Multinational Production and Intra-firm Trade *

Vanessa Alviarez †
University of Michigan

Ayhab Saad ‡
University of Michigan

April 4, 2014

Abstract

A salient empirical regularity of multinational production (MP) is that foreign affiliate sales are decreasing in trade costs. As a response, intra-firm trade, from parents to foreign affiliates, has been combined with standard models of *horizontal* MP to generate complementarities between trade and multinational activity that deliver gravity-style predictions for foreign affiliates' sales. However, intra-firm trade is not common across foreign affiliates but rather concentrated among a small set of large multinational firms (Ramondo et al., 2014). In addition, we document that not only firms in the upper-tail of the firm's size distribution, which are more likely to conduct intra-firm transactions, are subject to gravity forces; but also sales of relatively small foreign affiliates are significantly affected by geographical barriers. Two puzzles emerge: (i) why intra-firm trade is concentrated among the largest multinational firms? and (ii) what are the mechanisms that drive affiliates' sales in the lower tail of the distribution to obey gravity forces, even in the absence of intra-firm trade? In this paper we deliver a framework to explain theses two phenomenas. An affiliate's marginal cost is affected by the parent decision regarding the method of knowledge transfer. Exporting intermediate inputs embodying knowledge to an affiliate is subject to the standard iceberg-type trade costs and fixed costs of intra-firm trade. The costs of direct knowledge transfer are also increasing with geographical barriers but rises less than the costs of exporting intermediate inputs. Because of the fixed costs of intra-firm trade, only the most productive multinational firms choose to export to its affiliates. Moreover, the share of imported intermediate inputs to the affiliate's total costs is increasing with firm's productivity. We show that, in equilibrium and taking into account both the intensive and the extensive margins, foreign affiliates' sales for both the affiliates who import and do not import from parents suffer from gravity forces.

JEL-Classification: F12, F23
Keywords: Intra-firm trade, Multinational Production

*The authors thank Andrei Levchenko, Alan Deardorff, Robert Franzese, Linda Tesar, and Kyle Handley for valuable discussions and helpful suggestions.

†University of Michigan. Economics Department. Email: alviarez@umich.edu
‡University of Michigan. Economics Department. Email: ayfsaad@umich.edu

1 Introduction

The proximity-concentration tradeoff constitutes the basis of one of the most important theories of multinational production.[1] In a stark contrast with the prediction of the workhorse monopolistic competition model of trade and foreign direct investment, several new empirical papers document that total foreign affiliates' sales are subject to gravity-style forces akin to those observed for aggregate exports (Yeaple, 2009; Keller and Yeaple, 2013; Irarrazabal et al., 2013). That is, rather than overcome the transportation costs associated to exports, multinational sales also decrease with remoteness and other geographical variables.

A natural explanation for the observed patterns of bilateral foreign affiliate sales is the existence of trade in intermediate inputs across countries within the boundaries of the firmintra-firm trade. The usage of intermediate inputs produced by the parent introduces a source of complementarities between trade and multinational production, given that in order to produce overseas, foreign affiliates have to import intermediate inputs from their home market.

Intra-firm trade is an important component of U.S. international trade. In particular, exports of manufacturing goods from U.S. parents to their cross border network of affiliates account for 20 percent of U.S. exports; and intra-firm imports by foreign controlled U.S. affiliates from their foreign parent groups have generally accounted for 20-25 percent of total U.S. imports. Thus, to bridge the theory of horizontal multinational activity with the new empirical facts, new models have incorporated intra-firm transactions in the workhorse framework of trade and foreign direct investment (FDI) (Helpman et al., 2004. HMY, henceforth).

A striking feature of intra-firm trade is its pronounced heterogeneity across firms; not only at the aggregate level, but also at the sector-destination country level. In particular, using a detailed data from the Bureau of Economics Analysis, Ramondo et al. (2014) have documented that intra-firm trade is concentrated among a small number of large affiliates and it only represents a very small fraction of their input and their total sales. For example, in 2004, the median manufacturing affiliate received none of its inputs from their parent firm,[2] and sold 91 percent of its production

[1] There are two main competing international trade theories that aim to explain the behavior of multinational corporations. The first of these theories is based on the so call *horizontal* multinational production, under which the primary motive for expanding operations overseas is to satisfy foreign final consumers. Under this theoretical approach the multinational firm faces a concentration-proximity tradeoff. On the one hand, firms that choose to produce at home market and sell internationally through exports, while taking advantage of the economies of scale of concentrating operations in one location face higher marginal cost of selling to foreign markets because of the associated transportation cost. On the other hand, the firm could take advantage of the proximity to foreign markets when it sets operations abroad and saves the transportation cost associated to exports (Brainard, 1997; Markusen, 2004). The second theory is based on the *vertical* multinational activity, under which the primary motive of setting operations abroad is to take advantage of price differences of factors of production across countries, to produce intermediate inputs at a lower cost. These inputs could be sell back to the parent company or to a third affiliate in another country to advance further stages of the production process (Helpman, 1984).

[2] Of course, this does not rule out the possibility that an affiliate is importing intermediate inputs from another affiliate who is part of the international production chain. Unfortunately, such flows are not recorded in any of the available datasets. Nevertheless, the fact that the vast majority of affiliates sell their output to unrelated parties, alleviate part of these concerns. We discuss these issues in more details in a later section.

to unrelated parties, mostly in the host country.[3] The observed selection to engage in intra-firm trade, the skewness of intra-firm flows towards large corporations and the bias of multinationals sales to local final consumers are robust to the country of destination or the industry of operations.

These findings pose new challenges to existing FDI-intrafirm models. Firstly, existent models take it for granted that all affiliates import from their parents and therefore are silent about the selection and the skewness observed in intra-firm flows. Secondly, if importing from parents is what derives the affiliate's sales to be declining in trade frictions, then FDI-gravity shall be present only for those firms in the upper tail of the firm's size distribution but not for the relatively small size firms.[4,5] Using Orbis dataset, we present evidence showing that the gravity of foreign sales could not solely be explained by intra-firm trade. In fact, we show that even firms that likely do not engage in intra-firm trade exhibit a significant resistance to geographical barriers. This is a finding that contrasts with the predictions of new FDI-intrafrirm models, where the only source of gravity for MP sales is the complementarity between trade and MP imposed by intra-firm flows. It has been well established in the literature that multinational sales, on average, follow the law of gravity that has very well characterized bilateral trade in the geographical space. Alternatively, we divide the sample of firms by its likelihood of engaging in intra-firm transaction. This is, we show that the standard gravity variables (i.e., distance, common border, common language, and regional trade agreements) play a significant role in diminishing the observed foreign affiliates sales, for the lower and the upper tails of the firm's size distribution. As expected, gravity forces diminish the sales of the very large affiliateswhich previews evidence found often involve in intra-firm trade. Surprisingly, the gravity frictions also negatively and significantly affect the sales of the relatively small foreign affiliates, which often do not trade with their parents and sell the vast majority of their output to unrelated parties in the host market. This results are very similar when the sample of firms is divided below and above of the fifty percentile of the firm's size distribution and when only the firms below the twenty five percentile to represent small firms while those firms above the seventy five percentile to represent larger affiliates. The results of the impact of gravity on multinational

[3]Notice that the fraction of foreign affiliates that receive or not intermediate inputs from their parents, and sell the vast majority of their output to unrelated parties (in the local market or in a third market), are the observations that lie in this category of horizontal MP with vertical linkages. This observation could be potentially important, because if there is circular intra-firm trade, the parent could provide intermediate inputs to an affiliate that is part of an international production chain, not to satisfy final consumers but to produce intermediate inputs to continue the production process in other countries.

[4]Of course, the fact that intra-firm trade is concentrated among the largest multinational corporationsgranularity of intra-firm tradecould be enough to generate FDI-gravity on the entire sample of multinational firms.

[5]The stylized facts about intra-firm trade and the gravity of multinational activity also impose important challenges to the theory of multinational production based on vertical integration. In contrast with horizontal intrafirm-FDI- models, where firms could engage or not in intra-firm flows from parents to affiliates, in models of vertical multinational production, intra-firm transactions are a necessary condition to the existence of foreign affiliates, whose main role is to provide cheaper intermediate inputs to their parents and to other affiliates within the corporation. Therefore, conditional on being a multinational, the vertical integration theory of multinational activity could not rationalize the observed absence of intra-firm flows among firms within a corporation that instead sell the majority of their output to unrelated parties in the host market.

activity are robust to different econometric specifications.

To capture the former stylized fact and to account for the extensive margin of intra-firm trade flows, this paper develops a novel multi-country model of heterogeneous firms, in which parent firms decide whether or not to supply foreign affiliates with intermediate inputsinstead of let them produce those in the host marketand if so, optimally decide the fraction of intermediate inputs that should be imported from the parent company. The proposed theoretical framework matches the distribution of multinational sales as well as the intra-firm trade patterns observed in the data: the less productive firms do not import at all from their parent, whereas the most productive ones engage in intra-firm trade. In the model, the selection is explained by the irreversible investment in which a multinational corporation has to incur in order to establish an adequate platform to carry cross borders transactions within the boundaries of the firm on regular bases. The high cost associated to these important coordination efforts is a fact well explored in the international management literature (Seuring and Goldbach, 2002).

This paper contributes to previous efforts to rationalize intra-firm trade patterns. Irarrazabal et al. (2013) propose a HMY model of horizontal multinational production with intra-firm trade from parents to affiliates. Their model assumes that the final good produced by affiliates is assembled in a Cobb Douglas fashion using local labor and intermediate inputs produced and shipped by the parent. As a consequence, all firms engage in intra-firm trade and the share of imported intermediate inputs from the parent in total cost is the same for all firms regardless of their productivity level. Similarly, Ramondo and Rodriguez-Clare (2013) develop a general equilibrium model of trade and multinational production under perfect competition in which foreign affiliates use an international input bundle in production, where a fraction is obtained in the local market and the rest comes from the parent firm. This paper differs from these approaches in that it endogenizes the existence as well as the degree in which intra-firm trade occurs.

Following Keller and Yeaple (2013), we assume that when a firm produces overseas, either requires to establish communication with the headquarters to receive the instructions to produce the intermediates inputs—direct knowledge transfer or alternatively, the multinational can transfer knowledge across borders by exporting intermediate inputs embodying the technology—indirect knowledge transfer. When the firm produces its intermediate inputs, it faces the costs of transfer knowledge across countries; but if instead, the firm buy the intermediates from its parent, it faces the associated transportation costs. Therefore, under this framework, multinational sales of more knowledge intensive firms will suffer more strongly from gravity, precisely because these companies face relatively high costs of direct knowledge transfer, reducing the elasticity of intra-firm trade to changes in transportation cost. Notice that in Keller and Yeaple (2013) model, although affiliates differ in their share of imported intermediate inputs from the parent, it assumes that all foreign affiliates buy at least some inputs from their headquarters. Nonetheless, in the data, only a small fraction of firms, often relatively large, buy from their parent while a vast majority of them report zero intra-firm flows.[6]

Our paper improves the previous theoretical framework in several dimensions in oder to reproduce several of the recent uncover stylized facts of intra-firm trade. First, knowledge intensity is more heterogeneous across firms within an industry than it is across industries. Based on this fact, we choose to model knowledge intensity as firm-specific rather than sector-specific. A more knowledge-intensive affiliate is so because it requires relatively more knowledge-intensive intermediate inputs than a less knowledge-intensive firm. Second, the knowledge intensity of production affects not only the composition or degree of in-house production versus imported intermediate inputs, but also the existence of intra-firm trade itself. That is, in the current model some firms could find it optimal not to engage in intra-firm trade because developing a distribution channel entails to incur in a fixed cost.[7] Third, the share of imported intermediate inputs increases with firm's productivity. Fourth, communication costs are higher, the higher the knowledge intensity of the intermediate inputs and the larger the distance between the parent and the affiliate firm.[8]

Notice that in this framework the usage of intermediate inputs has a meaning that differs from most trade models. In Keller and Yeaple (2013) as well as in the model developed in this paper, intermediate inputs are a channel of technology transfer between parent and affiliates to achieve the same productivity level. These intermediate inputs are firm-specific and we assume that they should be produced inside the boundaries of the firm. In this environment, the affiliate faces only two options: either it produces the intermediate inputs by itself, in which case it requires indirect transfer of technology in the form of communication to receive the necessary instruction from its parent firm; or it can import these intermediates from headquarters, saving the communication cost but facing the transportation cost of exporting.[9]

From an empirical perspective, it has been a challenge for the literature of multinational production and intra-firm trade to unveil its determinants. One of the reasons is the limited information

[6]In the appendix of the paper, Keller and Yeaple sketch an extension of the original model in which firms have the option of paying a fixed cost of investing in information and communication technology in order to lower the efficiency cost of knowledge transfers by reducing the efficiency loss of remote production. An implication of this extension is that only the most productive firms produce a larger fraction of their intermediates in the host market, reducing their reliance in imported intermediate inputs, given that, only those affiliates are able to incur in the fixed cost. However, this prediction is contrary to the patterns observed in the data in which only the most productive firms engage in intra firm trade. Instead, our model proposes a very different type of fixed cost that is able to deliver the observed selection. In our model to engage in intra-firm trade, the corporation has to build a complex distribution network that allows frequent trade between related affiliates operating in different countries.

[7]It is possible to think that the need of the affiliate of receiving instructions from the parent firm decreases with time, as the affiliate moves along in the learning curve. Nonetheless the multinational firms develop innovation at a high rate and constantly those new techniques will be passed through the affiliates in the form of direct technology transfer.

[8]For example, communications tend to be more cumbersome when firms are located in different time zones, in countries with different languages, among others.

[9]In most models of trade and multinational production, subsidiaries adopt the same technology level as their foreign market by construction; instead in this model, the firm buys or produces intermediate inputs that allow it to reproduce the activity of the parent firm. Some of these models incorporate a measure of productivity losses, to capture difference in productivity between the headquarters and foreign affiliates and they often use "standard" inputs—not firm-specific—which can be outsourced from any market.

in the available datasets in order to distinguish vertical and horizontal FDI. Given that there is no direct measures of the underlying motives of MP, researches have used intra-firm transactions to obtain some information about the nature of multinational firms. Very often the existence of intra-firm trade has been interpreted as evidence for vertical MP and against horizontal MP.[10] This is because under vertical MP multinational firms decide to produce cheaper intermediate inputs abroad within the boundaries of the firm in order to internalize any product contractibility issue and potential spillovers of proprietary knowledge that could emerge from outsourcing. Under vertical MP, foreign production will always result in intra-firm flows because affiliates are meant to produce firm-specific intermediate inputs, which will be exported to the parent firmor to another affiliate-but not to satisfy the final foreign consumption. In fact, horizontal MP can rationalize intra-firm transactions from the parent to the affiliate when the last one imports knowledge embedded in the form of intermediate inputs. Nonetheless, the workhorse model of horizontal MP will not require intra-firm trade in order to absorb the productivity of the parent, but rather it assumes that the affiliate replicates the same organizational structure of the parent.

Therefore, the magnitude and direction of intra-firm flows could shed light to understand the relative importance of these two alternative theories, given that horizontal MP will be compatible with intra-firm transactions from the parent to the affiliate, while vertical MP will be compatible with transactions in both direction; although more strongly with trade of intermediates from the affiliates to the parent firms or to other affiliates within the corporation.[11] Even though we recognize the richness of intra-firm transactions, which includes sales from affiliates to parents as well as from affiliates to affiliates, the model developed in this paper adopts a horizontal perspective of multinational activity. Intra-firm trade from the parent to their network of foreign affiliates is quan-

[10] For instance, Keane and Feinberg (2006) find that the reason of the dramatic growing in intra-firm trade flows among U.S. and Canada over the 1984-1995 period were due to the intensive, rather than the extensive margin. They attribute the low contribution of the extensive margin to the fact that the modest tariff reductions were not sufficient to justify fixed costs of overhauling international supply chains. This is, the extensive margin of intra-firm trade is often associated with the existence of MP. This is a very narrow interpretation given that in fact the majority of foreign affiliates establish operations abroad without establishing any kind of trade with the parent or any other affiliate.

[11] For the U.S. there are two main sources of information. One of them is the U.S. Census Bureau who administrates the information contained in the custom declaration that includes information on ownership ties between the foreign and domestic parties involved in any transactions. The other main source is U.S. Commerce Department's Bureau of Economics Analysis, which conducts extensive surveys to multinational firms that includes questions on the value of specific trade flows between foreign and domestic units of the firm. The Census data is comprehensive in the detail of the traded goods between U.S. parents and affiliates overseas and also between foreign parents and their affiliates operating in the U.S. It allows to identify not only the type of relationship between the parts involved in the international transaction, but also the nationality of the parts. Unfortunately, the Census data does not collect further information of the activity of the foreign part of the transaction, which is relevant to characterize the behavior of multinational corporations. On the other hand, the BEA data offers in-depth information about affiliates operations, including total assets, sales, net income, employment, $R\&D$. It also has information about the international transaction between and the affiliate and related and unrelated parties, with the parent company, with the host market, or with third markets. Unfortunately, affiliates do not report the related parties sales to third market disaggregated by countries. For this reason, even when it is possible to track the intra-firm transaction between parents and affiliates, it is not the case for the trade among foreign affiliate that are part of the same company.

titatively important and it is also consistent with a model of horizontal multinational production and gravity on foreign affiliate sales.[12]

The reminder of the paper is organized as follows. Section 2 discusses the main source of data in our analysis and it also describes the main characteristics of multinational sales at the firm level. Section 3 presents the three main stylized fact that support our contribution to the literature, both in term of our model and the evidence presented. Section 4 lays out the theoretical framework and derives the analytical implications for intra-firm flows and multinational sales. Section 5 discuses the parametrization, the functional forms and the estimation strategy of some of the key model's parameters. Section 6 presents the general equilibrium and the gravity equation of affiliate sales for firms that do engage in intra-firm trade and those that do not. Section 7 concludes.

2 Data

In this section, we explain in detail the source and characteristics of the dataset used in the analysis. The primary source of information is Orbis, which gathers firm-level information across a wide range of countries. In particular, it contains relevant information about the ownership structure of the firm, with a detailed list of direct and indirect subsidiaries and stockholders, the company's degree of independence, its ultimate owner and other companies in the same family.[13] Orbis does not have information about the transaction between parents and affiliates firms, as the Census Bureau data does, instead it offers more information about the foreign affiliates' operations, including financial statements as well as a comprehensive set of indicators of economic activity.[14]

For the purpose of this analysis we have constructed two samples. The first sample is comprised of U.S. affiliate firms operating outside the United States, but whose ultimate owners—or parents—are located in the United States. The second sample contains information of foreign affiliates firms operating in the U.S., but whose parents are located outside the U.S. Notice that in both samples we only account for those affiliates that have a global ultimate owner (GUO), this is a company who exercises the greater degree of control over the affiliate and that owns at least 50 percent of the shares; thus, a firm is considered foreign owned if it is majorityor wholly owned by a foreign multinational firm.[15] Regarding the sample, it is important to mention that only those firms for

[12] Zeile (2003) uses detailed data from the Census Bureau and find U.S. intra-firm exports mainly consist of shipments from U.S. parent companies to their foreign affiliates, and U.S. intra-firm imports mainly consist of shipments from foreign parent groups to U.S. affiliates.

[13] Alfaro and Chen (2012) have assess the extent and coverage of this data set using more aggregated information for alternative sources. Because we concentrate here on affiliates owned by U.S. parent firms, as well as U.S. affiliates owned by foreign parents, we have used the aggregate values in the BEA data to evaluate the accuracy of the information provided by Orbis.

[14] The best characterization of the intra-firm trade can be obtained from the Census Bureau, but it lacks of information about the activity of the affiliates, including the type and destination of its exports, which is contained in the BEA data.

[15] We also consider a company to be an ultimate owner (UO) if it has no identified shareholders or if its shareholder's percentages are not know. It is worthwhile to mention that we consider only Global, rather than Domestic ultimate

which operating revenue is known for at least one of the years in the sample period (2004-2013), are considered in the analysis.

The analysis focuses on manufacturing industries. It covers more than 9 thousand U.S. own affiliates operating in 35 developed countries, and it also covers the foreign own affiliates operating in the U.S. from these 35 countries. Four categories of information are used for each firm: (a) industry information including the 4-digit NACE code of the primary industry in which the establishments operate, (b) location information; (c) non-consolidated financial information including operating revenue, employment, assets, investment, wages, material cost, among others; and (d) degree of ownership and detailed information about the global ultimate owner, including a compressive set of financial information of the parents firms which includes the industry of operation, revenue, employment, assets, research and development expenditures, and number of patents among others.

In order to construct a useful sample, the data was subjected to an extensive cleaning up process in which we eliminate firms whose operating revenue is below one million dollar and with less than 15 employees. Furthermore, to alleviate the problem from potential outliers, we eliminate firms below the 0.1th percentile and above the 99.9th percentile in the distribution of sales. The final sample comprises 8,572 foreign affiliates and 2,210 parents, covering 261 manufacturing industries for the period 2004-2013.

3 Stylized Facts

In this section we introduce some key regularities about the foreign sales and the location patterns of U.S. multinational firms. First, we show that the knowledge intensity of the U.S. parent firms is very heterogeneous across firms, even within very narrow defined sectors. Second, we show some evidence of the granularity of multinational activity from the parent as well as from the affiliate perspective. Multinationals are a relatively rare type of firms. Despite of the disproportionately contribution of U.S. multinationals to total output and trade, they represent less than 1 percent of all U.S. companies. Moreover, the vast majority of U.S. parents only operate in one foreign market regardless of the manufactured industry; and also for any given market-sector pair the market share of U.S. foreign production is concentrated in a very small set of affiliates. Third, we present some initial empirical evidence that intra-firm trade alone is not enough to reconcile the underlying incentives in *horizontal* FDI models and the observed strong dampening effect of distance on MP. Overall, this section gives the grounds of our motivation and provides support for the building blocks of the model proposed in section 4.

owners. The Domestic UO is the highest company in the path between a foreign affiliate and its Global UO but that is located in the same country as the affiliate firm. Thus, an affiliate will be considered domestic, rather than foreign, when the GUO and the DUO are both in the same country. The definition of Global Ultimate Owner, with a minimum of fifty percent ownership adopted in this paper, is also the one followed by international agencies and by the U.S. Bureau of Economic Analysis.

Fact 1: *Research and Development intensity is highly heterogeneous across multinational firms within a narrow defined industry.* The average research intensity varies significantly across parent firms, even in considerable narrow defined industries. Borga and Zeile (2004) find that foreign manufacturing affiliates have a greater propensity to source intermediate goods from their U.S. parents is increasing in their parent R&D and capital intensity. This suggest that the propensity of affiliates to source intermediate inputs from their parents is related to the level of intangible assets embodied in the inputs traded within the firm.

Figure 4 shows the density of the parent's share of R&D expenditure for the pool of U.S. parents in the sample regardless of the industry classification. As can be observed the expenditure in research and development is remarkably higher among the most productive U.S. parent firms. In fact, more than eighty percent of the R&D expenditures in a given industry is in hands of few but very large firms. Figure 5 shows the density of parent's R&D expenditure share for four selected three-digit level NACE sector classification: (1) manufacturing of parts and accessories for motor vehicles—NACE 293 (top-left panel), (2) manufacture of other special-purpose machinery—NACE 289 (top-right panel), (3) manufacture of basic pharmaceutical products—NACE 211 (bottom-left panel), and (4) manufacture of air and spacecraft and related machinery—NACE 303 (bottom-right panel). The share of R&D is calculated as the fraction of the research and development expenditures of the firm relative to the total R&D expenditures of all U.S. parents firms operating in the same three digit sectoral classification. It is clear, that the concentration of R&D expenditures in few large parents is not being driven by sector- specific characteristics. The results are qualitatively similar even when considering only those the firms belonging to a given sector.

However, only one third of the U.S. ultimate owner with at least one foreign affiliate in the sample (U.S. parent firms) have information about R&D expenditures. In order to address how previous results can be affected for the lack of more complete information on R&D expenditures, Figure 6 shows the density of the productivity for two groups of parent firms: those for which Orbis data contains information regarding the expenditures in research and development activities; and those parent firms that contain missing values for the R&D variable. The productivity density is shown for both groups in the same industries used in the analysis above. Firm's productivity is measured by the output per worker of the U.S. parent. Figure 6 highlights that those firms for which Orbis does not record information about R&D expenditures are on average less productive than firms for which it does. Therefore, we conclude from this evidence that even when multinationals are responsible for the majority of the private R&D activities, the largest share of the R&D expenditures in any given industry is being mainly carried on by few but very productive U.S. parent firms.

Fact 2: *The distribution of foreign affiliate sales is fat tailed, for each country and sector pair.*

A well documented fact is that firm sales follow a Zipf Law distribution (Gabaix, 2009 and di Giovanni and Levchenko, 2012). In addition, Ramondo et al. (2014) show that intra-firm trade is concentrated among a small number of large affiliates. In particular, firms below the mean of the size distribution do not trade with their parent firms at all. In this section, we show that the distribution of sales of U.S. foreign affiliates—as well as the sales of foreign affiliates in U.S.— is very fat tailed. Not just at overall, within an industry, or within a country; but also for a given country-sector pair.

Figure 2 shows the distribution of the market share of each affiliate in each country-sector pair. As can be observe, most of firms represent a very small share in each market, and only a small fraction firms have remarkable large market share. Figure 3 evaluates the participation of U.S. parents on foreign markets. Each parent produces on average in two foreign economies, but fifty percent of the parents only produce in one market besides United States. Strikingly, the mean coincide with the number of markets penetrated by a firm in the 75 percentile of the distribution. Ten percent of the parent firms produce in more than four markets and only five percent of all firms set operations in seven or more foreign countries.

Fact 3: *Intra-firm trade alone cannot explain the observed gravity of multinational sales*

Intra-firm trade from the parent to the cross border network of foreign affiliates has been the approach used in the literature to rationalize the gravity of multinational production; meaning that aggregate foreign affiliate sales fall with geographical barriers. Nevertheless, only the most productive foreign affiliates buy intermediate inputs from their parent firms in the U.S.; a fact that is robust across countries and also across industries (Ramondo et al., 2014). From the perspective of the existing models, this implies that only sales of foreign affiliates located at the upper-tail of the size distribution should suffer from gravity. Conversely, in this subsection we present some evidence showing that the gravity of foreign sales could not solely be explained by intra-firm trade. In fact, we show that even firms that likely do not engage in intra-firm trade exhibit a significant resistance to geographical barriers. Because most models of horizontal multinational production that feature intra-firm trade fail to account for the observed selection of intra-firm trade—assuming that all firms will require some fraction of the intermediate inputs from the parents— in section 4 we propose a model to account for the intensive and the extensive margin of intra-firm trade.

Ideally, we would like to test the existence of gravity for two groups of firms: those that participate in intra-firm trade transactions and those that do not. Unfortunately, for this paper we do not have access to intra-firm trade data at the firm level.[16] Instead, we proceed to divide the U.S. affiliate firms by their size in two groups for any given host country-sector pair. First, we split the whole sample of firms in two subsamples by the 50th percentile of the affiliates' size distribution. This criteria is based on Ramondo et al. (2014) that found that none of the affiliates

[16]In oder to overcome this limitation in the near future, we will merge Orbis firm level data with the Census Bureau data, to get a perspective of the transaction between U.S. multinationals and their foreign affiliates as well as of the economic activity of these affiliates overseas.

below the median import intermediate inputs from their parent firms.[17,18] Second, we divide the sample of firms by those that belong to the lower-tail (below 25th percentile) and the upper-tail (above 75th percentile) of the firm size distribution, in each country-sector pair. Taking only the extremes of the firm size distribution reduces the likelihood that the so called small firms could engage in intra-firm trade, and increases the likelihood that the very large firms do.

Below we present the results of the gravity equation that comes from different specifications and samples. Table 1 presents the results of the regression for the intensive and extensive margin of multinational activity (column 1 and 2) as well as for the extensive margin only (column 3 and 4). It includes all U.S. multinational firms in our sample, and the data has been aggregated up to the country-sector level. As a proxy of geographical barriers we have included the log of physical distance ($ln\left(dist_{i,ua}\right)$), a set of dummy variables indicating whether countries share a common border ($Border_{i,us}$), common language ($Language_{i,us}$), belong to a given regional trade agreement ($RTA_{i,us}$) and whether they had a colony relationship ($Colony_{i,us}$). In column 1 and 3 we also control for some key characteristics of the host country that could determine the scale of foreign operations and so directly affect the volume of local sales as well as the intra-firm trade. These controls include the capital endowment relative to the U.S., a measure of the size of the market (GDP per capita) as well as a proxy of institutional quality measured by the Rule of Law variable from the Worldwide Governance Indicators database of the World Bank. In order to account for other country characteristics that are potential determinants of FDI and that are not included in our regression, such as relative technology differences and skill endowments among others, in column 2 and 3, we include instead country fixed effects to control for any country specific characteristic that could affect the gravity of multinational production. Notice that in both specifications sector fixed effects are included to control for the great observed heterogeneity of MP at the sectoral level that can affect the impact of gravity variables on MP sales as well as on the number of firms that produce overseas.

Consistent with previous studies, both the total affiliates sales and the number of U.S. parents are decreasing with trade barriers, and in particular, they are declining in distance from the United Sates in both specifications. Having a common language positively affect both margins as well; nevertheless, it loses statistically significance once we control for country fixed effect. The existence of a trade agreement between the United States and the host economy significantly increases the affiliates' sales but negatively affects the number of firms that engage in foreign production. A potential explanation is that trade agreements increase the sales of the U.S. firms in the foreign

[17]Our criteria differs from Ramondo et al. (2014) in that they show that their finding is established for the median firm in a given industry and for the median firm in a given region. Instead, we split the sample based on the median firm in each country-sector pair.

[18]As discussed in the introduction of this paper, the fact that firms are not receiving intermediate inputs from the parent firm, does not mean they are not engage in intra-firm trade with other affiliates within the same corporation. Ramondo et al. (2014) find that regardless of its size, the majority of firms sell their output to unaffiliated parties in the host country, which although indirectly, could partially alleviate concerns about trade among foreign affiliates. Unfortunately, intra-firm imports from other than the U.S. parents is not captured by the available data sources.

market by facilitating intra-firm transactions with the parent firm, but reduce the number of firms that find it profitable to engage in MP given that exporting gets more attractive. Of the hots-country specific variables, the size of the host country market (GDP) and the level of capital were significant and of the expected sign. On the other hand, foreign affiliate sales fall in host country institutional quality.[19]

To further explore whether the negative effect of geographical barriers on MP for the whole sample are not only driven by those firms who engage in intra-firm trade, Table 2 and Table 3 present the results of gravity on number of firms and MP sales, respectively, but this time dividing the sample on firms below and above the median of the firm size distribution in each country-sector pair. Column 1 and 2 show that, for both groups, the number of firms and the MP sales decrease with distance, showing a negative and statistically significant coefficient for firms below and above the median. Notice that the coefficient of the distance variable in Table 3 are very similar when country fixed effects are included. But, in order to evaluate the aggregate effect of the variables associated with gravity, we compute the bilateral MP costs based on the estimated coefficients to evaluate the distribution of MP cost for both groups.[20]. Figure 7 shows the density of MP costs for both groups. As can be observed the mean of both groups is similar but the variance of MP cost is considerable higher for smaller firms, showing that some country-sector pairs of this group are strongly affected by MP costs.

Given that we are relying on the size of the firm as a proxy of its participation in intra-firm trade, we reproduce the above exercise but this time we only consider firms under the 25th percentile, which most likely do not to engage in intra-firm trade, as well as firms above the 75th percentile of the distribution, which most likely conduct intra-firm transactions.[21] Table 2 reproduces the gravity regression for the group of firms in the tails of the firms size distribution. Consistent with previous results for both groups of firms, foreign sales and number of firms are significantly lower in countries far from the U.S.

So far, the evidence shows that regardless of their size—and therefore on whether or not firms do intra-firm trade—multinational sales are significant affected by gravity forces, either measured

[19] At a first glance, this seems a very surprisingly result but it is possible that it is driven by the fact that U.S. has less room to exploit its institutional comparative advantage in countries with high law enforcement level. In the light of the theory of the boundaries of the firm, this finding could go with both of the leading theories in this vein: transaction costs theory and property rights theory. According to transaction costs theory, better institutional setting reduces the need for vertical integration-reducing the number of majority own affiliates. Incorporating this finding with the property rights theory is more subtle: if contractibility is a more of issue on the investments carried by the headquarter, then the result is consistent with the property rights theory; as institutions advance, the need to provide more incentives to headquarter declines, leading to less vertical integration.

[20] We calculate the following equation based on the gravity estimated coefficients for each country-sector pair:

$$\widehat{\tau}^{mp}_{i,us} = \widehat{\beta}_d \times lndist_{i,us} + \widehat{\beta}_b \times border_{i,us} + \widehat{\beta}_{language} \times lan_{i,us} + \widehat{\beta}_{RTA} \times RTA_{i,us} + \widehat{\beta}_c \times colony_{i,us}$$

[21] The fact that the firm size follows a Zipf law distribution, and also that the distribution of foreign sales is very fat tailed, could induce that some firms above the median are not large enough to trade within the firm.

by distance only or distance plus other gravity variables. Therefore the collected evidence shows that the data rejects a model in which the only source of MP gravity comes from intra-firm trade. The model presented in the next section attempts to address two important aspects of the data. Fist, the observed selection in intra-firm trade: only very few and large firms conduct intra-firm transaction across border within the firm. Second, it proposes another source of gravity to capture the fact that also multinational firms that do not trade with the affiliates are significantly affected by the gravity forces.

4 The Model

Our model is based on Helpman et al. (2004). Firms are heterogeneous in term of their productivities. Goods are horizontally differentiated with each variety produced by a firm that acts as a monopolist. A firm can enter the foreign market via exporting or by opening a foreign affiliate in the destination market (FDI). As is well known, in choosing between either mode of entry, a firm faces a proximity-concentration trade off: establishing a foreign affiliate is associated with lower variable trade costs but higher fixed cost of conducting multinational production. The model predicts a definitive firms' hierarchy: least productive firms do not produce, low productive firms only sell to the domestic market, medium productive firms export, and most productive firm turn into multinational corporations. Furthermore, as in Irarrazabal et al. (2013) and Keller and Yeaple (2013), we introduce parent-to-affiliate intra-firm trade to generate FDI-gravity akin to the standard trade-gravity.

The model contributes to the literature in many ways. First, in order to be consistent with the stylized facts (i.e., intra-firm trade is concentrated among the very most productive multinational corporations, with the majority of FDI firms report zero intra-firm), we introduce fixed cost of intra-firm trade. Second, in contrast to Irarrazabal et al. (2013), the share of imported intermediate inputs to the total intermediate inputs costs is not constant and varies with firm's size. Unlike Keller and Yeaple (2013) and consistent with the empirical fact that the share of intermediate inputs to total input costs is also increasing with firm size, we tie firm's productivity to firm's knowledge intensity (R&D) to associate intra-firm with firm's size. Finally, we show that FDI-gravity style forces are presented in the model even for the foreign affiliates that do not import from their parents.

4.1 Consumer Demand

The world economy consists of N countries (indexed by i, n). Each country is populated by L_n utility-maximizer consumers, with each consumer inelastically supplying one unit of labor (the only factor of production). A representative consumer in country n derives her utility from the consumption of a homogenous good Q_0 and a continuum of differentiated goods that belong to the differentiated sector Q_n. Consumer's preferences between the homogenous good and the differentiated goods sector are represented by the Cobb-Douglas utility function with an income fraction μ spent on the differentiated goods

$$U_n = Q_0^{1-\mu} Q_n^{\mu}, \quad \mu \in (0,1) \tag{1}$$

Preferences over the differentiated goods are CES with elasticity of substitution $\sigma > 1$. The consumption of each variety ω in the set of all available varieties in country n, Ω_n (endogenously determined); $q^d(\omega)$, enters the CES aggregation symmetrically

$$Q_n = \left[\int_{\omega \in \Omega_n} q^d(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}}. \tag{2}$$

As is well know, the demand for each variety in country n is given by: $q^d(\omega) = A_n p_n(\omega)^{-\sigma}$. Here, $p_n(\omega)$ denotes the price of variety ω in country n. The index of market size in country n, A_n is exogenous from the point of view of consumers and individual producers.[22]

4.2 Production and Market Structure

The market for the homogenous good is perfectly competitive, and the production technology of the homogenous product is linear in labor: w_n units of labor are required to produce one unit of the homogenous good in country n. The homogenous good is freely traded in the world economy. So long as $1-\mu$, L_n, and the variable trade costs are large enough, the production of the homogenous good Q_0 in country $n \in \{1, 2, \ldots N\}$ is strictly positive. The price of the homogenous good is normalized to one; in effect, the wage in country n is pinned down by the numeraire and is equal to w_n.[23]

Each country n is endowed with exogenously given potential number of firms (producers) J_n. Each firm produces a unique variety using a variety-specific composite intermediate input. Productivity $\varphi \in R_{++}$ is a firm-specific that is drawn from a known cumulative distribution $G(\varphi)$ with probability density distribution $g(\varphi)$. Since φ is firm-specific, and each firm produces a unique variety we index goods with φ instead of ω. A firm with productivity draw φ requires $\frac{1}{\varphi}$ units of the firm-specific composite intermediate input M_φ to produce one unit of variety $\omega(\varphi)$. The composite intermediate input is produced under a CES aggregation of a continuum of intermediate

[22] $A_n \equiv \mu \frac{X_n}{P_n^{1-\sigma}}$. The aggregate price level of the differentiated goods sector in country n is denoted by P_n, and X_n represents the total expenditures in country n.

[23] The incomplete specialization assumption has been used by many researches for tractability and simplification purposes (for example, see Chaney, 2008). Proceeding without the outside sector will not alter the results presented in the paper, however.

inputs with elasticity of substitution $\eta \geq 1$:[24]

$$M_\varphi = \left(\int_0^\infty \beta(z|\varphi)^{\frac{1}{\eta}} m(z)^{\frac{\eta-1}{\eta}} dz \right)^{\frac{\eta}{\eta-1}}. \tag{3}$$

A couple of notes warrant attention here: (i) $m(z)$ is the quantity of an intermediate input of knowledge intensity z, with higher z indicating higher knowledge intensity, (ii) $\beta(z|\varphi)$ is the cost share of intermediate input z to the total cost of intermediate input bundle specific to φ-firm, and $\int_0^\infty \beta(z|\varphi)dz = 1$ for any φ, (iii) $\beta(z|\varphi)$ is log-supermodular in z and φ. That is, while all firms employ the same CES aggregation and use the same continuum of intermediate inputs, the share of each intermediate input z to the total cost of intermediate composite is firm specific. To be precise, $\beta(z|\varphi)$ is log-supermodular in z and φ if for $z^{'} > z^{''}$ and $\varphi^1 > \varphi^2$, $\beta(z^{'}|\varphi^1)\beta(z^{''}|\varphi^2) > \beta(z^{'}|\varphi^2)\beta(z^{''}|\varphi^1)$. In words, firm φ^1 is more knowledge-intensive because it requires relatively more knowledge-intensive intermediate inputs relative the low productivity firm φ^2,[25], and (iv) production technology of producing intermediate inputs is common across all firms: one unit of labor is needed to produce one unit of z.

4.3 Mode of Entry

A domestic firm gains access to the domestic market in country n after incurring a fixed cost of production f_{nn} units of labor. Country $i \neq n$ exporters to country n are subject to both fixed export cost f_{ni}[26] units of country i labor, and iceberg-type variable trade costs, $\tau_{ni} - 1 > 0$. Country i firms can also serve country n via FDI: pay a fixed cost of FDI, f_{ni}^{fdi} units of country i labor, and start serving n via its affiliates there. In so doing, a firm avoids the transportation costs associated with shipping the final good, but conveys an additional fixed cost of opening an affiliate in country n. Conditional on establishing a foreign affiliate in country n, a parent firm in country i has an option to let its affiliate to produce all intermediate inputs composite M (standard HMY setting), or chooses to ship intermediate inputs to its affiliate (intra-firm trade) where the fraction of inputs offshored and the volume of the intrafirm trade are endogenous. If a parent in country i decides to engage in zero intra-firm trade with its affiliate (i.e., let the affiliate produce all the intermediate inputs and the final good), and since M_φ is firm-specific, the affiliate is not as efficient as its parent; consequently, an affiliate needs $t_{ni}(z) > 1$ units of labor to produce one unit of intermediate input z.[27] In the case of intra-firm trade, a parent firm incurs a fixed cost

[24] It can be shown that the limit of the CES aggregation as η approaches one is a Cobb-Douglas.

[25] The intermediate composite aggregation and the notion of log-supermodularity were outsourced from Keller and Yeaple (2013). In contrast to Keller and Yeaple (2013), knowledge-intensity is defined on the firm level, not the industry level; a propriety that enables us to generate firm-level prediction regarding intra-firm trade. For a formal treatment of the log-supermodular assumption and its usage in the international trade context, see Costinot (2009).

[26] First subscript refers to the destination market and the second one to the origin country.

[27] A very crucial assumption to our model is that $t_{ni}(\varphi)$ is a function of trade frictions τ_{ni}. Nevertheless, the impact of the distance between the two countries, language, time zone and borders on trade frictions τ_{ni} is stronger

of initiating an intrafrim trade, f_{ni}^{int} units of country i labor; the shipped intermediate inputs are subject to the standard iceberg-type trade costs τ_{ni}, while the intermediate inputs produced by an affiliate are subject to productivity loss that is intermediate input specific $t_{ni}(z) \geq 1$ (we call it $t(z)$ for notational simplicity).[28]

4.3.1 Intra-firm Trade and Knowledge Transfer

The production of intermediate input with knowledge intensity z is firm-knowledge-specific. Moving knowledge over geographic space is costly. Transferring the knowledge required to produce intermediate input z to an affiliate entails, for example, communication cost, mis-implementation and mis-interpretation. Differently put, knowledge is not perfectly codified and therefore any knowledge transfer between a parent and its affiliate is subject to errors. Intuitively, the higher the knowledge intensity of the intermediate input z, the higher are the cost of transferring knowledge from a parent to the affiliate. Knowledge transfer takes two forms (i) *disembodied knowledge transfer*: parent firms directly transfer the necessary knowledge of producing input z to their affiliates who use the transmitted knowledge to produce that particular intermediate input. If this is the case, as mentioned above, the knowledge transfer costs are denoted by $t(z)$. To capture the idea that the cost of moving knowledge over space is increasing with knowledge intensity z, we assume that $t(0) = 0$, $\lim_{z\to\infty} t(z) > \tau_{ni}$ and $t'(z) > 0$,[29], and (ii) *embodied knowledge transfer*: simply, a parent produces intermediate input z and ships it to the affiliate in country n.

Finally, the production technology of the final good is invariant to the location of the producer (parent vs affiliate): regardless who produces the final good (parent or affiliate), $\frac{1}{\varphi}$ units of M_φ are needed to produce one unit of the final good. The decisions whether to export, to open an affiliate, and to outsource intermediate inputs production impact the production of final good only through its impact on the production of the composite of intermediate input M_φ.

4.4 Partial Equilibrium

First, we characterize the geography of input sourcing. The decision whether to outsource the production of intermediate input z is simply pinned down by comparing the cost of embodied knowledge transfer $w_i\tau_{ni}$ and disembodied knowledge transfer $w_n t(z)$. The cost of obtaining input z of a foreign affiliate is $c(z) = \min\{w_n t(z), w_i\tau_{ni}\}$. Given our assumption on the function $t(z)$, there exists an intermediate input with knowledge intensity $\tilde{z}$ such that : for any $z < \tilde{z}$, $t(z) < \varpi\tau_{ni}$, and for $z > \tilde{z}$, $t(z) > \varpi\tau_{ni}$. Then, we define $\tilde{z}(\tau_{ni}\varpi) = t^{-1}(\tau_{ni}, \varpi)$, where $\varpi \equiv \frac{w_i}{w_n}$. Conditional

in the case of physical shipping compared with its effect on the cost of knowledge transfer. Formally, $0 < \frac{\partial t_{ni}}{\partial \tau} < 1$. We elaborate more in this point and the former one below.

[28] Again, trade costs rise faster with distance and other trade frictions than does $t_{ni}(z)$.

[29] Notice that the cost of knowledge transfer is not firm-specific; however, the aggregate cost of disembodied knowledge transfer for a given fraction of the intermediate inputs varies across firms because of the log-supermodulity assumption.

on serving market n by FDI, we characterize the cost of the composite intermediate input to an affiliate with productivity draw φ,[30]

$$C_{ni}^{M}(\tau_{ni}, \mathcal{I}(\varphi), \varphi, \varpi) = \begin{cases} w_n \bar{t} & \text{if} \quad \mathcal{I}(\varphi) = 0, \\ \left(\int_0^{\tilde{z}(\tau_{ni},\varpi)} \beta(z|\varphi)(t(z)w_n)^{1-\eta} dz + (\tau_{ni} w_i)^{1-\eta} \int_{\tilde{z}(\tau_{ni},\varpi)}^{\infty} \beta(z|\varphi) dz \right)^{\frac{1}{1-\eta}} & \text{if} \quad \mathcal{I}(\varphi) = 1. \end{cases} \tag{4}$$

The indicator function $\mathcal{I}(\varphi)$ equals one if an affiliate outsources some of the intermediate inputs from its parent and zero otherwise. As we show below, the indicator function depends on firm's productivity draw φ. Indeed, $C_{ni}^{M}(\tau_{ni}, \mathcal{I}(\varphi) = 1, \varphi) < C_{ni}^{M}(\tau_{ni}, \mathcal{I}(\varphi) = 0, \varphi)$.[31] The elasticity of $C_{ni}^{M}(\tau_{ni}, \mathcal{I}(\varphi) = 1, \varphi, \varpi)$ with respect to trade costs τ_{ni}, $\varepsilon^{MC}(\tau_{ni}, \varphi, \varpi)$ is given by

$$\varepsilon^{MC}(\tau_{ni}, \varphi, \varpi) = \frac{(w_i \tau_{ni})^{1-\eta} \int_{\tilde{z}(\tau_{ni},\varpi)}^{\infty} \beta(z|\varphi) dz}{\int_0^{\tilde{z}(\tau_{ni},\varpi)} \beta(z|\varphi)(t(z)w_n)^{1-\eta} dz + (\tau_{ni} w_i)^{1-\eta} \int_{\tilde{z}(\tau_{ni},\varpi)}^{\infty} \beta(z|\varphi) dz}. \tag{5}$$

In order to show that within all firms that decide to enter country n by establishing a foreign affiliate, only a subset of those firms (the most productive) choose to ship intermediate inputs to its affiliates, we introduce the following lemmas,

Lemma 1 *The elasticity of marginal cost of composite intermediate input with respect to trade costs τ_{ni} is increasing in firm's productivity φ. For $\varphi^1 > \varphi^2$, $\varepsilon^{MC}(\tau_{ni}, \varpi, \varphi^1) > \varepsilon^{MC}(\tau_{ni}, \varpi, \varphi^2) > 0$.*

Lemma 2 *let $\theta(\tau_{ni}, \varphi, \varpi)$ be the share of imported inputs $M(\tau_{ni}, \varphi, \varpi)$ in total composite intermediate input costs $TC(\tau_{ni}, \varphi, \varpi)$. Then, $\theta(\tau_{ni}, \varphi, \varpi) = \frac{M(\tau_{ni},\varphi,\varpi)}{T(\tau_{ni},\varphi,\varpi)} = \varepsilon^{MC}(\tau_{ni}, \varphi, \varpi)$ is (i) increasing in φ, (ii) the import cost share is declining in trade costs for all firms, and (iii) the rate of decline in the import cost share is slower in the more knowledge intensive firms.*

Despite the fact that all firms choose to import the same range of intermediate inputs (notice that $t(z)$ and τ_{ni} are not firm-specific), the share of the imported intermediate inputs to the total composite intermediate input costs varies across firms in a way consistent with the log-supermodularity assumption. Accordingly, all the variations in the share of the imported intermediate inputs to the total costs are on the intensive margin not the extensive margin.[32] Lemma 1 is of great importance in the current setting: more knowledge intensive firms are more vulnerable

[30] $\bar{t} \equiv \int_0^{\infty} \beta(z|\varphi) t(z)^{1-\eta} dz$.

[31] The results emanates from firm's optimization and the definition of $\tilde{z}(\tau, \varpi)$. In fact, on might even argue that $\bar{t} \leq \left(\int_0^{\infty} \beta(z|\varphi)(w_n t(z))^{1-\eta} dz \right)^{\frac{1}{1-\eta}}$, since an affiliate that was assigned to produce all the intermediate inputs will be on average more efficient in producing any intermediate input z compared to an affiliate that produces a fraction of the intermediate inputs. This could be as a result of external return to scale or knowledge spillover and learning by doing hypothesis. Regardless, $\bar{t}$ is strictly higher than $C_{ni}^{M}(\tau, \varphi, \mathcal{I})$. Moreover, an important assumption has to be made here:t $w_n \bar{t} < w_i \tau_{ni}$. Otherwise no firm chooses FDI without intrafirm over exporting.

[32] If we let $t(z)$ be dependent on firm's productivity, both the extensive and the intensive margin of imported inputs will vary across firms. All the results presented in the paper will be reinforced.

to trade costs because they are more dependent on the imported intermediate inputs from their parents (Lemma 2). Reframing, the firm-level gains from trade liberalization (savings in marginal cost) are positively related to firm's knowledge intensity, i.e., productivity. The second part of Lemma 2 is trivial and intuitive. The third part of the same Lemma spawns from Lemma 1.

To sum up, the two lemmas above highlight the role of firm's knowledge intensity (productivity), trade impediments and the interaction between the two in shaping intrafirm trade on the firm level. More knowledge-intensive firms are so because they require more knowledge-intensive intermediate inputs. A more knowledge-intensive affiliate imports higher share of its intermediate inputs from its parent, and consequently an increase in trade costs raises the marginal cost of composite intermediate input of more knowledge-intensive affiliate proportionally more than less knowledge-intensive firms. Changes in trade costs impact firms' decision regarding embodied and disembodied knowledge transfer; yet the degree of substitution between them is significantly less for more knowledge intensive firms. An increase in trade costs, for example, leads to less decrease in the share of imported inputs to aggregate composite intermediate input costs for high knowledge-intensive affiliate since the more knowledge-intensive affiliate's ability to substitute embodied with disembodied knowledge transfer is constrained by the large demand for the highly knowledge-intensive inputs.

Embodied vs Disembodied Knowledge Transfer: Given the isoelastic demand facing each working firm in country n, profits for an affiliate in country n and a parent in country i can be written as,[33]

$$\pi_{ni}^{aff} = \varphi^{\sigma-1} B_n C_{ni}^M(\tau_{ni}, \mathcal{I}(\varphi), \varphi, \varpi)^{1-\sigma} - w_i(f_{ni}^{fdi} + \mathcal{I}(\varphi) f_{ni}^{int}), \tag{6}$$

An affiliate chooses to outsource intermediate inputs from parent if and only if the increase in its profits due to the decrease in the marginal cost of composite intermediate input is large enough to cover the fixed cost of intrafirm;

$$\varphi^{\sigma-1} B_n \left[\Delta C_{ni}^M(\tau_{ni}, \mathcal{I}(\varphi), \varphi, \varpi)\right] \geq w_i f_{ni}^{int}, \tag{7}$$

where $\Delta C_{ni}^M(\tau_{ni}, \mathcal{I}(\varphi), \varphi, \varpi) \equiv C_{ni}^M(\tau_{ni}, \mathcal{I}(\varphi) = 1, \varphi)^{1-\sigma} - C_{ni}^M(\tau_{ni}, \mathcal{I}(\varphi) = 0, \varphi)^{1-\sigma}$ denotes the gains in variable profits as a result of the decline in the marginal cost of composite intermediate input once an affiliate starts intrafirm trade with its parent. In the Appendix, we show that the left hand side of Equation (7) is continuos and strictly increasing in φ. As a result, there exists a productivity cutoff φ_{ni}^{int} such that all affiliates with productivity above it choose to import a fraction of its intermediate inputs from their parents, whereas, conditional on FDI, firms with productivity below it do not import from parents.

[33] $B_n \equiv \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} A_n$. Notice that the marginal cost of producing the final good is given by $\frac{C_{ni}^M(\varphi,.)}{\varphi}$, which we require to be strictly decreasing in φ. This can be done by imposing a specific functional form on $C_{ni}^M(\varphi)$ such that the marginal cost of the final good is decreasing in φ or, equivalently, we assume that firm's draw φ is transformed to actual firm's productivity via a strictly increasing function $f(\varphi)$ such that the marginal cost of the final good is decreasing in φ.

Proposition 1 *There exists a productivity cutoff* φ_{ni}^{int} *such that*

$$\mathcal{I}(\varphi) = \begin{cases} 1 & \text{if} \quad \varphi \geq \varphi_{ni}^{int} \\ 0 & \text{otherwise} \end{cases}$$

That is, only the most productive foreign affiliates in country n *engage in intrafirm trade with their parents (import intermediate inputs from their parents).*

The productivity cutoff φ_{ni}^{int} is simply pinned down from equation (7):

$$(\varphi_{ni}^{int})^{\sigma-1} B_n \left[\Delta C_{ni}^M(\tau_{ni}, \mathcal{I}(\varphi), \varphi_{ni}^{int}, \varpi)\right] = w_i f_{ni}^{int}. \tag{8}$$

The FDI cutoff $\varphi_{ni}^{fdi} < \varphi_{ni}^{int}$ is found by the usual way: equating export profits $\pi_{ni}(\varphi)$ with FDI profits without intrafirm π_{ni}^{fdi}

$$(\varphi_{ni}^{fdi})^{\sigma-1} B_n \left[C_{ni}^M(\tau_{ni}, \mathcal{I}(\varphi) = 0, \varphi_{ni}^{fdi}, \varpi)^{1-\sigma} - (w_i \tau_{ni})^{1-\sigma}\right] = w_i(f_{ni}^{fdi} - f_{ni}). \tag{9}$$

Exporting cutoff to country n is given by;,

$$\varphi_{ni}^{\sigma-1} B_n (w_i \tau_{ni})^{1-\sigma} - w_i f_{ni} = 0 \tag{10}$$

To complete the characterization of varieties produced and consumed in country n, the zero profit cutoff (ZPC) is as usual,

$$\varphi_{nn}^{\sigma-1} B_n w_n^{1-\sigma} - w_n f_{nn} = 0 \tag{11}$$

Parameter Restrictions and Firms Hierarchy: Consistent with the literature we impose the following restrictions on the model's parameters to sustain firms' hierarchy in the HMY.

- Exporters are more productive than nonexporters: $\varphi_{ii} < \varphi_{ni}$; if, under symmetric countries, $f_{ni} > \tau_{in}^{1-\sigma} f_{ii}$.
- Exporters are less productive than multinational firms: $\varphi_{ni} < \varphi_{ni}^{fdi}$; if $f_{ni}^{fdi} > (\tau_{ni}\varpi)^{\sigma-1}\bar{t}^{1-\sigma} f_{ni}$, and $\bar{t} < \varpi\tau_{ni}$.
- Multinational firms with nonzero intrafirm are more productive than multinational with zero intarfirm: $\varphi_{ni}^{int} > \varphi_{ni}^{fdi}$; if $f_{ni}^{int} > 0$.[34]

The geography of foreign affiliate sales: A country i foreign affiliate sales in country n, $r_{ni}^{aff}(\varphi)$ are given by

$$r_{ni}^{aff}(\varphi) = \sigma\varphi^{\sigma-1} B_n \left[C_{ni}^M(\tau_{ni}, \mathcal{I}(\varphi), \varphi, \varpi)\right]^{1-\sigma} \tag{12}$$

[34]In fact, f_{ni}^{int} has to be greater than $f_{ni}^{fdi} - f_{ni}$.

Proposition 2 ***(Gravity)****: Country i foreign affiliate sales (conditional on opening an affiliate) in country n,* $r_{ni}^{aff}(\varphi)$ *are decreasing in trade costs* τ_{ni}*. Let* $\varepsilon_{ni}^{r}(\varphi, \tau_{ni}) < 0$ *be the elasticity of affiliate sales with respect to trade costs, then the absolute value of* $\varepsilon_{ni}^{r}(\varphi, \tau_{ni})$ *is increasing in* φ*. In words, the sales of more knowledge intensive firms (affiliates) are more sensitive to trade costs. That is,* ***FDI-Gravity*** *is more pronounce for more knowledge intensive parents-affiliates.*

5 Parameterization, Functional Forms and Estimation

First, we provide functional forms of the log-supermodular function $\beta(z|\varphi)$, the cost of disembodied knowledge transfer, and the distribution of productivity draw. Before proceeding further, we set $\eta = 1$, and therefore M_φ is a Cobb-Douglas composite intermediate input: $M_\varphi = \mathcal{C}.exp\left\{\int_0^\infty \beta(z|\varphi)\ln m(z)dz\right\}$.[35] The correspondent cost function of the intermediate input composite: $C = exp\left\{\int_0^\infty \beta(z|\varphi)\ln w_z dz\right\}$. In our context, assuming $w_i = 1$ for $i \in \{1, 2, \ldots, N\}$, domestic producers composite intermediate input cost is given by $C_{nn}^M = 1$, while

$$C_{ni}^M(\tau_{ni}, \varphi, \mathcal{I}) = \begin{cases} \bar{t} & \text{if} \quad \mathcal{I} = 0, \\ \exp\left\{\int_0^{\tilde{z}} \beta(z|\varphi)\ln t(z)dz + \int_{\tilde{z}}^\infty \beta(z|\varphi)\ln \tau_{ni} dz\right\} & \text{if} \quad \mathcal{I} = 1 \end{cases}$$

Following Keller and Yeaple (2013), we set the knowledge transfer function $t(z) = \exp\{z\}$. Let $\phi(\varphi)$ denote φ-firm's knowledge intensity where $\phi(\varphi)$ is weakly increasing in φ. In order to simplify the analysis, we assume that $\phi(\varphi)$ takes two values low and high: $\phi(\varphi) \in \{\phi^l, \phi^h\}$. We adopted a very simple reduced form to connect the well documented relationship between firm's size (productivity) and knowledge intensity; specifically, for any $\varphi(\phi) > \varphi_{ni}^{int}$, $\phi = \phi^h$ and $\phi = \phi^l$ otherwise. This greatly simplifies the analysis without altering our results regarding the correlation between intrafirm trade and firm's knowledge-intensity. We still able to use this simple functional form to compare intrafirm trade across firms with different knowledge intensity. Accordingly, we change the notation slightly: we use $\beta(z|\phi)$ instead of $\beta(z|\varphi)$. The cost share function $\beta(z|\phi)$ is log-supermodular in z and ϕ; therefore, we let $\beta(z|\phi)$ be an exponential with parameter $\frac{1}{\phi}$.[36]

We additionally assume that the costs of disembodied technology transfer also vary with destination-original pair characteristics. Broadly, the factors that are widely used in estimating trade costs between countries are also expected to affect the costs of disembodied technology transfer but with less order of magnitude: $t_{ni}(z) = g_{ni}t(z)$. Hence, $\bar{t}_{ni} = g_{ni}\exp\left\{\int_0^\infty \beta(z|\phi)\ln t(z)dz\right\}$.[37]

[35] $\mathcal{C} \equiv \int_0^\infty \beta(z|\varphi)\ln \beta(z|\varphi)dz$ is constant.

[36] $\beta(z|\phi) = \frac{1}{\phi}\exp\left\{\frac{-z}{\phi}\right\}$. It is straightforward to check that $\log\beta(z|\phi)$ is supermodular and $\int_0^\infty \beta(z|\phi)dz = 1$.

[37] $1 < g_{ni} < \tau_{ni}$. Akin to τ_{ni}, g_{ni} denotes the costs of disembodied knowledge transfer as a function of distance, common border and language, the time zone of n, i, colonial origins,...

To operationalize the model we let $g_{ni} = \tau_{ni}^{\alpha}$, where $\alpha \in (0,1)$. With the functional forms at hand, the marginal cost of obtaining the composite intermediate input for an affiliate with knowledge intensity $\phi \in \{\phi^l, \phi^h\}$ is

$$C_{ni}^{M}(\tau_{ni}, \mathcal{I}, \phi) = \begin{cases} \bar{t} = \tau_{ni}^{\alpha}\text{exp}\{\phi\} & \text{if} \quad \mathcal{I} = 0, \\ \text{exp}\left\{\phi\left(1 - \tau_{ni}^{\frac{\alpha-1}{\phi}}\right) + \alpha\text{ln}\tau_{ni}\right\} & \text{if} \quad \mathcal{I} = 1 \end{cases} \tag{13}$$

Providing that $\tau_{ni} > g_{ni}\text{exp}\{\phi^l\}$.[38]

5.1 Foreign affiliate's sales: firm-level gravity

Foreign affiliate's sales are given by equation (12). Given the functional forms provided in this section, we have:

$$r_{ni}^{fdi} = \sigma\varphi^{\sigma-1}B_n\left(\tau_{ni}^{\alpha}exp(\phi)\right)^{1-\sigma}, \tag{14}$$

and

$$r_{ni}^{int} = \sigma\varphi^{\sigma-1}B_n\left(exp\left\{\phi(1 - \tau_{ni}^{\frac{\alpha-1}{\phi}}) + \alpha\text{ln}\tau_{ni}\right\}\right)^{1-\sigma}. \tag{15}$$

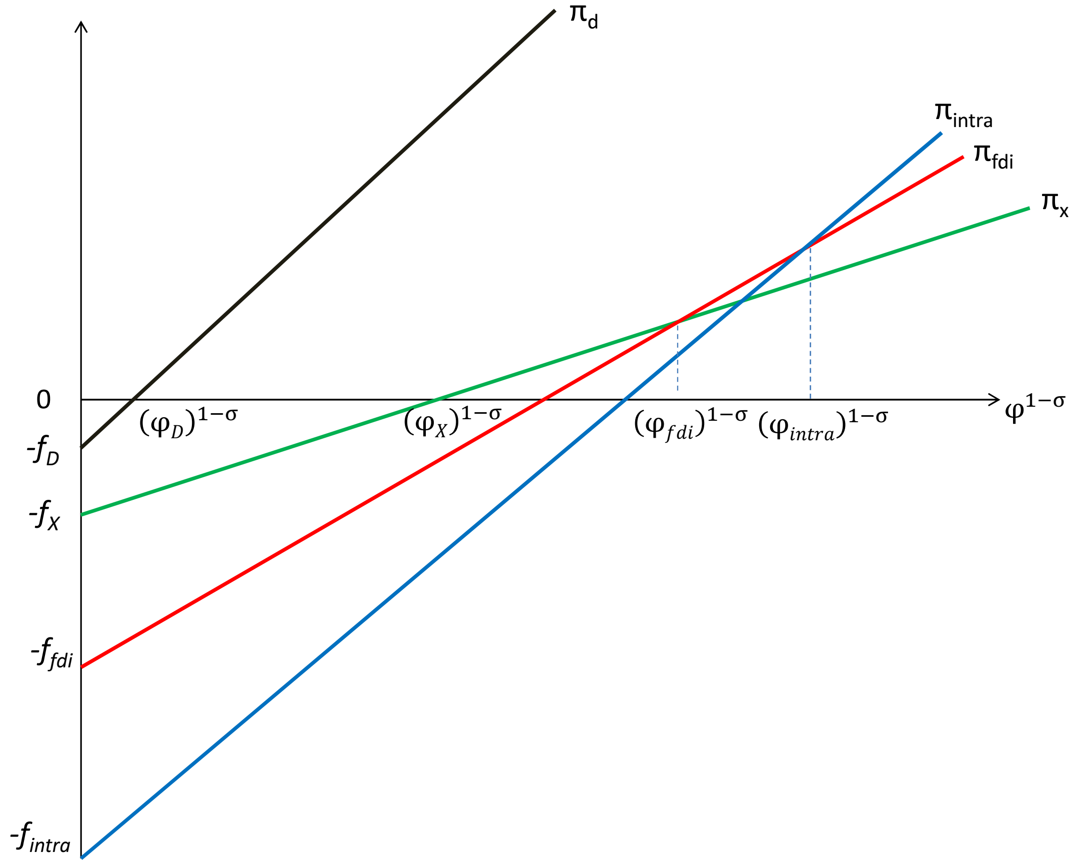
Accordingly, the elasticity of foreign affiliate's sales with respect to trade costs is given by;

$$\varepsilon_{ni}^{r}(\phi, \tau_{ni}, \mathcal{I}) = \begin{cases} (1-\sigma)\alpha < 0, & \text{if} \quad \mathcal{I} = 0, \\ (1-\sigma)\left((1-\alpha)\tau_{ni}^{\frac{\alpha-1}{\phi}} + \alpha\right) < 0, & \text{if} \quad \mathcal{I} = 1 \end{cases} \tag{16}$$

It is straightforward to verify that the sales of affiliates who import from their parents respond relatively more than the sales of affiliates who do do important from parents. Furthermore, for affiliates who import from their parents, their sales are more responsive to change in trade costs the higher the knowledge intensity: $\frac{\partial\varepsilon_{ni}^{r}(\phi,\tau_{ni})}{\partial\phi} < 0$.

[38]This assumption is needed in order for the FDI cutoff to be well defined. ϕ^l is very small such that $\text{exp}(\phi^l) \approx 1$.

Figure 1. Profit from domestic sales, exports, FDI and intra-firm trade



Notes: This figure shows the different productivity cutoff for different firms, where $(\varphi_D)^{1-\sigma}$ represents the cutoff for domestic producers, $(\varphi_X)^{1-\sigma}$ represents the cutoff for exporters, $(\varphi_{fdi})^{1-\sigma}$ represents the cutoff for firms engaging in multinational production, and $(\varphi_{intra})^{1-\sigma}$ represents the cutoff for foreign affiliates that also engage in intra-firm trade

6 General Equilibrium

Using the functional forms in the previous section, here we solve the economy general equilibrium aggregate variables. Consistent with the literature, we assume that firm's productivity is distributed Pareto with shape parameter κ,[39]

$$G(\varphi) = 1 - \varphi^{-\kappa}, \quad \text{for} \quad \varphi > 1, \quad \text{and} \quad \kappa > \sigma - 1.$$

The relevant cutoffs for country-pair (n, i) are given as follows[40]

$$\begin{aligned}
\text{Zero profit cutoff ZPC} : \varphi_{nn}^{\sigma-1} &= \frac{f_{nn}}{B_n} \\
\text{Export cuttoff} : \varphi_{ni}^{\sigma-1} &= \frac{f_{ni}}{B_n}\tau_{ni}^{\sigma-1} \\
\text{FDI cutoff} : (\varphi_{ni}^{fdi})^{\sigma-1} &= \frac{f_{ni}^{fdi} - f_{ni}}{B_n C_{1ni}} \\
\text{Intrafirm cutoff} : (\varphi_{ni}^{int})^{\sigma-1} &= \frac{f_{ni}^{int}}{B_n C_{2ni}}
\end{aligned}$$

As we show in Figure 1, the logic of the standard HMY model is strongly presented in our framework. At the heart of it is the proximity-concentration trade-off (in figure proximity is represented by the slope of each profit's line, while concentration is represented by y-axis intersection). However, in HMY model, the line representing the profits for affiliates who import π_{int} does not exist, and the line denoted by π_{fdi} is parallel to the domestic profits line. In a model with FDI and Intra-firm Irarrazabal et al. (2013), the line representing the profits for affiliates who do not import from parents π_{fdi} is missing since by default all affiliates import from parents.

[39]The assumption that $\kappa > \sigma - 1$ ensures the the distribution of firm's size has a finite mean. In general, $G(\varphi) = 1 - \left(\frac{\varphi_{min}}{\varphi}\right)^{\kappa}$, and $\kappa > 2$. We work with $\varphi_{min} = 1$. In this section we assume that all firms in a given sector share the same knowledge-intensity given by the mean of knowledge-intensity of all firms operating in that particular sector. Our main goal in this sectionderiving sectoral aggregate affiliates' sales (sectoral aggregate gravity) is preserved under this simplification. In this vein, our model becomes very similar to Keller and Yeaple (2013) with an exception of affiliates' endogenous selection into importers and non-importers from their parents.

[40]$C_{1ni} \equiv \tau_{ni}^{\alpha(1-\sigma)} \exp\{\phi(1-\sigma)\} - \tau_{ni}^{1-\sigma} > 0$.
$C_{2ni} \equiv \exp\{\phi(1 - \tau_{ni}^{\frac{\alpha-1}{\phi}}) + \alpha \ln \tau_{ni}\}^{1-\sigma} - \tau_{ni}^{\alpha(1-\sigma)} \exp\{\phi(1-\sigma)\} > 0$.

Aggregate price index: The aggregate price index in country n is given by;

$$P_n^{1-\sigma} = J_n \int_{\varphi_{nn}}^{\infty} p_{nn}(\varphi)^{1-\sigma} dG(\varphi) + \sum_{i \neq n}^{N} J_i \int_{\varphi_{ni}}^{\infty} p_{ni}(\varphi)^{1-\sigma} dG(\varphi), \tag{17}$$

$$p_{ni}(\varphi) = \begin{cases} \frac{\sigma}{\sigma-1} \frac{\tau_{ni}}{\varphi} & \text{if} \quad \varphi_{ni} < \varphi < \varphi_{ni}^{fdi} \\ \frac{\sigma}{\sigma-1} \frac{\tau_{ni}^{\alpha} exp(\phi)}{\varphi} & \text{if} \quad \varphi_{ni}^{fdi} < \varphi < \varphi_{ni}^{int} \\ \frac{\sigma}{\sigma-1} \frac{exp(\phi(1-\tau_{ni}^{\frac{\alpha-1}{\phi}}) + \alpha \ln \tau_{ni})}{\varphi} & \text{if} \quad \varphi_{ni}^{int} < \varphi \end{cases}$$

Evaluating the integration and using the Pareto distribution assumption,[41]

$$P_n^{-\kappa} = \frac{\kappa}{\kappa - (\sigma - 1)} \left(\frac{\sigma}{\sigma - 1} \right)^{-\kappa} \left(\frac{\mu X_n}{\sigma} \right)^{\frac{\kappa - (\sigma - 1)}{\sigma - 1}} \Xi_n \tag{18}$$

Indeed, X_n is an endogenous variable. Since the mass of firms is exogenously given, the aggregate profits of country n firms, including affiliates' profits, are strictly positive. Accordingly, total income/expenditure in country n is the sum of labor income and aggregate profits of all country n firms: $X_n = w_n L_n + \Pi_n$. As in Chaney (2008), we assume that each consumer in country n holds w_n shares in a completely diversified mutual global fund with s dividends per share in terms of the numeriare. Additionally, as in Eaton and Kortum (2002) and Chaney (2008), J_i is proportional to the size of labor force in country n; $J_n = w_n L_n$. Therefore, $X_n = w_n L_n (1 + s)$, and $J_n = \frac{X_n}{1+s}$. In the Appendix, we show that s is a function of the model's exogenous parameters: $s = \frac{\sigma - 1}{\sigma(\kappa - 1) + 1}$.

The aggregate equilibrium price level in country n is the solution of Equation (18) in terms of the model's exogenous parameters. Once P_n obtained, we can retrieve all the relevant cutoffs, trade flows, foreign affiliates' sales, and economic welfare.

6.1 Aggregate Sales: Gravity Equations

The model delivers three gravity equations: (i) Aggregate export sales from country i to country n: X_{ni}, (ii) Country i foreign affiliates' sales in country n, with no intrafirm between parents and affiliates; X_{ni}^{fdi}, and (iii) Country i foreign affiliates' sales in country n, for affiliates that import

[41] $\Xi_n \equiv \sum_{i=1}^{N} J_i \left(\tau_{ni}^{-\kappa} f_{ni}^{\frac{\sigma-1-\kappa}{\sigma-1}} + \mathcal{I}_{i \neq n} \left\{ (f_{ni}^{fdi} - f_{ni})^{\frac{\sigma-1-\kappa}{\sigma-1}} C_{1ni}^{\frac{\kappa}{\sigma-1}} + (f_{ni}^{int})^{\frac{\sigma-1-\kappa}{\sigma-1}} C_{2ni}^{\frac{\kappa}{\sigma-1}} \right\} \right)$. The indicator function $\mathcal{I}_{i \neq n} = 1$ if $i \neq n$ and zero otherwise.

from parents; X_{ni}^{int}.[42]

$$X_{ni} = \frac{\mu X_n X_i \tau_{ni}^{-\kappa}\, \delta_{ni}}{\Xi_n} \tag{19}$$

$$X_{ni}^{fdi} = \frac{\mu X_n X_i \left\{\tau_{ni}^{\alpha} exp(\phi)\right\}^{-\kappa} \lambda_{ni}}{\Xi_n} \tag{20}$$

$$X_{ni}^{int} = \frac{\mu X_n X_i exp\left\{\phi\left(1-\tau_{ni}^{\frac{\alpha-1}{\phi}}\right) + \alpha(\ln\tau_{ni})\right\}^{-\kappa} \vartheta_{ni}}{\Xi_n} \tag{21}$$

Ξ_n is a reminiscent of the multilateral resistance term in Eaton and Kortum (2002). It is a measure of country n attractiveness (remoteness) taking into account all trading countries. The bilateral terms δ_{ni}, λ_{ni}, and ϑ_{ni} depend only on country i and country n parameters.[43] Relative to the standard gravity equation (e.g., Melitz-Chaney style model with no FDI), the impact of variable trade costs on country i exporters to country n is more involved. Without FDI sales, country i aggregate exports to country n can be decomposed into the intensive and the extensive margins, with the average exporter's sales being invariant to variable trade costs and the mass of exporting firms negatively associated with trade costs. In the presence of FDI sales, variable trade costs impact both the mass of exporters and the average export sales per firm. In Chaney (2008), for instance, δ_{ni} is a function of fixed costs of export f_{ni}, and does not depend on τ_{ni}. Here, λ_{ni} is a function of τ_{ni}, and therefore the response of X_{ni} to changes in τ_{ni} depends on changes in δ_{ni} and $\tau_{ni}^{-\kappa}$. Formally, let $\xi_{X,\tau}$ be the elasticity of aggregate exports sales between countries i and n with respect to variable trade costs τ_{ni}, and $\xi_{\delta,\tau}$ is the elasticity of δ with respect to τ, then[44]

$$\xi_{X,\tau} = -\kappa - |\xi_{\delta,\tau}| < 0, \tag{22}$$

Likewise, the elasticity of of aggregate foreign affiliate sales for affiliates that do not import from their parents with respect to variable trade costs, and the elasticity of aggregate foreign affiliates'

[42] With a slight abuse of notation, we redefine $\Xi_n \equiv \sum_{i=1}^{N} L_i(1+s)\left(\tau_{ni}^{-\kappa} f_{ni}^{\frac{\sigma-1-\kappa}{\sigma-1}} + \mathcal{I}_{i\neq n}\left\{(f_{ni}^{fdi}-f_{ni})^{\frac{\sigma-1-\kappa}{\sigma-1}} C_{1ni}^{\frac{\kappa}{\sigma-1}} + (f_{ni}^{int})^{\frac{\sigma-1-\kappa}{\sigma-1}} C_{2ni}^{\frac{\kappa}{\sigma-1}}\right\}\right)$.

[43] $\delta_{ni} \equiv f_{ni}^{\frac{\sigma-1-\kappa}{\sigma-1}} - \left[\frac{f_{ni}^{fdi}-f_{ni}}{\tau_{ni}^{(1-\sigma)(\alpha-1)} exp(\phi(1-\sigma))-1}\right]^{\frac{\sigma-1-\kappa}{\sigma-1}}$, $\lambda_{ni} \equiv \left[\frac{f_{ni}^{fdi}-f_{ni}}{1-\tau_{ni}^{(1-\sigma)(1-\alpha)} exp(\phi(\sigma-1))}\right]^{\frac{\sigma-1-\kappa}{\sigma-1}} - \left[\frac{f_{ni}^{int}}{(\tau_{ni}^{\alpha} exp(\phi))^{\sigma-1} C_{2ni}}\right]^{\frac{\sigma-1-\kappa}{\sigma-1}}$, and $\vartheta_{ni} \equiv \left[\frac{f_{ni}^{int}}{\left(\exp((\phi(1-\tau_{ni}^{\frac{\alpha-1}{\phi}})+\alpha\ln\tau_{ni})\right)^{\sigma-1} C_{2ni}}\right]^{\frac{\sigma-1-\kappa}{\sigma-1}}$. Our assumptions about firms hierarchy and the necessary parameter restrictions to maintain it are sufficient for both δ_{ni} and ϑ_{ni} to be positive. On the other hand λ_{ni} is positive if $f_{ni}^{int} > (f_{ni}^{fdi} - f_{ni})\frac{C_{2ni}}{C_{1ni}}$.

[44] $\xi_{\delta,\tau} = -\frac{\kappa-(\sigma-1)}{\sigma-1}\left[\frac{f_{ni}^{fdi}-f_{ni}}{\tau^{\sigma-1}C_{1ni}}\right]^{\frac{\sigma-1-\kappa}{\sigma-1}-1}\left[\frac{(1-\sigma)(\alpha-1)\tau^{(1-\sigma)(\alpha-1)-1}exp(\phi(1-\sigma)}{(\tau^{\sigma-1}C_{1ni})^2}\right]\frac{\tau}{\delta} < 0$.

sales for affiliates the import from their parents are, respectively, given by[45]

$$\xi_{X^{fdi},\tau} = -\alpha\kappa + \xi_{\lambda,\tau} < 0 \tag{23}$$

$$\xi_{X^{int},\tau} = -\left[\tau_{ni}^{\frac{\alpha-1}{\phi}}(1-\alpha) + \alpha\right]\kappa + \xi_{\vartheta,\tau} < 0 \tag{24}$$

Aggregate affiliates' sales (for importer affiliates) decrease as trade costs increase. It is straightforward to show this since the second term of Equation (24) is negative for any $\alpha \in (0,1)$. The finding that foreign affiliates' sales are negatively correlated with trade costs for the affiliates who import from their parents is not surprising and consistent with the models that introduce intrafirm trade between affiliates and parents Irarrazabal et al. (2013) and Keller and Yeaple (2013). We are mainly interested in the gravity equation for affiliates who report zero intrafirm with their parents. The intrafirm trade mechanism that puts gravity forces in play is ceased in the case of small affiliates who never import from parents. Nonetheless, as we show in equation (23), the sales of non-importer affiliates are still suffering from gravity forces (see the Appendix for formal derivations and the conditions for DI gravity to hold). In our context, affiliates need the knowledge-specific to produce the final good, which it can obtain through importing intermediate inputs from parents-embodying knowledge- or through direct knowledge transfer, which is not observed in the data. Since trade frictions impact the cost of knowledge transfer, affiliates' marginal cost and sales are negatively affected by the distance from headquarter and other common trade frictions.

In order to comment on the role of intensive and extensive margins in the gravity equations above, in a line with Chaney (2008), we formally introduce the impact of changing variable trade costs on the intensive margin (sales of existing firms) and the extensive margin (sales of new entrants). By differentiating the expression for aggregate exports from country i to country n $X_{ni} = J_i \int_{\varphi_{ni}}^{\varphi_{ni}^{fdi}} r_{ni}(\varphi)dG(\varphi)$, the following expression for the elasticity of X_{ni} with respect to τ_{ni} is obtained,[46]

$$\xi_{X,\tau} = \overbrace{(1-\sigma)}^{\text{Intensive margin}} + \overbrace{\frac{\kappa-(\sigma-1)}{\varphi_{ni}^{\sigma-1-\kappa} - (\varphi_{ni}^{fdi})^{\sigma-1-\kappa}}\left[\xi_{\varphi^{fdi},\tau}(\varphi_{ni}^{fdi})^{\sigma-1-\kappa} - \varphi_{ni}^{\sigma-1-\kappa}\right]}^{\text{Extensive margin}}, \tag{25}$$

where, $\xi_{\varphi^{fdi},\tau}$ denotes the elasticity of FDI cutoff with respect to variable trade costs. If $\xi_{\varphi^{fdi},\tau}$ is negative then both the sales of existing exporters and the sales of new exporters decrease with trade costs. By contrast, if α is large enough, $\xi_{\varphi^{fdi},\tau}$ is positive; yet it is still small enough such that the extensive margin continues to be negative. In fact, $\xi_{\varphi^{fdi},\tau} < 1$ for any value of

[45] $\xi_{\vartheta,\tau} = (\kappa-(\sigma-1))\left[\tau^{\alpha(1-\sigma)}\frac{exp(\phi(1-\sigma))}{C_{2ni}}\left((\alpha-1)\tau^{\frac{\alpha-1}{\phi}}\right)\right] < 0$. Deriving the sign of $\xi_{\lambda,\tau}$ involves a tremendous algebra and is not trivial. In general, $\xi_{\lambda,\tau}$ is negative if α and f_{ni}^{int} are large enough (see the Appendix for details). Nonetheless, $\xi_{X^{fdi},\tau}$ is negative as long as α is not very close to zero.

[46] We use Leibniz integral rule to differentiate the aggregate exports expression.

$\alpha \in (0,1)$.[47] Consistent with our finding that the number of foreign affiliates in the lower tail of firm's size distribution decreases as the distance from headquarter increases, we proceed with positive elasticity of FDI cutoff with respect to trade costs, $0 < \xi_{\varphi^{fdi},\tau} < 1$ (i,e., α is large enough). Interestingly, even if the FDI cutoff is increasing in τ, as in HMY, the ratio of the number of multinational firms to the number of exporters increases as trade costs increase. Clearly, if FDI cutoff is ∞, the model collapses to Chaney's model and $\xi_{X,\tau} = -\kappa$.

The same analysis for the aggregate sales of affiliates who do not import from parents, X_{ni}^{fdi} is executed,

$$\xi_{X^{fdi},\tau} = \overbrace{\alpha(1-\sigma)}^{\text{Intensive margin}} + \overbrace{\frac{\kappa-(\sigma-1)}{(\varphi_{ni}^{fdi})^{\sigma-1-\kappa} - (\varphi_{ni}^{int})^{\sigma-1-\kappa}} \left[\xi_{\varphi^{int},\tau}(\varphi_{ni}^{int})^{\sigma-1-\kappa} - \xi_{\varphi^{fdi},\tau}(\varphi_{ni}^{fdi})^{\sigma-1-\kappa}\right]}^{\text{Extensive margin}} \tag{26}$$

The elasticity of intra-firm cutoff with respect to variable trade costs is denoted by $\xi_{\varphi^{int},\tau}$.[48] In the Appendix we show that if the fixed cost of intrafirm trade is sufficiently hight, the impact of trade costs on the extensive margin is negative as well.[49]

The impact of variable trade costs on the intensive and the extensive margins for affiliates who import from their parents is as follows

$$\xi_{X^{int},\tau} = \overbrace{(1-\sigma)\left[(1-\alpha)\tau_{ni}^{\frac{\alpha-1}{\phi}} + \alpha\right]}^{\text{Intensive margin}} - \overbrace{(\kappa-(\sigma-1))\xi_{\varphi^{int},\tau}}^{\text{Extensive margin}}. \tag{27}$$

Both sales per existing affiliates and the sales of new importer affiliates decline as trade costs increase. An intriguing result here is that although the impact of trade costs on the intensive margin unambiguously larger for importer-affiliates than non-importer affiliates, the relative impact on the extensive margin for non-importer affiliates relative to importer affiliates is ambiguous: the sales of new entrants/existing non-importer affiliates might decline more than its counterpart for importer-affiliates as trade costs increase. In effect, the overall impact of trade costs on the aggregate sales of non-importer affiliates might even be stronger than its impact on the overall sales of importer affiliates because of the extensive margin responses to increasing trade costs. In other word, gravity forces could be stronger for affiliates who do not report intrafirm relative to affiliates who import from parents.

[47] Specifically, $\xi_{\varphi^{fdi},\tau} = \frac{\alpha exp(\phi(1-\sigma)) - \tau_{ni}^{(1-\sigma)(1-\alpha)}}{exp(\phi(1-\sigma)) - \tau_{ni}^{(1-\sigma)(1-\alpha)}} < 1$.

[48] $\xi_{\varphi^{int},\tau} = \frac{1}{1-\sigma}\frac{\partial \ln C_{2ni}}{\partial \ln \tau} > 0$.

[49] In fact, we also show the conditions under which FDI gravity equation holds even with positive extensive margin. In general this will be the case for a wide range of parameter values.

7 Conclusion

This paper starts by documenting an empirical regularity that cannot be fully taken into account by existing theoretical frameworks: foreign affiliates' sales are decreasing in trade costs even for those affiliates who do not engage in intra-firm transactions. In order to close this gap, we propose a new theoretical framework to rationalize this finding together with another stylize fact: the majority of firms do not engage in intra-firm transactions and even among those that do, intra-firm trade is highly concentrated in a small set of large multinational firms. Internalizing these regularities into an unified model improves our understanding of the nature and structure of multinational firms and the complex network connections between parents and affiliates. In addition, it provides a guide to further develop a quantitative framework that allows us to measure the welfare gains associated to reduction in trade barriers in a granular economy, where not only exports and multinational activity are subject to selection and are concentrated in a few big firms, but also the intra-firm transaction across borders.

This paper is part of a larger research agenda which attempts to quantify the potentially large gains from trade as well as the gains from multinational production that take place in an economy where trade liberalization will not only impact physical trade but also transfer of knowledge across countries. This could affect the employment in the host and the home country, and consequently could have sizable implications in the skilled composition of workers in both economies. Moreover, the interaction between trade costs and knowledge transfer across firms might be an useful tool to advance the theory of the boundaries of multinational firms.

Finally, relying on the quantitative extension of the presented theoretical framework, we will proceed to evaluate the impact of idiosyncratic firms' shocks on aggregate outcomes, in an economy in which production and trade—both arm's length and intra-firm—are concentrated among a small number of large multinational corporations. In particular, compared with a counterfactual scenario in which producing overseas is prohibitively costly for all firms, the observed aggregated volatility is expected to be significantly lower. This can be explained by the fact that in granular economies a large fraction of the small set of firms that dominate the market are indeed foreign affiliate firms rather than solely local exporters. The empirical analysis will use firm-level data to track the transactions of U.S. parents with their foreign multinationals and to explore the heterogeneity of intra-firm trade in the upper tail of the distribution and the persistence of gravity along firms of all sizes.

References

Alfaro, Laura and Maggie X. Chen, "Market Reallocation and Knowledge Spillover: The Gains from Multinational Production," Harvard Business School Working Papers 12-111, Harvard Business School June 2012.

Borga, Maria and William J. Zeile, "International Fragmentation of Production and the Intrafirm Trade of U.S. Multinational Companies," BEA Papers 0035, Bureau of Economic Analysis January 2004.

Brainard, Lael S., "An Empirical Assessment of the Proximity-Concentration Trade-off between Multinational Sales and Trade," *American Economic Review*, September 1997, *87* (4), 520–44.

Chaney, Thomas, "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, September 2008, *98* (4), 1707–21.

Costinot, Arnaud, "An Elementary Theory of Comparative Advantage," *Econometrica*, july 2009, *77* (4), 1165–1192.

di Giovanni, Julian and Andrei A. Levchenko, "Country Size, International Trade, and Aggregate Fluctuations in Granular Economies," *Journal of Political Economy*, 2012, *120* (6), 1083 – 1132.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, September 2002, *70* (5), 1741–1779.

Gabaix, Xavier, "Power Laws in Economics and Finance," *Annual Review of Economics*, may 2009, *1* (1), 255–294.

Helpman, Elhanan, "A Simple Theory of International Trade with Multinational Corporations," *Journal of Political Economy*, June 1984, *92* (3), 451–71.

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

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

Keane, Michael P. and Susan E. Feinberg, "Accounting for the Growth of MNC-Based Trade Using a Structural Model of U.S. MNCs," *American Economic Review*, December 2006, *96* (5), 1515–1558.

Keller, Wolfgang and Stephen R. Yeaple, "The Gravity of Knowledge," *American Economic Review*, June 2013, *103* (4), 1414–44.

Markusen, James R., *Multinational Firms and the Theory of International Trade*, MIT press, 2004.

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

_ , Veronica Rappoport, and Kim J. Ruhl, "Horizontal versus Vertical Foreign Direct Investment: Evidence from U.S. Multinationals," Working Papers, New York University 2014.

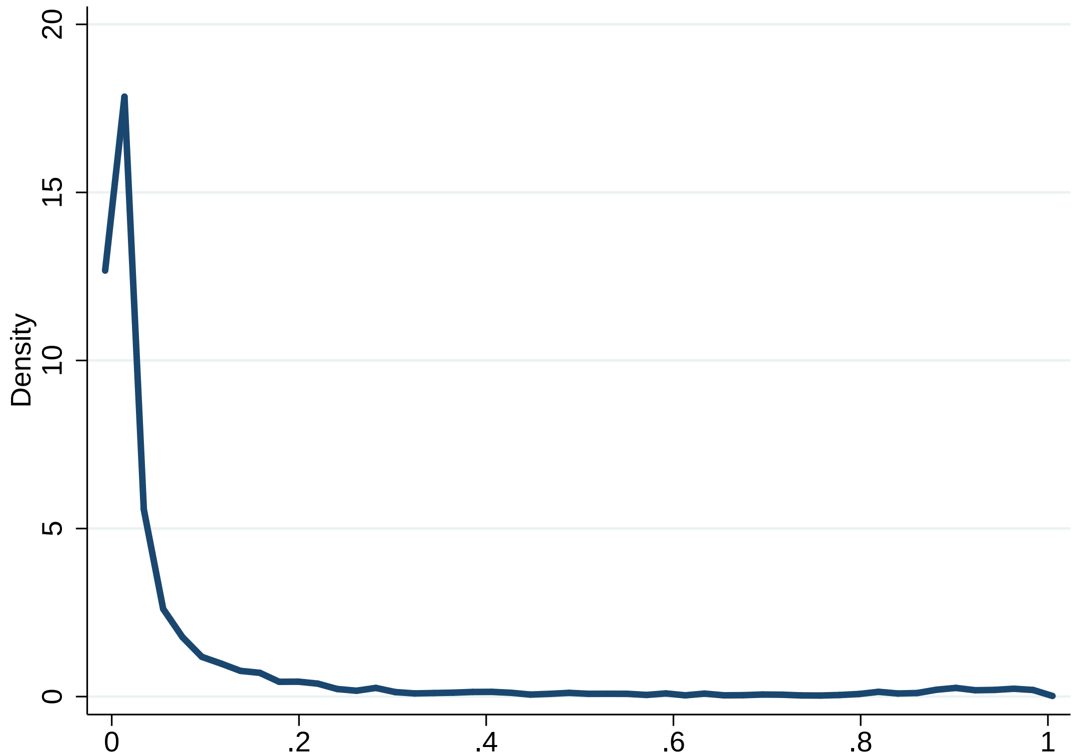
Seuring, Stefan and Maria Goldbach, *Cost Management in Supply Chains*, Springer, 2002.

Yeaple, Stephen R., "Firm Heterogeneity and the Structure of U.S. Multinational Activity," *Journal of International Economics*, July 2009, *78* (2), 206–215.

Zeile, William J., "Trade in Goods Within Multinational Companies: Survey-Based Data and Findings for the United States of America," BEA Papers 0033, Bureau of Economic Analysis November 2003.

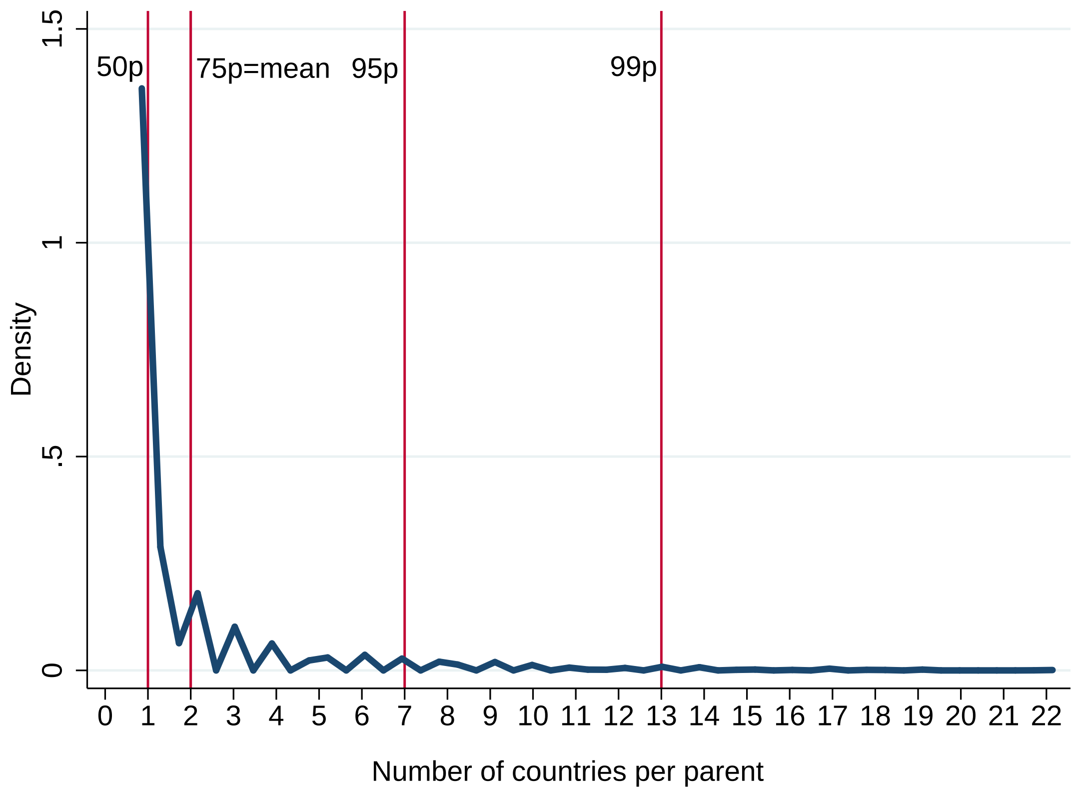
Tables and Figures

Figure 2. Density of U.S Foreign Affiliate Sales



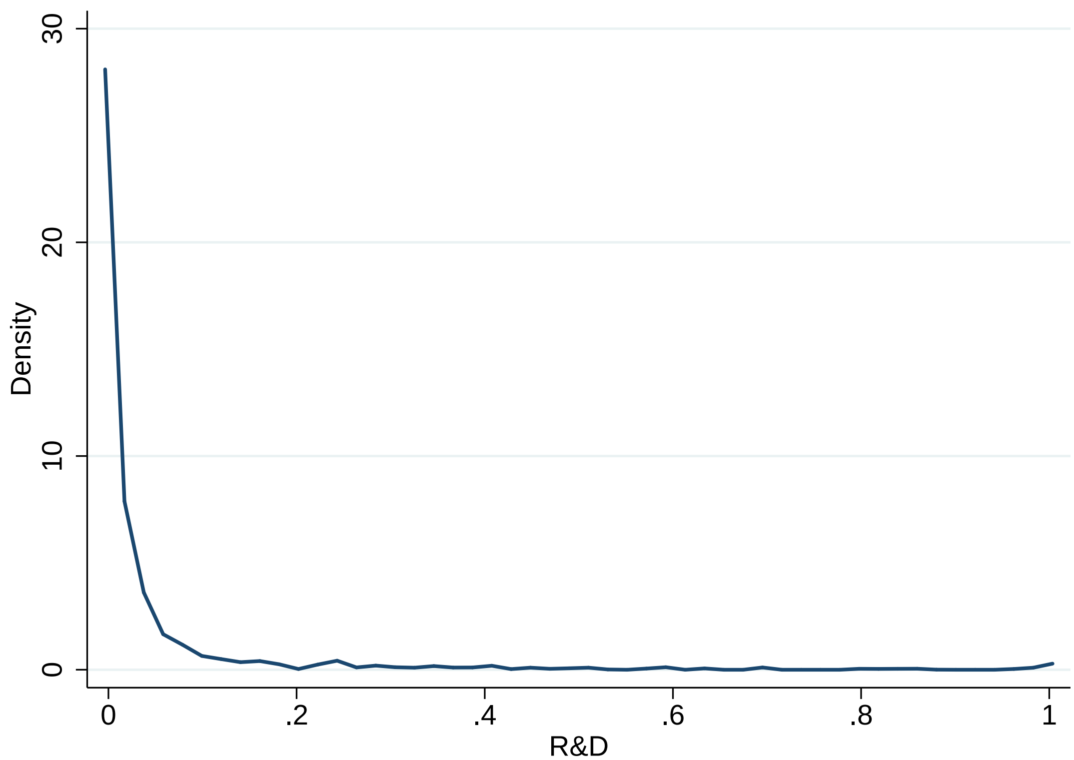
This figure represents the density of U.S Foreign Affiliate Sales as a fraction of parent's sales. As can be observe the distribution of foreign affiliate sales is dominated by small affiliates, with the presence of only few but large affiliates firms that account for the majority of the U.S. multinational production.

Figure 3. Market Penetration



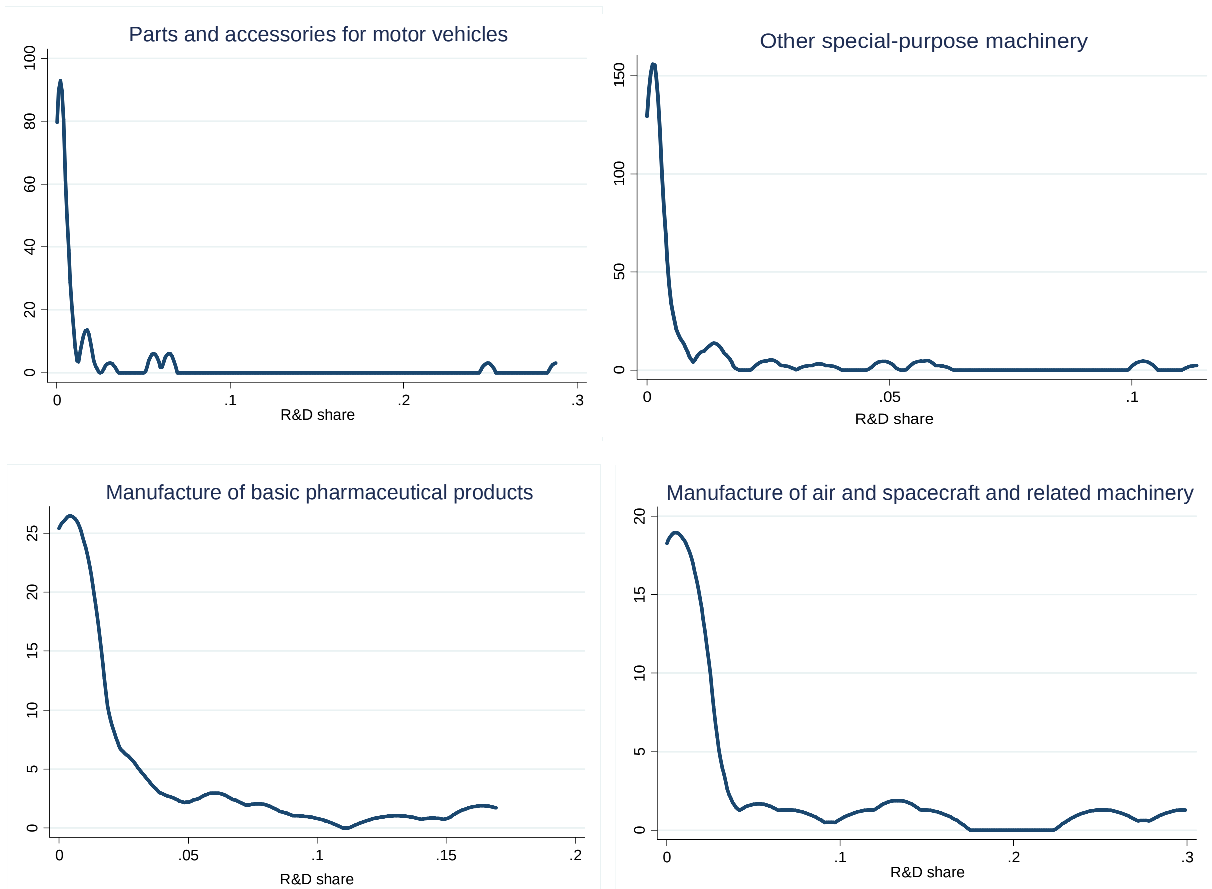
This figure represents the density of the number of markets in which U.S. parent firms produce. The vertical lines represent the cutoff the the 50, 75, 95 and 99 percentile, respectively. Half of the firms only have operations in only two or one foreign country. Only few parents engage in multinational activity in more than seven foreign markets.

Figure 4. Research and Development Share



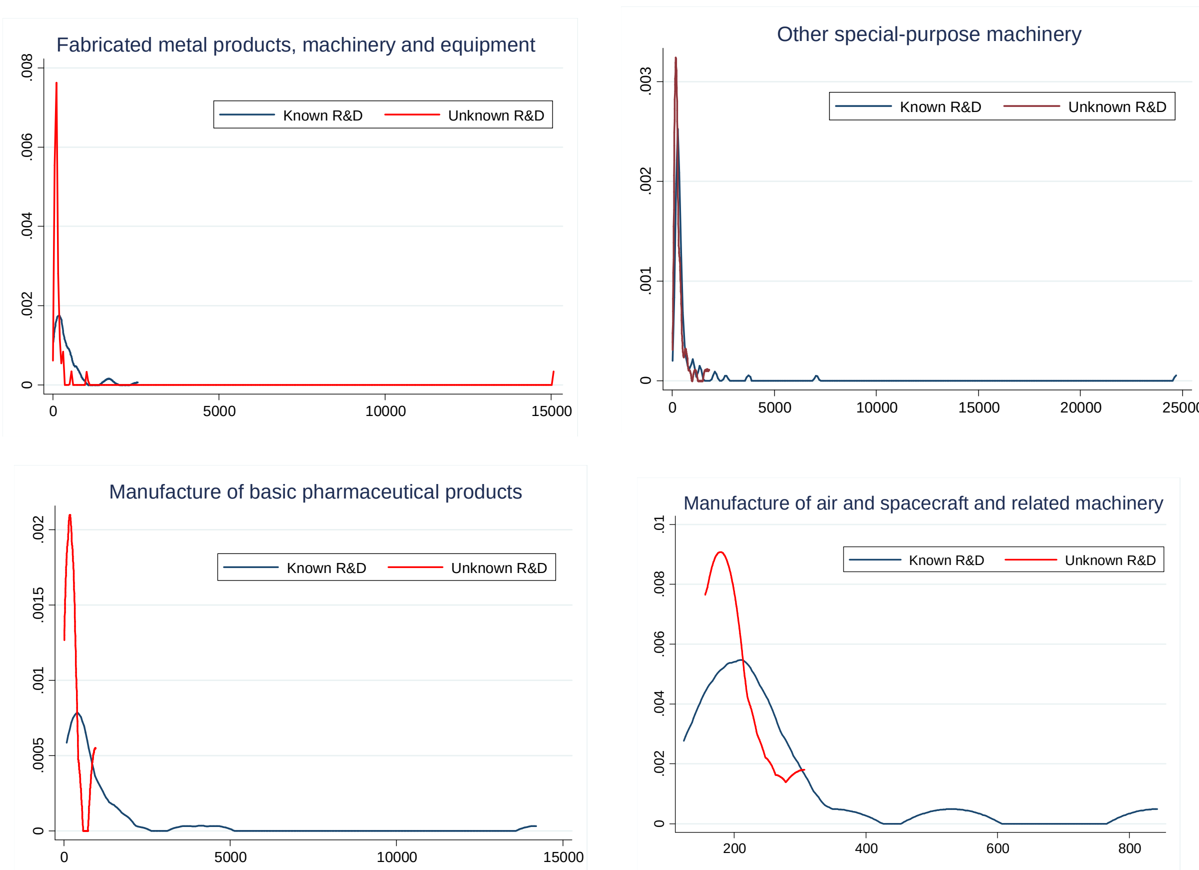
This figure shows the density of the parent's R&D expenditure share in each of the 104 manufactured sectors of the NACE classification at three digit level of disaggregation. The share of R&D is calculated as the fraction of the total Research and Development expenditure of the firm relative to the total R&D expenditure of all U.S parents firms operating the the same NACE3 sectoral classification. The density is showed for the pool of firms-sectors in the sample

Figure 5. Density of Firms' R&D shares for selected industries



This figure shows the density of the parent's share of R&D expenditure share for four selected three-digit level of NACE sector classification: 1) manufacturing of parts and accessories for motor vehicles—NACE 293 (top-left panel); 2) manufacture of other special-purpose machinery—NACE 289 (top-right panel); 3) manufacture of basic pharmaceutical products—NACE 211 (bottom-left panel), and 4) manufacture of air and spacecraft and related machinery—NACE 303 (bottom-right panel). The share of R&D is calculated as the fraction of the total Research and Development expenditure of the firm relative to the total R&D expenditure of all U.S parents firms operating the the same 3 digit sectoral classification.

Figure 6. Density of Fimrs' Producivity by R&D group



This figure shows the density of the productivity for two groups of parent firms: those for which Orbis data contains information regarding the expenditure in research and development activities (Known R&D); and those parent firms that contain missing values for R&D (Unknown R&D). The productivity density is show for both groups in four selected industries at three-digit level of NACE sector classification: 1) manufacturing of parts and accessories for motor vehicles—NACE 293 (top-left panel); 2) manufacture of other special-purpose machinery—NACE 289 (top-right panel); 3) manufacture of basic pharmaceutical products—NACE 211 (bottom-left panel), and 4) manufacture of air and spacecraft and related machinery—NACE 303 (bottom-right panel). Firms productivity is measured by the output per worker of the U.S parent. Only one third of the U.S parent firms (that at least have one affiliate overseas) show positive values of R&D expenditures.

Table 1. Gravity Equation of MP (country-sector level)

Dep. Variable	MP sales		N. of firms	
	(1)	(2)	(3)	(4)
ln *Dist*	–0.6999***	–0.7044***	–0.5525***	–0.0893**
	(0.2165)	(0.1190)	(0.0687)	(0.0427)
Border	–2.8115***	–0.6675	–0.0831	1.7173***
	(0.6920)	(0.6752)	(0.2231)	(0.2990)
Language	0.3939*	1.0690	0.3643**	0.3681
	(0.2069)	(0.9952)	(0.0711)	(0.3664)
Colony	–0.0449	1.1371***	0.5632**	1.9199***
	(0.1822)	(0.3812)	(0.0648)	(0.1366)
RTA	1.5951***	1.5730**	–0.1314	–0.3411*
	(0.3972)	(0.7277)	(0.1237)	(0.1896)
Capital (relative US)	2.1510***		2.1014***	
	(0.6134)		(0.2844)	
ln *GDPperc*	0.5664*		0.7646***	
	(0.3928)		(0.1184)	
Rule of Law	–0.0198**		–0.0196**	
	(0.0094)		(0.0028)	
Country FE	No	Yes	No	Yes
Sector FE	Yes	Yes	Yes	Yes
N.Observations	1111	1111	1194	1194

Notes: Dependent variables: foreign affiliates sales relative to parent's sales operating in each host country-sector pair in column (1) and (2); Number of US parents with at least one affiliates in each host country-sector pair in column (3) and (4). The regressors include the natural log of the distance between U.S and the host market (ln *Dist*); a dummy for the participation of the host market in a regional trade agreement (*RTA*), a dummy of common border (*border*), common language (*language*) and whether or not the host market and U.S. had a colonial relationship (*colony*). Other controls includes the level of capital endowment (*Capital*), the natural log of GDP per capita and a measure of the institutional quality of the host country (*Rule of Law*). Robust standard errors reported in parentheses. Significance is denoted: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 2. Gravity Equation of MP (country-sector level)

Dep. Variable	Number of firms			
	(<50p)	(>50p)	(<50p)	(>50p)
ln *Dist*	–0.3536***	–0.5443***	–0.4954***	–0.4774***
	(0.0573)	(0.0819)	(0.0743)	(0.0672)
Border	n/a	n/a	–0.3300	-0.1060
			(0.2440)	(0.2142)
Language	0.1747	0.7719***	0.2371***	0.3315***
	(0.1282)	(0.1938)	(0.0790)	(0.0738)
Colony	1.3362***	0.7263**	0.4838***	0.3913***
	(0.1330)	(0.2360)	(0.0651)	(0.0608)
RTA	0.5152***	–0.2478	0.0628	–0.1443
	(0.1543)	(0.2291)	(0.1388)	(0.1243)
Capital (relative US)			1.8563***	1.6531***
			(0.2701)	(0.2720)
ln *GDPperc*			0.5076***	0.4968***
			(0.1232)	(0.1153)
Rule of Law			–0.0129***	–0.0136***
			(0.0031)	(0.0029)
Country FE	Yes	Yes	No	No
Sector FE	Yes	Yes	Yes	Yes
N.Observations	938	942	938	942

Notes: Dependent variables: Number of US parents with at least one affiliates in each host country-sector pair. The regressors include the natural log of the distance between U.S and the host market (ln *Dist*); a dummy for the participation of the host market in a regional trade agreement (*RTA*), a dummy of common border (*border*), common language (*language*) and whether or not the host market and U.S. had a colonial relationship (*colony*). Other controls includes the level of capital endowment (*Capital*), the natural log of GDP per capita and a measure of the institutional quality of the host country (*Rule of Law*). Robust standard errors reported in parentheses. Significance is denoted: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. n/a mean not available, because a variable is collinear with country fixed effects

Table 3. Gravity Equation of MP (country-sector level)

Dep. Variable	MP Sales			
	(<50p)	(>50p)	(<50p)	(>50p)
ln *Dist*	–0.8568***	–0.8399***	–0.2594	–0.4007*
	(0.1601)	(0.1235)	(0.0687)	(0.2660)
Border	–0.9984**	–0.3156	–1.4844	–2.3741***
	(0.4678)	(0.7528)	(0.2231)	(0.8410)
Language	1.6716	1.1554	0.3216**	0.3004
	(1.4587)	(0.9505)	(0.0711)	(0.2294)
Colony	0.3474	1.2135**	0.5915**	–0.0531
	(1.3918)	(0.4905)	(0.0648)	(0.1904)
RTA	0.8893	1.5286**	–0.7275	1.7651*
	(1.3877)	(0.6900)	(0.1237)	(0.1896)
Capital (relative US)			1.4399***	1.8031***
			(0.2844)	(0.6635)
ln *GDPperc*			0.7646***	0.9009**
			(0.1184)	(0.4539)
Rule of Law			–0.0196**	-0.0197*
			(0.0028)	(0.0102)
Country FE	Yes	Yes	No	No
Sector FE	Yes	Yes	Yes	Yes
N.Observations	875	904	875	904

Notes: Dependent variables: foreign affiliates sales relative to parent's sales operating in each host country-sector pair. The regressors include the natural log of the distance between U.S and the host market (ln *Dist*); a dummy for the participation of the host market in a regional trade agreement (*RTA*), a dummy of common border (*border*), common language (*language*) and whether or not the host market and U.S. had a colonial relationship (*colony*). Other controls includes the level of capital endowment (*Capital*), the natural log of GDP per capita and a measure of the institutional quality of the host country (*Rule of Law*). Robust standard errors reported in parentheses. Significance is denoted: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 4. Gravity Equation of MP (country-sector level)

Dep. Variable	Number of firms			
	(<25p)	(>75p)	(<25p)	(>75p)
ln *Dist*	–0.2780***	–0.4082***	–0.3604***	–0.2499***
	(0.0289)	(0.0479)	(0.0674)	(0.0600)
Border	n/a	n/a	–0.1449	0.1369
			(0.2123)	(0.1782)
Language	0.0830	0.3979***	0.2649***	0.3179***
	(0.2251)	(0.0976)	(0.0732)	(0.0662)
Colony	0.2217	0.4821***	0.3449***	0.2496***
	(0.2114)	(0.1272)	(0.0616)	(0.0534)
RTA	0.3339***	–0.4197***	–0.0920	–0.1666*
	(0.1089)	(0.1226)	(0.1129)	(0.1011)
Capital (relative US)			1.3084***	1.2075***
			(0.2844)	(0.2343)
ln *GDPperc*			0.3277***	0.2043**
			(0.1117)	(0.0926)
Rule of Law			–0.0080**	-0.0061***
			(0.0029)	(0.0023)
Country FE	No	Yes	No	Yes
Sector FE	Yes	Yes	Yes	Yes
N.Observations	762	780	762	780

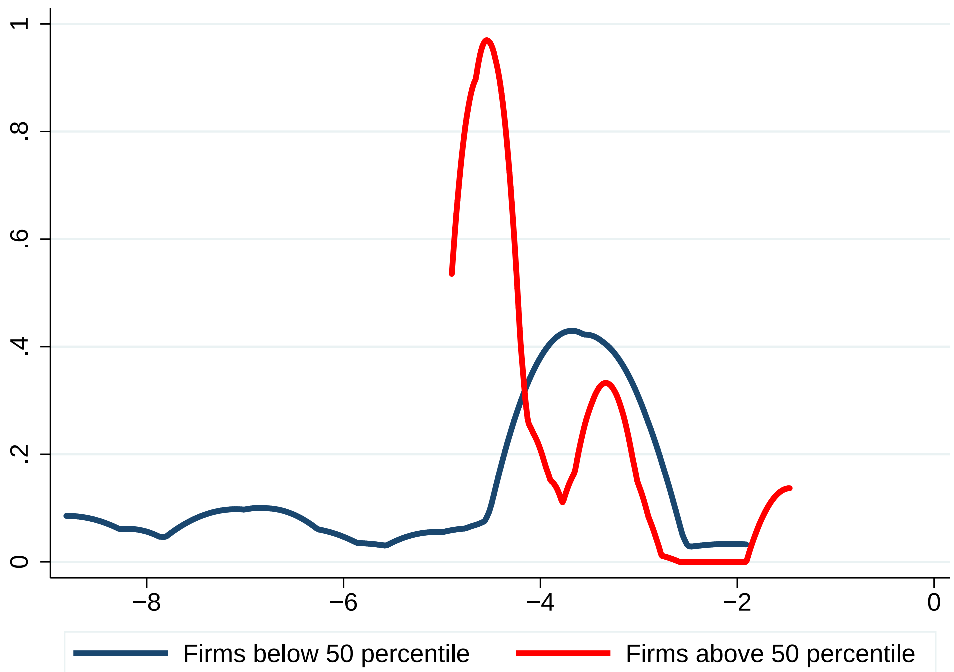
Notes: Dependent variables: Number of US parents with at least one affiliates in each host country-sector pair. The regressors include the natural log of the distance between U.S and the host market (ln *Dist*); a dummy for the participation of the host market in a regional trade agreement (*RTA*), a dummy of common border (*border*), common language (*language*) and whether or not the host market and U.S. had a colonial relationship (*colony*). Other controls includes the level of capital endowment (*Capital*), the natural log of GDP per capita and a measure of the institutional quality of the host country (*Rule of Law*). Robust standard errors reported in parentheses. Significance is denoted: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. n/a mean not available, because a variable is collinear with country fixed effects

Table 5. Gravity Equation of MP (country-sector level)

Dep. Variable	MP Sales			
	(<25p)	(>75p)	(<25p)	(>75p)
ln *Dist*	–1.0994***	–0.7857***	–0.6170*	–0.8584***
	(0.1897)	(0.1464)	(0.3553)	(0.2971)
Border	–5.0072**	–0.2373	–1.8248	–3.4116***
	(0.2.1279)	(0.7628)	(1.2137)	(0.9378)
Language	3.8818*	1.7033	0.5525**	0.6566**
	(1.1.7287)	(1.1007)	(0.2874)	(0.2631)
Colony	2.1475	1.8609	–1.0250***	–0.0164
	(1.7119)	(1.2184)	(0.2286)	(0.2220)
RTA	2.1991	2.1108**	0.3802	1.8904***
	(1.6473)	(0.9123)	(0.6421)	(0.4956)
Capital (relative US)			1.0083***	2.5191***
			(0.9921)	(0.7881)
ln *GDPperc*			–0.4518	1.1348**
			(0.5731)	(0.5132)
Rule of Law			0.0083**	–0.0318***
			(0.0146)	(0.0114)
Country FE	No	Yes	No	No
Sector FE	Yes	Yes	Yes	Yes
N.Observations	703	753	703	753

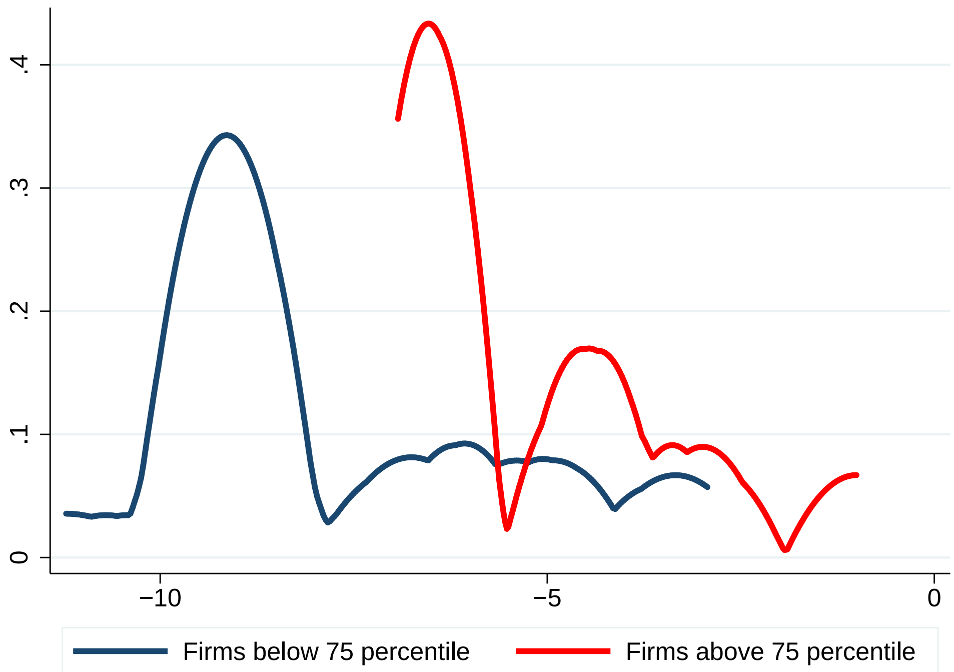
Notes: Dependent variables: foreign affiliates sales relative to parent's sales operating in each host country-sector pair. The regressors include the natural log of the distance between U.S and the host market (ln *Dist*); a dummy for the participation of the host market in a regional trade agreement (*RTA*), a dummy of common border (*border*), common language (*language*) and whether or not the host market and U.S. had a colonial relationship (*colony*). Other controls includes the level of capital endowment (*Capital*), the natural log of GDP per capita and a measure of the institutional quality of the host country (*Rule of Law*). Robust standard errors reported in parentheses. Significance is denoted: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$..

Figure 7. Distribution of Estimated MP Cost



Notes: This figures shows the combined effect of gravity variables on MP sales for firms below and above the median in each country-sector pair. For each country sector-pair and firm group in the sample we compute the following equation: $\widehat{\tau}^{mp}_{i,us} = \widehat{\beta}_d \times lndist_{i,us} + \widehat{\beta}_b \times border_{i,us} + \widehat{\beta}_{language} \times lan_{i,us} + \widehat{\beta}_{RTA} \times RTA_{i,us} + \widehat{\beta}_c \times colony_{i,us}$

Figure 8. Distribution of Estimated MP Cost



Notes: This figures shows the combined effect of gravity variables on MP sales for firms below the 25th percentile and those firms above the 75th percentile in each country-sector pair. For each country sector-pair and firm group in the sample we compute the following equation:

$\widehat{\tau}^{mp}_{i,us} = \widehat{\beta}_d \times lndist_{i,us} + \widehat{\beta}_b \times border_{i,us} + \widehat{\beta}_{language} \times lan_{i,us} + \widehat{\beta}_{RTA} \times RTA_{i,us} + \widehat{\beta}_c \times colony_{i,us}$

Appendix A: Proofs

Lemma 1: *The elasticity of marginal cost of composite intermediate input with respect to trade costs τ_{ni} is increasing in firm's productivity φ. For $\varphi^1 > \varphi^2$, $\varepsilon^{MC}(\tau_{ni}, \varpi, \varphi^1) > \varepsilon^{MC}(\tau_{ni}, \varpi, \varphi^2) > 0$.*
Proof: The proof is based on Keller and Yeaple (2013). By contradiction method, assume that $\varepsilon^{MC}(\tau_{ni}, \varpi, \varphi^1) < \varepsilon^{MC}(\tau_{ni}, \varpi, \varphi^2)$. Then,

$$\int_{\tilde{z}}^{\infty} \beta(z|\varphi^1)dz \int_0^{\tilde{z}} \beta(z|\varphi^2)t(z)^{1-\eta}dz < \int_{\tilde{z}}^{\infty} \beta(z|\varphi^2)dz \int_0^{\tilde{z}} \beta(z|\varphi^1)t(z)^{1-\eta}dz. \tag{.1}$$

Without loss of generality we set $\varpi = 1$. By definition, if $\beta(z|\varphi)$ is log-supermodular in z and α, then for $z' > z''$,

$$\beta(z'|\varphi^1)\beta(z''|\varphi^2)t(z)^{1-\eta} > \beta(z'|\varphi^2)\beta(z''|\varphi^1)t(z)^{1-\eta}. \tag{.2}$$

Integrate with respect to z'' over $[0, z')$ and with respect to z' over $[z', \infty)$, and replace z' with $\tilde{z}$ we get

$$\int_{\tilde{z}}^{\infty} \beta(z|\varphi^1)dz \int_0^{\tilde{z}} \beta(z|\varphi^2)t(z)^{1-\eta}dz > \int_{\tilde{z}}^{\infty} \beta(z|\varphi^2)dx \int_0^{\tilde{z}} \beta(z|\varphi^1)t(z)^{1-\eta}dz \tag{.3}$$

Contradiction ■

Lemma 2: *let $\theta(\tau_{ni}, \varphi, \varpi)$ be the share of imported inputs $M(\tau_{ni}, \varphi, \varpi)$ in total composite intermediate input costs $TC(\tau_{ni}, \varphi, \varpi)$. Then, $\theta(\tau_{ni}, \varphi, \varpi) = \frac{M(\tau_{ni}, \varphi, \varpi)}{T(\tau_{ni}, \varphi, \varpi)} = \varepsilon^{MC}(\tau_{ni}, \varphi, \varpi)$ is i) increasing in φ, ii) the import cost share is declining in trade costs for all firms, and iii) the rate of decline in the import cost share is slower in the more knowledge intensive firms.*
Proof: Part i) follows immediately Lemma 1. For part two, the elasticity of $\theta(\tau_{ni}, \varphi)$ with respect to τ_{ni} is given by (w.l.o. $\partial t(z)/\partial\tau = 1$)

$$\xi_{\theta,\tau} = -(\eta - 1)(1 - \theta(\tau, \varphi)) - \frac{\partial\tilde{z}(\tau)}{\partial\tau} \frac{\beta(z|\varphi)\tau}{\int_{\tilde{z}(\tau)}^{\infty} b(z|\varphi)dz} < 0. \tag{.4}$$

The third part is implied by the monotone likelihood ratio property: $\frac{\beta(z|\varphi^i)}{\int_{\tilde{z}}^{\infty} \beta(z|\varphi^1)dx} < \frac{\beta(z|\varphi^2)}{\int_{\tilde{z}}^{\infty} \beta(z|\varphi^2)dz}$, and $\theta(\tau, \varphi^1) > \theta(\tau, \varphi^2)$ ■

Proposition 1: *There exists a productivity cutoff φ_{ni}^{int} such that*

$$\mathcal{I}(\varphi) = \begin{cases} 1 & \textit{if} \quad \varphi \geq \varphi_{ni}^{int} \\ 0 & \textit{otherwise} \end{cases}$$

That is, only the most productive foreign affiliates in country n engage in intrafirm trade with their parents (import intermediate inputs form their parents).
Proof: An affiliate chooses to import from its parent if,

$$\varphi^{\sigma-1} B_n \left[\Delta C_{ni}^M(\tau_{ni}, \mathcal{I}(\varphi), \varphi, \varpi)\right] \geq w_i f_{ni}^{int}, \tag{.5}$$

The first term in the left hand side of the equation above $\varphi^{\sigma-1}$ is increasing in φ. The second term $\Delta C_{ni}^M(\tau_{ni}, \mathcal{I}(\varphi), \varphi, \varpi) \equiv C_{ni}^M(\tau_{ni}, \mathcal{I}(\varphi) = 1, \varphi)^{1-\sigma} - C_{ni}^M(\tau_{ni}, \mathcal{I}(\varphi) = 0, \varphi)^{1-\sigma}$ is also increasing in φ.

Notice that $C(\mathcal{I}=0,\varphi^1) > C(\mathcal{I}=0,\varphi^2)$,[50] whereas $C(\tau=1,\mathcal{I}=1,\varphi^1) = C(\tau=1,\mathcal{I}=1,\varphi^2)$. By Lemma 1, $\varepsilon^{MC}(.,\varphi^1) > \varepsilon^{MC}(.,\varphi^2)$, then moving from no intra-firm trade to importing any fraction of intermediate inputs from parents yields larger saving in the cost of producing the intermediate composite input for the higher knowledge-intensive firm (more productive). ■

Proposition 2: *Country i foreign affiliate sales (conditional on opening an affiliate) in country n, $r_{ni}^{aff}(\varphi)$ are decreasing in trade costs τ_{ni}. Let $\varepsilon_{ni}^{r}(\varphi,\tau_{ni}) < 0$ be the elasticity of affiliate sales with respect to trade costs, then the absolute value of $\varepsilon_{ni}^{r}(\varphi,\tau_{ni})$ is increasing in φ. In words, the sales of more knowledge intensive firms (affiliates) are more sensitive to trade costs. That is, **FDI-Gravity** is more pronounce for more knowledge intensive parents-affiliates.*

Proof: Notice that

$$\varepsilon_{ni}^{r}(\varphi,\tau_{ni},\mathcal{I}) = (1-\sigma)\varepsilon_{ni}^{MC}(\varphi,\tau_{ni},\mathcal{I}) \tag{.6}$$

The proof then follows immediately from the properties of $\varepsilon_{ni}^{MC}(\varphi,\tau_{ni},\mathcal{I})$. Moreover, when $\mathcal{I}=0$, as explained in the text $\bar{t}$ is increasing with τ_{ni}. Thus the proof is complete. ■

Appendix B: Detail Derivations

Dividends per share s: In the text we claim that $s = \frac{\sigma-1}{\sigma(\kappa-1)+1}$. Let Π_n be the aggregate profits of all firms in country n, including foreign affiliates profits,

$$\Pi_n = \sum_{i=1}^{N} J_n \left\{ \int_{\varphi_{in}}^{\varphi_{in}^{fdi}} \pi_{in}(\varphi)dG(\varphi) + \int_{\varphi_{in}^{fdi}}^{\varphi_{in}^{int}} \pi_{in}^{fdi}(\varphi)dG(\varphi) + \int_{\varphi_{in}^{int}}^{\infty} \pi_{in}^{int}(\varphi)dG(\varphi) \right\}, \tag{.7}$$

and $\varphi_{nn}^{fdi} = \varphi_{nn}^{int} = \infty$. The domestic/export profits, non-importer foreign affiliates profits and importer affiliates profits are denoted by $\pi_{in}(\varphi), \pi_{in}^{fdi}(\varphi)$, and $\pi_{in}^{int}(\varphi)$, respectively. Using the functional forms of the profits, the Pareto distribution, the cutoffs' equations and integrating, we get

$$\Pi_n = \frac{\sigma-1}{\sigma\kappa} \sum_{i=1}^{N} R_{in} + R_{in}^{fdi} + R_{in}^{int} \tag{.8}$$

R_{in}, R_{in}^{fdi} and R_{in}^{int} denote the values of the aggregate sales of exporting to country i, the aggregate foreign affiliates sales-who do not import-, and the importer aggregate affiliate sales, respectively. Indeed, $R_{nn}^{fdi} = R_{nn}^{int} = 0$. Let Π denote the world aggregate profits: $\Pi = \sum_{n \in N} \Pi_n$, then

$$\Pi = \frac{\sigma-1}{\sigma\kappa} \sum_{n\in N}\sum_{i\in N} R_{in} + R_{in}^{fdi} + R_{in}^{int} \tag{.9}$$

$$= \frac{\sigma-1}{\sigma\kappa} Y \tag{.10}$$

Here, Y is the world total sales/expenditures. World's total profits Π is also given by the dividends per share times the total number of shares. Thus, $\Pi = \sum_{n\in N} sL_n = \frac{\sigma-1}{\sigma\kappa}Y = \frac{\sigma-1}{\sigma\kappa}\sum_{n\in N} L_n(1+s)$,

[50] Notice that $C(\mathcal{I}=0,\varphi) = \bar{t} = \int_0^\infty \beta(z|\varphi)t(z)^{1-\eta}dz$, which is indeed increasing in φ under the assumptions about $\beta(z|\varphi)$ and $t(z)$.

where the last equality follows from balanced trade and the fact that $X_n = L_n + \Pi_n = L_n + sL_n$. Then,

$$s = \frac{\Pi}{\sum_{n \in N} L_n} = \frac{\sigma - 1}{\sigma\kappa}(1+s)$$
$$\rightarrow s = \frac{\sigma - 1}{\sigma(\kappa - 1) + 1}$$

Derivation of Gravity Equations

Aggregate exports from country i to country n is given by[51]

$$X_{ni} = J_i \int_{\varphi_{ni}}^{\varphi_{ni}^{fdi}} \sigma\varphi^{\sigma-1} \left(\mu X_n / P_n^{1-\sigma}\right) \tau_{ni}^{1-\sigma} dG(\varphi) \tag{.11}$$

Evaluating the integration, using the formula for the aggregate price level, and substituting out the cutoffs and $J_i = X_i/(1+s)$, we obtain the gravity equation derived in the text. Similarly, non-importer affiliates' aggregate sales and importer affiliates' aggregate sales can by expressed by

$$X_{ni}^{fdi} = J_i \int_{\varphi_{ni}^{fdi}}^{\varphi_{ni}^{int}} \sigma\varphi^{\sigma-1} \left(X_n / P_n^{1-\sigma}\right) \left[\tau_{ni}^{\alpha} exp(\phi)\right]^{1-\sigma} dG(\varphi) \tag{.12}$$

$$X_{ni}^{int} = J_i \int_{\varphi_{ni}^{int}}^{\infty} \sigma\varphi^{\sigma-1} \left(X_n / P_n^{1-\sigma}\right) \left(exp(\phi(1 - \tau_{ni}^{\frac{\alpha-1}{\phi}}) + \alpha \ln)\right)^{1-\sigma} dG(\varphi) \tag{.13}$$

Using the same steps as before, we get the gravity equations for non-importer affiliates' sales and importer affiliates' sales.

FDI- Gravity: Affiliates who do not import from parents:

In the text we claimed that the sales of non-importer decrease in trade frictions; equation (23). In order to prove this formally we use our analysis of the intensive/extensive margin. Remember that we can disentangle the impact of trade costs on affiliates' sales into the intensive and the extensive margins;

$$\xi_{X^{fdi},\tau} = \overbrace{\alpha(1-\sigma)}^{\text{Intensive margin}} + \overbrace{\frac{\kappa - (\sigma - 1)}{(\varphi_{ni}^{fdi})^{\sigma-1-\kappa} - (\varphi_{ni}^{int})^{\sigma-1-\kappa}} \left[\xi_{\varphi^{int},\tau}(\varphi_{ni}^{int})^{\sigma-1-\kappa} - \xi_{\varphi^{fdi},\tau}(\varphi_{ni}^{fdi})^{\sigma-1-\kappa}\right]}^{\text{Extensive margin}} \tag{.14}$$

The extensive margin is negative if and only if, $\xi_{\varphi^{fdi},\tau}(\varphi^{fdi})^{\sigma-1-\kappa} > \xi_{\varphi^{int},\tau}(\varphi^{int})^{\sigma-1-\kappa}$. This will be the case if, $\frac{\alpha C^M(.,\mathcal{I}=0)^{1-\sigma} - \tau^{1-\sigma}}{\varepsilon^{MC} C^M(.,\mathcal{I}=1)^{1-\sigma} - \alpha C^M(.,\mathcal{I}=0)^{1-\sigma}} > \left(\frac{f_{ni}^{fdi} - f_{ni}}{f_{ni}^{int}}\right)^{(1-\sigma)(\sigma-1-\kappa)} \left(\frac{C_{2ni}}{C_{1ni}}\right)^{(1-\sigma)(\sigma-1-\kappa)}$. For FDI cutoff be well defined, we require $f_{ni}^{int} > (f_{ni}^{fdi} - f_{ni})\frac{C_{2ni}}{C_{1ni}}$. If f_{ni}^{int} is way larger than the last term then the last term of the previous inequality becomes very small and approaches zero as $f_{ni}^{int} \rightarrow \infty$. Therefore, there exists $f_{ni}^{int} < \infty$ such that the extensive margin is negative. If this condition does

[51] Notice that we do not include the intrafirm export in the total exports. It is easy to show that total intrafirm exports is constant share of the importer total affiliates sales.

not hold, all what we need to have FDI-gravity is $-\frac{\xi_{\varphi^{int},\tau}(\varphi_{ni}^{int})^{\sigma-1-\kappa}-\xi_{\varphi^{fdi},\tau}(\varphi_{ni}^{fdi})^{\sigma-1-\kappa}}{\varphi_{ni}^{fdi})^{\sigma-1-\kappa}-(\varphi_{ni}^{int})^{\sigma-1-\kappa}} < \frac{\alpha(\sigma-1)}{\kappa-(\sigma-1)}$, which is easily satisfied for reasonable parameter values. If either of these two conditions is satisfied, FDI sales must be negatively correlated with trade frictions.

Derivation of the marginal cost of intermediate input composite: equation (13)

$$C_{ni}^{M}(\tau,\phi,\mathcal{I}) = \begin{cases} \tau^{\alpha}\exp\left\{\int_0^{\infty}\frac{1}{\phi}exp(-z/\phi)zdz\right\} & \text{if} \quad \mathcal{I}=0 \\ \exp\left\{\int_0^{\tilde{z}}\frac{1}{\phi}exp(-z/\phi)(\alpha\ln\tau+z)dz + \ln\tau\int_{\tilde{z}}^{\infty}\frac{1}{\phi}exp(-z/\phi)dz\right\} & \text{if} \quad \mathcal{I}=1 \end{cases} \tag{.15}$$

Integrating by parts and substituting out $\tilde{z}=(1-\alpha)\ln\tau_{ni}$, the required results are obtained.