Heavier than Air? Knowledge Transmission within the Multinational Firm*

Dany Bahar†

Harvard University

April 7, 2014

[Download Newest Version]

JOB MARKET PAPER

Keywords: multinational, firms, corporations, knowledge, FDI, horizontal expansion

JEL Classification Numbers: F21, F23, D83, L22, L23

*I thank my extraordinary advisors Laura Alfaro, Pol Antras, Ricardo Hausmann and Elhanan Helpman for their encouragement, counsel, valuable time and thoughtful comments throughout this process. This paper also benefited from helpful comments at various stages from Martin Abel, Sam Asher, Sebastian Bustos, Michele Coscia, Marc Melitz, Paul Novosad, Nathan Nunn, Lant Pritchett, Ran Shorrer, Muhammed Yildirim, Richard Zeckhauser, Oren Ziv and the participants of numerous seminars at Harvard Economics Department, Harvard Business School and Harvard Kennedy School, as well as the NBER Productivity Lunch. Jean-Paul Leidenz Font provided excellent research assistance. I acknowledge financial support from Harvard Center for International Development and Harvard Economics Department. All errors are my own.

†dbahar@fas.harvard.edu; http://scholar.harvard.edu/dbaharc

Abstract

To what extent do barriers to knowledge transmission influence a firm's decision to expand? Using a worldwide dataset on foreign subsidiaries, I show that multinational corporations are, on average, about 12% less likely to horizontally expand a sector that is one standard deviation above the mean in the knowledge intensity scale. Evidence shows that when firms do expand their knowledge-intensive activities they tend to do so at shorter geographic distances. Locating a foreign subsidiary in the same time zone as its headquarters tends to reduce barriers to knowledge transmission by easing communication and effectively reducing the distance between them by, on average, 3500 Km. The empirical results can be explained through an expansion of the theoretical framework developed by Helpman, Melitz and Yeaple (2004). The new model incorporates the cost of knowledge transmission for firms engaged in foreign direct investment, which affects the mechanisms of the proximity-concentration hypothesis.

1 Introduction

About fifty percent of cross-country income variation is explained by differences in productivity.[1] This begs the question: if productivity-inducing knowledge[2] is available in some places, why isn't it available in others? Arrow (1969) suggests that the transmission of knowledge is difficult and costly. These difficulties arise because effective knowledge transmission involves human interaction, which cannot be fully replaced with written words[3] (e.g., even in today's world, business trips have not been fully replaced by emails). A firm, as any other economic agent, also faces difficulties when transferring knowledge among different divisions and affiliates. When a firm operates across borders, different time zones, languages and cultures can raise knowledge transmission costs further. This study contributes to the literature by addressing the effects of knowledge transmission costs on the expansion of multinational corporations (MNCs).

This paper presents two main empirical findings and formalizes them in a theoretical framework. First, MNCs are less likely to horizontally expand their knowledge intensive activities to foreign locations, compared to non knowledge intensive industries. Second, when they do expand, they tend to do so at a shorter geographic distance. Interestingly, however, geographic distance becomes less relevant for horizontal expansion when a firm and its subsidiary are located within the same time zone, and thus able to communicate in real time.

These findings cannot be explained by most theoretical models on MNC fragmentation, which implicitly or explicitly assume zero cost, or costs orthogonal to distance, of transferring knowledge between headquarters and subsidiaries (i.e., Helpman 1984; Markusen 1984; Brainard 1993; Markusen

[1]e.g., Caselli 2005, Hall and Jones 1999

[2]The Merriam-Webster dictionary defines knowledge as the set of information, understanding, and/or skills that one gets from experience or education.

[3]Knowledge that resides in human minds is usually referred to as tacit (Polanyi 1966). Tacit knowledge is information that cannot be easily explained, embedded or written down.

et. al. 1996; Markusen 1997; Carr et. al. 2001; Helpman, Melitz and Yeaple 2004; Keller and Yeaple 2013). A number of empirical studies have tested the validity of these models' predictions, but there has been little or no emphasis on testing the assumption that knowledge transmission is costless.[4]

Thus, to explain the results, I augment the model by Helpman, Melitz and Yeaple (2004) by incorporating the marginal cost of knowledge transfer faced by firms engaged in foreign direct investment (FDI). The augmented model's main assumption is the existence of a marginal cost of knowledge transmission that increases with the level of knowledge intensity and the distance between headquarters and subsidiary. This departs from the traditional view, which assumes that the fixed costs of initial setup are the only costs incurred by the headquarters when creating a foreign affiliate. In reality, the costs of maintaining and interacting with a foreign subsidiary are present throughout the subsidiary's lifetime, and do not end the day the plant is built.

For the empirical analysis, I use a sample derived from the Worldbase dataset by Dun & Bradstreet,[5] comprising more than 60,000 foreign subsidiaries of MNCs with information on their physical location and primary economic activity, as defined by the 1987 Standard Industry Classification (SIC). From this dataset, I identify those foreign subsidiaries that represent a horizontal expansion of the associated MNC. Using geocoded location data, I measure the precise distance between each foreign affiliate and its MNC global headquarters. I then compute industry-specific knowledge intensity measures. These indicators reflect the accumulated experience and training of workers in any given industry, using occupational characteristics defined in the O*NET project dataset. I link these indicators to the industry reported by each foreign subsidiary in the dataset. Finally, I exploit variation in the

[4]e.g., Brainard 1997; Carr et. al. 2001; Markusen and Maskus 2002.

[5]The dataset was privately acquired from D&B and is not publicly accessible. It has been previously used in the literature by Lipsey (1978), and more recently by Harrison et. al. (2004), Black and Strahan (2002), Acemoglu, Johnson & Mitton (2009), Alfaro and Charlton (2009), Alfaro and Chen (2012).

knowledge intensity of subsidiaries to study the model's predictions about the nature of knowledge transmission for horizontal foreign subsidiaries as opposed to the MNC's domestic affiliates as well as non-horizontal foreign subsidiaries.

The data reveals that firms are less likely to have horizontal foreign subsidiaries producing knowledge intensive goods. This result controls for transportation costs, characteristics of the host country relative to the headquarters country, and MNC fixed effects. More specifically, manufacturing industries that are one standard deviation above the knowledge intensity mean are, on average, about 3.6 percentage points less likely to be replicated abroad as a horizontal foreign subsidiary, or 12% based on the actual proportion of foreign affiliates in the sample. For example, a semiconductor manufacturing plant is about 30 percentage points less likely to be replicated abroad than a meat packing plant.

Using the whole portfolio of foreign subsidiaries of a firm, which include both horizontal and non-horizontal subsidiaries, I find that horizontal subsidiaries are characterized by being located at shorter geographic distances to the headquarters, a result that supports the model's assumption that the cost of transferring knowledge to horizontal foreign subsidiaries increases with distance. In light of this result, I explore the relationship between distance and knowledge intensity for horizontal subsidiaries. The assumptions of the model imply that firms, in order to maximize profits, face a tradeoff that drives them to locate foreign horizontal subsidiaries nearby—especially when they produce a knowledge intensive good. This is supported by the data, which shows a negative partial correlation between knowledge intensity and the distance between a headquarters and its horizontal foreign subsidiaries. For instance, an American MNC with a meat packing horizontal subsidiary located in Turkey would locate its horizontal semiconductors plant in Ireland.

Much of the literature would posit that these results are driven by transportation costs of intermediate goods, which are assumed to be more prevalent for knowledge intensive industries (Irrazabal et. al. 2013; Keller and Yeaple 2013).[6] I find evidence in the data to rule out this mechanism. More specifically, I find that when a headquarters and its subsidiaries are located in the same time zone, distance losses relevance in a firm's decision to expand horizontally. This implies that real-time communication decreases the cost of transferring knowledge by effectively "reducing" the distance between headquarters and subsidiary by two thirds, or by 3500 Km. for the average foreign subsidiary. This suggests that the cost of shipping intermediate goods (which would be just as relevant within the same time zone, because north-south shipping is equally as expensive as east-west shipping), is not the only factor driving a firm's location decisions. Rather, the evidence suggests that the cost of transferring knowledge plays an important role by incentivizing firms to locate their knowledge intensive subsidiaries at shorter distances. Speaking a common language also seems to effectively reduce distance between a headquarters and a subsidiary (though not as much), while the existence of a non-stop commercial flight between a headquarters and its subsidiary does not. This implies that real-time remote interaction and cultural similarities are more important in lowering the cost of knowledge transfer than the ease of face-to-face interaction.

The results show that the cost of knowledge transmission is a determinant of MNC activity. These findings have larger implications for a number of yet-unanswered questions in economics. For instance, high barriers to knowledge transmission may explain persistent differences in productivity levels between countries and the divergence of their incomes over time (e.g., Pritchett 1997, Hall and Jones 1999), because productivity-inducing knowledge does not diffuse easily.

The rest of the paper is divided as follows. Section 2 summarizes the re-

[6]Keller and Yeaple (2013) assume that knowledge is substitutable with intermediate goods, inducing intra-firm trade in knowledge intensive sectors and thus worsening the performance of distant foreign affiliates.

lated literature. Section 3 outlines a theoretical framework that explores how knowledge transmission can affect MNC decisions and provides guidance for the empirical analysis. Section 4 describes the dataset and the construction of relevant variables. Section 5 discusses the empirical strategy and presents results and their interpretation, while Section 6 concludes and addresses areas for future research regarding the role of knowledge in economic activity.

2 Related Literature

The determinants of MNC expansion and fragmentation have been explored in the literature for years.[7] Helpman (1984) suggests vertical fragmentation is motivated by differences in factor abundance between the host and recipient country. Markusen (1984) models the case when horizontal expansion can arise between two identical countries, based on the assumption that a headquarters' activities can be geographically distant from production processes. Brainard (1993) modeled the "proximity-concentration hypothesis," in which both transportation costs and increasing returns play a role in international horizontal expansion of MNCs.

Carr, Markusen and Maskus (2001) –building on Markusen et. al. (1996) and Markusen (1997)– endogenize the vertical and horizontal decisions of firm in what is known as the "knowledge capital model." The model is based on three critical assumptions. First, knowledge-based assets may be fragmented from production; second, knowledge-based assets are skilled labor intensive; and third, the services of knowledge-based assets are (at least partially) joint inputs (i.e., homologous to a public good within the firm) into multiple production facilities. In this model, vertical fragmentation arises from the first two assumptions, while horizontal expansion is a result of the third one.

Papers such as Brainard (1997), Carr et. al. (2001), Markusen and

[7]See Antras and Yeaple (2013) for recent review on this topic.

Maskus (2002) empirically test for the predictions of the above mentioned models, with little emphasis on testing the validity of the zero-cost assumption concerning knowledge transmission.

In the literature on heterogenous firms, Helpman, Melitz and Yeaple (2004) present a model in which horizontal FDI substitutes for exports. In it, a firm's potential profit determines such tradeoff, based on fixed costs and transportation costs. More recently, Keller and Yeaple (2013) augment this model by adding the knowledge component. In their model, once the firm expands horizontally beyond its borders, it faces the tradeoff between creating an upstream plant in that remote location (which locally provides the knowledge) or, alternatively, shipping the knowledge-embedded intermediate good from the headquarters' site (being the main assumption that knowledge can be fully embedded into an intermediate good). The model predicts that a firm will decide to do the latter for knowledge-intensive products. Under this framework, the lower profitability that characterizes distant subsidiaries active in more knowledge intensive industries is explained by intra-firm trade. That is, firms face higher trade costs for knowledge intensive industries given their optimal choice of importing the "ready-to-go" knowledge embedded in intermediate goods from its headquarters. They present empirical evidence supporting this hypothesis.

I introduce tacit knowledge as an additional component to this discussion. Michael Polanyi (1966) referred to tacit knowledge as information that is difficult to transfer: it cannot be easily explained, embedded or written down.[8] Firms possess tacit knowledge both in their specific processes and in the minds of their employees (e.g., Kogut and Zander 1992). It is in the interest of the firm to transfer this knowledge, as efficiently as possible, to all of its subsidiaries. However, the tacit character of this knowledge implies it cannot be embedded in intermediate goods, and that there are difficulties

[8]Others in the management and strategy literature have referred to this type of knowledge as "sticky information" (e.g., Von Hippel 1994; Szulanski 1996, 2002).

associated with its transmission. If these difficulties are large enough, we would expect them to have an impact on the pattern of MNCs' decisions regarding foreign expansion.

Is it reasonable to think that such difficulties exist? In fact, the consensus in the existing literature on the economics of knowledge is that the transmission of knowledge is not immediate, and that knowledge diffusion strongly decays with distance. For instance, the paper by Jaffe, Trajtenberg and Henderson (1993) was among the first to make this claim, showing that patent citations are more frequent within the same geographic area. Bottazi and Peri (2003) followed up using European data. Along the same lines, Keller (2002) showed that knowledge spillovers decrease with distance by looking at productivity changes as explained by foreign R&D investment. He documents that the half-life of such spillovers is 1200Km. More recently, Bahar et. al. (2014) show that a country is 65% more likely to add a new product to its export basket whenever a geographic neighbor is a successful exporter of the same good, a finding that is attributed to the local character of knowledge diffusion.[9]

In this context, this paper aims to contribute to the literature by presenting unexplored evidence on the role tacit knowledge transmission plays in the activity of MNCs, focusing on horizontal expansion.

3 Conceptual Framework

In this section I augment the model by Helpman, Melitz and Yeaple (2004) – referred to as HMY hereafter – by including a new parameter capturing the intra-firm cost of transmitting knowledge between headquarters and foreign subsidiaries. This extension allows us to understand how the cost of knowledge transmission faced by firms affects their decision to serve foreign markets. First the common set-up is described and then the proper adapta-

[9] See Keller (2004) for a review of this literature.

tion is incorporated.

As in HMY, there are N countries producing H+1 sectors with labor as the only input of production. H sectors (indexed 1, 2,...,H) produce a differentiated good, while the other sector (indexed 0) produces an homogenous good (which serves as the numeraire). In any given country, individuals spend a share $\beta_h > 0$ of their income on sector h, such that $\sum_{0 \leq h \leq H} \beta_h = 1$. Country i is endowed with L^i units of labor and the wage rate in this country is denoted by w^i.

Consider now a particular differentiated sector, h. For simplicity of notation, the index h is dropped in the next few paragraphs, but it is implicit that all sector specific variables may vary across sectors.[10] In order to enter the industry in country i a firm bears a fixed and sunk cost f_E denominated in units of labor. After bearing this cost, the potential entrant learns its labor-per-unit cost, a, drawn from a common and known distribution $G(a)$. Upon observing this cost, the firm may choose not to enter, and thus bear no additional costs and receive no revenues. If it chooses to produce, however, an additional cost of f_D units of labor is incurred. There are no other fixed costs if the firm chooses to produce and sell in the local market only.

The firm can choose to serve a foreign market either by exporting or creating a foreign subsidiary. If the firm chooses to export, it bears an additional cost of f_X (per country it exports to). If it chooses to create a foreign affiliate, it incurs an additional cost of f_I for every foreign market it chooses to serve this way. Similar to HMY, f_X can be interpreted as the cost of forming a local distribution and service network in the foreign market, and f_I includes all of these costs, as well as the cost of forming a subsidiary in the foreign country and the overhead production costs embodied in f_D. Hence, $f_I > f_X > f_D$.

The homogenous good is freely traded at no cost.[11] Differentiated goods

[10]Some sector-specific variables are explicitly kept in the notation, such as t and k, since these variables will be relevant in the empirical analysis.

[11]Thus, as long as the numeraire good is produced in all countries the wage rate is

that are exported from country i to country j are subject to a "melting-iceberg" transport cost $\tau(t, d_{ij})$ which is an increasing function of the per unit shipping cost of the good (denoted by t, and proxies for weight or other good specific characteristics) and the distance between countries i and j (denoted by d_{ij}). It is assumed that that $\tau(t, d_{ij}) > 1$. That is, a firm in country i has to ship τ units of a good for 1 unit to arrive in country j.

Analogously, serving a foreign market through an affiliate is subject to a marginal cost $\kappa(k, d_{ij})$ related to the transfer of knowledge. $\kappa(k, d_{ij})$ is assumed to be an increasing function of both the knowledge intensity of the good (represented by k) and distance (d_{ij}). The cost of transferring knowledge includes resources and time used for communicating with foreign affiliates to transmit proper knowledge required for efficient production. It is assumed that $\tau(t, d_{ij}) > \kappa(k, d_{ij}) > 1$ for all goods. The last inequality implies that for a multinational corporation, the cost of selling 1 unit of a good through a foreign affiliate is $\kappa(k, d_{ij})$.

The cost of knowledge transmission in knowledge intensive sectors being higher is justified given that these sectors require higher interaction and communication among their workers. Thus, firms pay for business travel and communication services that occur more often within these sectors. In addition, and perhaps more importantly, knowledge intensive activities usually encompass tasks with higher probability of failure and thus requiring trained and experienced workers. This too raises operational costs.

Assuming that knowledge transmission costs are increasing in distance is consistent with empirical evidence (e.g., Jaffe, Trajtenberg and Henderson 1993, Bottazi and Peri 2001, Keller 2002, Keller 2004, Bahar et. al. 2014). This evidence is reviewed in the previous section.

All the producers which serve a market engage in monopolistic competition. Consumer preferences across varieties of a differentiated product h have the standard CES form, with an elasticity of substitution $\varepsilon = \frac{1}{1-\alpha} > 1$.

equalized.

It is well known that these preferences generate a demand function $A^i p^{-\varepsilon}$ for each product in the industry in country i, where $A^i = \frac{\beta}{\int_0^{n^i} p^i(s)^{1-\varepsilon} ds} E^i$, n^i is the measure of firms active in the industry in country i, and $p^i(s)$ is the consumer price for a product indexed s.

In this setting, an active producer with labor requirement of a optimally sets a price of $\frac{w^i a}{\alpha}$. Consequently, the price of a locally produced good is $\frac{w^i a}{\alpha}$, the price of a good which is exported to country j is $\frac{\tau(t,d_{ij}) w^j a}{\alpha}$, and the price of a good that is sold by a foreign affiliate in country j is $\frac{\kappa(k,d_{ij}) w^j a}{\alpha}$, where a is the labor required for the producer to manufacture one unit of the product.

In what follows, it is shown that the balance of forces ruling the tradeoff of serving a foreign market through exports or FDI is influenced by the knowledge intensity of the product.

The assumption that the numeraire good is produced in each country simplifies the analysis, as it implies that the wage rate is equalized across all countries and is equal to 1. Hence, the operating profit for a firm in country i with a labor coefficient of a from serving the domestic market maybe expressed as $\pi_D^i = a^{1-\varepsilon} B^i - f_D$, where $B^i = (1-\alpha) \frac{A^i}{\alpha^{1-\varepsilon}}$. The additional profits from exporting to country j are $\pi_X^i = (\tau(t, d_{ij}) \cdot a)^{1-\varepsilon} B^j - f_X$ and those from selling in country j through a foreign affiliate are $\pi_I^i = (\kappa(k, d_{ij}) \cdot a)^{1-\varepsilon} B^j - f_I$. B^i represents demand parameters for country i and are considered exogenous to each individual firm.

Hence, in this setting, the productivity parameter a will be critical for a firm's decision of whether to serve the local market only or to serve foreign markets, either through exports or FDI. The sorting pattern is similar to the one in HMY and is based in the following equations:

$$(a_D)^{1-\varepsilon} \cdot B^i = f_D, \qquad \forall i \tag{1}$$

$$(\tau(t, d_{ij}) \cdot a_X)^{1-\varepsilon} \cdot B^j = f_X, \quad \forall i, \forall j \neq i \tag{2}$$

$$\left[\kappa(k, d_{ij})^{1-\varepsilon} - \tau(t, d_{ij})^{1-\varepsilon}\right] \cdot a_I^{1-\varepsilon} \cdot B^j = f_I - f_X, \quad \forall i, \forall j \neq i \tag{3}$$

Similar to HMY, the first two equations define the productivity thresholds after which firms will sell domestically or export, respectively. The minimum productivity threshold after which firms will engage in FDI is derived from Equation (3).[12] This threshold is defined as:

$$a_I^{1-\varepsilon} = \frac{f_I - f_X}{\left[(\kappa(k, d_{ij}))^{1-\varepsilon} - (\tau(t, d_{ij}))^{1-\varepsilon}\right] B^j}, \forall i, \, \forall j \neq i \tag{4}$$

Predictions derived from this model will serve as the basis for the empirical analysis. The implications of the original HMY model are straightforward. An increase in $\tau(t, d_{ij})$, either through an increase in either t or d_{ij}, will result in lower π_E making it more likely to substitute exports with FDI. This is part of the mechanism of the concentration-proximity tradeoff. However, with the inclusion of $\kappa(k, d)$ into the model, some new predictions arise, assuming full symmetry in fixed costs and demand variables for all sectors and countries. The propositions are presented in terms of $\phi(a_I) = a_I^{1-\varepsilon}$.

Proposition 1. *As k increases, the profitable FDI threshold ($a_I^{1-\varepsilon}$) increases, implying fewer firms will substitute exports towards FDI.*

$$\frac{\partial \phi(a_I)}{\partial k} = \frac{\partial \phi(a_I)}{\partial \kappa} \cdot \frac{\partial \kappa}{\partial k} > 0 \tag{5}$$

Proposition (1) is a direct consequence of adding κ into the model. Thus, *ceteris paribus*, FDI will be less likely for sectors with higher k. The graphical representation of the model in Figure 1 shows the case for two sectors that differ in their knowledge intensity, $\underline{k}$ and $\overline{k}$ (where $\overline{k} > \underline{k}$). Notice that the profit functions for both sectors originate in the same fixed cost value f_I, but the function is flatter for the sector $\overline{k}$ (dashed line). Hence, the productivity threshold required for a firm to substitute exports with FDI becomes higher for sectors with higher levels of k. That is, $(a_I^{ij,\overline{k}})^{1-\varepsilon} > (a_I^{ij,\underline{k}})^{1-\varepsilon}$.

[12]Condition (3) will have a positive solution if we assume $\kappa(k, d_{ij})^{\varepsilon-1} f_I > \tau(t, d_{ij})^{\varepsilon-1} f_X > f_D$, which is homologous to condition (1) in HMY (with equal wages across countries), but including κ.

[Figure 1 about here.]

Proposition 2. *As d increases, the change in $a_I^{1-\varepsilon}$ is ambiguous.*

$$\frac{\partial\phi(a_I)}{\partial d} = \frac{\partial\phi(a_I)}{\partial\tau}\cdot\frac{\partial\tau}{\partial d} + \frac{\partial\phi(a_I)}{\partial\kappa}\cdot\frac{\partial\kappa}{\partial d} = \begin{cases} \geq 0, & \left[\frac{\tau(t,d)}{\kappa(k,d)}\right]^{\varepsilon-1} \geq \frac{\epsilon_{\tau,d}}{\epsilon_{\kappa,d}} \\ < 0, & otherwise \end{cases} \tag{6}$$

To understand Proposition (2) suppose there are two foreign countries h and j such that $d_{ij} > d_{ih}$. In the original HMY model, longer distances will reduce π_E hence making it *always* more profitable for a given level of a to engage in FDI instead of exports. However, with the inclusion of κ in the model, longer distances will reduce both π_E and π_I. Thus, the equilibrium point can shift either way, depending on the elasticity of profits with respect to distance. The left panel of Figure 2 shows the case when $(a_I^{ih})^{1-\varepsilon} > (a_I^{ij})^{1-\varepsilon}$. Intuitively this happens whenever π_E is more elastic to changes in distance than π_I (or given the condition stated in Equation (6), where ϵ represents elasticity; see Appendix Section A.1 for more details on this condition). This case is qualitatively the same result as in the HMY model. The right panel of Figure 2, however, shows another possibility. In it $(a_I^{ih})^{1-\varepsilon} < (a_I^{ij})^{1-\varepsilon}$. In this case, FDI will be less profitable for longer distances hence resulting in *fewer* firms substituting exports with FDI.

[Figure 2 about here.]

The predictions coming out of the model following the inclusion of an intra-firm cost of transferring knowledge (κ) have testable implications in the data. First, *ceteris paribus*, industries with higher levels of knowledge intensity will be *less likely* to expand horizontally to foreign destinations. Second, horizontal expansion will be *less likely* in foreign locations that are located at longer distances under certain conditions. Regarding Proposition (2), given there are no empirical priors on whether the stated condition holds, letting the data speak will provide guidance on the assumptions of the developed model. That is, if horizontal FDI correlates negatively with longer distances, then there is empirical support to assume that $\partial\kappa/\partial d > 0$.

The next section presents the sample and the variables used to perform the empirical analysis.

4 Data and Definitions of Variables

4.1 Worldbase dataset by Dun & Bradstreet

This paper uses the Worldbase dataset by Dun & Bradstreet (from May 2012) as its main data source. The dataset has information on more than one hundred million establishments worldwide. Each establishment is uniquely identified and linked to its global headquarters (referred to as the "global ultimate"). For this study I focus on foreign plants engaged in manufacturing industries (SIC codes 2000 to 3999) owned by MNCs. As suggested by Caves (1971), an MNC is "an enterprise that controls and manages production establishments – plants – located in at least two countries."[13]

Two different samples are obtained from the dataset. The first one, uses both domestic and horizontal foreign subsidiaries of MNCs. The second one, exclusively uses the complete portfolio of foreign subsidiaries of MNCs, which include horizontal and non-horizontal subsidiaries.

The first sample includes about 64,462 subsidiaries, both domestic and foreign (the latter defined as being in a different country than their global ultimate). The second sample consists of 60,621 foreign subsidiaries. Overall, headquarters are scattered across 89 countries while subsidiaries are in over 100 countries.

For the analysis I will use the reported main SIC code as the only indicator

[13]I exclude MNCs for which 99% of their subsidiaries or employees are in the home country, besides them having plants in two or more countries. This drops a small number of Chinese MNCs with one or two subsidiaries in Hong Kong and the rest in China.

of a plant's economic activity. There are about 450 unique SIC 4-digit codes (in manufacturing) reported by subsidiaries as their main economic activity in the dataset (see Appendix Section A.2 for more details on this).

In order to obtain the precise location of each plant I geocode the dataset using Google Maps Geocoding API to find the exact latitude and longitude of its headquarters and each one of its foreign subsidiaries. With this I computed the exact distance between each headquarters and its foreign subsidiaries. Figure 3 maps the unique locations of all foreign subsidiaries (dots) and headquarters (triangles) in the sample.

[Figure 3 about here.]

For instance, Figure 4 shows the headquarters and subsidiaries of an American car manufacturing multinational firm. The firm, headquartered in the US, has a number of foreign subsidiaries on different continents. The lines originating from the headquarters represent the geographic distance to each subsidiary.

[Figure 4 about here.]

4.2 Definitions of Variables

4.2.1 Horizontal Foreign Subsidiary

I define a foreign subsidiary as a horizontal expansion based on its SIC code vis-à-vis all the SIC codes reported by the firm, in all of its domestic subsidiaries in the home country. This resolves the data issues that arise when the economic activity of the headquarters does not necessarily represent the main business of the firm. For instance, in the dataset, the headquarters of a well known worldwide multinational in the cosmetic world is defined under SIC code 6719 ("holding company"). However, many of its domestic subsidiaries are classified under SIC code 2844 (perfumes, cosmetics, and other toiletries), which would be a more natural classification for the firm as a whole. Hence, by limiting the definitions to the global ultimate's SIC category only, horizontal relationships would be underestimated.

Following the methodology used by Alfaro and Charlton (2009) I exclude from the definition of horizontal expansions those foreign subsidiaries that fall in both horizontal and vertical classification (see Appendix Section A.3 for more details).

When limiting the sample to domestic and foreign horizontal affiliates only, the latter are about 29% of all plants. When looking at the broader foreign affiliates portfolio of a MNC, (which includes all types of foreign subsidiaries, and excludes domestic subsidiaries), around 34% of all foreign subsidiaries are classified as horizontal expansions. Of the remaining 64% of non-horizontal affiliates, only a handful can be classified as vertical foreign subsidiaries,[14] while the majority are subsidiaries classified in industries that are unrelated to the firms' core business, as measured by the sectors the firm is producing at home.[15]

4.2.2 Knowledge Intensity Measures

In order to estimate the knowledge intensity of industries I construct indicators that measure the accumulated experience and training required for optimal performance of the different occupations associated with each industry. These measures attempt to proxy for the knowledge parameter k referred to in the theoretical framework above.

Knowledge is defined as the set of information, skills and understanding that one acquires through experience and education. The tacit component of knowledge is the one that resides mostly in people's brains, and cannot

[14] 3,062 observations can be classified as vertical foreign subsidiaries, while 8,108 are classified as "complex", implying they fall in both horizontal and vertical categories. These "complex" subsidiaries are considered neither horizontal nor vertical. Appendix Section A.3 expands on this discussion.

[15] This is an interesting finding in and of itself, and is also noted by Alfaro and Charlton (2009). While attempting to explain this finding is out of the scope of this paper, it is a part of the future research agenda.

be codified. Thus, in order to quantify the intensity of the tacit knowledge that characterizes an specific industry I compute the average experience and training of that industry's representative workforce. This differs from other measures that would capture only the codified component of knowledge such as patent counts or years of schooling of workers. To the best of my knowledge, these are the first measures that attempt to capture the tacit knowledge intensity of an industry.

To construct the knowledge intensity measures I use data from the Occupational Employment Statistics (OES) from the Bureau of Labor Statistics,[16] and occupational profiles compiled by the Occupational Information Network (O*NET) project.[17] OES breaks down the composition of occupations for each industry code,[18] based on a list of about 800 occupations. These occupations can be linked to occupational profiles generated by O*NET, which includes results from a large number of survey questions on the characteristics of each occupation.

The relevant questions in the survey that capture the learning component of the workers, as mentioned above, are the ones related to experience and training. The exact form of the questions from O*NET are:

- How much related experience (in months) would be required to be hired to perform this job?
- How much "on-site" or "in-plant" training (in months) would be required to be hired to perform this job?

[16] Data from 2011, downloadable from ftp://ftp.bls.gov/pub/special.requests/oes/oesm11in4.zip

[17] O*NET is the successor of the US Department of Labor's Dictionary of Occupational Titles (DOT). I use the O*NET database version 17, downloadable from http://www.onetcenter.org/download/database?d=db_17_0.zip. Costinot et. al. (2011) also use O*NET to create an industry level measure of task routineness for 77 sectors. Keller and Yeaple (2013) also present results making use of knowledge intensity variables constructed with O*NET in the web appendix.

[18] I used Pierce & Schott (2012) concordance tables to convert industry codes from NAICS to 1987 SIC. The concordance table is downloadable from http://faculty.som.yale.edu/peterschott/files/research/data/appendix_files_20111004.zip.

- How much "on-the-job" training (in months) would be required to be hired to perform this job?

Using these questions I generate the main knowledge intensity measure that I will be using in the empirical analysis section.[19] The measure, which I refer to it as *"Experience plus training"* throughout the paper, is constructed by measuring the (wage-weighted) average months of experience plus on-site and on-the-job training required to work in each industry.

Using this measure, industries related to legal, financial and engineering services rank highly in the list among the knowledge intensive industries. In the manufacturing sector, industries ranking highly are computer related (SIC 3573, 3571 and 3572), communications equipment (SIC 3669, 3663 and 3661) and electronics and semiconductors (SIC 3672, 3674 and 3676). Appendix Section A.5 expands on this discussion.

One limitation of this measure is that it is based on US data. Full precision would require to compute these weighted averages using data on occupations per industry for each country separately. However, this data is unavailable, and I will assume the ranking in the knowledge intensity of industries based on US data proxies that of the rest of the world.

I find that this measure correlates positively with other knowledge intensity measures used in the literature, such as the average R&D share of sales per industry (e.g., Nunn & Trefler 2008; Keller & Yeaple 2013), as evidenced in Table 1.[20]

[Table 1 about here.]

[19] Appendix Section A.6.2 presents robustness tests of the empirical analysis using a measure averaging the experience indicators only (excluding the training indicators).

[20] It also correlate positively with other measures that could proxy for knowledge intensity or complexity. The correlation coefficient with the share of non-production workers in total employment, from the NBER-CES Manufacturing Industry Database (Becker et. al. 2013), is 0.68. Similarly, the correlation coefficient with the Product Complexity Index, developed by Hausmann et. al. (2011), is 0.49.

The R&D based measures, however, have three main shortcomings that could generate significant biases. First, these measures assign a zero value to about half of the industries, because most firms within those industries have no R&D investment whatsoever. For these industries in the lower end of the distribution, the intensity of their knowledge is indistinguishable.[21] Second, since these measures are computed by averaging across each industry the R&D share of sales reported by a (random or not) sample of firms, they are likely to favor industries in which larger firms are more prevalent. This might happen in industries for which the barriers to entry are higher, and not necessarily knowledge intensive industries. Third, R&D investment might not be equally accounted for across all industries.

The O*NET based measures solve these issues. Their distribution is smoother (see Appendix Section A.5), they do not rely on a sampling of firms, and they use the same standardized measure for all industries. Hence, I use these indicators as the main proxies for knowledge intensity throughout the paper.

4.2.3 Unit shipping costs

Unit shipping costs for SIC manufacturing industries are computed using data from Bernard, Jensen & Schott (2006).[22] This industry-level measure aims to proxy for t, referred to in the theoretical framework as the unit shipping cost variable, which accounts for how costly it is to transport one unit of that good irrespective of industry. For instance, goods with the highest unit shipping costs in the dataset include ready-mixed concrete and ice, which require special forms of transportation.

The variable measures the amount of US dollars required to transport 1$ worth of a good per every 100Km. It is computed by averaging the same measure per industry across all countries exporting to the US in year 2005.

[21] See Appendix Section A.4.

[22] Downloadable from http://faculty.som.yale.edu/peterschott/files/research/data/xm_sic87_72_105_20120424.zip

To deal with long tails, this variable will be used in a logarithmic scale in all the different empirical specifications.

4.2.4 Ease of Communication Proxies

In order to proxy for the ease of communication between a subsidiary and its headquarters, I use three variables: non-stop flights, working hours overlap and common language.

The first variable is used because the existence of non-stop flights would proxy for the ease of managers and workers to do more frequent business trips, given the convenience of a direct flight. Business trips, by allowing face-to-face interaction, would facilitate the transmission of tacit knowledge. However, it is important to note that business trips, even if convenient, happen much less often than phone calls due to the elevated costs associated with them. In order to compute the existence of a non-stop air route between a headquarters and its subsidiary, I matched all the existing airports within a 100Km radius (conditional on being in the same country), using the geocoded latitude and longitude. The data for airports (with their respective coordinates) and active air routes come from OpenFlight.com.[23] Through this matching I create a dummy variable which takes the value of 1 if there is a non-stop flight between the headquarters and its subsidiary.[24]

The second variable, overlap in working hours, aims to capture the "real-time" communication ability between managers and workers in the two plants. Being in the same time zone allows workers to use phone or videoconference communication more frequently (substituting partially for means such as fax or email). This allows for better transmission of tacit knowledge, which is valuable for troubleshooting or crisis solving. In order to compute the over-

[23]Data downloadable from http://openflights.org/data.html. Downloaded in June 2013.

[24]I also compute the minimum number of non-stop flights required to travel between two given airports by using the shortest path algorithm. The results using this measure, however, are qualitatively the same as the ones that use the non-stop flight dummy. Thus, this measure is omitted in the analysis.

lap in working hours I use the geocoded longitude of each subsidiary to find its time zone, and compare it to that of its headquarters. Assuming that working hours run from 8:00am to 6:00pm (10 hours in total), the variable measures, for a single day, the number of hours that overlap in the working schedule of both the headquarters and its subsidiary.

Finally, a common language captures cultural proximity, and also better ability to communicate between workers in both locations of the same firm. The common language comes from CEPII's GeoDist database (Mayer and Zignago, 2011). Two countries have a common language if at least 8% of the population in both countries speak such language.

The sample is merged with data that proxies for the ease of communication between the headquarters and its foreign subsidiaries: the existence of a non-stop air route between their nearby airports, the number of overlapping working hours in a given day and whether there is a common language spoken in their respective countries.

5 Empirical Analysis

This section first discusses the broad empirical strategy and then presents descriptive statistics from the sample. The following subsection presents results of the empirical analysis that test the propositions presented in the theoretical framework section. The remaining subsections present additional evidence consistent with the assumption that the barriers of transferring knowledge are higher for longer geographic distances.

5.1 Empirical Strategy

The conceptual framework outlined above is useful to understand the determinants of horizontal expansion for MNCs. The empirical section focuses on understanding the role of knowledge in particular.

In spite of the lack of firm-level export data in the sample, I test for the implications of the model through *reduced forms* that look at the determinants of horizontal FDI, as compared to both domestic subsidiaries and non-horizontal foreign subsidiaries.

The first empirical exercise will deal with testing Proposition (1) of the conceptual framework: are knowledge intensive activities less likely to be replicated abroad? To do so, I will look at a sample of domestic and foreign (horizontal) subsidiaries, and estimate the likelihood of an industry being replicated abroad given its knowledge intensity.

In order to test Proposition (2), that is, whether longer distances makes the knowledge transmission process more costly for firms, I rely on the complete portfolio of foreign affiliates of the MNCs in the sample. Thus, the question the empirical specification asks is: conditional on having a foreign plant in a given industry and location, is it likely to be an horizontal one, given its distance to the headquarters and the knowledge intensity of its economics activity? More broadly, the goal of the exercise is to test whether the patterns for horizontal subsidiaries in the data are consistent with the mechanisms described in the model. The underlying assumptions for this analysis to serve as proof of the raised question are described in the next section.

Following this, I relax the assumptions of the previous analysis, and use only the horizontal foreign subsidiaries to test a deviation of the conceptual framework. More specifically, I test for a negative correlation between distance and knowledge intensity, which follows the model's prediction.

It is important to clarify that this exercise does not substitute for, nor it intends to, using firm-level export data as part of the identification strategy. Yet, the exercise adds value by presenting stylized facts that are robust, and at the same time, consistent with the mechanisms of the conceptual framework with respect to the role of knowledge transmission in explaining the existence and location of horizontal subsidiaries of a MNC.

5.2 Descriptive statistics

This section provides descriptive details about the sample, in terms of the distribution of foreign affiliates across regions of the world and developing vs. developed countries.

Table 2 presents descriptive statics which compare domestic to horizontal foreign subsidiaries in the sample. This sample includes domestic subsidiaries and foreign subsidiaries that replicate production abroad (i.e. an horizontal expansion). In total, there are 64,462 subsidiaries that are owed by 1540 MNCs. Domestic subsidiaries tend to be more numerous than foreign ones (on average, 29% of these subsidiaries are foreign). The table also includes the knowledge intensity variable measured in standard deviations from the mean (denoted by KI), averaged over domestic and over foreign subsidiaries. The last column presents the difference, with stars denoting the correspondent p-value level.

[Table 2 about here.]

As it can be seen, on average, industries of the foreign subsidiaries are roughly as half as knowledge-intensive as the industries manufactured by their domestic counterparts. The same pattern holds for all presented cuts of the data, besides for few firms based on non-OECD countries (with a p-value of 0.10), and for few firms based on Western Europe, for which the difference is not statistically significant. This statistic is consistent with Proposition (1) of the conceptual framework. I use this sample in the next subsection to analyze this proposition further.

Table 3 summarizes the number of records in the sample that includes only foreign subsidiaries, both horizontal and non-horizontal. Overall, there are 8,266 MNC firms which have 60,621 foreign subsidiaries. Of those subsidiaries, 34% are defined as horizontal expansions, while the rest could be vertical subsidiaries or, simply, a foreign subsidiary in a non-related industry[25]. The average distance in the sample between headquarters and subsidiaries is 5,152Km. Regarding communication proxies, a subsidiary and its headquarters overlap, on average, 7.3 working hours in a given day, and for about 25% of subsidiaries there exists a commercial non-stop flight from their headquarters. The following rows present the same statistics across different cuts of the sample, based on the headquarters' country. For instance, most of the foreign subsidiaries are located in OECD countries (49,936 vs. 10,685). Similarly, the table shows that most of the foreign subsidiaries in the sample are located in Western Europe and North America.

[Table 3 about here.]

I also present results of the distribution of sectors among foreign affiliates, to understand whether in the sample there are some sectors that are more likely to appear (i.e. be reported) than others. In terms of industries, the distribution of different sectors in the sample is not homogenous, as can be seen in Figure 5. Some sectors are more prevalent than others in the data. The industries that appear the most in the data are Ready-Mixed Concrete (SIC 3273), Pharmaceutical Preparations (SIC 2834) and Motor Vehicles Parts (SIC 3714). To alleviate concerns on how this distribution could affect the results, all the standard deviations calculations allow for clustering at the industry level.

[Figure 5 about here.]

In addition, it is worth emphasizing that each foreign subsidiary in the sample manufactures a specific product. Hence, if a MNC has several foreign subsidiaries, then each one of those could be manufacturing a different product (in its 4 digit classification). The sample that a single MNC that has more than one foreign subsidiary could be manufacturing more than one product.

[25] For instance, this could be the result of a MNC diversifying its portfolio by acquiring foreign firms.

Figure 6 shows that larger MNCs (as measured by number of affiliates) tend to make a larger number of different products.

[Figure 6 about here.]

5.2.1 Notes on the Reliability of the Data

The Worldbase dataset collected by Dun & Bradstreet is sourced from a number of reliable organizations all over the world, including public registries. According to Dun & Bradstreet's website, "the data undergoes a thorough quality assurance process to ensure that our customers receive the most up-to-date and comprehensive data available".[26] However, it is important to acknowledge that, given the lack of access to public registries for every country, it is not possible to asses with full accuracy the representativeness of the data. Alfaro and Charlton (2009) compare the dataset with the US multinational firms sample by the US Bureau of Economic Analysis, and find consistencies between the two datasets. Moreover, the regional breakdown of foreign subsidiaries presented in Table 3 below seems to be consistent with aggregate figures of FDI inflows across world regions.

Some basic relationships drawn from the sample also behave as expected. For instance, the number of countries in which an MNC has foreign affiliates is related to the overall size of the MNC. Figure 7 presents the relationship between the size of firms (in number of establishments in the left panel, and in total number of employees in the right panel[27]) against the number of foreign countries in which their subsidiaries are located (on the vertical axis). Each observation in the scatterplot is an MNC labeled with its headquarters' country ISO3 code. The figure shows smaller MNCs are present in fewer countries, while larger MNCs tend to be more spread out in terms of the number of countries they have a presence in.

[26] http://dnb.com.au/Credit_Reporting/The_quality_of_DandBs_data/index.aspx

[27] Including their domestic plants for both.

[Figure 7 about here.]

Focusing the analysis on the within-firm dimension significantly diminishes the sampling concerns further. This is because, while methods for gathering information may not be symmetric across countries, they would not systematically differ by firm or by industry. The per-country likelihood of missing data would be the same for all firms and industries, controlling for the location of the MNC. Thus, concerns regarding biases caused by possible sampling asymmetries are not particularly large for the purpose of this empirical exercise.

5.3 Knowledge intensity as a determinant of horizontal expansion

The first empirical exercise deals with understanding the determinants of horizontal expansion, with guidance of the theoretical model outlined above. It uses the sample that includes both domestic and foreign horizontal subsidiaries (described in Table 2). Thus, the unit of analysis is a subsidiary. The empirical specification is a *reduced form* of the exports-FDI tradeoff. That is, the analysis aims to understand the differential patterns between domestic and foreign horizontal subsidiaries. In terms of the theory presented above, the underlying assumption is that, whenever a firm sell a particular product to a foreign market, domestic subsidiaries are associated with exports whereas foreign subsidiaries are associated with FDI, and thus substitute for exports. The question asked in this exercise is, what characteristics of an industry make it more likely to be replicated abroad (i.e. that exist as a foreign subsidiary)?

According to Equation (3), $a_I^{1-\varepsilon}$, which represents the productivity threshold after which a firm engages in FDI, is a function of $\tau(t,d)$ and $\kappa(t,d)$ as well as the fixed costs and demand variables. It is important to clarify that industries may vary in their $a_I^{1-\varepsilon}$ threshold, and its value will determine the likelihood of horizontal expansion for that industry (given the distribution of productivity for firms within each sector). That is, controlling for demand variables and fixed costs, industries with a lower $a_I^{1-\varepsilon}$ will be more likely to be horizontally expanded, and vice-versa.

In this context, for a given firm and location, if a subsidiary is replicating production abroad (i.e. foreign and horizontal), it implies that the productivity level of such MNC goes beyond the minimum industry-specific threshold $a_I^{1-\varepsilon}$ for which FDI becomes more profitable than exports, in that industry. Controlling for MNC productivity, thus, exploiting variation in observed variables will shed light on the the determinants of the $a_I^{1-\varepsilon}$ value for different industries, or alternatively, the likelihood of horizontal expansion[28]:

$$Foreign_s = \beta_k \cdot k_s + \beta_t \cdot log(t_s) + controls_{h,s} + \varphi_h + e_{h,s} \tag{7}$$

Where the independent variable is a dummy which takes the value 1 if the subsidiary is a foreign horizontal affiliate of the firm (and 0 if it is a domestic one). k_s is a measure of knowledge intensity of the economic activity (i.e., the manufactured good or product) of the foreign subsidiary. t_s is the unit shipping cost for the good manufactured in the foreign subsidiary. $controls_{h,s}$ is a vector of variables that control for the size of the market and factor endowments of the host country relative to that of the country of the headquarters,[29] which controls for aggregate demand and cost of producing in the host country. φ_h represents MNC fixed effects, which controls for the productivity level a of the firm. It is worth mentioning that subsidiaries within a single MNC might differ in their economic activity, thus allowing for within-firm variation in the right hand side variables of the empirical specification (see Figure 6). Finally, $e_{h,s}$ is the error term.

According to the theoretical framework presented above, we expect the following β_k to be negative (see Equation (5) and Figure 1).

[28] That is, $Prob(a^{1-\varepsilon} > a_I^{1-\varepsilon})$.
[29] i.e., $y_{h,s} = log(y_h) - log(y_s)$.

The results are presented in Table 4. All the columns include the control variables. The table uses the *experience plus training* measure discussed above as a proxy for k, which is measured in standard deviations from the mean.

[Table 4 about here.]

Column 1 presents the complete specification. The results suggest that, everything else equal, industries that are one standard deviation above the mean in terms of their knowledge intensity, are 3.6 percentage points less likely to be replicated abroad. This represents a reduction of about 12% given the unconditional probability of being a foreign affiliate in the sample (which is 29% as shown in Table 2). For instance, according to this estimation, semiconductors (SIC 3674), which is characterized by having workers with an average of over 80 months of required experience plus training[30], is about 30 percentage points less likely to be replicated abroad than a meat packing plant (SIC 2011), which its workers have, on average, 37 months of experience plus training.

The estimator for β_k is robust across all specifications. This result controls for the unit shipping cost, and for the size of the market and factor endowments of the host country relative to that of the country of the headquarters[31]. According to the theoretical framework above, this is a straightforward result from the assumption that ${}^{\partial\kappa}/_{\partial k} > 0$

The inclusion of host country fixed effects in Column 4 rules out other potential stories that could be driving the results. For instance, poor intellectual property regulation in different countries.[32] The estimate of β_k is robust to the inclusion of this set of fixed effects in terms of its magnitude,

[30] see Appendix Section A.5

[31] This is 1 for all domestic subsidiaries, naturally.

[32] Appendix Section A.6.4 presents results excluding China from the sample, to alleviate possible biases this country might generate in the results due to IP concerns. The results are robust to the exclusion of China.

negative sign and its statistical significance.[33]

The analysis presented in Table 4 seems to support Proposition (1) of the model. The next subsection focuses on Proposition (2).

5.4 Is the cost of knowledge transfer increasing in distance?

The previous sample is not useful to test the implications of distance, because there is no information on where are the domestic subsidiaries exporting to, if at all. Thus, to test the implications of Proposition (2), I will use a dataset that includes only foreign affiliates. The idea is to understand whether there are differential patterns in the data for horizontal affiliates (i.e. replication of production), using as a comparison group the non-horizontal subsidiaries. The underlying assumption of using non-horizontal foreign subsidiaries as a counterfactual is that the marginal cost of transferring knowledge is zero (or very little) for non-horizontal subsidiaries, where as they are alike in all unobservables after the controls. While there is likely a fixed cost of transferring knowledge to non-horizontal subsidiaries when they are created or acquired, the assumption of zero *marginal* costs for this type of subsidiaries relates to the intuition that there is less the headquarters can do to offer ongoing troubleshooting or to train workers in these plants when it comes to production lines that are essentially different from the ones that exist at home. Therefore, controlling for variables that would explain the decision of a firm to locate a foreign subsidiary in a given location (regardless of whether it is horizontal or not), the residual differences could be attributed to the cost of transferring knowledge, and more so if they are consistent with the conceptual framework. Yet, given the assumption is not testable, Section 5.4.1 below relaxes it, and find consistent results.

[33]The results are robust to using parent industry (2-digit) interacted with host country fixed effects, to allow for differential policies at the country level for different types of industries.

The empirical specification for this exercise is described in equation 8:

$$HOR_s = \beta_k \cdot k_s + \beta_d \cdot log(d_{h,s}) + \beta_t \cdot log(t_s) + controls_{h,s} + \varphi_h + e_{h,s} \quad (8)$$

Where the independent variable is a dummy which takes the value 1 if the subsidiary (indexed by s) in that observation is a horizontal foreign affiliate, and 0 if, instead, is a non-horizontal foreign subsidiary. Again, k_s is a measure of knowledge intensity in standard deviations from the mean, associated with the foreign subsidiary. $d_{h,s}$ is the distance between the headquarters and the foreign subsidiary. t_s is the unit shipping cost for the good manufactured in the foreign subsidiary. $controls_{h,s}$ is the same vector as in specification 7. Similarly to before, φ_h represents MNC fixed effects and $e_{h,s}$ is the error term. If the mechanisms of the model are in place, we could expect a negative β_d, which could *only* be explained if κ increases with distance

The results are presented in Table 5. All the columns include the control variables. The table uses the *experience plus training* measure discussed above as a proxy for k.

[Table 5 about here.]

Column 1 presents the complete specification, while the other columns vary in the number of variables used in the regression. The estimator for β_k is negative and statistically significant; the estimator for β_d is also negative and statistically significant; and the estimator for β_t has the expected positive sign, but lacks statistical significance.

Before turning into the coefficient of interest for this exercise (β_d), it should be noted that the negative sign for the estimator of β_k is consistent with the previous results in Table 4 and Equation (5) of the theoretical framework. More specifically, an industry with a knowledge intensity measure one standard deviation above the mean is about 8.7 percentage points less likely to be horizontally expanded, compared to non-horizontal affiliates. Hence, once again, the data suggests that the barriers associated with the transmission of knowledge from the headquarters to the subsidiaries are an important determinant of horizontal expansion.

Across all specifications that include $log(d)$, the estimator for β_d remains negative and statistically significant. As explained above, in a model that ignores the cost of transferring knowledge, an increase in distance will unequivocally *increase* the incentives for horizontal FDI. However, only the inclusion of κ in the model as an increasing function of distance would explain the obtained results, which suggests that an increase in d would *reduce* the likelihood of horizontal FDI.[34]

What does $\beta_d < 0$ imply? The theoretical model, as summarized in the right panel of Figure 2, contemplates a case in which a firm serving a further away market would be better-off by exporting than by setting up a foreign affiliate, because transmitting knowledge to this remote location will significantly lower profits from FDI relative to profits from exports. However, given that the empirical specification is a reduced-form of the theoretical implications, it is not possible to distinguish between the case in which the firm effectively *substitutes* FDI with exports, or alternatively, the case in which the firm *reduces* its horizontal FDI in absolute terms, driven by a reduction of total sales (both through FDI and exports) in a further away location. In both cases, though, the negative sign of β_d implies that the cost of transferring knowledge is increasing with distance. In fact, given that the control group includes vertical subsidiaries, there are less reasons to expect this result. In theory, vertical subsidiaries are located closer to the headquarters as compared to horizontal subsidiaries (given the transportation costs associated with importing the intermediate goods from the vertical subsidiary to the headquarters). Therefore, a negative estimator for β_d is even more striking.

[34] See Appendix Section A.1 for more details on the theoretical conditions for this to happen.

Finally, the estimator for β_t is positive in sign, though statistically insignificant across all specifications. The positive sign is consistent with the proximity-concentration hypothesis: firms will tend to serve foreign markets through foreign affiliates for goods with larger trade costs (e.g., Brainard 1993, 1997; Helpman, Melitz and Yeaple 2004).

Similarly to Table 4, Column 4 includes host country fixed effects, which would control for poor intellectual property regulation in different countries. It is important to acknowledge that the specification lacks variables that control for industry-specific fixed costs of exporting and creating new plants. It can be argued that, as long as fixed costs are not dependent on k or d, then the results are indicative of the explained mechanisms. If fixed costs are the same across industries and countries, then their exclusion should not bias the results. If the fixed costs are country-dependent, then the controls included in Column 4 would account for them.

5.4.1 A trade-off between distance and knowledge intensity

This subsection relaxes the underlying assumption which was required to compare horizontal to non-horizontal foreign subsidiaries stated above, which is critical to correctly interpret the estimation of β_d.

The theoretical framework above provides guidance to address this question in a different way. A firm's total profit when it engages in FDI is $\pi = \pi_D + \pi_I$. Given that π_I is subject to cost κ, and κ increases both in k and d, then $\partial\pi/\partial k < 0$ and $\partial\pi/\partial d < 0$. Figure 8 abstracts from the model the expected relationship between d and k that drives a firm's decision to engage in FDI. The figure includes the case that assumes linear relationships. In it, each line represents a profit function. The curve $\pi_D + \pi_E$ represents total profits for an exporting firm, while the curve $\pi_D + \pi_I$ represent total profits for a firm engaging in FDI instead. The profit for an exporting firm does not vary with the level of knowledge intensity (k) of the good, whereas the profit for the same firm if it engages in FDI instead does vary with k. Both profits functions decrease in distance, represented by d.

In all cases, however, it can be seen that for higher levels of k and d (i.e., knowledge intensity and distance, respectively) firms would be better off by substituting exports with FDI. The opposite happens for cases in which both k and d are low. Moreover, even when firms engage in FDI, their profits decrease with both distance and knowledge intensity. Thus, MNCs in knowledge intensive products would be better off by locating their foreign subsidiaries at closer geographic distance.

[Figure 8 about here.]

I explore whether relationship between k and d described above is seen using only the horizontal foreign subsidiaries in the data. That is, conditional on being an horizontal foreign subsidiary, do we see a clear negative relationship between the knowledge intensity of its sector, and the distance to its headquarters? The proper way to do this is to analyze these variables after controlling for the regressors in Specification (8). Hence, this exercise has two steps.

First, I decompose distance and knowledge intensity and keep the part that is not explained by these other regressors (i.e., the residuals). That is, I define:

$$\begin{aligned} U\left[log(d)\right] &= log(d) - \gamma_t^1 log(t) - controls'_{h,s}\gamma_c^1 - \varphi_h \\ U\left[k\right] &= k - \gamma_t^2 log(t) - controls'_{h,s}\gamma_c^2 - \varphi_h \end{aligned}$$

Where the γ coefficients are estimated by regressing $log(d)$ and k on the regressors, limiting the sample to horizontal foreign subsidiaries only. Notice that the inclusion of MNC firm fixed effect, imply that the residuals will contain within firm variation only.

The second step is to estimate $U\left[log(d)\right]$ and $U[k]$ using the sample, and to find a functional form that properly fits the relationship under consideration. As explained above, we expect this relationship to be negative. Figure 9 presents the results of this exercise, using the *experience plus training* indicator as a proxy for k. The left column performs a linear fit between k and d while the right column performs a quadratic firm between the two.

[Figure 9 about here.]

The linear fit shows a monotonic decreasing relationship between k and d, as depicted in Figure 8. In its linear form, the calculation suggests that the distance to the headquarters is shorter by 7.8% for every standard deviation above the mean in knowledge intensity. This implies that, for an American MNC, a meat packing subsidiary would be located in Turkey (approximately, 10000Km from USA), while a semiconductor plant would be located in Ireland (approximately 6500km from USA), *ceteris paribus*. The evidence hence suggests, that, indeed, firms face a trade-off between distance and knowledge intensity, providing further evidence on the fact that the cost of knowledge transmission is a function of both these terms.

Interestingly, the quadratic fit suggests an inverted U-shaped relationship. That is, the estimated quadratic relationship does not seem to be monotonically decreasing for the lower values of k (although a flat or even negative slope in that area cannot be rejected in the data either). However, and perhaps more importantly, for higher levels of knowledge intensity there is a clear negative relationship with distance. This result is qualitatively important, given that it would be consistent with the idea that distance appears to matter much more for higher levels of knowledge intensity. Intuitively, this means that after certain level of knowledge intensity, the more sophisticated products are the closer the foreign subsidiaries will be located to the headquarters. The negative second derivative implied in the fit suggests that the documented negative relationship intensifies with the level of knowledge intensity.

Overall, the data supports the existence of a trade-off between distance and knowledge intensity for horizontal foreign subsidiaries. This implies that, if MNC do expand horizontally, the foreign affiliates will tend to be geographically closer to the headquarters the more knowledge intensive the product under consideration is.[35]

These results, in their reduced form, are consistent with the ones found by Keller and Yeaple (2013). They find that distant foreign affiliates in knowledge intensive sectors perform worse. The authors, however, attribute these results to additional trade costs due to the substitution of transferring knowledge with intermediate goods. The framework and results presented above present an alternative explanation to this finding, in which the performance of subsidiaries is highly affected by the inefficiencies of transferring knowledge at longer distances, and not higher intra-firm cost induced by intermediate goods. The next section explores this issue further.

5.5 Ease of communication and knowledge transmission

The empirical results, while consistent with the theoretical framework, might be driven by factors other than knowledge not accounted for, in the presence of omitted variable bias. For instance, a conventional explanation in the literature would be that knowledge intensive sectors are associated with higher intra-firm trade of intermediate goods, making it less profitable to locate those plants in far away locations (Irrazabal et. al. 2013; Keller and Yeaple 2013).

Keller and Yeaple, in particular, assume that knowledge can be fully embedded in intermediate goods, that are in turn shipped to remote locations. However, this assumption is not feasible for tacit knowledge. Thus, it could well be that it is the cost of transmitting tacit knowledge which drives the documented relationship.

This subsection performs a test that disentangles between both explana-

[35]Appendix Section A.6.5 replicates these results excluding foreign subsidiaries located in Western Europe owned by a Europe-based MNC, given the relative shorter distances within the continent. Results are robust to the exclusion of these firms.

tions. If the cost of transferring knowledge is indeed an increasing function of distance – as argued – and thus, a determinant in the location decisions of firms, then easier communication between headquarters and subsidiaries would work as a cost-reducing mechanism for the purpose of transmitting knowledge. This would be hard to explain with the intra-firm trade mechanism, given that the ease of communication is orthogonal to the transportation costs of intermediate goods.

I test for this hypothesis by estimating an extended version model (8) which includes variables that proxy for the ease of communication within the firm. These variables, all measured for each subsidiary and its headquarters, are (1) the existence of a commercial non-stop air route (between airports within 100Km); (2) the number of overlapping working hours in a business day; and (3) a binary variable indicating whether the countries of both the headquarters and the subsidiary speak a common language[36] (see Section 4.2.4 for more details on the construction of these variables).

The purpose of utilizing these variables is to proxy for forms of communication that allow for the transmission of tacit knowledge, though they are quite different between themselves. As explained above, business travel provides the opportunity to work face-to-face, though it occurs with less frequency, given the high costs of traveling.[37] Being in the same time zone allows for convenient real-time, day-to-day, communication, significantly reducing waiting time between the two ends for problem solving or consulting about specific tasks.[38] Lastly, if two countries speak a common language, it is more likely that the workers in both the local and remote locations of the same firm can communicate more easily, either in person or remotely, and better communicate (and more often) with each other.

[36] Defined as a language that is spoken by 8% or more of the population in both countries.

[37] Giroud (2012) finds that the existence of commercial air routes between subsidiaries and headquarters positively affects the profitability of the former.

[38] Stein and Daude (2007) find that time zone is an important determinant of aggregate FDI flows, which they attribute to better monitoring.

The results are presented in Table 6. All columns use the *experience plus training* indicator to proxy for k.

[Table 6 about here.]

Column 3 of Table 6 shows that the estimator for β_d is reduced in magnitude by two thirds of its original value, and becomes statistically insignificant when using the number of overlapping working hours as a control (as compared to Column 1, which replicates the first specification of Table 5). This result suggests that being in the same time zone reduces the barriers to transferring knowledge induced by the distance component (given that the estimator for β_k maintains its magnitude and negative sign in those specifications, implying only the distance channel in $\kappa(k, d)$ is affected). That is, real-time communication effectively "reduces" the distance between the headquarters and its subsidiaries, by about two thirds. For the average foreign subsidiary, being in the same time zone is equivalent to being geographically closer to the headquarters by about 3500 Km. The ability to communicate on real-time for troubleshooting or other purposes, avoiding long waiting times, seems to ease knowledge transmission more than the ease of face-to-face interaction. The costs associated with the knowledge intensity component still seem play a role, regardless of communication.

An alternative explanation of the previous results that relies on shipping costs of intermediate goods can be ruled out: transportation costs should be just as expensive north to south as they are east to west.

In terms of the other variables, it can be seen in Column 2 that the existence of a non-stop commercial air route seems not to change the original results, thus hinting that face-to-face interaction plays a lesser role in the stated mechanisms. However, Column 4 shows that having a common language also effectively reduces the distance between a headquarters and its subsidiary by less than half.

Figure 10 replicate Figure 9, this time adding as additional controls all the ease of communication variables that are included in Table 6. The left panel shows the linear fit with the original controls, while the right panel controls also for the ease of communication. It can be seen that the slope that defines the relationship between k and d after controlling for the ease of communications is about 33% flatter. While the negative relationship still seems to hold, the reduction in the slope is consistent with the results presented in this section.

[Figure 10 about here.]

These findings are insightful on their own. The results suggest that being in the same time zone and speaking a common language seems to facilitate the transmission of knowledge. The ability of managers in the headquarters to communicate with colleagues in foreign locations, for troubleshooting or consulting on an open-ended range of issues, is more efficient when communication happens in real time, without long waiting times. This might be even more relevant for transmitting tacit knowledge, given that complicated problems would require real-time interaction, and not just explanations being sent through fax or email. Furthermore, this logic could also serve as an example for arguing that the barriers of transmitting knowledge is increasing with distance: managers and workers in the headquarters might require working extra hours to communicate with their peers in foreign subsidiaries, incurring additional compensation and operational costs.

6 Concluding Remarks

This paper has provided evidence on the important role of knowledge, and the difficulties associated with its transmission in the day-to-day activities of MNCs. Sizable costs of transferring knowledge, even within firms, would have an impact on their strategies to either export or undertake foreign investment, directly affecting the global economy in terms of trade and capital flows. Furthermore, the empirical analysis presents evidence of a tradeoff firms face, which drives them to locate foreign subsidiaries producing knowledge intensive goods in geographic locations that are closer to the headquarters. Thus, knowledge is not lighter than air. Rather, its diffusion is difficult and costly, and hence it has implications on economic activity.

These findings are not inconsistent with the mechanism of the proximity-concentration hypothesis (e.g., Brainard 1997), yet they introduce a new and unexplored dimension. The cost of transferring knowledge plays a role that counteracts the incentives to engage in FDI driven by transportation costs. Hence, FDI does not necessarily become more profitable than exports for all industries with high transportation costs.

More generally, the fact that geographic distance hinders the process of knowledge transmission is a result that defies the traditional way economists have thought about FDI and MNC activity. In most of the international economics research, it is taken as a given that knowledge is fully transferrable without incurring any costs whatsoever – not even for different types of technologies or goods. However, if one takes into account the large variety of different industries that exist in the world, and how they can dramatically differ in almost any dimension, it follows that we can expect each firm to set a strategy that is dependent on the types of products they produce and sell. In a globalized economy, being able to sell products at a global scale requires a minimum level of productivity, which firms achieve by acquiring productive knowledge. The way firms acquire and maintain this knowledge is through their workers in the headquarters and all of its relevant subsidiaries (domestic and foreign). The finding that knowledge transmission incurs costs that are dependent on distance would thus have a significant impact on the expansion decisions of MNCs.

Nonetheless, this paper has left open some other specific questions that will shed light on our general understanding of knowledge. For instance, is the cost of knowledge transmission a relevant determinant for service provider firms, as it is for manufacturing firms? Given the difference in the nature of services vs. manufacturing industries in terms of their tradability, we can expect different patterns in the data. Also, how does the knowledge intensity of the good relate to the existence of regional hubs, as opposed to different plants serving every foreign market? What tools and means are at a firm's disposal to enhance the process through which it transfers knowledge to its subsidiaries and workers? These and other questions are an essential part of the future research agenda.

Naturally, this research agenda also contains questions that have relevant policy implications. While governments intend to develop their private sectors by attracting foreign investment, designing an effective policy should answer questions such as: is there enough infrastructure in place to allow effective communication for foreign firms? Should the focus be on specific types of firms and specific industries for which knowledge transmission will be easier? Do all types of products have the potential to generate productivity spillovers to domestic firms, or only those for which the cost of knowledge transmission is low?

All in all, despite the fact that productivity outweighs factor accumulation in growth accounting exercises (Hall and Jones 1999, Caselli 2005), the process through which knowledge is accumulated by economic agents is still an under-researched area. However, a better understanding of this process is critical to answering open questions in economics. The difficulties associated with transferring and acquiring knowledge, which translates into productivity shifts, are not unique to MNCs. They can also relate to domestic firms (e.g., Bloom et. al. 2013; Kalnins and Lafontaine 2013), investors (e.g., Coval and Moskowitz 2001), innovation (e.g., Kerr 2008) and even countries' export baskets diversification (Bahar et. al. 2014). At a larger scale, the documented evidence reinforces the importance of knowledge transmission in overall economic activity. Thus, understanding the ways knowledge affects economic activity lies at the core of important and unanswered questions on convergence, development and growth. Knowledge and its diffusion, after all, are significant phenomena that can alter global economic patterns in as-of-yet unexplored ways.

7 References

References

Acemoglu, Daron, Simon Johnson, and Todd Mitton. "Determinants of Vertical Integration : Financial Development and Contracting Costs." *The Journal of Finance* LXIV, 3: (2009) 1251–1290.

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

Alfaro, Laura, and Maggie Xiaoyang Chen. "Surviving the Global Financial Crisis: Foreign Ownership and Establishment Performance." *American Economic Journal: Economic Policy* 4, 3: (2012) 30–55.

Arrow, Kenneth J. "Classificatory Notes on the Production and Transmission of Technologcal Knowledge." *The American Economic Review* 59, 2: (1969) 29–35.

Bahar, Dany, Ricardo Hausmann, and Cesar A. Hidalgo. "Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion?" *Journal of International Economics* 92, 1: (2014) 111–123.

Becker, Randy, Wayne Gray, and Jordan Marvakov. "NBER-CES Manufacturing Industry Database: Technical Notes.", 2013.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott. "Trade costs, firms and productivity." *Journal of Monetary Economics* 53, 5: (2006) 917–937.

Black, SE, and PE Strahan. “Entrepreneurship and bank credit availability.” *The Journal of Finance* LVII, 6: (2002) 2807–2833.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. “Does Management Matter? Evidence from India.” *The Quarterly Journal of Economics* 128, 1: (2012) 1–51.

Bottazzi, Laura, and Giovanni Peri. “Innovation and spillovers in regions : Evidence from European patent data.” *European Economic Review* 47: (2003) 687–710.

Brainard, SL. “A simple theory of multinational corporations and trade with trade-off between proximity and concentration.” *NBER Working Paper Series* , 4269.

———. “An empirical assessment of the proximity-concentration tradeoff between multinational sales and trade.” *The American Economic Review* 87, 4: (1997) 520–544.

Carr, DL, JR Markusen, and KE Maskus. “Estimating the knowledge-capital model of the multinational enterprise.” *The American Economic Review* 91, 3: (2001) 693–708.

Caselli, Francesco. “Accounting for cross-country income differences.” *Handbook of Economic Growth* 1, 05: (2005) 679–741.

Caves, RE. “International Corporations: The Industrial Economics of Foreign Investment.” *Economica* 38, 149: (1971) 1–27.

Costinot, Arnaud, Lindsay Oldenski, and James Rauch. “Adaptation and the boundary of multinational firms.” *The Review of Economics and Statistics* 93, February: (2011) 298–308.

Coval, Joshua D., and Tobias J. Moskowitz. "The Geography of Investment: Informed Trading and Asset Prices." *Journal of Political Economy* 109, 4: (2001) 811–841.

Fan, JPH, and LHP Lang. "The Measurement of Relatedness : An Application to Corporate Diversification." *The Journal of Business* 73, 4: (2000) 629–660.

Giroud, Xavier. "Proximity and Investment: Evidence from Plant-Level Data." *The Quarterly Journal of Economics* 128, 2: (2012) 861–915.

Hall, R. E., and C. I. Jones. "Why do Some Countries Produce So Much More Output Per Worker than Others?" *The Quarterly Journal of Economics* 114, 1: (1999) 83–116.

Harrison, Ann E., Inessa Love, and Margaret S. McMillan. "Global capital flows and financing constraints." *Journal of Development Economics* 75, 1: (2004) 269–301.

Hausmann, Ricardo, César A Hidalgo, Sebastián Bustos, Michele Coscia, Sarah Chung, Juan Jímenez, Alexander Simoes, and Muhammed A. Yildirim. *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. Cambridge, MA, 2011.

Helpman, E. "A Simple Theory of International Trade with Multinational Corporations." *The Journal of Political Economy* 92, 3: (1984) 451–471.

Helpman, Elhanan, Marc J Melitz, and Stephen R Yeaple. "Export Versus FDI with Heterogeneous Firms." *American Economic Review* 94, 1: (2004) 300–316.

Hippel, E Von. "'Sticky information' and the locus of problem solving: implications for innovation." *Management science* 40, 4: (1994) 429–439.

Irarrazabal, Alfonso, Andreas Moxnes, and LD Opromolla. "The Margins of Multinational Production and the Role of Intrafirm Trade." *Journal of Political Economy* 121, 1: (2013) 74–126.

Jaffe, A.B., M. Trajtenberg, and R. Henderson. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics* 108, 3: (1993) 577.

Kalnins, Arturs, and Francine Lafontaine. "Too Far Away? The Effect of Distance to Headquarters on Business Establishment Performance." *American Economic Journal: Microeconomics* 5, 3: (2013) 157–179.

Keller, Wolfgang. "Geographic localization of international technology diffusion." *American Economic Review* 92, 1: (2002) 120–142.

———. "International Technology Diffusion." *Journal of Economic Literature* XLII, September: (2004) 752–782.

Keller, Wolfgang, and Stephen Ross Yeaple. "The Gravity of Knowledge." *American Economic Review* 103, 4: (2013) 1414–1444.

Kerr, W.R. "Ethnic scientific communities and international technology diffusion." *The Review of Economics and Statistics* 90, 3: (2008) 518–537.

Kogut, B, and U Zander. "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology." *Organization science* 3, 3: (1992) 383–397.

Lipsey, RE. "The Creation of Microdata Sets for Enterprises and Establishments." *Annales de l'INSEE* , 30: (1978) 395–422.

Markusen, James R. "Multinationals, multi-plant economies, and the gains from trade." *Journal of international economics* 16, 1984: (1984) 205–226.

Markusen, James R., and Keith E. Maskus. “Discriminating among alternative theories of the multinational enterprise.” *Review of International Economics* 10, 4: (2002) 694–707.

Markusen, JR. “Trade versus Investment Liberalization.” *NBER Working Paper Series* , 6231.

Markusen, JR, AJ Venables, DE Konan, and KH Zhang. “A unified treatment of horizontal direct investment, vertical direct investment, and the pattern of trade in goods and services.” *NBER Working Paper Series* , 5696.

Mayer, Thierry, and Soledad Zignago. “Notes on CEPII distances measures : The GeoDist database.” *CEPII Working Paper* , 25.

Nunn, Nathan, and D Trefler. “The boundaries of the multinational firm: an empirical analysis.” In *The Organization of Firms in a Global Economy*, edited by E Helpman, D Marin, and T Verdier, Cambridge, MA: Harvard University Press, 2008, 55–83.

Pierce, JR, and PK Schott. “A concordance between ten-digit US Harmonized System Codes and SIC/NAICS product classes and industries.” *Journal of Economic and Social Measurement* , 301.

Polanyi, M. *The Tacit Dimension.* Chicago; London: University of Chicago Press, 1966, 2009 edition.

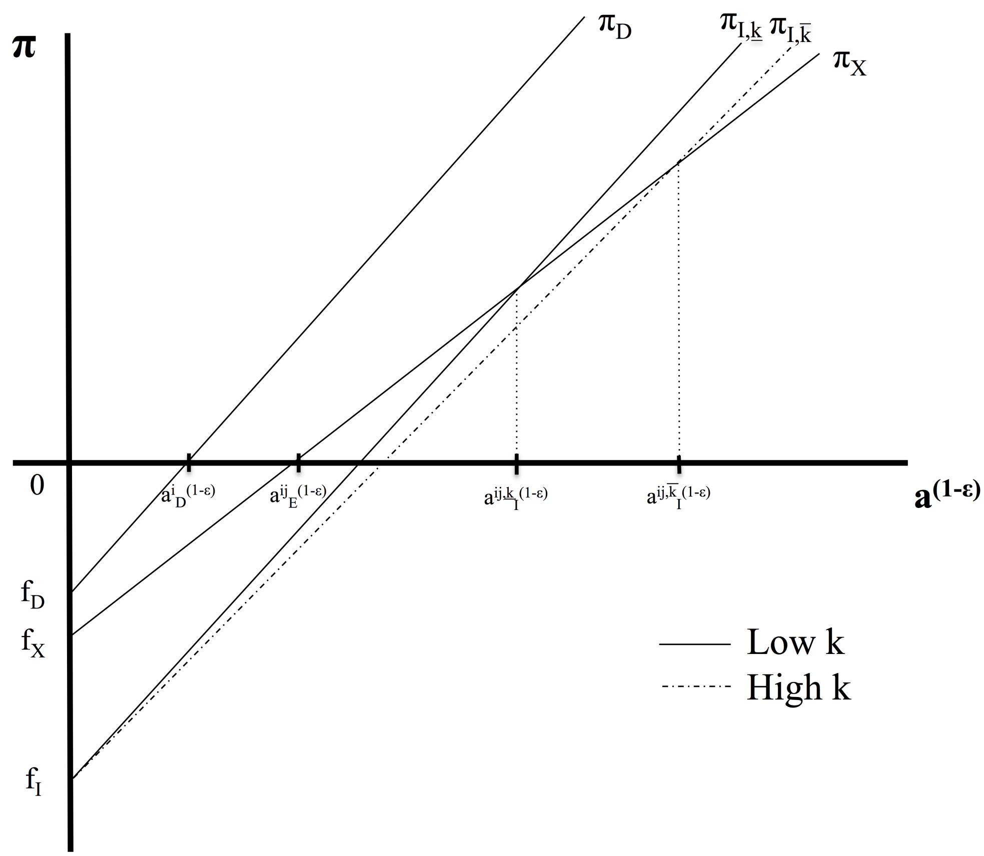
Pritchett, Lant. “Divergence, big time.” *The Journal of Economic Perspectives* 11, 3: (1997) 3–17.

Stein, Ernesto, and Christian Daude. “Longitude matters: Time zones and the location of foreign direct investment.” *Journal of International Economics* 71, 1: (2007) 96–112.

Szulanski, G. *Sticky Knowledge: Barriers to Knowing in the Firm.* SAGE Strategy series. SAGE Publications, 2002.

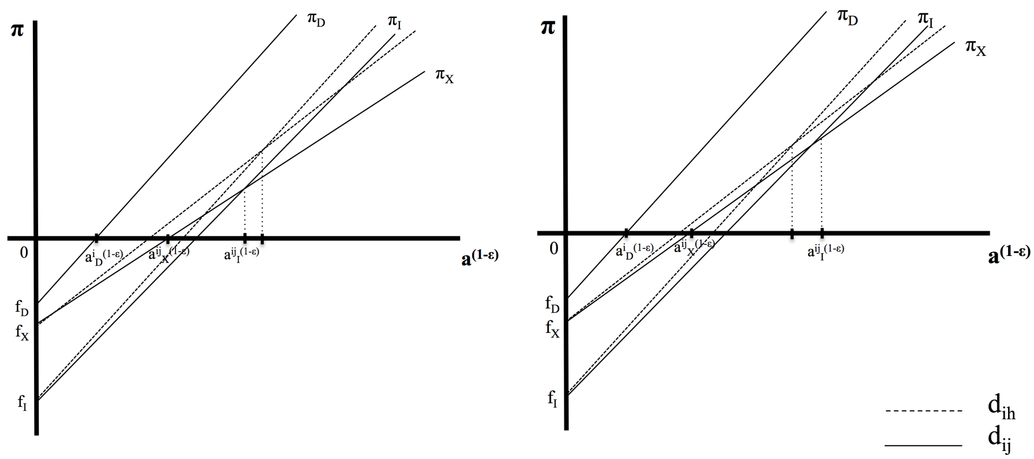
Szulanski, Gabriel. "Exploring Internal Stickiness: Impediments to the Transfer of Best Practice Within the Firm." *Strategic management journal* 17: (1996) 27–43.

Figure 1: Increase in k (knowledge intensity)



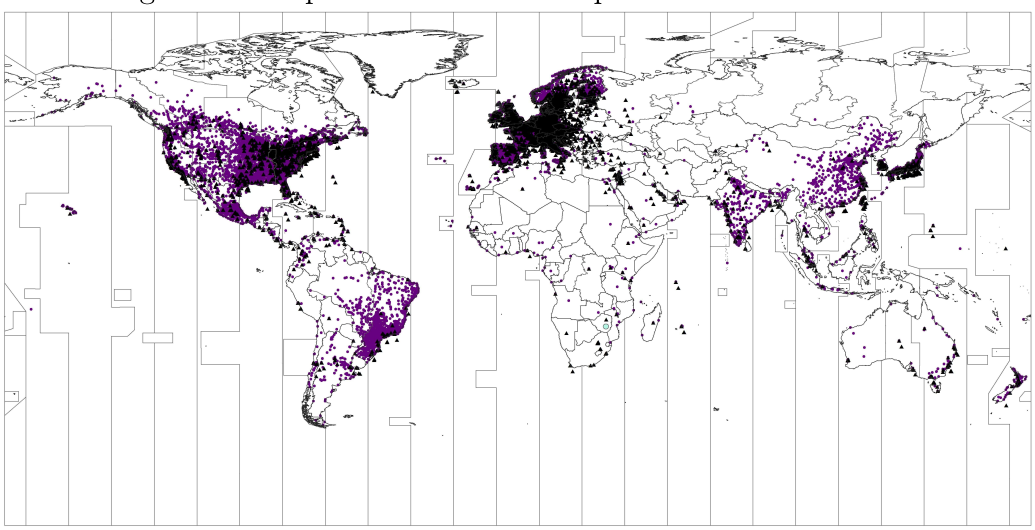
Graphical representation of the model, for a case considering two sectors with different levels of k, where $\overline{k} > \underline{k}$. The result suggests that the threshold a_I is an increasing function of k. Thus, FDI will be less likely for sectors with higher k.

Figure 2: Increase in d (distance)



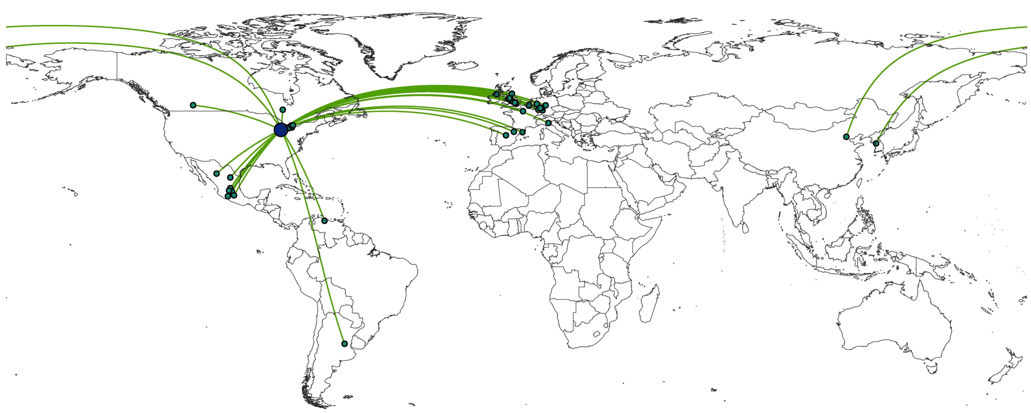
Graphical representation of the model, for a case considering a firm serving two foreign markets h and j, where $d_{ij} > d_{ih}$. The left panel shows the case where the threshold a_I is a decreasing function of d, while the right panel shows the case where the threshold a_I is an increasing function of d. The case in the left panel assumes that π_E is more elastic to changes in distance than π_I, while the case in the right panel assumes otherwise.

Figure 3: Unique locations of headquarters and subsidiaries



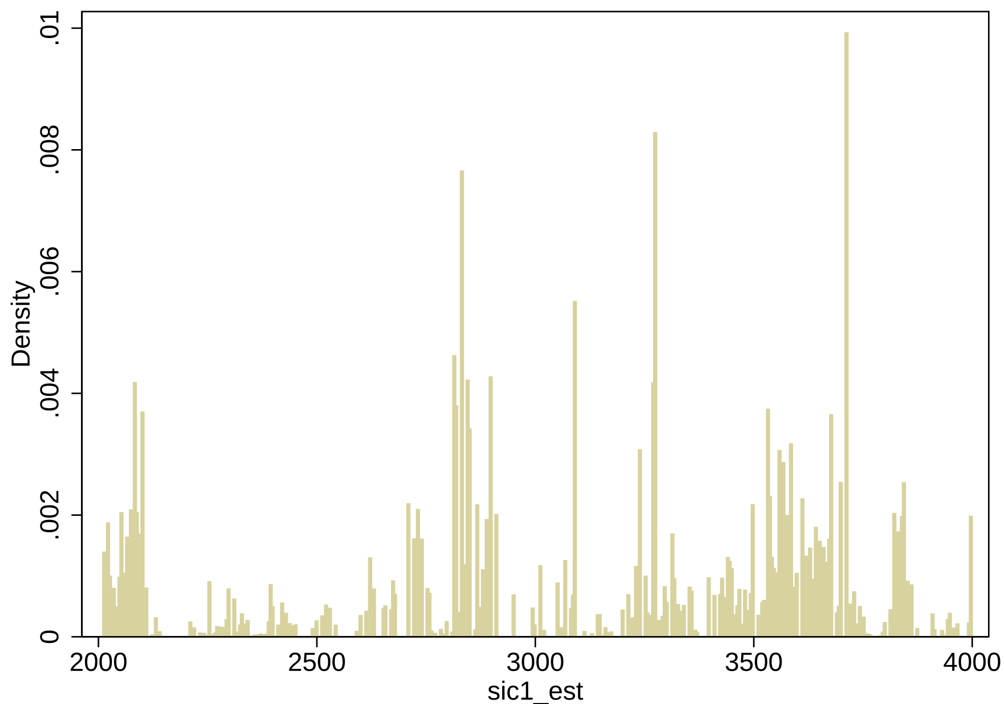
The figure shows a World map with the geocoded location of all the headquarters (triangles) and foreign subsidiaries (dots) in the sample.

Figure 4: Headquarters and foreign subsidiaries of an American MNC



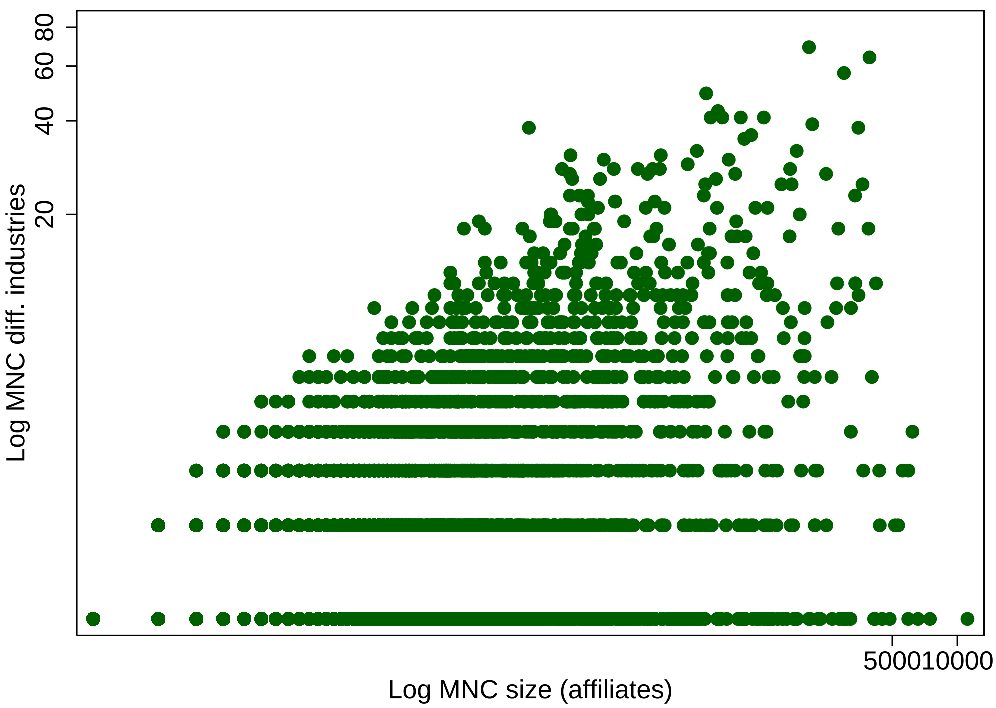
The figure is an example of the resolution of the data. It shows a World map with the geocoded location of the headquarters of an American car manufacturing firm and all of its subsidiaries.

Figure 5: Histogram of SIC codes in the sample



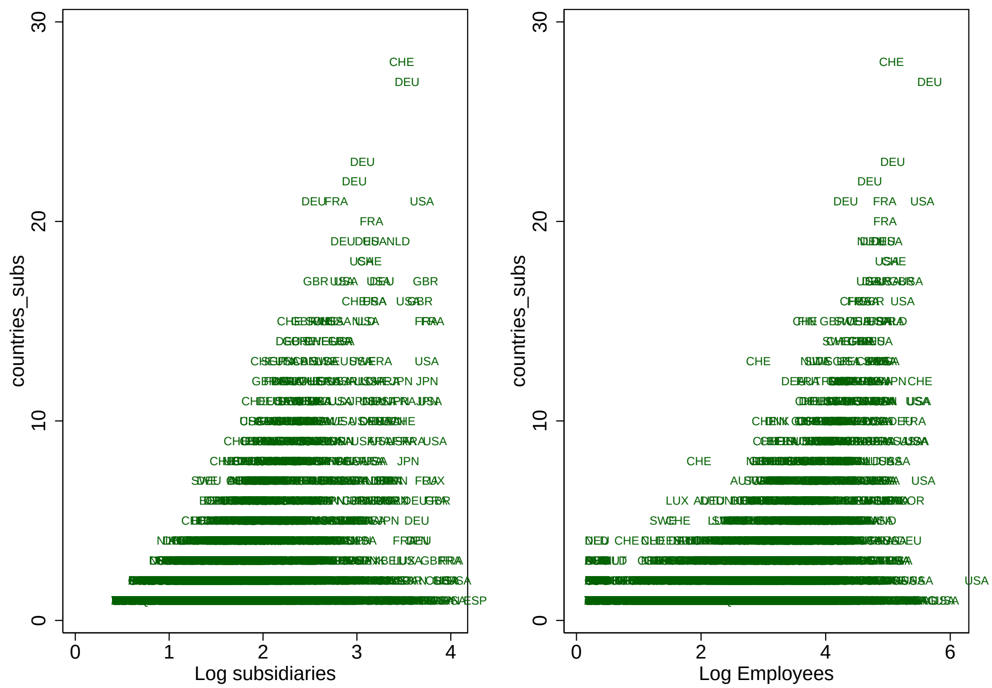
The figure is an histogram of the SIC industries reported in the dataset. Each bin represents the frequency of a particular SIC code within the manufacturing sector. Notice that the SIC classification is not fully continuos, what explains the zero values in the figure.

Figure 6: Number of different industries Vs. MNC size



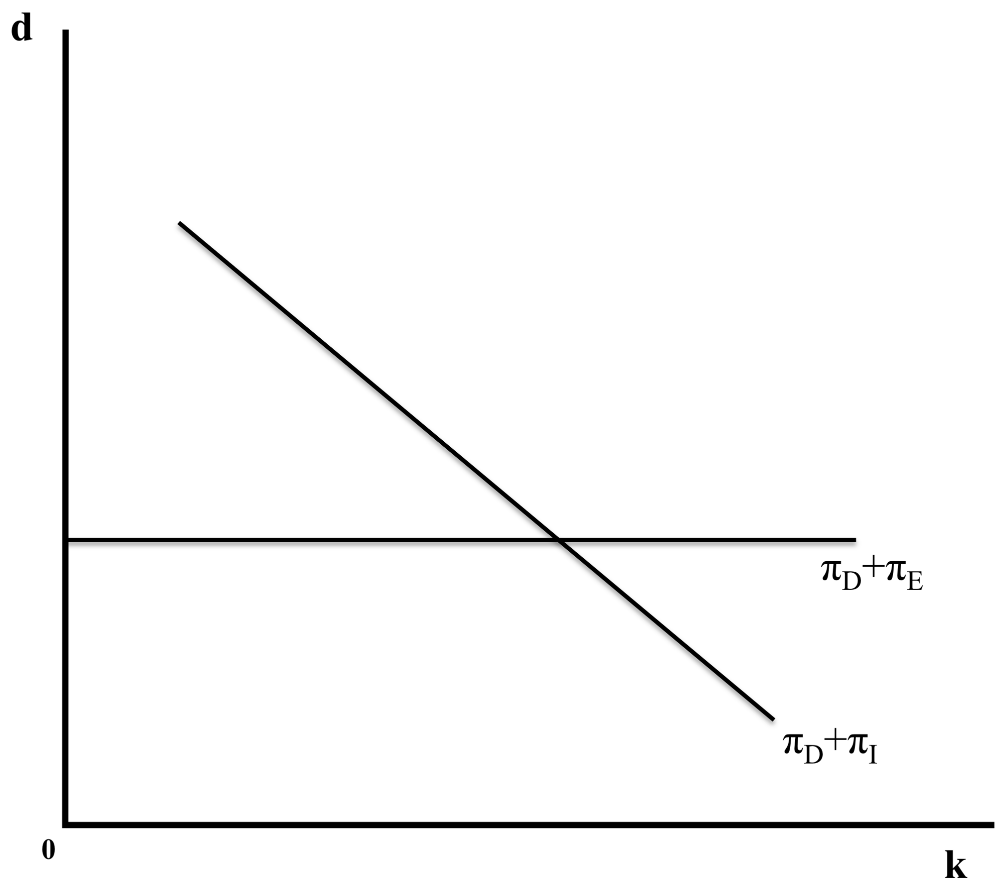
The figure plots the relationship between MNC size and total number of (different) industries the MNC is active in through its foreign affiliates. The figure reveals that larger MNCs (measured in terms of number of subsidiaries) tend to make a larger number of different products.

Figure 7: MNC size vs. number of countries



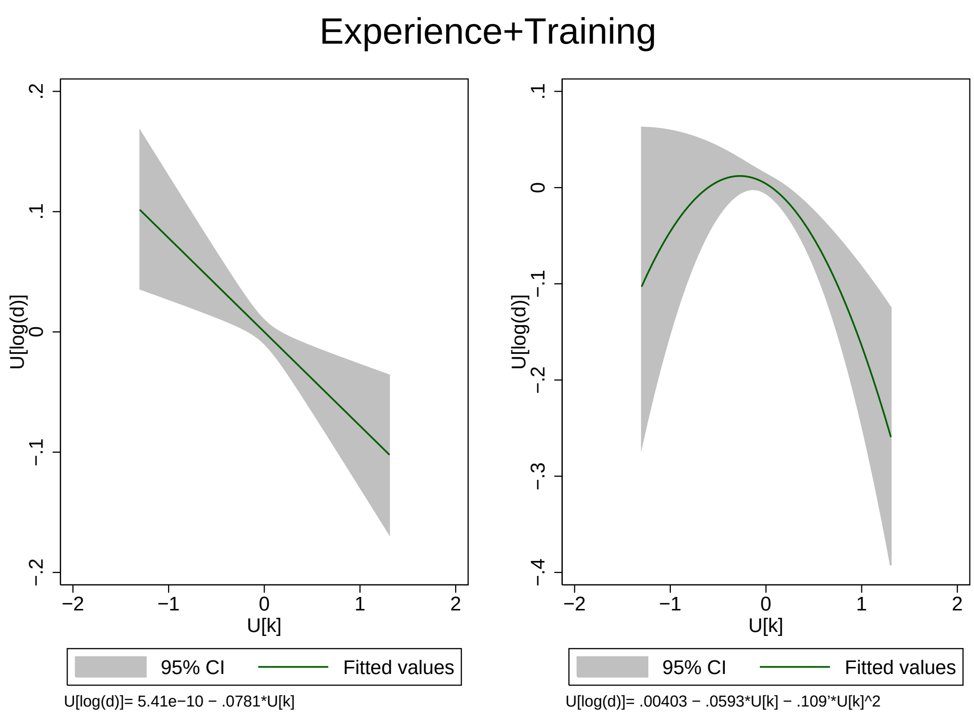
The figure shows the relationship between the size of MNC (horizontal axis) and the number of foreign countries they are active in (vertical axis). In the scatterplots, each observation is an MNC, labeled with the ISO3 code of the country where its headquarters is located. The left panel measures the firms' size by the total number of subsidiaries it has (both domestic and foreign), while the right panel uses the total employees (both in domestic and foreign plants).

Figure 8: Profit curves, in the k and d dimension



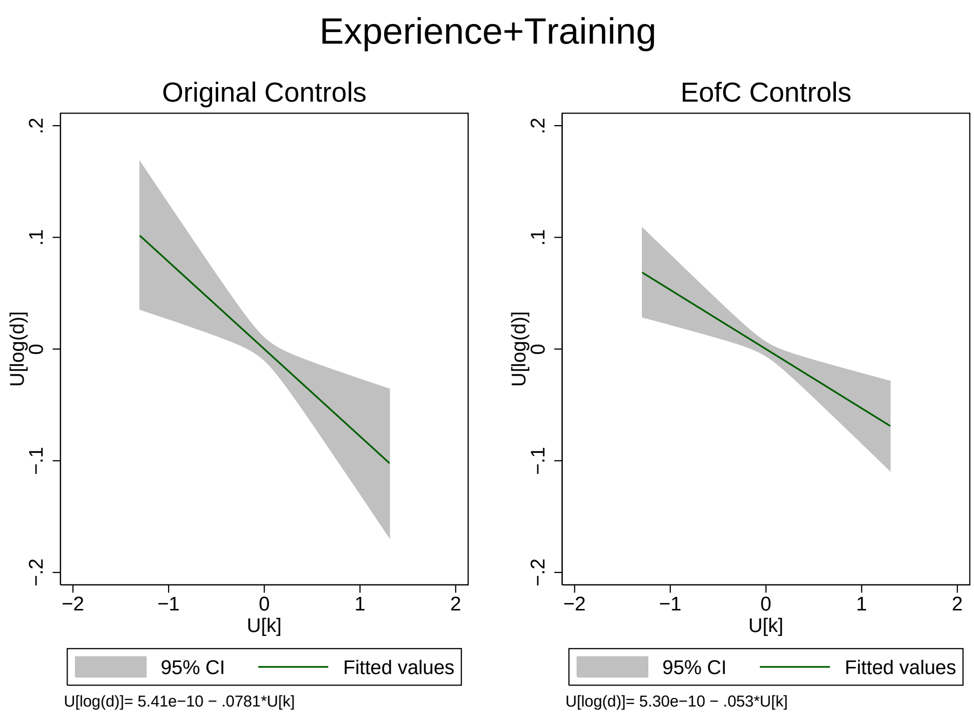
The figure is a graphical representation of a firm's profit as a function of k and d. The curve $\pi_D + \pi_E$ represents total profits for an exporting firm, while the curve $\pi_D + \pi_I$ represent total profits for a firm engaging in FDI instead.

Figure 9: Estimated relationship of $U[k]$ and $U[log(d)]$



The figure presents the empirical fit for the relationship between d and k (the latter proxied by the *experience plus training* measure). The left column performs a linear fit between k and d while the right column performs a quadratic firm between the two. The grey area represents the 95% confidence interval for the estimated relationship.

Figure 10: Estimated relationship of $U[k]$ and $U[log(d)]$



The figure presents the empirical fit for the relationship between d and k (the latter proxied by the *experience plus training* measure). The left column performs a linear fit between k and d using the original controls, while the right panel repeats the exercise adding the ease of communication variables as controls. The grey area represents the 95% confidence interval for the estimated relationship.

Table 1: KI Measures Correlations

Variables	Experience + Training	R&D share (N&T)	R&D share (K&Y)
Experience + Training	1.000		
R&D share (N&T)	0.354	1.000	
R&D share (K&Y)	0.420	0.682	1.000

The table shows the Pearson correlation coefficients between the O*NET based measures of knowledge intensity and R&D share in sales, used previously in the literature as proxies of knowledge intensity by Nunn and Trefler (2008) and Keller and Yeaple (2013).

Table 2: Descriptive Statistics (Domestic Vs. Foreign Subsidiaries)

	MNC	# Subs	Foreign (%)	$KI_{Foreign}$	$KI_{Domestic}$	Δ
All Observations	1540	64462	.29	.19	.33	-.14***
Non OECD	28	958	.12	.42	.34	.086*
OECD	1512	63504	.29	.19	.33	-.14***
East Asia & Pacific	306	17008	.11	.37	.46	-.087***
Latin America & Caribbean	18	1920	.58	-.38	-.3	-.083***
North America	508	24891	.2	.24	.33	-.089***
South Asia	15	370	.16	.17	.29	-.12*
Western Europe	693	20273	.51	.2	.18	.014

The table presents descriptive statistics from the sample. It presents for different cuts of the sample, based on the home country of the MNC, the total number of MNC firms, the number of subsidiaries, the proportion of those subsidiaries that are foreign (horizontal) subsidiaries, the average knowledge intensity of the foreign subsidiaries, the average knowledge intensity for the domestic subsidiaries, and the difference between these averages, denoted by Δ. Stars represent statistical significance of the difference: $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 3: Descriptive Statistics (Foreign Subsidiaries)

	MNC	# Subs	H %	Dist	WH	DF
All Observations	8266	60621	34	5152	7.3	.25
Non OECD	2590	10685	32	7697	7.5	.2
OECD	6520	49936	35	4608	7.3	.27
East Asia & Pacific	2074	5560	26	7329	6.1	.24
Eastern Europe	68	125	30	1738	9.2	.2
Latin America & Caribbean	981	7394	36	8453	7.5	.093
Middle East & N. Africa	67	93	40	7549	7.6	.19
North America	2246	16944	37	6698	6.1	.02
South Asia	410	2405	45	7877	6.4	.2
Sub-Saharan Africa	33	51	65	7832	7.7	.098
Western Europe	4969	28049	33	2686	8.3	.44

The table presents descriptive statistics from the sample. It presents for different cuts of the sample the total number of MNC, foreign subsidiaries (Sub), the percentage of subsidiaries classified as horizontal expansion (H%), the average distance in kilometers between subsidiaries and headquarters (Dist), the average number of overlapping working hours between the subsidiaries and the headquarters (WH) and the proportion of subsidiary-headquarter pairs that have a direct flight in between them (DF).

Table 4: Determinants of Foreign Replication of Production

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0359	-0.0230		-0.0348
	(0.017)**	(0.013)*		(0.014)**
log(t)	-0.0235		-0.0049	-0.0198
	(0.023)		(0.019)	(0.022)
GDP per capita ratio	-0.3952	-0.4008	-0.3943	0.3863
	(0.131)***	(0.128)***	(0.131)***	(0.131)***
Population ratio	0.0848	0.0866	0.0852	-0.0668
	(0.019)***	(0.019)***	(0.019)***	(0.028)**
Capital per worker ratio	0.3299	0.3326	0.3294	-0.2328
	(0.080)***	(0.078)***	(0.080)***	(0.069)***
Human Capital ratio	0.9534	0.9499	0.9547	0.0294
	(0.180)***	(0.176)***	(0.180)***	(0.072)
Land per worker ratio	-0.1029	-0.0994	-0.1030	0.1103
	(0.018)***	(0.018)***	(0.018)***	(0.045)**
Constant	0.2220	0.2614	0.2480	0.9500
	(0.046)***	(0.007)***	(0.040)***	(0.104)***
N	61410	64389	61410	61410
R-squared	0.52	0.51	0.52	0.56
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (7) using a sample of domestic and foreign subsidiaries that replicate home production. The left hand side variable is a binary variable that takes the value 1 if the subsidiary is foreign. The variables in the right hand side include the unit shipping cost associated with the industry, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parentheses.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 5: Determinants of Horizontal FDI

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0877	-0.0872		-0.0898
	(0.043)**	(0.043)**		(0.043)**
log(d)	-0.0242		-0.0239	-0.0230
	(0.009)**		(0.009)**	(0.010)**
log(t)	0.0233	0.0239	0.0611	0.0229
	(0.066)	(0.066)	(0.056)	(0.065)
GDP per capita ratio	0.1297	0.1311	0.1282	0.9039
	(0.056)**	(0.057)**	(0.056)**	(0.243)***
Population ratio	0.0128	0.0210	0.0142	0.3127
	(0.007)*	(0.007)***	(0.007)**	(0.077)***
Capital per worker ratio	-0.0833	-0.1005	-0.0792	-0.6599
	(0.045)*	(0.047)**	(0.046)*	(0.182)***
Human Capital ratio	-0.0052	0.0370	-0.0098	-0.4287
	(0.063)	(0.063)	(0.064)	(0.282)
Land per worker ratio	-0.0131	-0.0106	-0.0131	0.0124
	(0.007)*	(0.008)	(0.007)*	(0.054)
Constant	0.6264	0.4426	0.6745	0.7577
	(0.161)***	(0.135)***	(0.148)***	(0.226)***
N	55136	55137	55136	55136
R-squared	0.47	0.47	0.47	0.47
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as a horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary (in logs), the unit shipping cost (in logs), knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parentheses.

${}^{*}p < 0.10, {}^{**}p < 0.05, {}^{***}p < 0.01$

Table 6: Determinants of Horizontal FDI, Ease of Communication

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0877	-0.0878	-0.0868	-0.0875
	(0.043)**	(0.043)**	(0.043)**	(0.043)**
log(d)	-0.0242	-0.0254	-0.0076	-0.0187
	(0.009)**	(0.009)***	(0.014)	(0.008)**
log(t)	0.0233	0.0232	0.0239	0.0233
	(0.066)	(0.066)	(0.066)	(0.066)
Non-stop Flight		-0.0084		
		(0.010)		
Working hours overlap			0.0091	
			(0.005)*	
Common Language				0.0497
				(0.024)**
GDP per capita ratio	0.1297	0.1302	0.1271	0.1279
	(0.056)**	(0.056)**	(0.056)**	(0.055)**
Population ratio	0.0128	0.0131	0.0123	0.0139
	(0.007)*	(0.007)*	(0.007)*	(0.007)**
Capital per worker ratio	-0.0833	-0.0841	-0.0892	-0.0800
	(0.045)*	(0.046)*	(0.044)**	(0.044)*
Human Capital ratio	-0.0052	-0.0023	-0.0033	-0.0107
	(0.063)	(0.062)	(0.062)	(0.064)
Land per worker ratio	-0.0131	-0.0124	-0.0079	-0.0075
	(0.007)*	(0.007)*	(0.007)	(0.008)
Constant	0.6264	0.6383	0.4316	0.5718
	(0.161)***	(0.160)***	(0.217)**	(0.158)***
N	55136	55136	55136	55132
R-squared	0.47	0.47	0.47	0.47
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	N

The table presents results for the estimation of Specification (8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as a horizontal expansion. The variables on the right hand side include the distance from the MNC headquarters to the foreign subsidiary (in logs), the unit shipping cost (in logs), knowledge intensity measures (in standard deviations from the mean), and other controls. The right hand side also includes variables measuring the ease of communication between a headquarters and its subsidiaries. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parentheses.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

A Appendix

A.1 Condition for ${}^{\partial\phi(a_I)}/_{\partial d} > 0$

Horizontal FDI will be less profitable at longer distances if ${}^{\partial\phi(a_I)}/_{\partial d} > 0$ (where $\phi(a_I) = a_I^{1-\varepsilon}$). For simplicity, I compute the conditions for which ${}^{\partial log(\phi(a_I))}/_{\partial d} > 0$ instead:

$$\frac{\varepsilon - 1}{\kappa(k,d)^{1-\varepsilon} - \tau(t,d)^{1-\varepsilon}} \cdot \left[\kappa(k,d)^{-\varepsilon}\frac{\partial\kappa}{\partial d} - \tau(t,d)^{-\varepsilon}\frac{\partial\tau}{\partial d}\right] > 0$$

Given that the left term will always be positive (given the assumption that $\tau(t,d) > \kappa(k,d)$ and $\varepsilon > 1$, the conditions for the inequality to hold are derived from the right term only. Hence, we have:

$$\begin{aligned} \kappa(k,d)^{-\varepsilon} \cdot \frac{\partial\kappa}{\partial d} &> \tau(t,d)^{-\varepsilon} \cdot \frac{\partial\tau}{\partial d} \\ \frac{1}{\kappa(k,d)^{\varepsilon-1}} \cdot \epsilon_{\kappa,d} &> \frac{1}{\tau(t,d)^{1-\varepsilon}} \cdot \epsilon_{\tau,d} \end{aligned}$$

Where $\epsilon_{\kappa,d}$ and $\epsilon_{\tau,d}$ are the elasticity of κ and τ with respect to distance d, respectively.

Hence for the condition to hold it must be that:

$$\left[\frac{\tau(t,d)}{\kappa(k,d)}\right]^{\varepsilon-1} > \frac{\epsilon_{\tau,d}}{\epsilon_{\kappa,d}}$$

There is no reason to believe that this condition is not economically feasible.

A.2 Heterogeneity in the number of reported SIC industries in the dataset

While the dataset has information on up to six industries per plant (a main one plus five other) the number of establishments that report more than one activity varies dramatically per country. The left panel of Figure A1 shows the average number of reported industries across all subsidiaries per country, while the right panel shows, per country, the percentage of firms reporting one, two, three, four, five or six industries. In most countries, the average number of reported firms is below two; and the majority of firms in more than half the countries report only one SIC code.

[Figure A1 about here.]

A.3 Using the input-output table to define vertical relationships

In order to filter out from the definition of horizontal those links that could also be defined as vertical, either upstream or downstream, I use the US input-output provided by Fan and Lang (2000). I follow the methodology suggested by Alfaro and Charlton (2009) and Acemoglu et. al. (2009) to define vertical relationships.

More in general, the diagram in Figure A2 is useful to understand how horizontal and vertical links are defined in the dataset. Within a single MNC firm, an horizontal link is defined as a foreign subsidiary that is classified under the same SIC code as any of its domestic subsidiaries. Then I use the US I/O table by Fan & Lang (2000) to define vertical relationships, both downstream and upstream. A subsidiary is defined as upstream vertical if its main economic activity is an input of $0.05 or more per each dollar of output of any of the domestic subsidiaries of the firm. Similarly, a subsidiary is defined as downstream vertical if any of the domestic subsidiary provides an input to it of $0.05 or more per each dollar of output.

After such classification, those subsidiaries that fall into both categories (horizontal and vertical) are filtered out from the horizontal classification. This implies that the sample classifies as horizontal only final goods, which is the matter of study of the theoretical framework presented.

[Figure A2 about here.]

Appendix Section A.6.1 presents robustness tests of all tables using alternative thresholds (0.01 and 0.10). The use of $0.05 in the main body of the paper follows the precedent set by Alfaro & Charlton (2009).

A limitation of this methodology is that technologies might vary across countries, and hence, the US I/O table would loss some validity in defining upstream or downstream relationships. While acknowledging this limitation I assume that the US I/O table is a good proxy for measuring vertical links, regardless of the country, in line with the previous literature.

A.4 Limitations of the R&D intensity measures

Nunn & Trefler (2008) and Keller & Yeaple (2013) use the average R&D share of firms' sales as their measure of knowledge intensity. Nunn & Trefler use firm-level data from Orbis, while Keller & Yeaple use data from COMPUSTAT.

The two measures are skewed towards the few industries with large R&D investment, while the zeros or very small values are highly abundant (see Figure A3). In fact, for Nunn & Trefler half of the industries have an R&D intensity measure below 0.2%, while the largest industries have a value of 190%. In Keller & Yeaple's measure the median is 0.7% while the most knowledge intensive industry has a share of R&D over sales of over 1000%.

[Figure A3 about here.]

A.5 O*NET knowledge intensity measures

Figure A4 presents the distribution of the knowledge intensity measure used in the paper: experience plus training (based on experience plus on-site and on-the-job training figures of workers in each industry). As opposed to the R&D investment based variables used in the literature (see Section A.4), the distribution of the O*NET based variables is smoother, and behaves more like a normal probability density function. Figure A5 presents the same graphs limiting the sample to manufacturing industries only.

[Figure A4 about here.]

[Figure A5 about here.]

Tables A1 presents the top and bottom ten products in the manufacturing division (SIC codes 2000 to 3999) ranked by the knowledge intensity measure.

[Table A1 about here.]

A.6 Robustness Tests

A.6.1 Varying thresholds in the definition of horizontal subsidiaries

As explained in Section A.3, subsidiaries that classify both as horizontal and vertical (according to the I/O table) are not considered horizontal. The intuition for such approach is to limit the analysis of horizontal to final goods only.

To do so, a threshold of $0.05 per each $ of output, was selected in order to define vertical relationships. This section presents the robustness test varying such threshold, for all tables in the main body of the paper.

Tables A2-A4 replicate all results using threshold 0.01, while tables A5-A7 replicate all results using the threshold 0.1.

Varying the input-output threshold is robust to the results presented in the main body of the paper.

[Table A2 about here.]

[Table A3 about here.]

[Table A4 about here.]

[Table A5 about here.]

[Table A6 about here.]

[Table A7 about here.]

A.6.2 Additional measures of knowledge intensity

In the main body of the paper I perform the analysis using one constructed measure of knowledge intensity denominated *experience plus training.* In this section I use instead a modification of such measure which only takes into account the accumulated experience of the workers in the industry (excluding the on-site and on-the-job training component). The results are robust to this other measure as can be seen in Tables A8, A9 and A10, as well as in Figure A6.

[Table A8 about here.]

[Table A9 about here.]

[Figure A6 about here.]

[Table A10 about here.]

A.6.3 Non-linear effects of distance

Is the negative relationship between distance and the likelihood of a foreign subsidiary being horizontal linear? I test for that substituting in the estimation of specification 8 the continuos measure of distance ($log(d)$) by a set of dummies, each one representing a 500km interval in the distance between the headquarters and the foreign subsidiary. The results are presented in table A11. As it can be seen, the negative correlation becomes larger in magnitude the further away the headquarters is from the location of the foreign subsidiary.

The results suggest that up to 8000Km the correlation between distance and the existence of an horizontal foreign subsidiary is negative and increasing in magnitude (besides the 3000-4000km bucket, which present a positive, though non-statistically significant coefficient). Only after 8000 km the coefficients are reduced in terms of magnitude, while still negative.

Figure A7 looks at the non-linearity of the distance effect. The Figure reflects a monotonically decreasing relationship between distance and the likelihood of horizontal foreign subsidiaries, in general, up to 8000Km. This is consistent with the linear fit shown in the main body of the paper. Given the standard errors, however, there is little we can say about a U-shaped non-linear form. Yet, for longer distances, the coefficients are strictly negative. It is important to note that after 8000Km there are considerably less observations in each one of those buckets.

[Table A11 about here.]

[Figure A7 about here.]

A.6.4 Intellectual Property Rights: Excluding China

To alleviate concerns that the results are driven by the lack of intellectual property rights in China, I replicate Table 5 excluding China from the sample.

When excluding China from the sample, however, the results are robust to the ones presented in the main body of the paper, as can be seen in Table A12.

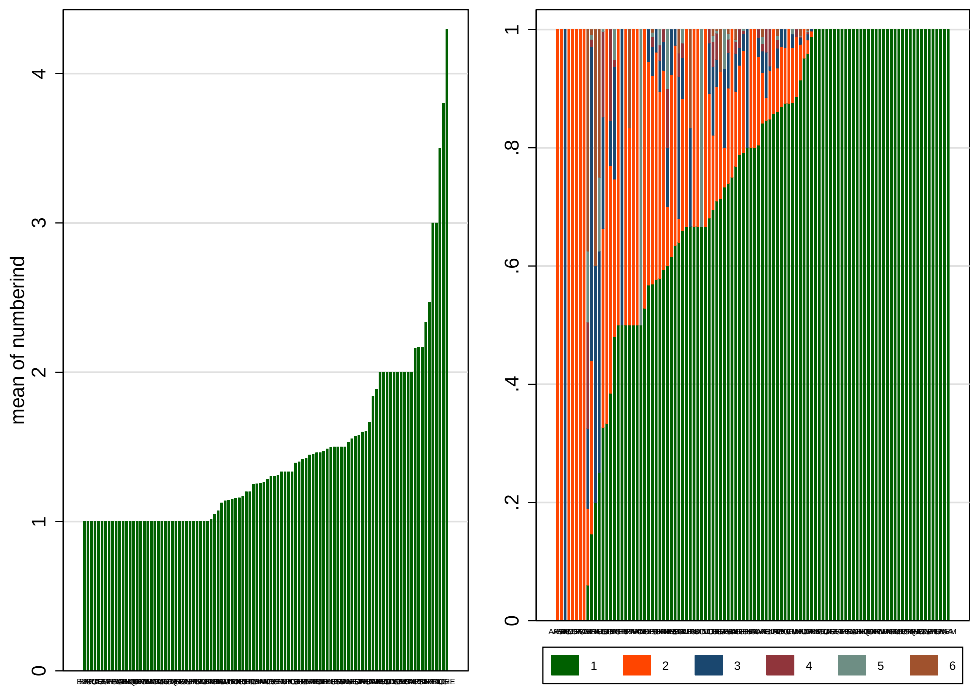
[Table A12 about here.]

A.6.5 Excluding European Firms

Given the large number of European firms in the sample, and the short distances in the continent, this raises concerns about the validity of the analysis in terms of the tradeoff MNC face in locating their knowledge intensive subsidiaries at shorter distances. Hence, I repeat the corresponding analysis excluding all foreign subsidiaries located in Western Europe that belong to a European MNC (i.e., for which its headquarters is located in Western Europe). The results can be seen in in Figure A8. As can be seen, the results are robust to the exclusion of these observations from the sample.

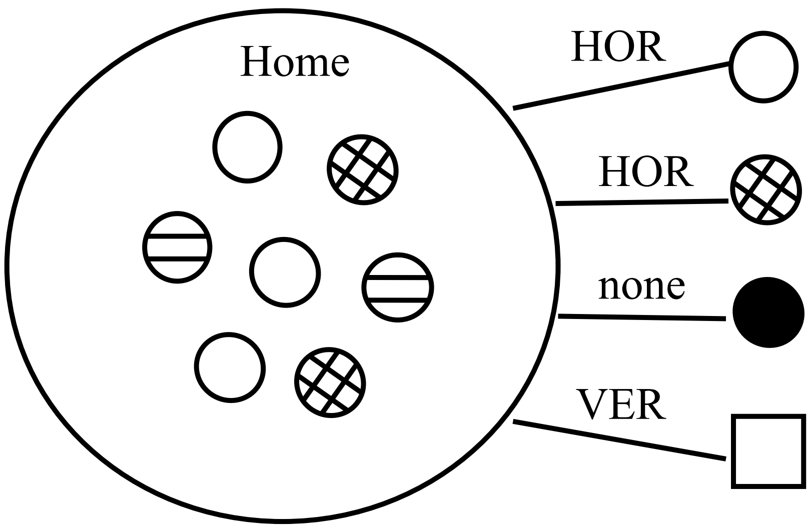
[Figure A8 about here.]

Figure A1: Distribution of reported SIC codes by plant, per country



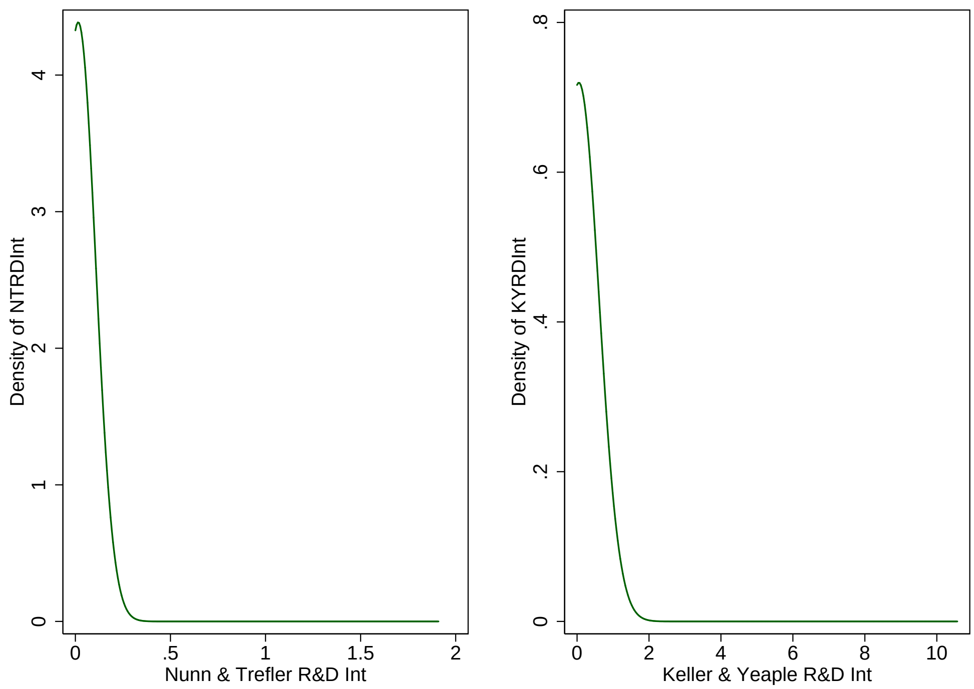
The figure describe the distribution of number of industries reported by establishment in the sample. The left panel shows the average number of reported industries across all subsidiaries per country, while the right panel shows, per country, the percentage of firms reporting one, two, three, four, five or six industries.

Figure A2: Definition of Horizontal and Vertical



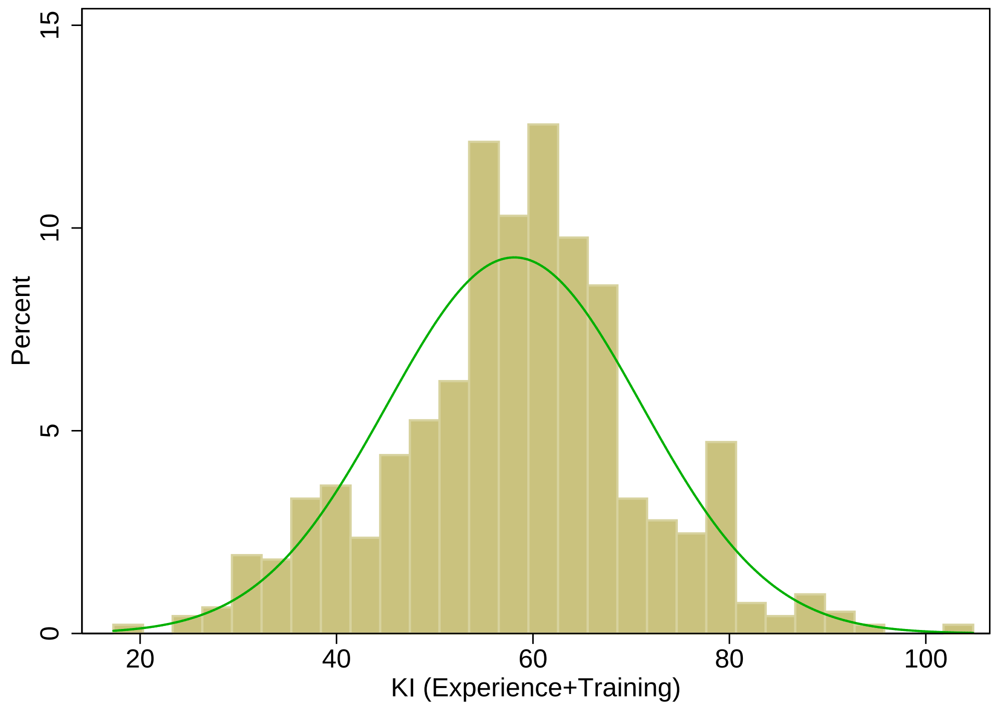
The diagram describes the methodology used to classify foreign subsidiaries as horizontal expansions based on their reported economic activity vis-a-vis the economic activity of the MNC in its home country.

Figure A3: Fitted distribution of R&D Measures



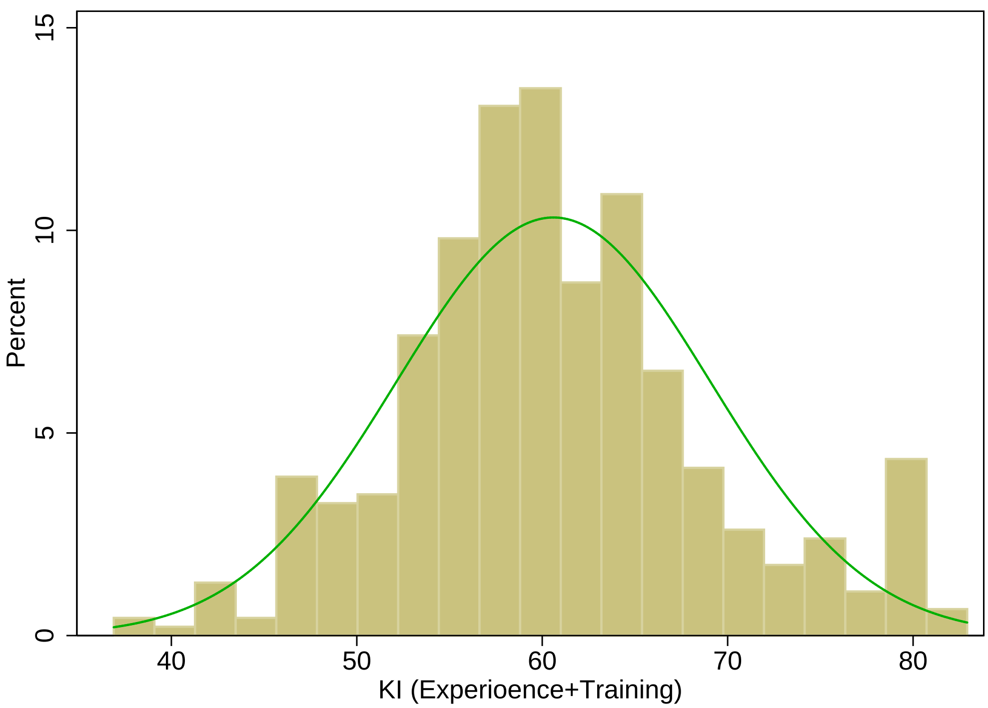
The figure shows the fitted distribution for the industry level R&D investment as share of sales, compiled from firm level datasets by Nunn & Trefler (2008) and Keller & Yeaple (2013) in the left and right panel respectively.

Figure A4: Histogram O*NET-based KI (All Industries)



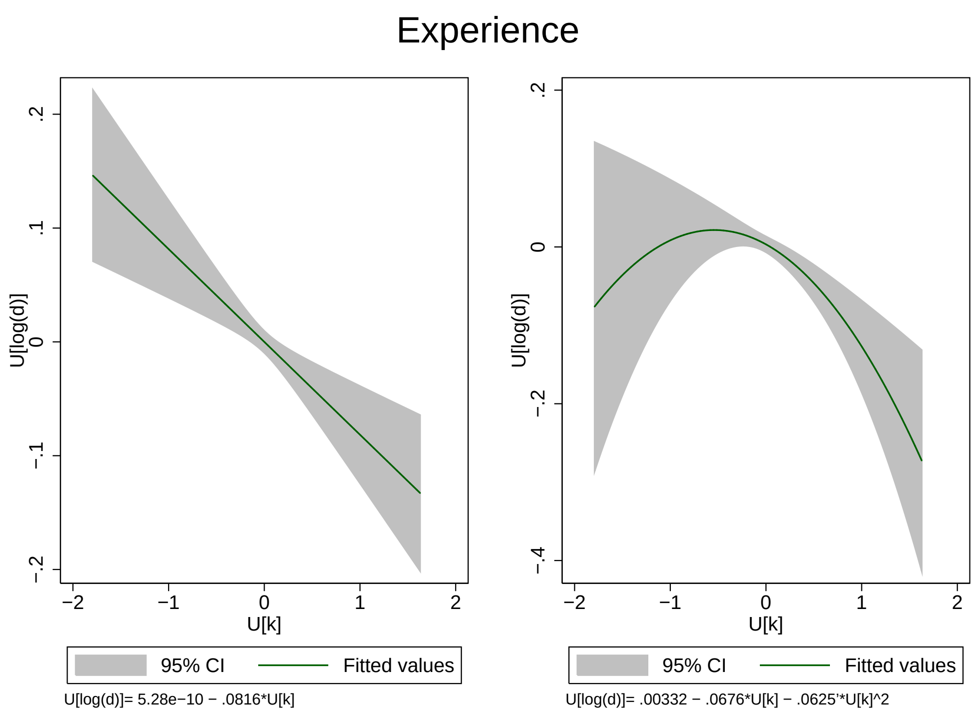
The figure shows the fitted distribution for the computed "experience plus training" O*NET-based knowledge intensity measures for all industries. Industries are defined in SIC 1987 4-digit industries.

Figure A5: Histogram O*NET-based KI (Manufacturing Only)



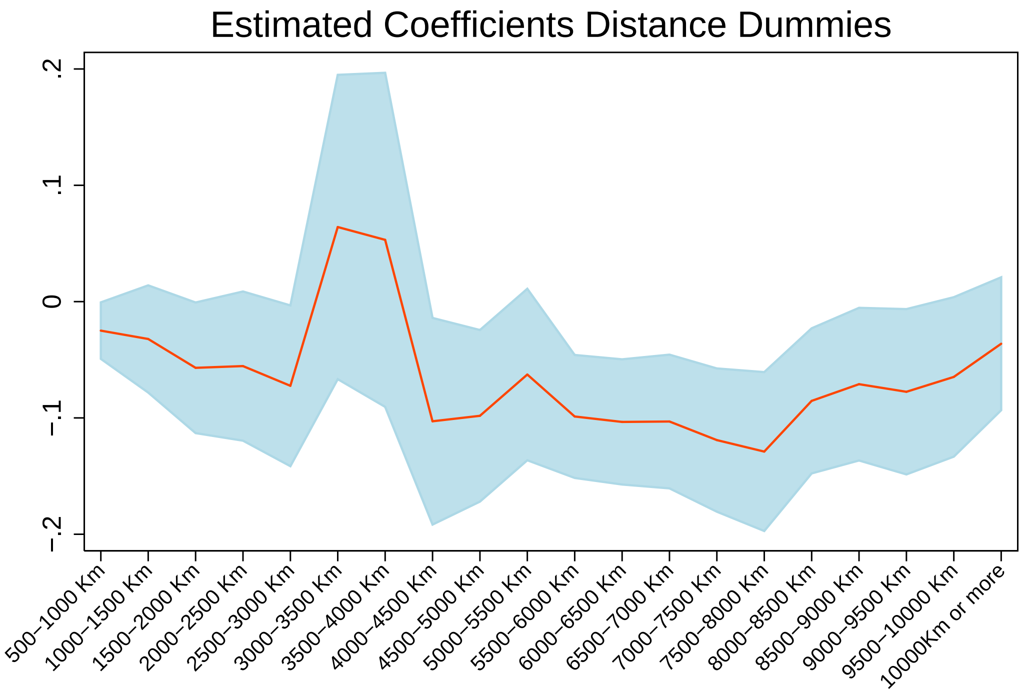
The figure shows the fitted distribution for the computed "experience plus training" O*NET-based knowledge intensity measures for manufacturing industries only. Industries are defined in SIC 1987 4-digit industries.

Figure A6: $U[k]$ vs. $U[log(d)]$, (KI: experience)



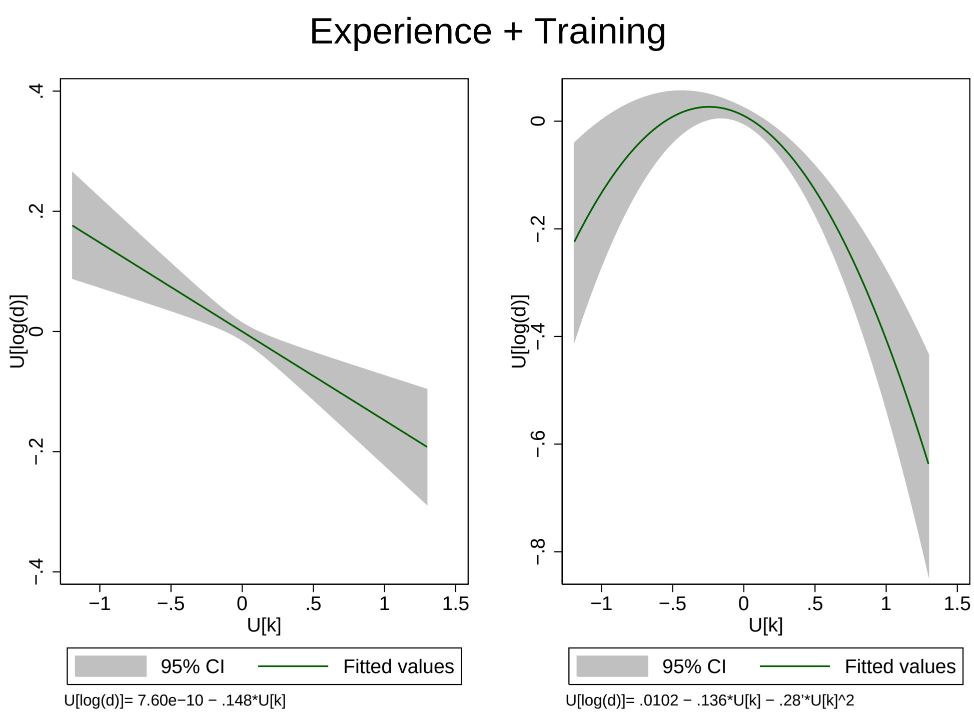
The figure presents the empirical fit for the relationship between d and k (the latter proxied by the *experience* measure). The left column performs a linear fit between k and d while the right column performs a quadratic firm between the two. The grey area represents the 95% confidence interval for the estimated relationship. The sample excludes foreign subsidiaries located in Europe owned by a European firm.

Figure A7: Distance Intervals Estimators



The figure presents the empirical estimation for the distance intervals coefficients from Table A11. The grey area represents 95% confidence intervals.

Figure A8: $U[k]$ vs. $U[log(d)]$, excluding Europe



The figure presents the empirical fit for the relationship between d and k (the latter proxied by the *experience plus training* measure). The left column performs a linear fit between k and d while the right column performs a quadratic firm between the two. The grey area represents the 95% confidence interval for the estimated relationship. The sample excludes foreign subsidiaries located in Europe owned by a European firm.

Table A1: Top and bottom 10 manufacturing products, ranked by KI

Rank	**SIC**	**Name**	**Value**
		Ranking by Experience + Training, Top 10	
1	3669	Communications Equipment, NEC	82.92
2	3663	Radio and Television Broadcasting and Communications Equipment	82.92
3	3661	Telephone and Telegraph Apparatus (except consumer external modems)	81.45
4	3677	Electronic Coils, Transformers, and Other Inductors	79.97
5	3676	Electronic Resistors	79.97
6	3678	Electronic Connectors	79.97
7	3675	Electronic Capacitors	79.97
8	3671	Electron Tubes	79.97
9	3672	Printed Circuit Boards	79.97
10	3674	Semiconductors and Related Devices	79.97
		Ranking by Experience + Training, Bottom 10	
459	2013	Sausages and Other Prepared Meat Products (except lard made from purchased materials)	36.89
458	2011	Meat Packing Plants	36.89
457	2411	Logging	39.79
456	2077	Animal and Marine Fats and Oils (animal fats and oils)	41.39
455	2053	Frozen Bakery Products, Except Bread	41.53
454	2045	Prepared Flour Mixes and Doughs	41.53
453	2098	Macaroni, Spaghetti, Vermicelli and Noodles	41.53
452	2051	Bread and Other Bakery Products, Except Cookies and Crackers	41.53
451	2015	Poultry Slaughtering and Processing (poultry slaughtering and processing)	42.72
450	2052	Cookies and Crackers (unleavened bread and soft pretzels)	45.04

The table presents the top and bottom 10 manufacturing sectors ranked by the "experience plus training" O*NET based knowledge intensity measure.

Table A2: Determinants of Foreign Replication of Production (threshold 0.01)

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0416	-0.0162		-0.0405
	(0.016)**	(0.012)		(0.013)***
log(t)	-0.0553		-0.0321	-0.0512
	(0.020)***		(0.017)*	(0.017)***
GDP per capita ratio	-0.4123	-0.3943	-0.4115	-0.0361
	(0.177)**	(0.173)**	(0.178)**	(0.061)
Population ratio	0.0621	0.0635	0.0627	0.0641
	(0.025)**	(0.024)***	(0.025)**	(0.016)***
Capital per worker ratio	0.2818	0.2769	0.2813	0.0482
	(0.076)***	(0.075)***	(0.076)***	(0.035)
Human Capital ratio	1.1475	1.1184	1.1484	
	(0.237)***	(0.230)***	(0.237)***	
Land per worker ratio	-0.1017	-0.0985	-0.1023	0.1951
	(0.022)***	(0.022)***	(0.022)***	(0.039)***
o.Human Capital ratio				0.0000
				(.)
Constant	0.1114	0.2133	0.1453	0.7139
	(0.038)***	(0.007)***	(0.035)***	(0.070)***
N	47657	50096	47657	47657
R-squared	0.52	0.52	0.52	0.57
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (7) using a sample of domestic and foreign subsidiaries that replicate home production. The left hand side variable is a binary variable that takes the value 1 if the subsidiary is foreign. The variables in the right hand side include the unit shipping cost associated with the industry, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parentheses.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A3: Determinants of Horizontal FDI (threshold 0.01)

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0859	-0.0855		-0.0870
	(0.030)***	(0.030)***		(0.030)***
log(d)	-0.0202		-0.0199	-0.0190
	(0.006)***		(0.006)***	(0.006)***
log(t)	-0.0308	-0.0302	0.0063	-0.0306
	(0.043)	(0.043)	(0.037)	(0.043)
GDP per capita ratio	0.0909	0.0920	0.0894	0.7052
	(0.050)*	(0.051)*	(0.050)*	(0.196)***
Population ratio	-0.0072	-0.0004	-0.0058	0.2853
	(0.005)	(0.005)	(0.006)	(0.072)***
Capital per worker ratio	-0.0733	-0.0876	-0.0693	-0.3497
	(0.040)*	(0.042)**	(0.041)*	(0.117)***
Human Capital ratio	-0.0150	0.0202	-0.0195	-0.7377
	(0.055)	(0.053)	(0.055)	(0.303)**
Land per worker ratio	-0.0177	-0.0157	-0.0178	0.0634
	(0.007)***	(0.007)**	(0.007)***	(0.035)*
Constant	0.3534	0.2002	0.4005	0.2565
	(0.092)***	(0.080)**	(0.085)***	(0.151)*
N	55136	55137	55136	55136
R-squared	0.48	0.48	0.48	0.49
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Table A4: Ease of Communication (threshold 0.01)

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0859	-0.0861	-0.0858	-0.0857
	(0.030)***	(0.030)***	(0.030)***	(0.030)***
log(d)	-0.0202	-0.0219	-0.0177	-0.0146
	(0.006)***	(0.006)***	(0.010)*	(0.005)***
log(t)	-0.0308	-0.0310	-0.0307	-0.0308
	(0.043)	(0.043)	(0.043)	(0.043)
Non-stop Flight		-0.0122		
		(0.009)		
Working hours overlap			0.0013	
			(0.004)	
Common Language				0.0500
				(0.024)**
GDP per capita ratio	0.0909	0.0916	0.0905	0.0890
	(0.050)*	(0.050)*	(0.051)*	(0.049)*
Population ratio	-0.0072	-0.0068	-0.0073	-0.0062
	(0.005)	(0.005)	(0.005)	(0.005)
Capital per worker ratio	-0.0733	-0.0745	-0.0742	-0.0700
	(0.040)*	(0.040)*	(0.039)*	(0.038)*
Human Capital ratio	-0.0150	-0.0108	-0.0147	-0.0205
	(0.055)	(0.054)	(0.055)	(0.055)
Land per worker ratio	-0.0177	-0.0167	-0.0170	-0.0121
	(0.007)***	(0.007)**	(0.006)***	(0.007)*
Constant	0.3534	0.3705	0.3249	0.2985
	(0.092)***	(0.093)***	(0.138)**	(0.086)***
N	55136	55136	55136	55132
R-squared	0.48	0.48	0.48	0.49
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	N

The table presents results for the estimation of Specification (8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures (in standard deviations from the mean) and other controls. The right hand side also includes variables measuring the ease of communication between a headquarters and its subsidiaries. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A5: Determinants of Foreign Replication of Production (threshold 0.1)

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0334	-0.0255		-0.0324
	(0.016)**	(0.012)**		(0.014)**
log(t)	-0.0111		0.0059	-0.0077
	(0.021)		(0.017)	(0.020)
GDP per capita ratio	-0.3745	-0.3818	-0.3740	0.4705
	(0.124)***	(0.122)***	(0.124)***	(0.124)***
Population ratio	0.0894	0.0914	0.0898	-0.0498
	(0.018)***	(0.017)***	(0.018)***	(0.025)**
Capital per worker ratio	0.3138	0.3202	0.3135	-0.2595
	(0.077)***	(0.075)***	(0.076)***	(0.066)***
Human Capital ratio	0.9593	0.9491	0.9613	-0.0210
	(0.167)***	(0.163)***	(0.167)***	(0.069)
Land per worker ratio	-0.0994	-0.0968	-0.0995	0.1379
	(0.017)***	(0.017)***	(0.017)***	(0.041)***
Constant	0.2536	0.2695	0.2772	0.8667
	(0.041)***	(0.007)***	(0.036)***	(0.099)***
N	65058	68207	65058	65058
R-squared	0.52	0.51	0.52	0.55
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (7) using a sample of domestic and foreign subsidiaries that replicate home production. The left hand side variable is a binary variable that takes the value 1 if the subsidiary is foreign. The variables in the right hand side include the unit shipping cost associated with the industry, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parentheses.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A6: Determinants of Horizontal FDI (threshold 0.1)

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0937	-0.0932		-0.0957
	(0.042)**	(0.042)**		(0.042)**
log(d)	-0.0244		-0.0241	-0.0216
	(0.009)***		(0.009)***	(0.010)**
log(t)	0.0489	0.0496	0.0893	0.0484
	(0.062)	(0.062)	(0.052)*	(0.061)
GDP per capita ratio	0.1267	0.1281	0.1250	0.9220
	(0.055)**	(0.056)**	(0.056)**	(0.246)***
Population ratio	0.0153	0.0236	0.0169	0.3181
	(0.007)**	(0.007)***	(0.007)**	(0.078)***
Capital per worker ratio	-0.0837	-0.1010	-0.0794	-0.6704
	(0.045)*	(0.046)**	(0.046)*	(0.187)***
Human Capital ratio	0.0068	0.0494	0.0019	-0.4480
	(0.062)	(0.062)	(0.063)	(0.285)
Land per worker ratio	-0.0169	-0.0144	-0.0169	0.0110
	(0.007)**	(0.007)**	(0.007)**	(0.055)
Constant	0.7121	0.5271	0.7636	0.8320
	(0.151)***	(0.127)***	(0.138)***	(0.222)***
N	55136	55137	55136	55136
R-squared	0.48	0.48	0.48	0.48
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A7: Ease of Communication (threshold 0.1)

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0937	-0.0938	-0.0928	-0.0935
	(0.042)**	(0.042)**	(0.042)**	(0.042)**
log(d)	-0.0244	-0.0253	-0.0077	-0.0197
	(0.009)***	(0.009)***	(0.014)	(0.008)**
log(t)	0.0489	0.0488	0.0495	0.0489
	(0.062)	(0.062)	(0.062)	(0.062)
Non-stop Flight		-0.0063		
		(0.010)		
Working hours overlap			0.0092	
			(0.005)**	
Common Language				0.0420
				(0.024)*
GDP per capita ratio	0.1267	0.1270	0.1240	0.1251
	(0.055)**	(0.056)**	(0.055)**	(0.054)**
Population ratio	0.0153	0.0155	0.0149	0.0162
	(0.007)**	(0.007)**	(0.007)**	(0.007)**
Capital per worker ratio	-0.0837	-0.0843	-0.0896	-0.0809
	(0.045)*	(0.045)*	(0.044)**	(0.044)*
Human Capital ratio	0.0068	0.0089	0.0087	0.0021
	(0.062)	(0.061)	(0.062)	(0.063)
Land per worker ratio	-0.0169	-0.0164	-0.0117	-0.0121
	(0.007)**	(0.007)**	(0.007)*	(0.007)*
Constant	0.7121	0.7210	0.5171	0.6661
	(0.151)***	(0.150)***	(0.206)**	(0.149)***
N	55136	55136	55136	55132
R-squared	0.48	0.48	0.48	0.48
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	N

The table presents results for the estimation of Specification (8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures (in standard deviations from the mean) and other controls. The right hand side also includes variables measuring the ease of communication between a headquarters and its subsidiaries. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A8: Determinants of Foreign Replication of Production (KI: experience)

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0300	-0.0186		-0.0291
	(0.014)**	(0.010)*		(0.012)**
log(t)	-0.0244		-0.0049	-0.0206
	(0.024)		(0.019)	(0.022)
GDP per capita ratio	-0.3948	-0.4005	-0.3943	0.3856
	(0.131)***	(0.128)***	(0.131)***	(0.131)***
Population ratio	0.0849	0.0866	0.0852	-0.0673
	(0.019)***	(0.019)***	(0.019)***	(0.028)**
Capital per worker ratio	0.3295	0.3323	0.3294	-0.2326
	(0.080)***	(0.078)***	(0.080)***	(0.069)***
Human Capital ratio	0.9535	0.9500	0.9547	0.0290
	(0.180)***	(0.176)***	(0.180)***	(0.072)
Land per worker ratio	-0.1030	-0.0995	-0.1030	0.1095
	(0.018)***	(0.018)***	(0.018)***	(0.045)**
Constant	0.2214	0.2620	0.2480	0.9514
	(0.046)***	(0.007)***	(0.040)***	(0.103)***
N	61410	64389	61410	61410
R-squared	0.52	0.51	0.52	0.56
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (7) using a sample of domestic and foreign subsidiaries that replicate home production. It uses an O*NET-based indicator for knowledge intensity based on workers' accumulated experience (excluding training). The left hand side variable is a binary variable that takes the value 1 if the subsidiary is foreign. The variables in the right hand side include the unit shipping cost associated with the industry, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parentheses.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A9: Determinants of Horizontal FDI (KI: experience)

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0628	-0.0625		-0.0645
	(0.037)*	(0.037)*		(0.037)*
log(d)	-0.0241		-0.0239	-0.0230
	(0.009)**		(0.009)**	(0.010)**
log(t)	0.0275	0.0281	0.0611	0.0271
	(0.067)	(0.067)	(0.056)	(0.066)
GDP per capita ratio	0.1309	0.1323	0.1282	0.8952
	(0.056)**	(0.057)**	(0.056)**	(0.241)***
Population ratio	0.0131	0.0212	0.0142	0.3121
	(0.007)*	(0.007)***	(0.007)**	(0.077)***
Capital per worker ratio	-0.0836	-0.1007	-0.0792	-0.6494
	(0.046)*	(0.047)**	(0.046)*	(0.180)***
Human Capital ratio	-0.0071	0.0350	-0.0098	-0.4310
	(0.063)	(0.063)	(0.064)	(0.283)
Land per worker ratio	-0.0132	-0.0107	-0.0131	0.0122
	(0.007)*	(0.008)	(0.007)*	(0.053)
Constant	0.6318	0.4487	0.6745	0.7556
	(0.162)***	(0.136)***	(0.148)***	(0.227)***
N	55136	55137	55136	55136
R-squared	0.47	0.47	0.47	0.47
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (8) using a sample of foreign subsidiaries of MNCs. It uses an O*NET-based indicator for knowledge intensity based on workers' accumulated experience (excluding training). The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A10: Ease of Communication (KI: experience)

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0628	-0.0629	-0.0621	-0.0629
	(0.037)*	(0.037)*	(0.037)*	(0.037)*
log(d)	-0.0241	-0.0252	-0.0072	-0.0185
	(0.009)**	(0.009)***	(0.014)	(0.008)**
log(t)	0.0275	0.0274	0.0281	0.0273
	(0.067)	(0.067)	(0.067)	(0.067)
Non-stop Flight		-0.0080		
		(0.010)		
Working hours overlap			0.0093	
			(0.005)**	
Common Language				0.0503
				(0.024)**
GDP per capita ratio	0.1309	0.1314	0.1282	0.1290
	(0.056)**	(0.056)**	(0.056)**	(0.055)**
Population ratio	0.0131	0.0133	0.0126	0.0142
	(0.007)*	(0.007)*	(0.007)*	(0.007)**
Capital per worker ratio	-0.0836	-0.0844	-0.0895	-0.0803
	(0.046)*	(0.046)*	(0.044)**	(0.044)*
Human Capital ratio	-0.0071	-0.0043	-0.0051	-0.0126
	(0.063)	(0.062)	(0.063)	(0.064)
Land per worker ratio	-0.0132	-0.0126	-0.0079	-0.0075
	(0.007)*	(0.007)*	(0.007)	(0.008)
Constant	0.6318	0.6431	0.4340	0.5765
	(0.162)***	(0.161)***	(0.219)**	(0.160)***
N	55136	55136	55136	55132
R-squared	0.47	0.47	0.47	0.47
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	N

The table presents results for the estimation of Specification (8) using a sample of foreign subsidiaries of MNCs. It uses an O*NET-based indicator for knowledge intensity based on workers' accumulated experience (excluding training). The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures (in standard deviations from the mean) and other controls. The right hand side also includes variables measuring the ease of communication between a headquarters and its subsidiaries. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A11: Determinants of Horizontal FDI, Distance Dummies

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable			
	(1)	(2)	(3)
k	-0.0868		-0.0894
	(0.043)**		(0.043)**
log(t)	0.0225	0.0598	0.0217
	(0.066)	(0.056)	(0.065)
500-1000Km	-0.0249	-0.0246	-0.0234
	(0.012)**	(0.012)**	(0.013)*
1000-1500Km	-0.0321	-0.0310	-0.0230
	(0.023)	(0.024)	(0.021)
1500-2000Km	-0.0569	-0.0570	-0.0411
	(0.029)**	(0.028)**	(0.024)*
2000-2500Km	-0.0554	-0.0546	-0.0309
	(0.033)*	(0.033)*	(0.029)
2500-3000Km	-0.0724	-0.0728	-0.0391
	(0.035)**	(0.035)**	(0.031)
3000-3500Km	0.0641	0.0636	0.0907
	(0.067)	(0.066)	(0.060)
3500-4000Km	0.0531	0.0502	0.0830
	(0.073)	(0.073)	(0.067)
4000-4500Km	-0.1029	-0.1031	-0.0707
	(0.045)**	(0.045)**	(0.048)
4500-5000Km	-0.0982	-0.1009	-0.0867
	(0.038)***	(0.037)***	(0.036)**
5000-5500Km	-0.0627	-0.0661	-0.0604
	(0.037)*	(0.037)*	(0.036)*
5500-6000Km	-0.0987	-0.1008	-0.0951
	(0.027)***	(0.026)***	(0.027)***
6000-6500Km	-0.1034	-0.1037	-0.0994
	(0.027)***	(0.027)***	(0.027)***
6500-7000Km	-0.1031	-0.1037	-0.0957
	(0.029)***	(0.029)***	(0.030)***
7000-7500Km	-0.1191	-0.1190	-0.1068
	(0.031)***	(0.031)***	(0.031)***
7500-8000Km	-0.1289	-0.1284	-0.1097
	(0.035)***	(0.034)***	(0.033)***
8000-8500Km	-0.0853	-0.0863	-0.0653
	(0.032)***	(0.031)***	(0.034)*
8500-9000Km	-0.0710	-0.0705	-0.0475
	(0.033)**	(0.033)**	(0.037)
9000-9500Km	-0.0775	-0.0766	-0.0528
	(0.036)**	(0.036)**	(0.037)
9500-10000Km	-0.0647	-0.0596	-0.0350
	(0.035)*	(0.036)*	(0.036)
10000Km+	-0.0362	-0.0349	-0.0087
	(0.029)	(0.029)	(0.033)
Constant	0.4849	0.5342	0.5921
	(0.140)***	(0.127)***	(0.218)***
N	55137	55137	55137
R-squared	0.47	0.47	0.47
MNC FE	Y	Y	Y
Host Cntry FE	N	N	Y

The table presents results for the estimation of Specification (8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary in dummies each representing a 500km interval, the unit shipping cost, knowledge intensity measures. All specifications include a vector of controls which include the ratio of GDP per capita, population, human capital, physical capital and land between the home and recipient country of the investment. All columns also include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A12: Determinants of Horizontal FDI, excluding China

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0869	-0.0864		-0.0890
	(0.044)**	(0.044)**		(0.044)**
log(d)	-0.0257		-0.0254	-0.0249
	(0.010)***		(0.010)***	(0.011)**
log(t)	0.0256	0.0261	0.0625	0.0253
	(0.066)	(0.067)	(0.057)	(0.065)
GDP per capita ratio	0.1303	0.1269	0.1287	0.9143
	(0.055)**	(0.055)**	(0.055)**	(0.243)***
Population ratio	0.0127	0.0219	0.0141	0.3109
	(0.007)*	(0.007)***	(0.007)**	(0.077)***
Capital per worker ratio	-0.0834	-0.1030	-0.0793	-0.6143
	(0.047)*	(0.048)**	(0.048)*	(0.182)***
Human Capital ratio	-0.0064	0.0531	-0.0106	-0.5025
	(0.066)	(0.065)	(0.067)	(0.292)*
Land per worker ratio	-0.0131	-0.0120	-0.0132	-0.0179
	(0.007)*	(0.008)	(0.007)*	(0.052)
Constant	0.6420	0.4466	0.6891	0.6876
	(0.161)***	(0.135)***	(0.149)***	(0.214)***
N	54259	54260	54259	54259
R-squared	0.47	0.47	0.47	0.48
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (8) using a sample of foreign subsidiaries of MNCs, excluding subsidiaries in China. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$