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 University, the AEA meeting, the ETSG meeting, the EIIT meeting, and the LACEA Trade, Integration and Growth Meeting for helpful feedback. We also thank William Kerr for discussions on the agglomeration indices and kindly providing us the patent concordance data, Bill Simpson for excellent advice on computing the indices, Francisco Pino for assistance with the GIS software, Dun & Bradstreet and Dennis Jacques for help 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 October 2012
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 offshore clusters of multinationals are not a simple reflection of domestic 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

An 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 cross-border production, investment, and innovation led by multinational corporations (MNCs). Multinational affiliate sales as a share of world GDP have more than doubled in the past two decades, increasing from 27 percent in 1990 to 58 percent in 2007.[1] 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).

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 domestic industrial clusters? The agglomeration of economic activity, as recognized in the urban economics literature and, more recently, by trade economists in the New Economic Geography (NEG), is one of the salient features of economic development.[2]

Our goal in this paper is to address the above questions by examining the patterns and determinants underlying the global agglomeration of multinational production. In contrast to domestic production, which emphasizes domestic geography and natural advantage, multinational production stresses foreign market access and comparative advantage. Further, a growing literature led by Helpman, Melitz, and Yeaple (2004) shows multinationals to exhibit different attributes compared to average domestic firms. They are often the most productive, capital intensive, and innovative in an industry and, by nature, the most geographically mobile. Moreover, MNCs tend to incur large trade costs in sourcing intermediate inputs and reaching downstream buyers. They also face large entry costs when locating abroad due to, for example, limited supply of qualified labor and capital goods, both of which raise the value of external scale economy. Known for their technology intensity, MNCs also tend to be the largest sources and recipients of technology spillovers. All these attributes suggest that the clustering patterns of MNCs are likely influenced by forces different from those that affect domestic firms.

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"). Location fundamentals, in the context of MNCs, consist of primarily foreign market size, trade costs, and comparative advantage. Agglomeration economies, the study of which dates from Marshall (1890), build on location fundamentals and stress the benefits of geographic proximity between firms. These include not only the benefits that accrue to vertical production linkages and labor market externality, the forces most studied in the literature, but more important with respect to the decision to engage in MNC activities, the benefits that derive

[1] Source: UNCTAD, World Investment Report (2009).

[2] See Ottaviano and Puga (1998), Head and Mayer (2004), Ottaviano and Thisse (2004), and Rosenthal and Strange (2004) for excellent reviews of these literatures.

from capital-good market externality and technology diffusion. We assess in this paper how the above economic factors affect the agglomeration decisions of MNCs relative to those of domestic plants. While existing studies show multinationals to tend to cluster within an individual country, there is relatively little evidence on the global significance, characteristics, and determinants of MNC agglomeration.

An evaluation of the patterns and determinants of MNC agglomeration faces 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 amount of activity (e.g. the number of firms or production volume) located in a particular geographic unit omit, by treating space as discrete areas, agglomerative activities separated by administrative and geographic borders. The indices can also be affected by the extent of industrial concentration, misidentifying agglomeration for industries with small numbers of establishments. Second, disentangling the effects of location fundamentals and agglomeration economies is complicated by the difficulty of measuring them quantitatively. Further, their common propensity to lead MNCs to locate next to each other makes it challenging to identify their relative effects. Third, to quantify global patterns of MNC agglomeration requires cross-country data documenting multinational activities at the establishment, instead of firm, level so that agglomeration can be measured across countries and with the most disaggregated information.

To overcome the above challenges, our empirical analysis proceeds in three steps. First, we construct a spatially continuous index of agglomeration for pairwise industries (also referred to as coagglomeration). We obtain latitude and longitude codes for each establishment in the data based on plant-level physical location information and compute the distance between each pair of establishments. Following a new empirical methodology introduced by Duranton and Overman (2005) (henceforth, DO) and Ellison, Glaeser, and Kerr (2010) (henceforth, EGK), we then employ a Monte-Carlo approach that compares the actual geographic density of plants in each industry pair with the densities of counterfactuals to separate agglomeration from the general geographic concentration of multinationals and deal with the effect of industrial concentration. Industry pairs that exhibit greater geographic density than the counterfactuals are considered to exhibit significant evidence of agglomeration.

We construct the agglomeration index for each pairwise industries to disentangle the effects of location fundamentals and various agglomeration forces.[3] As noted by EGK, while location fundamentals and all agglomeration economies tend to predict agglomeration among firms in the same industry, their predictions of which industry pairs should agglomerate vary significantly. Compared to firms in the same industries, firms from different industry pairs exhibit greater variation in their relatedness in production, factor markets, and technology space, thereby displaying different agglomeration incentives. For example, firms in the automobile industry may

[3] We use the term "agglomeration" broadly to refer to both within- and between-industry agglomeration (the latter sometimes referred to as "coagglomeration"). The broad usage of the term "agglomeration" is fairly common in the literature.

agglomerate because of both location fundamentals and any of the agglomeration economies whereas firms in the automobile and steel industries are likely motivated to agglomerate mainly because of their 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, after computing the actual agglomeration index of MNCs, we construct an expected index of agglomeration to capture the effect of FDI location fundamentals. This index reflects the geographic distribution 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 industry pairs' common market access and comparative advantage motives.

Third, controlling for the agglomeration predicted by location fundamentals, we 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 spillover) and a new measure of capital-good market externality, explain the variations in the agglomeration index of multinational firms. We construct the proxies of agglomeration forces using lagged, disaggregated U.S. industry account data to mitigate the potential endogeneity concern, as it is not very likely that U.S. industries' production, factor, and technology linkages are a result of MNCs' agglomeration around the world.

We quantify the global agglomeration of MNCs using WorldBase, a worldwide plant-level dataset that provides detailed location, ownership, and activity information for establishments in more than 100 countries. Its broad cross-country coverage enables us to depict worldwide patterns of MNC agglomeration. Moreover, the dataset's detailed location and operation information for over 43 million plants, including multinational and domestic, offshore and headquarters establishments, makes it possible to compare the agglomeration of different types of establishments. More important, the database reports the physical location of each establishment which allows us to construct indices of agglomeration using precise latitude and longitude codes for each plant and the distance and 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 MNCs follow distinctively different agglomeration patterns offshore than their domestic counterparts. The agglomeration indices of MNC foreign subsidiaries, MNC headquarters, and domestic plants exhibit limited correlations. For some industry pairs, the clusters of MNC foreign subsidiaries resemble those of headquarters and domestic plants, while for others, the different types of establishments exhibit sharply different agglomeration patterns.

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, are crucial determinants of MNCs' overseas location decisions. 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 by the agglomeration patterns, the relative importance of location fundamentals and agglomeration economies varies significantly between MNC offshore and domestic plant agglomeration and between MNC offshore and headquarters agglomeration. Comparing MNC offshore agglomeration with the general agglomeration of domestically owned plants, we find MNC plants, reflecting their high capital and innovation intensity, to be significantly more influenced than non-MNC plants by capital-good market and technological agglomeration factors. The under-provision of capital goods in many host countries increases MNCs' incentives to locate proximate to one another overseas and take advantage of agglomeration economies. Moreover, 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.

Fourth, to further alleviate concerns regarding endogenous agglomeration economy measures, we examine regional agglomeration patterns from which the United States is excluded. If U.S. domestic industry-pair relationships could be affected by the agglomeration of MNCs in the U.S., then one would expect that the former would not be affected by the agglomeration of MNCs located in other regions like Europe. We find the results to be qualitatively similar.

Finally, we investigate not just the pattern, but also the process, of agglomeration. Exploring the dynamics in MNCs' offshore agglomeration sheds light on the formation of MNC clusters and mitigates the possibility of reverse causality between our measures of location fundamentals and agglomeration economies and MNCs' agglomeration patterns. We find the results to be robust. Moreover, multinational entrants display stronger propensities to cluster with incumbent multinationals as opposed to local incumbent plants, again, especially when there are relatively strong capital-good market externalities and technology diffusion benefits.

These findings convey implications central to academic and policy debates on FDI. Growing evidence suggests that multinationals play a significant role in the performance of local economies, raising local wages (see, for example, Aitken, Harrison, and Lipsey, 1996) and generating productivity spillovers (see, for example, 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 is critical to designing these economic policies (Harrison and Rodríguez-Clare, 2009).

Our paper is closely related to three separate strands of literature. First, we build on an extensive literature in international trade that examines MNCs' decision to invest abroad (Helpman, 1984; Markusen and Venables, 1998, 2000; Markusen, 2002). In this paper, we investigate the extent to which market access and comparative advantage explain the clustering of MNC cross-border activities. However, we go beyond the emphasis on location fundamentals and introduce a separate category of factors, namely, agglomeration economies.

Exploring the role of agglomeration economies relates the present paper to another literature in international trade that emphasizes the advantage of proximity between customers and suppliers. Several studies (see, for example, Head, Ries, and Swenson, 1995; Bobonis and Shatz, 2007) show that MNCs with vertical production linkages tend to agglomerate regionally within a country. Our analysis extends this literature by investigating the global patterns of MNC offshore and headquarters agglomeration using a spatially continuous index that addresses the challenges faced by previous measures. Our results show that capital-good externality and technology diffusion, factors that have not been emphasized in this literature, exert an important effect on the agglomeration of MNCs. Omitting these agglomeration forces could potentially lead to a biased estimation of the importance of production linkages in MNCs' agglomeration decisions.

Our study is also related to a broader literature in urban economics which assesses the importance of Marshallian agglomeration economies in domestic economic geography. A seminal study in this literature, Ellison and Glaeser (1997), proposes a spatial concentration index that takes into account the effect of industrial concentration in each industry. DO's recent study of the United Kingdom contributes to the literature by developing a novel index that is independent of the level of geographic disaggregation. EGK, adapting DO's index to evaluate the coagglomeration of U.S. industries, find input-output relationships and labor market pooling to play an important role. Our analysis of 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 relative to non-multinationals. In contrast to the urban literature's focus on domestic markets and natural advantage, our paper re-considers definitions of agglomeration determinants in the context of multinational firms. Our results indicate that MNCs' offshore production exhibits sharply different agglomeration motives than their domestic counterparts. 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 describes the cross-country establishment data. Section 3 discusses the methodology used in this paper to construct pairwise industry agglomeration indices and summarizes the agglomeration patterns observed worldwide. Section 4 describes the methodology used to measure location fundamentals and agglomeration economies. Sections 5 and 6 report the econometric evidence on the determinants of MNC agglomeration and the comparison with non-MNC plants, respectively. Sections 7 and 8 present additional analyses that examine, respectively, agglomeration patterns in Europe and the agglomeration process of MNCs to address endogeneity concerns. The last section concludes.

2 Data: The WorldBase Database

Our empirical analysis employs a unique worldwide establishment dataset, WorldBase, that covers more than 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. D&B compiles data from a wide range of sources including 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.[4] All information collected by D&B is verified centrally via a variety of manual and automated checks.[5]

2.1 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 offering several distinct advantages over alternative data sources.

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 in light of the increasing geographic agglomeration occurring across regional and country borders. Table A.1 shows that more than 20 percent of pairs of multinationals located within 200 km are in 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.

[4] For more information, see: http://www.dnb.com/us/about/db_database/dnbinfoquality.html. The dataset employed in this paper was acquired from D&B with disclosure restrictions.

[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 United States., 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.

Third, the WorldBase database reports the physical address and postal code of each plant, whereas 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. We apply the Haversine formula to the geocode data to compute the great circle distance between each pair of establishments. To account for other forms of trade barriers, such as border, language, and tariffs, we further obtain an estimated measure of trade cost between each pair of plants. The distance and the trade cost information is used to construct an index of agglomeration following the empirical methodology described in Section 3.

2.2 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

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

[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).

with fewer than 10 observations.[8]

Our final sample includes 32,427 MNC offshore manufacturing plants. 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 United States, United Kingdom, Canada, France, Poland, Czech Republic, and Mexico.

To examine the coverage of our MNC establishment data, we compared U.S. owned subsidiaries in the WorldBase database with the U.S. Bureau of Economic Analysis' (BEA) Direct Investment Abroad Benchmark Survey, a legally mandated confidential survey conducted every five years that covers virtually the entire population of U.S. MNCs. The comparison revealed similar accounts of establishments and activities between the two databases. We also compared WorldBase with UNCTAD's Multinational Corporation Database. These 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 detailed discussion of the WorldBase data and comparisons with other data sources.

3 Quantifying the Global Agglomeration of Multinational Firms

In this section, we describe the empirical methodology used to quantify the global agglomeration of multinational firms and present the stylized agglomeration patterns. 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).[9]

3.1 Econometric Methodology

Many previous indices have tended to equalize agglomeration with activities located in the same administrative or geographic region (measured by number of firms or volume 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 the geographic concentration of manufacturing industry due to location attractiveness from agglomeration. Third, previous

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

[9] Ottaviano and Puga (1998), Head and Mayer (2004), Ottaviano and Thisse (2004), and Rosenthal and Strange (2004) provide excellent reviews of this literature.

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. These indices' accuracy is thus dependent on the scale of geographic units. Ellison and Glaeser (1997) develop an index that solves the first two problems. DO 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 within 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 adapted by EGK to investigate the coagglomeration of U.S. industries. We extend this index to a global context to measure the degree of coagglomeration of multinational firms worldwide. Because it accounts for continuity in space, the index is well suited 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.[10]

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

The empirical procedure to construct the index involves three steps. To compare 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 density 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 travel 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.

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

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., $K_{k\widetilde{k}}(\tau)$):

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

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 data reflected around zero and the bandwidth set to minimize the mean integrated squared error.[11] This step generates a kernel estimator for each of the $7,875$ ($= 126 \times 125/2$) manufacturing industry pairs in our data.[12]

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, given by

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

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

Step 2: Counterfactuals and global confidence bands To obtain counterfactual estimators, we estimate the geographic distribution of the manufacturing multinationals as a whole in order 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. Given our goal of obtaining, in this step, the overall agglomeration patterns of MNCs, the random samples are drawn from the entire set of MNC establishment locations.[13] 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, unweighted or weighted by employment, for each of the 7,875,000 samples. This gives $1,000$ kernel estimators for each of the $7,875$ industry pairs.

[11] Although we follow DO and EGK in obtaining kernel estimators, a less computationally intensive approach that yields similar properties would be to look at cumulative distances.

[12] Identical industry pairs (126 observations) are dropped from the analysis because, as explained earlier, we rely on industry-pair variations in relatedness in production, factor demand, and technology to disentangle the effects of location fundamentals and various agglomeration economies. Identical industry pairs exhibit all dimensions of relatedness and lack the needed variation. Moreover, as we explain in Section 4, the measures of location fundamentals and agglomeration economies used in this paper, by design, capture only between-industry relationships. The main empirical analysis is performed at the SIC 3-digit level. This level of industry disaggregation is dominated by the availability of control variables, as described in Section 4.

[13] An alternative approach would be to use all existing, including domestic and MNC, establishment locations as the counterfactuals. This would help to control for the effect of general location factors instead of those that affect primarily MNCs' location decisions. In Section 6, we perform an analysis in that direction by employing domestic establishments as the benchmark and comparing the agglomeration patterns of MNC and domestic plants.

We compare the actual and counterfactual kernel estimators at various distance thresholds, including 200, 400, 800, and 1,600 kilometers (the maximum threshold being 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. Following EGK, for each industry pair k and $\widetilde{k}$, we obtain

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

or employment-weighted

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

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 greater than that of counterfactuals.

3.2 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.[14]

[Table 1 about here]

MNC offshore agglomeration For MNC foreign subsidiaries, the average agglomeration index value at the 200 km level is close to 0.1 percent, with the maximum value reaching 2.5 percent. In fact, only around 30 percent of industry pairs have a positive 200 km agglomeration index, that is, showing statistically significant evidence of agglomeration at 200 km. For the rest

[14] The scale of the agglomeration index is driven by the scope of the dataset and the empirical methodology. Because we take into account the distance of all establishment pairs worldwide (the maximum distance being around 20,000 km), 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.

of the industry pairs, MNC foreign subsidiaries do not display systematic patterns of agglomeration. At the more aggregate 400 km level, the average agglomeration index value increases to 0.2 percent, with the maximum reaching 5.4 percent, and nearly a third of industry pairs have a positive agglomeration index.

Industry pairs that exhibit some of the highest offshore agglomeration index values, reported in Table A.2, 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 As shown in Table 1, the average degree is slightly lower for the agglomeration of MNC subsidiary workers than for the agglomeration of individual subsidiaries. For example, at the 400 km level, the average value of the MNC worker agglomeration index is 0.18 percent, as opposed to 0.21 percent in the case of the MNC subsidiary agglomeration index. About 24 percent of industry pairs have a positive agglomeration index 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).

Correlations of the subsidiary and 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 more than 40 percent higher for MNC headquarters (0.13) than for foreign subsidiaries (0.09). This is consistent with the knowledge capital theory of multinational firms (see Markusen, 2002), which predicts that MNC headquarters locate in skilled-labor abundant countries and subsidiaries dispersedly distribute across host regions based on markets and comparative advantage.

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 for some industry pairs the clusters of MNC subsidiaries resemble those of headquarters, while for others the two types of establishments exhibit distinctly different agglomeration patterns. The emerging offshore clusters of MNCs are not merely a projection of their domestic clusters at home. The driving forces of MNCs' offshore agglomeration are likely to vary from those of headquarters, as we explore in Section 5.

4 Measuring FDI Location Fundamentals and Agglomeration Economies

After computing the agglomeration indices, we now turn to the economic factors that could account for the observed agglomeration patterns of MNCs. 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. We describe below how each of the above factors is measured in the empirical analysis.

4.1 FDI Location Fundamentals

We construct a measure of FDI location fundamentals, including market size, comparative advantage, and trade costs, by adopt a two-step procedure. We first estimate a conventional empirical equation following Carr, Markusen and Maskus (2001), Yeaple (2003a), and Alfaro and Charlton (2009). Specifically, we consider the following specification:

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

where $y_{c\widetilde{c}k}$ denotes either the number or the total employment of subsidiaries in country $\widetilde{c}$ and industry k owned by MNCs in country c, $marketsize\_ave_{c\widetilde{c}}$ is average market size proxied by the GDP of home and host countries,[15] $distance_{c\widetilde{c}}$ is the distance, $skill\_diff_{c\widetilde{c}}$ represents the difference in skill endowment, measured by average years of schooling, between the home and the host countries (i.e., $skill_{\widetilde{c}} \square skill_c$), $skillintensity_k$ is the skilled labor intensity proxied by share of non-production workers for each industry, $tariff_{c\widetilde{c}k}$ is the level of tariff set by the host country $\widetilde{c}$ on the home country c in industry k, and $\varepsilon_{c\widetilde{c}k}$ are the residuals. In addition to the above variables, host-country characteristics such as institutional and physical infrastructure could also affect multinationals' location decisions.[16] We thus include vectors of country-industry

[15] We consider, in addition to GDP, market potential which is the sum of domestic and distance-weighted export market size of the home and host countries.

[16] As noted by Helpman (2006), firms' sorting patterns and organization choices are dependent on the characteristics of the firms and the contractual environment (see, for example, Antras, 2003; Grossman and Helpman, 2002). Existing empirical evidence also suggests that institutional development (such as the rule of law) exerts a positive effect on the receipt of foreign investment (see Bénassy-Quéré, Coupet, and Mayer, 2007; Alfaro, Kalemli-Ozcan,

dummies, μ_{ck} and $\mu'_{\widetilde{c}k}$, to control for all country-industry specific factors such as institutional quality, physical infrastructure, domestic industry size, and economic policies.[17]

We estimate equation (5) using Poisson quasi-MLE (QMLE).[18] If market access is a significant motive in MNCs' investment decisions, we expect the effects of market size and trade cost (measured by distance and tariff) to be positive, that is, $_1 > 0$, $_2 > 0$, and $_5 > 0$. If comparative advantage is a significant motive, we expect the effect of trade cost to be negative and the effect of difference in skilled labor endowment to be negative for unskilled-labor intensive industries, that is, $_2 < 0$, $_4 > 0$, and $_5 < 0$. Our estimates are largely consistent with the literature (see, for example, Yeaple, 2003a; Alfaro and Charlton, 2009). Consistent with the market access motive, MNCs are found more likely to invest in countries with a larger market size ($_1 > 0$). Consistent with the comparative advantage motive, MNCs have a greater probability of investing in unskilled labor abundant countries ($_3 < 0$), especially in unskilled-labor intensive industries ($_4 > 0$), and trade cost exerts a negative effect on MNCs' investment decisions ($_2 < 0$ and $_5 < 0$).[19]

Based on the estimates of equation (5), we obtain and sum, for each host country $\widetilde{c}$ and industry k, the 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 density function based exclusively on the estimated effects of location characteristics including market size, comparative advantage, and trade costs. We compare in Section 5 the role of these characteristics relative to that of agglomeration forces in determining the spatial patterns of multinational firms.

and Volosovych, 2008, among others).

[17] 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.

[18] 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 estimated, respectively, the decision to trade and volume of trade, the results were similar.

[19] Results are suppressed because of space considerations and available upon request.

4.2 Agglomeration Economies

In addition to the location fundamentals of FDI, agglomeration economies also can 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 supplies of capital goods. Further, given their technology intensity, technology diffusion from closely linked industries can be particularly attractive to MNCs. We discuss below the role of each agglomeration economy in multinational firms' location choices and the proxies used to represent each force.

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.[20] 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.[21]

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 Data (specifically, the Detailed-Level Make, Use and Direct Requirement Tables) published by the Bureau of Economic Analysis, and define $IOlinkage_{k\widetilde{k}}$ as the share of industry k's inputs that come directly 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. Because firms' proximity to one another shields workers from the vicissitudes of firm-specific shocks, workers in locations in which other firms stand ready to hire them are often willing to accept lower wages.[22] Externalities can also occur as workers move from

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

[21] 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.

[22] 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 benefit because multiple firms offer protection against ex post appropriation of investments in human capital.

one job to another. This is especially true between MNCs which are characterized by similar skill requirements and large expenditures 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.).[23]

To examine labor market pooling forces, we follow EGK in measuring each industry pair's similarity in occupational labor requirements. We use the Bureau of Labor Statistics' 2006 National Industry-Occupation Employment Matrix (NIOEM), which reports industry-level employment across detailed occupations (e.g., Assemblers and Fabricators, Metal Workers and Plastic Workers, Textile, Apparel, and Furnishings Workers, Business Operations Specialists, Financial Specialists, Computer Support Specialists, and Electrical and Electronics Engineers). We convert occupational employment counts into occupational percentages for each industry, map the BLS industries to the SIC3 framework, and 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 arise as well in capital-good markets. This force, under-stressed in the literature, 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 (e.g. producers of specialized components and providers of machinery maintenance) and reduce the risk of investment (due, for example, to the existence of resale markets).[24] Local expansion of capital intensive activities can consequently lead to expansion of 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.[25]

Technology diffusion A fourth motive relates to the diffusion of technologies. Technology can

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

[24] 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.

[25] Note that 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.

diffuse from one firm to another through movement of workers between companies, interaction between those who perform similar jobs, or direct interaction between firms through 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., "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, for example, Jaffe et al., 2000; EGK), using patent citation flow data taken from the NBER Patent Database. The data, compiled by Hall et al. (2001), includes detailed records for all patents granted by the United States Patent and Trademark Office (USPTO) from January 1975 to December 1999. Each patent record provides information about the invention (e.g., technology classification, citations of prior art) and inventors submitting the application (e.g., name and city). We construct the technology diffusion variable, that is, $technology_{k\widetilde{k}}$, by measuring the extent to which technologies in industry k cite technologies in industry $\widetilde{k}$, and vice versa.[26] 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 maximum 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 while our analysis covers multinational activity around the world further mitigates the possibility of endogenous production, factor, and technology 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.

Table A.3 reports the summary statistics of industry-level control variables. Table A.4 presents the correlation matrix. For example, the correlation between industry-pair production linkage and similarity in capital-good demand is about 0.19, the correlation between production linkage and technology diffusion 0.29.[27]

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

[27] The table also shows the mean and 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.

5 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 4.1, and (ii) proxies for agglomeration forces described in Section 4.2 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 ($technology_{k\widetilde{k}}$). In addition to the location fundamentals and the agglomeration economies considered above, multinationals might also agglomerate because of factors like shared natural advantage and externality in institutional and physical infrastructure investment. We account for these factors using an industry fixed effect. Specifically, we include α_K, a vector of industry dummies that takes the value of 1 if either industry k or $\widetilde{k}$ corresponds to a given industry, and zero otherwise. These industry dummies control for all industry-specific factors and agglomeration patterns.

MNC offshore agglomeration We 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.[28] 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 a 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 aggregate geographic level.

[Table 3 about here]

[28] 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. Tables showing univariate results are suppressed from the paper due to space considerations but available upon request.

The lower panel of Table 3 reports the normalized beta coefficients.[29] 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, which suggests 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 MNCs' offshore agglomeration. The parameter of labor-market correlation is insignificant in the multivariate regressions.[30]

Comparing the estimates across distance thresholds, we find that at more aggregate geographic levels, the impact of technology diffusion diminishes and the effect of capital-good market externalities rises while the role of vertical production linkages remains mostly constant. 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 importance of location fundamentals and agglomeration economies, we find the effect of location fundamentals, which is significant at 1600 km, to dominate 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 the cumulative effect of agglomeration forces is 0.08 standard deviation. At the more disaggregated geographic levels, however, location-fundamental considerations do not appear to have a statistically significant effect and agglomeration forces become the driving forces. Table 4 reports a similar analysis excluding the location fundamental variable. The coefficients and statistical significance of the agglomeration forces remain largely unchanged.[31]

[Table 4 about here]

MNC offshore worker agglomeration We have examined MNC offshore agglomeration thus far using subsidiary as the unit of observation. We now take into account the different employment sizes of multinational subsidiaries, which 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. Note 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 exhibit significantly higher level of employment agglomeration.

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

[30] 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.

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

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, at more aggregate geographic levels, the effects of labor market externalities and technology spillovers diminish, while capital-good market correlation exerts a significant and positive effect. In contrast to the case for agglomeration of subsidiaries, the location fundamental variable plays a significant role at all distance thresholds, continuing to exert a stronger impact than agglomeration forces. At 200 km, a one-standard-deviation increase in location fundamentals and all agglomeration forces increases MNC subsidiary employment agglomeration by 0.31 and 0.17 standard deviation, respectively. Also noteworthy is that the impact of location fundamentals falls, and the importance of agglomeration forces rises slightly, at more disaggregated geographic levels, which suggests that the explanatory power of location fundamentals is greater for cross-country patterns of multinational activities than for the agglomeration of MNCs at the localized level, and vice versa for agglomeration economies.

[Table 5 about here]

MNC domestic headquarters agglomeration We next examine the determinants of MNC headquarters clusters relative to 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 4.1, 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, which suggests an important role for the characteristics of headquarter countries including market size, skilled labor endowment, and access to host countries. At 200 km, both technology diffusion and labor market correlation play a positive and significant role, with a cumulative effect of about 0.06. Beyond 200 km, the effect of labor market becomes insignificant and 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 on the agglomeration of MNC headquarters and, further, input-output relationships affect MNC subsidiaries but not headquarters. These results suggest that MNC subsidiary agglomeration is more influenced by market access and comparative advantage motives, capital market externalities, and vertical production linkages, whereas agglomeration of headquarters, with their specialization in R&D, management, and the provision of other services, is more influenced by technology diffusion than by production linkages.

6 Comparing the Agglomeration of MNC and non-MNC Plants

Having established the agglomeration patterns of MNCs, we now compare the effects of location fundamentals and agglomeration economies between multinational and non-multinational plants.

Conducting an empirical analysis of all domestic manufacturing plants is infeasible given the size of the entire WorldBase dataset and computational intensity of the procedure. Consequently, 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 yields in a final sample of 127,897 domestically owned plants.[32]

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 correlation of the MNC and domestic plant agglomeration indices is relatively low at around 0.2 at 200 km suggesting sharply different spatial patterns for multinational and non-multinational plants.

Next we formulate a counterpart of equation (6) for domestic plants. Taking the difference of the two equations 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 in 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 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

[32] A similar random sampling strategy was used in EGK.

aggregate geographic levels (e.g., 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.

7 Sensitivity Analysis: The Endogeneity of Agglomeration Economies

A potential concern with our analysis thus far is that the agglomeration economy measures might endogenously reflect the agglomeration patterns of multinational firms. For example, the input-output linkage between the apparel and cotton industries may reflect not just the inherent characteristics of apparel manufacturing, but also the agglomeration of the two industries due, for example, to availability of raw materials leading apparel manufacturers to favor cotton over other types of fabrics. Similarly, the technology spillover between the telecommunication and computer industries might be due not only to the intrinsic technological relationship between the two industries, but also to a historical factor that led the two industries to locate together and subsequently become familiar with each other's technologies.

This concern is mitigated in our paper by three factors. First, our analysis controls for the role of location fundamentals and industry-specific characteristics. This enables us to separate industries' geographic concentration due to location attractiveness from agglomeration activities driven by agglomeration economies. Second, our measures of agglomeration economies are constructed using U.S. industry account data while the paper examines global agglomeration patterns. U.S. industries' input-output linkages, factor market correlations, and technology spillovers are not very likely a result of agglomeration around the world. Third, the focus on MNCs reduces the possibility of reverse causation, as MNCs constitute a small subset of firms in each industry and the agglomeration economy measures are built with industry wide data that include information on domestic firms.

We nevertheless perform an additional exercise to further alleviate concerns about endogeneity. Because the global agglomeration patterns of multinational firms include the agglomeration of MNCs in the United States, we examine regional agglomeration for which the U.S. is excluded. If U.S. domestic industry-pair relationships are affected by the agglomeration of MNCs in the United States, then one would expect the former to be less likely to be affected by the agglomeration of MNCs located in other regions such as Europe. In this case, the agglomeration economy measures constructed with U.S. industry account data are orthogonal to the agglomeration patterns observed in Europe.[33]

[33] In examining the agglomeration of U.S. firms, EGK address the endogeneity of the U.S. agglomeration economy measures by instrumenting the variables with the U.K. counterpart measures. But using another country's data to instrument the agglomeration economy variables would not alleviate the potential for endogeneity in our analysis because it would face issues similar to the U.S. data. Using the U.S. agglomeration economy measures to predict

We proceed by repeating the procedure described in Section 3.1 to construct the agglomeration indices for MNCs located in Europe. These indices capture the degree to which MNCs in a given industry pair agglomerate in Europe at various threshold distances.

[Table 8 about here]

The results are reported in Table 8. We find the estimates to be qualitatively similar to those reported in Tables 4 and 5.[34] Multinational subsidiaries in industries with greater labor market correlation and technology spillover are found to have a higher level of agglomeration, especially at the 200 and 400 km levels. Input-output production linkage and capital-good market correlation also exert a significant effect on the agglomeration of MNCs in Europe. Consistent with the earlier results, we find the effects of labor market externalities and technology spillovers to diminish at more aggregate geographic levels. Further, labor market externality appears to be the strongest agglomeration force at disaggregated distance levels.

8 The Process of MNC Agglomeration

To shed further light on the formation of MNC clusters, in particular, the spatial interdependence between incumbents and entrants, we now turn from geographic patterns to the process of multinational agglomeration. Doing so also helps to 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 could give rise to the potential for reverse causality between MNC location patterns and measures of economic fundamentals and agglomeration economies. 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.[35] Although the low correlation between the indices of MNC agglomeration and domestic plant agglomeration reported in Section 5 suggests that this is not likely to be a significant issue, we take a further measure to address the concern.

Consequently, we explore in this section the dynamics of location decisions. Specifically, we distinguish new from incumbent plants and assess new MNC plants' propensity to agglomerate

the agglomeration patterns in a non-U.S. region would, however, mitigate the possibility of reverse causation and help identify the causal effects of agglomeration forces.

[34] Because we are now examining regional, instead of global, agglomeration, we consider only threshold distances up to 800 km.

[35] A related concern here is that when multinational establishments come into existence as a result of cross-border acquisitions, their agglomeration patterns might simply reflect the agglomeration patterns of domestic establishments. We argue that MNCs' acquisition decisions, like MNC location choices in general, are dependent on location fundamentals and agglomeration economies. Moreover, the option to restructure (including to retain or shut down) acquired plants further enables 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.

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 3, 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 9 about here]

We compare the agglomeration index for MNC entrants against two benchmarks. First, as in Section 5, we adopt domestic plants as the benchmark and compare how MNCs agglomerate with incumbent MNCs relative to the clustering of domestic plants. Table 9 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 technology spillover benefits, capital-good market externalities, and vertical production linkages are relatively stronger. 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 measures 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.

9 Conclusion

The emergence of new multinational clusters is 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. Our analysis, using a worldwide plant-level dataset and a novel index of agglomeration, yields a number of new insights into the industrial landscape of multinational production.

First, offshore clusters of MNCs are not simply a reflection of the domestic industrial clusters. Multinationals follow distinctively different agglomeration patterns offshore than their domestic counterparts. Second, FDI location fundamentals, although they play a significant role in explaining the agglomeration of multinational firms, 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. Third, the importance of location fundamentals and agglomeration economies varies significantly between MNCs' offshore agglomeration and the agglomeration of MNC headquarters and domestic plants. For example, MNCs' offshore plants are significantly more influenced than non-MNC plants by capital-good market and technological agglomeration factors. Finally, in the process of agglomeration, multinational entrants display stronger propensities to cluster with incumbent multinationals as opposed to incumbent local plants. Again, this is especially the case when capital-good market externalities and technology diffusion benefits are strong.

Two potential extensions of our analysis are worthy of particular attention. First, 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 yield additional policy insights. A second direction for future research involves micro patterns of agglomeration. Our analysis, like most existing research on agglomeration, has not explored potential heterogeneity within individual industrial clusters and how the role of firm heterogeneity might shape the formation of industrial clusters. Given such heterogeneous characteristics of firms as size and foreign ownership, the level of agglomeration centering on each firm could 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] Alfaro, Laura, Kalemli-Ozcan, Sebnem, and Volosovych, Vadym (2008) Why Doesn't Capital Flow from Rich to Poor Countries? An Empirical Investigation. *Review of Economics and Statistics* 90: 347-368.

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

[6] Antràs, Pol (2003) Firms, Contracts and Trade Structure. *Quarterly Journal of Economics* 118: 1375-1418.

[7] Antràs, Pol and Helpman, Elhanan (2008) "Contractual Frictions and Global Sourcing," in Elhanan Helpman, Dalia Marin, and Thierry Verdier (Eds.), *The Organization of Firms in a Global Economy*, Cambridge: Harvard University Press, forthcoming.

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

[9] Bénassy-Quéré Agnes, Coupet Maylis, and Mayer Thierry (2007) Institutional Determinants of Foreign Direct Investment. *World Economy* 30(5): 764-782.

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

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

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

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

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

[15] 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.

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

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

[18] Ellison, Glenn, Glaeser, Edward and Kerr, William (2010) What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review* 100: 1195–1213.

[19] Grossman, Gene M. and Helpman, Elhanan (2002) Integration Versus Outsourcing In Industry Equilibrium. *Quarterly Journal of Economics* 117(1): 85-120.

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

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

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

[23] Head, Keith and Mayer, Theirry (2004) The Empirics of Agglomeration and Trade. In: Henderson, J.V., and Thisse, J.-F. (Eds.), *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: Elvieser, pp. 2609-2669.

[24] 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.

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

[26] Helpman, Elhanan (2006) Trade, FDI and the Organization of Firms. *Journal of Economic Literature* 44(3): 589-630.

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

[28] 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.

[29] 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.

[30] 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.

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

[32] Lipsey, Robert E. (1978) The Creation of Microdata Sets for Enterprises and Establishments. *Annales de l'INSEE* 30(31): 395-422.

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

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

[35] 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.

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

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

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

[39] Ottaviano, Gianmarco I. P. and Puga, Diego, (1998) Agglomeration in the Global Economy: A Survey of the "New Economic Geography." *World Economy* 21(6): 707-731.

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

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

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

[43] 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.

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

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

[46] 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.

[47] 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 3 to a measure of global agglomeration that accounts for other forms of trade barriers including 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.[36]

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

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

and substitute the distance, contiguity, language, and PTA information for each pair of establishments into the equation to compute the fitted trade cost $\widehat{\tau}_{ij}$.[37]

[36] The estimation results are available upon request.

[37] 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). Because 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 3, we construct a agglomeration index based on the generalized measure of trade costs (instead of distance). We can see from Table A.5 which 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 ease of trading intermediate inputs and final goods due to low tariffs, country contiguity, and low language barriers are taken into account. For agglomeration forces to be meaningful, goods and factors must have little tradability (e.g., knowledge and physical capital) or, more generally, face high trade and movement barriers.

[Table A.5 about here]

Table 1: Descriptive Statistics for Multinational Agglomeration Indices

	# Obs.	Mean	Std. Dev.	Min.	Max.
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

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. 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 Endogeneity of Agglomeration Economy Measures: The Agglomeration Patterns of MNCs in Europe

	T= 200 kms	T= 400 kms	T= 800 kms
IO Linkages	0.104	0.248*	0.454**
	(0.079)	(0.157)	(0.209)
Capital Good	0.008	0.031*	0.044*
	(0.010)	(0.019)	(0.026)
Labor	0.031***	0.032*	0.036
	(0.008)	(0.018)	(0.030)
Technology	0.335**	0.514**	0.715**
	(0.151)	(0.262)	(0.393)
Location Fundamentals	-0.001	-0.004	-0.003
	(0.003)	(0.005)	(0.004)
# Obs.	7166	7166	7166
R^2	0.635	0.717	0.853
		Beta Coefficients	
IO Linkages	0.009	0.009	0.008
Capital Good	0.014	0.021	0.014
Labor	0.055	0.023	0.012
Technology	0.030	0.019	0.013
Location Fundamentals	-0.158	-0.087	-0.076

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 9: 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 ≤ 200	28.3	91.6	5.6	0.2	131.4
dist ≤ 400	54.8	194.1	24.5	0.4	268.7
dist ≤ 800	124.2	423.0	85.6	0.7	510.9
dist ≤ 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: Descriptive Statistics for Agglomeration Economies

	# Obs.	Mean	Std. Dev.	Min.	Max.
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: 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 A.4: 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.5: 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.