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, James Harrigan, Keith Head, Tarun Khanna, Jim Markusen, Keith Maskus, Mike Moore, Henry Overman, John Ries, Roberto Samaneigo, Stephen Yeaple, Tony Yezer, and seminar and conference participants at Harvard Business School, University of Virginia Darden School of Business, George Washington Univeristy, the AEA meeting, the ETSG meeting, 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 computing the agglomeration indices, 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 dataset 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 November 2011
JEL No. D2,F2,R1,R3

ABSTRACT

The explosion of multinational activities in recent decades is rapidly transforming the global landscape of industrial production. But are the emerging clusters of multinational production the rule or the exception? What drives the offshore agglomeration of multinational firms? Using a unique worldwide plant-level dataset that reports detailed location, ownership, and operation information for plants in over 100 countries, we construct a spatially continuous index of agglomeration and investigate the patterns and determinants underlying the global economic geography of multinational firms. Our analysis shows that the emerging o¤shore clusters of multinationals are not a simple reflection of traditional industrial clusters. Location fundamentals including market access and comparative advantage and under-emphasized agglomeration economies including capital-good market externality and technology diffusion play a particularly important role in multinationals' offshore 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

Exponential increase in flows of goods, capital, and ideas is one of the most prominent economic trends in recent decades. A key driver of this phenomenon is the cross-border production, investment and innovation led by Multinational Corporations (MNCs). Multinational affiliate sales as a share of world GDP has more than doubled in the past two decades, increasing from 27 percent in 1990 to 58 percent in 2007. This explosion of MNC activities is rapidly transforming the global landscape of industrial production, precipitating 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). Learning the patterns and the determinants of these emerging MNC clusters becomes central to understanding the world's industrial growth and economic development.

Are the new MNC clusters the rule or the exception? What drives the current, offshore agglomeration of MNCs? Are they a mirror projection of the traditional industrial clusters? In this paper, we address these questions by examining the patterns and the determinants underlying the global agglomeration of multinational production. These questions convey implications central to academic and policy debates on foreign direct investment (FDI). Growing evidence suggests that multinationals play a significant role in the performance of local economies, raising local wages (see, e.g., Aitken, Harrison, and Lipsey, 1996) and generating productivity spillovers (see, e.g., Javorcik, 2004). Recognizing these benefits, many countries, including both FDI source and destination nations, have long offered lucrative incentives to MNCs in the hope of building and sustaining industrial clusters. Understanding the location interdependence of multinational firms and how they agglomerate with one another can help better design these economic policies.

The first goal of this paper is to characterize the patterns of MNC agglomeration, both offshore and at headquarters. While existing studies show multinationals tend to cluster within an individual country, there is relatively little evidence on the global significance and characteristics of MNC agglomeration. Our second goal, after identifying the patterns, is to investigate the determinants of MNC agglomeration. A growing literature led by Helpman, Melitz and Yeaple (2004) shows that multinationals exhibit sharply different attributes compared to average domestic firms. Multinationals are often the most productive, capital intensive, and innovative in an industry and, by nature, the most geographically mobile. These attributes determine that the clustering patterns of MNCs are likely influenced by different forces than those of domestic firms. In this paper, we disentangle the relative importance of two distinct categories of economic factors: *location fundamentals* (also referred to as "first nature") and *agglomeration economies* (also known as "second nature"). In contrast to domestic production which emphasizes primarily domestic geography and natural advantage, economic fundamentals of multinational production stress foreign market access and country comparative advantage. Agglomeration economies, which originates from Marshall (1890), build on location fundamentals but focus instead on the benefits of geographic proximity between firms. These benefits include not only those operating through vertical production linkages and labor market externality, forces most documented, but more importantly to MNCs, gains from capital-good market externality and technology diffusion.

Evaluating the patterns and the determinants of MNC agglomeration poses several key challenges. First, the measurement of agglomeration has been a central challenge in the economic geography literature. Traditional indices that define agglomeration as the size of activities (the number of firms or the size of production) located in the same geographic unit provide computationally simple measures of agglomeration, but omit agglomeration activities separated by administrative and geographic borders by treating space as discrete areas. The indices can also be affected by the extent of industrial concentration and misidentify agglomeration for industries with a small number of establishments. Second, disentangling the effects of location fundamentals and agglomeration economies on the spatial patterns of MNC agglomeration is a difficult task. Both location fundamentals and the various agglomeration economies are difficult to measure quantitatively. Further, they can all lead MNCs to locate next to each other, thereby making it challenging to identify their relative effects. Third, quantifying the global patterns of MNC agglomeration requires cross-country data that document activities of multinationals offshore and at headquarters. It also requires detailed location and operational information at the establishment, instead of firm, level so agglomeration can be measured with most disaggregated information.

In this paper, we measure the global agglomeration of MNCs by constructing a spatially continuous index of agglomeration following a new empirical methodology introduced by Duranton and Overman (2005) (henceforth, DO). This index, computed based on establishment-level data, captures the degree of agglomeration at the industry level. The index treats space as continuous and measures agglomeration using actual proximity. The index also employs a Monte-Carlo approach which compares the actual geographic distributions of plants with the distributions of counterfactuals to deal with the effect of industrial concentration and separate agglomeration from the general geographic concentration of manufacturing industries. We adopt the index in this paper to characterize the global patterns of MNC agglomeration. In doing so, we take advantage of the spatial continuity of the index and capture the increasing agglomeration activities occurring across regional and national borders.

To disentangle the effects of location fundamentals and various agglomeration forces, we proceed in three steps. First, we construct the agglomeration index for pairwise industries (also referred to as coagglomeration). As Ellison, Glaeser, and Kerr (2009) note, compared to firms in the same industries, firms from different industry pairs often exhibit greater variations in their relatedness in production, factor markets, and technology space. These variations lead industry pairs to have different agglomeration incentives. For example, while firms in the automobile industry can agglomerate because of both location fundamentals and any of the agglomeration economies, firms in the automobile and steel industries are likely motivated to agglomerate mainly because of vertical production linkages. Exploring the pairwise industry agglomeration of MNCs thus makes it possible to separate the effects of location fundamentals and the various agglomeration economies.

Second, we capture the effect of FDI location fundamentals by constructing an expected index of MNC agglomeration. This index is estimated using the geographic distributions of MNC plants predicted exclusively by FDI location fundamentals including foreign market size, trade costs, and comparative advantage. Specifically, we invoke a two-step procedure. In the first step, we estimate a conventional gravity-type FDI equation and examine the effects of market access and comparative advantage on multinationals' location decisions. Based on the estimates, we obtain the location patterns of MNC establishments predicted by the location fundamentals and, in the second step, construct an index of agglomeration using the predicted, instead of actual, locations. This index represents the expected degree of pairwise industry agglomeration based on the industry pairs' common market access and comparative advantage motives.

Third, controlling for the agglomeration predicted by location fundamentals, we then examine the degree to which proxies of agglomeration forces, including both proxies previously considered (that is, between-industry input-output linkages, labor demand similarity, and technology diffusion) and a new measure of capital-good market externality, explain the extent of agglomeration among multinational firms. We construct the proxies of agglomeration economies using the U.S. industry account data such as the Benchmark Input-Output Accounts and the Capital Flow Tables. This approach is motivated by several considerations. In particular, it makes sure that our proxies of agglomeration economies reflect industries' intrinsic production technology characteristics that are relatively stable over time, limiting the potential for the measures to endogenously respond to MNC agglomeration. Further, using the U.S. as the reference country and imposing the U.S. production technology on the other countries also mitigates the possibility of endogenous measures.

To quantify the global patterns of MNC agglomeration, we perform the analysis using a worldwide plant-level dataset, WorldBase, that provides detailed location, ownership, and activity information for establishments in over 100 countries. The dataset offers several distinct advantages compared to alternative data sources. Its broad cross-country coverage enables us to go beyond individual countries and depict the global patterns of MNC agglomeration. In addition, the dataset reports detailed location and operation information for over 43 million plants, including multinational and domestic, offshore and headquarters establishments, making it possible to compare the agglomeration of different types of establishments. More importantly, the WorldBase database reports the physical location of each establishment, based on which we are able to construct indices of agglomeration using precise latitude and longitude codes for each plant and the distance and the trade cost between each pair of establishments.

Our analysis presents a rich array of new findings that shed light on the global agglomeration of MNCs. First, we show that multinationals follow distinctively different agglomeration patterns offshore than they did at headquarters. The offshore locations of MNCs are, on average, more dispersedly distributed compared to their domestic headquarters locations, suggesting that the offshore clusters of MNCs are not a simple projection of the traditional industrial clusters.

Second, FDI location fundamentals including market access and comparative advantage, although playing a significant and vital role, are not the only driving forces in the patterns of MNC offshore agglomeration. Agglomeration economies, especially capital-good market externality and technology diffusion, exert a crucial effect on MNCs' location decisions overseas. When comparing the relative importance of location fundamentals and agglomeration economies, we find the effect of location fundamentals to exceed the cumulative impact of agglomeration forces. A one-standard-deviation increase in the former is associated with a 0.31 standard-deviation increase in the extent of MNCs' offshore agglomeration at the 200 km level, whereas the cumulative effect of agglomeration economies is around 0.17.

Third, as suggested in the agglomeration patterns, the relative importance of location fundamentals and agglomeration economies varies significantly between MNCs' offshore and headquarters agglomeration. Location fundamentals and capital-good market externality exert a stronger effect on the offshore agglomeration of MNC subsidiary establishments while technology diffusion and labor market externalities are the leading forces behind the agglomeration of headquarters. Vertical production linkages, in contrast, matter for offshore clustering only. These results are consistent with the market seeking and input sourcing motives of cross-border production and the emphasis of headquarters on knowledge intensive activities such as R&D, management, and services. The under-provision of capital goods in many host countries also raises MNCs' incentives to locate proximate to one another overseas and take advantage of agglomeration economies. When comparing MNC offshore agglomeration with the general agglomeration of domestically owned plants, we find, in accordance with their high capital and innovation intensity, MNC plants are significantly more influenced by capital-good market and technological agglomeration factors than non-MNC plants.

Finally, we examine not just the pattern, but also the process, of agglomeration. While our proxies of agglomeration forces, as described earlier, are less likely to be influenced by the agglomeration of MNCs, exploring the dynamics in MNCs' offshore agglomeration helps mitigate the concern of potential reverse causality between countries' economic fundamentals and MNCs' agglomeration patterns. We find our results to remain qualitatively similar. Further, multinational entrants display stronger propensities to cluster with incumbent multinationals as opposed to local incumbent plants. This, again, is especially true when there are relatively strong capital-good market externalities and technology diffusion benefits.

Our paper is closely related to three separate strands of literature. First, we build on an extensive literature in international trade that examines the decision of MNCs to invest abroad. The literature stresses two main FDI motives, market access, whereby firms choose to produce overseas to avoid trade costs, and comparative advantage, whereby firms choose to locate each stage of production in a country where the factor used intensively in that stage is abundant (Helpman, 1984; Markusen and Venables, 1998, 2000; Markusen, 2002). In this paper, we consider market access and comparative advantage as the location fundamentals of multinational firms. We investigate the extent to which sharing common market access and comparative advantage motives explains the agglomeration of MNC cross-border activities. However, we also go beyond the emphasis on location fundamentals and introduce a separate category of factors, agglomeration economies.

By exploring the role of agglomeration economies, the paper is related to another literature in international trade that emphasizes the advantage of proximity between customers and suppliers. Several studies (see, e.g, Head, Ries, and Swenson, 1995; Blonigen, Ellis, and Fausten, 2005; Bobonis and Shatz, 2007; Amiti and Javorcik, 2008) show that MNCs with vertical production linkages tend to agglomerate regionally within a country. Our analysis extends this strand of literature by investigating the global patterns of MNC offshore and headquarters agglomeration. To establish the patterns, we construct a spatially continuous index of agglomeration that addresses the challenges faced by previous measures. We also examine the relative importance of a variety of agglomeration forces instead of focusing on vertical production linkages. Our results show that forces such as capital-good externality and technology diffusion exert an important effect on the agglomeration of MNCs. Omitting these forces could potentially lead to a biased estimation of the importance of production linkages in the agglomeration decisions of MNCs.

Our study is also related to a broader literature in urban economics which assesses the importance of Marshallian agglomeration forces in domestic economic geography. A seminal study in this literature, Ellison and Glaeser (1997), proposes an index of spatial concentration that takes into account the effect of industrial concentration in each industry, an issue noted to affect the accuracy of previous indices. The recent study by DO of the United Kingdom extends the literature by developing a spatially continuous concentration index that is independent of the level of geographic disaggregation. Ellison, Glaeser, and Kerr (2009) apply DO's index to evaluate the coagglomeration of U.S. industries and find input-output relationships and labor market pooling to play an important role. Our analysis, assessing the patterns and causes of MNCs' global agglomeration, contributes to this literature in several important ways. The study offers a perspective on the structure of industrial agglomeration around the world and investigates how the most mobile and distinctive group of firms—the multinationals—agglomerate domestically and overseas. In contrast to the focus of the urban literature on domestic markets and natural advantage, our paper re-considers definitions of agglomeration determinants in the context of multinational firms. Our results indicate that offshore production leads MNCs to exhibit sharply different agglomeration motives overseas than their domestic counterparts, playing a central role in the evolution of industrial landscape around the world. This result is also true when we explore the entry patterns of MNCs to offer new insights into the process of agglomeration.

The rest of the paper is organized as follows. Section 2 discusses the methodology used in this paper to construct pairwise industry agglomeration indices. Section 3 describes the cross-country establishment data and Section 4 summarizes the agglomeration patterns observed worldwide. Section 5 describes the methodology used to measure location fundamentals and agglomeration economies. Sections 6 and 7 report the econometric analysis. The last section concludes.

2 Quantifying the Global Agglomeration of Multinational Firms

In this section, we describe the empirical methodology used to quantify the global agglomeration patterns of multinational firms. As noted in Head and Mayer (2004), 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. One of the latest progresses in this literature is Duranton and Overman (2005).[1]

When measuring agglomeration, many previous 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. Third, previous indices, by equating agglomeration with activities in the same region, can omit agglomerating activities separated by administrative or geographic borders while overestimating the degree of agglomeration within the same administrative or geographic units. The accuracy of the indices is thus dependent on the scale of geographic units. Ellison and Glaeser (1997) develop an index that solves the first two problems. DO (Duranton and Overman, 2005) address the remaining issue of the dependence of existing measures on the level of geographic disaggregation by developing a "continuous-space concentration index."

DO's index 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. The index has then been applied by Ellison, Glaeser, and Kerr (2009) to investigate the coagglomeration of U.S. industries. We extend this index to a global context and measure the degree of coagglomeration of multinational firms around the world. Because it accounts for the continuity in space, the index offers an ideal measure for cross-country studies. We also expand the original index's focus on distance as the main form of trade cost to a measure that accounts for various forms of trade costs.[2]

There are two requirements for the construction of this index. First, availability of physical

[1] See Head and Mayer (2004), Ottaviano and Thisse (2004), and Rosenthal and Strange (2004) for excellent reviews of this literature.

[2] In the main empirical analysis, we construct measures of agglomeration using distance. In Appendix A, we consider alternative measures using estimated trade costs.

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.

The empirical procedure to construct the index consists of three steps. Given our interest in comparing global location patterns of MNC subsidiaries and headquarters, 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. Note that although the locations of nearly all establishments in our data are known with a high degree of precision, distance (as well as estimated trade cost) is an approximation of the true trade cost between establishments. One source of systematic error, for example, is that journey time for any given distance might differ between low- and high-density areas. Given the potential noise in the measurement of trade costs, we follow DO in adopting kernel smoothing when estimating the distribution function.

Let τ_{ij} denote the distance between establishment i and j. For each industry pair k and $\widetilde{k}$, we obtain a kernel estimator at any point τ (i.e., $K_{k\widetilde{k}}(\tau)$):

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

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 a kernel estimator for each of the $7,875$ $(= 126 \times 125/2)$ manufacturing industry pairs in our data.[3]

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

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

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.

[3] Identical industry pairs are dropped from the analsysis because, as described earlier, we reply on industry-pair variations in their relatedness in production, factor demand, and technology to disentangle the effects of location fundamentals and various agglomeration economies. Identical industry pairs typically exhibit all dimensions of relatedness and lack such variations. The level of industry disaggregation in our analysis is dominated by the availability of control variables, as we explain in Section 5.

Step 2: Counterfactuals and global confidence bands Next we obtain counterfactual estimators. This step obtains the geographic distribution of the manufacturing multinationals as a whole, making it possible to control for factors that affect all manufacturing multinational 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 has the same number of observations as the actual data. We then calculate the bilateral distance between 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.

We compare the actual and counterfactual kernel estimators at various 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). 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 $\overline{f}_{k\widetilde{k}}(\tau)$ to denote the upper global confidence band of industry pair k and $\widetilde{k}$. When $f_{k\widetilde{k}}(\tau) > \overline{f}_{k\widetilde{k}}(\tau)$ for at least one $\tau \in [0, T]$, the industry pair is considered to agglomerate at 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

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

or employment-weighted

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

The index measures the extent to which establishments in industries k and $\widetilde{k}$ agglomerate at 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.

3 Data: The WorldBase Database

Our empirical analysis employs a unique worldwide establishment dataset, WorldBase, that covers over 43 million public and private establishments in more than 100 countries and territories. WorldBase is compiled by Dun & Bradstreet (D&B), a leading source of commercial credit and marketing information since 1845.[4] D&B compiles data from a wide range of sources including

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

public registries, partner firms, telephone directory records, and websites, presently operating in over a dozen countries either directly or through affiliates, agents, and associated business partners. All information collected by D&B is verified centrally via a variety of manual and automated checks.[5]

3.1 Cross-Country 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 several distinct advantages compared to alternative data sources used in previous studies.

First, its broad cross-country coverage enables us to examine agglomeration on a global and continuous scale. Examining the global patterns of agglomeration allows us to offer a systematic perspective that takes into account nations at various stages of development. Viewing agglomeration on a continuous scale is important for accounting for the increasing geographic agglomeration occurring across regional and country borders. Table A.1 shows that over 20 percent of pairs of multinationals located within 200 km 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' growing participation in regional trading blocs and rapid declines in cross-border trade costs.

Second, the database reports detailed information for multinational and non-multinational, offshore and headquarters establishments. This makes it possible to compare agglomeration patterns across different types of establishments and investigate how the economic geography of production evolves with forms of firm organization.

Third, 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 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.

[5] Early uses of D&B data include, for example, Lipsey's (1978) 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 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.

We apply the Haversine formula to the geocode data to compute the great-circle distance between each pair of establishments. We also obtain, in addition to distance, an estimated measure of trade cost between each pair of plants to account for other forms of trade barriers such as border, language and tariffs. The distance and the trade cost information is then used to construct an index of agglomeration following the empirical methodology described in Section 2.

3.2 MNC Establishment Data

Our main empirical analysis is based on MNC manufacturing establishments in 2005.[6] 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, employment, and year started.

An establishment is deemed as an MNC foreign subsidiary 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. The parent is defined as an entity that has legal and financial responsibility for another establishment.[7] We drop establishments with zero or missing employment values and industries with fewer than 10 observations.[8]

There are in total $32,427$ MNC offshore 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). Top host countries include 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

[6] In the last part of Section 6, we expand the analysis to include domestic firms to compare the agglomeration patterns of MNC and non-MNC plants.

[7] 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 (2003b) for theoretical work in this area and Chen (2011) 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).

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

five years that covers virtually the entire population of U.S. MNCs. The comparison shows that the two databases have similar accounts of establishments and activities. 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.

4 Patterns of MNC Offshore and Headquarters Agglomeration

Table 1 presents the descriptive statistics of the agglomeration indices for MNC offshore subsidiaries, offshore subsidiary workers, and domestic headquarters, respectively. As shown in the table, there exist significant variations in the level of agglomeration across both industry pairs and different types of establishments.[9]

MNC offshore agglomeration For MNC foreign subsidiaries, about 30 percent of industry pairs exhibit statistically significant evidence of agglomeration at 200 km (that is, with a positive 200-km agglomeration index) and nearly a third of industry pairs show evidence of clustering at 400 km. For the rest of the industry pairs, MNC foreign subsidiaries do not display systematic patterns of agglomeration.

The upper panel of Figure 1 plots a network view of agglomerating industry pairs in MNCs' offshore activities. In this figure, each node represents an individual 3-digit SIC industry and each link indicates the existence of a positive agglomeration index value at 200 km level (i.e., statistically significant agglomeration at 200 km) with the weight of each link increasing with the value of the agglomeration index. The size of each node is proportional to the number of industries that agglomerate with a given industry. Industries represented by the larger nodes are hence more centrally located than industries represented by smaller nodes. It is clear that industries are far from equal in the extent of agglomeration. 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.

[Figure 1 about here]

[9] The scale of the agglomeration index is driven by the scope of the dataset and the 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 offshore 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).

MNC offshore worker agglomeration The average degree of agglomeration is slightly lower, as shown in Table 1, for the agglomeration of MNC subsidiary workers. About 24 percent of the industry pairs exhibit evidence of clustering at 200 km and 28 percent at 400 km.

Industry pairs that exhibit some of the highest offshore worker agglomeration are reported in Table A.2. The industries, all highly labor intensive, include Dolls, Toys, Games and Sporting (394) and Footwear, Except Rubber (314), Dolls, Toys, Games and Sporting (394) and Boot and Shoe Cut Stock and Findings (313), and Knitting Mills (225) and Footwear, Except Rubber (314).

The correlations of the subsidiary and the subsidiary worker agglomeration indices are reported in Table 2. The correlation of the two types of agglomeration is around 0.42 at 200 km and rises with distance thresholds.

[Table 2 about here]

MNC headquarters agglomeration The degree of pairwise-industry agglomeration is, on average, higher among MNC headquarters than across MNC subsidiaries. As shown in Table 1, the average value of the agglomeration index is 50 percent higher for MNC headquarters at 200 km than for foreign subsidiaries. This is consistent with the knowledge capital theory of multinational firms (see, Markusen, 2002), which predicts MNC headquarters to be located in skilled-labor abundant countries and subsidiaries to be dispersedly distributed across host regions based on markets and comparative advantages.

The lower panel of Figure 2 depicts a network view of MNC headquarters agglomeration. Like the subsidiaries, industries such as Paperboard Mills (263), Newspaper Publishing, Publishing and Printing (271), Miscellaneous Publishing (274), Miscellaneous Primary Metal Products (339), Miscellaneous Transportation Equipment (379), and Watches, Clocks, Clockwork Operated Devices and Parts (387), exhibit particularly strong propensities to agglomerate with other industries. But compared to MNC subsidiaries, the extent of agglomeration in these industries is significantly greater for MNC headquarters.

The agglomeration patterns of MNC headquarters and foreign subsidiaries are correlated with a coefficient of 0.41 at 200 km, as shown in Table 2. This suggests that while for some industry pairs the clusters of MNC subsidiaries resemble those of headquarters, for other industries the two types of establishments exhibit distinctively different agglomeration patterns. The emerging offshore clusters of MNCs are not merely a projection of the traditional clusters in industrial countries. The driving forces of MNCs' offshore agglomeration are likely to vary from those of headquarters, as we explore in Section 6.

5 Measuring FDI Location Fundamentals and Agglomeration Economies

Now we turn to the economic factors that could account for the observed agglomeration patterns of MNCs and how each of them is measured in the empirical analysis. The location decision of multinational firms can be viewed as a function of two categories of factors. One consists of location fundamentals of FDI that motivate MNCs to invest in a given country, namely, market access and comparative advantage; the other consists of agglomeration forces including (i) vertical production linkages, (ii) externality in labor markets, (iii) externality in capital-good markets, and (iv) technology diffusion.

5.1 FDI Location Fundamentals

To control for FDI motives such as market size, comparative advantage, and trade costs, we adopt a two-step procedure. First, we estimate a conventional empirical equation following Carr, Markusen and Maskus (2001), Yeaple (2003a), and Alfaro and Charlton (2009). Specifically, we consider the following specification:

$$\begin{aligned} y_{c\tilde{c}k} = {}& {}_0 + {}_1 marketsize\_size_{c\tilde{c}} + {}_2 distance_{c\tilde{c}} + {}_3 skill\_diff_{c\tilde{c}} \\ & + {}_4 skill\_diff_{c\tilde{c}} \times skillintensity_k + {}_5 tariff_{c\tilde{c}k} + \mu_{ck} + \mu'_{\tilde{c}k} + \varepsilon_{c\tilde{c}k} \end{aligned} \tag{5}$$

where $y_{c\tilde{c}k}$ denotes either the number or the total employment of subsidiaries in country $\tilde{c}$ and industry k owned by MNCs in country c, $marketsize\_ave_{c\tilde{c}}$ is average market size proxied by the GDP of home and host countries,[10] $distance_{c\tilde{c}}$ is the distance, $skill\_diff_{c\tilde{c}}$ represents the difference in skill endowment, measured by average years of schooling, between the home and the host countries (i.e., $skill_{\tilde{c}} \square skill_c$), $skillintensity_k$ is the skilled labor intensity proxied by share of non-production workers for each industry, $tariff_{c\tilde{c}k}$ is the level of tariff set by the host country $\tilde{c}$ on the home country c in industry k, μ_{ck} and $\mu'_{\tilde{c}k}$ are vectors of country-industry dummies that control for all country-industry specific factors such as size of domestic industries, country institutional characteristics, and economic policies, and $\varepsilon_{c\tilde{c}k}$ are the residuals.[11]

[10] In addition to GDP, we also considered the market potential, i.e., the sum of domestic and distance-weighted export market size, of home and host countries.

[11] Note that the effect of agglomeration forces such as the size of upstream and downstream industries is controlled for in equation (5) by country-industry dummies. Ideally we would like to estimate equation (5) at more disaggregated geographic levels such as cities and provinces, but the explanatory variables in equation (5) are mostly available only at the country level.

We estimate equation (5) using Poisson quasi-MLE (QMLE).[12] If market access is a significant motive in MNCs' investment decisions, we expect ${}_1 > 0$, ${}_2 > 0$, and ${}_5 > 0$. If comparative advantage is a significant motive, we expect ${}_2 < 0$, ${}_4 > 0$, and ${}_5 < 0$. 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. All host-country characteristics are lagged by 5 years to mitigate reverse causality. In Section 7, we further address the issue by exploring the entry of MNC activities. Our estimates are largely consistent with those of Yeaple (2003a) and Alfaro and Charlton (2009), and suggest significant effects of both market access (${}_1 > 0$) and comparative advantage (${}_2 < 0$, ${}_3 < 0$, ${}_4 > 0$, and ${}_5 < 0$) motives.[13]

Based on the estimates of equation (5), we obtain and sum, for each host country $\widetilde{c}$ and industry k, values of $y_{c\widetilde{c}k}$ predicted by market access and comparative advantage factors. 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 $\widehat{y}_{\widetilde{s}\widetilde{k}}$) as the weight when estimating the kernel function. This generates, for each pair of industries, an expected geographic agglomeration function 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 spatial patterns of multinational firms.

5.2 Agglomeration Economies

In addition to the location fundamentals of FDI, agglomeration economies can also affect multinationals' location choices. The advantage of proximity can differ dramatically between multinational corporations and domestic firms and between MNC foreign subsidiaries and domestic headquarters. For instance, 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 diffusion from closely linked industries can

[12] Santos Silva and Tenreyro (2006) point out that Poisson QMLE can be more attractive than least-square estimators when the variance of the error term is a function of the covariates, in which case the conditional expectation of the logged error term in the log-form estimation equation will not be zero. Head and Ries (2008) further show that estimates produced in this method are smaller than the least-square estimates and remarkably robust to the treatment of zeros and missing values. We also considered a two-step Heckman selection procedure following Helpman et al. (2008) in which we estimate, respectively, the decision to trade and the volume of trade and found the results similar.

[13] Because of space consideration, the results are suppressed from the paper and available from request.

be particularly attractive to MNCs. We review below the role of each agglomeration economy in multinational firms' location choices.

Vertical production linkages 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. This can be especially true for MNCs given their large volumes of sales and intermediate inputs.[14] 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.[15]

To determine the importance of customer and supplier relationships in multinationals' agglomeration decisions, we construct a variable, $IOlinkage_{k\widetilde{k}}$, 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 $IOlinkage_{k\widetilde{k}}$ as the share of industry k's inputs that come from industry $\widetilde{k}$, and vice versa. 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 the mean of the input and output relationships for each pair of industries.

Externality 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.[16] 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.).[17]

To examine 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

[14] For FDI theoretical literature in this area, see, for example, Krugman (1991), Venables (1996), and Markusen and Venables (2000).

[15] 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.

[16] This argument has been formally considered in Marshall (1890), Krugman (1991), and Helsley and Strange (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.

[17] The flow of workers can also lead to technology diffusion, another Marshallian force discussed further below.

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.

Externality in capital-good markets External scale economies can similarly arise in the capital-good markets. This is a force that has been under-stressed in the literature, but has particular relevance to multinational firms given their 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).[18] 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 costs.

To evaluate the role of capital-good market externalities, we construct a new 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-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. We measure each using-industry pair's similarity in capital-good demand structure, denoted by $capitalgood_{k\widetilde{k}}$, using the correlation of investment flow vectors.[19]

Technology diffusion A fourth motive relates to the diffusion of technologies. Technology 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 technology spillovers, which can then be transferred to other parts of the company.

To capture this agglomeration force, we construct a proxy of technology diffusion frequently considered in the knowledge spillover literature (see, e.g., Jaffe et al., 2000; Ellison, Glaeser, and Kerr, 2009), using patent citation flow data taken from the NBER Patent Database. The data,

[18] 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-good demand, in which case the estimated parameters of the variables would represent the net effect of similar factor demand structures on agglomeration decisions.

[19] Note this measure captures a different dimension of industry-pair relatedness than vertical production linkages. Unlike vertical production linkages, industry-pair correlations in capital-good demand reflect industry pairs' similarity in capital-good demand and thus scope for externality in capital-good markets.

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 technology diffusion variable, i.e., $technology_{k\widetilde{k}}$, by measuring the extent to which technologies in industry k cite technologies in industry $\widetilde{k}$, and vice versa.[20] In practice, there is little directional difference in $technology_{k\widetilde{k}}$ due to the extensive number of citations within a single technology field. We obtain both max and mean for each set of pairwise industries.

Constructing the proxies of agglomeration economies using the U.S. industry account data is motivated by three considerations. First, compared to firm-level input-output, factor demand, or technological information which is typically unavailable, industry-level production, factor and technology linkages reflect standardized production technologies and are relatively stable over time, limiting the potential for the measures to endogenously respond to MNC agglomeration. Second, using the U.S. as the reference country further mitigates the possibility of endogenous production linkage measures, even though the assumption that the U.S. production structure carries over to other countries could potentially bias our empirical analysis against finding a significant relationship. Third, the U.S. industry accounts are more disaggregated than most other countries, enabling us to dissect linkages between disaggregated product categories.

6 Evaluating the Role of FDI Location Fundamentals and Agglomeration Economies

We now examine the role of FDI location fundamentals and agglomeration economies in explaining the pairwise-industry agglomeration of MNCs. Formally, we estimate the following empirical specification:

$$\begin{aligned} agglomeration_{k\widetilde{k}}(T) = {} & \alpha_K + \beta_1 fundamentals_{k\widetilde{k}} \\ & + \beta_2 IOlinkage_{k\widetilde{k}} + \beta_3 labor_{k\widetilde{k}} + \beta_4 capitalgood_{k\widetilde{k}} + \beta_5 technology_{k\widetilde{k}} + \varepsilon_{ij}, \end{aligned} \tag{6}$$

where $agglomeration_{k\widetilde{k}}(T)$ is the agglomeration index of industry pairs k and $\widetilde{k}$ at threshold distance T (relative to the counterfactuals) and the right-hand side includes (i) the agglomeration patterns predicted by FDI location fundamentals ($fundamentals_{k\widetilde{k}}$) as 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-good market similarities ($labor_{k\widetilde{k}}$ and $capitalgood_{k\widetilde{k}}$), and technology diffusion ($technology_{k\widetilde{k}}$). We also use an industry fixed effect by including α_K, a

[20] The concordance between the USPTO classification scheme and SIC3 industries is adopted in the construction of the variable.

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 the summary statistics of industry-level control variables. Table A.3 presents the correlation matrix. For example, the correlation between industry-pair production linkage and similarity in capital-good demand is about 0.19 and the correlation between production linkage and technology diffusion is 0.29.[21]

MNC offshore agglomeration Consider first the agglomeration of MNC subsidiaries. Table 3 reports the multivariate regression results. Agglomeration forces including vertical production linkages, capital-good market correlation, and technology diffusion all play a significant role and display the expected signs.[22] For example, at 400 km a 10-percentage-point increase in the level of technology diffusion, that is, the percentage of patent citations between two industries, leads to a 0.117-percentage-point increase in the level of the agglomeration index between industries. This is equivalent to an 60-percent improvement over the average (0.2). The location fundamental variable is significant at 1600 km influencing the spatial patterns of MNCs at a relatively large geographic scale.

[Table 3 about here]

The lower panel of Table 3 reports the normalized beta coefficients.[23] Comparing the standardized coefficients of agglomeration forces, we find the effects of technology diffusion and capital-good market correlation to outweigh that of vertical production linkages, suggesting that given the technology and capital intensive characteristics of multinational firms it is important to take into account not only vertical production linkages, but also technology and capital-good market externalities, in explaining the offshore agglomeration of multinational firms. The parameter of labor-market correlation is insignificant in the multivariate regressions.[24]

Comparing the estimates across distance thresholds, we find the impact of technology diffusion diminishes at more aggregate geographic levels while the effect of capital-good market

[21] The table also shows the mean and the maximum measures of production linkages and technology diffusion to be highly correlated. We used the mean values in our analysis, but obtained similar results when we used the maximum measure.

[22] 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-good market correlation had the greatest impact across all distance thresholds, followed by labor-demand correlation, technology diffusion, and production linkages. The table (and similarly all other tables showing univariate results) is suppressed from the paper due to space consideration but available upon request.

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

[24] We also considered excluding the capital-good market correlation variable. We found the technology diffusion and production linkage variables to remain positive and significant and the labor correlation coefficient to remain insignificant. This result suggests that the capital-good variable is capturing agglomeration incentives not represented by the other variables.

externalities rises. The role of vertical production linkages, on the other hand, remains mostly constant across distance thresholds. The stronger effect of technology diffusion at shorter distance levels suggests that compared to the other agglomeration economies, benefits from technology diffusion tend to be localized geographically.

Comparing the relative importances of location fundamentals and agglomeration economies, we find that at 1600 km where the effect of location fundamentals is significant, its impact dominates the cumulative importance of agglomeration economies. A one-standard-deviation increase in location fundamentals leads to a 0.33 standard-deviation increase in the level of agglomeration while a one-standard-deviation increase in proxies of agglomeration economies are associated with about a 0.08 standard-deviation increase in agglomeration intensity. At the more disaggregated geographic levels, however, location-fundamental considerations do not appear to have a statistically significant effect; agglomeration forces, in turn, become the driving forces. Table 4 performs similar analysis excluding the location fundamental variable. The coefficients and the statistical significance of the agglomeration forces remain broadly unchanged.[25]

[Table 4 about here]

MNC offshore worker agglomeration So far we have examined MNC offshore agglomeration using subsidiary 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. Technology diffusion, another force of agglomeration that involves close labor interaction and mobility, also plays a significant role in explaining the agglomeration of MNC subsidiary workers between industries. In fact, technology spillover appears to be the strongest agglomeration factor at most distance thresholds. Further, while the effects of labor-market externalities and technology spillovers diminish at more aggregate geographic levels, capital-good market correlation exerts a significant and positive effect at larger distance thresholds. Unlike agglomeration of subsidiaries, the location fundamental 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 location fundamental 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 location fundamentals falls, and the importance of agglomeration forces rises slightly, at

[25] Appendix A reports the estimation results for an alternative agglomeration index constructed using estimated trade costs.

more disaggregated geographic levels, suggesting that location fundamentals have more explanatory power in explaining the aggregate, cross-country patterns of multinational activities than the agglomeration of MNCs at the localized level while the contrary is true for agglomeration economies.

[Table 5 about here]

MNC domestic headquarters agglomeration Now we examine the determinants of MNCs' headquarters clusters in comparison with MNC clusters overseas. To control for the role of location fundamentals in explaining the agglomeration of MNC headquarters, we follow the procedure described in Section 5.1, but obtain the level of MNC activities predicted for each MNC home country and construct the expected distribution and agglomeration of MNC headquarters following the rest of the procedure.

Table 6 reports the estimation results. All variables except vertical production linkages exert a significant effect. A one-standard-deviation increase in the location fundamental variable is associated with a 0.21 standard-deviation increase in MNC headquarters agglomeration, suggesting an important role for the characteristics, including market size, skilled labor endowment, and access to host countries, of headquarters countries. At 200 km, both technology diffusion and labor market correlation play a positive and significant role, with a cumulative effect around 0.06. Beyond 200 km, the effect of labor market becomes insignificant while the importance of capital-good market 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]

Comparing Table 6 with Table 3, we find that location fundamentals and capital market externality exert a stronger effect on MNCs' offshore agglomeration than the agglomeration of MNC headquarters and, further, input-output relationships affect MNC subsidiaries but not headquarters. These results suggest that the determinants of MNC subsidiary agglomeration are at variance with those of headquarters. The former is more influenced by market access and comparative advantage motives, capital market externalities, and vertical production linkages, while the headquarters, with specialization in R&D, management, and the provision of other services, place less emphasis on production linkages and more on technology diffusion.

MNC v.s. non-MNC plants After establishing the agglomeration patterns of MNCs, we now compare how the effects of location fundamentals and agglomeration economies might vary between multinational and non-multinational plants. This comparison will help us better understand the distinct economic geography of multinational production and the role of MNCs in shaping the world's industrial landscape.

Conducting our empirical analysis for all domestic manufacturing plants is infeasible given the size of the dataset and the computational intensity of the empirical procedure. To keep the analysis feasible, we adopt a random sampling strategy. For each SIC 3-digit industry with more than 1,000 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.[26]

Comparing the index of MNC agglomeration with that of domestic plants, we 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. We further find that the correlation of the MNC and domestic plant agglomeration indices is around 0.2 at 200 km. The relatively low correlations suggest sharply different spatial patterns for multinational and non-multinational plants.

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

$$
\begin{aligned}
& agglomeration^m_{k\widetilde{k}}(T) - agglomeration^d_{k\widetilde{k}}(T) \\
& = (\beta^m_1 - \beta^d_1) fundamentals_{k\widetilde{k}} + (\beta^m_2 - \beta^d_2) IOlinkage_{k\widetilde{k}} + (\beta^m_3 - \beta^d_3) labor_{k\widetilde{k}} \\
& + (\beta^m_4 - \beta^d_4) capitalgood_{k\widetilde{k}} + (\beta^m_5 - \beta^d_5) technology_{k\widetilde{k}} + \varepsilon_{ij},
\end{aligned} \tag{7}
$$

where $agglomeration^m_{k\widetilde{k}}(T) - agglomeration^d_{k\widetilde{k}}(T)$ represents the difference between the MNC and domestic pairwise-industry agglomeration indices, and the coefficient vector, $\beta^m - \beta^d$, represents the difference of the effects of the covariates on multinational foreign subsidiaries and domestic plants.

[Table 7 about here]

The results are reported in Table 7. We find that proxies for capital-good market externalities and technology diffusion 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. Location fundamental variables including market size and comparative advantage, on the other hand, exert a stronger impact on domestic plants. These findings are consistent with the characteristics of multinational firms: relative to their domestic counterparts, multinationals exhibit greater participation in technology and physical capital intensive activities and thereby stronger agglomeration economies in technology and capital-good markets.

[26] A similar random sampling strategy was used in Ellison, Glaeser, and Kerr (2009).

7 The Process of MNC Agglomeration

Now we turn from the geographic patterns to the process of multinational agglomeration to shed further light on the formation of MNC clusters, in particular, the spatial interdependence between incumbents and entrants. It will also help address two potential econometric concerns in evaluating the determinants of agglomeration. First, the different establishment dates of plants. Our estimates thus far take into account not only 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.[27] 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.[28] Although the low correlation between the indices of MNC agglomeration and domestic plant agglomeration reported in Section 6 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 in location decisions. 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 location fundamentals and agglomeration economies in MNCs' entry decisions. Repeating the procedure described in Section 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, 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 technology

[27]Reverse causality between proxies for agglomeration forces and MNC agglomeration patterns is less likely because, as described in Section 5.2, the proxies used in this paper are constructed based on industry-level production technology characteristics which are less likely to change significantly over time.

[28]A related potential concern here is that when multinational establishments come into existence as a result of cross-border acquisitions, their agglomeration patterns can simply reflect the agglomeration patterns of domestic establishments. We argue that MNCs' acquisition decisions, like the general MNC location choices, are dependent on location fundamentals and agglomeration economies. Moreover, the option to restructure (including to keep or shut down) acquired plants further allows MNCs to optimize their location decisions in response to location factors. The fact that we observe a low correlation between the agglomeration indices of MNCs and domestic plants suggests that MNCs' agglomeration patterns do not simply reflect the agglomeration patterns of domestic plants. But to provide further assurance that our analysis captures the agglomeration incentives of multinationals, we explore in this section the entry patterns of new Greenfield FDI.

spillover benefits, capital-good market externalities, and vertical production linkages. Labor-market and location-fundamental 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 plants than with incumbent domestic plants. Moreover, the estimated effects of the location fundamentals and agglomeration economies remain largely similar.

8 Conclusion

The emergence of new multinational clusters represents one of the most notable phenomena in the process of globalization. We examine in this paper the global patterns and forces of MNC agglomeration both offshore and at headquarters. Using a worldwide plant-level dataset and a spatially continuous index of agglomeration, our analysis presents several new insights into the industrial landscape of multinational production.

First, the offshore clusters of MNCs are not a simple reflection of the traditional industrial clusters. Multinationals follow distinctively different agglomeration patterns overseas than they did at headquarters. The geographic distribution of MNC subsidiaries tends to be relatively more dispersed than that of headquarters. Second, while FDI location fundamentals play a significant role in explaining the agglomeration of multinational firms, they are not the only driving force. In addition to market access and comparative advantage motives, multinationals' location choices are significantly affected by agglomeration economies including not only vertical production linkages but also technology diffusion and capital-market externalities, two traditionally under-emphasized forces. Third, the importance of location fundamentals and agglomeration economies varies significantly between MNCs' offshore and headquarters agglomeration. Location fundamentals and capital-good market externality exert a stronger effect on MNCs' offshore agglomeration and vertical production linkages matter for offshore clustering only. Finally, in the process of agglomeration, multinational entrants display stronger propensities to cluster with incumbent multinationals as opposed to incumbent local plants. The pattern, again, is especially true when there are strong capital-good market externalities and technology diffusion benefits.

Two potential extensions of our analysis 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 not explored potential heterogeneity within each industrial clusters and 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, leading to a hub-and-spoke agglomeration pattern.

References

[1] Acemoglu, Daron, Johnson, Simon and Mitton, Todd. (2009) Determinants of Vertical Integration: Financial Development and Contracting Costs. *Journal of Finance* 63, 1251-1290.

[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] 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.

[5] Anderson, James and Eric van Wincoop (2003) Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review* 93(1): 170-192.

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

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

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

[9] 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.

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

[11] Chen, Maggie (2011) Interdependence in Multinational Production Networks. *Canadian Journal of Economics*, forthcoming.

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

[13] 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.

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

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

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

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

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

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

[20] Head, Keith and Theirry Mayer (2004) The Empirics of Agglomeration and Trade. In: Henderson, J.V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics, Volume 4. Edited by J. V Elsevier, Amsterdam, 2609-2669.

[21] 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.

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

[23] Helpman, Elhanan, Melitz, Marc and Rubinstein, Yona (2008) Estimating Trade Flows: Trading Partners and Trading Volumes. *Quarterly Journal of Economics* 123(2): 441-487.

[24] 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.

[25] Jaffe, Adam, Trajtenberg, Manuel and Fogarty, Michael (2000) Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors. *American Economic Review Paper and Proceedings* 90: 215-218.

[26] 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.

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

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

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

[30] 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.

[31] Markusen, James (2002) Multinational Firms and the Theory of International Trade. Cambridge: MIT Press.

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

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

[34] Ottaviano, Gianmarco and Thisse, Jacques-François (2004). Agglomeration and economic geography. In: Henderson, J.V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics, vol. 4. Elsevier, Amsterdam, pp. 2563-2608.

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

[36] Rosenthal, Stuart and Strange, William (2004) Evidence on the nature and sources of agglomeration economies. In: Henderson, J.V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics, vol. 4. Elsevier, Amsterdam, pp. 2119-2171.

[37] 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.

[38] Santos Silva, J. M. C. and Tenreyro, Silvana (2006) The Log of Gravity. *Review of Economics and Statistics* 88(4): 641-658

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

[40] Yeaple, Stephen (2003a). The Role of Skill Endowments in the Structure of U.S. Outward Foreign Direct Investment. *Review of Economics and Statistics* 85(3): 726-734.

[41] Yeaple, Stephen (2003b) 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.

Appendix A: Accounting for Trade Costs

In this appendix, we extend the agglomeration index constructed in Section 2 to a measure of global agglomeration that accounts for other forms of trade barriers such as border, tariffs and language. The role of location fundamentals and agglomeration economies in explaining this index can be potentially different because, for example, intermediate inputs and final goods can be more tradable than physical and knowledge capital.

We employ a two-step procedure to estimate a comprehensive measure of trade costs for each pair of establishments. 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. As in Chen (2009), 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 $\lambda_4 > 0$. Following Santos Silva and Tenreyro (2006), we estimate the gravity equation using Poisson quasi-MLE (QMLE).

A dataset that covers trade flows between 80 countries is used in the estimation. 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 taken from the Tuck Trade Agreements Database and the 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.[29]

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

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

and substitute the distance, contiguity, language, and PTA information for each pair of establishments into the equation to compute the fitted trade cost $\widehat{\tau}_{ij}$.[30]

[29] The estimation results are available upon request.

[30] To account for home bias in intra-national trade costs, we add a positive constant to $\widehat{\tau}_{ij}$ for establishments located in the same country based on home bias estimates reported in Anderson and van Wincoop (2003). Since estimating home bias for each country in our sample requires intra-national trade flow data for many countries and is beyond the scope of this analysis, we used Anderson and van Wincoop's (2003) U.S. estimates.

Repeating this methodology described in Section 2, we construct a agglomeration index based on the generalized measure of trade costs (instead of distance). Table A.4 reports the multivariate regression results for the agglomeration index constructed based on estimated trade costs (instead of distance). We find that technology diffusion and capital market externalities have a positive and significant effect while the effects of the labor and production linkages variables are insignificant. These results suggest that vertical production 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 already taken into account. For agglomeration forces to be meaningful, goods and factors must have little tradability (e.g., knowledge and physical capital) or, more generally, face high trade and movement barriers.

[Table A.4 about here]

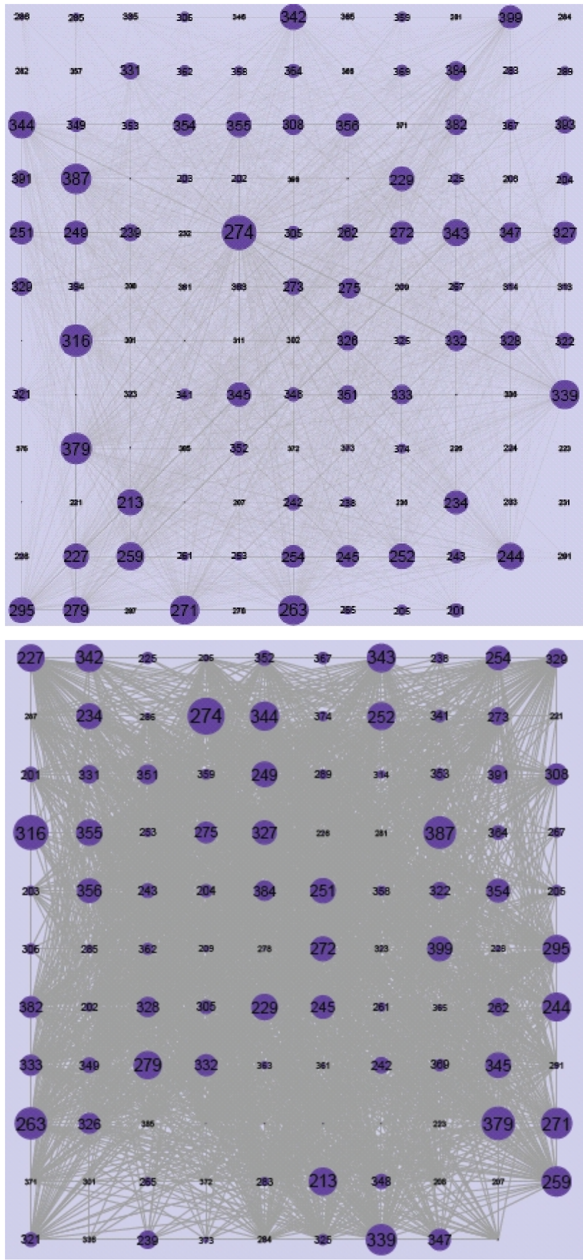

Figure 1: The Agglomeration Patterns of MNC Subsidiaries (top) and Headquarters (bottom) (Notes: Each node represents an SIC 3-digit manufacturing industry. Industry pairs in which there is significant agglomeration at 200 km are linked. The size of each node is proportional to the number of agglomerating 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 Workers (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 Good	7875	0.476	0.209	-0.004	1.000
Labor	7875	0.333	0.227	0.014	1.000
Technology	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, 800 km, and 1600 km. Input-Output (IO) Linkages, Capital Good, Labor, and Technology correspond to the industry-level variables employed to proxy for the various agglomeration economies: vertical production linkages, externalities in factor markets including labor and capital goods, and technology diffusion. 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 Workers

	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. See text for detailed descriptions of the variables.

Table 3: Location Fundamentals, 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 Good	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)
Technology	0.609**	1.178**	2.521**	4.395**
	(0.293)	(0.546)	(1.117)	(2.371)
Location Fundamentals	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 Good	0.035	0.039	0.043	0.046
Labor	-0.002	-0.007	-0.015	-0.023
Technology	0.031	0.027	0.025	0.022
Location Fundamentals	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. See text for detailed descriptions of the variables.

Table 4: Agglomeration Economies and MNC Subsidiary Agglomeration (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 Good	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)
Technology	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 Good	0.034	0.038	0.042	0.045
Labor	0.005	-0.001	-0.009	-0.015
Technology	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. See text for detailed descriptions of the variables.

Table 5: Location Fundamentals, Agglomeration Economies, and MNC Subsidiary Worker 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 Good	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)
Technology	2.262***	3.957***	6.243***	9.333***
	(0.516)	(0.867)	(1.613)	(2.356)
Location Fundamentals	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 Good	0.033	0.045	0.066	0.065
Labor	0.042	0.039	0.027	0.016
Technology	0.100	0.091	0.073	0.061
Location Fundamentals	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. See text for detailed descriptions of the variables.

Table 6: Location Fundamentals, 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 Good	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)
Technology	0.793***	1.727***	3.870***	6.935***
	(0.241)	(0.477)	(1.153)	(1.735)
Location Fundamentals	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 Good	0.017	0.024	0.032	0.033
Labor	0.030	0.020	0.003	-0.007
Technology	0.028	0.027	0.027	0.028
Location Fundamentals	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. 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 Good	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)
Technology	-1.214	2.823*	24.272***	62.572***
	(0.839)	(1.706)	(3.409)	(6.220)
Location Fundamentals	-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 Good	0.047	0.067	0.085	0.086
Labor	-0.034	-0.065	-0.099	-0.084
Technology	-0.020	0.021	0.086	0.126
Location Fundamentals	-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. See text for detailed descriptions of the variables.

Table 8: The Process of Agglomeration – 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 Good	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)
Technology	0.878	6.603***	33.455***	84.362***
	(0.781)	(1.655)	(3.244)	(6.295)
Location Fundamentals	-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 Good	0.028	0.041	0.053	0.063
Labor	-0.060	-0.088	-0.122	-0.112
Technology	0.015	0.055	0.130	0.181
Location Fundamentals	-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. 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 Worker 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 Good	Labor	Technology	Technology (max.)
IO Linkages	1.000					
IO Linkages (max.)	0.973	1.000				
Capital Good	0.191	0.189	1.000			
Labor	0.232	0.225	0.567	1.000		
Technology	0.291	0.284	0.230	0.331	1.000	
Technology (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 technology diffusion. See text for detailed descriptions of the variables.

Table A.4: Multinational Subsidiary Agglomeration Index
with an Estimated 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 Good	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)
Technology	6.932**	6.943**	7.998**	8.145***
	(3.321)	(2.917)	(3.154)	(2.702)
Location Fundamentals	-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 Good	0.028	0.033	0.030	0.031
Labor	-0.005	-0.005	-0.001	-0.001
Technology	0.108	0.105	0.099	0.097
Location Fundamentals	-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. See text for detailed descriptions of the variables.