of existing foreign affiliates to be more likely to be selected by multinationals. Crozet, Mayer and Mucchielli (2004) and Bobonis and Shatz (2007) study the determinants of location choices by foreign investors in France and in the U.S., respectively, and find similar evidence of clustering. Recent work by Amiti and Javorcik (2008) and Debaere, Lee and Park (2010) that examines the entry decisions of foreign multinationals in China also shows market and supplier accesses to be important factors in location choices.

A third strand of related studies is the urban economics literature that examines patterns of domestic agglomeration. A number of papers in this area, including Ellison and Glaeser (1997), Rosenthal and Strange (2001), DO, and Ellison, Glaeser and Kerr (2009), assess the importance of agglomeration forces in the industrial localization in the U.S. and U.K. Ellison and Glaeser (1997) propose an index of concentration to measure the level of spatial concentration within each subnational administrative unit such as county and state. This index takes into account the effect of industrial concentration in each industry, an issue noted to affect the accuracy of previous indices. Rosenthal and Strange (2001) adopt the Ellison and Glaeser (1997) index and evaluate the importance of agglomeration forces in explaining the localization of U.S. industries. Regressing the new index on a large number of industry characteristics, Rosenthal and Strange (2001) find both labor market pooling and input-output linkages to have a positive impact on agglomeration. The effect of knowledge spillovers is also significant, but mostly at the local level. The recent study by DO extends this literature by developing a spatially continuous concentration index that is independent of the level of geographic disaggregation (see Section 4 for a detailed description). Ellison, Glaeser and Kerr (2009), applying both this and the Ellison and Glaeser (1997) index, evaluate the coagglomeration of U.S. industries and find, similar to Rosenthal and Strange (2001), a particularly important role for input-output relationships and labor market pooling.

3 Data

3.1 The WorldBase database

Our empirical analysis employs WorldBase, a new worldwide establishment dataset complied by Dun & Bradstreet (D&B) that covers more than 43 million public and private establishments in over 100 countries and territories.

Dun & Bradstreet is the leading source of commercial credit and marketing information since 1845.[4] D&B presently operates in more than a dozen countries and territories either directly or through affiliates, agents, and associated business partners, and compiles data from a wide range of sources including public registries, partner firms, telephone directory records, and websites. All information is verified centrally via a variety of manual and automated checks. Early uses of D&B data include Caves' (1975) size and diversification pattern comparisons between Canadian and U.S. domestic plants and subsidiaries of U.S. multinationals in Canada, and Lipsey's (1978)

[4] For more information, see: http://www.dnb.com/us/about/db_database/ dnbinfoquality.html.

comparisons of the D&B data with existing sources with regard to the reliability of U.S. data. More recently, Harrison, Love, and McMillian (2004) use D&B's cross-country foreign ownership information. Other research that has used D&B data includes Black and Strahan's (2002) study of entrepreneurial activity in the U.S., Rosenthal and Strange's (2003) analysis of micro-level agglomeration in the U.S., Acemoglu, Johnson, and Mitton's (2009) cross-country study of concentration and vertical integration, and Alfaro and Charlton's (2009) analysis of vertical and horizontal activities of multinationals.

WorldBase reports for each establishment in the dataset detailed information on location, ownership, and activities. Four categories of information are used in this paper: (i) industry information including the four-digit SIC code of the primary industry in which each establishment operates; (ii) ownership information including headquarters, domestic parent, global parent, status (joint venture, corporation, partnership), and position in the hierarchy (branch, division, headquarters); (iii) detailed location information for both establishment and headquarters; and (iv) operational information including sales and employment, and year started.

Our main empirical analysis is based on manufacturing establishments in 2005, including both MNC and domestic plants. An establishment is deemed MNC owned if it satisfies two criteria: (i) it reports to a global parent firm, and (ii) the headquarters or parent firm is located in a different country. A parent is defined as an entity that has legal and financial responsibility for another establishment.[5] We drop establishments with zero or missing employment values and industries with fewer than 10 observations.[6]

There are in total $32,427$ MNC owned manufacturing plants in our final sample. Top industries include Electronic Components and Accessories (367), Miscellaneous Plastics Products (308), Motor Vehicles and Motor Vehicle Equipment (371), General Industrial Machinery and Equipment (356), Laboratory Apparatus and Analytical, Optical, Measuring, and Controlling Instruments (382), Drugs (283), Metalworking Machinery and Equipment (354), Construction, Mining, and Materials Handling (353), and Special Industry Machinery except Metalworking (355). Among the top host countries are China, the U.S., the U.K., Canada, France, Poland, Czech Republic, and Mexico.

To examine the coverage of our MNC establishment data, we compared U.S. owned subsidiaries in the WorldBase database with the U.S. Bureau of Economic Analysis' (BEA) Direct Investment Abroad Benchmark Survey, a legally mandated confidential survey conducted every five years that covers virtually the entire population of U.S. MNCs. The comparison shows that

[5] There are, of course, establishments that belong to the same multinational family. Although separately examining the interaction of these establishments is beyond the focus of this paper, we expect the Marshallian forces to have a similar effect here. For example, subsidiaries with an input-output linkage should have incentives to locate near one another independent of ownership. See Yeaple (2003) for theoretical work in this area, and Chen (2009b) for supportive empirical evidence. One can use a similar methodology (estimating geographic distributions of establishments that belong to the same firm and comparing them with distributions of counterfactuals) to study intra-firm interaction (see Duranton and Overman, 2008).

[6] Requiring positive employment helps to exclude establishments registered exclusively for tax purposes.

the two databases have similar accounts of establishments and activities across countries and industries. We also compared WorldBase with UNCTAD's Multinational Corporation Database. The two databases differ in that the former reports at the plant level and the latter at the firm level. For the U.S. and other major FDI source countries, the number of firms is similar between the two databases, but WorldBase contains more plants. See Alfaro and Charlton (2009) for a more detailed discussion of the WorldBase data and comparisons with other data sources.

Conducting our empirical analysis for all domestic manufacturing plants is infeasible given the size of the dataset and computational intensity of the empirical procedure (described in detail in Section 4). To keep the analysis feasible, we adopt a random sampling strategy. For each SIC 3-digit industry with more than 1000 observations, we obtain a random sample of 1,000 plants. For industries with fewer than 1,000 observations, we include all domestic plants. This results in a final sample of 127,897 domestically owned plants. A similar random sampling strategy was used in Ellison, Glaeser and Kerr (2009).

3.2 Global coverage and geocode information

D&B's WorldBase is, in our view, an ideal data source for the research question proposed in this study. It offers two distinct advantages compared to alternative data sources used in previous studies. First, WorldBase's global coverage makes it possible to examine agglomeration on a global and continuous scale. As described earlier, a significant fraction of geographic agglomeration occurs across country borders. Table A.1 shows that more than 20 percent of pairs of multinationals located within 200 km (approximately the distance between New York City and Washington, DC) are in two different countries. The percentage rises to 45 percent at 400 km and 70 percent at 800 km. This is not surprising given countries' increasing integration in regional trading blocs. A growing number of clusters have formed, for example, on the borders of EU member countries and the U.S.-Mexico border. This means that limiting analysis to a particular country can omit a significant share of agglomeration.

Second, the WorldBase database reports the physical address and postal code of each plant while most existing datasets report business registration addresses. The physical location information enables us to obtain precise latitude and longitude information for each plant in the data and compute the distance (and, more generally, trade cost) between each establishment pair. Existing studies have tended to use distance between administrative units, such as state distances, as a proxy for distance of establishments. In doing so, establishments proximate in actual distance but separated by administrative boundaries (e.g., San Diego and Phoenix) can be considered dispersed. Conversely, establishments far in distance but located in the same administrative unit (e.g., San Diego and San Francisco) can be counted as agglomeration.

We obtain latitude and longitude codes for each establishment using a geocoding software (GPS Visualizer). This software uses Yahoo's and Google's Geocoding API services, well known as the industry standard for transportation data. It provides more accurate geocode information than most alternative sources. The geocodes are obtained in batches and verified for precision. We apply the Haversine formula to the geocode data to compute the great-circle distance between each pair of establishments. The distance information is then used to construct an index of agglomeration following the empirical methodology described below.

4 Constructing the Agglomeration Index

As noted by Head and Mayer (2004b), measurement of agglomeration is a central challenge in the economic geography literature. Continuous effort has been devoted to designing an index that accurately reflects the agglomeration of economic activities. The latest progress in the literature is Duranton and Overman (2005).

4.1 Issues in constructing the agglomeration indices

Most existing indices have tended to equalize agglomeration with activities located in the same administrative or geographic region (measured by number of firms or size of production in the region). Three issues arise with such measures. First, these indices can be strongly driven by industrial concentration. Industries with a small number of establishments may appear agglomerative when they are not. Second, many indices cannot separate general geographic concentration of the manufacturing industry due to location attractiveness from agglomeration. Manufacturing plants attracted to the same location because of location characteristics can be interpreted as agglomeration. The index developed by Ellison and Glaeser (1997) resolves these issues.

Duranton and Overman (2005) address the unresolved issue of the dependence of existing measures on the level and method of geographic disaggregation by developing a "continuous-space concentration index". Previous indices, by equating agglomeration with activities in the same region, omit agglomerating activities separated by administrative or geographic borders while overestimating the degree of agglomeration within the same administrative or geographic units.

DO's (2005) index addresses these issues and exhibits five important properties essential to agglomeration measures. First, the index is comparable across industries and captures cross-industry variation in the level of agglomeration. Second, it controls for industrial concentration in each industry. Third, the index is constructed based on a counterfactual approach and controls for the effect of location factors such as market size, natural resources, and policies that apply to all manufacturing plants. Fourth, by taking into account spatial continuity, the index is unbiased with respect to the scale and aggregation of geographic units. Finally, the index offers an indication of the statistical significance of agglomeration.

DO construct this index to measure the significance of same-industry agglomeration in the U.K. We expand this index to a global context and measure the degree of between-industry agglomeration around the world. Because it accounts for the continuity in space, the index offers an ideal measure for cross-country studies. We also extend the original index's focus on distance as the main form of trade cost to a measure of global agglomeration that accounts for various forms of trade costs.

There are two requirements for the construction of this index. First, availability of physical location information for each establishment at the most detailed level. The WorldBase dataset, supplemented by a geocoding software, satisfies this requirement. Second, as described below, the empirical procedure adopted to construct the index uses a simulation approach that is extremely computationally intensive, especially for cross-country studies and large datasets.

4.2 Empirical procedure

The empirical procedure consists of three steps. Given our interest in comparing global location patterns across different types of establishments, that is, MNC headquarters versus MNC subsidiaries and multinational versus domestic plants, we repeat the procedure for each type of establishment.

Step 1: Kernel estimator We first estimate an actual geographic distribution function for each pair of industries using the distance data. Note that although the locations of nearly all establishments in our data are known with a high degree of precision, distance is an approximation of the true physical distance between establishments. One source of systematic error, for example, is that journey times for any given distance might differ between low- and high-density areas. Given the potential noise in the measurement of distances, we follow DO in adopting kernel smoothing when estimating the distribution function of distance. In the appendix, we consider a generalized measure of trade cost to take into account the effects of tariff, language, and border. We follow the same procedure described here and construct measures of agglomeration that account for other forms of trade costs.

Let d_{ij} denote the distance between establishment i and j. For each industry pair k and $\widetilde{k}$, we obtain a kernel estimator of bilateral distances at any point d (i.e., $f_{k\widetilde{k}}(d)$):

$$f_{k\widetilde{k}}(d) = \frac{1}{n_k n_{\widetilde{k}} h} \sum_{i=1}^{n_k} \sum_{j=1}^{n_{\widetilde{k}}} K\left(\frac{d \square d_{ij}}{h}\right), \tag{1}$$

where n_k and $n_{\widetilde{k}}$ are the number of plants in industries k and $\widetilde{k}$, respectively, h is the bandwidth, and K is the kernel function. We use Gaussian kernels with the bandwidth set to minimize the mean integrated squared error. This step generates $7,875$ kernel estimators for the $7,875$ $(= 126 \times 125/2)$ manufacturing industry pairs in our data.[7]

[7]Since we focus on between-industry agglomeration (i.e., industry coagglomeration) to disentangle the effects of different agglomeration forces, same-industry pairs (at SIC 3 digit level) are excluded in the analysis. The level

In addition to estimating the geographic distribution of establishment pairs, we can also treat each worker as the unit of observation and measure the level of agglomeration among workers. To proceed, we obtain a weighted kernel estimator by weighing each establishment by employment size. This is given by

$$f^{w}_{k\widetilde{k}}(d) = \frac{1}{h\sum_{i=1}^{n_k}\sum_{j=1}^{n_{\widetilde{k}}}(r_i r_j)}\sum_{i=1}^{n_k}\sum_{j=1}^{n_{\widetilde{k}}} r_i r_j K\left(\frac{d - d_{ij}}{h}\right) \tag{2}$$

where r_i and r_j represent, respectively, the number of employee in establishments i and j. We do this for each of the $7,875$ industry pairs.

Step 2: Counterfactuals and global confidence bands Next, we obtain counterfactual estimators. This step obtains the geographic distribution of the manufacturing industry as a whole, making it possible to control for factors that affect all manufacturing plants. We proceed by drawing, for each of the 7,875 industry pairs, 1,000 random samples each of which includes two counterfactual industries. Note that to control for the potential effect of industry concentration, it is important that the counterfactual industry in each sample have a number of observations similar to the actual data. We then calculate the bilateral distance of each pair of establishments and obtain a kernel estimator, either unweighted or weighted by employment, for each of the 7,875,000 samples. This gives us $1,000$ kernel estimators for each of the $7,875$ industry pairs.[8]

To identify agglomeration, we compare the actual and counterfactual kernel estimators at various distance thresholds T. Several distance thresholds, including 200, 400, 800, and 1,600 kilometers (the maximum threshold is roughly the distance between Detroit and Dallas and between London and Lisbon), are considered. We compute the 95% global confidence band for each threshold distance. Following DO, we choose identical local confidence intervals at all levels of distance such that the global confidence level is 5%. We use $\overline{f}_{k\widetilde{k}}(d)$ to denote the upper global confidence band of industry pair k and $\widetilde{k}$. When $f_{k\widetilde{k}}(d) > \overline{f}_{k\widetilde{k}}(d)$ for at least one $d \in [0, T]$, the industry pair is considered to agglomerate within distance T and exhibit greater agglomeration than counterfactuals. Graphically, it is detected when the kernel estimates of the industry pair lie above its upper global confidence band.

Step 3: Agglomeration index We now construct the agglomeration index. For each industry pair k and $\widetilde{k}$, we obtain

$$agglomeration_{k\widetilde{k}}(T) \equiv \sum_{d=0}^{T} \max\left(f_{k\widetilde{k}}(d) - \overline{f}_{k\widetilde{k}}(d), 0\right) \tag{3}$$

of industry disaggregation in our analysis is dominated by the availability of control variables, as we explain in Section 5.

[8] The Monte-Carlo nature of this approach makes it computationally intensive, especially given our worldwide dataset and focus on between-industry agglomeration. Repeating the procedure each time (as we examine, respectively, MNC headquarters, subsidiaries, subsidiary employment, and domestic plants) requires approximately one month of computing time utilizing 2 quad core 3.00 GHz processors and Windows 64-bit systems.

or employment-weighted

$$agglomeration^{w}_{k\widetilde{k}}(T) \equiv \sum\nolimits_{d=0}^{T} \max\left(f^{w}_{k\widetilde{k}}(d) \,\square\, \overline{f}^{w}_{k\widetilde{k}}(d), 0 \right). \tag{4}$$

The index measures the extent to which establishments in industries k and $\widetilde{k}$ agglomerate within the threshold distance T and the statistical significance thereof. When the index is positive, the level of agglomeration between industries k and $\widetilde{k}$ is significantly different from that of counterfactuals.

5 Determinants of Multinational Firm Agglomeration

After constructing the agglomeration index, we discuss the various factors that can lead to agglomeration and how each is measured in the empirical analysis. As described in Section 2, the location decision of multinational firms can be viewed as a function of two categories of factors. The first consists of first-nature location fundamentals that motivate firms to invest in a given location, namely, market access and comparative advantage; the second consists of second-nature agglomeration forces including (i) proximity to suppliers and customers, (ii) external scale economy in factor markets, and (iii) knowledge spillovers.[9]

5.1 First nature: location fundamentals

The theoretical FDI literature has identified two main motives to explain firms' foreign investment decisions. First, firms may choose to produce overseas to avoid trade costs. Referred to as the market access (or tariff jumping) motive, this strategy leads firms to expand horizontally duplicating their production process in foreign countries. Markusen and Venables (1998) offer a representative model of "horizontal FDI." Second, when the production process consists of various separable stages that require different factor intensities, firms may pursue a vertical type of FDI locating each stage in a country in which the factor used intensively is abundant. Helpman's (1984) classical model of "vertical FDI" reflects this comparative advantage motive.

Controlling for these motives enables us to separate geographic concentration of multinationals driven by shared, first-nature incentives such as market size, comparative advantage, and trade policies from the effect of second-nature agglomeration forces on multinationals' location decisions. We achieve this goal via a two-step procedure.

First, we estimate a conventional empirical equation following Carr, Markusen and Maskus (2001), Yeaple (2003), and Alfaro and Charlton (2009). Specifically, we consider the following

[9] In addition to benefits, agglomeration can also incur costs (diseconomies) including increasing land price, labor cost, congestion, and other negative externalities (such as pollution).

specification:

$$
\begin{aligned}
y_{c\widetilde{c}k} = {} & \beta_0 + \beta_1 marketsize\_ave_{c\widetilde{c}} + \beta_2 distance_{c\widetilde{c}} + \beta_3 skill\_diff_{c\widetilde{c}} \\
& + \beta_4 skill\_diff_{c\widetilde{c}} \times skillintensity_k + \beta_5 skillintensity_k + \beta_6 tariff_{c\widetilde{c}k} + \varepsilon_{c\widetilde{c}k}
\end{aligned} \tag{5}
$$

where $y_{c\widetilde{c}k}$ denotes either the number or total employment of subsidiaries in country $\widetilde{c}$ and industry k owned by MNCs in country c, $marketsize\_ave_{c\widetilde{c}}$ is average market size proxied by GDP of the home and host countries, $distance_{c\widetilde{c}}$ is the distance, $skill\_diff_{c\widetilde{c}}$ represents the difference in skill endowment, measured by average years of schooling, between the home and host countries (i.e., $skill_{\widetilde{c}} - skill_c$), $skillintensity_k$ is the skilled labor intensity proxied by share of non-production workers for each industry, $tariff_{c\widetilde{c}k}$ is the level of tariff set by the host country $\widetilde{c}$ on the home country c in industry k, and $\varepsilon_{c\widetilde{c}k}$ are the residuals.

If market access is a significant motive in MNCs' investment decisions, we expect $\beta_1 > 0$, $\beta_2 > 0$, and $\beta_6 > 0$. If comparative advantage is a significant motive, we expect $\beta_2 < 0$, $\beta_4 > 0$, and $\beta_6 < 0$. We obtain GDP data from the World Bank's WDI database, education information from Barro and Lee (2000), and tariff data from the TRAINS database, and construct skilled labor intensity from U.S. census data. Our estimates are largely consistent with those of Yeaple (2003) and Alfaro and Charlton (2009), and suggest significant effects of both market access ($\beta_1 > 0$) and comparative advantage ($\beta_2 < 0$, $\beta_3 < 0$, $\beta_4 > 0$, $\beta_5 < 0$, and $\beta_6 < 0$) motives.

Based on the estimates of equation (5), we obtain and sum, for each host country $\widetilde{c}$ and industry k, fitted values of $y_{c\widetilde{c}k}$. This gives $\widehat{y}_{\widetilde{c}k}$, the level of MNC activities in each host country and industry predicted exclusively by first-nature location fundamentals. Although ideally we would like to obtain predicted values at more disaggregated geographic levels such as cities and provinces, the explanatory variables in equation (5) are mostly available only at the country level. To construct predicted FDI activities at a more disaggregated location level, we use the actual share of multinationals in each city to capture cross-city variations in attractiveness (e.g., port access and favorable industrial policies). Multiplying the actual share by $\widehat{y}_{\widetilde{c}k}$ gives $\widehat{y}_{sk}$ for each city s and industry k.

In the second stage, we repeat step 1 of DO's procedure to obtain a geographic distribution function for each pair of industries k and $\widetilde{k}$. We use the predicted levels of MNC activity (either predicted number or total employment of MNCs) in each city and industry (i.e., $\widehat{y}_{sk}$ and $\widehat{y}_{s\widetilde{k}}$) as the weight when estimating the kernel function. This generates, for each pair of industries, an expected geographic agglomeration index based exclusively on the estimated effects of location characteristics including market size, comparative advantage, and trade costs. We compare in Section 6 the role of these characteristics relative to that of agglomeration forces in determining the global agglomeration patterns of multinational firms.

5.2 Second nature: agglomeration forces

In addition to first-nature variables, multinationals' location choice can also be affected by agglomeration forces. Since Marshall (1890), economists have long recognized the importance of agglomeration benefits, arguing that the domestic industrial clusters that emerged in many countries (e.g., Silicon Valley) can be explained by the cost and productivity advantages enjoyed by firms that locate near one another. These advantages include (i) proximity to suppliers and customers, (ii) external scale economy in factor markets, and (iii) knowledge spillovers.

The advantage of proximity can differ dramatically between multinational corporations and domestic firms. Multinationals often incur substantial trade costs in sourcing intermediate inputs and reaching downstream buyers. They also face significant market entry costs when relocating to a foreign country because of, for example, limited supply of capital goods. Further, given their technology intensity, technology spillovers from closely linked industries can be particularly attractive to MNCs. We review below the role of each Marshallian force in multinational firms' location choices.

Proximity to customers and suppliers Marshall (1890) argued that transportation costs induce plants to locate close to inputs and customers and determine the optimal trading distance between suppliers and buyers.[10] This can be especially true for MNCs given their large volumes of sales and intermediate inputs. Compared to domestic firms, multinationals are often the leading corporations in each industry. Because they tend to be the largest customers of upstream industries as well as the largest suppliers of downstream industries, the input-output relationship between MNCs (e.g., Dell and Intel, Ford and Delphi) can be far stronger than that between average firms.[11]

To determine the importance of customer and supplier relationships in multinationals' agglomeration decisions, we construct a variable, $linkage_{k\widetilde{k}}$, to measure the extent of the input-output relationship between each pair of industries. We use the 2002 Benchmark Input-Output Accounts published by the Bureau of Economic Analysis, and define $linkage_{k\widetilde{k}}$ as the share of industry k's inputs that come from industry $\widetilde{k}$, and vice versa.[12] These shares are calculated relative to all input-output flows including those to non-manufacturing industries and final consumers. As supplier flows are not symmetrical, we take either the maximum or mean of the input and output relationships for each pair of industries.

Note that the assumption that the U.S. IO structure (and similarly the structure of factor and technology demand we discuss next) carries over to other countries can potentially bias

[10] For FDI theoretical literature in this area, see, for example, Krugman (1991), Venables (1996), Krugman and Venables (1996), Puga and Venables (1997), Markusen and Venables (2000), and Baldwin and Ottaviano (2001).

[11] Head, Ries and Swenson (1995) note, for example, that the dependence of Japanese manufacturers on the "just-in-time" inventory system exerts a particularly strong incentive for vertically linked Japanese firms to agglomerate abroad.

[12] The D&B data use 1987 SIC; the 2002 Benchmark IO Accounts NAICS. We use the concordance from the U.S. Census Bureau taken from http://www.census.gov/epcd/naics02/S87TON02.HTM.

our empirical analysis against finding a significant relationship. But on the other hand, it also mitigates the possibility that our control variables are endogenous to the agglomeration of multinationals.

External scale economy in labor markets Agglomeration can also yield benefits through external scale economies in labor markets. Firms' proximity to one another shields workers from the vicissitudes of firm-specific shocks; as a result, workers in locations in which other firms stand ready to hire them are often willing to accept lower wages.[13] Externalities can also occur as workers move from one job to another. This is especially true between MNCs because of their similar skill requirements and large expenditure on worker training. MNCs can have a particularly strong incentive to lure workers from one another because the workers tend to receive certain types of training that are well suited for working in most multinational firms (business practices, business culture, etc.).[14]

To test labor market pooling forces, we follow Ellison, Glaeser and Kerr (2009) in measuring each industry pair's similarity in occupational labor requirements. We use the Bureau of Labor Statistics' 2006 National Industry-Occupation Employment Matrix (NIOEM) which reports industry-level employment across detailed occupations (e.g., Assemblers and Fabricators, Metal Workers and Plastic Workers, Textile, Apparel, and Furnishings Workers, Business Operations Specialists, Financial Specialists, Computer Support Specialists, and Electrical and Electronics Engineers). We convert occupational employment counts into occupational percentages for each industry and map the BLS industries to the SIC3 framework. We measure each industry pair's labor similarity, $labor_{k\widetilde{k}}$, using the correlation in occupational percentages.

External scale economy in capital-good markets External scale economies can similarly arise in the capital-good markets. This is a force that has not been emphasized in the literature, but has particular relevance given multinational firms' large involvement in capital-intensive activities. Geographically concentrated industries offer better support to providers of capital goods (such as producers of specialized components and providers of machinery maintenance), and reduce the risk of investment (due, for example, to the existence of resale markets). As a result, local expansion of capital intensive activities can lead to expansion in the supply of capital goods, thereby exerting a downward pressure on cost.

To evaluate the role of capital-market externalities, we construct a measure of industries' similarity in capital-good demand using capital flow data from the Bureau of Economic Analysis (BEA). The capital flow table (CFT), a supplement to the 1997 benchmark input-output (I-

[13] This argument has been formally considered in Marshall (1890), Krugman (1991), and Helsley and Strange (1990) and tested in Diamond and Simon (1990). Rotemberg and Saloner (2000), in a related motivation, argue that workers can also gain because multiple firms protect workers against ex post appropriation of investments in human capital.

[14] The flow of workers can also lead to knowledge spillover, another Marshallian force discussed further below.

O) accounts, shows detailed purchases of capital goods (e.g., motors and generators, textile machinery, mining machinery and equipment, wood containers and pallets, computer storage devices, wireless communications equipment) by using industry. As for the labor market variable, we measure each industry pair's similarity in capital structure, denoted by $capital_{k\widetilde{k}}$, using the correlation of investment flow vectors.[15]

Knowledge spillovers A third motive relates to the flow of ideas that facilitates innovation and the development of new technologies. Knowledge can diffuse from one firm to another through movement of workers between companies, interaction between people who perform similar jobs, or direct interaction between firms such as technology sourcing. This has been noted by Navaretti and Venables (2006), who predict that MNCs may benefit from setting up affiliates in proximity to other MNCs with advanced technology (i.e., "the so-called centers of excellence"). The affiliates can benefit from knowledge spillovers, which can then be transferred to other parts of the company.

To capture this agglomeration force, we construct a measure of knowledge spillovers between industries using patent citation flow data taken from the NBER Patent Database. The data, compiled by Hall et al. (2001), includes detailed records for all patents granted by the United States Patent and Trademark Office (USPTO) from January 1975 to December 1999. Each patent record provides information about the invention (e.g., technology classification, citations of prior art) and inventors submitting the application (e.g., name and city). We construct the knowledge spillovers variable, i.e., $knowledge_{k\widetilde{k}}$, by measuring the extent to which technologies in industry k cite technologies in industry $\widetilde{k}$, and vice versa.[16] In practice, there is little directional difference in $knowledge_{k\widetilde{k}}$ due to the extensive number of citations within a single technology field. We obtain both max and mean for each set of pairwise industries.

6 Main Empirical Evidence

In this section, we first discuss the pairwise-industry agglomeration patterns of multinational firms including, respectively, MNC subsidiaries, subsidiary employment, and headquarters. We then evaluate the role of first-nature fundamentals and second-nature factors in explaining each of these patterns, and contrast the clustering of MNC subsidiaries with that of domestic plants.

[15] Agglomeration can also induce costs by, for example, increasing labor and land prices. Like benefits, these costs can be potentially greater for industries with similar labor and capital demand, in which case the estimated parameters of the variables would represent the net effect of similar factor demand structures on agglomeration decisions.

[16] We use the concordance adopted in Ellison, Glaeser and Kerr (2009) (we thank William Kerr for providing the data). Concordances are developed between the USPTO classification scheme and SIC3 industries based on probabilistic mapping. We use patents filed by all nationalities.

6.1 Patterns of multinational-firm agglomeration

Table 1, which presents descriptive statistics of the agglomeration indices for MNC subsidiaries, subsidiary employment, and headquarters, reveals significant variation in the level of agglomeration across industry pairs.[17] In about 30 percent of industry pairs, multinational subsidiaries exhibit statistically significant evidence of agglomeration at 200 km relative to the manufacturing sector as a whole. In nearly a third of industry pairs, MNC subsidiaries show evidence of clustering at 400 km. The degree of pairwise-industry agglomeration is noticeably higher among MNC headquarters than across MNC subsidiaries, suggesting greater dispersion among the latter.

Table 2 reports the correlation of the indices. As shown, each index is highly correlated across the different distance thresholds. Between MNC subsidiaries and headquarters, the correlation is around 0.41 at 200 km and rises with the distance thresholds, reaching 0.59 at 1600 km. This suggests that while for some industry pairs the clusters of MNC subsidiaries are similar to those of headquarters, for other industries the two types of establishments exhibit distinctively different agglomeration patterns.

[Tables 1 and 2 about here]

Figure 1 plots a network view of agglomerating industry pairs. In this figure, each node represents an individual 3-digit SIC industry and each link indicates the existence of a positive agglomeration value at 200 km level (i.e., statistically significant agglomeration of the two industries at 200 km) with the weight of each link increasing with the value of the agglomeration index. The size of each node represents the number of industries that agglomerate with a given industry. Industries represented by the larger nodes are hence more central than industries represented by smaller nodes. It is clear that not all industries are equal. Some, such as Paperboard Mills (263), Newspaper Publishing, Publishing and Printing (271), Miscellaneous Publishing (274), Leather Products Luggage (316), Miscellaneous Primary Metal Products (339), Miscellaneous Transportation Equipment (379), and Watches, Clocks, Clockwork Operated Devices and Parts (387), agglomerate with a particularly large number of industries. The counterpart figure for MNC headquarters is shown in Figure 2. The degree of industry agglomeration is substantially higher for headquarters, but like the subsidiaries, some industries exhibit particularly strong propensities to agglomerate with other industries.

[Figures 1-2 about here]

[17] The scale of the agglomeration index is driven by the scope of the dataset and empirical methodology. That we take into account the distance of all pairs of establishments around the world (the maximum distance being around 20,000 km) determines that the kernel estimates at each distance level will be low. Adoption of the Monte Carlo approach also means that the indices are constructed based on differences from the 95% global confidence bands. A positive value represents statistically significant evidence of agglomeration.

Industry pairs that exhibit some of the highest agglomeration index values are reported in Table A.2. They include, for example, Footwear except Rubber (314) and Boot and Shoe Cut Stock and Findings (313), Knitting Mills (225) and Footwear except Rubber (314), Dolls, Toys, Games (394) and Sporting and Athletic and Footwear except Rubber (314), Miscellaneous Publishing (274) and Paperboard Mills (263), and Miscellaneous Publishing (274) and Miscellaneous Transportation Equipment (379).

6.2 Evaluating the role of first and second natures

We now examine the role of first-nature fundamentals and second-nature factors in explaining the pairwise-industry patterns of MNC agglomeration. Formally, we estimate the following empirical specification:

$$\begin{aligned} agglomeration_{k\widetilde{k}}(T) = \alpha_K + \beta_1 firstnature_{k\widetilde{k}} \\ + \beta_2 IOlinkage_{k\widetilde{k}} + \beta_3 labor_{k\widetilde{k}} + \beta_4 capital_{k\widetilde{k}} + \beta_5 knowledge_{k\widetilde{k}} + \varepsilon_{ij}, \end{aligned} \quad (6)$$

where $agglomeration_{k\widetilde{k}}(T)$ is the agglomeration index of industry pairs (relative to the counterfactuals) and the right-hand side includes (i) the first-nature predicted agglomeration patterns ($firstnature_{k\widetilde{k}}$) constructed in Section 5.1, and (ii) proxies for agglomeration forces described in Section 5.2 consisting of input-output linkages ($IOlinkage_{k\widetilde{k}}$), labor- and capital-market externalities ($labor_{k\widetilde{k}}$ and $capital_{k\widetilde{k}}$), and knowledge spillovers ($knowledge_{k\widetilde{k}}$). We also use an industry fixed effect by including α_K, a vector of industry dummies that takes the value of 1 if either industry k or $\widetilde{k}$ corresponds to a given industry and zero otherwise. These industry dummies control for industry-specific factors such as natural advantage and market structure which may affect the location patterns of each industry.

The lower panel of Table 1 reports summary statistics for the industry-level control variables. Table A.3 presents the correlation matrix. For example, the correlation between industry-pair input-output linkage and similarity in capital structure is about 0.19 and the correlation between input-output linkage and knowledge spillovers is 0.29.[18]

6.2.1 Multinational subsidiaries

We begin with the agglomeration of MNC subsidiaries. Table 3 reports the multivariate regression results. Agglomeration forces including IO linkages, capital-market correlation, and knowledge spillovers all play a significant role and display the expected signs.[19] For example,

[18] The table also shows the mean and maximum measures of IO linkages and knowledge spillovers to be highly correlated. We use average values in our analysis in Section 6, but obtain similar results when we use the maximum measure (available upon request).

[19] In univariate regression results for each of our main variables, all the agglomeration variables were found to be highly significant across the different distance threshold levels. The estimated effects also exhibited expected signs. Across agglomeration forces, capital-market correlation had the greatest impact across all distance thresholds,

at 400 km, a 10-percentage-point increase in the level of knowledge spillovers, that is, the percentage of patent citations between two industries, leads to a 0.117-percentage-point increase in the level of agglomeration between industries. This is equivalent to an 60-percent improvement over the average (0.2). The first-nature variable is significant at 1600 km explaining agglomeration of MNCs at a relatively large geographic scale.

[Table 3 about here]

The lower panel of Table 3 reports the normalized beta coefficients.[20] Comparing the standardized coefficients of agglomeration forces, we find the effects of knowledge spillovers and capital-market correlation to outweigh that of vertical production linkages. The parameter of labor-market correlation is insignificant in the multivariate regressions.[21] One possible explanation for this result is multinationals' motive to search for the cheapest production labor market, placing less emphasis on external scale economies in labor markets. These results differ sharply from existing findings on domestic industrial localization in the U.S. Ellison, Glaeser and Kerr (2009), for example, find input-output relationships to have the greatest effect of all the Marshallian factors considered in their study, followed by labor market pooling. Intellectual spillover, in contrast, plays a weaker role. Our findings suggest that given the knowledge and capital intensive characteristics of multinational firms, it is important to take into account not only vertical production linkages, but also knowledge spillovers and capital-market externalities, in explaining the agglomeration of multinational firms.

Comparing the estimates across distance thresholds, we find the impact of knowledge spillovers diminishes at more aggregate geographic levels while the effect of capital-market externalities rises. The role of vertical production linkages, on the other hand, remains mostly constant across distance thresholds. The stronger effect of knowledge spillovers at shorter distance levels suggests that compared to the other agglomeration benefits, benefits from knowledge spillovers tend to be localized geographically.

Comparing the relative importances of first and second natures, we find that at 1600 km where the effect of first-nature variables is significant, the relative importance of first nature dominates the cumulative importance of agglomeration forces. A one-standard-deviation increase in first-nature fundamentals leads to a 0.33 standard-deviation increase in the level of agglomeration while a one-standard-deviation increase in proxies of agglomeration forces are associated with

followed by labor-structure correlation, knowledge spillovers, and input-output linkages. The table (and similarly all other tables showing univariate results) is suppressed from the paper due to space consideration but available upon request.

[20] Standardized coefficients enable us to compare the changes in the outcome associated with the metric-free changes in each covariate.

[21] We also considered excluding the capital-market correlation variable. We found the knowledge spillover and IO linkage variables to remain positive and significant and the labor correlation coefficient to remain insignificant. This result suggests that the capital variable is indeed capturing agglomeration incentives not represented by the other variables.

about a 0.08 standard-deviation increase in agglomeration intensity. The estimated magnitude of the impact of first nature is in alignment with existing studies of domestic industrial concentration. For example, Ellison, Glaeser and Kerr (2009) show that a one-standard-deviation increase in natural advantage leads to 0.25 standard-deviation increase in the agglomeration of U.S. industries. Our proxy of first nature takes into account a large variety of economic fundamentals stressed in the existing empirical FDI literature, including market size, comparative advantage, and trade costs. It is hence not surprising that our first nature proxy explains a larger fraction of MNC agglomeration than the natural advantage proxy in Ellison, Glaeser and Kerr (2009) for domestic firm agglomeration, and more than agglomeration proxies. At the more disaggregated geographic levels, however, first-nature considerations do not appear to have a significant effect; agglomeration forces, in turn, become the driving forces. Table 4 performs similar analysis excluding the first nature variable. The coefficients and the statistical significance of the agglomeration forces remain broadly unchanged.

[Table 4 about here]

6.2.2 Multinational subsidiary employment

So far we have examined MNC agglomeration using establishment as the unit of observation. We now take into account the different employment sizes of multinational subsidiaries. This essentially treats the worker as the unit of observation, and measures the level of agglomeration among workers. This exercise, by differentiating the agglomeration incentives between individual establishments and workers, has implications for policy making targeted at influencing the geographic distribution of workers.

Table 5 reports the estimates. We notice that in contrast to Table 3, in which labor-market correlation does not exert a significant effect, multinational subsidiaries in industries with greater potential labor-market externalities are found to have a significantly higher level of employment agglomeration. Knowledge spillover, another force of agglomeration that involves close labor interaction and mobility, also plays a significant role in explaining the agglomeration of MNC subsidiary employment between industries. In fact, knowledge spillover appears to be the strongest agglomeration factor at most distance thresholds (with statistical significance at 1 percent). Whereas the effects of labor-market externalities and knowledge spillovers diminish at more aggregate geographic levels, capital-market correlation exerts a significant and positive effect at larger distance thresholds. Unlike agglomeration of subsidiaries, the first-nature variable plays a significant role at all distance thresholds continuing to exert a stronger impact than agglomeration forces. A one-standard-deviation increase in the first-nature variable leads to a 0.31 standard-deviation increase in the agglomeration of MNC subsidiary employment at 200 km, whereas the cumulative effect of agglomeration forces is about 0.17. Also noteworthy is that the impact of first nature falls, and the importance of agglomeration forces rises slightly, at more disaggregated geographic levels.

[Table 5 about here]

6.2.3 Multinational headquarters

Now we consider and examine whether MNC headquarters clusters are motivated by the same factors that motivate MNC subsidiary clusters. To control for the role of first-nature characteristics in explaining the agglomeration of MNC headquarters, we follow the procedure described in Section 5.1. Specifically, we obtain predicted levels of MNC activities by FDI home countries and then construct the expected distribution and agglomeration of MNC headquarters following the rest of the procedure.

Table 6 reports the estimation results. All variables except input-output linkages exert a significant effect. A one-standard-deviation increase in the first nature variable is associated with a 0.21 standard-deviation increase in MNC headquarters agglomeration. At 200 km, both knowledge spillovers and labor structure correlation play a positive and significant role, with a cumulative effect around 0.06. Beyond 200 km, the effect of labor structure becomes insignificant while the importance of capital structure correlation increases. Again, this result is consistent with the localized feature of labor markets and lower mobility of labor in comparison to capital goods.

[Table 6 about here]

The finding that input-output relationships affect MNC subsidiaries but not headquarters suggests that with the geographic separation of headquarters services and production activities, the determinants of MNC subsidiary location are at variance with those of headquarters. As the headquarters become increasingly specialized in management, research, marketing, and the provision of other services, the importance of vertical production linkages diminishes.

6.3 Multinational v.s. non-multinational plants

After establishing the agglomeration patterns of MNCs, we now compare how the effects of first-nature fundamentals and second nature agglomeration forces vary between multinational and non-multinational plants.

We proceed by constructing a similar index of agglomeration for domestic plants worldwide. As for the multinational establishments, we obtain geocode information for each domestic plant and compute the distance between each pair of plants. We then follow the procedure in Section 4.2 and estimate an index of agglomeration for each pair of domestic industries, that is, $agglomeration^{d}_{k\widetilde{k}}(T)$ where the superscript d denotes domestic.

Comparing the index of MNC agglomeration with that of domestic plants, we first find that at 200 km the index is higher for multinationals in 51 percent of industry pairs. At 400 km, multinationals exhibit stronger agglomeration intensities in 40 percent of the industry pairs. The percentage rises to 46 percent at the more aggregate geographic level 1600 km. We further find that the correlation of the MNC and domestic plant agglomeration indices is around 0.2 at 200 km and rises to 0.32 at 1600 km. These observations suggest that multinational and non-multinational plants follow distinctively different agglomeration patterns.[22]

We next formulate a counterpart equation of equation (6) for domestic plants and take the difference of the two equations. This gives us:

$$
\begin{aligned}
&agglomeration^{m}_{k\widetilde{k}}(T) \square agglomeration^{d}_{k\widetilde{k}}(T) \\
&= (\beta^{m}_{1} \square \beta^{d}_{1}) firstnature_{k\widetilde{k}} + (\beta^{m}_{2} \square \beta^{d}_{2}) IOlinkage_{k\widetilde{k}} + (\beta^{m}_{3} \square \beta^{d}_{3}) labor_{k\widetilde{k}} \\
&+ (\beta^{m}_{4} \square \beta^{d}_{4}) capital_{k\widetilde{k}} + (\beta^{m}_{5} \square \beta^{d}_{5}) knowledge_{k\widetilde{k}} + \varepsilon_{ij}, \qquad (7)
\end{aligned}
$$

where $agglomeration^{m}_{k\widetilde{k}}(T) \square agglomeration^{d}_{k\widetilde{k}}(T)$ represents the difference between the MNC and domestic pairwise-industry indices, and the coefficient vector, $\beta^{m} \square \beta^{d}$, represents the different effects of the first-nature and second-nature variables on multinational and domestic plants.

[Table 7 about here]

The results are reported in Table 7. We find that proxies for capital-market externalities and knowledge spillovers exert a stronger effect on multinationals than on domestic plants in the same industry pairs. The role of the input-output relationship is not significantly different between the two at disaggregated geographic levels, but is significantly stronger for multinationals at more aggregate geographic levels such as 800 km and 1600 km. First-nature variables including market size, trade costs and comparative advantage have a stronger impact on domestic plants, suggesting that domestic industrial clusters place greater emphasis on location fundamentals. These findings are consistent with the distinct characteristics of multinational firms: relative to their domestic counterparts, multinationals exhibit greater participation in knowledge and physical capital intensive activities. These results highlight the importance of distinguishing agglomeration incentives between MNCs and domestic plants.

7 Exploring Entry Patterns of Multinational Firms

In examining the role of location fundamentals and agglomeration forces in the agglomeration patterns of MNCs, two potential econometric concerns can arise. First, we have not taken into account the different timing of establishments. Our estimates thus far reflect how economic factors explain the geographic pattern of plants at a given time, taking into account not only

[22] This finding also mitigates a potential concern that can arise in the empirical analysis, namely, the possibility that the index of MNC agglomeration captures the agglomeration between MNCs and domestic plants. This potential concern is further addressed in Section 7.

new plants' entry decisions but also incumbents' decisions to continue in their current locations. But the mix of old and new plants gives rise to the potential for reverse causality between MNC location patterns and economic fundamentals.[23] Second, there is the possibility that our index of MNC agglomeration captures not only the agglomeration between MNCs, but also clustering between MNC and domestic plants.[24] Although the low correlation between the indices of MNC agglomeration and domestic plant agglomeration reported in Section 6.3 suggests that this is not likely to be a significant issue, we take a further measure to address the concern.

Consequently, we explore in this section the dynamics of location patterns and spatial interdependence between entrants and incumbents. Specifically, we distinguish new plants from incumbents in our data and assess new MNC plants' propensity to agglomerate with incumbents. This enables us to identify the roles of the first- and second-nature factors in MNCs' entry decisions. Repeating the procedure described in Section 4.2, we construct an index of agglomeration between MNC entrants in 2004-2005 and MNC incumbents established before 2004. For each industry pair k and $\widetilde{k}$, the index measures the propensity of new MNC subsidiaries in industry k to cluster with incumbent MNCs in industry $\widetilde{k}$, and vice versa.

[Table 8 about here]

We compare the agglomeration index for MNC entrants against two benchmarks. First, as in Section 6.3, we adopt domestic plants as the benchmark and compare how MNCs agglomerate towards incumbent MNCs relative to the clustering of domestic plants. Table 8 reports the estimates. The role of second-nature agglomeration forces remains robust in explaining the entry patterns of MNCs. Relative to domestic plants, multinational entrants display a stronger propensity to cluster with incumbent multinationals when there are relatively stronger knowledge spillover benefits, capital-market externalities, and vertical production linkages. Labor market pooling and first-nature variables, again, have a greater impact on the agglomeration of domestic plants.

To address the possibility that the index of MNC agglomeration reflects clustering with domestic plants, we construct an alternative benchmark, an agglomeration index to measure the propensity of new MNC subsidiaries to cluster with domestic plants. We find that for each industry pair, MNCs exhibit a stronger tendency to agglomerate with incumbent MNC

[23] Reverse causality between proxies for agglomeration forces and MNC agglomeration patterns is less likely given that the proxies used in this paper are constructed based on industry-level production technology characteristics, which are less likely to change significantly over time.

[24] A related potential concern here is that when multinational establishments came into existence as a result of cross-border acquisitions, their agglomeration patterns can simply reflect the agglomeration patterns of domestic establishments. We argue that because MNCs' acquisition choices are also dependent on location fundamentals and agglomeration economies, acquisitions patterns constitute part of MNC agglomeration patterns. The fact that we observe a low correlation between the agglomeration indices of MNCs and domestic plants suggests that MNCs do not simply follow the agglomeration patterns of domestic plants. But to provide further assurance that our analysis captures the agglomeration incentives of multinationals, we focus in this section on the entry patterns of new Greenfield FDI.

plants than with incumbent domestic plants. Moreover, the estimated effects of the first-nature fundamentals and second-nature agglomeration variables remain largely similar.

8 Conclusion

We examine in this paper the relative importance of first-nature location fundamentals and second-nature agglomeration economies in the global agglomeration patterns of multinational firms. Our analysis indicates that while first-nature economic characteristics play a significant role in explaining the agglomeration of multinational firms, they are not the only driving force. In addition to shared, first-nature motives of market access and comparative advantage, multinationals' location choices are significantly affected by a variety of agglomeration economies at the world level. Second-nature forces including not only vertical production linkages but also knowledge spillovers and capital-market externalities, two traditionally under-emphasized forces, all exert a significant effect on the clustering of MNCs. Labor market pooling plays a significant role in the clustering of MNC subsidiary employment.

Comparing their relative importances, we find the effect of first-nature factors to outweigh the cumulative impact of agglomeration economies. Across agglomeration economies, we find knowledge spillovers and capital-market externalities to have a particularly important effect. This finding is similarly true when we compare the clustering of multinational firms with that of domestic counterparts in the same industries. In comparison to domestic firms, multinationals are significantly more influenced by knowledge- and physical-capital considerations and less by labor market pooling.

The importance of agglomeration forces also varies between multinational headquarters and subsidiaries. Our results suggest that headquarters and subsidiaries follow distinct agglomeration patterns that reflect their increasing specialization in, respectively, headquarters services and production activities. Whereas agglomeration forces except vertical production linkages (namely, knowledge spillover and capital- and labor-market externalities) exert a significant effect on the clustering of MNC headquarters, all factors including vertical production linkages play a role in the clustering of MNC subsidiaries.

The role of second-nature forces remains robust when we explore the dynamics of multinational firm agglomeration. Examining the process of agglomeration enables us to address potential econometric concerns such as reverse causality and identify the causal effects of location fundamentals and agglomeration economies. Our results suggest that all second-nature forces except labor market pooling, namely, knowledge spillovers, capital-market externalities, and input-output production linkages, exert a significant effect on the propensity of new MNCs to cluster with MNC incumbents.

Two potential extensions are worthy of particular attention. First, the patterns of MNC agglomeration can vary across regions. For example, labor-market externalities can offer a stronger incentive for agglomeration in countries with more rigid and less mobile labor markets. Similarly, the varying quality of infrastructure across regions can affect the value of proximity for vertically linked industries. Firms are likely to have a stronger motive to cluster with suppliers and customers in countries with poorer infrastructure. Further analysis of the role of regional characteristics in determining the clustering of MNCs could provide additional policy insights. A second direction for future research involves micro patterns of agglomeration. Our analysis, like most of existing research on agglomeration, has explored neither potential heterogeneity within each industrial clusters nor how the role of firm heterogeneity might shape the formation of industrial clusters. Given the heterogeneous characteristics, such as size and foreign ownership, of firms, the level of agglomeration centering each firm can be different. Some may attract more agglomeration than others, leading to a hub-and-spoke location pattern. We address this question in a companion project which record significant heterogeneity within industrial clusters.

References

[1] Acemoglu, Daron, Johnson, Simon and Mitton, Todd. (2009) Determinants of Vertical Integration: Contracts, Finance, and Regulation. *Journal of Finance*, forthcoming.

[2] Aitken, Brian, Harrison, Ann and Lipsey, Robert (1996) Wages and foreign ownership A comparative study of Mexico, Venezuela, and the United States. *Journal of International Economics* 40(3-4): 345-371.

[3] Alfaro, Laura and Charlton, Andrew (2009) Intra-Industry Foreign Direct Investment. *American Economic Review* 99(5): 2096-2119.

[4] Alfaro, Laura, Chanda, Areendam, Kalemli-Ozcan, Sebnem and Sayek, Selin (2004) FDI and Economic Growth: the Role of Local Financial Markets. *Journal of International Economics* 64(1): 89-112.

[5] Amiti, Mary and Javorcik, Beata Smarzynska (2008) Trade Costs and Location of Foreign Firms in China. *Journal of Development Economics* 85(1-2): 129-149.

[6] Anderson, James E. and van Wincoop, Eric (2004) Trade Costs. *Journal of Economic Literature* 42(3): 691-751.

[7] Baldwin, Richard E. and Ottaviano, Gianmarco (2001) Multiproduct Multinationals and Reciprocal FDI Dumping. *Journal of International Economics* 54(2): 429-448.

[8] Barro, Robert J. and Lee, Jong-Wha (2000) International Data on Educational Attainment: Updates and Implications. CID Working Paper No. 42.

[9] Blonigen, Bruce A., Ellis, Christopher J. and Fausten, Dietrich (2005) Industrial Groupings and Foreign Direct Investment. *Journal of International Economics* 65(1): 75-91.

[10] Bobonis, Gustavo J. and Shatz, Howard J. (2007) Agglomeration, Adjustment, and State Policies in the Location of Foreign Direct Investment in the United States. *The Review of Economics and Statistics* 89(1): 30-43.

[11] Brainard, S. L. (1997) An empirical assessment of proximity-concentration trade-off between multinational sales and trade. *American Economic Review* 87 (4): 520-544.

[12] Carr L. David, James R. Markusen and Keith E. Maskus (2001) Estimating the Knowledge-Capital Model of the Multinational Enterprise. *American Economic Review* 91(3): 693-708.

[13] Caves, Richard E. (1975) Diversification, Foreign Investment and Scale in North American Manufacturing Industries. Ottawa: Economic Council of Canada.

[14] Chen, Maggie (2009a) Regional Economic Integration and Geographic Concentration of Multinational Firms. *European Economic Review* 53(3): 355-375.

[15] Chen, Maggie (2009b) Third-Country Effects in Multinational Production Networks. George Washington University, mimeo.

[16] Crozet, Matthieu, Mayer, Thierry and Mucchielli, Jean-Louis (2004) How Do Firms Agglomerate? A Study of FDI in France. *Regional Science and Urban Economics* 34: 27-54.

[17] Davis, Donald R. and Weinstein, David E. (2002) Bones, Bombs, and Break Points: The Geography of Economic Activity. *American Economic Review* 92(5): 1269-1289.

[18] Debaere, Peter, Joon Lee and Myungho Park (2010) Agglomeration, Backward and Forward Linkages: Evidence from South Korean Investment in China. *Canadian Journal of Economics*, forthcoming.

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

[20] Duranton, Gilles and Overman, Henry (2005) Testing for Localization Using MicroGeographic Data. *Review of Economic Studies* 72(4): 1077-1106.

[21] Duranton, Gilles and Overman, Henry (2008) Exploring the Detailed Location Patterns Of U.K. Manufacturing Industries Using Microgeographic Data. *Journal of Regional Science* 48(1): 213-243.

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

[23] Ellison, Glenn and Glaeser, Edward (1999) The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration? *American Economic Review* 89(2): 311-316.

[24] Ellison, Glenn, Glaeser, Edward, and Kerr, William (2009) What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review* (forthcoming).

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

[26] Harrison, Ann E. and Rodríguez-Clare, Andrés (2009) Trade, Foreign Investment, and Industrial Policy. *Handbook of Development Economics*, forthcoming.

[27] Harrison, Ann E., Love, Inessa and McMillan, Margaret S. (2004) Global Capital Flows and Financing Constraints. *Journal of Development Economics* 75(1): 269-301

[28] Head, Keith, Ries, John and Swenson, Deborah (1995) Agglomeration Benefits and Location Choice: Evidence from Japanese Manufacturing Investments in the United States. *Journal of International Economics* 38(3-4): 223-247.

[29] Head, Keith and Mayer, Thierry (2004a) Market Potential and the Location of Japanese Investment in the European Union. *Review of Economics and Statistics* 86(4): 959-972.

[30] Head, Keith and Mayer, Thierry (2004b) The Empirics of Agglomeration and Trade. In *Handbook of Regional and Urban Economics*, Volume 4. Edited by J. V. Henderson and J. E. Thisse. Elsevier: 2609-2669.

[31] Helpman, Elhanan (1984) A Simple Theory of Trade with Multinational Corporations. *Journal of Political Economy* 92(3): 451-471.

[32] Helpman, Elhanan, Melitz, Marc and Yeaple, Stephen (2004) Export versus FDI with Heterogeneous Firms. *American Economic Review* 94(1): 300-316.

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

[34] Javorcik, Beata S. (2004) Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In search of spillovers through backward linkages. *American Economic Review* 94 (3): 605-627.

[35] Krugman, Paul (1991) Increasing Returns and Economic Geography. *Journal of Political Economy* 99(3): 483-499.

[36] Krugman, Paul (1993) On the Number and Location of Cities. *European Economic Review* 37(2-3): 293-298.

[37] Krugman, Paul and Venables, Anthony J. (1996) Integration, Specialization, and Adjustment. *European Economic Review* 40(3-5): 959-967.

[38] Levinsohn, J. and Petrin, A. (2003) Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies* 70(2): 317-342.

[39] Markusen, James (1984) Multinationals, Multi-Plant Economies, and the Gains from Trade. *Journal of International Economics* 16(3-4): 205-226.

[40] Markusen, James and Venables, Anthony J. (1998) Multinational Firms and the New Trade Theory. *Journal of International Economics* 46(2): 183-203.

[41] Markusen, James and Venables, Anthony J. (2000) The Theory of Endowment, Intra-industry and Multi-national trade. *Journal of International Economics* 52(2): 209-234.

[42] Marshall, Alfred (1890) *Principles of Economics*. London, UK: MacMillan and Co.

[43] Navaretti, Giorgio Barba and Venables, Anthony J. (2006) *Multinational Firms in the World Economy*. Princeton University Press.

[44] Puga, Diego and Venables, Anthony J. (1997) Preferential Trading Arrangements and Industrial Location. *Journal of International Economics* 43(3-4): 347-368.

[45] Rosenthal, Stuart and Strange, William (2001) The Determinants of Agglomeration. *Journal of Urban Economics* 50(2): 191-229.

[46] Rosenthal, Stuart and Strange, William (2003) Geography, Industrial Organization, and Agglomeration. *Review of Economics and Statistics* 85(2): 377-393.

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

[48] Venables, Anthony J. (1996) Equilibrium Locations of Vertically Linked Industries. *International Economic Review* 37(2): 341-359.

[49] Yeaple, Stephen (2003) The Complex Integration Strategies of Multinational Firms and Cross-Country Dependencies in the Structure of Foreign Direct Investment. *Journal of International Economics* 60(2): 293-314.

[50] Yeaple, Stephen (2009) Firm Heterogeneity and the Structure of U.S. Multinational Activity: An Empirical Analysis. *Journal of International Economics* 78(2): 206-215.

Appendix: Extending the Spatial Agglomeration Index

In this appendix, we extend the spatial agglomeration index constructed in Section 4 to a measure of global agglomeration that accounts for various forms of trade costs including border, language, and tariffs. The extended index captures the agglomeration of MNCs in a generalized metric of trade costs. The role of agglomeration economies in explaining this index can be potentially different because, for example, intermediate inputs and final goods can be more tradeable than knowledge and physical capital.

Following Head and Mayer (2004a) and Chen (2009a), we employ a two-step procedure to estimate a comprehensive measure of trade costs for each pair of MNC subsidiaries. We first estimate a standard trade gravity equation given by

$$q_{ijt} = EX_{it} + IM_{jt} + \lambda Z_{ijt} + \varepsilon_{ijt}, \tag{8}$$

where the dependent variable is the natural log of imports of country j from country i denoted as q_{ijt}, EX_{it} denotes an exporter-year fixed effect, IM_{jt} represents an importer-year fixed effect, and $\lambda Z_{ijt} \equiv \lambda_1 \ln d_{ij} + \lambda_2 B_{ij} + \lambda_3 B_{ij} \times L_{ij} + \lambda_4 PTA_{ijt}$ with Z_{ijt} representing a vector of bilateral market access variables. In particular, Z_{ijt} includes $\ln d_{ij}$, the natural log of distance between the capital cities of the importer and exporter countries, B_{ij}, a dummy variable that equals 1 if the trading countries share a border and 0 otherwise, and L_{ij}, a dummy variable that equals 1 when the two countries share a common language. Following Head and Mayer (2004a) and Chen (2009a), the equation allows the border effect to differ across importing countries depending on whether they speak the same language as the exporting country. The expectations are $\lambda_1 < 0$, $\lambda_2 > 0$, $\lambda_3 > 0$, and $\lambda_4 > 0$.

We estimate the gravity equation using a dataset that covers trade flows between 80 countries. We obtain the trade data from the COMTRADE database, and geographic information, including distance, border, and language, from the CEPII distance dataset. The PTA information is from the Tuck Trade Agreements Database and WTO Regional Trade Agreements Dataset. Our estimates of the gravity equation are broadly consistent with the existing literature. All the bilateral market access variables exert an expected effect on trade volume.[25]

In the second stage, we use the estimated parameters of bilateral access variables, that is, λ_1-λ_4, to construct the generalized measure of trade cost. Specifically, we consider

$$\tau_{ij} = \square \widehat{\lambda}_1 \ln d_{ij} \square B_{ij}(\widehat{\lambda}_2 + \widehat{\lambda}_3 L_{ij}) \square \widehat{\lambda}_4 PTA_{ijt} \tag{9}$$

and substitute the distance, contiguity, language, and PTA information for each pair of subsidiaries into the equation to compute the fitted trade cost τ_{ij}.

Repeating this methodology described in Section 4, we construct a agglomeration index

[25] For a comprehensive review, see Anderson and van Wincoop (2004).

based on the generalized measure of trade costs (instead of distance). Table A.4 reports the multivariate regression results. Knowledge spillovers (0.108) and capital market externalities (0.028) have a positive and significant effect, while the effects of the labor and linkages variables are insignificant. These results suggest that IO linkages do not play a significant role in explaining the agglomeration of MNC subsidiaries when the ease of trading intermediate inputs and final goods due to low tariffs, country contiguity, and low language barriers are taken into account. For agglomeration forces to be meaningful, goods and factors must have little tradeability (e.g., knowledge and physical capital) or, more generally, face high trade and movement barriers.

[Table A.4 about here]

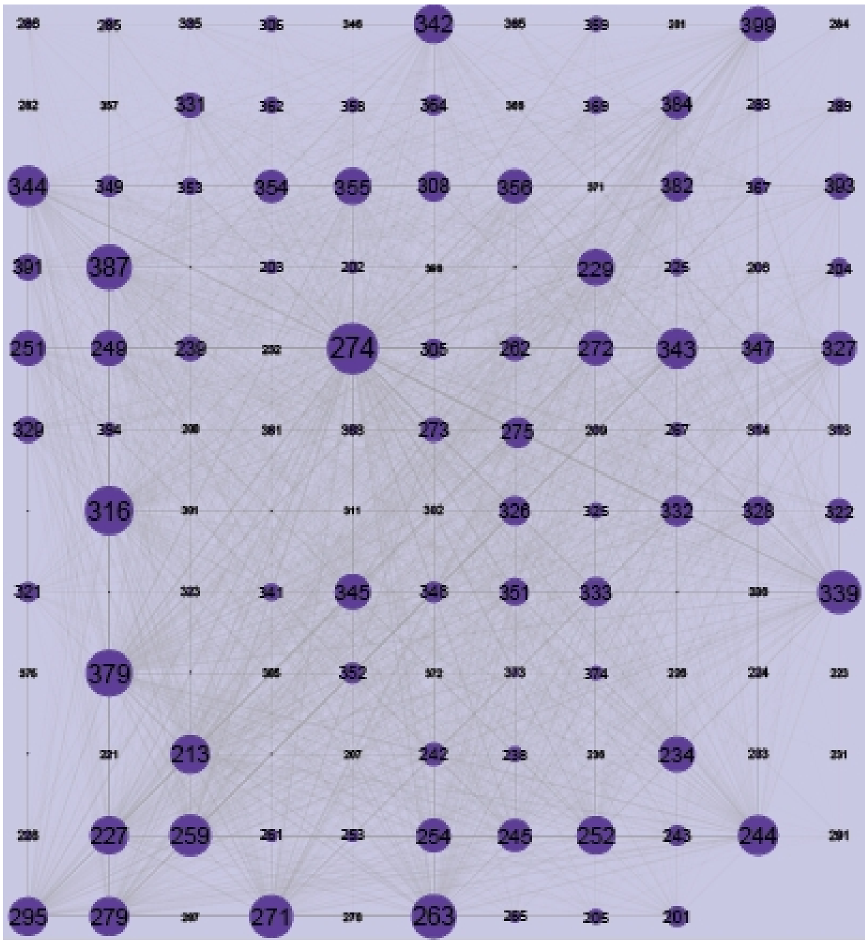

Figure 1: The agglomeration pattern of MNC subsidiaries

(Notes: Each node represents an SIC 3-digit manufacturing industry. Industries in which there is significant agglomeration at 200 km are linked. The size of each node is proportional to the number of agglomerating industries.)

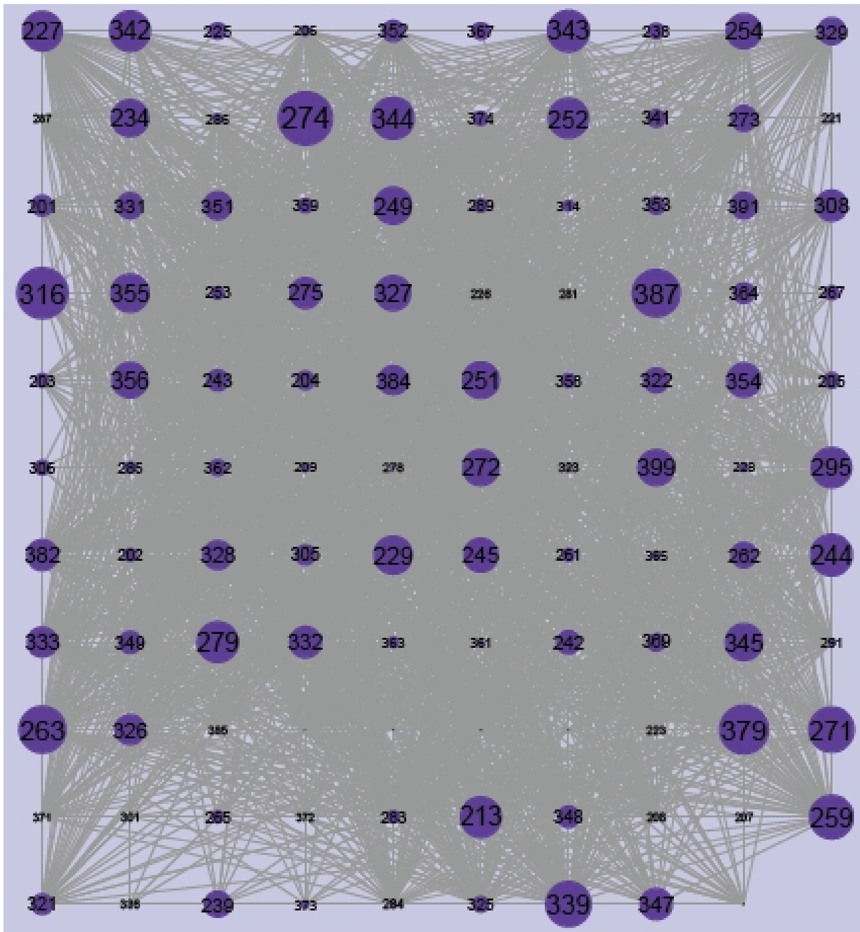

Figure 2: The agglomeration pattern of MNC headquarters

(Notes: Each node represents an SIC 3-digit manufacturing industry. Industries in which there is significant agglomeration at 200 km are linked. The size of each node is proportional to the number of agglomerating industries.)

Table 1: Descriptive Statistics for Multinational Agglomeration Indices and Agglomeration Economies

	# Obs.	Mean	Std. Dev.	Min.	Max.
Agglomeration Indices–Pairwise-Industry Level					
Subsidiaries (Percentage Points)					
Threshold (T) = 200 km	7875	0.095	0.230	0.000	2.538
T = 400 km	7875	0.213	0.505	0.000	5.453
T= 800 km	7875	0.506	1.174	0.000	11.856
T= 1600 km	7875	1.006	2.308	0.000	21.126
Subsidiary Employment (Percentage Points)					
Threshold (T) = 200 km	7875	0.090	0.262	0.000	2.997
T = 400 km	7875	0.186	0.505	0.000	5.523
T= 800 km	7875	0.402	0.997	0.000	10.140
T= 1600 km	7875	0.717	1.794	0.000	16.539
Headquarters (Percentage Points)					
Threshold (T) = 200 km	7875	0.135	0.327	0.000	3.249
T = 400 km	7875	0.315	0.735	0.000	6.889
T= 800 km	7875	0.761	1.681	0.000	14.806
T= 1600 km	7875	1.373	2.895	0.000	24.280
Agglomeration Economies–Pairwise-Industry Level					
Input Output (IO) Linkages	7875	0.003	0.012	0.000	0.193
Capital	7875	0.476	0.209	-0.004	1.000
Labor	7875	0.333	0.227	0.014	1.000
Knowledge	7875	0.007	0.012	0.000	0.179

Notes: The agglomeration indices are constructed by comparing the estimated distance kernel function of each industry pair with the 95 percent global confidence band of counterfactual kernel estimators at 200 km, 400 km, 800km, and 1600 km. Input Output (IO) Linkages, Capital, Labor, and Knowledge correspond to the industry-level variables employed to proxy for the various agglomeration economies: proximity to input suppliers or industrial customers, external scale economies in factor markets including labor and capital, and knowledge spillovers. Same industry pairs (SIC3) are excluded. See text for detailed descriptions of the variables.

Table 2: Correlation of MNC Agglomeration Indices

MNC Subsidiaries and Subsidiary Employment

	200 km (Subs.)	400 km (Subs.)	800 km (Subs.)	1600 km (Subs.)	200 km (Empl.)	400 km (Empl.)	800 km (Empl.)	1600 km (Empl.)
T = 200 km (Subs.)	1.000							
T = 400 km (Subs.)	0.993	1.000						
T = 800 km (Subs.)	0.962	0.986	1.000					
T = 1600 km (Subs.)	0.882	0.919	0.965	1.000				
T = 200 km (Empl.)	0.420	0.374	0.327	0.295	1.000			
T = 400 km (Empl.)	0.498	0.463	0.427	0.398	0.985	1.000		
T = 800 km (Empl.)	0.603	0.591	0.581	0.570	0.888	0.952	1.000	
T = 1600 km (Empl.)	0.616	0.619	0.633	0.662	0.769	0.852	0.955	1.000

MNC Subsidiaries and Headquarters

	200 km (Subs.)	400 km (Subs.)	800 km (Subs.)	1600 km (Subs.)	200 km (HQ)	400 km (HQ)	800 km (HQ)	1600 km (HQ)
T = 200 km (Subs.)	1.000							
T = 400 km (Subs.)	0.993	1.000						
T = 800 km (Subs.)	0.962	0.986	1.000					
T = 1600 km (Subs.)	0.882	0.919	0.965	1.000				
T = 200 km (HQ)	0.406	0.419	0.425	0.399	1.000			
T = 400 km (HQ)	0.421	0.438	0.450	0.429	0.993	1.000		
T = 800 km (HQ)	0.453	0.477	0.500	0.493	0.955	0.982	1.000	
T = 1600 km (HQ)	0.497	0.526	0.564	0.590	0.858	0.896	0.955	1.000

Notes: Obs=7875. Same industry pairs (SIC3) are excluded. See text for detailed descriptions of the variables.

Table 3: Agglomeration Economies and MNC Subsidiary Agglomeration

	T= 200 km	T= 400 km	T= 800 km	T= 1600 km
IO Linkages	0.265*	0.573*	1.331**	2.596**
	(0.147)	(0.306)	(0.656)	(1.296)
Capital	0.038***	0.093***	0.241***	0.506***
	(0.014)	(0.032)	(0.066)	(0.139)
Labor	-0.002	-0.015	-0.079	-0.231
	(0.016)	(0.035)	(0.068)	(0.160)
Knowledge	0.609**	1.178**	2.521**	4.395**
	(0.293)	(0.546)	(1.117)	(2.371)
First Nature	0.018	0.019	0.020	0.021*
	(0.025)	(0.019)	(0.022)	(0.012)
# Obs.	7875	7875	7875	7875
R^2	0.571	0.600	0.627	0.631
	Beta Coefficients			
IO Linkages	0.014	0.014	0.014	0.013
Capital	0.035	0.039	0.043	0.046
Labor	-0.002	-0.007	-0.015	-0.023
Knowledge	0.031	0.027	0.025	0.022
First Nature	0.266	0.264	0.279	0.333

Notes: Bootstrapped standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed effect. Normalized beta coefficients in lower panel. Same industry pairs (SIC3) are excluded. See text for detailed descriptions of the variables.

Table 4: Agglomeration Economies and MNC Subsidiary Agglomeration Index (Agglomeration Economies Only)

	T= 200 km	T= 400 km	T= 800 km	T= 1600 km
IO Linkages	0.250*	0.541*	1.252*	2.413*
	(0.140)	(0.309)	(0.664)	(1.351)
Capital	0.037***	0.092***	0.238***	0.499***
	(0.014)	(0.028)	(0.064)	(0.127)
Labor	0.005	-0.002	-0.045	-0.153
	(0.018)	(0.037)	(0.080)	(0.163)
Knowledge	0.574*	1.101*	2.330**	3.943*
	(0.309)	(0.608)	(1.143)	(1.992)
# Obs.	7875	7875	7875	7875
R^2	0.570	0.599	0.626	0.630
	Beta Coefficients			
IO Linkages	0.013	0.013	0.013	0.013
Capital	0.034	0.038	0.042	0.045
Labor	0.005	-0.001	-0.009	-0.015
Knowledge	0.029	0.025	0.023	0.020

Notes: Bootstrapped standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed effect. Normalized beta coefficients in lower panel. Same industry pairs (SIC3) are excluded. See text for detailed descriptions of the variables.

Table 5: Agglomeration Economies and MNC Subsidiary Employment Agglomeration

	T= 200 km	T= 400 km	T= 800 km	T= 1600 km
IO Linkages	-0.145	-0.256	-0.272	-0.750
	(0.209)	(0.403)	(0.683)	(1.160)
Capital	0.041*	0.109**	0.315***	0.557***
	(0.023)	(0.044)	(0.089)	(0.144)
Labor	0.048*	0.088*	0.120	0.128
	(0.026)	(0.048)	(0.104)	(0.162)
Knowledge	2.262***	3.957***	6.243***	9.333***
	(0.516)	(0.867)	(1.613)	(2.356)
First Nature	0.0004***	0.0004***	0.0004***	0.0004**
	(0.0001)	(0.0001)	(0.0001)	(0.0002)
# Obs.	7875	7875	7875	7875
R^2	0.327	0.327	0.363	0.402
	Beta Coefficients			
IO Linkages	-0.007	-0.006	-0.003	-0.005
Capital	0.033	0.045	0.066	0.065
Labor	0.042	0.039	0.027	0.016
Knowledge	0.100	0.091	0.073	0.061
First Nature	0.315	0.349	0.390	0.435

Notes: Bootstrapped standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed effect. Normalized beta coefficients in lower panel. Same industry pairs (SIC3) are excluded. See text for detailed descriptions of the variables.

Table 6: Agglomeration Economies and MNC Headquarters Agglomeration

	T= 200 km	T= 400 km	T= 800 km	T= 1600 km
IO Linkages	0.090	0.156	0.127	0.457
	(0.174)	(0.406)	(0.815)	(1.254)
Capital	0.026	0.084**	0.261***	0.459***
	(0.019)	(0.040)	(0.088)	(0.164)
Labor	0.043**	0.064	0.019	-0.085
	(0.021)	(0.044)	(0.104)	(0.180)
Knowledge	0.793***	1.727***	3.870***	6.935***
	(0.241)	(0.477)	(1.153)	(1.735)
First Nature	0.022**	0.023***	0.024*	0.019
	(0.009)	(0.009)	(0.013)	(0.018)
# Obs.	7875	7875	7875	7875
R^2	0.639	0.65	0.664	0.667
	Beta Coefficients			
IO Linkages	0.003	0.003	0.001	0.002
Capital	0.017	0.024	0.032	0.033
Labor	0.030	0.020	0.003	-0.007
Knowledge	0.028	0.027	0.027	0.028
First Nature	0.212	0.212	0.208	0.213

Notes: Bootstrapped standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed effect. Normalized beta coefficients in lower panel. Same industry pairs (SIC3) are excluded. See text for detailed descriptions of the variables.

Table 7: Comparing MNC Subsidiaries with Domestic Plants

	T= 200 km	T= 400 km	T= 800 km	T= 1600 km
IO Linkages	0.041	1.081	5.447**	10.876**
	(0.599)	(1.306)	(2.760)	(4.437)
Capital	0.162***	0.494***	1.335***	2.383***
	(0.051)	(0.113)	(0.220)	(0.366)
Labor	-0.110**	-0.443***	-1.430***	-2.130***
	(0.049)	(0.112)	(0.231)	(0.410)
Knowledge	-1.214	2.823*	24.272***	62.572***
	(0.839)	(1.706)	(3.409)	(6.220)
First Nature	-0.047***	-0.047***	-0.044***	-0.035***
	(0.003)	(0.002)	(0.002)	(0.002)
# Obs.	7875	7875	7875	7875
R^2	0.049	0.053	0.064	0.073
	Beta Coefficients			
IO Linkages	0.001	0.008	0.020	0.023
Capital	0.047	0.067	0.085	0.086
Labor	-0.034	-0.065	-0.099	-0.084
Knowledge	-0.020	0.021	0.086	0.126
First Nature	-0.213	-0.217	-0.219	-0.228

Notes: Bootstrapped standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Normalized beta coefficients in lower panel. Same industry pairs (SIC3) are excluded. See text for detailed descriptions of the variables.

Table 8: The Entry Decision – MNC Subsidiaries versus Domestic Plants

	T= 200 km	T= 400 km	T= 800 km	T= 1600 km
IO Linkages	0.818	2.424*	8.000***	16.045***
	(0.714)	(1.460)	(2.770)	(4.915)
Capital	0.094*	0.289***	0.789***	1.690***
	(0.056)	(0.096)	(0.228)	(0.397)
Labor	-0.183***	-0.571***	-1.692***	-2.797***
	(0.045)	(0.097)	(0.213)	(0.417)
Knowledge	0.878	6.603***	33.455***	84.362***
	(0.781)	(1.655)	(3.244)	(6.295)
First Nature	-0.040***	-0.038***	-0.033***	-0.027***
	(0.003)	(0.003)	(0.002)	(0.002)
# Obs.	6966	6966	6966	6966
R^2	0.04	0.043	0.054	0.068
		Beta Coefficients		
IO Linkages	0.015	0.021	0.032	0.036
Capital	0.028	0.041	0.053	0.063
Labor	-0.060	-0.088	-0.122	-0.112
Knowledge	0.015	0.055	0.130	0.181
First Nature	-0.186	-0.182	-0.170	-0.177

Notes: Bootstrapped standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Normalized beta coefficients in lower panel. Same industry pairs (SIC3) are excluded. See text for detailed descriptions of the variables.

Table A.1: Distribution of Establishment Pairs by Distance and Different Countries

	All pairs		Pairs located in two different countries		
	Pairs (mil)	Ave. dist (km)	Pairs (mil)	Percentage	Ave. dist (km)
dist $\leq$ 200	28.3	91.6	5.6	0.2	131.4
dist $\leq$ 400	54.8	194.1	24.5	0.4	268.7
dist $\leq$ 800	124.2	423.0	85.6	0.7	510.9
dist $\leq$ 1600	257.1	806.6	198.7	0.8	885.8

Notes: Authors' calculations.

Table A.2: Top Industry Pairs by MNC Subsidiary Agglomeration Index

MNC Subsidiary Agglomeration Index			
T = 200 km			
274	Miscellaneous Publishing	379	Miscellaneous Transportation Equipment
314	Footwear, Except Rubber	313	Boot And Shoe Cut Stock And Findings
225	Knitting Mills	313	Boot And Shoe Cut Stock And Findings
367	Electronic Components And Accessories	225	Knitting Mills
225	Knitting Mills	314	Footwear, Except Rubber
T = 400 km			
274	Miscellaneous Publishing	379	Miscellaneous Transportation Equipment
314	Footwear, Except Rubber	313	Boot And Shoe Cut Stock And Findings
225	Knitting Mills	313	Boot And Shoe Cut Stock And Findings
274	Miscellaneous Publishing	213	Chewing And Smoking Tobacco And Snuff
263	Paperboard Mills	213	Chewing And Smoking Tobacco And Snuff
MNC Subsidiary-Employment Agglomeration Index			
T = 200 km			
394	Dolls, Toys, Games And Sporting	314	Footwear, Except Rubber
394	Dolls, Toys, Games And Sporting	313	Boot And Shoe Cut Stock And Findings
225	Knitting Mills	314	Footwear, Except Rubber
314	Footwear, Except Rubber	313	Boot And Shoe Cut Stock And Findings
225	Knitting Mills	394	Dolls, Toys, Games And Sporting And Athletic
T = 400 km			
394	Dolls, Toys, Games And Sporting	314	Footwear, Except Rubber
394	Dolls, Toys, Games And Sporting	313	Boot And Shoe Cut Stock And Findings
225	Knitting Mills	314	Footwear, Except Rubber
314	Footwear, Except Rubber	313	Boot And Shoe Cut Stock And Findings
225	Knitting Mills	313	Boot And Shoe Cut Stock And Findings

Notes: Same industry pairs (SIC3) are excluded. See text for detailed descriptions of the variables.

Table A.3: Correlation of Agglomeration Economies

	IO Linkages	IO Linkages (max.)	Capital	Labor	Knowledge	Knowledge (max.)
IO Linkages	1.000					
IO Linkages (max.)	0.973	1.000				
Capital	0.191	0.189	1.000			
Labor	0.232	0.225	0.567	1.000		
Knowledge	0.291	0.284	0.230	0.331	1.000	
Knowledge (max.)	0.264	0.257	0.188	0.297	0.976	1.000

Notes: Obs=7875. Both average and maximum measures are obtained for IO linkages and knowledge spillovers. See text for detailed descriptions of the variables.

Table A.4: Multinational Subsidiary Agglomeration Index
with a Generalized Measure of Trade Cost

	T= 200 km	T= 400 km	T= 800 km	T= 1600 km
IO Linkages	-0.387	-0.333	-0.213	-0.142
	(0.431)	(0.444)	(0.753)	(0.657)
Capital	0.101*	0.123*	0.133	0.144*
	(0.060)	(0.069)	(0.083)	(0.085)
Labor	-0.016	-0.016	-0.003	-0.006
	(0.126)	(0.113)	(0.114)	(0.105)
Knowledge	6.932**	6.943**	7.998**	8.145***
	(3.321)	(2.917)	(3.154)	(2.702)
First Nature	-0.004	-0.003	0.003	0.002
	(0.037)	(0.013)	(0.006)	(0.003)
# Obs.	7875	7875	7875	7875
R^2	0.336	0.342	0.418	0.413
	Beta Coefficients			
IO Linkages	-0.006	-0.0051	-0.003	-0.002
Capital	0.028	0.033	0.030	0.031
Labor	-0.005	-0.005	-0.001	-0.001
Knowledge	0.108	0.105	0.099	0.097
First Nature	-0.017	-0.027	0.045	0.081

Notes: Bootstrapped standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed effect. Normalized beta coefficients in lower panel. Same industry pairs (SIC3) are excluded. See text for detailed descriptions of the variables.