The Global Networks of Multinational Firms*

Laura Alfaro†
Harvard Business School and NBER

Maggie Chen‡
George Washington University

December 2009

Abstract

In this paper we characterize the topology of global multinational networks and examine the macro and micro patterns of multinational activity. We construct indices of network density at both pairwise industry and establishment level and measure agglomeration in a global and continuous metric space. These indices exhibit distinct advantages compared to traditional measures of agglomeration including the independence on the level of geographic aggregation. Estimating the indices using a new worldwide establishment dataset, we investigate both the significance and causes of multinational firm co-agglomeration. In contrast to the conventional emphasis of the literature on the role of input-output linkages, we assess the effect of various agglomeration economies. We find that, relative to counterfactuals, multinationals with greater factor-market externalities, knowledge spillovers, and vertical linkages exhibit significant co-agglomeration. The importance of these factors differs across headquarters, subsidiary, and employment networks, but knowledge spillovers and capital-market externalities, two traditionally under-emphasized forces, exert consistently strong effects. Within each macro network, there is a large heterogeneity across subsidiaries. Subsidiaries with greater size and higher productivity attract significantly more agglomeration than their counterfactuals and become the hubs of the network.

JEL codes: F2, D2, R1, R3
Key words: multinational firm, agglomeration, network, input-output linkage, knowledge spillover, factor market externality

---

*We thank Jim Anderson, Bruce Blonigen, Tarun Khanna, Keith Maskus, Mike Moore, John Ries, Stephen Yeaple, Tony Yezer, and participants in the ETSG plenary session, EIIT meeting and GW seminar for helpful comments and suggestions. We also thank William Kerr for kindly providing us the patent concordance data, Bill Simpson for excellent advice in calculating the agglomeration index, and Francisco Pino for help with the GIS software. We are grateful to Dun & Bradstreet and Dennis Jacques for helping us with the D&B data set and HBS and GW for financial support.

†Email: lalfaro@hbs.edu; Phone: 617-495-7981.
‡Email: xchen@gwu.edu; Phone: 202-994-0192.

1 Introduction

The past decades have witnessed an explosion in the activities of multinational corporations (MNCs). Sharp declines in trade and telecommunication costs have led to increasing separation of management and production facilities within individual firms. The rise of multinational firms represents a particularly extreme example of expanding geographic distance between firm leadership and production. Firms that agglomerated in Silicon Valley and Detroit now have subsidiaries clustering in Bangalore (termed the Silicon Valley of India) and Slovakia (nicknamed Detroit of the East).

Although many studies have examined the location choices of MNCs, few empirical analyses have investigated the global pattern of multinational firm agglomeration. Little is known about the interdependent networks of MNCs, the most active actors of globalization, and how they compare to the traditional autarkic industrial landscape. Do multinationals agglomerate with one another overseas? Do they agglomerate in the same fashion abroad as they do at home? What motivates the agglomeration? Finally, are there any hub and spoke corporations? An answer to these questions is central to the long standing debate over the consequences of foreign direct investment (FDI). MNC recipient countries often offer lucrative incentives to attract foreign investment and justify on the grounds of positive spillovers. The industrial countries that are home to most of the multinationals, on the other hand, are increasingly concerned about the possibility of job losses as capital moves abroad. Understanding the interdependencies in the MNC networks and how multinationals influence one another in their activities at home and overseas is critical to these debates.

We address the above questions by constructing the topology of global MNC networks and examining the significance and causes of multinational agglomeration.[1] Using a new worldwide establishment dataset and novel indices of agglomeration, we characterize both the *macro* and *micro* patterns of multinational production activity.

This paper contributes to the existing literature in four ways. First, we examine the global network of multinationals. In contrast to the existing literature where the majority of studies focus on agglomeration within a country or a region, we investigate the location interdependence of multinationals in a global and continuous metric space. This enables us to analyze the world landscape of multinational production without any constraint on geographic region or industry. It also permits us to expand the definition of agglomeration beyond traditional country borders to take into account agglomerative activities separated by administrative boundaries but proximate in trade costs.

Second, we construct indices of network density to measure the extent and statistical significance of multinational agglomeration. These indices, embedded in a simple theoretical network framework, build on a new empirical methodology from urban economics introduced by Duran-

[1] To be precise, we investigate patterns of agglomeration between industries, also referred to in the urban economics literature as co-agglomeration; see, e.g., Ellison, Glaeser and Kerr (2009). We use the two terms interchangeably in the paper.

ton and Overman (2005, henceforth DO). Albeit computationally intensive, the indices exhibit distinct advantages over traditional measures of agglomeration. In particular, they overcome the dependence of previous indices on the level of geographical disaggregation.[2] They also control for the overall distribution of MNCs and the effect of factors common to all MNCs. We construct the indices by first estimating the distance (and trade cost) kernel function of MNCs for each pair of industries based on bilateral establishment distance (and trade cost) data. We then adopt a Monte Carlo approach and obtain counterfactual kernel estimators using the entire MNC establishment dataset. We draw, for each industry pair, 1000 random samples of MNC establishments and compute the 95% global confidence band of kernel estimators. Multinationals in a given industry pair are considered to agglomerate only when their kernel estimates are statistically greater than the counterfactuals within a threshold distance. This approach, in conjunction with the use of fixed effects, helps us assess the geographic distribution of each industry pair relative to the overall distribution of MNCs and separate agglomeration from geographic concentration that results from, for example, host-country attractiveness.

Third, we assess the relative importance of various agglomeration forces, including inter-industry IO linkages, factor-market externalities, and knowledge spillovers, in explaining the cross-industry variation in multinational co-agglomeration. This contrasts with the bulk of the multinational literature, which emphasizes primarily the role of input-output (IO) linkages and has been generally silent on the effect of factor-market externalities and knowledge spillovers. We also compare the structure of MNC subsidiary networks with that of their headquarters counterparts. This comparison helps us differentiate the incentives of headquarters and subsidiary agglomeration and examine whether the increasing separation of headquarters services and production activities has led to different geographic patterns.

Finally, in a sharp departure from the existing literature, which focuses largely on aggregate-level agglomeration and assumes firms are created equal, we do not treat the network as a homogeneous entity. We take into account the role of firm heterogeneity in the formation of networks and allow the incentives to agglomerate to vary across firms. Specifically, we develop a micro counterpart of the macro network indices described above to measure the extent of agglomeration centering each multinational corporation. Based on these micro indices, we identify the hubs and spokes in the network and examine how firm heterogeneity, reflected in size and productivity, influences the ability to attract agglomeration.

To achieve the above goals, we employ a new worldwide establishment dataset that provides detailed location, ownership, and activity information for establishments in more than 100 countries. This dataset constitutes an ideal source of information for MNC studies and offers two distinct advantages over existing firm-level datasets. First, it includes nearly the world popu-

[2] As Head and Mayer (2004a) note, "measuring spatial concentration of activity is a far less trivial exercise that might seem at first sight". Previous studies in urban economics have generally used other indices such as the one developed by Ellison and Glaeser (1997). The accuracy of these indices depends crucially on the level and method of geographical disaggregation. The index presented here does not have the above issue. It is, however, extremely computationally intensive given the Monte Carlo nature of the methodology, which we discuss in Section 4.

lation of MNCs, enabling us to go beyond the study of individual countries and examine the topology of global multinational networks. Second, the dataset reports the physical address and postal code of each establishment while most existing firm-level datasets report business registration addresses. The latter information is extremely valuable for constructing a true index of agglomeration using continuous metrics of space and, more generally, trade costs.

We find significant evidence of co-agglomeration among multinationals. These patterns are driven by factor-market externalities, knowledge spillovers, and vertical linkages. Relative to counterfactuals, multinationals sharing similar investment-good and labor demands, technological know-how or vertical production relationships tend to locate near one another in both headquarters and subsidiaries.

The importance of these factors differs across headquarters, subsidiary, and employment networks. All agglomeration economies except input-output linkages (namely, knowledge spillover and capital- and labor-market externalities) exert a significant effect on the co-agglomeration of MNC headquarters, whereas in the case of subsidiary co-agglomeration all factors but labor-market externalities play a significant role. Capital-market externalities, in particular, pose a strong effect in the subsidiary network. The driving forces in the MNC subsidiary-employment network are knowledge spillovers and labor-market externalities, factors that require high labor mobility and close labor interaction. These findings stress the importance of taking into account agglomeration incentives other than input-output relationships. They also imply increasing specialization of headquarters and subsidiaries within each multinational firm, and consequently distinct agglomeration patterns.

Multinationals are far from equal within each pairwise industry network. We record considerable heterogeneity in the extent of agglomeration across multinational subsidiaries. Larger and more productive MNC establishments attract significantly greater agglomeration than their smaller counterfactuals. Subsidiaries with a greater revenue or employment and higher productivity tend to become the hubs of networks; smaller, less productive subsidiaries emerge as spokes.

Our findings have first-order policy implications. They suggest, for example, that policies aimed at influencing multinationals' location choices should take into account the interdependence in MNC networks. This interdependence goes beyond vertical production linkages; it can arise as a result of externalities in knowledge, physical capital, and labor markets. These externalities can not only exacerbate the outward and inward movement of multinationals, but also magnify the effects of FDI in home and host countries on, for example, factor markets and technology spillover.[3]

Our paper is closely related to three strands of literature. First, a number of studies in international trade including Head, Ries and Swenson (1995), Head and Mayer (2004a), Crozet,

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

Mayer and Mucchielli (2004), Blonigen, Ellis and Fausten (2005), Bobonis and Shatz (2007), and Amiti and Javorcik (2008) have examined the role of linkages in individual multinationals' location decisions. The results of these studies, which suggest that MNCs with vertical linkages tend to agglomerate within a host country/region, shed light on the role of vertical production relationship in multinationals' location choice.[4] We contribute to this literature by examining agglomeration in a global context.[5] By employing a new worldwide dataset and novel empirical indices, we are able to examine the global agglomeration of multinational firms in continuous metric space and at both macro and micro levels. Moreover, we take into account agglomeration incentives other than input-output linkages. Our findings suggest that knowledge spillovers and factor-market externalities, two traditionally under-emphasized forces, play an equally, if not more, important role in multinationals' location interdependence.

Second, our paper builds on the existing urban economics literature that examines domestic agglomeration. A number of papers in this area including Ellison and Glaeser (1997), Ellison, Glaeser and Kerr (2009, henceforth EGK) and DO assess the relative importance of agglomeration forces in explaining industrial localization in the U.S. and U.K. Three factors have been emphasized by these studies: market access to suppliers and customers, labor market pooling, and technology spillover. Evidence of co-location in U.S. industries shows, for example, that firms tend to locate near supplier and customer industries. As mentioned, we construct agglomeration indices based on the empirical methodology of DO. But in contrast to existing studies that focus on domestic patterns, we analyze the global landscape of multinational production.[6] We extend the traditional domestic spatial agglomeration index to a measure of global agglomeration. We also derive micro counterparts of the industry-level agglomeration indices and examine the heterogeneous ability of individual firms to attract agglomeration. In a further departure from the literature, we stress the role of external economies of scale in capital, a potentially important

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

[5] Our paper is also related to a recent work by Blonigen, Davies *et al.* (2007) who take the important step of investigating the cross-country spatial interdependence of FDI flows in the same industry. Using U.S. sectoral outbound FDI data, they examine how investments in third countries affect a country's receipt of FDI from the U.S. and find that while the results are sensitive to the sample of host countries examined there is some evidence of negative interdependence between proximate host countries. We investigate in this paper the significance and causes of spatial interdependence of multinational firms around the world. Using worldwide establishment data and a continuous metric of agglomeration, we take into account both within- and between-country interdependence and examine the motives of multinational co-agglomeration.

[6] In this paper, we focus on the global agglomeration of multinationals. In their study of U.K. manufacturing firms, Duranton and Overman (2008) compare the location patterns of domestic and foreign-owned establishments in the same industry and find mixed evidence: the former can be more localized than the latter in some industries but more dispersed in others.

factor for MNCs.

Finally, this paper relates to a growing recent literature on networks that analyzes the organization of firms (see, for example, Rauch, 1999, 2001; Jackson, 2008; Chen, 2009b). We complement this literature by introducing the notion of networks to the theoretical and empirical analysis of multinational firms. We use indices of network density and characterize the underlying structure of MNC geographic distribution. This synthesis between business network and multinational corporations becomes crucial as firms employ increasingly complex integration and sourcing strategies extending the traditional autarkic business network across borders and firms.

The rest of the paper is structured as follows. In Section 2, we review conceptual issues in FDI agglomeration. In Section 3, we present a theoretical framework that serves as a guide to the empirical analysis. We describe the construction of pairwise-industry network densities and micro level counterparts in Section 4 and discuss the data in Section 5. In Sections 6 and 7 we report, respectively, industry- and micro-level econometric analyses and empirical evidence. Section 8 conclude.

2 Conceptual Issues

A firm becomes multinational when it establishes in two or more countries business enterprises over which it exercises some minimum level of ownership control. Although location patterns of multinationals have long been considered complex, for analytical simplicity multinational activity has usually been classified into horizontal and vertical FDI. A firm engages in horizontal FDI when it replicates a subset of its activities or production processes in another country, in other words, when the same stage of the production process is duplicated. The firm engages in vertical FDI when it segments its production across countries, that is, when it breaks the value added chain.

These two types of FDI are first formalized in Markusen (1984) and Helpman (1984). Building on these two seminal papers, the existing literature has generally viewed FDI location as a choice driven by either the proximity-concentration tradeoff or factor price differentials. Multi-plant firms are motivated by the potential to save on either trade or production costs.

In this paper, we consider, instead, the role agglomeration incentives play in determining the location of multinationals. Since Marshall (1920), economists have recognized the importance of agglomeration benefits and argued that the industrial clusters we observe (e.g., Silicon Valley) can be explained by the cost and productivity advantages firms enjoy when they locate near one another. These advantages include (i) proximity to suppliers and customers, (ii) external scale economy in factor markets, and (iii) knowledge spillovers.[7]

[7] In addition to benefits, agglomeration can also incur costs (diseconomies) including increasing land price, labor cost, congestion, and other negative externalities (such as pollution). We provide more discussion on this in Sections 3 and 5.

Most of the literature, however, is focused on whether firms localize in a given country (mainly due to data limitations). We investigate, in this paper, the significance of co-agglomeration incentives in the global environment and how they affect multinational firms. Compared to domestic firms, multinational corporations often incur large trade costs in sourcing intermediate inputs and reaching downstream buyers. In the meantime, they face large market entry cost when relocating to a foreign country because of, for example, limited supplies of qualified labor and capital goods, both of which raise the value of external scale economy. MNCs are also well known for their technology intensity. For these firms, technology spillovers from closely linked industries can be particularly attractive. We review below the role of each Marshallian force in multinational firms' location choice.

Proximity to customers and suppliers Marshall (1920) argued that transportation costs induce plants to locate close to their inputs and customers and determine the optimal trading distance between suppliers and buyers.[8] This is especially true for MNCs given their large volumes of sales and intermediate inputs. Compared to the other types of firms, multinationals are often the leading corporations of each industry and command dominant market shares. Because they tend to be the largest customers of upstream industries and the largest suppliers of downstream industries, the input-output relationships can be far stronger than for average firms. We consider how vertical linkages motivate multinationals worldwide to co-agglomerate globally.[9] Furthermore, we investigate how the role of input-output linkages compares with those of the other Marshallian forces that have received less attention in the literature.

External scale economies in factor markets: Labor and capital Agglomeration can also yield benefits through external scale economies in factor markets. Firms' proximity to one another shields workers from the vicissitudes of firm-specific shocks. As a result, workers in locations where other firms stand ready to hire them are often willing to accept lower wages.[10] Externalities also occur as workers move from one job to another. This is especially true between MNCs because of their similar skill requirements and large expenditure on worker training. MNCs can have a particularly strong incentive to lure workers from other MNCs 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.).[11]

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

[9] 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.

[10] This argument has been formally considered in Marshall (1920), Krugman (1991), and Helsley and Strange (1990) and tested in Diamond and Simon (1990). Rotemberg and Saloner (2000), in a related motivation, argue that workers can also gain because multiple firms protect workers against ex post appropriation of investments in human capital.

[11] The flow of workers can also lead to knowledge spillover, a third Marshallian force which we discuss below.

External scale economies in the capital market, a force not usually emphasized in the literature, can also lead to agglomeration, in particular of the many multinational firms involved in capital-intensive activities. Geographically concentrated industries offer better support to providers of capital goods (such as producers of specialized components and providers of machinery maintenance). They also reduce the risk of investment (due, for example, to the existence of resale markets). As a result, local expansion of capital intensive activity can lead to expansion in the supply of capital goods exerting a downward pressure on the cost. This benefit is especially valuable to multinationals in FDI destinations where initial stock of capital goods is limited.

Knowledge spillovers A third motive relates to the flow of ideas that facilitate innovation and the development of new technologies. Knowledge can diffuse from one firm to another as workers move between companies, through interaction between people who perform similar jobs for different companies, or consequent to direct interaction between firms in technology sourcing. This has been noted by Navaretti and Venables (2005), who predict that MNCs may benefit from setting up affiliates in proximity to other MNCs with advanced technology, i.e., "the so-called centers of excellence." The affiliates can benefit from knowledge spillovers, which can then be transferred to other parts of the company.

The existing FDI literature has emphasized primarily the role of customer-supplier linkages and not systematically examined the effect of various agglomeration economies. In this paper, we take into account all the foregoing incentives and assess their relative importance in multinational agglomeration. To fulfill this goal, we construct formal indices of agglomeration.

3 Theoretical Framework

In this section, we consider a simple theoretical framework adopted from the network literature as a guide to the empirical analysis. In departure from most previous theoretical studies, we invoke the notion of *network* to describe the interdependence of firms in location decisions. This approach informs a number of alternative empirical indices of agglomeration including that introduced by DO. It also motivates several micro-level indices that quantify the extent of agglomeration centering each firm. These empirical indices, as discussed in Section 5, account for spatial continuity and exhibit distinct advantages compared to traditional measures of agglomeration including, for example, independence on the level of geographic aggregation.

3.1 Setup

Networks We use a *network*, G, to represent the location distribution of firms.[12] There are K industries, $k = 1, 2, ..., K$; each industry has N_k firms. We denote the set of firms in each industry as $\mathcal{N}_k$ and firms as nodes. We use $\mathcal{N}$ to represent the universe of firms and G the grand network. The grand network is the list of all pairs of firms $\{i, j\}$ between which there is a link. We use ij to represent the link $\{i, j\}$, so $ij \in G$ indicates that i and j are linked in the grand network.

Within the grand network are industry pair specific "sub" networks. We use $g_{k\widetilde{k}} \equiv \{ij : i \in \mathcal{N}_k, j \in \mathcal{N}_{\widetilde{k}}, k \neq \widetilde{k}\}$ to denote the network of industries k and $\widetilde{k}$ in which firms from industries k and $\widetilde{k}$ are connected. There are in total $K \cdot (K-1)/2$ such networks. The cardinality of these networks is represented by $n(g_{k\widetilde{k}})$. We use g_k to denote the overall network of industry k and $n(g_k)$ the cardinality of network g_k. By definition, $g_{k\widetilde{k}}$ is a subnetwork of g_k, i.e., $g_{k\widetilde{k}} \subset g_k$. Finally, we define a network as an empty network g^e when $ij = 0$ for all $i \in \mathcal{N}_k, j \in \mathcal{N}_{\widetilde{k}}$ and a complete network g^c when $ij = 1$ for all $i \in \mathcal{N}_k, j \in \mathcal{N}_{\widetilde{k}}$.

Distance We use the physical distance between two firms i and j, d_{ij}, to represent the distance of nodes in each network $g_{k\widetilde{k}}$.[13] Given our interest in the geographic distribution of MNCs, we take the links of each network as given and endogenously determine the distance between each pair of nodes. Put differently, we allow firms, the network nodes, to be *mobile* and adjust their distance to one another.

This approach differs from previous theoretical studies in which the choice variable is a dummy representing the decision to enter a given location (a country or region).[14] As noted by DO, empirical indices developed based on discrete-choice frameworks do not fully account for the continuity of space and can be biased depending on the level of geographic disaggregation. To develop indices unbiased with respect to the scale of geographic unit, we treat each network as a continuous metric space and distance of firms as the choice variable.

Utility and efficiency Now consider the utility function of each node. Let $u_i(g_k)$ (defined below) describe the utility (or profit) of firm i (in industry k) under the network g_k. Each firm's utility is a function of the benefits it receives from sharing links with others.[15] We use $\delta_{ij} \cdot d_{ij}^{-\varepsilon}$ to represent the benefits of link ij, which we assume to deteriorate in the distance of ij, i.e., d_{ij}. As described in Section 2, the benefits of sharing a link with another firm consist of three categories: input-output linkage, factor-market externalities, and knowledge spillovers. We assume that δ_{ij}

[12] We use the term firm loosely in this section. Starting from the next section, we introduce the term establishment to acknowledge the establishment-level nature of our dataset.

[13] In the empirical analysis (Appendix A), we also consider a comprehensive measure of trade cost to take into account other forms of trade barriers such as tariff, language, and border.

[14] See, for example, Venables (1996), Krugman and Venables (1996), and Puga and Venables (1997).

[15] Our theoretical framework, with its focus on firm interdependence, does not model characteristics that exert a common effect on firms (e.g., host-country natural advantages). These characteristics are controlled for in our empirical investigation through counterfactuals and fixed effects.

follows a Pareto distribution given by $F(\delta) = 1 - (b_{k\widetilde{k}}/\delta)^s$, where $b_{k\widetilde{k}}$ is the minimum benefit and s the shape parameter.[16] Formally, we consider the following utility function:

$$u_i(g_k) = \left[\sum_{\widetilde{k}} \sum_{ij \in g_{k\widetilde{k}}} \left(\frac{\delta_{ij}}{d_{ij}{}^{\varepsilon}} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

where σ denotes a constant elasticity of substitution between links and $\varepsilon < \sigma/(\sigma-1)$. In addition to the benefits, there are also costs of sharing a close link including, for example, competition for resources and remoteness from the other nodes. We assume these costs to increase in the level of proximity and lead to a constraint $\sum_{\widetilde{k}} \sum_{ij \in g_{k\widetilde{k}}} c/d_{ij} = E$ for each firm. Firms maximize $u_i(g_k)$ subject to the constraint, choosing a vector of distances, i.e., $\{d_{ij}\}$. To complete the utility discussion, we define a network as efficient when it maximizes $\sum_i u_i(g)$.

Optimal distance Maximizing equation (1) with respect to $\{d_{ij}\}$ yields:

$$d^*_{ij} = A \cdot \delta_{ij}^{-\lambda} \tag{2}$$

where $A \equiv E/\left(c \sum_j \delta_{ij}^{\lambda}\right)$ and $\lambda \equiv (\sigma-1)/[\sigma - \varepsilon(\sigma-1)]$. The above equation suggests that the optimal distance between firms i and j is a decreasing function of δ_{ij}, the benefit of the link between i and j. For nodes where $\delta_{ij} = 0$, $d^*_{ij} = \infty$, in which case $ij = 0$. For nodes where $\delta_{ij} = \infty$, $d^*_{ij} = 0$.

Next we develop various parameters of network structure to characterize the extent of agglomeration at macro and micro levels.

3.2 Pairwise industry network

3.2.1 Network density

We start with pairwise industry networks. We obtain the probability that nodes in network $g_{k\widetilde{k}}$ are within a threshold distance T and define this probability as the density of network $g_{k\widetilde{k}}$:

$$density_{k\widetilde{k}}(T) = \Pr_{ij \in g_{k\widetilde{k}}} (d_{ij} < T), \tag{3}$$

[16] To keep the theoretical analysis tractable, we assume here that $b_{k\widetilde{k}}$ is symmetric for each industry pair. Although the potential asymmetry of benefits between each pair of industries and consequent asymmetric agglomeration incentives are a question of notable theoretical interest, it does not add significant empirical insights to our analysis for two reasons. First, we can only observe each industry pair's status of agglomeration not their incentives to agglomerate. Second, in the empirical analysis we allow $b_{k\widetilde{k}}$ to be asymmetric within each industry pair whenever applicable and consider both the mean and the max of $b_{k\widetilde{k}}$ and $b_{\widetilde{k}k}$. We find that the two measures are highly correlated and the estimation results are largely similar.

where d_{ij} is the distance between firms i and j. Given equation (2), the above equation is equivalent to:

$$density_{k\widetilde{k}}(T) = \Pr_{ij \in g_{k\widetilde{k}}}(\delta_{ij} - \gamma > 0) = \int_{\gamma}^{\infty} dF(\delta_{ij}), \tag{4}$$

where $\gamma \equiv (A/T)^{1/\lambda}$. With $F(\delta) = 1 - (b_{k\widetilde{k}}/\delta)^s$, we obtain

$$density_{k\widetilde{k}}(T) = \left(\frac{b_{k\widetilde{k}}}{\gamma}\right)^s. \tag{5}$$

The above equation suggests that networks with higher $b_{k\widetilde{k}}$ exhibit greater density at any given T.

3.2.2 Expected proximity

Next we construct a measure of expected proximity of nodes in each network. Formally, we consider

$$proximity_{k\widetilde{k}}(T) \equiv \underset{ij \in g_{k\widetilde{k}}, d_{ij} < T}{E}\left[\frac{1}{d_{ij}}\right], \tag{6}$$

where $1/d_{ij}$ represents the proximity of nodes i and j. The above equation can be rewritten as:

$$proximity_{k\widetilde{k}}(T) = \int_{\gamma}^{\infty} \frac{1}{A} \cdot \delta_{ij}^{\lambda} dF(\delta_{ij}), \tag{7}$$

which can simplified to

$$proximity_{k\widetilde{k}}(T) = \alpha \cdot \left(\frac{b_{k\widetilde{k}}}{\gamma}\right)^s, \tag{8}$$

where $\alpha \equiv s\gamma^{\lambda}/(A(s-\lambda))$. Nodes in networks with higher $b_{k\widetilde{k}}$ are predicted to have greater expected proximity.

3.2.3 Centrality (degrees)

Next we consider the centrality (or degrees) of each industry k, denoted as $centrality_k$. The centrality measures the number of industries that agglomerate with industry k within a threshold distance T, i.e.,

$$centrality_k(T) = \sum_{\widetilde{k}} I_{k\widetilde{k}}\ [density(T) > 0], \tag{9}$$

where $I_{k\widetilde{k}}\ [density(T) > 0]$ is an indicator variable that equals 1 if $density(T) > 0$ for industries k and $\widetilde{k}$ and 0 otherwise. This equation suggests that the centrality of each industry is an increasing function of $\sum_{\widetilde{k}} b_{k\widetilde{k}}^s$.

3.3 Micro network

We now consider a more disaggregated network, the micro-network of each node i denoted as $g^m_{i\widetilde{k}}$. The network $g^m_{i\widetilde{k}}$ is the set of links in industry $\widetilde{k}$ owned by firm i, i.e., $g^m_{i\widetilde{k}} \equiv \{ij : j \in \mathcal{N}_{\widetilde{k}}\}$. By definition, $g^m_{i\widetilde{k}}$ is a sub-network of $g_{k\widetilde{k}}$, i.e., $g^m_{i\widetilde{k}} \subset g_{k\widetilde{k}}$. Similar to $g_{k\widetilde{k}}$, we derive the density of network $g^m_{i\widetilde{k}}$. To proceed, we first make some explicit assumptions about the functional form of δ_{ij}. We assume that $\delta_{ij} \equiv \theta_i \theta_j \underline{b}_{k\widetilde{k}}$, where θ_i and θ_j denote the productivity of nodes (firms) i and j, respectively, and $\underline{b}_{k\widetilde{k}}$ the baseline benefit between industries k and $\widetilde{k}$. We also assume that θ_i (and analogously, θ_j) follows a Pareto distribution given by $F(\theta) = 1 - (z/\theta)^s$, where z is the minimum productivity and s the shape parameter.[17] This means $b_{k\widetilde{k}} = z^2 \underline{b}_{k\widetilde{k}}$.

3.3.1 Network density

Similar to the pairwise industry network density $density_{k\widetilde{k}}(T)$, the density of $g^m_{i\widetilde{k}}$ is defined as the probability that nodes in network $g^m_{i\widetilde{k}}$ are within a threshold distance T, i.e.,

$$density_{i\widetilde{k}}(T) = \Pr_{ij \in g^m_{i\widetilde{k}}} (d_{ij} < T), \tag{10}$$

where d_{ij} is the distance between firms i and j.

Given $\delta_{ij} \equiv \theta_i \theta_j \underline{b}_{k\widetilde{k}}$, we can rewrite the above equation as

$$density_{i\widetilde{k}}(T) = \Pr_{ij \in g^m_{i\widetilde{k}}} \left(\theta_i \theta_j \underline{b}_{k\widetilde{k}} - \gamma > 0\right) = \Pr_{ij \in g^m_{i\widetilde{k}}} \left(\theta_j > \frac{\gamma}{\theta_i \underline{b}_{k\widetilde{k}}}\right), \tag{11}$$

where $\gamma \equiv (A/T)^{1/\lambda}$. This is equivalent to

$$density_{i\widetilde{k}}(T) = \left(\frac{z\theta_i \underline{b}_{k\widetilde{k}}}{\gamma}\right)^s. \tag{12}$$

Nodes with higher productivity θ_i are predicted to exhibit greater network density.

3.3.2 Expected proximity

We next consider the expected proximity of nodes in each micro-network. At each threshold distance T, this is given by

$$proximity_{i\widetilde{k}}(T) \equiv \underset{ij \in g^m_{i\widetilde{k}}, d_{ij} < T}{E} \left[\frac{1}{d_{ij}}\right] = \int_{\gamma/(\theta_i \underline{b}_{k\widetilde{k}})}^{\infty} \frac{1}{A} \cdot \left(\theta_i \theta_j \underline{b}_{k\widetilde{k}}\right)^{\lambda} dF(\theta_j). \tag{13}$$

[17] Several recent studies including Helpman, Melitz and Yeaple (2004), Yeaple (2009) and Chen and Moore (2009) have examined the role of firm productivity in multinationals' location decisions. We follow these studies in assuming that each firm draws a distinct productivity from a Pareto distribution function.

Simplifying the above equation yields

$$proximity_{i\widetilde{k}}(T) = \alpha \cdot \left(\frac{z\theta_i \underline{b}_{k\widetilde{k}}}{\gamma} \right)^s, \tag{14}$$

which indicates that $proximity_{i\widetilde{k}}(T)$ is positively correlated with firm productivity θ_i.

3.3.3 Centrality (degrees)

Finally, we construct the centrality (degrees) of each node (firm) denoted as $centrality_i(T)$. The centrality of each node is given by

$$centrality_i(T) = \sum_{\widetilde{k}} I_{i\widetilde{k}} \ [density^m(T) > 0], \tag{15}$$

and measures the number of industries that agglomerate with firm i within a threshold distance T. Nodes with a higher θ_i exhibit higher centrality in the grand network.

In the next section, we describe empirical indices that correspond to $density_{k\widetilde{k}}(T)$, $proximity_{k\widetilde{k}}(T)$ at the pairwise industry level and $density_{i\widetilde{k}}(T)$ and $proximity_{i\widetilde{k}}(T)$ at firm level. We then examine, in Section 5, the roles of agglomeration benefits ($b_{k\widetilde{k}}$) and firm heterogeneity (θ_i) in explaining these indices by formally estimating equations (5), (8), (12), and (14).

4 Empirical Methodology

Before we describe the empirical methodologies, we first note that an important goal of the empirical analysis is to disentangle the attributes of industry pair and micro networks from the attributes of the grand network. This is necessary to isolate the effect of factors common to all multinationals such as market size, labor cost, and natural advantage and to establish the role of industry-pair characteristics, i.e., $b_{k\widetilde{k}}$ (the benefits of proximity including input-output linkage, factor-market externalities, and knowledge spillovers) in explaining co-agglomeration decisions.

This issue has been identified by DO as one of the key five properties that should be satisfied by any measure of agglomeration. First, the index must be comparable across industries. Second, it should control for the overall location pattern of manufacturing. Third, it should account for industrial concentration. Fourth, it should be unbiased with respect to scale and aggregation. Finally, it should give an indication of the significance of the results. The indices we construct follow the methodology introduced by DO and satisfy all five requirements.

We use a new worldwide establishment dataset to examine the co-agglomeration of multinationals. As we describe in detail in Section 5.1, this dataset provides distinct advantages over the other existing data sources including worldwide coverage and information of plants' physical locations, both of which are essential to constructing an unbiased measure of global agglomeration.

4.1 Pairwise-industry network indices

We start with the pairwise-industry network parameters. As we discuss in Section 6, we are interested in three types of MNC networks: the networks of MNC headquarters, subsidiaries, and subsidiary employment. For each type of network, we construct the network parameters following the procedure described below and proceeding in three steps as in DO and EGK .

Step 1: Kernel estimator First, we calculate the distance between each pair of establishments. Specifically, we apply the Haversine formula to each establishment pair's geocodes and compute the great-circle distance.[18] This generates in total $N \times (N-1)/2$ number of bilateral distances where N denotes the number of establishments in the dataset. As we note in Section 5.1, there are 32,427 manufacturing MNC subsidiaries in our final sample, which means there are $525,738,951$ $(= 32,427 \times 32,426/2)$ bilateral distances.

Note that although the locations of nearly all establishments in our data are known with a high degree of precision, distance is an approximation of the true physical distance between establishments. One source of systematic error, for example, is that journey times for any given distance might differ between low- and high-density areas. Given the potential noise in the measurement of distances, we follow DO and adopt kernel smoothing when estimating the distribution of bilateral distances. In addition to distance, we also consider, in Appendix A, a generalized measure of trade cost to take into account the effect of tariff, language, and border. We follow the same procedure described here and construct measures of agglomeration that account for various forms of trade costs.

Let d_{ij} denote the distance between establishment i and j. The kernel estimator of bilateral distances at any point d (i.e., $f_{k\widetilde{k}}(d)$) is:

$$f_{k\widetilde{k}}(d) = \frac{1}{n_k n_{\widetilde{k}} h} \sum_{i=1}^{n_k} \sum_{j=1}^{n_{\widetilde{k}}} K\left(\frac{d - d_{ij}}{h}\right), \tag{16}$$

where h is bandwidth and K the kernel function. We use Gaussian kernel with the bandwidth set to minimize the mean integrated squared error.

We estimate the above kernel function $f_{k\widetilde{k}}(d)$ for each of the $7,875$ $(= 126 \times 125/2)$ industry pairs in the data.[19]

Step 2: Counterfactuals and global confidence bands Next, we estimate the kernel function of the grand network. We draw, for each of the 7,875 industry pairs, 1,000 random samples from the entire multinational establishment dataset. Note to control for the potential effect of industry size, it is important that the counterfactual industry in each sample has a number of observations similar to the actual data. We then calculate the bilateral distance

[18] We discuss in Section 5 how we obtain geocode information for each establishment.

[19] Given our focus on co-agglomeration, we drop all the identical industry pairs.

of each pair of establishments and obtain the kernel estimator for all 7,875,000 samples. This gives us $1,000$ kernel estimators for each of the $7,875$ industry pairs, $7,875,000$ kernel estimators total.[20]

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

Step 3: Network density and expected proximity We now construct the index of network density. For each industry pair k and $\widetilde{k}$, we obtain

$$density_{k\widetilde{k}}(T) \equiv \sum\nolimits_{d=0}^{T} \max\left(f_{k\widetilde{k}}(d) - \overline{f}_{k\widetilde{k}}(d), 0\right). \tag{17}$$

This captures the probability that, relative to the manufacturing sector as a whole, establishments in industries k and $\widetilde{k}$ agglomerate within the threshold distance T. Similarly, we compute the expected proximity of nodes in each pairwise industry network by considering

$$proximity_{k\widetilde{k}}(T) \equiv \sum\nolimits_{d=0}^{T} 1/d \cdot \max\left(f_{k\widetilde{k}}(d) - \overline{f}_{k\widetilde{k}}(d), 0\right). \tag{18}$$

The above index estimates each industry pair's average proximity, relative to its counterfactuals, within each threshold distance T.

4.2 Micro network indices

Next, we examine the heterogeneity embedded in the macro networks and measure the extent of agglomeration centering on each establishment. To do so, we develop a procedure analogous to the methodology described above. The micro network indices constructed based on this procedure control for the effect of location as well as industry characteristics and enable us to focus on the role of establishment heterogeneity in explaining differential ability to attract agglomeration. Again, we proceed in three steps.

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

Step 1: Kernel estimator First, we obtain, for each establishment, the kernel estimator of bilateral distances at any point d (i.e., $f_i(d)$). To make computation feasible, we include all the establishments that have been found to agglomerate with the industry to which the establishment of interest belongs, i.e., $density_{k\tilde{k}}(T) > 0$. Formally, we obtain

$$f_i(d) = \frac{1}{n_i h} \sum_{j:density_{k\tilde{k}}(T)>0} K\left(\frac{d - d_{ij}}{h}\right), \tag{19}$$

where n_i is the cardinality of i's micro network, h is bandwidth, and K the kernel function. As before, all kernel estimates are calculated using a Gaussian kernel with the bandwidth set to minimize the mean integrated squared error.

Step 2: Counterfactuals In the second stage, we construct a counterfactual kernel estimator for each establishment, i.e., $\overline{f}_i(d)$. Unlike the industry level indices for which we rely on samples randomly generated from the entire MNC dataset, we adopt here the mean kernel estimates of each host country and industry as the counterfactual. This enables us to control for all factors common to establishments in the same industry and located in the same country and to focus on each establishment's deviation from its average counterpart.

Step 3: Network density and expected proximity Finally, we construct the density index of each micro network, i.e.,

$$density_i(T) \equiv \sum\nolimits_{d=0}^{T} \left(f_i(d) - \overline{f}_i(d)\right). \tag{20}$$

This index captures the relative probability that other establishments agglomerate with i, as opposed to i's counterfactuals, within distance T. Similarly, we compute the expected proximity of nodes in each micro network, i.e.,

$$proximity_i(T) \equiv \sum\nolimits_{d=0}^{T} \frac{1}{d} \left(f_i(d) - \overline{f}_i(d)\right). \tag{21}$$

Establishments with the greatest density and average proximity are the hubs of the multinational network whereas those with relatively low network density and proximity play the role of spokes.

5 Data

5.1 The WorldBase database

We employ a new worldwide establishment dataset, WorldBase, for our empirical analysis. The data, compiled by Dun & Bradstreet (D&B) for 2005, includes more than 43 million plant-level observations in more than 205 countries and territories.[21] The unit of observation in WorldBase is the establishment rather than the firm.

In our empirical analysis, we include all foreign-owned manufacturing establishments. We describe an establishment as foreign-owned if it satisfies two criteria (i) it reports to a global parent firm, and (ii) the headquarters or parent firm is located in a different country. Parents are defined in the data as entities that have legal and financial responsibility for another establishment.[22]

We use, for each establishment, four main categories of information, (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 the establishment and headquarters, and (iv) operational information including sales and employment.

We drop establishments with zero or missing employment values.[23] Industries with fewer than 10 observations are also dropped from the analysis. This results in a final sample of 32,427 multinational subsidiaries (distributed across 126 manufacturing industries and 108 host countries). Top industries include Electronic Components and Accessories (367), Miscellaneous Plastics Products (308), Motor Vehicles and Motor Vehicle Equipment (371), General Industrial Machinery and Equipment (356), Laboratory Apparatus and Analytical, Optical, Measuring, and Controlling Instruments (382), Drugs (283), Metalworking Machinery and Equipment (354), Construction, Mining, and Materials Handling (353), and Special Industry Machinery except Metalworking (355). Among the top host countries are China, U.S., U.K., Canada, France, Poland, the Czech Republic, and Mexico.

As we discuss in Section 7, we also match our data with Orbis, another worldwide establishment data, to obtain detailed time-series financial information of multinationals such as revenue, value added, material cost and asset. The matching is done based on either the firm DUNS number or business name and location.

[21] We compared the U.S. owned subsidiaries in the WorldBase data 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. Across industries, we find the two databases to have not only similar accounts of establishments and activities, but also consistent distribution across countries. We also compared WorldBase with UNCTAD's Multinational Corporation Database. The two databases differ in that the former reports at the plant level and the latter at the firm level. Our analysis requires plant-level data. For the U.S. and other major FDI source countries, the number of firms is similar between the two databases, but WorldBase shows more plants. UNCTAD's data is also inflated by a large number of Chinese observations, which represent all approved FDI projects registered in the Chinese government but overestimate the number of actual foreign firms. See Alfaro and Charlton (2009) for a more detailed discussion of the WorldBase data and comparisons with other data sources.

[22] 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-ouput linkage should have incentives to locate near one another independent of ownership. See Yeaple (2003) for theoretical work in this area, and Chen (2009b) for supportive empirical evidence. One can use a similar methodology (calculating densities for establishments that belong to the same firm and counterfactuals for industries with the same number of firms and distribution of establishments) to study intra-firm interaction (see Duranton and Overman, 2008).

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

5.2 Geocode data

A distinct feature of the WorldBase dataset, beyond its worldwide coverage, is that it reports physical location including country, state, city, street address, and postal code, for all establishments including both headquarters and subsidiaries. This allows us to obtain, using a geocoding software (Yahoo! Geocoding API), the latitude and longitude for each establishment.[24] We compute, using the geocode data and the Haversine formula, the great-circle distance of all establishment pairs. This information is crucial for examining agglomeration in a continuous metric space. The majority of existing studies have relied on the number of establishments in a given region as the measure of agglomeration. In doing so, the results can be highly sensitive to geographic scale and level of aggregation.

From the between-establishment distance data, we can construct co-agglomeration indices at both the pairwise industry and establishment levels using the methodology described in Section 4. The pairwise-industry indices are constructed for each pair of SIC 3-digit manufacturing industries. The level of industry disaggregation in our analysis is dominated by the availability of control variables, as we explain next, and is the same as reported by EGK. We do not include within-industry pairs as we focus on the interdependence of two industries in location choice. This gives us in total $126 \times 125/2 = 7875$ pairs of industries.

5.3 Control variables: Industry-level characteristics

We now describe the industry-level variables employed to proxy the various FDI agglomeration economies including (1) proximity to input suppliers or industrial customers, (2) external scale economies in factor markets including labor and capital, and (3) knowledge spillovers.

Proximity to suppliers and customers To determine the importance of customer and supplier relationships for pairwise industries, we use the 2002 Benchmark Input-Output Accounts published by the Bureau of Economic Analysis.[25] We define Input-Output (IO) $linkage_{k\widetilde{k}}$ as the share of industry k's inputs that come from industry $\widetilde{k}$, and vice versa.[26] These shares are calculated relative to all input-output flows including those to non-manufacturing industries and final consumers. As supplier flows are not symmetrical, we take either the maximum or mean of the input and output relationships for each pairwise industry combination.

[24] This software uses a powerful commercial street dataset, well known as the industry standard for transportation data. It also provides more accurate geocode information than most alternative sources.

[25] Note that the assumption that U.S. IO structure (and similarly the structure of factor and technology demand discussed next) carries over to other countries can potentially bias our empirical analysis against finding a significant relationship. On the other hand, it also mitigates the possibility that our control variables are endogenous to the co-agglomeration of multinationals.

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

External scale economies in factor markets

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

Capital As mentioned, we also emphasize the importance of capital-market externalities, a force generally ignored in previous studies. To capture the potential for capital-market externalities, we construct a measure of industries' similarity in capital-good demand using capital flow data from the Bureau of Economic Analysis (BEA). The capital flow table (CFT), a supplementary table 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. As for the labor market variable, we measure each industry pair's similarity in capital structure, denoted by $capital_{k\widetilde{k}}$, using the correlation of investment flow vectors.[27]

Knowledge spillovers We follow EGK in using patent citation flow data to measure knowledge spillovers between industries. The data, taken from the NBER Patent Database compiled by Hall et al. (2001), includes detailed records for all patents granted by the United States Patent and Trademark Office (USPTO) from January 1975 to December 1999. Each patent record provides information about the invention (e.g., technology classification, citations of prior art) and inventors submitting the application (e.g., name and city). We construct the knowledge spillovers variable, i.e., $knowledge_{k\widetilde{k}}$, by measuring the extent to which technologies in industry k cite technologies in industry $\widetilde{k}$, and vice versa.[28] In practice, there is little directional difference in $knowledge_{k\widetilde{k}}$ due to the extensive number of citations within a single technology field. We obtain both max and mean for each pairwise industries.[29]

[27] As mentioned in Section 2, agglomeration can also incur costs such as increasing labor and land prices. It is possible that these costs are greater for industries with similar labor and capital demand in which case the estimated parameters of these variables would represent the net effect of similar factor demand structures on co-agglomeration decisions.

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

[29] The lower panel of Table 1 reports the summary statistics of the industry-level control variables. Appendix Table 1 presents the correlation. The table shows both the mean and the maximum measures for IO linkages and knowledge spillovers. As seen in the table, the two measures are highly correlated. We use average values in our analysis in Section 6 but obtain similar results when using the maximum measure (available upon request).

6 Empirical Evidence: Pairwise Industry Networks

We begin our empirical analysis by first examining the attributes of pairwise industry networks. We then investigate in Section 7 the heterogeneity within each macro network and how multinationals differ in their ability to attract agglomeration.

As mentioned in the introduction, previous studies in economic geography have focused on explaining the industrial clusters that have arisen in developed economies (e.g., Silicon Valley and Detroit) or location interdependence of MNCs in a given host country or region (e.g., Japanese firms in the U.S. or Europe). One of the main goals of this paper is to investigate how the rise of MNCs led to the formation of industrial clusters around the world and how these clusters compare to their headquarters counterparts. To this end, we examine networks of both multinational headquarters and multinational subsidiaries and establish their respective attributes.

6.1 Descriptive statistics

We first discuss the statistical attributes of the networks, beginning with network density. As shown in Table 1, the average density of worldwide multinational headquarters at 200km—in comparison to the manufacturing sector as a whole—is 0.1 percent.[30] The density rises as we raise the level of distance threshold, reaching 0.8 at 800km and 1.4 at 1600km. There is significant variation across industry pairs, with a third of industry pairs exhibiting statistically significant evidence of co-agglomeration in the headquarters network and the rest not differing significantly from the manufacturing sector as a whole.

The network density within a threshold distance is noticeably lower when we examine the worldwide subsidiary network. This suggests that the MNC subsidiary networks are more dispersed around the world than their headquarters counterparts. The above observations are also true when we look at the average proximity of nodes in each network. Multinational headquarters are more proximate to one another than their foreign subsidiary counterparts. Appendix Table 2 lists industry pairs that exhibit the highest subsidiary network densities at different distance thresholds. Industry pairs, such as Footwear except Rubber (314) and Boot and Shoe Cut Stock and Findings (313), Knitting Mills (225) and Footwear except Rubber (314), Miscellaneous Publishing (274) and Paperboard Mills (263), and Miscellaneous Publishing (274) and Miscellaneous Transportation Equipment (379), are among the top co-agglomerative pairs.[31]

[30] The low magnitude of network density is driven by the worldwide scope of our data and the empirical methodology. That we take into account the distance of all pairs of establishments around the world (the maximum distance being around 20,000km) determines the low values of kernel estimates at relatively short distance thresholds. The adoption of the Monte Carlo approach also means that the densities are constructed as the difference from the 95% global confidence bands. A positive value would represent statistically significant evidence of agglomeration. For our empirical analysis, we focus on the cross-industry variation in network density and how it is explained by the different agglomeration incentives of pairwise industries.

[31] Although the focus of this paper is on MNC agglomeration and thus domestic establishments are excluded in the analysis, we did compare the location distribution of MNC establishments with that of domestically owned plants. We found an insignificant correlation between the two worldwide. Duranton and Overman (2008) show a similar finding for the UK manufacturing establishments. This mitigates the potential concern that our indices of

Table 2 reports the correlation of network density indices. As shown, the densities are highly correlated across the different distance thresholds, for both headquarters and subsidiaries. Between the headquarters and subsidiary networks, the correlation of network densities is around 0.41 at 200km and rises as we raise the distance thresholds, reaching 0.59 at 1600km. This suggests that many industry pairs that co-agglomerate in their headquarters also tend to cluster in their foreign subsidiary locations, but for the remaining industries the structure of headquarters and subsidiary networks can be different.

[Tables 1 and 2 about here]

Appendix Table 3 reports the correlation of the expected proximity indices (the inverse of distance). Like network density, the expected proximity indices are highly correlated within headquarters and subsidiary networks across different distance thresholds. Between the headquarters and subsidiary networks, the correlation of the expected proximity indices is around 0.39, similar to the correlation of network densities.

Finally, we note that industries differ considerably in the degree of centrality in the grand network. We plot in Figure 1 the pairwise-industry headquarters network wherein each node represents an individual SIC industry and each link indicates the existence of a positive network density at the 200km level (i.e., statistically significant co-agglomeration within 200kms). We use the size of each node to represent the number of agglomerative links (i.e., centrality/degrees) of each industry. Industries represented by larger nodes are the center of a multinational network. It is clear that not all industries are equal. Some, such as Paperboard Mills (263), Newspaper Publishing, Publishing and Printing (271), Miscellaneous Publishing (274), Leather Products Luggage (316), Miscellaneous Primary Metal Products (339), Miscellaneous Transportation Equipment (379), and Watches, Clocks, Clockwork Operated Devices and Parts (387), attract more co-agglomeration than others. This observation is true as well for the network of multinational subsidiaries in Figure 2. The centrality of industries varies considerably; some, many the hubs of the headquarters network, are more centered than the others. The degree of co-agglomeration is, however, substantially lower in the subsidiary network as noted in Table 1.

[Figures 1-2 about here]

We now investigate the role of agglomeration economies in explaining the attributes of multinational networks.

6.2 Econometric analysis: Network density

Formally, we estimate equation (5) of Section 3, summarized in the following empirical specification:

$$density_{k\widetilde{k}}(T) = \alpha + \beta_1 IOlinkage_{k\widetilde{k}} + \beta_2 capital_{k\widetilde{k}} + \beta_3 labor_{k\widetilde{k}} + \beta_4 knowledge_{k\widetilde{k}} + D_K + \varepsilon_{ij}, \quad (22)$$

MNC agglomeration capture the agglomeration between MNC and domestic firms.

where $density_{k\widetilde{k}}(T)$ is the network density of pairwise industries (relative to the counterfactuals) and the right-hand side includes the benefits of co-agglomeration, $\delta_{k\widetilde{k}}$, which consist of input-output linkages, factor-market externalities, and knowledge spillovers. We also use a "consolidated" industry fixed effect by including D_K, a vector of industry dummies that take the value of 1 if either industry k or $\widetilde{k}$ corresponds to a particular industry and zero otherwise. These industry dummies control for all industry-specific factors, such as natural advantage and market structure, that might affect the location patterns of each industry.[32] We estimate network density for both headquarters and subsidiaries. As suggested by Figures 1 and 2 and as can be seen in Tables 1 and 2 (the correlation matrix), the headquarters and subsidiary networks can display different properties.

6.2.1 Headquarters network

Table 3 presents univariate regression results for the headquarters network. As the table shows, all variables have a positive and significant effect on their own. All the co-agglomeration forces including input-output linkages, capital- and labor-market externalities and knowledge spillover play a significant and expected role in explaining the extent of co-agglomeration. For example, a 10-percentage-point increase in the level of knowledge spillovers, measured by two industries' share of patent citations from each other, yields a 0.13-percentage-point (an equivalence of 100 percent) increase in the level of network density at 200km and a 0.26-percentage-point (an equivalence of 84 percent) increase at 400km.

The lower panel of the table presents standardized coefficients.[33] At the 200 km threshold, for example, a one standard-deviation increase in the extent of knowledge spillover is associated with a 0.04 standard-deviation increase in network density. The potential benefit of labor-market externality, measured by two industries' correlation in labor demand structure, has the largest coefficient (0.06). We obtain similar rankings across the different thresholds.

Table 4 reports the multivariate estimation results. All variables except input-output linkages continue to exert a significant effect. At the 200km and 400km thresholds, both knowledge spillovers and labor structure correlation exert a positive effect on the extent of headquarters co-agglomeration. Beyond 400km, the effect of labor structure becomes insignificant whereas the importance of capital structure correlation rises. This result is consistent with the lower mobility of labor in comparison to capital goods: co-agglomeration within short distances is important to realizing external scale economies in labor market.

[Tables 3-4 about here]

[32] The literature has considered the role of "natural advantage" (Ellison and Glaeser, 1997, 1999), also known as "first nature" (Krugman, 1993) and "locational fundamentals" (Davis and Weinstein, 2002). The closely related factor proportion theory takes the location of productive resources as given and uses it to explain the geographic distribution of production. We use fixed effect to capture these forces.

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

To assess the relative importance of co-agglomeration forces in the headquarters network, we report in the lower panel of Table 4 the normalized beta coefficients. Correlation in labor demand has the strongest effect on headquarters agglomeration at 200km (0.034), followed by knowledge spillover (0.026) and capital demand correlation (0.016). At distances beyond 400km, correlation in capital demand structure becomes the strongest factor.

These results differ from EGK's findings for U.S. manufacturing.[34] EGK find input-output relationships have the largest effect of all the Marshallian factors considered in their study, followed by labor market pooling. Intellectual spillover, in contrast, plays a weaker role. Our results show that factor-market externalities and knowledge spillovers exert the largest impact on multinational headquarters whereas input-output linkages do not play a significant role. This difference in findings suggests that, with the geographic separation of headquarters services and production activities, the determinants of MNC headquarters location are at variance with those of average manufacturing firms. The former take greater account of externality benefits in knowledge and factor markets; the latter place greater weight on proximity to suppliers and buyers. This is also in alignment with the distinct characteristics of multinational firms relative to their domestic counterparts including greater participation in knowledge, human, and physical capital investments and more engagement in export and import activities for input sourcing and output sales. It highlights the importance of distinguishing the structure of MNC networks from that of traditional industrial production.

6.2.2 Subsidiary network

Moving on to the network of MNC subsidiaries, we examine how co-agglomeration forces affect multinationals' location interdependence overseas.

Table 5 reports univariate regression results for each of our main variables. All the explanatory variables are highly significant across the different distance threshold levels. The estimated effects also exhibit expected signs. For example, at 400km, a 10-percentage-point increase in the level of knowledge spillovers, that is, the percentage of patent citations within two industries, leads to a 0.17-percentage-point increase in the density of the subsidiary network. This is equivalent to an 80-percent improvement over the average (0.2). Similarly, a 10-percentage-point increase in the level of IO linkages gives rise to a 0.1-percentage-point (47 percent) increase in the density.

The lower panel of Table 5 reports the normalized beta coefficients. In contrast to the headquarters network in which labor-demand correlation and knowledge spillovers have the strongest effects at 200km and 400km, capital-market correlation exerts the greatest impact across all distance thresholds in the case of subsidiary networks, followed by labor-structure correlation, knowledge spillovers, and input-output linkages.

[Tables 5-6 about here]

[34] Our methodology in calculating the density index follows DO and differs slightly from EGK.

Table 6 reports the full regression results. When we include all the explanatory variables, IO linkages, capital-market correlation and knowledge spillovers all play a significant role. This finding contrasts with the bulk of the literature which emphasizes primarily the effect of IO linkages on multinationals' subsidiary location interdependence. Here, we find that other co-agglomeration benefits, i.e., knowledge spillovers and capital-market externalities, play an equally, if not more, important role. Ignoring these factors can potentially bias the understanding of the determinants of MNC co-agglomeration. The parameter of labor-market correlation becomes insignificant in the multivariate regressions. One possible explanation for this result is multinationals' concentration on production activities in foreign subsidiaries and the motive to search for the cheapest production labor market in which external scale economies play a less significant role.

Considering the standardized coefficients, we find that, among all the agglomeration forces, capital-market correlation has the strongest effect (around 0.04), followed by knowledge spillovers and IO linkages.[35] This ranking holds across the different distance thresholds. The relative importance of the factors differs between the headquarters and subsidiary networks. For the headquarters network, labor-market externalities and knowledge spillovers are the strongest factors at short distances, whereas capital-market externalities consistently have the largest effect on subsidiaries. Input-output relationships play a significant role in the subsidiary network, but not among headquarters. This confirms the distinct specialization of headquarters and subsidiaries: the former become increasingly specialized at management, research, marketing, and the provision of other services, whereas the latter concentrate on production activities.[36]

6.2.3 Subsidiary employment network

So far we have constructed the network indices using establishment as the unit of observation. We now consider an alternative measure that takes into account the different employment sizes of multinational subsidiaries. This essentially treats worker as the unit of observation and measures the level of co-agglomeration among workers who belong to establishments in different industries. This type of measures enables us to examine the structure of MNC subsidiary-employment network and compare its determinants with the determinants of the network of individual subsidiaries. It can also lead to useful implications for policy-making targeted at influencing the geographic distribution of workers.

To proceed, we first obtain a weighted kernel estimator where we weigh each establishment

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

[36] In Appendix A, we expand the analysis of the subsidiary network and re-construct the network densities based on a generalized measure of trade cost that takes into account the role of other trade barriers including language, tariffs, and contiguity.

by the size of employment. This is given by

$$f^{w}_{k\tilde{k}}(d) = \frac{1}{h\sum_{i=1}^{n_k}\sum_{j=1}^{n_{\tilde{k}}}(r_i r_j)}\sum_{i=1}^{n_k}\sum_{j=1}^{n_{\tilde{k}}} r_i r_j K\left(\frac{d-d_{ij}}{h}\right) \quad (23)$$

where r_i and r_j represent, respectively, the number of employee in establishments i and j. We do this for both the $7,875$ actual and $7,875,000$ counterfactual kernel estimations at the industry pair level. We then construct the employment network density for each of the pairwise-industry networks and re-estimate equation (22).

The average density of the employment network is 0.1 percent at 200km and 0.7 at 1600km. Appendix Table 4 presents the correlation of the employment network densities. As they were for the other network densities, the indices are highly correlated across the distance thresholds. Appendix Table 4 also compares employment network densities to those of the subsidiary network counterparts. The indices show a correlation of around 0.42 at 200km and 0.66 at 1600km. These correlations suggest that the co-agglomeration pattern of MNC subsidiary employment can be different from that of individual MNC subsidiaries, especially at shorter distances.

Appendix Table 5 lists the industry pairs that exhibit the highest employment network densities at different distance thresholds. Although the list includes some of the top co-agglomerating pairs listed in Appendix Table 2, other industries appear as well. For example, Dolls, Toys, Games and Sporting and Athletic (394) and Footwear, Except Rubber (314), Dolls, Toys, Games and Sporting and Athletic (394) and Boot and Shoe Cut Stock and Findings (313), and Knitting Mills (225) and Footwear except Rubber (314) are among the top co-agglomerative pairs in the subsidiary employment network.

Table 7 reports univariate regression results for each of our main variables. All of the explanatory variables are highly significant across all distance threshold levels. At 200 km, for example, knowledge spillover has the largest standardized coefficient when included alone, followed by labor-market correlation and capital-structure correlation. Table 8 shows the multivariate estimates. We notice that in contrast to Table 6 in which labor-market correlation does not exert a significant effect, multinational subsidiaries with greater potential labor-market externalities are found to have a significantly higher level of employment co-agglomeration. Knowledge spillover, another force of agglomeration that involves close labor interaction and mobility, also plays a significant role in explaining the density of MNC subsidiary-employment network. In fact, knowledge spillover appears to be the strongest factor relative to the other forces at most distance thresholds. Capital-market correlation, albeit less important in this context, especially at shorter distances, continues to exert a significant and positive effect on the extent of co-agglomeration.

[Tables 7-8 about here]

6.3 Econometric analysis: Expected proximity

We next examine another measure of agglomeration embedded in the theoretical framework, expected node proximity. In comparison to network density, which represents the probability of agglomeration within a certain distance threshold, this index measures the expected level of proximity within the distance threshold. It captures not only the statistical significance of agglomeration but also the average proximity, defined as the inverse of distance, of agglomerative establishments. We follow the empirical procedure described in Section 4 and construct an index of expected proximity for each of the $7,875$ industry pairs. We then estimate this index as a function of co-agglomeration economies using equation (22) where the dependent variable is now $proximity_{k\widetilde{k}}(T)$. Industry pairs that especially value close interaction are expected to exhibit greater expected proximity.

Headquarters network Table 9 presents the univariate results for the headquarters network. All variables of interest have a positive and significant effect on their own. The lower panel of the table presents standardized coefficients. Both the ranking and magnitude of the estimates are largely similar to those reported for network density (Table 3). At the 200 km threshold, for example, a one standard-deviation increase in the extent of knowledge spillovers is associated with a 0.04 standard-deviation increase in expected proximity. The labor correlation variable has the largest estimated effect (0.061), followed by knowledge spillovers and capital-market externalities.

Table 10 reports the multivariate estimation results. Labor and knowledge spillover variables exert a positive and significant effect at all distance thresholds. The standardized coefficients in the lower panel of Table 10 show labor-demand correlation to have the strongest effect on the expected headquarters proximity (e.g., 0.041 at 200km), followed by knowledge spillovers (0.027). These results further highlight the central role played by labor market and knowledge externalities in determining the structure of headquarters networks. Headquarters with strong interaction in knowledge and human capital market are not only more likely to co-agglomerate but also more proximate to one another. This finding is not surprising given the limited mobility of labor relative to final goods, intermediate inputs, and capital goods. The latter, being more mobile and tradeable, pose less of a constraint on geographic proximity.

[Tables 9-10 about here]

Subsidiary network We now consider the proximity of subsidiaries. Table 11 reports the univariate regression results for the subsidiary network. All variables are highly significant on their own and display the expected signs. At 200 km, for example, knowledge spillovers, labor, and capital exhibit estimated effects of around 0.051 whereas the effect of IO linkages is around 0.028. In the multivariate regression results shown in Table 12, knowledge spillovers (0.036), capital correlation (0.029), and IO linkages (0.013) all have a strong effect while labor has an insignificant parameter. These results are similar to those reported in Table 6, in which we show that industries that share a strong link through knowledge and physical capital or vertical input-output relationships tend to co-agglomerate in the subsidiary network. Here, we find that these industries are also more proximate, lending further support to the importance of geographic proximity in achieving externalities in knowledge and physical capital market.

[Tables 11-12 about here]

7 Empirical Evidence: Micro Networks

Having established the attributes and causes of industry co-agglomeration, we now examine multinational firms' heterogeneous ability to attract agglomeration. Section 3.3 predicts that firms with higher productivity become the hubs of the networks whereas the less productive emerge as spokes. To test this hypothesis, we estimate equation (12) from Section 3 using the following empirical specification:

$$density_i(T) = \alpha + \beta\theta_i + D_k + D_c + \varepsilon_i, \tag{24}$$

where $density_i(T)$ is the estimated density of establishment i's network that captures the probability of other establishments agglomerating around i, as opposed to i's counterfactuals in the same host country and industry, within a threshold distance T. We obtain estimates of $density_i(T)$ based on the methodology described in Section 4.2 for all subsidiaries for which industries are found in Section 6 to co-agglomerate with i's industry.[37] On the right hand side of equation (24) is the productivity of each subsidiary i denoted by θ_i; either subsidiary size or various measures of productivity are used to proxy for θ_i. In addition to θ_i, we include a vector of industry and country dummies, represented by D_k and D_c, to control for industry- and country-specific factors such as sectoral concentration, host-country natural advantage, and factor endowments. This enables us to focus on the effect of subsidiary heterogeneity in determining the extent of agglomeration. Table 13 reports the summary statistics for $density_i(T)$ and some of the establishment characteristics.

The estimation results based on subsidiary size and labor productivity are presented in Table 14. We find a positive and statistically significant relationship between subsidiary size, measured in terms of employment or revenue, and extent of co-agglomeration. Larger multinational subsidiaries attract significantly greater co-agglomeration than their smaller counterfactuals. Subsidiaries with higher labor productivity also tend to become centers of the networks as expected from Section 3.3. This result is true across all distance thresholds. In addition to network den-

[37] Subsidiaries in industries in which there is no evidence of co-agglomeration with the industry of interest are excluded here since the focus of the micro-level analysis is to explore the differential ability of subsidiaries in each industry to attract agglomeration as a function of subsidiary productivity. Industries that do not exhibit any evidence of co-agglomeration are not relevant to this analysis.

sity, we also considered expected proximity in each subsidiary's micro network. The finding is similar: larger, more productive subsidiaries evidence greater proximity to other establishments.

[Tables 13-14 about here]

So far we have used either subsidiary size or labor productivity as a proxy of θ_i. Two potential concerns can arise with respect to these measures. First, the two variables (in particular, subsidiary size), albeit closely related to productivity, may not fully capture subsidiaries' total factor productivity (TFP). Second, one may argue that there is a potential reverse causality between these characteristics and the level of agglomeration around each subsidiary leading to potential endogeneity.

To address these issues, we next consider a formal measure of total factor productivity for each multinational subsidiary. Specifically, we obtain for each subsidiary's headquarters firm detailed financial information from which we estimate the headquarters' total factor productivity. Instead of using the subsidiary's productivity information directly, we include spatial and time lags between the measures of productivity and agglomeration. We use the productivity of the headquarters firm in a lagged period as a proxy for subsidiary efficiency and assess whether subsidiaries with more productive headquarters are more capable of attracting agglomeration. By providing headquarters services such as management and R&D, the productivity of headquarters has a direct effect on the efficiency of subsidiaries. Using the former (from a lagged period) mitigates the potential endogeneity concern that arises with direct measures of subsidiary efficiency.

To proceed, we employ a complementary data source, Orbis, another worldwide establishment database that reports financial, ownership and location information for public and private companies in more than 100 countries. Compared to WorldBase, Orbis provides more comprehensive financial information and reports income and balance sheet variables. This information makes it possible to formally estimate total factor productivity. The challenge, however, is to match the two establishment databases given their different main identification systems.[38] We conduct the matching based on either the DUNS number (for establishments for which both databases report this information) or business name and address. We proceed manually when using the latter matching criteria. This gives us an approximately 70-percent match rate.[39]

For each matched subsidiary, we obtain detailed financial data including revenue, value added, material cost, employment, and capital for the period of 1998-2005. We then estimate, based on the information, the productivity of each multinational using the semiparametric estimator developed by Levinsohn and Petrin (2003). We use productivity estimates in 2000 to allow for

[38] The publisher of WorldBase, Dun & Bradstreet, introduced and uses the Data Universal Numbering System (the D-U-N-S number) to identify businesses numerically for data-processing purposes. The system supports the linking of plants and firms across countries and has been widely adopted. The publisher of Orbis, Bureau Van Dijk, employs a different identification system that assigns an account number to each establishment. It reports the DUNS numbers only for a select number of countries including the U.S. and some western European nations.

[39] Among the matched multinationals, we lost additional observations due to the lack of financial data in Orbis.

a five-year lag between the measures of headquarters productivity and estimated agglomeration at the subsidiary level as well as to reduce the possibility of reverse causality.

The results are qualitatively similar to those reported in Table 14. Subsidiaries are far from equal. Those with more productive headquarters attract significantly more co-agglomeration than their counterfactuals. A one standard-deviation increase in TFP leads to a 0.06 standard-deviation increase in the extent of agglomeration at 400km and a 0.08 standard-deviation increase at 800km. The more productive are clearly more centered than less productive spoke subsidiaries.

8 Conclusion

In this paper, we characterize the topology of global multinational networks. We construct indices of network density and measure the extent and statistical significance of multinational co-agglomeration in a continuous metric of space. Using a new worldwide establishment dataset that reports the physical locations of multinational establishments around the world, we examine both the macro and micro patterns of multinational production.

We find that, relative to counterfactuals, multinational subsidiaries with greater factor-market externalities, knowledge spillovers, and input-output linkages tend to agglomerate to one another. The importance of these agglomeration economies is, however, different across headquarters, subsidiary, and employment networks. All agglomeration motives except input-output linkages exert a significant effect on the co-agglomeration of MNC headquarters. In the case of subsidiary co-agglomeration, all factors but labor-market externalities play a significant role. Capital-market externalities, in particular, exert a strong effect. Knowledge spillovers and labor-market externalities are driving forces in explaining co-agglomeration in the MNC subsidiary-employment network. These results provide further evidence of the increasing separation of headquarters services and production activities within multinational firms and suggest that the differential specialization of headquarters and subsidiaries engenders distinct agglomeration patterns.

Within each industry network, we identify significant heterogeneity across multinational subsidiaries. We find a positive and significant relationship between subsidiary productivity and the extent of agglomeration: larger, more productive multinational establishments attract significantly greater agglomeration than their smaller counterfactuals.

These results suggest that more consideration should be given to the interdependence of multinational firms in FDI policy. A preferential policy scheme whereby favorable incentives are offered first to industries with the greatest positive externalities might be more effective than a uniform incentive system. But there are a number of reasons for interpreting potential policy implications of our results with caution. Because many factors play a role in firms' location decisions and the realization of Marshallian externalities is a complex process, it may not be possible, for example, for a country to duplicate the circumstances that led to agglomeration in other nations. Clearly, more research is needed to further clarify the role of policy.

Two potential extensions are worthy of particular attention. First, like the majority of the existing economic geography literature, we focus on examining patterns of agglomeration and have not investigated the dynamics of the agglomeration process. The network of multinationals is, however, likely to evolve over time in response to, for example, technological progress, financial development, and declining trade and telecommunication costs. Examining how the attributes of MNC networks vary with the proliferation of globalization, facilitated through the use of time-series establishment data, could shed additional light on the effect of economic policies. Second, the structure of MNC networks can vary across regions. Because more rigid labor markets impose greater constraints on labor mobility, labor-market externalities offer an especially strong incentive for agglomeration in countries with such 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 co-agglomerate with suppliers and customers in countries with poorer infrastructure. Further analysis of the role of regional characteristics in determining the structure of MNC networks could provide even more policy insights.

References

[1] Alfaro, Laura and Charlton, Andrew (2009) Intra-Industry Foreign Direct Investment. *American Economic Review* 99(5): 2128-2151.

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

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

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

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

[6] Blonigen, Bruce, Davies, Ronald, Waddell, Glenn and Naughton, Helen (2007) FDI in Space: Spatial Autoregressive Relationships in Foreign Direct Investment. *European Economic Review* 51(5): 1303-1325.

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

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

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

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

[11] Chen, Maggie and Moore, Michael (2009) Location Decision of Heterogeneous Multinational Firms. *Journal of International Economics*, forthcoming.

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

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

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

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

[16] 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.

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

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

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

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

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

[22] 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.

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

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

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

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

[27] 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.

[28] Jackson, Matthew (2008) Networks and Economic Behavior. *Annual Review of Economics* (forthcoming).

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

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

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

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

[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. (2000) The theory of Endowment, Intra-industry and Multi-national trade. *Journal of International Economics* 52(2): 209-234.

[35] Marshall, Alfred (1920) *Principles of Economics*. London, UK: MacMillan and Co.

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

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

[38] Rauch, James (1999) Networks versus Markets in International Trade. *Journal of International Economics* 48(1): 7-35.

[39] Rauch, James (2001) Business and Social Networks in International Trade. *Journal of Economic Literature* 39(4): 1177-1203.

[40] 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.

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

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

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

Appendix A: Agglomeration indices with a generalized measure of trade cost

As explained in the introduction, most of the FDI literature has focused on agglomeration within a host country or region. Agglomerative activities by plants located in two separate countries but proximate in distance and trade cost have been largely ignored. We addressed the issue by measuring agglomeration with continuous spatial metrics. We have not, however, taken into account the role of other trade barriers such as language and tariffs that can affect trade cost at any distance. In this section, we extend the spatial agglomeration indices constructed in Section 4 to a measure of global agglomeration that accounts for various forms of trade costs. This can result in potentially different effects of agglomeration economies as, for example, capital and intermediate inputs are more tradeable than labor and knowledge capital.

We follow Head and Mayer (2004b) and Chen (2009a) and employ a two-step procedure for estimating a comprehensive measure of trade cost for each pair of subsidiaries. We first estimate a standard trade gravity equation given as

$$Q_{ijt} = EX_{it} + IM_{jt} + \lambda Z_{ijt} + \varepsilon_{ijt}, \tag{25}$$

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

We use a dataset that covers trade flows between 80 countries to estimate the gravity equation. We obtain the trade data from the COMTRADE database and geographic information, including distance, border, and language, from the CEPII distance dataset. The PTA information is from the Tuck Trade Agreements Database and 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.[40]

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

$$\tau_{ij} = -\widehat{\lambda}_1 \ln d_{ij} - B_{ij}(\widehat{\lambda}_2 + \widehat{\lambda}_3 L_{ij}) - \widehat{\lambda}_4 PTA_{ijt} \tag{26}$$

and substitute the distance, contiguity, language, and PTA information for each pair of sub-

[40] For a comprehensive review in this area, see Anderson and van Wincoop (2004).

sidiaries into the equation to compute the fitted trade cost τ_{ij}.

We then repeat the methodology described in Section 4 to construct indices of agglomeration based on the generalized measure of trade costs. Appendix Table 6 reports the univariate results. All variables are highly significant on their own and display the expected signs. At 200 km, the knowledge spillover variable exhibits the greatest effect when estimated on its own (0.112), followed by labor and capital (0.064 and 0.058, respectively), and finally IO-Linkages (0.021). Appendix Table 7 shows the multivariate regression results. Knowledge spillovers (0.108) and capital market externalities (0.028) have a positive and significant effect. The labor and the linkages variables are not significant. These results suggest that when the ease of trading intermediate inputs, final goods, and labor due to low tariffs, country contiguity, and low language barriers are taken into account, labor and IO linkages do not play a significant role in explaining the co-agglomeration of MNC subsidiaries. For agglomeration forces to be meaningful, goods and factors must have little tradeability (such as knowledge and physical capital) or, more generally, face high trade and movement barriers.

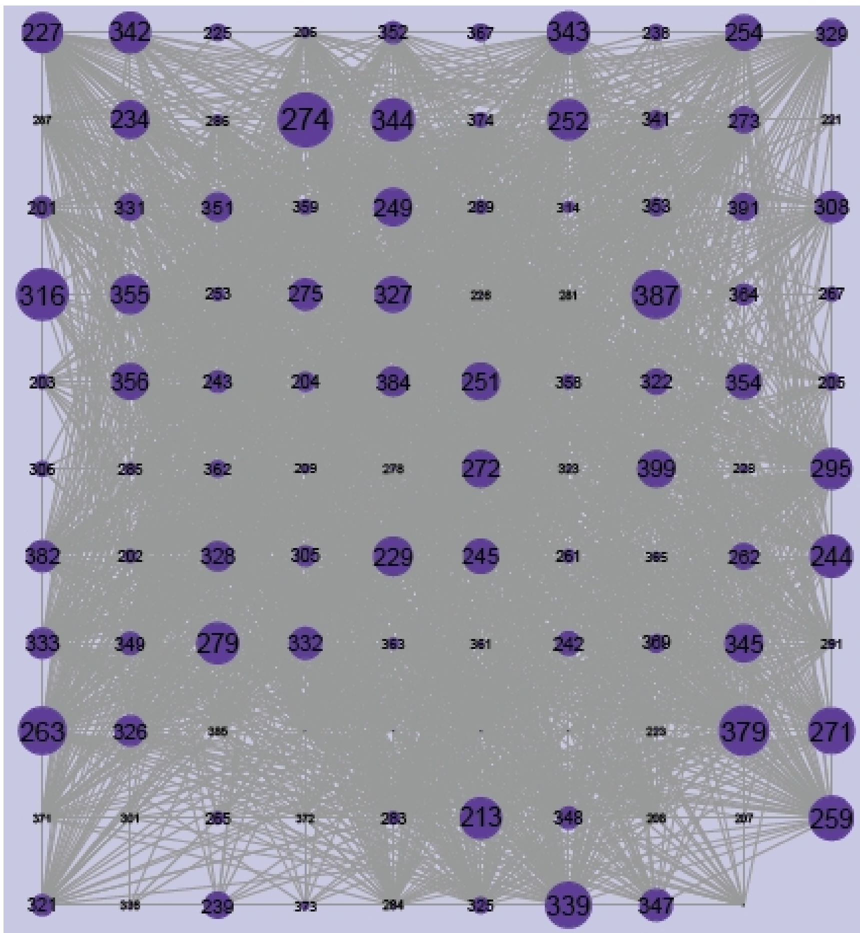

Figure 1: The network of headquarters

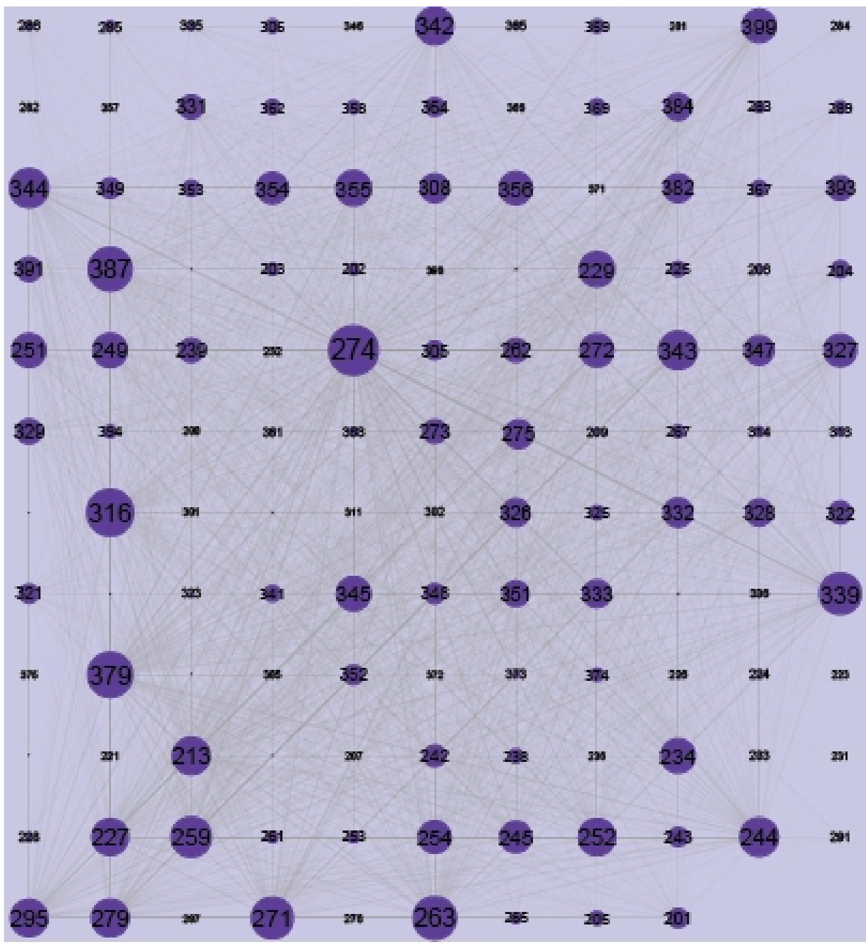

Figure 2: The network of subsidiaries

Table 1: Descriptive Statistics of Pairwise-Industry Multinational Activity (MNC)
Network Densities and Co-Agglomeration Economies

	# Obs.	Mean	Std. Dev.	Min.	Max.
		Network Densities--Pairwise-Industry Level			
Headquarters (Percentage Points)					
Threshold (T) = 200 kms	7875	0.1353	0.3266	0.0000	3.2486
T = 400 kms	7875	0.3153	0.7351	0.0000	6.8887
T= 800 kms	7875	0.7614	1.6815	0.0000	14.8063
T= 1600 kms	7875	1.3733	2.8954	0.0000	24.2803
Subsidiaries (Percentage Points)					
Threshold (T) = 200 kms	7875	0.0952	0.2296	0.0000	2.5381
T = 400 kms	7875	0.2125	0.5050	0.0000	5.4530
T= 800 kms	7875	0.5058	1.1744	0.0000	11.8555
T= 1600 kms	7875	1.0057	2.3082	0.0000	21.1262
		Co-Agglomeration Economies--Pairwise-Industry Level			
Input Output (IO) Linkages	7875	0.0033	0.0119	0.0000	0.1933
Capital	7875	0.4759	0.2086	-0.0041	1.0000
Labor	7875	0.3328	0.2268	0.0140	1.0000
Knowledge	7875	0.0074	0.0116	0.0001	0.1790

Notes: Same industry pairs (SIC3) excluded. The network density indices compare the estimated distance kernel function of each industry pair based on bilateral-establishment distance data to counterfactual kernel estimators based on a Monte Carlo approach at 200 kms, 400 kms, 800kms, and 1600 kms. Input Output (IO) Linkages, Capital, Labor, and Knowledge correspond to the industry-level variables employed to proxy the various FDI agglomeration economies: proximity to input suppliers or industrial customers; external scale economies in factor markets including labor and capital; and knowledge spillovers. See text for detailed descriptions of the variables.

Table 2: Correlation of Pairwise-Industry MNC Network Densities

	T = 200 kms (HQ)	T = 400 kms (HQ)	T = 800 kms (HQ)	T = 1600 kms (HQ)	T = 200 kms (Subs.)	T = 400 kms (Subs.)	T = 800 kms (Subs.)	T = 1600 kms (Subs.)
T = 200 kms (HQ)	1.0000							
T = 400 kms (HQ)	0.9926	1.0000						
T = 800 kms (HQ)	0.9554	0.9818	1.0000					
T = 1600 kms (HQ)	0.8580	0.8962	0.9553	1.0000				
T = 200 kms (Subs.)	0.4060	0.4210	0.4534	0.4970	1.0000			
T = 400 kms (Subs.)	0.4187	0.4380	0.4766	0.5258	0.9930	1.0000		
T = 800 kms (Subs.)	0.4247	0.4503	0.4998	0.5639	0.9618	0.9859	1.0000	
T = 1600 kms (Subs.)	0.3990	0.4291	0.4925	0.5896	0.8823	0.9186	0.9647	1.0000

Notes: Obs=7875. Correlation of network densities. See text for detailed descriptions of the variables.

Table 3: Co-Agglomeration Economies and MNC Headquarters Network Density

	T = 200 kms				T = 400 kms			
IO Linkages	0.5116***				1.0434***			
	(0.1683)				(0.3587)			
Capital		0.0747***				0.1698***		
		(0.0129)				(0.0284)		
Labor			0.0858***				0.1750***	
			(0.0143)				(0.0314)	
Knowledge				1.2611***				2.6553***
				(0.2063)				(0.4525)
# Obs.	7875	7875	7875	7875	7875	7875	7875	7875
R^2	0.637	0.638	0.638	0.638	0.648	0.649	0.649	0.649
	Beta Coefficients				Beta Coefficients			
	0.0187	0.0477	0.0596	0.0449	0.0169	0.0482	0.0540	0.0420
	T = 800 kms				T = 1600 kms			
IO Linkages	1.8403**				3.0292**			
	(0.7595)				(1.2624)			
Capital		0.3617***				0.5605***		
		(0.0630)				(0.1071)		
Labor			0.3087***				0.4325***	
			(0.0696)				(0.1190)	
Knowledge				5.3172***				8.6889***
				(0.9773)				(1.6958)
# Obs.	7875	7875	7875	7875	7875	7875	7875	7875
R^2	0.663	0.663	0.663	0.663	0.666	0.667	0.666	0.667
	Beta Coefficients				Beta Coefficients			
	0.0131	0.0449	0.0416	0.0368	0.0125	0.0404	0.0339	0.0349

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed-effects. See text for detailed descriptions of the variables.

Table 4: Co-Agglomeration Economies and MNC Headquarters Network Density

	T= 200 kms	T= 400 kms	T= 800 kms	T= 1600 kms
IO Linkages	0.0743	0.1201	0.0468	0.3167
	(0.1780)	(0.3819)	(0.8074)	(1.3376)
Capital	0.0253	0.0829**	0.2574***	0.4530***
	(0.0192)	(0.0419)	(0.0920)	(0.1546)
Labor	0.0496**	0.0795*	0.0534	-0.0243
	(0.0218)	(0.0476)	(0.1056)	(0.1816)
Knowledge	0.7534***	1.6369***	3.6666***	6.5762***
	(0.2198)	(0.4850)	(1.0721)	(1.9144)
# Obs.	7875	7875	7875	7875
R^2	0.639	0.650	0.664	0.667
	Beta Coefficients			
IO Linkages	0.0027	0.0020	0.0003	0.0013
Capital	0.0162	0.0235	0.0319	0.0326
Labor	0.0344	0.0245	0.0072	-0.0008
Knowledge	0.0268	0.0259	0.0253	0.0264

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed-effects. Robust normalized beta coefficients in lower panel. See text for detailed descriptions of the variables.

Table 5: Co-Agglomeration Economies and MNC Subsidiary Network Density

	T= 200 kms				T= 400 kms			
IO Linkages	0.5070***				1.0542***			
	(0.1385)				(0.2929)			
Capital		0.0551***				0.1196***		
		(0.0108)				(0.0228)		
Labor			0.0474***				0.0943***	
			(0.0117)				(0.0250)	
Knowledge				0.8651***				1.6987***
				(0.2607)				(0.5037)
# Obs.	7875	7875	7875	7875	7875	7875	7875	7875
R^2	0.569	0.570	0.569	0.569	0.598	0.598	0.598	0.598
	Beta Coefficients				Beta Coefficients			
	0.0264	0.0501	0.0469	0.0438	0.0249	0.0494	0.0424	0.0391
	T= 800 kms				T= 1600 kms			
IO Linkages	2.3052***				4.1910***			
	(0.6493)				(1.2593)			
Capital		0.2689***				0.5039***		
		(0.0505)				(0.0988)		
Labor			0.1874***				0.3081***	
			(0.0564)				(0.1117)	
Knowledge				3.5321***				5.9811***
				(0.9943)				(1.7687)
# Obs.	7875	7875	7875	7875	7875	7875	7875	7875
R^2	0.625	0.626	0.625	0.626	0.629	0.629	0.629	0.629
	Beta Coefficients				Beta Coefficients			
	0.0234	0.0478	0.0362	0.0350	0.0217	0.0456	0.0303	0.0301

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed-effects. See text for detailed descriptions of the variables.

Table 6: Co-Agglomeration Economies and MNC Subsidiary Network Density

	T= 200 kms	T= 400 kms	T= 800 kms	T= 1600 kms
IO Linkages	0.2499*	0.5413*	1.2525*	2.4133*
	(0.1427)	(0.3029)	(0.6762)	(1.3179)
Capital	0.0374***	0.0920***	0.2376***	0.4992***
	(0.0137)	(0.0290)	(0.0645)	(0.1273)
Labor	0.0046	-0.0018	-0.0453	-0.1528
	(0.0179)	(0.0374)	(0.0820)	(0.1603)
Knowledge	0.5738*	1.1014*	2.3300**	3.9433*
	(0.3119)	(0.5983)	(1.1604)	(2.0392)
# Obs.	7875	7875	7875	7875
R^2	0.570	0.599	0.626	0.630
	Beta Coefficients			
IO Linkages	0.0130	0.0128	0.0127	0.0125
Capital	0.0340	0.0380	0.0422	0.0451
Labor	0.0046	-0.0008	-0.0087	-0.0150
Knowledge	0.0291	0.0254	0.0231	0.0199

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed-effects. Robust normalized beta coefficients in lower panel. See text for detailed descriptions of the variables.

Table 7: Co-Agglomeration Economies and MNC Subsidiary Employment Network Density

	T = 200 kms				T = 400 kms			
IO Linkages	0.6778***				1.3183***			
	(0.1868)				(0.3429)			
Capital		0.1175***				0.2467***		
		(0.0171)				(0.0339)		
Labor			0.1349***				0.2657***	
			(0.0169)				(0.0330)	
Knowledge				2.7923***				5.0974***
				(0.4389)				(0.7577)
# Obs.	7875	7875	7875	7875	7875	7875	7875	7875
R^2	0.314	0.318	0.320	0.323	0.314	0.319	0.321	0.323
	Beta Coefficients				Beta Coefficients			
	0.0309	0.0935	0.1167	0.1238	0.0311	0.1018	0.1192	0.1172
	T = 800 kms				T = 1600 kms			
IO Linkages	2.6045***				3.6565***			
	(0.6546)				(1.1887)			
Capital		0.5265***				0.8416***		
		(0.0672)				(0.1159)		
Labor			0.5035***				0.7659***	
			(0.0656)				(0.1182)	
Knowledge				8.6713***				12.9516***
				(1.2215)				(1.9755)
# Obs.	7875	7875	7875	7875	7875	7875	7875	7875
R^2	0.351	0.357	0.357	0.357	0.393	0.398	0.397	0.397
	Beta Coefficients				Beta Coefficients			
	0.0312	0.1102	0.1146	0.1011	0.0243	0.0979	0.0968	0.0839

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed-effects. See text for detailed descriptions of the variables.

Table 8: Co-Agglomeration Economies and MNC Subsidiary Employment Network Density

	T= 200 kms	T= 400 kms	T= 800 kms	T= 1600 kms
IO Linkages	-0.1512	-0.2691	-0.299	-0.8005
	(0.1922)	(0.3591)	(0.6847)	(1.2314)
Capital	0.0400*	0.1064**	0.3081***	0.5434***
	(0.0225)	(0.0436)	(0.0851)	(0.1477)
Labor	0.0570**	0.1067**	0.1626*	0.2126
	(0.0254)	(0.0497)	(0.0981)	(0.1735)
Knowledge	2.2285***	3.8848***	6.0831***	9.0114***
	(0.5118)	(0.8953)	(1.4667)	(2.3301)
# Obs.	7875	7875	7875	7875
R^2	0.326	0.326	0.361	0.400
	Beta Coefficients			
IO Linkages	-0.0069	-0.0064	-0.0036	-0.0053
Capital	0.0318	0.0439	0.0645	0.0632
Labor	0.0493	0.0479	0.0370	0.0269
Knowledge	0.0988	0.0893	0.0709	0.0584

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed-effects. Robust normalized beta coefficients in lower panel. See text for detailed descriptions of the variables.

Table 9: Co-Agglomeration Economies and Expected Proximity in MNC Headquarters Network

	T = 200 kms				T = 400 kms			
IO Linkages	0.2856***				0.2978***			
	(0.0953)				(0.1003)			
Capital		0.0378***				0.0400***		
		(0.0072)				(0.0076)		
Labor			0.0479***				0.0498***	
			(0.0081)				(0.0085)	
Knowledge				0.6919***				0.7172***
				(0.1152)				(0.1216)
# Obs.	7875	7875	7875	7875	7875	7875	7875	7875
R^2	0.622	0.623	0.624	0.623	0.621	0.622	0.623	0.622
	Beta Coefficients				Beta Coefficients			
	0.0192	0.0445	0.0612	0.0453	0.0189	0.0444	0.0601	0.0444
	T = 800 kms				T = 1600 kms			
IO Linkages	0.3033***				0.3046***			
	(0.1045)				(0.1044)			
Capital		0.0420***				0.0414***		
		(0.0079)				(0.0079)		
Labor			0.0510***				0.0500***	
			(0.0089)				(0.0090)	
Knowledge				0.7199***				0.6872***
				(0.1276)				(0.1286)
# Obs.	7875	7875	7875	7875	7875	7875	7875	7875
R^2	0.619	0.619	0.619	0.606	0.606	0.606	0.607	0.606
	Beta Coefficients				Beta Coefficients			
	0.0185	0.0447	0.0590	0.0427	0.0188	0.0447	0.0587	0.0413

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed-effects. See text for detailed descriptions of the variables.

Table 10: Co-Agglomeration Economies and Expected Proximity in MNC Headquarters Network

	T= 200 kms	T= 400 kms	T= 800 kms	T= 1600 kms
IO Linkages	0.0475	0.0501	0.0505	0.0603
	(0.0994)	(0.1047)	(0.1087)	(0.1077)
Capital	0.0072	0.0086	0.0108	0.0110
	(0.0109)	(0.0115)	(0.0120)	(0.0120)
Labor	0.0323***	0.0329**	0.0330**	0.0324**
	(0.0124)	(0.0131)	(0.0138)	(0.0140)
Knowledge	0.4067***	0.4195***	0.4115***	0.3784***
	(0.1227)	(0.1296)	(0.1358)	(0.1369)
# Obs.	7875	7875	7875	7875
R^2	0.624	0.623	0.620	0.607
	Beta Coefficients			
IO Linkages	0.0032	0.0032	0.0031	0.0037
Capital	0.0085	0.0096	0.0115	0.0119
Labor	0.0412	0.0398	0.0381	0.0381
Knowledge	0.0266	0.0259	0.0244	0.0228

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed-effects. Robust normalized beta coefficients in lower panel. See text for detailed descriptions of the variables.

Table 11: Co-Agglomeration Economies and Expected Proximity in MNC Subsidiary Network

	T = 200 kms				T = 400 kms			
IO Linkages	0.3089***				0.3256***			
	(0.0827)				(0.0872)			
Capital		0.0314***				0.0332***		
		(0.0064)				(0.0067)		
Labor			0.0295***				0.0308***	
			(0.0069)				(0.0073)	
Knowledge				0.5719***				0.5944***
				(0.1681)				(0.1750)
# Obs.	7875	7875	7875	7875	7875	7875	7875	7875
R^2	0.534	0.534	0.534	0.535	0.538	0.539	0.539	0.539
	Beta Coefficients				Beta Coefficients			
	0.0284	0.0504	0.0516	0.0512	0.0282	0.0503	0.0507	0.0501

	T = 800 kms				T = 1600 kms			
IO Linkages	0.3437***				0.3458***			
	(0.0913)				(0.0906)			
Capital		0.0351***				0.0349***		
		(0.0070)				(0.0070)		
Labor			0.0319***				0.0310***	
			(0.0077)				(0.0079)	
Knowledge				0.6163***				0.6075***
				(0.1795)				(0.1757)
# Obs.	7875	7875	7875	7875	7875	7875	7875	7875
R^2	0.544	0.545	0.545	0.545	0.541	0.542	0.542	0.542
	Beta Coefficients				Beta Coefficients			
	0.0281	0.0502	0.0495	0.0491	0.0281	0.0496	0.0480	0.0481

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed-effects. See text for detailed descriptions of the variables.

Table 12: Co-Agglomeration Economies and Expected Proximity in MNC Subsidiary Network

	T= 200 kms	T= 400 kms	T= 800 kms	T= 1600 kms
IO Linkages	0.1459*	0.1558*	0.1679*	0.1739*
	(0.0848)	(0.0895)	(0.0938)	(0.0934)
Capital	0.0180**	0.0197**	0.0216**	0.0222**
	(0.0081)	(0.0086)	(0.0090)	(0.0090)
Labor	0.0053	0.0050	0.0043	0.0031
	(0.0107)	(0.0113)	(0.0119)	(0.0120)
Knowledge	0.4012**	0.4150**	0.4292**	0.4246**
	(0.2020)	(0.2103)	(0.2155)	(0.2108)
# Obs.	7875	7875	7875	7875
R^2	0.536	0.540	0.546	0.543
	Beta Coefficients			
IO Linkages	0.0134	0.0135	0.0137	0.0141
Capital	0.0290	0.0298	0.0309	0.0315
Labor	0.0093	0.0082	0.0067	0.0047
Knowledge	0.0359	0.0350	0.0342	0.0336

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed-effects. Robust normalized beta coefficients in lower panel. See text for detailed descriptions of the variables.

Table 13: Descriptive Statistics of Micro Network Densities and MNC Heterogeneity

	# Obs.	Mean	Std. Dev.	Min.	Max.
Network Densities--Subsidiary Level (Percentage Points)					
Subsidiaries (Count)					
Threshold (T) = 200 kms	27889	0.0000	6.5833	-46.2955	57.9774
T = 400 kms	27889	0.0000	8.6119	-65.0346	49.6032
T= 800 kms	27889	0.0000	11.1215	-73.0153	78.2981
MNC Heterogeneity--Subsidiary Level					
ln (Labor)	27889	4.0155	2.0788	0.0000	11.2253
ln (Sales)	27782	12.2908	7.1734	0.0000	21.4839
ln (Sales/Labor)	27782	8.2820	7.0229	-11.2253	21.4145

Notes: Establishment level data. The indices represent the deviation from the establishment's average counterpart in the host country/industry. See text for detailed descriptions of the variables.

Table 14: MNC Heterogeneity and Micro Network Density

	T= 200 kms	T= 200 kms	T= 200 kms	T= 400 kms	T= 400 kms	T= 400 kms	T= 800 kms	T= 800 kms	T= 800 kms
ln (Labor)	0.0810***			0.0545*			0.0746**		
	(0.0213)			(0.0285)			(0.0367)		
ln (Sales)		0.0191***			0.0175**			-0.0038	
		(0.0068)			(0.0084)			(0.0096)	
ln (Sales/Labor)			0.0139*			0.0147*			-0.0112
			(0.0074)			(0.0091)			(0.0104)
# Obs.	27860	27753	27753	27860	27753	27753	27860	27753	27753
R^2	0.631	0.631	0.631	0.596	0.596	0.596	0.623	0.624	0.624

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The indices represent the deviation from the establishment's average counterpart in the host country/industry. All regressions include industry and country fixed-effects. See text for detailed descriptions of the variables.

Appendix Table 1: Correlation of Co-Agglomeration Economies

	IO Linkages	IO Linkages (max.)	Capital	Labor	Knowledge	Knowledge (max.)
IO Linkages	1.0000					
IO Linkages (max.)	0.9732	1.0000				
Capital	0.1905	0.1886	1.0000			
Labor	0.2319	0.2249	0.5672	1.0000		
Knowledge	0.2905	0.2840	0.2301	0.3313	1.0000	
Knowledge (max.)	0.2640	0.2571	0.1879	0.2974	0.9763	1.0000

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

Appendix Table 2: Top Industry Pairs--Pairwise-Industry MNC Subsidiary Network Density

		200km	
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
		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
		800km	
274	Miscellaneous Publishing	379	Miscellaneous Transportation Equipment
274	Miscellaneous Publishing	263	Paperboard Mills
274	Miscellaneous Publishing	213	Chewing And Smoking Tobacco And Snuff
274	Miscellaneous Publishing	271	Publishing, Or Publishing And Printing
274	Miscellaneous Publishing	339	Miscellaneous Primary Metal Products
		1600km	
274	Miscellaneous Publishing	379	Miscellaneous Transportation Equipment
274	Miscellaneous Publishing	263	Paperboard Mills
274	Miscellaneous Publishing	213	Chewing And Smoking Tobacco And Snuff
274	Miscellaneous Publishing	271	Publishing, Or Publishing And Printing
263	Paperboard Mills	379	Miscellaneous Transportation Equipment

Notes: The indices exclude industry fixed-effects. See text for detailed description of the variables.

Appendix Table 3: Correlation of Expected Proximity in MNC Headquarters and Subsidiary Networks

	T = 200 kms (HQ E.P.)	T = 400 kms (HQ E.P.)	T = 800 kms (HQ E.P.)	T = 1600 kms (HQ E.P.)	T = 200 kms (Subs. E.P.)	T = 400 kms (Subs. E.P.)	T = 800 kms (Subs. E.P.)	T = 1600 kms (Subs. E.P.)
T = 200 kms (HQ E.P.)	1.0000							
T = 400 kms (HQ E.P.)	0.9996	1.0000						
T = 800 kms (HQ E.P.)	0.9972	0.9986	1.0000					
T = 1600 kms (HQ E.P.)	0.9900	0.9925	0.9970	1.0000				
T = 200 kms (Subs. E.P.)	0.3775	0.3783	0.3836	0.3897	1.0000			
T = 400 kms (Sub E.P.)	0.3800	0.3809	0.3865	0.3929	0.9997	1.0000		
T = 800 kms (Sub E.P.)	0.3835	0.3847	0.3909	0.3980	0.9981	0.9992	1.0000	
T = 1600 kms (Subs. E.P.)	0.3849	0.3865	0.3934	0.4017	0.9929	0.9949	0.9977	1.0000

Notes: Obs=7875. See text for detailed descriptions of the variables.

Appendix Table 4: Correlation of MNC Subsidiary and MNC Subsidiary Employment Network Densities

	T = 200 kms (Subs.)	T = 400 kms (Subs.)	T = 800 kms (Subs.)	T = 1600 kms (Subs.)	T = 200 kms (Empl.)	T = 400 kms (Empl.)	T = 800 kms (Empl.)	T = 1600 kms (Empl.)
T = 200 kms (Subs.)	1.0000							
T = 400 kms (Subs.)	0.9930	1.0000						
T = 800 kms (Subs.)	0.9618	0.9859	1.0000					
T = 1600 kms (Subs.)	0.8823	0.9186	0.9647	1.0000				
T = 200 kms (Empl.)	0.4200	0.3742	0.3272	0.2951	1.0000			
T = 400 kms (Empl.)	0.4983	0.4631	0.4265	0.3984	0.9849	1.0000		
T = 800 kms (Empl.)	0.6025	0.5906	0.5805	0.5697	0.8877	0.9517	1.0000	
T = 1600 kms (Empl.)	0.6164	0.6191	0.6334	0.6622	0.7692	0.8521	0.9554	1.0000

Notes: Obs=7875. See text for detailed descriptions of the variables.

Appendix Table 5: Top Industry Pairs--Pairwise-Industry MNC Employment Network Density

		200km	
394	Dolls, Toys, Games And Sporting And Athletic	314	Footwear, Except Rubber
394	Dolls, Toys, Games And Sporting And Athletic	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
		400 km	
394	Dolls, Toys, Games And Sporting And Athletic	314	Footwear, Except Rubber
394	Dolls, Toys, Games And Sporting And Athletic	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
		800km	
225	Knitting Mills	314	Footwear, Except Rubber
394	Dolls, Toys, Games And Sporting And Athletic	313	Boot And Shoe Cut Stock And Findings
225	Knitting Mills	313	Boot And Shoe Cut Stock And Findings
314	Footwear, Except Rubber	313	Boot And Shoe Cut Stock And Findings
394	Dolls, Toys, Games And Sporting And Athletic	314	Footwear, Except Rubber
		1600km	
225	Knitting Mills	314	Footwear, Except Rubber
271	Publishing, Or Publishing And Printing	379	Miscellaneous Transportation Equipment
225	Knitting Mills	313	Boot And Shoe Cut Stock And Findings
233	Women's, Misses', and Juniors' Outerwear ·	314	Footwear, Except Rubber
367	Electronic Components And Accessories	225	Knitting Mills

Notes: The indices exclude industry-fixed effects. See text for detailed descriptions of the variables.

Appendix Table 6: Co-Agglomeration Economies and MNC Subsidiary Network Density with Generalized Measure of Trade Cost

	T = 200 kms				T = 400 kms			
IO Linkages	1.3485***				1.4753***			
	(0.3709)				(0.4172)			
Capital		0.2083***				0.2315***		
		(0.0447)				(0.0489)		
Labor			0.2117***				0.2268***	
			(0.0468)				(0.0485)	
Knowledge				7.1818***				7.3196***
				(2.4190)				(2.3660)
# Obs.	7875	7875	7875	7875	7875	7875	7875	7875
R^2	0.328	0.329	0.330	0.336	0.334	0.336	0.336	0.342
	Beta Coefficients				Beta Coefficients			
	0.0216	0.0583	0.0644	0.1120	0.0228	0.0626	0.0666	0.1102
	T = 800 kms				T = 1600 kms			
IO Linkages	1.9062***				2.0303***			
	(0.6072)				(0.6449)			
Capital		0.2747***				0.2891***		
		(0.0632)				(0.0679)		
Labor			0.2799***				0.2915***	
			(0.0568)				(0.0604)	
Knowledge				8.5599***				8.7591***
				(2.3164)				(2.3124)
# Obs.	7875	7875	7875	7875	7875	7875	7875	7875
R^2	0.411	0.412	0.413	0.418	0.406	0.407	0.408	0.412
	Beta Coefficients				Beta Coefficients			
	0.0243	0.0612	0.0678	0.1063	0.0249	0.0618	0.0678	0.1044

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed-effects. See text for detailed descriptions of the variables.

Appendix Table 7: Co-Agglomeration Economies and MNC Subsidiary Network Density with Generalized Measure of Trade Cost

	T= 200 kms	T= 400 kms	T= 800 kms	T= 1600 kms
IO Linkages	-0.3843	-0.3277	-0.2232	-0.1611
	(0.4522)	(0.4829)	(0.6329)	(0.6621)
Capital	0.1012*	0.1237*	0.1328*	0.1434*
	(0.0610)	(0.0652)	(0.0814)	(0.0871)
Labor	-0.0178	-0.0184	0.0015	0.0021
	(0.1145)	(0.1139)	(0.1158)	(0.1190)
Knowledge	6.9390**	6.9555**	7.9741***	8.0982***
	(3.0433)	(2.9752)	(2.8834)	(2.8740)
# Obs.	7875	7875	7875	7875
R^2	0.336	0.342	0.418	0.413
	Beta Coefficients			
IO Linkages	-0.0062	-0.0051	-0.0028	-0.0020
Capital	0.0283	0.0334	0.0296	0.0307
Labor	-0.0054	-0.0054	0.0004	0.0005
Knowledge	0.1082	0.1047	0.0991	0.0965

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All regressions include industry fixed-effects. Robust normalized beta coefficients in lower panel. See text for detailed descriptions of the variables.