# The Hardships of Long Distance Relationships: Time Zone Proximity and the Location of MNC's KnowledgeIntensive Activities 

Dany Bahar

CID Research Fellow and Graduate Student
Working Paper No. 85
August 2017 (Revised August 2019)
© Copyright 2017, 2019 Bahar, Dany; and the President and Fellows of Harvard College


Working Papers
Center for International Development at Harvard University

# The hardships of long distance relationships: time zone proximity and the location of MNC's knowledge-intensive activities* 

Dany Bahar ${ }^{\dagger}$<br>The Brookings Institution<br>Harvard CID, CESifo $\xi^{\mathcal{G}}$ IZA

August 1, 2019


#### Abstract

Using a unique dataset on worldwide multinational corporations (MNCs) with the precise location of headquarters and affiliates, I present evidence of a trade-off between distance to the headquarters and the knowledge intensity of its subsidiaries' economic activity. This trade-off is strongly diminished the higher the overlap in working hours between the headquarters and its foreign subsidiary. In order to rule out biases arising from confounding factors, I implement a regression discontinuity framework to show that the economic activity of a foreign subsidiary located just across the time zone line which increases the overlap in working hours with its headquarters is, on average, roughly one percent higher on the knowledge intensity scale. The results are driven by horizontal foreign subsidiaries. Overall, the findings suggest that barriers to real-time communication within MNCs play an important role in the location decisions of their subsidiaries.


Keywords: time zones, multinational corporations, knowledge, distance, FDI
JEL Classification Numbers: F23, L22

[^0]
## 1 Introduction

Arrow (1969) suggests that the transmission of knowledge is difficult and costly. It is reasonable to assume that these difficulties arise, in part, because effective knowledge transmission involves human interaction, which cannot be fully replaced with written words, nor be embedded in goods that can be shipped at low costs ${ }^{1}$ A firm, as any other economic agent, faces difficulties when transferring knowledge among different divisions and affiliates, as has been extensively explored in the literature (e.g., Oldenski, 2012, Keller and Yeaple, 2013; Giroud, 2012). The ability of a multinational corporation (MNC) to transfer knowledge to its affiliates, however, should reflect the locations of the firm's headquarters and its subsidiaries. This paper explores the role of time zones in the location decisions of knowledge-intensive activities of MNCs.

This paper's contribution to the literature is threefold. First, using a highly detailed establishment-level worldwide dataset on MNCs I complement the existing empirical evidence of an existing trade-off between the level of knowledge intensity of a foreign subsidiary's economic activity and the geographic distance between such subsidiary and its global headquarters. Second, I show that this "knowledge and distance trade-off" significantly weakens, and even disappears, the more overlap in working hours there is between a subsidiary and its global headquarters. Third, and more importantly, I use a regression discontinuity design to exploit discrete spatial variation in time zones and show that subsidiaries located at roughly the same distance from their headquarters are active in economic activities that significantly differ in knowledge intensity depending on whether they are located in a time zone closer to the headquarters or not. To the best of my knowledge, this is the first time this methodology has been used in this context.

This trade-off between distance and knowledge intensity cannot be explained using earlier frameworks that looked at the fragmentation of MNCs which, implicitly or explicitly, assumed zero marginal cost or costs orthogonal to distance of transferring knowledge between headquarters and subsidiaries (i.e., Helpman, 1984, Markusen, 1984, Brainard, 1993, Markusen et al., 1996, Markusen, 1997, Carr et al., 2001, Helpman et al., 2004, Markusen and Maskus, 2002). ${ }^{2}$ However, such trade-off would emerge from any model that incorporates the idea of marginal costs of transferring knowledge increasing in geographic distance. For example, Ramondo and Rodriguez-Clare (2013) study the gains from openness in a comprehensive model that includes both multinational production and trade, and assume that all multinational production activity entails iceberg type efficiency losses that vary across country pairs. This marginal cost, they assume, is an increasing function of geographical and cultural distance. Ramondo (2014) models multinational production where foreign subsidiaries face a combi-

[^1]nation of variable and fixed costs when relying on technology or knowledge from their headquarters, capturing the idea of limited span of control for the headquarters, thus generating a loss of productivity at the affiliate level. Arkolakis et al. (2013) model an economy in which countries specialize either in innovation or in production. When a country specializes in innovation it implies that its firms open production subsidiaries abroad with a marginal cost that affects their productivity. These marginal costs are meant to capture various impediments that multinationals face when operating in a different economic, legal or social environment, as well as the various costs of technology transfer incurred by multinationals in different production locations. Tintelnot (2017) presents a model of multinational production in which foreign plants engage in variable costs that are determined, in part, by distance between the headquarters and the location of the foreign subsidiary. Keller and Yeaple (2013) provide an explanation as to why the marginal cost of knowledge transmission increases with distance by modeling such costs as shipping costs for intermediate goods embedding headquarters services ${ }^{3}$ In their model, shipping of intermediate goods is more prevalent for subsidiaries active in knowledge intensive activities.

This paper builds on the research by Keller and Yeaple (2013), in further exploring the "distance and knowledge (intensity) trade-off" emerging from the distance-increasing cost of knowledge transmission within the MNC. In this context, this study contributes to this literature by finding that this tradeoff is to some extent determined by the ability of workers in different MNCs affiliates to communicate in real time, as measured by the overlap in working hours between a headquarters and its subsidiaries, on their precise geographic location. That is, if ceteris paribus, time zones affect the location decision of MNCs, then trade costs of intermediate goods cannot serve as a sufficient explanation for the "distance and knowledge" trade-off ${ }^{4}$

An interpretation of this finding is that not all knowledge can be fully embedded in intermediate goods. Rather, location decisions of MNCs depend on aspects beyond transportation costs, such as the ease of communication between headquarters and foreign subsidiaries, in aspects related to management, monitoring, coordination, troubleshooting, etc. All these aspects crucial to MNCs are, arguably, forms of tacit knowledge, as they require human interaction (Arrow, 1969). Thus, the loss in efficiency that distant foreign affiliates face as evidenced in the literature (e.g., Giroud, 2012) is likely related to the difficulties in transferring tacit knowledge across long distances (Polanyi, 1962), and not only due to difficulties associated with costs related to transportation of intermediate goods.

In fact, the consensus in the existing literature on the economics of knowledge is that the transmission of knowledge is not immediate, and that knowl-

[^2]edge diffusion strongly decays with distance ${ }^{5}$ For instance, Jaffe et al. (1993) -among the first ones to make this claim-show that patent citations are more frequent within the same geographic area. 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 1200 Km . In the context of MNCs, it has also been shown that more complicated tasks require more time and effort for coordination and monitoring, and this becomes much more difficult at longer distances (e.g., Gumpert, 2018) ${ }^{6}$

The empirical exercise in this paper is based on a sample of about 70,000 domestic and 45,000 foreign horizontal subsidiaries belonging to over 3,200 MNCs from the Worldbase dataset by Dun \& Bradstreet ${ }^{7}$ I focus my attention on MNCs active in the manufacturing sector that have expanded into foreign countries $\int^{8}$ For each one of the subsidiaries in the sample I have information on their physical location and primary economic activity, as defined by the 1987 Standard Industry Classification (SIC). A novel source of variation this paper relies on is a industry-level knowledge intensity measure that I construct based on workerlevel characteristics for each industry, as opposed traditional measures that rely on firms' balance-sheet data, such as R\&D intensity. The index I use aims to capture the tacit knowledge intensity of an economic activity by averaging the accumulated experience and training of the workforce in an industry, using occupational characteristics defined in the O*NET project dataset ${ }^{9}$ The main results, however, are robust to alternative industry-level measures of knowledge intensity.

After computing precise distances between each affiliate and its MNC global headquarters following a geocoding process using Google Maps, and using the knowledge intensity measures I document a number of stylized facts using the dataset.

First, I find a negative partial correlation between knowledge intensity and the distance between a headquarters and its subsidiaries, a result that is par-

[^3]ticularly strong for foreign horizontal subsidiaries, consistently with Keller and Yeaple (2013) and their modeling of the proximity concentration hypothesis. For example, everything else equal, the estimation in the paper implies that an American MNC headquartered in Houston TX in the United States would locate a meat packing subsidiary -an economic activity with low knowledge intensity levels- in Kabul in Afghanistan (approximately 12000Km away from the headquarters) and a semiconductor plant -an economic activity with high levels of knowledge intensity- in Ireland (approximately 7000 Km away from the headquarters). Second, I descriptively show that the distance and knowledge trade-off is weakened when there is a larger overlap in working hours between the subsidiary and its headquarters. Thus, 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 enough to explain the fact that MNCs tend to locate geographically closer their subsidiaries active in knowledge intensive activities.

In order to rule out other confounding factors driving the time zone result, I estimate a regression discontinuity using discrete changes in the overlap of working hours across time zone lines around the globe. Using this methodology I find evidence that subsidiaries with a higher overlap in working hours with their headquarters are, on average, active in industries that are more intensive in knowledge. I find, however, that this discontinuity is present only for foreign horizontal subsidiaries, and not so for other subsidiary types. In particular, the estimations using regression discontinuity find that the economic activity of a foreign horizontal subsidiary located just across a time zone line closer to the headquarters -thus increasing the overlap in working hours- is 0.6 to 0.84 percent higher on the knowledge intensity scale. The average effect of one more hour of overlap is equivalent to a reduction of about 200 Km . between a headquarters and its foreign subsidiary, based on the estimated knowledge and distance trade-off. I further show that the discontinuity cannot be explained by observable determinants of the location decision of MNCs, and the results are robust to using different bandwidths, weights and alternative knowledgeintensity measures, such the industry-level R\&D intensity.

This is consistent with the idea that the transmission of knowledge is costly for the replication of production in other remote locations, and these costs can be reduced when monitoring and communication can happen in a similar working day schedule for different locations. On the same note, these costs might be less relevant for domestic subsidiaries, or other types of foreign subsidiaries, a finding that I discuss more in detail when presenting the main results. Overall, the results in this paper suggest that the costs associated with transferring tacit knowledge within a firm is an important determinant of the concentrationproximity hypothesis, and in particular, the knowledge and distance trade-off that emerges from it.

The rest of the paper is divided as follows. Section 2 describes the dataset and the construction of relevant variables. Section 3 presents descriptive evidence documenting, among other results, the distance and knowledge intensity trade-off that arises from plausibly existing costs knowledge transmission within

MNCs, and how this trade-off is weakened by a higher overlap in working hours between affiliates and their headquarters. Section 4 present the main results of the paper where I implement a regression discontinuity design to estimate whether knowledge intensity significantly differs for subsidiaries across different time zones. Section 5 concludes and addresses areas for future research regarding the role of tacit knowledge in economic activity ${ }^{10}$

## 2 Data and Definitions of Variables

### 2.1 Worldbase dataset by Dun \& Bradstreet

I use the Worldbase dataset by Dun \& Bradstreet as the main data source for the empirical exercise. The dataset has information on more than one hundred million establishments worldwide with data from year 2012 (when it was acquired). Each establishment is uniquely identified and linked to its global headquarters (referred to as the "global ultimate"). For this study I focus on 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. 11

The sample obtained from the dataset includes about 67.7 thousand domestic and 43.5 thousand foreign subsidiaries of MNCs active in the manufacturing sector and scattered over 100 countries, which report to over $3200 \mathrm{MNCs}{ }^{12}$ Domestic subsidiaries are defined as those who are located in the same country of the headquarters, while foreign subsidiaries are those located in a different country.

For the analysis, I will use the reported main SIC code as the only indicator of a plant's economic activity. There are about 450 unique SIC 4-digit codes reported by subsidiaries as their main economic activity in the dataset. The sample I use is significantly smaller than the overall Dun \& Bradstreet dataset (which originally has over 124 million establishments) for several reasons. First, I only include subsidiaries of multinational corporations. Second, I further limit the study to all domestic subsidiaries and to all foreign subsidiaries active in the manufacturing sector (as known, most of the MNCs are in the service sector, as shown by Alfaro and Charlton, 2009).

In order to obtain the precise location of each subsidiary I geocode the dataset using Google Maps Geocoding API to find the exact latitude and longitude of its headquarters and each one of its subsidiaries as well as the distance between them. Figure 1 maps the unique locations of all foreign subsidiaries (dots) and headquarters (triangles) in the sample.

[^4][Figure 1 about here.]
For instance, Figure 2 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 2 about here.]
Online Appendix Section A discusses this dataset more in detail.

### 2.2 Main Variables Definitions

### 2.2.1 Foreign Subsidiaries: Horizontal and Vertical

For the plant-level dataset I define a foreign subsidiary as a subsidiary located in a country other than the MNC's headquarters. I further classify foreign subsidiaries as horizontal or vertical. An horizontal expansion is the foreign subsidiary's SIC code vis-à-vis all the SIC codes reported by any of the domestic counterparts of the same company. 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.

I also classify foreign subsidiaries as vertical ones, based on input-output coefficients of the industry of each subsidiary vis-à-vis SIC codes of its domestic counterparts. In particular, 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 ${ }^{13}$ A vertical subsidiary is defined as upstream (alt. downstream) vertical if its main economic activity is an input (alt. output) of $\$ 0.05$ or more per each dollar for the economic activity of any of the domestic subsidiaries of the firm ${ }^{14}$

The diagram in Figure 3 is useful to understand how horizontal and vertical links are defined in the dataset. It shows how I use all domestic subsidiaries of a MNC as the benchmark to classify foreign subsidiaries as horizontal or vertical. Foreign subsidiaries that do not classify as neither horizontal nor vertical according to these definitions are kept as part of the sample, too.

[^5][Figure 3 about here.]
In the sample, about $24 \%$ of subsidiaries are foreign horizontal and $9 \%$ are vertical. Similarly to Alfaro and Charlton (2009), I find that some horizontal foreign affiliates also classify as vertical. I treat these subsidiaries as horizontal, given that they represent a foreign replication of production of some industry located in the home country, despite also being part of the supply chain of other industries existing in the home country of the $\mathrm{MNC}{ }^{15}$ There is a number of foreign subsidiaries, too, that do not classify as either. Throughout the paper I present results using different cuts of the sample based on the type of subsidiary: all (including both domestic and foreign subsidiaries), foreign (including horizontal, vertical and others), horizontal (including those that classify as both horizontal and vertical), and (strictly) vertical.

### 2.2.2 Knowledge Intensity Measures

In order to estimate the knowledge intensity of industries I create a new measure that aims to capture the tacit knowledge intensity for each industry. The measures 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 $\mathrm{O}^{*}$ NET, which includes results from a large number of survey questions on the characteristics of each occupation $\sqrt{19}$

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 $\mathrm{O}^{*}$ NET are:

- How much related experience (in months) would be required to be hired to perform this job?

[^6]- How much "on-site" or "in-plant" training (in months) would be required to be hired to perform this job?
- 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. 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. In particular, for each sector $s$ knowledge intensity is defined as:

$$
K I_{i}=\sum_{o} \omega_{o, i} \text { cumexp }_{o}
$$

Where cumexp ${ }_{o}$ is the sum of experience and training associated to occupation $o, \omega_{i, o}$ is the weight of occupation $o$ in industry $i$, measured by either share of employment or wage. Naturally, we have that $\sum_{o} \omega_{o, i}=1$ for every $i$. While I use within-sector wage share to define $\omega_{o, i}$, the results are robust to using the within-sector employment share instead.

Using this measure, manufacturing 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) ${ }^{20}$

I find this measure correlates positively with other (proxies of) knowledge intensity measures used in the literature, such as the average $R \& D$ share of sales per industry. Correlation coefficient is 0.13 with $\mathrm{R} \& D$ intensity computed using the Compustat dataset and compiled by Keller and Yeaple (2013) and 0.20 with R\&D intensity computed using the Orbis dataset and compiled by Nunn and Trefler (2008), using manufacturing NAICS 4-digit industries ${ }^{21}$

The R\&D based measures, however, have three main shortcomings that could generate significant biases. First, these measures are concentrated in lower values and have a long tail, as shown in Figure 4 . This is 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 almost indistinguishable. 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, which may not be knowledge intensive industries. Third, R\&D investment might not be equally accounted for across all industries.
[Figure 4 about here.]

[^7]The O*NET based measure solves these issues. As shown in Figure B1, the distribution is smoother and closer to a two-tailed normal distribution ${ }^{22}$ Moreover, this measure does not rely on a sampling of particular firms, and uses the same standardized inputs for all industries. Because of these reasons, I use these indicators as the main proxies for knowledge intensity throughout the paper, though I include robustness tests that estimates the main results using alternative knowledge intensity measures, including the R\&D based ones.

Unless otherwise noted, all the estimations in the paper use the experience plus training measure to quantify knowledge intensity.

### 2.2.3 Unit shipping costs

Unit shipping costs for SIC manufacturing industries are computed using data from Bernard et al. 2006. ${ }^{23}$ This industry-level measure aims to proxy 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 100 Km . It is computed by averaging the same measure per industry across all countries exporting to the US in year 2005. To deal with long tails, this variable will be used in a logarithmic scale in all the different empirical specifications.

Unit shipping cost figures are negatively correlated with the knowledge intensity measures, with a correlation coefficient of about -0.6.

## 3 Research question and descriptive evidence

### 3.1 Research question

A result that has emerged from the international economics literature is that, in the presence of distance-increasing marginal costs of transferring knowledge within MNCs, firms would be less likely to geographically expand their knowledge-intensive activities to further away locations. The second implication of the existence of these costs is that affiliates locate in closer geographic proximity to the headquarters the more intensive in knowledge their main economic activity is (e.g., Ramondo and Rodriguez-Clare, 2013; Arkolakis et al. 2013, Keller and Yeaple, 2013, Ramondo, 2014 Tintelnot, 2017). Most recently, this "distance-knowledge trade off" has been empirically shown by Keller and Yeaple (2013) using American MNCs. An accepted explanation of this trade-off is 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

[^8]locations (Keller and Yeaple, 2013, Irarrazabal et al., 2013). This explanation, however, assumes implicitly or explicitly that knowledge can be fully embedded in intermediate goods, that are in turn shipped to remote locations.

However, this assumption might not be proper for all types of knowledge. Tacit knowledge cannot be, by definition, embedded in intermediate goods. For instance, transmission of tacit knowledge within the firm could involve costs of monitoring by the management in the headquarters or also real-time problem solving efforts by the chief engineering team dealing with the foreign team in certain manufacturing processes ${ }^{24}$ Distance would play an important role in this matter, as longer distances difficult the ability of team members located in different locations to work together in real time through phone, video conferences, etc. The ability for coworkers in MNCs to communicate in real-time for coordination, troubleshooting or monitoring purposes, for example, could be critical for certain industries. Thus, it could well be that it is the cost of transmitting this type of knowledge which partly drives the documented relationship, and not only the costs associated with shipping intermediate goods.

At the core of this paper is an empirical exercise that aims to disentangle between both explanations. 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, arguably, any aspect that eases communication that does not affect geographic distance should be orthogonal to transportation costs of intermediate goods. In particular, my focus is on time zones differences, thus the overlap in working hours between a headquarters and its subsidiaries. The ability to communicate in real time during working hours that are common could be crucial for knowledge intensive activities. Thus, simply put, the question is whether the intensity of this distance-knowledge trade-off is affected by the overlap in working hours between a headquarters and its subsidiaries.

Fortunately, the global character of the dataset I use provides a proper setting to study this question, by applying a regression discontinuity approach. In particular, the dataset allows me to compare the knowledge intensity of the economic activity of subsidiaries that are almost equidistant to the headquarters, but differ in the number of overlap in working hours (as these subsidiaries are located at either side of a time zone line). If a larger overlap in working hours is in fact cost-reducing for the knowledge transfer within the MNC, then we would observe that the economic activity of subsidiaries in a closer time zone to the headquarters would be, on average, more knowledge intensive.

[^9]
### 3.2 Stylized facts: the knowledge-distance trade-off

Before getting into the specifics of the regression discontinuity, I present some stylized facts and descriptive evidence of the distance-knowledge trade-off making use of the above described dataset and measures. Table 1 provides descriptive details about the used sample, in terms of the distribution of domestic and foreign affiliates (which are typically located at greater distance from the headquarters than purely domestic ones) across regions of the world and developing vs. developed countries.

## [Table 1 about here.]

In total, the complete sample includes 111,172 subsidiaries that are owed by $3,229 \mathrm{MNCs}$. About $39 \%$ of these subsidiaries are foreign, which include both horizontal and vertical ones, as well as foreign subsidiaries that do not fit into any of these two categories as defined above. 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. On average, foreign subsidiaries are engaged in less knowledge-intensive activities their domestic counterparts. The same pattern holds for all presented cuts of the data, except for few firms based in Western European countries. This is consistent with the idea that foreign expansions are less likely in knowledge intensive activities, as shown by Keller and Yeaple (2013). Online Appendix Section (C) expands on this analysis, using regression analysis to explore these patterns in the dataset, finding consistent results with the literature.

The main focus of this paper, however, is on exploring the relationship between distance and knowledge intensity, for which the theory and existing evidence suggests the existence of a trade-off between the two. I estimate this trade-off using the dataset for several cuts of the sample, exploiting the knowledge intensity of the economic activity of subsidiaries and their geographic distance to headquarters. The estimation is based on the following specification:

$$
\begin{equation*}
\log \left(k_{s}\right)=\beta_{d} \cdot \log \left(d_{s}\right)+\beta_{t} \cdot \log \left(t_{s}\right)+\varphi_{h}+e_{s} \tag{1}
\end{equation*}
$$

In the specification, $s$ indexes a subsidiary. The left hand side variable, $\log \left(k_{s}\right)$ is the logarithmic transformation of knowledge intensity associated to the economic activity of the subsidiary. The main right hand side variable is the geographic distance between the location of the headquarters and that of the foreign subsidiary, in $\operatorname{logs}$, denoted by $\log \left(d_{s}\right)$. In an effort to reduce possible omitted variable bias when estimating $\beta_{d}$, I control for the log unit shipping cost $\left(t_{s}\right)$, as well MNC fixed effects $\left(\varphi_{h}\right)$. The inclusion of shipping costs as control are particularly important for foreign horizontal subsidiaries, where trade costs are a crucial input in the decision between exporting or reproducing production in foreign markets and save such trade costs (i.e., Helpman et al., 2004). They might also be crucial for upstream vertical foreign subsidiaries, as higher shipping costs directly increase production costs. For consistency,
however, I control for these shipping cuts in every estimation using different sub-samples.

Evidence of the existence of a distance-knowledge trade-off would be supported by $\beta_{d}<0$ : foreign subsidiaries locate closer to the headquarters the more knowledge intensive they are. Note that the inclusion of MNC fixed effects imply that the exploited variation within firm.

I estimate $\beta_{d}$ for different cuts of the dataset for which different characteristics could play a role in the existence and intensity of a trade-off between knowledge and distance. First, I use the sample of all subsidiaries, which includes both foreign and domestic subsidiaries alike. In principle, even though we could expect an international border to increase the difficulties in the transmission of knowledge between a headquarters and its subsidiaries, such trade-off could also exist within a country, but might not show up in the estimation if the costs of transferring knowledge from a headquarters to its domestic subsidiaries are considerably smaller.

Second, I limit the sample to only foreign subsidiaries for which the trade-off would also exist, though it might depend on the type of subsidiary; therefore I explore this further in the following cuts.

Third, I estimate the relationship using horizontal foreign expansions. Foreign horizontal expansions replicate production in other locations, in order to substitute for trade. If replicating production requires constant interaction with headquarters, particularly for knowledge intensive economic activities -thus increasing marginal costs-, then we expect a negative estimator of $\beta_{d}$.

Fourth, I limit the sample to vertical foreign expansions, for which such a trade-off would exist if, similarly to what we would expect in horizontal expansions, MNCs regularly transfer knowledge from their headquarters to foreign subsidiaries that are on either side of the value chain.

Table 2 presents summary statistics for the relevant variables across the different sample cuts used in the exercise. It includes statistics for distance between each subsidiary and its headquarters, the knowledge intensity of the economic activity of the subsidiary and its associated unit shipping cost, as well as the overlap in working hours. In order to compute the overlap 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 $6: 00 \mathrm{pm}$ ( 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.

As shown in the table, the sample cut that includes "all" subsidiaries (that is, both domestic and foreign) has lower average values for distance and higher for overlap in working hours, as would be expected given that domestic subsidiaries are typically geographically closer to the headquarters. The "foreign" cut is composed by both horizontal and vertical foreign subsidiaries, as well as others that do not fit into any of those two definitions. All in all, the foreign subsidiaries (whether it includes all of them, or only the horizontal or the vertical ones) have similar average values in terms of distance to the headquarters, knowledge intensity and working hours overlap. Note, however, that average unit shipping
costs are lower in the "vertical" sample, and the minimum value for distance to the headquarters is almost as twice for the same sample as compared to foreign and to horizontal subsidiaries.
[Table 2 about here.]
The results of the estimation for $\beta_{d}$ are plotted in Figure 5, for each one of the four samples explained above. The results are indicative of the tradeoff between distance $(d)$ and knowledge intensity $(k)$. Note that it is only for horizontal foreign subsidiaries for which the trade-off is well defined and statistically significant. For the sample that pools all foreign subsidiaries there is no evidence of such trade-off, which might be because fundamental difference in the levels of the variables under consideration between subsidiaries that are foreign but cannot be classified as horizontal nor vertical and those that can. These differences might be distorting the relationship, thus not reflecting the expected downward sloping relationship. In fact, when looking at both horizontal and vertical foreign subsidiaries, separately, the point estimates for $\beta_{d}$ are negative. A thorough discussion on possible explanations behind the heterogeneity of results is included in Section 4, when presenting the main results of the paper ${ }^{25}$
[Figure 5 about here.]
All in all, however, the results are indicative of the trade-off between distance $(d)$ and knowledge intensity $(k)$, in particular for foreign horizontal subsidiaries (Panel 3), which is consistent with many of the aforementioned studies that introduce the cost of knowledge transmission within models of the proximityconcentration hypothesis. According to the estimation, a $10 \%$ increase in distance to the headquarters is associated with economy activity that is up to 0.017 percent lower in the knowledge intensity scale, depending on the used sample ${ }^{26}$ When focusing on foreign horizontal subsidiaries, for which the results are statistically significant, the corresponding figure is 0.014 , which while small in magnitude, matters for long distances such as the ones in that cut of the sample (the average distance for foreign horizontal subsidiaries is about 3650 Km.).

[^10]A back-of-the-envelope calculation exemplifies the economic significance of the point estimate. For instance, suppose an American MNC headquartered in the Houston, Texas. According to the estimated distance-knowledge trade-off, this MNC would locate its low knowledge intensive activity, such as a meat packing subsidiary, in Kabul, Afghanistan (approximately, 12600Km away), while an economic activity that is roughly twice as knowledge intensive, for example, a semiconductor plant, in Dublin, Ireland (approximately 7200km away) ${ }^{27}$

## Ease of communication and the knowledge-distance trade-off

I start the exploration of the main research question by providing more descriptive evidence surrounding the distance-knowledge trade-off once we take into account variables that ease the communication between a subsidiary and its headquarters. In particular I include in Specification (1) the overlap in working hours between each subsidiary and its headquarters, interacted with distance. This variable proxies for the ease of communication within the firm and its interaction allow us to understand whether the distance-knowledge trade-off is different for subsidiaries that are equidistant to the headquarters but with larger overlap in working hours. In a sense, this variable aims to capture "realtime" communication ability between workers in the two plants. 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. Stein and Daude (2007), for instance, find that time zone is an important determinant of aggregate FDI flows, which they attribute to better monitoring ${ }^{28}$

Naturally, one concern that might arise before even going into the results it the high multicollinearity between distance and overlap in working hours. Fortunately, while there is an obvious correlation between the two variables, the global nature of the dataset provides enough variation across these two

[^11]$$
\log \left(d_{s}\right)=\beta_{k} \log \left(k_{s}\right)+\beta_{t} \log \left(t_{s}\right)+\varphi_{h}+\varepsilon_{s}
$$

[^12]dimensions that essentially allows for the comparison between subsidiaries that, despite having a similar distance to their global headquarters, have significant differences in the overlap in working hours, as shown in Figure 6 .
[Figure 6 about here.]
The results for the estimation of Specification (1) including the overlap in working hours as an additional variable are presented in Table 3. All columns use the experience plus training indicator to proxy for $k$. Column 1 presents results using all subsidiaries, column 2 limits the sample to only foreign subsidiaries, while columns 3 and 4 does it for foreign horizontal and foreign vertical, respectively.
[Table 3 about here.]
First, across all specifications, the estimates for $\beta_{d}$ are negative and statistically significant, consistent with the previous evidence on the existence of a trade-off between distance and knowledge intensity ${ }^{29}$ In fact, note that the point estimate for $\beta_{d}$, across the board, is significantly larger maintaining its negative sign implying that when there is no overlap in working hours, the distance and knowledge intensity trade-off becomes much more pronounced. In addition, and perhaps more importantly, the interaction term $\log (d) \times$ whoursoverlap is estimated as positive and statistically significant, implying that the knowledgedistance trade-off weakens the more overlap in working hours. It is also worth noting that across all specifications the overlap in working hours coefficients are estimated to be negative, simply because the longer the distance between the two locations the smaller the overlap tends to be.

Figure 7 presents a graphical representation of the distance and knowledge intensity trade-off for different possible values that define the overlap in working hours between the foreign subsidiary and its headquarters. Each row in the figure plots the result for each one of the different cuts in the sample as columns in Table3. The left panel plots the estimated slope of such trade-off as a function of overlap in working hours. When the overlap is 0 , the slope of the trade-off is highly negative with magnitudes that are often 10 times as much as the slopes estimated in Figure5. As the overlap becomes larger, the trade-off weakens and even disappears (i.e., losses statistical significance). The right panel, plots the trade-off for different values of overlap in working hours (normalizing the $\log (k)$ to be zero at its smallest value of the distribution, for comparison purposes), showing that the trade-off is very close to zero when there are 10 hours of overlap between the two locations.

I also compute the trade-off allowing for a non-linear effect of working hours overlap, and find results that are consistent with allowing for a trade-off that varies linearly with one more hour of overlap. In addition, I reestimate the tradeoff excluding extreme values of the knowledge intensity measure as well as using

[^13]alternative measures of knowledge intensity, and, for the most part, I find results that are consistent with the ones presented so far (in particular when it comes to foreign horizontal subsidiaries). The results for foreign vertical subsidiaries are also robust to a more conservative definition of vertical subsidiaries, which uses a $10 \%$ threshold to define upstream and downstream relationships using input-output tables. Finally, I also find that the role of working hours overlap in reducing the estimated trade-off is particularly important for MNCs for which their main economic activity is high in the knowledge intensity scale. For more details on these robustness checks refer to Online Appendix Section E.
[Figure 7 about here.]
Thus, the elaborated stylized facts show that the trade-off seems to be heavily influenced by the overlap in working hours, suggesting that time zones play an important role in the location decisions of MNCs. Yet, these results rely on partial correlations and are, at most, descriptive. In the next section I implement a regression discontinuity design exploiting discrete changes in time zones across geographic distances, in an effort to deal with endogeneity concerns.

## 4 Regression discontinuity results

So far, the stylized facts presented suggest that time zones are a determinant of the location of MNCs' knowledge intensive activities. However, in order to rule out biases arising from confounding factors, I implement a regression discontinuity (RD) framework exploiting the fact that time zones vary discretely with distance. By doing so, I effectively compare the knowledge intensity of subsidiaries that belong to the same MNC but that are on different sides of a time zone line ${ }^{30}$ To the best of my knowledge, using a regression discontinuity design exploiting time zone differences is novel in the context of international economics, and in particular, when studying patterns of behavior of multinational corporations.

As a first step it is important to define the running variable centered around each time zone line: it corresponds to the distance (in kilometers) from every subsidiary to the nearest time zone line. A MNC will often have subsidiaries in located in time zones that are both eastwards and westwards relative to the

[^14]headquarters location. The "treatment" in this setting is being in a time zone that allows for a higher overlap in working hours between the headquarters and the subsidiary. Figure 8 clarifies how the running variable is defined and normalized using an hypothetical example. For instance, subsidiaries 1 and 2 are located east of their headquarters' location, while subsidiaries 3 and 4 are located west of it. The running variable is based on the distance from each one of these subsidiaries to the nearest time zone line. In that sense, because subsidiary 1 is located on the side of the line closest to the headquarters (e.g., higher overlap in working hours), its distance is marked as positive. On the other hand, since subsidiary 2 is located on other side of the line, then the running variable in that area is negative. Same concept applies to the running variable for subsidiaries 3 and 4 .
$$
\text { [Figure } 8 \text { about here.] }
$$

Each observation in the sample corresponds to a subsidiary, and thus the treatment can be defined as being in a closer time zone to its headquarters. Each headquarters, of course, has subsidiaries that are near several distinct time zone lines. Thus, in this setting it is very important to make sure that the regression discontinuity design is indeed comparing subsidiaries that belong to the same MNC as well as near the same time zone line to one side or the other. Thus, I estimate the following specification which include fixed effects for every MNC and every time-zone line, as follows:

$$
\begin{align*}
& \log \left(k_{s}\right)=\gamma_{1} \operatorname{closer} T Z_{s, t z}+\gamma_{2} \operatorname{distT}^{2} \text { line }_{s, t z}+  \tag{2}\\
& \quad \gamma_{3} \operatorname{closer} T Z_{s, t z} \times \operatorname{distTZline}_{s, t z}+\varphi_{h}+\tau_{t z}+e_{s}
\end{align*}
$$

Where $s$ represents a subsidiary and $t z$ represents a time zone line which is the nearest one to subsidiary $s . \log \left(k_{s}\right)$ is the knowledge intensity of the economic activity of the foreign subsidiary. The variable is a dummy which is defined as closerT$Z_{s, t z}=1\left\{\right.$ distTZline $\left._{s, t z}>0\right\}$. distTZline ${ }_{s, t z}$ is the running variable, defined as distance from subsidiary $s$ to its closest time zone line $t z$ following the guidelines explained before and represented in Figure 8 , rescaled and expressed in hundreds of kilometers. The interaction closer $T Z_{s, t z} \times$ $\operatorname{distT}^{\text {Z }}$ line $_{s, t z}$ allows more flexibility when estimating the slopes before and after the cutoff (e.g., the time zone line). $\varphi_{h}$ and $\tau_{t z}$ are MNC and time zone line fixed effects, respectively. By adding these fixed effects I make sure that we compare foreign subsidiaries belonging to the same MNC and that are across the same time zone line ${ }^{31}$

Figure 9 graphically represents the estimation of Specification (2) using the package rdplot (Calonico et al., 2014, 2017) to estimate the optimal bandwidth. It also uses a linear estimator as well as the "mimicking variance evenly-spaced"

[^15]method to define the number of bins, which result in relatively small bin sizes, reducing the possibility that the discontinuity is driven by few outliers on either side ${ }^{32}$ Similarly to the previous section, I plot the discontinuity for four different cuts of the sample: all, foreign, (foreign) horizontal and (foreign) vertical subsidiaries.
[Figure 9 about here.]
It is clear from the figure that evidence of discontinuity is only present for foreign horizontal subsidiaries (Panel 3) and not for any other sub-sample ${ }^{33}$ The discontinuity for the sub-sample of foreign horizontal affiliates is present when plotting the discontinuity using both a linear and a quadratic estimator, as shown in Figure 10. Both fits show that, indeed, horizontal foreign subsidiaries located just across the time zone line that increases the overlap in working hours with their headquarters (above zero on the horizontal axis) are active in economic activities higher, on average, in the knowledge intensity scale, as compared to those who are on the other side of the time zone line (below zero on the horizontal axis). This graph uses observations based on the optimal bandwidth which corresponds to 299.16 Km . at either side of the time zone lines, using the methodology by Cattaneo et al. (2018) who builds on the work by Imbens and Kalyanaraman (2012) ${ }^{34}$ As described above, the number of bins is based on the "mimicking variance evenly-spaced" method, though varying the number of bins does not alter the result (see Figure F2 in Online Appendix Section $\bar{F})$. Note that the vertical axis has values below and above zero because for this plot I use the residual of $\log \left(k_{s}\right)$ after controlling for $\varphi_{h}$ and $\tau_{t z}$.
[Figure 10 about here.]
The fact that the discontinuity is graphically present for foreign horizontal subsidiaries and not for other types of subsidiaries deserves a discussion. First, this result is consistent with the stylized fact documented in Figure 5 The distance-knowledge trade-off seems to be particularly pronounced for foreign horizontal subsidiaries, and not for any of the other sub-samples, which suggests that that the transmission of knowledge is less costly -or not increasing in distance- for other types of subsidiaries. When it comes to the lack of a discontinuity at the time zone line, however, this suggests that overlap in working hours is not a significant determinant of the location of affiliates of MNCs, other than foreign horizontal ones.

In fact, when it comes to domestic subsidiaries, one would think that the cost of knowledge transmission within the MNC is less than for foreign subsidiaries,

[^16]perhaps explaining why results are inconclusive when including in the sample domestic subsidiaries (see Figure 5). Note too that while there are about 35 thousand domestic subsidiaries in the RD sample, there is significantly less time zone variation for domestic subsidiaries (as shown in Figure 6) which difficulties the precision required for this exercise. In fact, $75 \%$ of the domestic subsidiaries in the RD sample have an overlap in working hours of 10 (the maximum value).

But why should we expect the results differ for different types of foreign subsidiaries? That is matter for further empirical research, though the previous literature provides frameworks that shed light on some answers to this important question. The results using all foreign subsidiaries incorporate horizontal, vertical but also others that do not classify as neither (for which communication might be less critical). Thus, it is more insightful to discuss the results that look at the particular type of foreign subsidiaries.

When it comes to foreign horizontal subsidiaries, the existence of a distanceknowledge trade-off is consistent with several works in the literature that incorporate the cost of transferring knowledge, often increasing in distance, to explain the proximity-concentration hypothesis. In particular, when a MNC replicates its production abroad to substitute for exports, communication with the headquarters arguably represents a marginal cost: the product produced at home has to have the same or a similar process of production in the foreign plant, for which there is the need for constant interaction. Thus, if being able to communicate in real time reduces these costs, which in turn are much higher for knowledge intensive activities, the existence of a discontinuity makes sense.

For vertical subsidiaries, there are a number of possibilities on why we do not see a clear trade-off in Figure 5 and a discontinuity in Figure 9. If the processes in each location are different, one could argue that the need for constant interaction between the teams in different location might be less critical or time sensitive. Naturally, one could think of a different scenario, where communicating with subsidiaries that are either upstream or downstream is as critical as for horizontal subsidiaries, since lack of coordination puts in risk the efficiency of the whole supply chain. Yet, as shown in previous work, often the categorization of vertical subsidiaries using input-output tables is not reflected in actual intra-firm trade (e.g., Atalay et al. 2014, Ramondo et al. 2015), thus these foreign subsidiaries might not be playing a significant role in the supply chain of the domestic affiliates, and thus interaction with the headquarters is less relevant. Note, too, that the baseline sample of strictly vertical foreign subsidiaries is relatively small: about 2,000 observations overall (see note in Figure 5), therefore this sample has little variation left after the inclusion of MNC fixed effects, which plays an important role in the precision of the estimation (both for the descriptives and the RD). While I lack proper data to explore these and other alternative explanations, I take the heterogeneity in results for different subsidiary types not as an avenue for further research.

In theory, the existence of a knowledge-distance trade-off that could be weakened by a higher overlap in working hours would come down to two conditions. First, the existence of a marginal cost of knowledge transmission within the firm that exists in distance; and second, that easing the communication between the
locations reduces the cost. The evidence presented thus far supports the idea that these two conditions are met by horizontal foreign subsidiaries, but less so for other subsidiary types.

In the rest of this section, I limit the RD results to foreign horizontal subsidiaries, the only sample for which I can find a graphical discontinuity in reduced form, unless otherwise noted.

The actual estimation of Specification (2) using data on horizontal foreign subsidiaries is presented in Table 4 Each column in the table presents results using a different bandwidth definition which range from 150 to 350 kilometers as well as the optimal bandwidth in the last column, which corresponds to 299.16 Km . The estimation uses a triangular weight scheme, giving higher weight to observations closer to the cutoff point ${ }^{35}$

## [Table 4 about here.]

The results suggest that a foreign subsidiary located just across a time zone line closer to the headquarters -thus increasing the overlap in working hoursis active in an industry that is, on average, 0.6 to 0.84 percent higher in the knowledge intensity scale, depending on the bandwidth used. At first this might look like a very small number. But when comparing this to our overall estimation of the knowledge-distance trade-off (see Figure 5), the point estimate from the RD is 5 to 6 times larger in magnitude ( -0.00137 vs. 0.0066 to 0.0084 ). Thus, according to a back-of-the-envelope calculation, an additional hour of overlap would offset the "negative" impact that roughly 200 Km . has in the distanceknowledge trade-off estimated above (see Figure 5) ${ }^{36}$

Note that we should not expect a "large" effect: This exercise compares foreign subsidiaries located very close to one another, some of which, the treated, have just an extra hour of work overlap with their colleagues at the headquarters (or sometimes even less, as some time zones change in increments of half an hour). Also, this is an average treatment effect for all possible different treatments (e.g., increasing the overlap in working hours from 2 to 3 , or from 6 to 7 , for example) and as such is a weighted average of different treatment effects which could be larger or smaller, depending on the relative improvement. Yet, at most we will always be limited by a "small" treatment. Thus, the results are qualitatively important taking into consideration that the only thing which differs between subsidiaries located at either side of the time zone line is having one more hour of overlap with their colleagues working at the headquarters location, even if their external validity cannot be proven beyond the estimation presented in Table 3, with all the limitations such approach has ${ }^{37}$

[^17]
## Robustness

In order for the results to be interpreted as unbiased, it is crucial that the discontinuity being exploited is not explanatory of other observable variables that in turn could be correlated with knowledge intensity of the subsidiary's economic activity. In order to explore whether that is the case, I reestimate Specification (2) replacing the dependent variable with other observable measures considered determinants of the location decision of foreign subsidiaries. These are the trade unit cost of the economic activity of the subsidiary (which would be an important determinant particularly for foreign horizontal subsidiaries), as well as variables that characterize the country where the subsidiary is located (income per capita, capital per worker and human capital) ${ }^{38}$ Since some of the horizontal foreign subsidiaries in the sample could also be classified as being upstream or downstream to the portfolio of economic activities produced at home by the MNC, I also test for dummies indicating those alternative classifications. The results can be found in Table 5 where each column uses a different dependent variable in the context of the regression discontinuity (using the optimal bandwidth and a triangular weighting scheme). As shown, the treatment effect is not statistically different from zero across all of these variables, reducing remaining concerns of other confounding factors driving the results. That is, time zone changes cannot explain differential patterns in terms of trade costs (Column 1), nor whether the foreign subsidiary is misclassified (Columns 2 to 4 , where $V$ stands for vertical, $U S$ for upstream and $D S$ for downstream, the latter two being subsets of the first one), nor the factor endowments of the country where the subsidiary is located (Columns 5, 6 and 7 ). The fact that the discontinuity is non-existent for other observable characteristics of the economic activity of the foreign subsidiary, and characteristics of the host country, reduces concerns that the main results are driven by confounding factors not accounted for.
[Table 5 about here.]
Another important aspect to consider is whether these results are robust to alternative measures of knowledge intensity. Figure 11 presents discontinuity plots and Table 6 estimates Specification (2), using alternative measures of

[^18]knowledge intensity as the dependent variable. Given that the used measure for knowledge intensity is not a common one, I test for whether the results are robust to alternative measures. First, in the column titled KI ${ }^{\text {Trimmed }}$ of Table 6. the estimation excludes uses the main knowledge intensity measure excluding industries that are below the 5th percentile and above the 95 th percentile of the distribution, to make sure the results are not being driven by outliers. The columns titled $R \& D(N T)$ and $R \& D(K Y)$ replicate the results using R\&D intensity computed using the Orbis dataset compiled by Nunn and Trefler (2008) and R\&D intensity computed using the Compustat dataset and compiled by Keller and Yeaple (2013), respectively. The column titled NPW Wages uses the share of non-production workers' wages in total payroll from the NBERCES Manufacturing Industry Database (Becker et al., 2013). The column titled Knowhow replicates results using another knowledge-intensity measure I constructed following the same steps outlined in Section 2.2 .2 but using as main input another question from $\mathrm{O}^{*} \mathrm{NET}$ : the extent to which workers in that industry "use and update relevant knowledge". Finally, the last column titled ProvSolv is another measure constructed, too, using a different input from O*NET: the intensity with which workers in that industry "make decisions and solve problems". Graphically, it is easy to notice in Figure 11, a statistically significant discontinuity in these measures on either side of a time zone line. In terms of the analytical results presented in Table 6, they are qualitatively consistent with the main results, though the magnitude of the point estimates -as expected- varies.
[Figure 11 about here.]
[Table 6 about here.]
A number of additional robustness tests are presented in Online Appendix Section F. In particular, I present results using a uniform weighting scheme and find that results are robust to this change (see Table F1). In addition, I reestimate the main RD specification including a more stringent set of fixed effects: one for every MNC and time zone line simultaneously ( $\varphi_{h} \times \tau_{t z}$ ). Table F2 presents this estimation which results in point estimates that -despite the sever lack of within-cell variation- are positive, thus reassuring, albeit smaller in magnitude and much less precise (e.g., weaker in terms of statistical significance). All in all, the different robustness tests show that the main RD results are not driven by peculiarities of the data, the choice of main right-hand-side variable or the specification used.

## 5 Concluding Remarks

In order to operate smoothly, MNCs make investments so that workers located in the headquarters and in foreign subsidiaries can work together to maintain uninterrupted business operations. "Working together" could imply many things, such as managing, coordinating, monitoring, troubleshooting, etc. One broad
way to denominate all these activities is under the umbrella of knowledge transmission. Knowledge transmission can certainly be costly and more intense for certain industries. These costs, of course, are an important aspect in determining location decisions of MNCs. This reasoning helps explain the existing trade-off between distance to the headquarters and the knowledge intensity level of a foreign subsidiary, a result consistent with the work of many others (e.g., Ramondo and Rodriguez-Clare, 2013, Ramondo, 2014, Arkolakis et al., 2013 Tintelnot, 2017; Keller and Yeaple, 2013).

The main contribution of this paper, however, is to show that this trade-off significantly weakens when time zones are taken into consideration: knowledge intensity is, on average, higher for roughly equidistant foreign subsidiaries that are closer to the headquarters in terms of in time zones. This result is at odds with models that suggest that larger intra-firm trade costs (between the headquarters and its knowledge intensive subsidiaries) can fully explain the aforementioned trade-off. Simply put, this result implies that the cost of shipping intermediate goods (or transporting people), 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 enough to explain the fact that MNCs tend to locate their foreign subsidiaries engaged in knowledge intensive activities geographically nearby.

This paper also contributes to the literature by adapting a regression discontinuity design that exploits multiple spatial differences across time zone lines in the context of researching the behavior of MNCs. To the best of my knowledge, this is the first time this methodology has been used in this context.

Nonetheless, the hope is that this paper has left open some other specific questions that will shed light on our general understanding of how the transmission of knowledge plays a role in the location decisions and the performance of MNCs. Moreover, the difficulties associated with transferring and acquiring knowledge, which translates into productivity shifts, are not unique to MNCs. They can also relate to purely domestic firms (e.g., Bloom et al., 2012; Kalnins and Lafontaine, 2013), and to investors (e.g., Coval and Moskowitz, 2001), among other activities. 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.

## 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, Pol Antràs, Davin Chor, and Paola Conconi. "Internalizing

Global Value Chains: A Firm-Level Analysis." Technical Report July, National Bureau of Economic Research, Cambridge, MA, 2015.

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.

Arkolakis, Costas, Natalia Ramondo, Andrés Rodríguez-Clare, and Stephen Yeaple. "Innovation and Production in the Global Economy." Technical Report 18792, National Bureau of Economic Research, Cambridge, MA, 2013.

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

Atalay, Enghin, Ali Hortaçsu, and Chad Syverson. "Vertical Integration and Input Flows." American Economic Review 104, 4: (2014) 1120-1148. http: //www. aeaweb.org/articles?id=10.1257/aer.104.4.1120.

Autor, David H, Frank Levy, and Richard J Murnane. "The Skill Content of Recent Technological Change: An Empirical Exploration." The Quarterly Journal of Economics 118, 4: (2003) 1279-1333.

Bahar, Dany. "Measuring knowledge intensity in manufacturing industries: a new approach." Applied Economics Letters 1-4.

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.

Bahar, Dany, and Hillel Rapoport. "Migration, Knowledge Diffusion and the Comparative Advantage of Nations." The Economic Journal 128, 612: (2018) F273-F305.

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

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott. "Trade costs, firms and productivity." Journal of Monetary Economics 53, 5: (2006) 917937.

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.

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

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. "Rdrobust: Software for regression-discontinuity designs." Stata Journal 17, 2: (2017) 372-404.

Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik. "Robust datadriven inference in the regression-discontinuity design." Stata Journal 14, 4: (2014) 909-946.

Campante, Filipe, and David Yanagizawa-Drott. "Long-Range Growth: Economic Development in the Global Network of Air Links." NBER Working Paper No. 22653: (2016) 50.

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.

Cattaneo, Matias D, Nicolas Idrobo, and Rocío Titiunik. "A Practical Introduction to Regression Discontinuity Designs." Cambridge Elements: Quantitative and Computational Methods for Social Science - Cambridge University Press I.

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.

Gumpert, Anna. "The Organization of Knowledge in Multinational Firms." Journal of the European Economic Association .

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

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.

Imbens, Gguido, and Karthik Kalyanaraman. "Optimal bandwidth choice for the regression discontinuity estimator." Review of Economic Studies 79, 3: (2012) 933-959.

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.

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.

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.

Oldenski, Lindsay. "Export Versus FDI and the Communication of Complex Information." Journal of International Economics 87, 2: (2012) 312-322.

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. Personal knowledge: Towards a post-critical philosophy. London, UK: Routledge, 1962.
——. The Tacit Dimension. Chicago; London: University of Chicago Press, 1966, 2009 edition.

Ramondo, Natalia. "A quantitative approach to multinational production." Journal of International Economics 93, 1: (2014) 108-122. http://dx.doi. org/10.1016/j.jinteco.2014.01.004.

Ramondo, Natalia, Veronica Rappoport, and Kim J Ruhl. "Intrafirm trade and vertical fragmentation in U.S. multinational corporations." Journal of International Economics .

Ramondo, Natalia, and Andrés Rodriguez-Clare. "Trade, Multinational Production, and the Gains from Openness." Journal of Political Economy 121, 2: (2013) 273-322.

Shirotori, M, B Tumurchudur, and O Cadot. "Revealed Factor Intensity Indices at the Product Level." Policy Issues in International Trade and Commodities 2010, 44.

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.

Tintelnot, Felix. "Global Production with Export Platforms." The Quarterly Journal of Economics 132, 1: (2017) 157-209.

Figure 1: 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 2: 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 3: 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 4: Distribution R\&D Measures


The figure shows the fitted distribution for the $R \& D$ Intensity measures used in the literature. The left panel corresponds to the industry-specific R\&D intensity computed using the Compustat dataset and compiled by Keller and Yeaple (2013) and the right panel corresponds to the same measure computed using the Orbis dataset and compiled by Nunn and Trefler (2008). Industries are defined in NAICS 4-digit industries.

Figure 5: Estimated relationship of $k$ and $d$


The figure presents the empirical fit for the relationship between $\log (k)$ and $\log (d)$ (the former proxied by the experience plus training measure), after controlling for unit trade costs and MNC fixed effects. The figure documents the estimation for all subsidiaries (panel 1), foreign subsidiaries (panel 2), foreign horizontal subsidiaries (panel 3), and foreign vertical subsidiaries (panel 4 ). $95 \%$ confidence intervals marked in grey dashed lines.

Figure 6: Distance and working hours in the sample


The figure plots, for each one of the foreign subsidiaries in the sample, the geographic distance to and the overlap in working hours with its global headquarters.

Figure 7: Knowledge and distance trade-off, by working hours overlap









- Point Estimate $\longmapsto 90 \% \mathrm{Cl}$


The figure presents the empirical fit for the relationship between $\log (d)$ and $\log (k)$ (the latter proxied by the experience plus training measure), after controls specified in Specification (1), for different levels of overlap in working hours between the headquarters and its affiliate. Each row presents results using different cuts of the sample (all, foreign, horizontal and vertical). The left panel shows the value of the slope, adjusted for different levels of the working hours overlap; when the overlap is 0 the slope of the trade-off is highly negative, and as the overlap becomes larger, the trade-off weakens. The right panel plots the aver3ge knowledge-distance trade-off based on different levels of working hours overlap between headquarters and foreign subsidiaries.

Figure 8: Running variable definition


The figure presents a graphical representation to understand how the running variable (distance to the nearest time zone line) is centered around the time zone line.

Figure 9: Regression discontinuity plots


The figure presents a graphical representation of the regression discontinuity design in Specification 2 for different cuts of the sample (all, foreign, horizontal and vertical subsidiaries). Bin sizes are defined on the "mimicking variance evenly-spaced" method (denoted by emsv in the rdplot package).

Figure 10: RD plots (Horizontal)


The figure presents a graphical representation of the regression discontinuity design in Specification 2 for foreign horizontal subsidiaries. The left panel presents a linear estimation while the right panel plots a quadratic one. Bin sizes are defined on the "mimicking variance evenly-spaced" method (denoted by emsv in the rdplot package).

Figure 11: RD plots, alt. knowledge intensity measures


The figure presents a graphical representation of the regression discontinuity design in Specification 2 for the sub-sample of foreign horizontal subsidiaries, using alternative knowledge intensity measures of its economic activity. The knowledge intensity measures include: (i) a trimmed version of the main knowledge intensity measure used throughout the paper excluding industries below the 5 th and above the 95 th percentile of the distribution (lncumexpWagetr); (ii) R\&D intensity computed using the Orbis dataset compiled by Nunn and Trefler (2008) (lnRDNT); (iii) R\&D intensity computed using the Compustat dataset and compiled by Keller and Yeaple (2013) (lnRDKY); (iv) share of non-production workers' wages in total payroll (lnnpwshareWage); (v) a knowledge-intensity measure I construct following the same steps outlined in Section 2.2 .2 using another input from O*NET: the extent to which workers in that industry "use and update relevant knowledge" (lnknowhow); and (vi) a knowledge intensity measure constructed, too, following the same steps outlined in Section 2.2 .2 using a different input from $\mathrm{O}^{*} \mathrm{NET}$ : the intensity with which workers in that industry "make decisions and solve problems" (lnprobsolv). All knowledge intensity variables have been transformed into logarithmic form. Bin sizes are defined on the "mimicking variance evenly-spaced" method (denoted by emsv in the rdplot package).

Table 1: Descriptive Statistics (Domestic Vs. Foreign Subsidiaries)

|  | MNC | \# Subs | Foreign (\%) | KI $_{\text {Foreign }}$ | KI $_{\text {Domestic }}$ | $\Delta$ |
| :--- | ---: | ---: | ---: | :---: | :---: | :---: |
| Panel A: Domestic vs. Foreign Affiliates |  |  |  |  |  |  |
| All Observations | 3229 | 111172 | .39 | .3 | .35 | $-.051^{* * *}$ |
| Non OECD | 121 | 3469 | .2 | .073 | .38 | $-.31^{* * *}$ |
| OECD | 3108 | 107703 | .4 | .3 | .35 | $-.046^{* * *}$ |
| East Asia \& Pacific | 623 | 26016 | .17 | .45 | .44 | .012 |
| Latin America \& Caribbean | 47 | 2975 | .59 | -.36 | -.27 | $-.088^{* * *}$ |
| North America | 1056 | 39826 | .29 | .31 | .38 | $-.072^{* * *}$ |
| South Asia | 70 | 2123 | .16 | .36 | .49 | $-.13^{* * *}$ |
| Western Europe | 1433 | 40232 | .63 | .31 | .2 | $.12^{* * *}$ |
| The table presents descriptive statistics from the sample. It presents for different cuts of the sample, |  |  |  |  |  |  |

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 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 $\Delta$. Stars represent statistical significance of the difference: ${ }^{*} p<0.10,{ }^{* *} p<0.05,{ }^{* * *} p<0.01$

Table 2: Summary statistics for subsidiaries by sub-sample

| Variable | Mean | sd | Min | Max |
| :--- | :---: | :---: | :---: | :---: |
| Sample: All (N=105183; MNCs=3145) |  |  |  |  |
| Distance, km (log) | 6.605 | 2.12 | 0.0 | 9.9 |
| Knowledge Intensity (log) | 4.128 | 0.14 | 3.6 | 4.4 |
| Unit Shipping Cost (log) | -2.048 | 0.66 | -3.6 | -0.1 |
| Working Hours Overlap | 8.608 | 2.11 | 0.0 | 10.0 |
| Sample: Foreign (N=40490; $M N C s=1902)$ |  |  |  |  |
| Distance, km (log) | 8.201 | 1.15 | 1.1 | 9.9 |
| Knowledge Intensity (log) | 4.120 | 0.14 | 3.6 | 4.4 |
| Unit Shipping Cost (log) | -1.969 | 0.70 | -3.6 | -0.1 |
| Working Hours Overlap | 6.967 | 2.45 | 0.0 | 10.0 |
| Sample: Horizontal (N=25265; MNCs=1204) |  |  |  |  |
| Distance, km (log) | 8.204 | 1.15 | 1.1 | 9.9 |
| Knowledge Intensity (log) | 4.112 | 0.13 | 3.6 | 4.4 |
| Unit Shipping Cost (log) | -1.895 | 0.76 | -3.6 | -0.1 |
| Working Hours Overlap | 6.962 | 2.46 | 0.0 | 10.0 |
| Sample: Vertical (N=1973; | $M N C s=292)$ |  |  |  |
| Distance, km (log) | 8.154 | 1.25 | 2.5 | 9.8 |
| Knowledge Intensity (log) | 4.157 | 0.15 | 3.6 | 4.4 |
| Unit Shipping Cost (log) | -2.170 | 0.58 | -3.6 | -0.9 |
| Working Hours Overlap | 6.936 | 2.59 | 0.0 | 10.0 |

The table presents descriptive statistics from the sample of subsidiaries for distance in between the headquarters and the subsidiary, the knowledge intensity of the foreign subsidiary, the unit trade cost and the overlap in working hours. The statistics are presented for different cuts of the sample used in the analysis: all subsidiaries, foreign subsidiaries, foreign horizontal subsidiaries and foreign vertical subsidiaries.

Table 3: Distance, knowledge intensity and overlap in working hours

| Dependent Variable: Log Knowledge Intensity |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
|  | $(1)$ | $(2)$ | $(3)$ | $(4)$ |
| $\log (\mathrm{d})$ | -0.0151 | -0.0214 | -0.0199 | -0.0672 |
|  | $(0.006)^{* *}$ | $(0.010)^{* *}$ | $(0.009)^{* *}$ | $(0.021)^{* * *}$ |
| $\log (\mathrm{~d})$ Xwhoursoverlap | 0.0015 | 0.0019 | 0.0017 | 0.0062 |
|  | $(0.001)^{* *}$ | $(0.001)^{* *}$ | $(0.001)^{*}$ | $(0.002)^{* * *}$ |
| whoursoverlap | -0.0138 | -0.0192 | -0.0166 | -0.0603 |
|  | $(0.006)^{* *}$ | $(0.009)^{* *}$ | $(0.008)^{* *}$ | $(0.019)^{* * *}$ |
| $\log (\mathrm{t})$ | -0.0961 | -0.0872 | -0.0801 | -0.1059 |
|  | $(0.008)^{* * *}$ | $(0.008)^{* * *}$ | $(0.012)^{* * *}$ | $(0.019)^{* * *}$ |
| Constant | 4.0705 | 4.1509 | 4.1451 | 4.5536 |
|  | $(0.057)^{* * *}$ | $(0.090)^{* * *}$ | $(0.082)^{* * *}$ | $(0.195)^{* * *}$ |
| N |  |  |  |  |
| Adj. R2 | 105183 | 40490 | 25265 | 1973 |
| $\varphi_{h}$ | 0.81 | 0.81 | 0.88 | 0.91 |

The table presents results for the estimation of Specification 1) using different cuts of the sample of MNCs' subsidiaries, including controls and interactions of distance between the subsidiary and its global headquarters (in logs) and the overlap in working hours between them. The left hand side variable is the knowledge intensity of the economy activity of the foreign subsidiary (in logs). The estimation controls for the unit shipping cost (in logs) of the good produced by the subsidiary. All specifications include MNC fixed effects. Robust standard errors clustered at the industry and MNC level are presented in parentheses. ${ }^{*} p<0.10,{ }^{* *} p<0.05,{ }^{* * *} p<0.01$

Table 4: RD estimation

| Dependent Variable: Foreign subsidiary's knowledge intensity (log) |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
|  | 150 | 250 | 350 | Optimal |
| closerTZ | 0.0084 | 0.0060 | 0.0075 | 0.0066 |
|  | $(0.003)^{* *}$ | $(0.003)^{*}$ | $(0.002)^{* * *}$ | $(0.002)^{* * *}$ |
| distTZzero100s | -0.0034 | -0.0028 | -0.0015 | -0.0018 |
|  | $(0.001)^{* * *}$ | $(0.001)^{* * *}$ | $(0.000)^{* * *}$ | $(0.000)^{* * *}$ |
| closerTZ $\times$ distTZzero100s | -0.0020 | 0.0024 | -0.0004 | 0.0005 |
|  | $(0.003)$ | $(0.002)$ | $(0.001)$ | $(0.001)$ |
| Constant | 4.1045 | 4.1020 | 4.1047 | 4.1055 |
|  | $(0.001)^{* * *}$ | $(0.001)^{* * *}$ | $(0.001)^{* * *}$ | $(0.001)^{* * *}$ |
| N |  |  |  |  |
| Adj. R2 | 5232 | 9702 | 14179 | 12367 |

The table presents results for the estimation of Specification 2 using a sample of foreign horizontal subsidiaries of MNCs estimated using several bandwidths for the running variable specified in each column. The last column uses the optimal bandwidth computed using the methodology described in Cattaneo et al. (2018) who build on the work by Imbens and Kalyanaraman (2012). The estimation uses a triangular weight scheme, giving higher weight to observations closer to the cutoff point. All specifications include MNC fixed effects and time zone line fixed effects. Robust standard errors clustered at the MNC and time zone line level are presented in parentheses.
${ }^{*} p<0.10,{ }^{* *} p<0.05,{ }^{* * *} p<0.01$
Table 5: RD estimation, various observable variables

| Dependent Variable: Various observable variables |  |  |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\log \left(t_{s}\right)$ | $V_{s}$ | $U S_{s}$ | $D S_{s}$ | $\log \left(G D P p c_{s}\right)$ | $\log \left(K p w_{s}\right)$ | $\log \left(H p w_{s}\right)$ |
| closerTZ | -0.0240 | 0.0214 | 0.0246 | 0.0186 | 0.0028 | -0.0359 | -0.0114 |
|  | $(0.018)$ | $(0.018)$ | $(0.017)$ | $(0.012)$ | $(0.028)$ | $(0.052)$ | $(0.022)$ |
| distTZzero100s | 0.0082 | -0.0104 | -0.0116 | -0.0051 | -0.0255 | -0.0070 | -0.0095 |
|  | $(0.004)^{*}$ | $(0.005)^{* *}$ | $(0.005)^{* *}$ | $(0.003)^{*}$ | $(0.012)^{* *}$ | $(0.019)$ | $(0.004)^{* *}$ |
| closerTZ $\times$ distTZzero100s | -0.0086 | 0.0159 | 0.0177 | -0.0004 | 0.0164 | -0.0255 | -0.0005 |
|  | $(0.005)^{*}$ | $(0.008)^{* *}$ | $(0.008)^{* *}$ | $(0.004)$ | $(0.012)$ | $(0.027)$ | $(0.007)$ |
| Constant | -1.8530 | 0.2669 | 0.2545 | 0.1372 | 10.3290 | 11.7521 | 2.2462 |
|  | $(0.006)^{* * *}$ | $(0.011)^{* * *}$ | $(0.010)^{* * *}$ | $(0.004)^{* * *}$ | $(0.014)^{* * *}$ | $(0.035)^{* * *}$ | $(0.010)^{* * *}$ |
| N |  |  |  |  |  | 12309 | 12323 |
| Adj. R2 | 12006 | 12367 | 12367 | 12367 | 12323 |  |  |

The table presents results for the estimation of Specification 2] using as dependent variable a number of covariates linked to the economic activity of the foreign horizontal subsidiary or its host country. They are: unit shipping costs of the economic activity of the subsidiary (Column 1); whether the subsidiary also classifies as vertical (Column 2), whether as upstream (Column 3) or as downstream (Column 4); as well as variables defined at the level of the host country of the foreign affiliate: income per capita (Column 5), physical capital per worker (Column 6), and human capital per worker (Column 7). The fact that the discontinuity is non-existent for other observable characteristics of the economic activity of the foreign subsidiary, and characteristics of the host country, reduces concerns that the main results are driven by confounding factors not accounted for. The sample is based on the optimal bandwidth computed using the methodology described in Cattaneo et al. (2018) who build on the work by Imbens and Kalyanaraman (2012). The estimation uses a triangular weight scheme, giving higher weight to observations closer to the cutoff point. All specifications include MNC fixed effects and time zone line fixed effects. Robust standard errors clustered at the MNC and time zone line level are presented in parentheses.
${ }^{*} p<0.10,{ }^{* *} p<0.05,{ }^{* * *} p<0.01$
Table 6: RD estimation, alt. knowledge intensity measures

| Dependent Variable: Knowledge intensity measures |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $K I^{\text {Trimmed }}$ | $R \& D(N T)$ | $R \& D(K Y)$ | $N P W^{\text {Wages }}$ | Knowhow | ProbSolv |
| closerTZ | $\begin{gathered} 0.0053 \\ (0.001)^{* * *} \end{gathered}$ | $\begin{gathered} 0.0013 \\ (0.001)^{* *} \end{gathered}$ | $\begin{gathered} 0.0049 \\ (0.002)^{* *} \end{gathered}$ | $\begin{gathered} 0.0059 \\ (0.003)^{*} \end{gathered}$ | $\begin{gathered} 0.0036 \\ (0.002)^{* *} \end{gathered}$ | $\begin{gathered} 0.0035 \\ (0.001)^{* * *} \end{gathered}$ |
| distTZzero100s | $\begin{gathered} -0.0014 \\ (0.001)^{* *} \end{gathered}$ | $\begin{aligned} & -0.0002 \\ & (0.000) \end{aligned}$ | $\begin{gathered} -0.0011 \\ (0.001)^{*} \end{gathered}$ | $\begin{aligned} & -0.0016 \\ & (0.001) \end{aligned}$ | $\begin{gathered} -0.0012 \\ (0.000)^{* *} \end{gathered}$ | $\begin{gathered} -0.0015 \\ (0.000)^{* * *} \end{gathered}$ |
| closerTZ $\times$ distTZzero100s | $\begin{aligned} & 0.0002 \\ & (0.001) \end{aligned}$ | $\begin{aligned} & -0.0003 \\ & (0.000) \end{aligned}$ | $\begin{aligned} & 0.0002 \\ & (0.001) \end{aligned}$ | $\begin{aligned} & -0.0001 \\ & (0.001) \end{aligned}$ | $\begin{aligned} & 0.0001 \\ & (0.001) \end{aligned}$ | $\begin{aligned} & 0.0006 \\ & (0.001) \end{aligned}$ |
| Constant | $\begin{gathered} 4.1063 \\ (0.001)^{* * *} \end{gathered}$ | $\begin{gathered} 0.0204 \\ (0.000)^{* * *} \end{gathered}$ | $\begin{gathered} 0.0376 \\ (0.001)^{* * *} \end{gathered}$ | $\begin{gathered} 0.4154 \\ (0.001)^{* * *} \end{gathered}$ | $\begin{gathered} 1.3489 \\ (0.001)^{* * *} \end{gathered}$ | $\begin{gathered} 1.3727 \\ (0.001)^{* * *} \end{gathered}$ |
| N | 11264 | 13696 | 11934 | 12277 | 10184 | 9208 |
| Adj. R2 | 0.87 | 0.77 | 0.69 | 0.82 | 0.87 | 0.88 |

The table presents results for the estimation of Specification 2] using as dependent variable alternative measures of knowledge intensity of the economic activity of the foreign horizontal subsidiary. The knowledge intensity measures are: (i) a trimmed version of the main knowledge intensity measure used throughout the paper excluding industries below the 5 th and above the 95 th percentile of the distribution; (ii) R\&D intensity computed using the Orbis dataset compiled by Nunn and Trefler (2008); (iii) R\&D intensity computed using the Compustat dataset and compiled by Keller and Yeaple (2013); (iv) share of non-production workers' wages in total payroll; (v) a measure of the extent to which workers in that industry "use and update relevant knowledge"; and (vi) a measure of the intensity with which workers in that industry "make decisions and solve problems". The sample is based on the optimal bandwidth computed using the methodology described in Cattaneo et al. (2018) who build on the work by Imbens and Kalyanaraman (2012).
 include MNC fixed effects and time zone line fixed effects. Robust standard errors clustered at the MNC and time zone line level are presented in parentheses.
${ }^{*} p<0.10,{ }^{* *} p<0.05,{ }^{* * *} p<0.01$

# Appendix for <br> The hardships of long distance <br> relationships: time zone proximity and the location decisions of MNC's knowledge-intensive activities 

August 1, 2019

## A On the Dun \& Bradstreet Dataset

## A. 1 Reliability

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" ${ }^{1}$ 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), however, compare the dataset with the US multinational firms sample by the US Bureau of Economic Analysis, and find consistencies between the two datasets.

Some basic relationships drawn from the sample 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, as can be seen in Figure A1. In particular, the figure shows the relationship between the size of MNC firms (in number of establishments in the left panel, and in total number of employees in the right pane ${ }^{2}$ 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.

## [Figure A1 about here.]

Focusing the analysis on the within-MNC dimension, adding MNC fixed effects to all specifications significantly diminishes the sampling concerns. This is because while methods for gathering information may not be symmetric across

[^19]countries, they would not systematically differ by firm. It is reasonable to assume that the per-country likelihood of missing data would be the same for all firms, controlling for the location of the headquarters of the MNC. Thus, concerns regarding biases caused by possible sampling asymmetries are not particularly large for the purpose of this empirical exercise.

Figure A1: 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).

## A. 2 Industries

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.]

I also present results on the distribution of sectors among foreign affiliates, to understand whether in the sample there are some sectors that are more likely to appear or be reported than others. In terms of industries, the distribution of sectors in the sample is not homogenous, as can be seen in Figure A2 Some sectors are more prevalent than others. 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 A2 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). Figure A3 shows that larger MNCs (as measured by number of affiliates) tend to make a larger number of different products.
[Figure A3 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: 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 continuous, what explains the zero values in the figure.

Figure A3: 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.

## B $\mathrm{O}^{*}$ NET knowledge intensity measure

This section describes characteristics of the knowledge intensity measure based on Bahar (2018). Figure B1 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 for workers in each industry). As opposed to the $\mathrm{R} \& \mathrm{D}$ investment based variables used in the literature, the distribution of the $\mathrm{O}^{*}$ NET based variables is smoother, and behaves more like a normal probability density function. Figure B2 presents the same graphs limiting the sample to manufacturing industries only.
[Figure B1 about here.]
[Figure B2 about here.]
Tables B1 presents the top and bottom ten products in the manufacturing division (SIC codes 2000 to 3999) ranked by the knowledge intensity measure.
[Table B1 about here.]

Figure B1: Histogram O*NET-based KI (All Industries)


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

Figure B2: Histogram O*NET-based KI (Manufacturing Only)


The figure shows the fitted distribution for the computed "experience plus training" O*NET-based knowledge intensity measure for manufacturing industries only. Industries are defined in SIC 1987 4 -digit industries.
Table B1: 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 |

## C Foreign expansion of knowledge-intensive activities

Perhaps a more rigorous approach to understand the patterns seen in Table 1 is to explore partial correlations in the sample. As described above, two of the main results that emerge from models that incorporate the cost of transferring knowledge within MNCs are that (i) foreign expansion of knowledge-intensive activities is more costly, and therefore less likely, and (ii) there is a trade-off between distance and knowledge intensity of the economic activity of the foreign subsidiary. In order to test for these expected results I estimate the following specification:

$$
\begin{equation*}
\text { Foreign }_{h, s}=\beta_{k} \cdot \log \left(k_{s}\right) \times \log \left(d_{s}\right)+\beta_{t} \cdot \log \left(t_{s}\right)+\varphi_{h}+\eta_{s}+e_{h, s} \tag{C1}
\end{equation*}
$$

Where $s$ indexes a subsidiary and $h$ its headquarters. The independent variable is a dummy which takes the value 1 if the subsidiary is a foreign 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. Importantly, the specification also includes $\varphi_{h}$, which represents MNC fixed effects. That is, the estimations use only the variation within each MNC in the sample, across its different subsidiaries. Thus, it controls the overall productivity level of the MNC. Naturally, different subsidiaries within a single MNC might differ in their reported 4-digit economic activity, thus allowing for within-firm variation in the right hand side variables of the empirical specification (see Online Appendix Section A. 2 for more information on the within-firm variation in economic activity). $\eta_{s}$ represents a fixed effect for the country where the subsidiary is located. $e_{h, s}$ is the error term.

I estimate Specification (C1 using a different sub-sample of foreign subsidiaries in every estimation. First, using all foreign subsidiaries; second, using only horizontal foreign subsidiaries; and third, using only vertical foreign subsidiaries. The nature of the specification also allows to estimate whether these foreign expansions are located at closer geographic distance whenever they are in knowledge-intensive economic activities, through the interaction term. The results are presented in Table C1.
[Table C1 about here.]
The results in Columns 1, 3 and 5 show mixed evidence that the expansion abroad is less likely for knowledge intensive activities. It is only in Column 3 that we find such result, when focusing only on horizontal foreign subsidiaries, consistently with what we would expect from models that incorporate the cost of transferring knowledge in the proximity-concentration hypothesis. Column 2 suggests that, everything else equal, industries for which their knowledge intensity is $10 \%$ higher are about 1.2 percentage points less likely to be replicated
abroad. However, according to Column 5, vertical foreign subsidiaries tend to be active in economic activities that are more knowledge intensive than their domestic counterparts. However, the results of the even columns (2, 4 and 6 ) control for geographic distance of each subsidiary to its global headquarters. There, the results of Column 4 that focuses on foreign horizontal subsidiaries remain consistent with the proximity-concentration hypothesis: knowledge intensive activities are less likely to be expanded abroad and, according to the $\log (d) \times \log (k)$ interaction, subsidiaries that are active in knowledge intensive tend to locate closer to the headquarters. The evidence of this when expanding the sample to all foreign subsidiaries (Column 2) or limiting to only vertical foreign subsidiaries (Column 6) is inconclusive ${ }^{3}$

It is important to mention that the point estimate for the trade cost variable is negative, though not statistically significant. We would have expected a positive estimate according to the proximity-concentration hypothesis, at least for Columns 3 and 4 . Yet, it is important to notice that this unexpected result is probably due to the inclusion of multinational fixed effects, given that most of the variation of the trade cost variable is across MNCs and not within. Thus, the inclusion of MNCs fixed effects comes at the price of losing the ability to exploit the variation of this variable. Alternative standardizations of the trade cost variable often provide positive point estimates, though still statistically insignificant. With respect to the controls defining whether the subsidiary is horizontal or represents a vertical linkage, besides allowing for different constants, there is not a straightforward interpretation for them.

These results are also consistent when focusing on the intensive margin: instead of using a dummy on the left hand side defining whether the subsidiary is foreign, I use the subsidiary's number of employees, only for foreign subsidiaries, after logarithmic transformation. Note that the number of employees in the dataset is noisy and often non-existent, which explains the much lower number of observations. The results are presented in Table C2. While it is reassuring that the point estimates are consistent in terms of sign with the main results, they often lack statistical significance.
[Table C2 about here.]
While the results are consistent with what we would expect, we must acknowledge the limitations of this estimation: first and foremost, the absence of exports by each MNC-product combination to particular destinations as a control. This is particularly important for horizontal expansions given that without data on exports by the same MNC to a given location, it is very difficult to say something precise about the proximity-concentration trade-off (as it is impossible to know whether the lack of horizontal expansion is happening in a place where the same product is being exported to). Yet, I take this as an exercise that sheds light on whether there is a differential pattern in terms of foreign expansion with respect to knowledge intensity, with the value added that it uses

[^20]a worldwide dataset.
Below, however, I follow Keller and Yeaple (2013) and use aggregated data at the industry level on exports and sales of foreign affiliates of US companies, and show consistent results with my measure of knowledge intensity.

## Proxying for demand: knowledge transmission in the proximityconcentration trade-off

The results on the determinants of horizontal expansion are not fully satisfactory since they do not account for the demand factor (i.e. it is not possible to see all locations where the MNCs faced a trade-off between exports and foreign affiliates, and decided on the former). However, in order to explore the role of knowledge intensity in the likelihood of horizontal expansion we follow the guidelines of Helpman, Melitz and Yeaple (2004) using BEA data on American MNCs for years 1999-2011 to show that sales of foreign affiliates decreases with the knowledge intensity level of the industry (using my own measures of knowledge intensity), after controlling for export volume for those industries. That is, by including exports in the specification I aim to control for sector specific demand of American exporters in that location. Thus, the specification I estimate is:

$$
\begin{equation*}
\log \left(\text { ForeignSales }_{s, y}\right)=\beta_{k} \log \left(k_{s}\right)+\beta_{t} \log \left(t_{s}\right)+\log \left(\text { Exports }_{s, y}\right)+\alpha_{y}+\varepsilon_{s, y} \tag{C2}
\end{equation*}
$$

Where $s$ indexes for industry and $y$ for year. The dependent variable is the sales of foreign affiliates of US multinationals in the rest of the world for sector $s$, in millions of dollars. The right hand side includes the knowledge intensity (denoted by $k$ ) and the unit shipping cost (denoted by $t$ ) of sector $s$. It also includes US export volumes of industry $s$ to the world in millions of dollars. There are in total seven sectors, and each is defined as a 3 -digit NAICS code. The specification also includes year dummies (denoted by $\alpha_{y}$ ). The results of this estimation are presented in Table C3.
[Table C3 about here.]
Column 1 estimates a linear regression where the dependent variable is the sales of foreign affiliates (in logs), controlling for total exports to the same destination in the same industry code. Column 2 estimates a linear regression where the dependent variable is the ratio of sales of foreign affiliates to exports, similarly to Helpman, Melitz and Yeaple (2004). Column 3 replicates Column 2 but excludes outliers in the sales to exports distribution. It can be seen that the estimator for $\beta_{k}$ is negative and statistically significant, regardless of whether the specification uses the total sales of foreign affiliates controlling for exports (Column 1), or the ratio of foreign sales to exports (Column 2 and 3). Similarly, $\beta_{t}$ is estimated to be positive and significant, as expected (i.e. industries with larger trade cost will generate incentives for firms to create foreign affiliates abroad to substitute for exports). Figure C1 plots the estimation of $\beta_{k}$ in the first column of Table C3.
[Figure C1 about here.]

Figure C1: Sales of Foreign Affiliates for American MNCs and Knowledge Intensity


Note: estimates control for trade costs and year fixed effects
The vertical axis is the log-sales of foreign affiliates and the horizontal axis is the log knowledge intensity measure.

Table C1: Determinants of foreignness

| Dependent Variable: Foreign Subsidiary (Binary Variable) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | All |  | Horizontal |  | Vertical |  |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| $\log (\mathrm{k})$ | $\begin{aligned} & 0.0268 \\ & (0.061) \end{aligned}$ | $\begin{aligned} & \hline 0.1138 \\ & (0.084) \end{aligned}$ | $\begin{gathered} -0.1213 \\ (0.069)^{*} \end{gathered}$ | $\begin{gathered} 0.2369 \\ (0.097)^{* *} \end{gathered}$ | $\begin{gathered} 0.1410 \\ (0.040)^{* * *} \end{gathered}$ | $\begin{aligned} & \hline 0.0224 \\ & (0.057) \end{aligned}$ |
| $\log (\mathrm{k}) \mathrm{Xlog}(\mathrm{d})$ |  | $\begin{aligned} & -0.0128 \\ & (0.014) \end{aligned}$ |  | $\begin{gathered} -0.0506 \\ (0.016)^{* * *} \end{gathered}$ |  | $\begin{aligned} & 0.0197 \\ & (0.013) \end{aligned}$ |
| $\log (\mathrm{d})$ |  | $\begin{gathered} 0.1253 \\ (0.057)^{* *} \end{gathered}$ |  | $\begin{gathered} 0.2851 \\ (0.065)^{* * *} \end{gathered}$ |  | $\begin{gathered} -0.0654 \\ (0.051) \end{gathered}$ |
| $\log (\mathrm{t})$ | $\begin{aligned} & -0.0080 \\ & (0.013) \end{aligned}$ | $\begin{aligned} & -0.0096 \\ & (0.011) \end{aligned}$ | $\begin{aligned} & -0.0097 \\ & (0.021) \end{aligned}$ | $\begin{aligned} & -0.0085 \\ & (0.017) \end{aligned}$ | $\begin{aligned} & 0.0084 \\ & (0.007) \end{aligned}$ | $\begin{aligned} & 0.0068 \\ & (0.007) \end{aligned}$ |
| Horizontal Subsidiary | $\begin{gathered} 0.5775 \\ (0.025)^{* * *} \end{gathered}$ | $\begin{gathered} 0.4597 \\ (0.025)^{* * *} \end{gathered}$ |  |  |  |  |
| Vertical Linkage | $\begin{gathered} 0.2292 \\ (0.024)^{* * *} \end{gathered}$ | $\begin{gathered} 0.1762 \\ (0.020)^{* * *} \end{gathered}$ | $\begin{gathered} 0.6168 \\ (0.038)^{* * *} \end{gathered}$ | $\begin{gathered} 0.4800 \\ (0.032)^{* * *} \end{gathered}$ |  |  |
| Constant | $\begin{aligned} & 0.1040 \\ & (0.237) \end{aligned}$ | $\begin{gathered} -0.7023 \\ (0.341)^{* *} \end{gathered}$ | $\begin{gathered} 0.7188 \\ (0.259)^{* * *} \end{gathered}$ | $\begin{gathered} -1.2275 \\ (0.390)^{* * *} \end{gathered}$ | $\begin{gathered} -0.5303 \\ (0.155)^{* * *} \end{gathered}$ | $\begin{aligned} & -0.1355 \\ & (0.232) \end{aligned}$ |
| N | 105159 | 105158 | 89182 | 89181 | 65285 | 65285 |
| Adj. R2 | 0.74 | 0.79 | 0.63 | 0.71 | 0.55 | 0.57 |
| $\varphi_{h}$ | Y | Y | Y | Y | Y | Y |
| $\eta_{s}$ | Y | Y | Y | Y | Y | Y |

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 and other controls. Each pair of columns uses a different sample of foreign subsidiaries and all domestic ones. All specifications include MNC fixed effects. Robust standard errors clustered at the industry and MNC level are presented in parentheses. ${ }^{*} p<0.10,{ }^{* *} p<0.05,{ }^{* * *} p<0.01$

Table C2: Determinants of foreignness, size of affiliate

| Dependent Variable: Number of employees (log) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | All |  | Horizontal |  | Vertical |  |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| $\log (\mathrm{k})$ | $\begin{gathered} -0.7012 \\ (0.398)^{*} \end{gathered}$ | $\begin{aligned} & 2.9002 \\ & (2.010) \end{aligned}$ | $\begin{gathered} \hline-0.6358 \\ (0.871) \end{gathered}$ | $\begin{aligned} & \hline 3.8366 \\ & (2.761) \end{aligned}$ | $\begin{gathered} 0.8216 \\ (1.336) \end{gathered}$ | $\begin{gathered} 9.9685 \\ (11.047) \end{gathered}$ |
| $\log (\mathrm{k}) \mathrm{X} \log (\mathrm{d})$ |  | $\begin{gathered} -0.4330 \\ (0.253)^{*} \end{gathered}$ |  | $\begin{aligned} & -0.5428 \\ & (0.351) \end{aligned}$ |  | $\begin{gathered} -1.0609 \\ (1.240) \end{gathered}$ |
| $\log (\mathrm{d})$ |  | $\begin{gathered} 1.7266 \\ (1.029)^{*} \end{gathered}$ |  | $\begin{aligned} & 2.1719 \\ & (1.427) \end{aligned}$ |  | $\begin{aligned} & 4.3656 \\ & (5.000) \end{aligned}$ |
| $\log (\mathrm{t})$ | $\begin{gathered} -0.2533 \\ (0.108)^{* *} \end{gathered}$ | $\begin{gathered} -0.2545 \\ (0.107)^{* *} \end{gathered}$ | $\begin{gathered} -0.2772 \\ (0.249) \end{gathered}$ | $\begin{aligned} & -0.2767 \\ & (0.246) \end{aligned}$ | $\begin{aligned} & 0.2539 \\ & (0.329) \end{aligned}$ | $\begin{aligned} & 0.2186 \\ & (0.319) \end{aligned}$ |
| Downstream Linkage | $\begin{aligned} & 0.0114 \\ & (0.117) \end{aligned}$ | $\begin{aligned} & 0.0203 \\ & (0.117) \end{aligned}$ | $\begin{gathered} -0.0319 \\ (0.165) \end{gathered}$ | $\begin{array}{r} -0.0129 \\ (0.168) \end{array}$ | $\begin{aligned} & 0.1767 \\ & (0.346) \end{aligned}$ | $\begin{aligned} & 0.1478 \\ & (0.338) \end{aligned}$ |
| Upstream Linkage | $\begin{gathered} 0.2521 \\ (0.093)^{* * *} \end{gathered}$ | $\begin{gathered} 0.2522 \\ (0.090)^{* * *} \end{gathered}$ | $\begin{gathered} 0.3040 \\ (0.139)^{* *} \end{gathered}$ | $\begin{gathered} 0.2944 \\ (0.140)^{* *} \end{gathered}$ | $\begin{gathered} 1.2671 \\ (0.384)^{* * *} \end{gathered}$ | $\begin{gathered} 1.2651 \\ (0.384)^{* * *} \end{gathered}$ |
| Horizontal Subsidiary | $\begin{aligned} & 0.0373 \\ & (0.066) \end{aligned}$ | $\begin{aligned} & 0.0359 \\ & (0.065) \end{aligned}$ |  |  |  |  |
| Constant | $\begin{gathered} 5.5037 \\ (1.509)^{* * *} \end{gathered}$ | $\begin{aligned} & -8.8495 \\ & (8.232) \end{aligned}$ | $\begin{aligned} & 5.2169 \\ & (3.271) \end{aligned}$ | $\begin{aligned} & -12.6607 \\ & (11.256) \end{aligned}$ | $\begin{aligned} & -0.6408 \\ & (5.034) \end{aligned}$ | $\begin{aligned} & -38.3045 \\ & (44.496) \end{aligned}$ |
| N | 17330 | 17330 | 11069 | 11069 | 654 | 654 |
| Adj. R2 | 0.31 | 0.31 | 0.32 | 0.32 | 0.24 | 0.24 |
| $\varphi_{h}$ | Y | Y | Y | Y | Y | Y |
| $\eta_{s}$ | Y | Y | Y | Y | Y | Y |

The left hand side variable is the number of employees of the foreign subsidiary (in logs). The variables in the right hand side include the unit shipping cost associated with the industry, knowledge intensity measures and other controls. Each pair of columns uses a different sample of foreign subsidiaries, excluding domestic ones. All specifications include MNC fixed effects. Robust standard errors clustered at the industry and MNC level are presented in parentheses. ${ }^{*} p<0.10,{ }^{* *} p<0.05,{ }^{* * *} p<0.01$

Table C3: Sales of Foreign Affiliates for American MNCs and Knowledge Intensity

| Dependent Variable: Log Sales of Foreign Affiliates |  |  |  |
| :--- | :---: | :---: | :---: |
|  | $\log ($ SalesFA $)$ | Ratio | Ratio (no outliers) |
| $\log (\mathrm{k})$ | -1.2664 | -1.5221 | -1.0200 |
|  | $(0.293)^{* * *}$ | $(0.654)^{* *}$ | $(0.300)^{* * *}$ |
| $\log (\mathrm{t})$ | 0.7290 | 1.6101 | 1.7935 |
|  | $(0.105)^{* * *}$ | $(0.230)^{* * *}$ | $(0.096)^{* * *}$ |
| $\log (\exp )$ | 1.1858 |  |  |
|  | $(0.048)^{* * *}$ |  |  |
| Constant | 5.3929 | 11.9843 | 10.2212 |
|  | $(0.630)^{* * *}$ | $(2.223)^{* * *}$ | $(1.085)^{* * *}$ |
|  |  |  |  |

The table presents results for the estimation of Specification using the Bureau of Economic Activity's data on American MNCs for years 1999-2011. The left hand side variable is the log of foreign sales of American MNCs in each 3-digit NAICS code (Column 1) or the ratio of foreign sales to exports (Columns 2 and 3 ). Column 3 exclude outlier observations in terms of the ratio. The variables in the right hand side include the unit shipping cost associated with the industry, a knowledge intensity measure and the exports of the US to the rest of the world in each 3-digits NAICS category (Column 1 only). All specifications include year fixed effects. Robust standard errors clustered at the year level are presented in parentheses.

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

## D Non-stop flights and the distance-knowledge trade-off

Another possible mechanisms through which MNCs can experience lower costs of transferring knowledge is through the existence of a non-stop flight in between the subsidiary and its global headquarters. The existence of non-stop flights would proxy for the ease of more frequent business trips between headquarters and subsidiaries, given the convenience of a direct flight. Business trips, by workers and managers, could facilitate the transmission of tacit knowledge. In this context, Giroud (2012) has shown that the existence of commercial air routes between subsidiaries and headquarters positively affects the profitability of the former.

To determine the existence of a non-stop air route between a headquarters and its subsidiary, I identify all the existing airports within a 100 Km radius using the geocoded latitude and longitude ${ }^{4}$ The data for airports (with their respective coordinates) and active air routes come from the OAG Flights $5^{5}$ 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.

Table D1 replicates the estimation in Table 3 but using the existence of a non-stop flight instead of the overlap in working hours.
[Table D1 about here.]
Across all different cuts of the sample, I am unable to find strong evidence that the existence of a non-stop flight significantly alters the distance-knowledge trade-off. These findings are insightful on their own. Another plausible explanation of the results by Keller and Yeaple (2013) is that the distance-knowledge trade-off responds to cost of transportation of managers or workers to the subsidiaries, as an "intermediate good". These travels would be easier through the existence of non-stop flights. Yet, the lack of results suggests that easing business travel cannot differentially explain the documented trade-off.

[^21]Table D1: Distance, knowledge intensity and non-stop flights

| Dependent Variable: Log Knowledge Intensity |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
| All |  |  |  |  |
| $\log (\mathrm{d})$ | Foreign | Horizontal | Vertical |  |
|  | -0.0002 | -0.0002 | -0.0014 | -0.0008 |
| $\log (\mathrm{~d})$ XdirectFlight | $(0.000)$ | $(0.001)$ | $(0.001)$ | $(0.003)$ |
|  | 0.0002 | 0.0009 | -0.0001 | -0.0006 |
| Non-stop Route | $(0.000)$ | $(0.001)$ | $(0.001)$ | $(0.002)$ |
|  | 0.0000 | -0.0062 | 0.0003 | 0.0093 |
| $\log (\mathrm{t})$ | $(0.002)$ | $(0.009)$ | $(0.008)$ | $(0.018)$ |
|  | -0.0961 | -0.0875 | -0.0803 | -0.1079 |
| Constant | $(0.008)^{* * *}$ | $(0.008)^{* * *}$ | $(0.012)^{* * *}$ | $(0.019)^{* * *}$ |
|  | 3.9315 | 3.9495 | 3.9714 | 3.9269 |
|  | $(0.017)^{* * *}$ | $(0.019)^{* * *}$ | $(0.025)^{* * *}$ | $(0.044)^{* * *}$ |
| N |  |  |  |  |
| Adj. R2 | 105183 | 40490 | 25265 | 1973 |
| $\varphi_{h}$ | 0.81 | 0.80 | 0.88 | 0.91 |

The table presents results for the estimation of Specification 1] using different cuts of the sample of MNCs' subsidiaries, including controls and interactions of distance between the subsidiary and its global headquarters (in logs) and the existence of a non-stop flight between them (using nearby airports). The left hand side variable is the knowledge intensity of the economy activity of the foreign subsidiary (in logs). The estimation controls for the unit shipping cost (in logs) of the good produced by the subsidiary. All specifications include MNC fixed effects. Robust standard errors clustered at the industry and MNC level are presented in parentheses.

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

## E Further exploring the distance-knowledge tradeoff after accounting for overlap in working hours

The results presented in Table 3 and Figure 7 use as a proxy for knowledge intensity the measures explained in Section 2 In this section I present further evidence that these results are robust to alternative specifications and not driven by outliers in the distribution of such measure. I also present evidence that these results are robust to using alternative measures of knowledge intensity and to a different definition of vertical subsidiaries.

## Using alternative specifications

The role of an additional hour of overlap between the subsidiary and its global headquarters might not explain changes in the estimation of the knowledgedistance trade-off in a linear way. In this section I reestimate the estimations depicted in Figure 7 with two main variations. First, I allow for an additional interaction between distance and working hours overlap in a quadratic form. The results are documented in Figure E1. According to this quadratic specification, the point estimates suggest that while the trade-off is weaker when accounting for a higher overlap in working hours, there is a decreasing marginal "effect" ${ }^{6}$ for every extra hour as compared to the linear specification presented in the main body of the paper, though the standard errors are much larger. However, all in all, the results are qualitatively and quantitatively similar to the linear case.
[Figure E1 about here.]
Second, I divide overlap in working hours in three categories: low ( 0 to 4 hours), medium ( 5 to 8 ) and high ( 9 or more). The reason to group them in 3 categories instead of more disaggregated categories is that by doing the latter I have several categories with very little observations (particularly for different cuts of the dataset) and the results are highly distorted by noise. The results for this exercise are depicted in Figure E2 In this figure, we can also see based on the points estimates that the trade-off estimator weakens for subsidiaries with "high" overlap in working hours with their headquarters, as opposed to those with "low" overlap in working hours, though the standard errors are also quite large. While the results for vertical foreign subsidiaries seem to be inconclusive, overall, the results are consistent with the linear specification.
[Figure E2 about here.]

## Using alternative knowledge intensity measures

I also test for whether the results are robust to alternative measures of knowledge intensity. First, Figure (E3) replicates the results but excluding industries that

[^22]are below the 5 th and above the 95 th percentile in terms of knowledge intensity, to make sure the results are not being driven by outliers. As the figure shows, the results are essentially identical.
[Figure E3 about here.]
Figure E4 replicates the results using R\&D intensity computed using the Compustat dataset and compiled by Keller and Yeaple (2013), and Figure E5 does so for R\&D intensity computed using the Orbis dataset and compiled by Nunn and Trefler (2008). Using both measures the resulting point estimates, while noisier (particularly for Orbis-based R\&D), seem to support the idea that the knowledge-distance trade-off weakens for subsidiaries in locations that have higher overlap in working hours with their headquarters.
[Figure E4 about here.]
[Figure E5 about here.]
I also reestimate the relationships based on other knowledge intensive measures, and find that the are qualitatively consistent to the main knowledge intensity measure used in the paper. Figure (E6) replicates the results using the share of non-production workers' wages in total payroll from the NBERCES Manufacturing Industry Database (Becker et al., 2013). Figure E7 replicates results using another knowledge-intensity measure I constructed following the same steps outlined in Section 2.2 .2 , this time using another input from O*NET: the extent to which workers in that industry "use and update relevant knowledge". Finally, Figure E8 does the same, but using a different input from O*NET: the intensity with which workers in that industry "make decisions and solve problems".
[Figure E6 about here.]
[Figure E7 about here.]
[Figure E8 about here.]
All in all, while results are often more noisy, I usually find that they are consistent with the main specification, in particular when it comes to foreign horizontal subsidiaries.

## Using a different definition for vertical subsidiaries

Given that in the main body of the paper the definition of a strictly vertical subsidiary is somewhat arbitrary, based on a $5 \%$ threshold, I present results on the distance-knowledge trade-off for vertical subsidiaries defined with a higher threshold. In particular, Figure E1replicates the results for foreign subsidiaries that represent $10 \%$ of the final output relative to industries that are upstream or downstream in which domestic subsidiaries of the same MNC are active
in, instead of $5 \%$, the threshold used in the main body of the paper. Note that, consistently with the main body of the paper, this sub-sample excludes any foreign subsidiaries that could also classify as horizontal. The results are consistent with those presented in Figure 7.
[Figure E9 about here.]

## Using different cuts based on average knowledge-intensity of MNC

It is interesting to further explore whether the results significantly vary based on the overall knowledge intensity of the MNC. Measuring sophistication of the MNC, however, is not trivial because (as explained in Section 2) often the SIC code of the headquarters is defined as a "holding company", which is not reflective of the main economic activity of the MNC. Thus, in order to deal with this, I compute the knowledge intensity of the most common economic activity based on all domestic subsidiaries of the MNC (in the case of several modes, I take the largest value). I then divide MNCs in terciles based on low, medium and high knowledge intensity.

Figures E10, E11 and E12 replicate the results of Figure 7 for the subsample of MNCs active in low, medium and high economic activities, respectively. While the results, overall, are consistent with the main ones presented in the main body of the paper, I find much more precise results for MNCs active in medium and high knowledge intensive activities, while for MNCs active in low knowledge intensive activities the results are much weaker.
[Figure E10 about here.]
[Figure E11 about here.]
[Figure E12 about here.]

Figure E1: Knowledge and distance trade-off, by working hours overlap (quadratic)









| - | 0 | - | 2 | - | 4 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| - | 6 | - | - | 8 | - | by the experience plus training measure), after controls specified in Specification 11, for different levels of overlap in working hours between the headquarters and its affiliate, allowing for a quadratic relationship. The left panel shows the value of the slope, adjusted for different levels of the working hours overlap; when the overlap is 0 the slope of the trade-off is highly negative, and as the overlap becomes larger, the trade-off weakens. The right panel plots the average knowledge-distance tradeoff based on different levels of working hours overlap between headquarters and foreign subsidiaries.

Figure E2: Knowledge and distance trade-off, by working hours overlap (nonlinear)


The figure presents the empirical fit for the relationship between $\log (d)$ and $\log (k)$ (the latter proxied by the experience plus training measure), after controls specified in Specification (1), for different levels of overlap in working hours between the headquarters and its affiliate categorized in three groups: low ( $0-4$ hours), medium ( 5 to 8 ) and high ( 9 or more). The left panel shows the value of the slope, adjusted for different levels of the wworking hours overlap; when the overlap is 0 the slope of the trade-off is highly negative, and as the overlap becomes larger, the trade-off weakens. The right panel plots the average knowledge-distance trade-off based on different levels of working hours overlap between headquarters and foreign subsidiaries.

Figure E3: Knowledge and distance trade-off, by working hours overlap (excl. outliers)









| - | 0 | - | 2 | - | 4 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| - | 6 | - | 8 | - |  |

The figure presents the empirical fit for the relationship between $\log (d)$ and $\log (k)$ (the latter proxied by the experience plus training measure, excluding outliers), after controls specified in Specification (1), for different levels of overlap in working hours between the headquarters and its affiliate. The left panel shows the value of the slope, adjusted for different levels of the working hours overlap; when the overlap is 0 the slope of the trade-offis highly negative, and as the overlap becomes larger, the trade-off weakens and even disappeats. The right panel plots the average knowledgedistance trade-off based on different levels of working hours overlap between headquarters and foreign subsidiaries; the trade-off is slightly positive when there are 10 hours or more of overlap between the two locations.

Figure E4: Knowledge and distance trade-off, by working hours overlap (R\&D)


The figure presents the empirical fit for the relationship between $\log (d)$ and $\log (k)$ (the latter proxied by R\&D intensity computed by Keller and Yeaple, 2013), after controls specified in Specification (1), for different levels of overlap in working hours between the headquarters and its affiliate. The left panel shows the value of the slope, adjusted for different levels of the working hours overlap; when the overlap is 0 the slope of the trade-off is highly negative, and as the overlap becomes larger, the trade-off weakens and even disappears. The right panel plots the average knowledge-distance tradeoff based on different levels of working hours ove 18 p between headquarters and foreign subsidiaries; the trade-off is slightly positive when there are 10 hours or more of overlap between the two locations.

Figure E5: Knowledge and distance trade-off, by working hours overlap (R\&D)


The figure presents the empirical fit for the relationship between $\log (d)$ and $\log (k)$ (the latter proxied by R\&D intensity computed by Nunn and Trefler, 2008), after controls specified in Specification (1), for different levels of overlap in working hours between the headquarters and its affiliate. The left panel shows the value of the slope, adjusted for different levels of the working hours overlap; when the overlap is 0 the slope of the trade-off is highly negative, and as the overlap becomes larger, the trade-off weakens and even disappears. The right panel plots the average knowledge-distance tradeoff based on different levels of working hours ove 19 p between headquarters and foreign subsidiaries; the trade-off is slightly positive when there are 10 hours or more of overlap between the two locations.

Figure E6: Knowledge and distance trade-off, by working hours overlap (nonproduction workers)









| - | 0 | - | 2 | - | 4 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| - | 6 | - | - | 8 | - |

The figure presents the empirical fit for the relationship between $\log (d)$ and $\log (k)$ (the latter proxied by share of non-production workers' wages in total payroll), after controls specified in Specification 11, for different levels of overlap in working hours between the headquarters and its affiliate. The left panel shows the value of the slope, adjusted for different levels of the working hours overlap; when the overlap is 0 the slope of the trade-of is highly negative, and as the overlap becomes larger, the trade-off weakens and even disappeary. The right panel plots the average knowledgedistance trade-off based on different levels of working hours overlap between headquarters and foreign subsidiaries; the trade-off is slightly positive when there are 10 hours or more of overlap between the two locations.

Figure E7: Knowledge and distance trade-off, by working hours overlap (knowhow)









| - | 0 | - | 2 | - | 4 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| - | 6 | - | - | 8 | - |

The figure presents the empirical fit for the relationship between $\log (d)$ and $\log (k)$ (the latter proxied by the extent to which workers in that industry use and update relevant knowledge), after controls specified in Specification 11, for different levels of overlap in working hours between the headquarters and its affiliate. The left panel shows the value of the slope, adjusted for different levels of the working hours overlap; when the overlap is 0 the slope of the trade-off is highly negative, and as the overlap becomes larger, the trade-off meakens and even disappears. The right panel plots the average knowledge-distance trade-off based on different levels of working hours overlap between headquarters and foreign subsidiaries; the trade-off is slightly positive when there are 10 hours or more of overlap between the two locations.

Figure E8: Knowledge and distance trade-off, by working hours overlap (problem solving)


The figure presents the empirical fit for the relationship between $\log (d)$ and $\log (k)$ (the latter proxied by the extent to which workers in that industry make decisions and solve problems), after controls specified in Specification 1], for different levels of overlap in working hours between the headquarters and its affiliate. The left panel shows the value of the slope, adjusted for different levels of the working hours overlap; when the overlap is 0 the slope of the trade-off is highly negative, and as the overlap becomes larger, the trade-off weakens and even disappears. The right panel plots the average knowledge-distance trade-off based on different levels of working hours overlap between headquarters and foreign subsidiaries; the trade-off is slightly positive when there are 10 hours or more of overlap between the two locations.

Figure E9: Knowledge and distance trade-off, by working hours overlap (Vertical, $10 \%$ threshold)


The figure presents the empirical fit for the relationship between $\log (d)$ and $\log (k)$ (the latter proxied by the experience plus training measure), after controls specified in Specification (1), for different levels of overlap in working hours between the headquarters and its affiliate. The estimation uses a sample of foreign vertical subsidiaries, defined as foreign subsidiaries that represent $10 \%$ of the final output relative to industries that are upstream or downstream in which domestic subsidiaries of the same MNC are active in (instead of $5 \%$, used in the main body of the paper). The left panel shows the value of the slope, adjusted for different levels of the working hours overlap; when the overlap is 0 the slope of the trade-off is highly negative, and as the overlap becomes larger, the tradeoff weakens and even disappears. The right panel plots the average knowledge-distance trade-off based on different levels of working hours overlap between headquarters and foreign subsidiaries; the trade-off is slightly positive when there are 10 hours or more of overlap between the two locations.

Figure E10: Knowledge and distance trade-off, by working hours overlap (low KI MNC)


The figure presents the empirical fit for the relationship between $\log (d)$ and $\log (k)$ (the latter proxied by the experience plus training measure), after controls specified in Specification (1), for different levels of overlap in working hours between the headquarters and its affiliate. The estimation uses only MNCs for which most of its economic activities at home are in the bottom tercile in terms of knowledge intensity. The left panel shows thg yalue of the slope, adjusted for different levels of the working hours overlap; when the overlap is the slope of the trade-off is highly negative, and as the overlap becomes larger, the trade-off weakens and even disappears. The right panel plots the average knowledge-distance trade-off based on different levels of working hours overlap between headquarters and foreign subsidiaries; the trade-off is slightly positive when there are 10 hours or more of overlap between the two locations.

Figure E11: Knowledge and distance trade-off, by working hours overlap (medium KI MNC)









| - | 0 | - | 2 | - | 4 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| - | 6 | - | - | 8 | - |

The figure presents the empirical fit for the relationship between $\log (d)$ and $\log (k)$ (the latter proxied by the experience plus training measure), after controls specified in Specification (1), for different levels of overlap in working hours between the headquarters and its affiliate. The estimation uses only MNCs for which most of its economic activities at home are in the middle tercile in terms of knowledge intensity. The left panel shows the yalue of the slope, adjusted for different levels of
 as the overlap becomes larger, the trade-off weakens and even disappears. The right panel plots the average knowledge-distance trade-off based on different levels of working hours overlap between headquarters and foreign subsidiaries; the trade-off is slightly positive when there are 10 hours or more of overlap between the two locations.

Figure E12: Knowledge and distance trade-off, by working hours overlap (high KI MNC)








$\left.\begin{array}{|llllll|}\hline- & 0 & - & 2 & - & 4 \\ - & 6 & - & - & 8 & -\end{array}\right) 10+4$

The figure presents the empirical fit for the relationship between $\log (d)$ and $\log (k)$ (the latter proxied by the experience plus training measure), after controls specified in Specification (1), for different levels of overlap in working hours between the headquarters and its affiliate. The estimation uses only MNCs for which most of its economic activities at home are in the top tercile in terms of knowledge intensity. The left panel shows the yalue of the slope, adjusted for different levels of
 as the overlap becomes larger, the trade-off weakens and even disappears. The right panel plots the average knowledge-distance trade-off based on different levels of working hours overlap between headquarters and foreign subsidiaries; the trade-off is slightly positive when there are 10 hours or more of overlap between the two locations.

## F Robustness tests on the regression discontinuity design

Figure F1 presents evidence that the choice of a different threshold to define vertical subsidiaries does not change the fact that there is no discontinuity in knowledge intensity for this type of subsidiaries. In particular, Figure F1 is a graphical representation of the regression discontinuity design in Specification (2) for foreign subsidiaries that represent $10 \%$ of the final output relative to industries that are upstream or downstream in which domestic subsidiaries of the same MNC are active in (instead of $5 \%$, the threshold used in the main body of the paper). Note that, consistently with the main body of the paper, this subsample excludes any foreign subsidiaries that could also classify as horizontal. Similarly as in Figure 9, the results show that there is no clear discontinuity.

## [Figure F1 about here.]

The main results of the paper defines based the number and sizes of bins based on the "mimicking variance evenly-spaced method" (see Calonico et al. 2014, 2017), which often results in slightly different number of bins in either side of the discontinuity. However, the graphical representation of the discontinuity (for foreign horizontal subsidiaries) is robust to the choosing bins through other methods, including manually. Figure F2 presents RD plots, analogous to Figure 10. where the number of bins are chosen manually and such number is the same on both sides of the discontinuity. The figure uses 50,100 and 150 bins on either side. The results are robust to this modification.
[Figure F2 about here.]
Table F1 repeats the estimation of Specification (2), using a uniform weighting scheme which provides equal weight to all observations within the selected bandwidth. The estimation uses the subsample of foreign horizontal subsidiaries. Results are robust to those presented in Table 4 in the main body of the paper.
[Table F1 about here.]
Table F2 presents results using a more conservative set of fixed effects: one for every MNC and time zone line simultaneously $\left(\varphi_{h} \times \tau_{t z}\right)$. This would be in fact comparing two horizontal foreign subsidiaries that belong to the same MNC and that are at either side of the same time zone line. The point estimates are positive, thus reassuring, albeit smaller in magnitude and without statistical significance. However, it is important to realize that this specification has very little variation to exploit in the sample: a quarter of the combined MNC and time zone lines pairs have 4 observations or less, half of them have 15 observations or less, and 75 percent of them have 46 observations or less. In an effort to exploit more variation, the estimations uses larger bandwidths (in each column), though even with relatively large bandwidths, there is simply not enough variation within each one of these cells contributing to the lack of precision of the estimate.
[Table F2 about here.]

Figure F1: RD plots (Vertical, 10\% threshold)


The figure presents a graphical representation of the regression discontinuity design in Specification (2), for foreign vertical subsidiaries that represent $10 \%$ of the final output relative to industries upstream or downstream in which domestic affiliates are active in. The left panel presents a linear estimation while the right panel plots a quadratic one. Bin sizes are defined on the "mimicking variance evenly-spaced" method (denoted by emsv in the rdplot package).

Figure F2: RD plots (Horizontal), choosing symmetric number of bins

## 50 Bins



100 Bins



## 150 Bins




The figure presents a graphical representation of the regression discontinuity design in Specification (22, for foreign horizontal subsidiaries. The graphs on the left are based on a linear estimation and the graphs on the right are based on a quadratic estimation. The number of bins are chosen manually (50, 100 and 150) and are symmetric on both sides of the discontinuity.

Table F1: RD estimation, uniform weighting scheme

| Dependent Variable: Foreign subsidiary's knowledge intensity (log) |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
|  | 150 | 250 | 350 | Optimal |
| closerTZ | 0.0084 | 0.0060 | 0.0075 | 0.0066 |
|  | $(0.003)^{* *}$ | $(0.003)^{*}$ | $(0.002)^{* * *}$ | $(0.002)^{* * *}$ |
| distTZzero100s | -0.0034 | -0.0028 | -0.0015 | -0.0018 |
|  | $(0.001)^{* * *}$ | $(0.001)^{* * *}$ | $(0.000)^{* * *}$ | $(0.000)^{* * *}$ |
| closerTZ $\times$ distTZzero100s | -0.0019 | 0.0024 | -0.0004 | 0.0005 |
|  | $(0.003)$ | $(0.002)$ | $(0.001)$ | $(0.001)$ |
| Constant | 4.1045 | 4.1020 | 4.1047 | 4.1055 |
|  | $(0.001)^{* * *}$ | $(0.001)^{* * *}$ | $(0.001)^{* * *}$ | $(0.001)^{* * *}$ |
| N |  |  |  |  |
| Adj. R2 | 5232 | 9702 | 14179 | 12367 |

The table presents results for the estimation of Specification 2 using a sample of foreign horizontal subsidiaries of MNCs estimated using several bandwidths for the running variable specified in each column. The last column uses the optimal bandwidth computed using the methodology described in Cattaneo et al. (2018) who build on the work by Imbens and Kalyanaraman (2012). The estimation uses a uniform weight scheme, giving same weight to all observations within the bandwidth. All specifications include MNC fixed effects and time zone line fixed effects. Robust standard errors clustered at the MNC and time zone line level are presented in parentheses. ${ }^{*} p<0.10,{ }^{* *} p<0.05,{ }^{* * *} p<0.01$

Table F2: RD estimation, $\varphi_{h} \times \tau_{t z}$ fixed effects

| Dependent Variable: Foreign subsidiary's knowledge intensity (log) |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
|  | 250 | 350 | 450 | 550 |
| closerTZ | 0.0011 | 0.0032 | 0.0026 | 0.0027 |
|  | $(0.002)$ | $(0.003)$ | $(0.002)$ | $(0.002)$ |
| distTZzero100s | -0.0013 | -0.0006 | -0.0003 | -0.0004 |
|  | $(0.000)^{* * *}$ | $(0.000)^{* *}$ | $(0.000)$ | $(0.000)^{*}$ |
| closerTZ $\times$ distTZzero100s | 0.0027 | 0.0000 | 0.0002 | 0.0005 |
|  | $(0.002)$ | $(0.001)$ | $(0.001)$ | $(0.001)$ |
| Constant | 4.1008 | 4.1030 | 4.1018 | 4.1015 |
|  | $(0.001)^{* * *}$ | $(0.001)^{* * *}$ | $(0.001)^{* * *}$ | $(0.001)^{* * *}$ |
| N |  |  |  |  |
| Adj. R2 | 8861 | 12924 | 16091 | 18985 |

The table presents results for the estimation of Specification 2, using a sample of foreign horizontal subsidiaries of MNCs estimated using several bandwidths for the running variable specified in each column. The last column uses the optimal bandwidth computed using the methodology described in Cattaneo et al. (2018) who build on the work by Imbens and Kalyanaraman (2012). The estimation uses a triangular weight scheme, giving higher weight to observations closer to the cutoff point. All specifications include MNC fixed effects and time zone line fixed effects. Robust standard errors clustered at the MNC and time zone line level are presented in parentheses.
${ }^{*} p<0.10,{ }^{* *} p<0.05,{ }^{* * *} p<0.01$


[^0]:    *This paper benefited from helpful comments at various stages from Martin Abel, Laura Alfaro, Pol Antras, Sam Asher, Sebastian Bustos, Marcus Casey, Matias Cattaneo, Michele Coscia, Ricardo Hausmann, Elhanan Helpman, Bill Kerr, Marc Melitz, Paul Novosad, Nathan Nunn, Lant Pritchett, Krista Rasmussen, Brina Seidel, Ran Shorrer, Ernesto Stein, Stephen Yeaple, Muhammed Yildirim, Richard Zeckhauser, Oren Ziv and participants in different seminars at Harvard University, the NBER Productivity Lunch, Indiana University, Vanderbilt University, Georgetown University, the World Bank, and others. Jean-Paul Leidenz Font and Samuel Rosenow provided excellent research assistance. All errors are my own.
    $\dagger$ Author's email and website: db21@post.harvard.edu, http://www.danybahar.com

[^1]:    ${ }^{1}$ 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.
    ${ }^{2}$ 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.

[^2]:    ${ }^{3}$ See Irarrazabal et al. (2013) for a similar setting which does not include the knowledge dimension.
    ${ }^{4}$ Note that this also applies to human capital as an intermediate good (in the form of managers visiting foreign subsidiaries, for example): the cost of transportation -for goods or for people-should not differ that much with distance whether it is within or across time zones.

[^3]:    ${ }^{5}$ See Keller $\sqrt{2004}$ for a review of this literature.
    ${ }^{6}$ 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. Bahar and Rapoport (2018), consistent with the idea that the transmission of tacit knowledge requires human interaction, show that migrants are an important driver of knowledge diffusion across nations.
    ${ }^{7}$ The dataset was privately acquired from $\mathrm{D} \& \mathrm{~B}$ and is not publicly accessible. It has been previously used in the literature by Lipsey (1978), and more recently by Black and Strahan (2002); Harrison et al. (2004); Acemoglu et al. (2009); Alfaro and Charlton (2009); Alfaro and Chen (2012) and Alfaro et al. (2015).
    ${ }^{\circ}$ Often, throughout the paper, 1 also use domestic subsidiaries.
    ${ }^{9}$ Previous studies have also used the O*NET database to construct industry level measures. For example, Oldenski (2012 measures the importance of communication with the headquarters and importance of the communication with the customer, for each industry. Costinot et al. (2011) uses O*NET to create an industry level measure of task routine-ness for 77 sectors. Keller and Yeaple (2013) also present results making use of knowledge intensity variables constructed with O*NET in their Appendix. Autor et al. (2003) use O*NET predecessor, DOT, to construct measures for routine and non-routine tasks.

[^4]:    ${ }^{10}$ The paper is accompanied by an Online Appendix which is refer to throughout the text.
    ${ }^{11}$ 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.
    ${ }^{12}$ I performed on the dataset an algorithm that would group multinationals not only based on their assigned number, but also on their names when the differences are small (e.g., Sony Corporation vs. Sony Corp).

[^5]:    ${ }^{13}$ 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.
    ${ }^{14}$ The use of $\$ 0.05$ in the main body of the paper follows the precedent set by Alfaro and Charlton, 2009 but its choice is not determinant for the results of the paper. In fact, I often present robustness tests using a threshold of $\$ 0.1$ and the results are unchanged.

[^6]:    ${ }^{15}$ I acknowledge the limitation of defining vertical subsidiaries using input-output tables instead of using intra-firm trade which I lack (e.g., Atalay et al. 2014 Ramondo et al. 2015). Yet, it is important to note that by "favoring" the definition of horizontal over vertical for those foreign subsidiaries that classify as both, my results are less likely to be biased due to oversampling vertical subsidiaries. This is particularly so because the main the results -as I discuss below- are driven by foreign horizontal subsidiaries, for which the classification does not rely on input-output data.
    ${ }^{16}$ Data from 2011, downloadable from ftp://ftp.bls.gov/pub/special.requests/oes/oesm11in4.zip
    ${ }^{17} \mathrm{O}^{*}$ NET is the successor of the US Department of Labor's Dictionary of Occupational Titles (DOT). I use the $\mathrm{O}^{*}$ 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 $\mathrm{O}^{*} \mathrm{NET}$ in the Appendix.
    ${ }^{18} \mathrm{I}$ used Pierce and 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.
    ${ }^{19}$ Since this measures is based on US data only, I will assume the US ranking in the knowledge intensity of industries proxies that of the rest of the world.

[^7]:    ${ }^{20}$ See Bahar (2018) and Online Appendix Section B for more details.
    ${ }^{21}$ It also positively correlates with other less popular measures that could proxy for knowledge intensity. 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 . These correlations were computed using SIC 4 digits codes.

[^8]:    ${ }^{22}$ According to the Skewness/Kurtosis test for normality, we cannot reject the hypothesis that these measures are normally distributed.
    ${ }^{23}$ Downloadable from http://faculty.som.yale.edu/peterschott/files/research/ data/xm_sic87_72_105_20120424.zip

[^9]:    ${ }^{24}$ In this context, Giroud (2012) has shown that the existence of commercial air routes between subsidiaries and headquarters positively affects the profitability of the former. 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.

[^10]:    ${ }^{25}$ One of the possible reasons that the results for foreign vertical affiliates are not very precise, is the relatively small number of affiliates in the sample (less than 2000), leaving for very little variation left after controlling for firm fixed effects. However, there could be as well other explanations, which are discussed below.
    ${ }^{26}$ Allowing for a more flexible fit, such as a quadratic one, suggests that the estimated relationship does not seem to be monotonically decreasing for the lower values of knowledge intensity (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.

[^11]:    ${ }^{27}$ Given the low estimated magnitude for $\beta_{d}$, since the within-firm variance for knowledge intensity (in logs) is much lower -in fact it corresponds to $1 / 350$ - than the within-firm variance of distance (in logs), it is easier for interpretation purposes to rely on the point estimate of $\beta_{k}$ in the 'mirror' specification:

[^12]:    The point estimate of $\beta_{k}$ corresponds to -0.49 , with the exact same statistical significance as the estimated $\beta_{d}$. Using this estimation one can compute: $\Delta \log \left(d_{s}\right)=-0.49 \times \Delta \log \left(k_{s}\right) \approx$ $-0.49 \times 0.77 \approx-0.38$, where 0.77 is the difference (in logs) of the knowledge intensity values for a semiconductor plant and a meat packaging plant. Thus, according to this estimation, a semiconductor plant would be in a location that is $40 \%$ closer to the headquarters than a plant in the meat packaging business. In fact, this relative difference would roughly correspond to locations such as Dublin and Kabul, the former location being about $40 \%$ closer to a headquarters in Houston than the latter.
    ${ }^{28}$ I also explore how is the knowledge-distance trade-off affected by having a non-stop flight in between the headquarters and the subsidiary. I find no evidence that the trade-off is reduced, as it is the case with time zone differences. This is despite the fact that Giroud (2012) has shown that the existence of commercial air routes between subsidiaries and headquarters positively affects the profitability of the former. For more on this, please see Online Appendix Section (D.

[^13]:    ${ }^{29}$ All specifications control for trade unit costs, and the estimator for $\beta_{t}$ is also negative, implying that the partial correlation between knowledge intensity and shipping costs is negative.

[^14]:    ${ }^{30}$ I refrain from presenting results using a fuzzy regression discontinuity design using the existence of a non-stop flight based on the work by Campante and Yanagizawa-Drott 2016), who show that the likelihood of having a non-stop flight experiences a jump at the 6000 miles (or 9600 kilometers) of distance. The reason not to present those results are twofold. First, the estimations shown in Online Appendix Section D finds no results when including a nonstop flight as an ease of communication variable. Second, while the reduced form (using the 6000 mile threshold) shows results consistent with what is expected (e.g., observations just below the threshold are more knowledge intensive than those just above it), the results are not robust when using the non-stop flight in a fuzzy regression discontinuity setting. This is probably because the sample does not include all the existing flights, but rather only those flying to and from airports near a headquarters or subsidiaries. Results of this, however, are available upon request.

[^15]:    ${ }^{31}$ Unfortunately the dataset does not have enough variation for an estimation that adds a joint fixed effect for every MNC and time-zone line. For more on this please refer to Online Appendix Section F

[^16]:    ${ }^{32}$ This method may result in a slightly different number of bins on either side of the discontinuity. In Online Appendix Section $F$ I show that choosing the number of bins manually and forcing the same number on both sides of the discontinuity does not alter the results.
    ${ }^{33}$ Online Appendix Section F includes a plot for foreign vertical subsidiaries using a higher input-output threshold, for which presumably real-time communication is more critical. Using this sub-sample I am unable to find evidence of discontinuity, either.
    ${ }^{34}$ In the regression analysis I present results using many different bandwidth sizes, finding consistent results.

[^17]:    ${ }^{35}$ Online Appendix Section F replicates the estimation using a uniform weighting scheme. The results are robust to this alternative weighting scheme.
    ${ }^{36}$ Given that a $1 \%$ increase in distance -roughly 36.5 Km . based on the sample averageis associated with a value in the knowledge intensity scale that is lower by $0.00137 \%$, then, based on the RD estimates that are 5 to 6 time larger in magnitude, one more hour of overlap is equivalent, on average, to being closer in distance by about 182.5 to 219 Km .
    ${ }^{37}$ One aspect that could raise concerns from the results is that the coefficient for the running variable (distTZzero100s) is negative. However, it is important to consider that this is not

[^18]:    the distance to the headquarters, as estimated in Figure 5 and Table 3 but rather the running variable, which measures distance to the nearest time zone line. Yet, the estimator suggests that the further away the foreign subsidiary is from the time zone line in the direction opposite of the headquarters (e.g., in a time zone that reduces the working hours overlap), the knowledge intensity of the foreign subsidiary tends to be larger. This, at first, is inconsistent with the overall results. This is, however, the result of a local linear regression, using a limited sample. The point of this exercise is rather to focus on the discontinuity, where distTZzero100s $=0$. Therefore, this highly local result -as seen before- does not represent the pattern in the overall data, which is explored in the previous section. Note, too, that when it comes to the relationship between distance and knowledge intensity on the right margin of the discontinuity, the standard errors of the estimation are large enough so that we cannot reject a flat slope for the estimator, or even a slightly positive one (as can also be evidenced in Figure 10 .
    ${ }^{38}$ The sources for these country-level variables are World Bank's World Development Indicators (i.e., income per capita), as well as from Shirotori et al. 2010) (i.e., measures of factor endowments such as stock of physical capital, human capital and land, all in per worker form).

[^19]:    ${ }^{1}$ http://dnb.com.au/Credit_Reporting/The_quality _of_DandBs_data/index.aspx
    ${ }^{2}$ Including their domestic plants for both.

[^20]:    ${ }^{3}$ When using a more conservative definition for vertical subsidiaries, such as those products that represent 10 cents of a dollar of input/output, the results are unchanged.

[^21]:    ${ }^{4}$ I condition the airport to be in the same country as the headquarters or the subsidiary.
    ${ }^{5}$ The dataset was privately acquired in January 2015.

[^22]:    ${ }^{6}$ I use the word "effect" for lack of a better term, baring the understanding that this is not a causal estimation, but rather a partial correlation.

