

Multinational Firms and the Structure of International Trade*

Pol Antràs^{*,†,‡} and Stephen R. Yeaple^{†,§}

^{*}Harvard University, Cambridge, MA, USA

[†]National Bureau of Economic Research, Cambridge, MA, USA

[‡]Centre for Economic Policy Research, London, UK

[§]Pennsylvania State University, University Park, PA, USA

Abstract

This chapter reviews the state of the international trade literature on multinational firms. This literature addresses three main questions. First, why do some firms operate in more than one country while others do not? Second, what determines in which countries production facilities are located? Finally, why do firms own foreign facilities rather than simply contract with local producers or distributors? We organize our exposition of the trade literature on multinational firms around the workhorse monopolistic competition model with constant-elasticity-of-substitution (CES) preferences. On the theoretical side, we review alternative ways to introduce multinational activity into this unifying framework, illustrating some key mechanisms emphasized in the literature. On the empirical side, we discuss the key studies and provide updated empirical results and further robustness tests using new sources of data.

Keywords

Multinationals, Foreign direct investment, Vertical FDI, Horizontal FDI, Complex FDI, Internalization, Offshoring, Outsourcing, Related party trade, Cross-border acquisitions, Greenfield FDI

JEL classification codes

F23, F12, F2, F1, F61, D21, D23, D22, D2, L2

1. INTRODUCTION

Over the last two decades, international trade theory has undergone a steady transformation that has placed firms rather than countries or industries as the central unit of

* The statistical analysis of firm-level data on U.S. multinational corporations reported in this study was conducted at the U.S. Bureau of Economic Analysis under arrangements that maintained legal confidentiality requirements. Views expressed are those of the authors and do not necessarily reflect those of the Bureau of Economic Analysis. We are grateful to our discussants, John McLaren and Esteban Rossi-Hansberg for their insightful and incisive comments at the handbook conference held in Cambridge in September of 2012. We also benefitted from valuable feedback from the editor, Elhanan Helpman. We also thank Ruiqing Cao and Yang Du for outstanding research assistance, and Davin Chor, Federico Díez, Willi Kohler, Nathan Nunn, Natalia Ramondo, Marcel Smolka, and Dan Trefler for their help with data sources. The Online Appendix mentioned in the text is available in the NBER Working Paper version of this chapter.

analysis. This transformation has been fueled by micro-level empirical studies that have shown international activity to be concentrated within a handful of very large firms that produce in multiple countries and multiple industries. In 2000, for instance, the top 1% U.S. exporters accounted for 81% of U.S. exports (Bernard et al., 2009). The involvement of these large firms in the world economy goes well beyond the mere act of selling domestically produced goods to foreign consumers. According to 2009 data from the U.S. Bureau of Economic Analysis, the sales of domestically produced goods to foreign customers account for only 25% of the sales of large American firms. The remaining 75% (nearly U.S. \$5 trillion) is accounted for by the sales of foreign affiliates of American multinationals (Yeaple, 2013). Furthermore, data from the U.S. Census Bureau indicates that roughly 90% of U.S. exports and imports flow through multinational firms, with close to one-half of U.S. imports transacted *within* the boundaries of multinational firms rather than across unaffiliated parties (Bernard et al., 2009).

This chapter reviews the state of the international trade literature on multinational firms. Before we begin, a few definitions are in order. In his encyclopedic monograph on the subject, Caves (2007, p. 1) defines a multinational firm as “an enterprise that controls and manages production establishments—plants—located in at least two countries.” While the corporate structure of multinational firms can be complicated, it is useful to define two types of entities within a multinational firm, the parent and the affiliate. *Parents* are entities located in one country (the source country) that control productive facilities, while *affiliates* are located in other countries (host countries). The notion of control is a judgmental one but it is often associated with ownership. Such ownership is the result of foreign direct investments, which can alternatively involve the acquisition of a controlling interest in an existing foreign firm (cross-border acquisitions) or the establishment of an entirely new facility in a foreign country (greenfield investment).

The positive theory of the multinational firm revolves around three main questions. First, why do some firms find it optimal to operate in more than one country while others do not? Second, what determines in which countries production facilities are located and in which they are not? Finally, why do firms own foreign facilities rather than simply contract with local producers or distributors?

The modern literature’s focus on the firm contrasts sharply with the traditional theory that made little distinction between foreign direct investment and international portfolio investment flows. According to the traditional theory, multinational firms were simply arbitrageurs that moved capital from countries where returns were low to countries where returns were high.¹ The genesis of the modern approach was Hymer’s (1960) seminal Ph.D. thesis. Hymer pointed out that the traditional international-finance approach was inconsistent with several features of foreign direct investment (FDI) data. He proposed a new, industrial-organization approach based on the notion that some firms own special assets that confer a strategic advantage over indigenous firms in foreign markets. In some

¹ This interpretation is implicit, for instance, in Mundell (1957).

cases, market imperfections preclude the use of these assets by foreign unaffiliated entities, thereby generating the need for a direct involvement of the asset owner. In sum, Hymer envisioned a world in which real (not financial) factors shape the location of multinational activity and financial flows are a mere consequence of the financial structure decisions of multinational firms.

Hymer's approach was later refined by several authors, including [Kindleberger \(1969\)](#), [Caves \(1971\)](#), [Buckley and Casson \(1976\)](#), and [Rugman \(1981\)](#), and culminated with [Dunning's \(1981\)](#) eclectic OLI framework, where OLI is an acronym for Ownership, Location, and Internalization. Put succinctly, the emergence of the multinational firm is explained by an **O**wnership advantage stemming from firm-specific assets that allow firms to compete in unfamiliar environments, a **L**ocation advantage that makes it efficient to exploit the firm assets in production facilities in multiple countries, and an **I**nternalization advantage that makes the within-firm exploitation of assets dominate exploitation at arm's length. The mainstream interpretation of the ownership advantage relates it to a proprietary technology or reputation that provides its owner with some market power or cost advantage over indigenous producers. The location advantage is often associated with the idea that the development of these assets (tangible or intangible) entails significant fixed costs, but these assets can then be used in different locations simultaneously in a nonrival manner. This allows economies of scale to be exploited efficiently within multinational firms, especially when trade frictions inhibit such exploitation via exporting. Another branch of the literature has related location advantages to situations in which production is amenable to geographical fragmentation, thus allowing different parts of the production process to be undertaken in the location where it is most cost-effective to do so. Finally, the internalization advantage is attributed to market failures in the transfer of technology—related to the partial nonexcludable, nonrival, and noncodifiable nature of technology—and to inefficiencies associated with market exchanges of highly customized intermediate inputs.

As insightful as the OLI literature is, it took some time before it was absorbed by international trade theory because a widely accepted general-equilibrium modeling of increasing returns to scale, product differentiation, and imperfect competition did not become available until the late 1970s and early 1980s, and because contract theory was still in its infancy in the 1970s. The modeling of product differentiation and market structure originally developed by [Dixit and Stiglitz \(1977\)](#) and later adopted by [Krugman \(1979, 1980\)](#) served the important role of providing a common language for researchers in the field to communicate among themselves, and opened the door to formally modeling multinational firms within general-equilibrium analysis.²

² Although the heavy use of specific functional forms for preferences and technology was viewed with some reservation by the old guard in the field, the publication of the landmark manuscript by [Helpman and Krugman \(1985\)](#) established the generality of most of the insights from Krugman's work and showed how the new features of New Trade Theory could be embedded into Neoclassical Trade Theory.

Following this tradition, we will organize our exposition of the trade literature on multinational firms around the classical [Krugman \(1980\)](#) model with constant-elasticity-of-substitution (CES) preferences, and the seminal variant of the model incorporating firm heterogeneity developed by [Melitz \(2003\)](#). On the theoretical side, we will review alternative ways to introduce multinational activity into the framework, while trying to illustrate some of the key mechanisms emphasized in the literature, even when these were developed under different modeling assumptions. Although some important papers in the trade literature on multinational firms adopt alternative modeling approaches to imperfect competition (most notably, [Markusen, 1984](#); [Horstmann and Markusen, 1987a](#); [Markusen and Venables, 1998](#)) or make less restrictive assumptions on technology and preferences (e.g., [Helpman, 1984](#)), we think that the advantage of a common framework outweighs the benefits of comprehensiveness and generality.

Given the space constraint, we will impose several limits on the scope of our review. First, we will not review in great detail certain branches of the literature that, although being associated with aspects of multinational activity, do not model multinational firms explicitly. Second, we will focus almost exclusively on positive issues related to the rationale for multinational activity and will thus not discuss the effects of multinational firms on goods and factor markets or the policy implications of these effects. The only exception to this rule is our brief discussion of the effects of vertical fragmentation on labor markets in [Section 5.2](#). Third, our emphasis will be on qualitative analysis, though we will briefly review recent advances in quantitative analysis in [Section 6](#). Fourth, although we will sometimes refer to the internalization decision of multinational firms as an organizational decision, we will not review the broader literature on the international organization of production, which is concerned not only with multinational firm boundaries, but also with incentive provision, delegation, and hierarchical structure within and across firms in the global economy.³ Fifth, we will restrict ourselves to discussing the operational decisions of multinational firms (such as those related to employment, production levels, location, and ownership), thus omitting a treatment of the financial aspects of these firms, which are important for understanding the relationship between multinational activity and FDI flows.⁴ For more encyclopedic treatments of multinational firms, we refer the reader to the monographs by [Caves \(2007\)](#), [Markusen \(2002\)](#), and [Barba-Navaretti and Venables \(2004\)](#), while an overview of the literature on the international organization of production can be found in [Antràs and Rossi-Hansberg \(2009\)](#).⁵

By limiting our focus, we will be able to take our survey beyond a simple enumeration of the various theoretical and empirical results in the literature. On the theoretical side, we

³ For recent work in this area, see [Grossman and Helpman \(2004\)](#) on incentive provision, [Marin and Verdier \(2009\)](#) and [Puga and Trefler \(2010\)](#) on delegation, and [Antràs et al. \(2006\)](#) and [Caliendo and Rossi-Hansberg \(2012\)](#) on hierarchical structure.

⁴ See [Klein et al. \(2002\)](#), [Desai et al. \(2004\)](#), and [Antràs et al. \(2009\)](#) for work on financial aspects of multinational activity.

⁵ See [Helpman \(2006\)](#) and [Spencer \(2005\)](#) for recent alternative reviews of this literature.

will explicitly derive a series of analytical results within a unified framework. While most of these results are not new, some had only been illustrated numerically and in somewhat disjointed frameworks. On the empirical side, our review will provide updated empirical results and further robustness tests using new sources of data.

The remainder of this chapter is organized as follows. [Section 2](#) briefly describes the data available to analyze the global operations of multinational firms and provides a list of “stylized” facts about the multinational firm. [Section 3](#) introduces the benchmark model that we will use to guide our overview of the theoretical literature. [Sections 4](#) and [5](#) focus on the integration of the multinational firms into our benchmark models and cover the relevant empirical literature that both informs the design of these models and tests these models’ predictions. [Section 6](#) outlines the approaches taken by a recent literature that explores the global structure of trade in multinational production in multicountry models. The internalization advantage of multinationals is taken as given in [Sections 4](#) through [6](#), but is explicitly modeled and empirically assessed in [Section 7](#). [Section 8](#) offers concluding remarks.

2. STYLIZED FACTS

In this section, we develop six stylized facts that describe broad features of the structure of multinationals’ global operations. We do so using three types of data. First, we use foreign direct investment data from the balance of payments. Foreign direct investment (FDI) occurs when a firm from one country obtains an operating stake (usually 10%) in an enterprise in another country or when a financial flow occurs between parties that are resident in different locations but related by ownership. Second, we use government survey data that distinguish between national firms and the parents and affiliates of multinational firms. We rely particularly on census data on U.S. parents and their foreign affiliates collected by the Bureau of Economic Analysis (BEA) of the United States. Finally, we use U.S. Related Party Trade data collected by the U.S. Bureau of Customs and Border Protection. This source provides data on related and nonrelated party U.S. imports and exports at the six-digit Harmonized System (HS) classification and at the origin/destination country level.⁶

We first document regularities in the countries that are the source of FDI and countries that are the destination of this FDI. FDI flows measure changes in the holdings of controlling interests in equity capital between countries. By aggregating flows over time by country one can obtain crude measures of how important countries are as hosts to parent firm operations (outward stocks) and as hosts to affiliate operations (inward stocks).

⁶ Related party trade means a minimum ownership stake between trading parties of 6% for imports and 10% for exports.

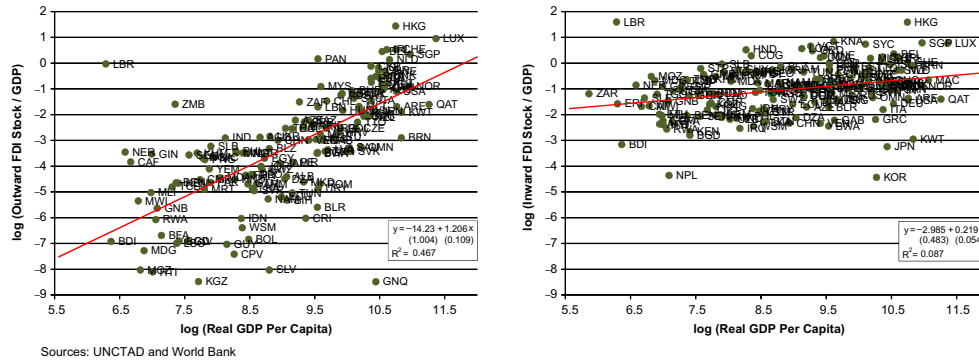


Figure 2.1 Aggregate FDI Stocks and Development

Figure 2.1 presents outward and inward FDI stocks for a large number of countries. In the left-hand panel, the logarithm of FDI stocks held by the sending country and normalized by the sending country's GDP is plotted against the logarithm of the sending country's GDP per capita. In the right-hand panel the stocks of inward FDI by destination country, normalized by destination country GDP, are plotted against the logarithm of GDP per capita by destination country. In both panels the best linear predictor is displayed as a line with the associated coefficients shown in the bottom right-hand corner.

These data show that developed countries are more engaged in both outward and inward flows than less developed countries, but the positive relationship is much more pronounced for outward flows. The outliers in both figures illustrate some of the deficiencies of FDI data as a measure of real production activity. For instance, Hong Kong, Singapore, Luxembourg, and Liberia have high levels of both inward and outward FDI stocks which reflect in part firms' efforts to park ownership of global assets in low-tax and weak-regulation countries.

Figure 2.1 suggests that there must be substantial two-way flows between countries. The extent of these two-way flows can be measured by Grubel-Lloyd indices

$$GL_{ij} = 100 \times \left(1 - \frac{|S_{ij} - S_{ji}|}{S_{ij} + S_{ji}} \right),$$

where S_{ij} is the stock of foreign direct investment owned by country i firms in country j . This index takes on a value of 100 when $S_{ij} = S_{ji}$ and a value of 0 when the stock is one-way. High values of GL_{ij} are associated with high levels of two-way flows. When computed for OECD data, GL among developed country partners ranges from 45 to 50, indicating a high level of two-way FDI, while the average GL between developed and developing countries tends to be only one-half as large.

The patterns of multinational production revealed by the FDI data can be corroborated by less comprehensive data on real activity from the United States. Most of the sales of

U.S. affiliates are made in developed countries while most of the affiliates active in the United States are owned by parents located in developed countries. Furthermore, U.S. related party trade data reveal that most of the trade between affiliates and their parents into and out of the United States occurs between developed rather than developing countries. We summarize these empirical regularities as the following fact:

Fact One: Multinational activity is primarily concentrated in developed countries where it is mostly two-way. Developing countries are more likely to be the destination of multinational activity than the source.

We now turn from patterns in aggregate multinational activity to patterns across industries. There is substantial variation across industries in the share of real activity contributed by multinational firms. For instance, data from the OECD reveal that foreign affiliates account for approximately a quarter of French manufacturing employment; however, the employment share is well over a third in capital intensive and R&D intensive industries like Chemicals and Machinery, while it is less than one-eighth in labor-intensive Food and Textile industries. This pattern repeats itself across developed countries.

The tendency of multinational activity to be concentrated in certain industries is also evident in the share of international trade that occurs between parties related by ownership. The left panel of [Figure 2.2](#) is constructed using the U.S. related-party data and measures of physical capital intensity from the NBER Manufacturing database. It shows that the share of U.S. imports that occurs within the boundaries of multinational firms is highly correlated with the logarithm of the capital-labor ratio in that industry. Imports of labor-intensive products, such as apparel (North American Industry Classification System (NAICS) 3159) or footwear (NAICS 3162), are transacted mostly at arm's-length, while imports of heavily capital-intensive products, such as motor vehicles (NAICS 3361) and pharmaceuticals (NAICS 3254), are traded within firm boundaries. The right panel of [Figure 2.2](#) depicts a similar strong positive correlation between the share of intrafirm trade and R&D intensity, despite the existence of some notable outliers, such as motor vehicle manufacturing (NAICS 3361).⁷

Another indication of the concentration of FDI in certain industries is the intra-industry nature of the two-way FDI flows observed with aggregate data. When computing Grubel-Lloyd indices of intraindustry FDI into and out of the United States for four-digit U.S. NAICS industries in 2010, one finds indices that generally exceed one-half and are of a similar order of magnitude to those computed with trade data. We summarize the evidence concerning the cross industry structure of the magnitude of FDI as the following fact:

⁷ R&D intensity is the ratio of R&D expenditures to sales, constructed by [Nunn and Treffer \(2008\)](#) using Orbis data. We exclude from the sample in the right panel those industries for which the ratio of R&D expenditures to sales exceeds one, but the positive and statistically significant correlation with intrafirm trade is robust to their inclusion.

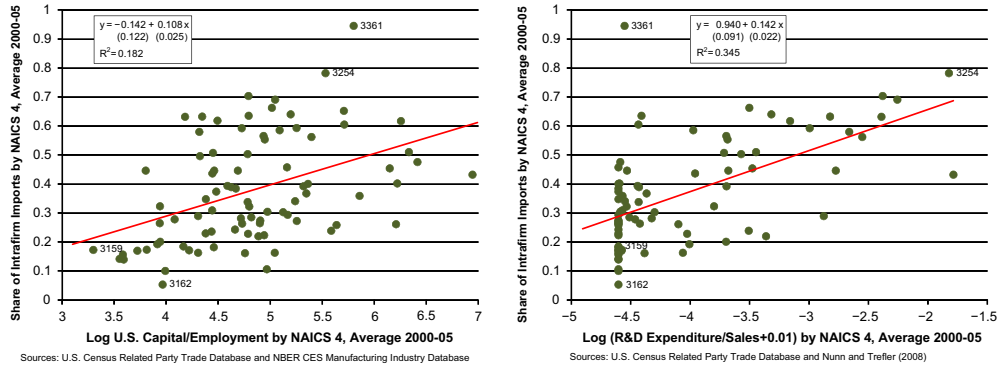


Figure 2.2 The Share of Intrafirm Imports, Capital Intensity, and R&D Intensity

Fact Two: The relative importance of multinationals in economic activity is higher in capital-intensive and R&D intensive goods, and a significant share of two-way FDI flows is intraindustry in nature.

A key issue in international economics is the role of geography in affecting the structure of international commerce. It is well known that a gravity equation fits international trade volumes well. The same is true for many aspects of multinational activity. In serving foreign markets, firms may choose to export their products from their source country s to a destination country d . Let X_{ds} be the aggregate value of exports from s to d . An alternative to exporting their product is to establish a foreign affiliate in country d to serve the local market. Let AS_{ds} be the sales of the affiliates located in country d that are owned by parents in country s . The left-hand panel of Figure 2.3 plots the logarithm of AS_{ds} normalized by source and destination GDP against the logarithm of the distance between s and d , while the right-hand panel plots the logarithm of AS_{ds}/X_{ds} between the two countries against the logarithm of distance.⁸

The left-hand panel shows that a 1% increase in distance is associated with 0.57 fall in affiliate sales. The right-hand panel shows that as distance increases affiliate sales are falling less rapidly than trade volumes, so while gravity holds for affiliate sales, its effect on trade volumes is stronger.

The stronger effect of distance on trade relative to foreign affiliate activity also appears within firms. Parent firms often produce components that are then shipped to foreign affiliates for further processing. According to BEA data, in 2009 the aggregate value added of U.S. manufacturing affiliates was U.S. \$474.5 billion while the imports of intermediate inputs by these affiliates from their parents totaled U.S. \$104 billion.⁹ Figure 2.4 shows

⁸ Affiliate sales data are from Ramondo (2013). We thank her for sharing her data with us.

⁹ According to data highlighted by Ramondo et al. (2012) at the affiliate level much of these imports are highly concentrated in a small number of highly integrated firms.

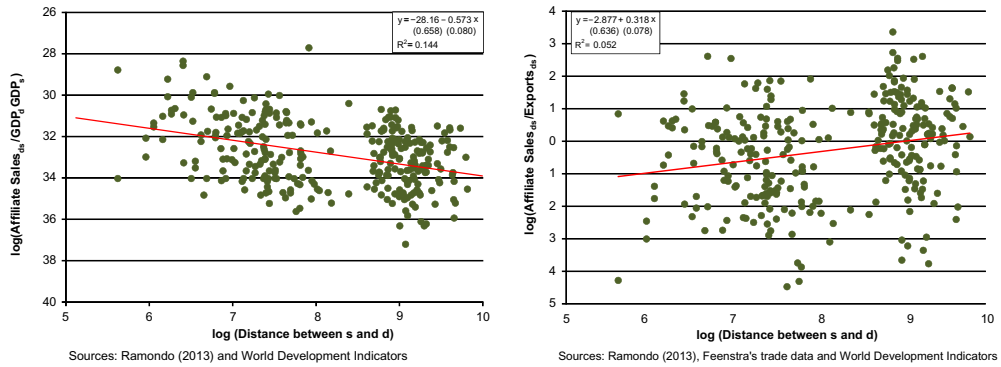


Figure 2.3 Gravity, FDI Sales, and Trade Flows

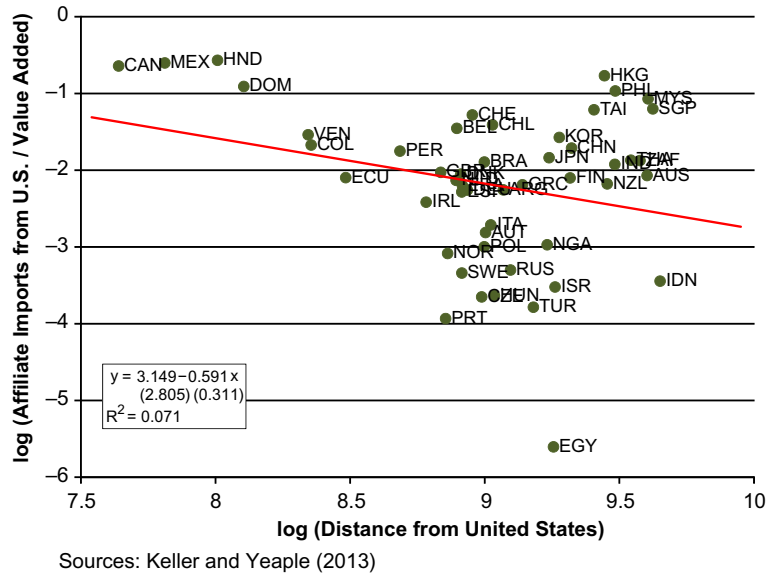


Figure 2.4 Affiliate Imports from U.S. relative to Local Value-Added

that the logarithm of the ratio of the value of imported intermediates to the sum of local value-added plus imported intermediates is declining in the distance between the affiliate and the U.S. parent. This suggests that vertical specialization is harder at long distances, but it is interesting to note that the very open Asian economies of Hong Kong, Malaysia, the Philippines, Singapore, and Taiwan import large amounts of intermediate inputs despite their distance from the United States.

The effect of distance on the structure of international trade and foreign affiliate activity is summarized as the following fact:

Fact Three: The production of the foreign affiliates of multinationals falls off in distance, but at a slower rate than either aggregate exports or parent exports of inputs to their affiliates.

We now compare the performance of multinationals to that of firms that do not own foreign operations. As noted in the introduction there are two distinct entities within multinational firms, the parents and the affiliates. The BEA and the U.S. Census collect a large amount of information about the structure of the operations of U.S.-based firms. From the data it is possible to compare the activities of U.S.-based parents with other firms that are active in the United States. In manufacturing sectors, U.S. parents account for less than one-half of 1% of enterprises but account for over 62% of value-added and 58% of employment ([Barefoot and Mataloni, 2011](#)). These numbers imply that labor productivity is higher in multinational parents relative to nonmultinationals. In addition, these parents account for almost three-quarters of private R&D conducted in U.S. manufacturing. This latter observation is in part accounted for by the fact that many parent firms are concentrated in R&D intensive industries. [Mayer and Ottaviano \(2007\)](#) provide similar evidence on the superior performance of parent firms in Germany, France, Belgium, and Norway.

We now make a similar comparison of the activities of foreign affiliates with other firms in their host country. [Table 2.1](#) shows the share of economic activity (rows) of foreign affiliates in their host country (columns) in manufacturing industries for a number of OECD countries. The first row shows that the share of foreign affiliates in the total number of manufacturing enterprises is typically very small. The share of affiliates in manufacturing employment, output, and R&D expenditures, shown in the next three rows, is larger by an order of magnitude. The last line shows that the share of foreign affiliates in total manufacturing exports is even larger than their share in total sales and employment.

Table 2.1 Affiliates Relative to Local Firms (Percentage Accounted for by Affiliates)

	Finland	France	Ireland	Holland	Poland	Sweden
Enterprises	1.6	2.0	13.4	3.4	16.0	2.8
Employment	17.2	26.2	48.0	25.1	28.1	32.4
Sales	16.2	31.8	81.1	41.1	45.2	39.9
R&D Expenditure	13.1	27.4	77.3	35.8	20.9	52.0
Exports	17.5	39.5	92.3	60.0	69.1	45.8

Source: OECD (2007).

Table 2.2 Destination of Affiliate Sales (%) by Industry

	Host Country	Other Foreign	United States
Total manufacturing	55	34	11
Textile and Apparel	45	35	19
Metals and Minerals	60	32	8
Chemicals and Plastics	58	36	6
Machinery	49	36	15
Computers and Electronics	40	43	16
Electronic Equipment	47	40	13
Transport Equipment	47	35	19
Other	66	26	8

Source: 2009 Benchmark Survey of U.S. Direct Investment Abroad, BEA.

These observations lead to the following fact:

Fact Four: Both the parents and the affiliates of multinational firms tend to be larger, more productive, more R&D intensive, and more export oriented than nonmultinational firms.

The large volumes of international trade between parents and affiliates are consistent with vertical specialization across parents and their affiliates. To explore the nature of this specialization we compare the share of U.S. parent firms in the total activity of U.S. multinationals' global operations. In 2009, BEA data reveal that U.S. parents accounted for 65% of total sales, 68% of value-added and employment, and 84% of research and development expenditures. One decade earlier, U.S. parents accounted for 74% of total sales, 78% of value-added, and 87% of R&D expenditures. The high and persistent concentration of R&D services in the parent firm suggests that parents are in the business of creating "ownership advantages."

If parents disproportionately provide R&D services to the multinational firm, what is the role played by the affiliates? This question can be answered in part by observing the destination of affiliate sales. The BEA collects data on how much the foreign affiliates of U.S. parents sell in their host country, how much they export to other foreign countries, and how much they export to the United States. The percentage breakdown of sales by these three categories aggregated across countries by major industry group for 2009 is shown in Table 2.2. The first row shows that 55% of the value of the output of U.S. affiliates stayed in the country of the affiliate, 34% was sold to foreign customers outside the U.S., and 11% was exported back to the United States. Hence, the primary purpose of multinational affiliates is to serve foreign markets rather than to provide inputs or final goods to the source country. However, in industries such as Textiles and Apparel, Machinery, Electronic and Transport Equipment, and especially computers, affiliates are more export oriented.

We summarize this information in the next fact:

Fact Five: Within multinational enterprises, parents are relatively specialized in R&D while affiliates are primarily engaged in selling goods in foreign markets, particularly in their host market.

One last empirical regularity involves the manner in which multinationals obtain foreign production facilities. A firm can obtain production facilities abroad by either opening a new plant (greenfield investment) or by acquiring an existing plant (cross-border merger and acquisition). UNCTAD data show that across all countries in 2007 the value of recorded mergers and acquisitions stood at over 50% of total FDI flows. Furthermore, for developed countries cross-border mergers and acquisitions accounted for 68% of FDI flows while for developing countries the number was only 18%. We summarize this information as the following fact:

Fact Six: Cross-Border Mergers and Acquisitions make up a large fraction of FDI and are a particularly important mode of entry into developed countries.

3. BENCHMARK MODEL: AN EXTENDED KRUGMAN (1980) MODEL

In this section, we describe the framework that we will build on to navigate through the literature. The framework is a strict generalization of the two-country [Krugman \(1980\)](#) model. We depart from [Krugman \(1980\)](#) in allowing for the existence of multiple sectors (including a homogeneous-good sector) and multiple factors of production (for concreteness, capital, and labor) available in fixed supply. Factors of production are internationally immobile but freely mobile across sectors. Our framework is closely related to the one developed by Melitz and Redding in [Chapter 1](#) of this handbook.

The world consists of two countries, H and F , that produce goods in $J + 1$ sectors. One sector produces a homogeneous good z , while the remaining J sectors produce a continuum of differentiated products. Preferences are identical everywhere in the world and given by:

$$U = \beta_z \log z + \sum_{j=1}^J \beta_j \log Q_j, \quad \text{with } \beta_z + \sum_{j=1}^J \beta_j = 1, \quad (1)$$

and with

$$Q_j = \left(\int_{\omega \in \Omega_j} q_j(\omega)^{(\sigma-1)/\sigma} d\omega \right)^{\sigma/(\sigma-1)}, \quad \sigma > 1.$$

Where desirable, it is straightforward to simplify the product space by setting $\beta_z = 0$ (i.e., dropping the outside good), or by focusing on the case in which there is only one differentiated-good sector (i.e., $J = 1$), as in [Krugman \(1980\)](#). Our interest is on the

behavior of firms in a given differentiated goods industry j . Maximizing (1) subject to the budget constraint yields the following demand for variety ω in industry j and country i :

$$q_j^i(\omega) = \frac{\beta_j E^i}{\int_{\omega \in \Omega_j^i} p_j^i(\omega)^{1-\sigma} d\omega} p_j^i(\omega)^{-\sigma} = A_j^i p_j^i(\omega)^{-\sigma}, \quad i = H, F. \quad (2)$$

Good z can be produced in any country under a constant-returns-to-scale technology which combines capital and labor in country i and which we represent by the linearly homogeneous unit cost function $c_z^i = c_z^i(w^i, r^i)$, with $i = H, F$. The unit cost of production can vary across countries due to factor price differences or due to technological differences. This homogeneous-good sector is perfectly competitive and we shall assume throughout that z can be costlessly traded across countries and serves as the numeraire, so that

$$c_z^i(w^i, r^i) \geq 1, \quad z^i > 0, \quad \text{complementary slack, for } i = H, F.$$

Whenever both countries produce good z we thus have $c_z^H(w^H, r^H) = c_z^F(w^F, r^F) = 1$.

Technology in the differentiated-good sectors features increasing returns to scale. As in [Krugman \(1980\)](#) and [Helpman and Krugman \(1985\)](#), the creation and production of differentiated varieties involve a fixed cost and a constant marginal cost in terms of a linearly homogeneous function of capital and labor, which we denote by $c_j^i(w^i, r^i)$. In particular, total costs associated with producing q_j^i units of sector j output in country i are given by

$$C_j^i(w^i, r^i, q_j^i) = \left(f_j + \frac{q_j^i}{\varphi} \right) c_j^i(w^i, r^i). \quad (3)$$

The parameter f_j in (3) captures the extent to which technology features economies of scale, and is assumed common across varieties. As for the marginal cost parameter φ , we will consider the case in which it is also common across firms within an industry, as in [Krugman \(1980\)](#), as well as the case in which it varies across firms within an industry, as in [Melitz \(2003\)](#). In the latter case, we will denote by $G(\varphi)$ the cumulative probability distribution from which firms draw their particular marginal cost parameter. The market structure in all differentiated-good sectors is monopolistic competition. International trade in differentiated goods is costly and entails an iceberg (*ad valorem*) trade cost such that τ_j units of sector j 's varieties need to be shipped from one country for 1 unit to reach the other country. Furthermore, in some cases, we also consider the existence of a fixed cost of exporting equal to f_X units of the composite factor represented by the above function $c_j^i(w^i, r^i)$.

It is straightforward to show (see [Chapter 1](#) by Melitz and Redding in this handbook) that this general framework encompasses, under alternative assumptions, the ones in [Krugman \(1980\)](#), [Melitz \(2003\)](#), [Chaney \(2008\)](#), and [Bernard et al. \(2007\)](#).

4. THE PROXIMITY-CONCENTRATION HYPOTHESIS

In the benchmark model above, firms were allowed to serve foreign markets only via exports. The evidence suggests instead that firms frequently choose to service foreign markets through local production by a subsidiary rather than through exporting, thus becoming multinational firms. In this section we focus on models that feature a proximity-concentration tradeoff to explain why multinational firms often replicate the same production activities in multiple countries. In terms of the OLI framework, foreign countries are attractive production locations to serve the local market when shipping costs and tariffs are high. Multiplant operation comes at the cost of failing to fully exploit increasing returns to scale in production. Because these models feature firms that replicate the same activity across countries, they are often referred to as horizontal FDI models.

4.1. Homogeneous Firms

We begin with the case in which firms within an industry are homogeneous and countries are identical to fix ideas with respect to the industry characteristics that affect firm choice of international organization. We then extend the analysis to allow for substantial differences across countries in terms of their endowment. We conclude this subsection with a discussion of evidence concerning the implications of this class of models.

4.1.1. A Symmetric Model

In order to understand the horizontal motive for multinational activity, it is simplest to start with the one-sector ($J = 1$), one-factor model in [Krugman \(1980\)](#), with two symmetric countries each endowed with L units of labor.¹⁰ We drop all subscripts j and set the wage rate to 1 in both countries without loss of generality, so that (3) reduces to

$$C(q) = f + \frac{q}{\varphi}.$$

For reasons that will become apparent below, we will distinguish two activities that jointly contribute to the fixed cost f . First, the invention or process of differentiation of a variety requires a fixed cost of f_E units of labor: f_E is a measure of economies of scale *at the firm level* and the level of this expenditure is unaffected by the number and location of plants that end up producing the variety (see [Markusen, 1984](#); [Helpman, 1984](#)).¹¹ This fixed cost f_E may include a variety of activities such as R&D expenditures, brand development, accounting, and finance operations. The second component of the fixed cost is denoted by f_D and captures overhead costs necessary for the manufacturing or assembly of the product. Unlike in the case of f_E , the fixed cost f_D needs to be incurred every time

¹⁰ The model in this section is a simplified version of the one in [Brainard \(1993, 1997\)](#).

¹¹ In the literature, one often encounters more general specifications for the technology representing the process of creation of intangibles. For instance, [Helpman \(1984\)](#) allows for a variable choice of firm intangibles with larger levels of intangibles reducing the marginal cost of downstream activities.

a plant is set up, and thus this parameter captures economies of scale *at the plant level*. Quite naturally, the distinction between f_E and f_D only becomes relevant in multiplant environments, such as the one we are about to consider.

The main innovation relative to [Krugman \(1980\)](#)'s model is that we do not force firms entering one country to service the other country via exporting. Instead, we allow firms to incur the plant-level fixed costs twice and service consumers abroad via local sales, thus avoiding the iceberg-type transportation costs $\tau > 1$. Under which conditions will firms find it optimal to deviate from the “pure-exporting” equilibrium that was the focus of [Krugman \(1980\)](#)? In order to answer this question, we first need to construct such an equilibrium and then explore the circumstances under which our suggested deviation is profitable.

Consider then the problem faced by a firm based in country $i = H, F$ that produces variety ω and services both markets from a plant located in i . Given the isoelastic demand schedule in (2), an exporter from $i = H, F$ will choose a domestic and foreign price that is a constant markup over marginal cost, thus yielding a total profit flow equal to

$$\pi_X^i = \varphi^{\sigma-1} B^i + \varphi^{\sigma-1} \tau^{1-\sigma} B^{-i} - f_E - f_D, \quad (4)$$

where

$$B^i = \frac{1}{\sigma} \left(\frac{\sigma-1}{\sigma} \right)^{\sigma-1} A^i. \quad (5)$$

In the industry equilibrium, free entry ensures that $\pi_X^H = \pi_X^F = 0$ and thus we must have:

$$B^H = B^F = B_X = \frac{f_E + f_D}{\varphi^{\sigma-1} (1 + \tau^{1-\sigma})}, \quad (6)$$

from which, using (5) and (2), one can back out the number of entrants in each country.

Having described the equilibrium, we next check whether a firm might have an incentive to deviate from this equilibrium and set up a plant in the foreign country. A potentially deviating firm engaged in FDI would obtain a profit flow equal to

$$\pi_I = \varphi^{\sigma-1} B^H + \varphi^{\sigma-1} B^F - f_E - 2f_D, \quad (7)$$

regardless of where the firm incurred its entry cost. Given the standard continuum assumption in monopolistic competition models, this deviator would have no impact on the demand level in either country and hence, plugging (6) into (7), we find that such deviation will be unprofitable (i.e., $\pi_I^i < 0$) whenever

$$\frac{2f_D}{f_E} > \tau^{\sigma-1} - 1. \quad (8)$$

As long as condition (8) holds, one can thus safely ignore the possibility of multinational activity and focus on equilibria with exporting. It is clear that this is more likely to be the

case, the higher are plant-specific economies of scale relative to firm-specific economies of scale (f_D/f_E), the lower are transport costs τ , and the lower is the elasticity of substitution σ . The intuition behind the effects of f_D and τ on the tradeoff between exporting and affiliate sales is rather straightforward. The effects of f_E and σ in (8) are more subtle, and to understand them it is important to bear in mind that from the point of view of a firm, the choice of exporting versus affiliate sales boils down to the choice between a selling strategy associated with low fixed costs but higher marginal costs (exporting) and another one associated with high fixed costs and low marginal cost (affiliate sales). Quite naturally, the latter option might look appealing when the firm expects a large volume of sales abroad and when profits are more sensitive to marginal costs. With that in mind, the positive effect of f_E on the likelihood of a deviation simply reflects the fact that in industries in which the process of innovation is more costly, the industry will end up being populated by a relatively low number of firms, and thus each firm will end up with a higher residual demand level (6). Finally, in industries in which varieties are relatively substitutable with each other, profits will tend to be particularly sensitive to marginal costs, and thus a deviation from a pure exporting equilibrium will be attractive.¹²

Up to now, we have focused on describing the circumstances under which a pure exporting equilibrium of the type studied by Krugman (1980) is robust to the possibility of servicing offshore markets via affiliate sales. A natural question is then: what happens when condition (8) is violated? It is straightforward to show that when the inequality in (8) is reversed then the unique equilibrium will be one with “pure affiliate activity” in the sense that no firm will have an incentive to sell abroad via exporting. This equilibrium features two-way FDI flows between countries that is a key feature of the data as summarized by Stylized Facts One and Two in Section 2. In the knife-edge case in which the left- and right-hand sides of (8) are equal to each other, there might be a coexistence of exporters and multinational firms within an industry, but their relative shares remain indeterminate.¹³

4.1.2. Country Asymmetries

We now consider an extension of the model above that incorporates a role for factor endowment differences to shape trade and multinational enterprise (MNE) activity flows across countries. We build on the work of Markusen and Venables (2000), whose

¹² Note that the common market size L does not enter the key condition (8) because, with CES preferences and free entry, the residual demand level faced by firms is independent of market size. With the linear demand system with horizontal product differentiation developed by Ottaviano and Thisse (2002), a symmetric increase in market size would tighten condition (8), thus making an equilibrium with pure exporting less likely. Below, we will illustrate nonneutral effects of asymmetric changes in size even in the CES case.

¹³ When the number of firms in an industry is fixed (entry is restricted), the share of firms that engage in FDI relative to exports can be determined for a wide range of parameter values as the price index is influenced only by firms' mode choices and not due to a free entry condition.

framework also falls under the umbrella of our benchmark model in [Section 3](#). In particular, they consider a model in which there are two sectors, a homogeneous-good sector ($\beta_z > 0$) and a differentiated-goods one ($J = 1$), and goods are produced using capital and labor, and the endowments of these factors can vary across countries. Their framework abstracts from intraindustry heterogeneity (φ is common for all firms) and trade costs are only variable in nature ($f_X = 0$).¹⁴ We will come back to their two-sector, two-factor model later in the section, but we will first demonstrate that some of their key results continue to apply in a simpler one-sector, one-factor model, with the benefit that these results can be demonstrated analytically.

With that in mind, let us revisit the model developed in the last section, but now let countries differ in their endowment of labor, with $L^H > L^F$ without loss of generality. It is worth emphasizing that one can interpret labor as a composite factor of production and that, given the absence of technology differences across countries, it is advisable to treat L as an efficiency-adjusted endowment of this factor. Thus the larger “size” of Home might reflect both differences in population and differences in the productivity of labor, the latter stemming from technology or from the availability of other factors that are complementary to labor. The key implication of introducing differences in country size is that, with only one sector in the economy, wages across countries will no longer be equalized, and as shown by [Krugman \(1980\)](#), in the absence of MNE activity, it is necessarily the case that the wage rate is higher in the larger country ($w^H > w^F$).

How do these factor price differences (and, indirectly, size differences) affect the likelihood of equilibria with MNE activity? To simplify matters, we follow [Markusen and Venables \(2000\)](#) in assuming that fixed costs of innovation are a function of factor prices in all countries in which production takes place. In the case of exporters, these fixed costs are simply given by $w^i f_E$, where $i = H, F$ is the country where the variety is produced. For multinational firms (i.e., firms with plants in both countries), these costs are given by $\frac{1}{2} (w^H + w^F) f_E$. The latter assumption implies that fixed costs in terms of labor are evenly split across locations, and thus the model does not allow one to distinguish between the “home” and “host” country of a multinational firm (or between the location of the parent company and that of the affiliate). This is a clear limitation of the analysis but it has the benefit of considerably simplifying the description of the parameter space in which MNE activity will arise in equilibrium. Notice that in the absence of factor price differences and setting labor as the numeraire, the fixed costs of entry for MNEs are equal to f_E , as in the symmetric model above.¹⁵

¹⁴ Strictly speaking, [Markusen and Venables \(2000\)](#) allow for general homothetic preferences over the homogeneous good and the CES aggregator of differentiated varieties, but their results are all numerical in nature and their simulations rely on a Cobb-Douglas aggregator analogous to the one we have imposed in (1).

¹⁵ [Markusen and Venables \(2000\)](#) consider alternative formulations of the fixed costs of entry that allow for substitution between Home and Foreign labor. This does not appear to affect the qualitative nature of the results, but it greatly complicates the analytical derivation of the results below.

We will next describe an equilibrium of the Krugman-type, with single-plant firms in both countries servicing offshore consumers via export sales. After straightforward manipulations analogous to those in the previous section, profits for an exporter from country $i = H, F$ can be expressed as:

$$\pi_X^i = \varphi^{\sigma-1} (w^i)^{1-\sigma} B^i + \varphi^{\sigma-1} (\tau w^i)^{1-\sigma} B^i - w^i (f_D + f_E),$$

where B^i is defined in (5). If both types of exporters are to break even B^H and B^F must be such that $\pi_X^H = \pi_X^F = 0$. Solving for these equilibrium values for B^H and B^F , and plugging them into the profit function of a potential MNE deviator yields

$$\pi_I = \varphi^{\sigma-1} (w^H)^{1-\sigma} B^H + \varphi^{\sigma-1} (w^F)^{1-\sigma} B^F - (w^H + w^F) \left(f_D + \frac{1}{2} f_E \right).$$

It thus follows that, for an equilibrium with no MNE activity to exist (i.e., $\pi_I < 0$), it needs to be the case that

$$\frac{\left(\frac{w^H}{w^F} \right) \left(\tau^{\sigma-1} - \left(\frac{w^H}{w^F} \right)^{-\sigma} \right) + \tau^{\sigma-1} - \left(\frac{w^H}{w^F} \right)^{\sigma}}{\left(\frac{w^H}{w^F} + 1 \right) (\tau^{\sigma-1} - 1)} \frac{f_D + f_E}{f_D + \frac{1}{2} f_E} < \frac{\tau^{\sigma-1} + 1}{\tau^{\sigma-1}}. \quad (9)$$

This expression is more cumbersome than the one in (8), but unambiguous comparative statics can still be obtained. First, holding constant the relative wage w^H/w^F , it can be verified that condition (9) is more likely to hold whenever f_D/f_E is high and whenever τ is low. Hence, one can still interpret the choice between exporting and FDI sales as reflecting a proximity-concentration tradeoff even in the presence of country asymmetries. The main novelty relative to the analogous condition (8) in the symmetric case is the fact that the tradeoff now also depends on the ratio of the Home and Foreign wages. Simple (though tedious) differentiation demonstrates that the left-hand-side attains a unique maximum at $w^H/w^F = 1$ (see the Online Appendix). In other words, the incentive for firms to deviate from a pure exporting equilibrium is highest when factor price differences across countries are small. Evaluating (9) at $w^H/w^F = 1$, we find that multinational firms can only arise in equilibrium whenever

$$\tau^{\sigma-1} - 1 > \frac{2f_D}{f_E} \quad (10)$$

just as in the symmetric model. The introduction of size asymmetries implies, however, that even when this inequality holds, as we shall assume for the remainder of this section, the equilibrium may be one with no MNE activity when factor prices are sufficiently dissimilar across countries.

Why do factor price differences favor an equilibrium with pure exporting? Intuitively, whereas exporting firms only use labor from one country, multinational firms use labor from both countries. Thus, the higher are factor price differences, the higher is the cost disadvantage faced by multinational firms relative to exporters from the low-wage country.

We have thus far treated wages as given, but they are naturally pinned down in equilibrium by labor-market clearing. In a Krugman-type equilibrium with no MNEs, the relative wage is implicitly given by

$$L^H \left(\frac{\tau w^F}{w^H} \right)^{\sigma-1} + L^F = \frac{w^H}{w^F} \left(L^H + L^F \left(\frac{\tau w^H}{w^F} \right)^{\sigma-1} \right). \quad (11)$$

As is well known, (11) implies that $w^H > w^F$ whenever $L^H > L^F$. Furthermore, the relative wage w^H/w^F is increasing in the relative size L^H/L^F of the Home country, and converges to $\tau^{(\sigma-1)/\sigma}$ when $L^H/L^F \rightarrow \infty$ (this is not proven in Krugman, 1980, but we do so in the Online Appendix).

An implication of this result is that even when (10) holds, the equilibrium will still necessarily only feature exporting when the relative size L^H/L^F of Home becomes sufficiently large.¹⁶ The threshold relative size $\bar{\lambda} > 0$ above which the equilibrium is of the type described in Krugman (1980) is implicitly defined by (11) and by condition (9) when it holds with equality.

What does the equilibrium look like for relative sizes L^H/L^F below the threshold $\bar{\lambda}$ and above 1? Our previous discussion implies that such equilibria will necessarily feature a positive mass of multinational firms as long as (10) holds. In the Online Appendix, we rule out the existence of “pure MNE” equilibria without exporters. Furthermore, numerical simulations suggest that that only two other types of equilibria can arise for $L^H/L^F > 1$. For an interval of relative sizes right below the threshold $\bar{\lambda}$, that is $L^H/L^F \in (\underline{\lambda}, \bar{\lambda})$, the equilibrium is one with MNEs, Home exporters, and Foreign exporters all coexisting. In that interval, the relative wage is implicitly given by (9) holding with equality independently of L^H/L^F , with changes in the relative supply of labor of each country being absorbed via changes in the mix of firms. If the ratio L^H/L^F is even closer to 1 and below a second threshold $\underline{\lambda}$, then the equilibrium is one in which multinational firms and Home exporters coexist, but Foreign exporters cannot profitably operate. Equilibria without exporters from the relatively large country (i.e., Home) do not appear to exist although we have not been able to prove this analytically.

We have focused on the case in which asymmetries arise because Home is larger than Foreign. The case $L^H/L^F < 1$ can be analyzed similarly and delivers analogous results. For a sufficiently small L^H/L^F , we have a pure exporting equilibrium, and the remaining

¹⁶ To be precise, when transport costs are so large that $\tau^{\frac{\sigma-1}{\sigma}} - 1 > 2f_D/f_E$, the equilibrium features MNE activity for all $L^H/L^F > 0$ (see the Online Appendix). We will abstract from this uninteresting case below.

equilibria feature some MNE activity and some exporting from the larger country, in this case Foreign.

We argued above that one should not interpret variation in L across countries as merely reflecting differences in population, as it might also reflect variation in the availability of factors that are complementary to labor. This brings us precisely to the contribution of [Markusen and Venables \(2000\)](#), who study a two-sector model in which production uses labor and a complementary factor (capital) which, crucially, are combined under different factor intensities in the two sectors, so that L can no longer be treated as a composite factor. In terms of our notation in the benchmark model, the factor cost w^i in the unique differentiated sector is now denoted by $c_1^i = c_1^i(w^i, r^i)$, while the unit cost of production in the homogeneous-good sector is given by $c_z^i = c_z^i(w^i, r^i)$. It should be clear that our derivations above regarding the likelihood of an equilibrium with pure exporting remain unaltered, with c_1^i replacing w^i everywhere. Hence, it continues to be the case that the likelihood of an equilibrium with MNE activity is highest when cost differences are lowest.

Unfortunately, the general equilibrium of this two-sector, two-factor model becomes analytically intractable, so [Markusen and Venables \(2000\)](#) have to resort to numerical analysis from the outset. Their simulations suggest that, in the absence of multinationals, factor price differences across countries appear to be large whenever relative and absolute factor endowment differences across countries are large. This resonates with the effects of absolute differences in L on equilibrium wages in the [Krugman \(1980\)](#) model. When introducing the possibility of MNE activity, [Markusen and Venables \(2000\)](#) graphically illustrate that as long as condition (10) holds, there is a two-dimensional region in the endowment space (which includes the point $L^H = L^F$ and $K^H = K^F$) in which only multinational firms will operate in the differentiated-good sector.¹⁷ Further away from the midpoint of the endowment space, they find that there are two-dimensional regions of the endowment space in which the equilibrium is mixed, with both MNEs and exporting firms being active. Their results also suggest that equilibria with differentiated-good sector exporters from a given country are more likely the larger the size of that country, and in particular the larger that country's endowment of the factor used intensively in the differentiated-good sector. The prediction of the model that FDI is more likely to arise between similarly endowed countries is consistent with the broad facts reported in [Section 2](#).

4.1.3. Evidence

The proximity-concentration tradeoff is intuitive and has a variety of implications for the cross-country and cross-industry structure of trade and FDI, but how well does it explain the data? The most well-known paper assessing the proximity-concentration framework is [Brainard \(1997\)](#) who relates the cross-country and cross-industry structure of trade

¹⁷ One way to rationalize the fact that their model is able to feature pure FDI equilibria (and the one-sector, one-factor model above does not) is that the presence of the homogeneous-good sector provides firms with a channel to repatriate profits across asymmetric markets.

and multinational production to proxies for various features of the industry's technology and various features of a country's endowment and policy environment. In this section, we provide estimates of the coefficients of Brainard's econometric model and theory-inspired extensions to that model using more up-to-date data than was available when the original paper was written.

The basic structure of Brainard's econometric model relates the propensity of U.S. firms in industry j to serve country i by exports relative to serving country i via a local affiliate as given by the equation:

$$\log \left(\frac{X_j^i}{X_j^i + S_j^i} \right) = \alpha_0 + \alpha_1 \text{Freight}_j^i + \alpha_2 \text{Tariff}_j^i + \alpha_3 \text{PlantSC}_j + \alpha_4 \text{CorpSC}_j + \alpha_5 \mathbf{C}^i + \mu_{ij},$$

where X_j^i is the exports by the United States in industry j to country i and S_j^i is the sales by U.S. affiliates in industry j located in country i . The explanatory variables Freight_j^i and Tariff_j^i are the logarithm of an *ad valorem* measure of shipping costs in industry j from the U.S. to country i and the logarithm of average tariffs facing U.S. firms in industry j and country i . The explanatory variables PlantSC_j and CorpSC_j are logarithms of measures of plant scale economies (number of production workers per representative plant) and corporate scale economies (the number of nonproduction workers per representative firm). Finally, \mathbf{C}^i is a vector of country controls that include a variable GDP/POP^i , which is the logarithm of the absolute difference in GDP per worker in the United States and in country i . This variable controls for factor endowment differences across countries.¹⁸

The dependent variable here captures the tendency of firms to substitute local production for exports while controlling for industry and country characteristics that jointly explain both the level of trade and the level of affiliate sales. As noted in [Section 2](#), trade and FDI flows are highly correlated both in gross and net terms so these factors that determine both the levels of trade and FDI are in fact important.

Brainard estimates this equation using U.S. trade data and affiliate sales data from the 1989 Benchmark Survey of U.S. Direct Investment Abroad conducted by the Bureau of Economic Analysis (BEA). She finds that α_1 , α_2 , and α_4 are negative and statistically significant while α_3 is positive and statistically significant. This is the best known evidence in favor of the proximity-concentration framework. She also finds that the coefficient on GDP/POP is positive and statistically significant, which suggests that FDI is primarily directed toward developed countries. If one treats GDP/POP as a proxy for wage differences across countries, the positive coefficient is supportive of our theoretical result above regarding the positive effect of wage differences on the likelihood of an equilibrium with pure exporting.

¹⁸ Other country controls include TRADE^i , an index of policy openness to trade, FDI^i , an index of policy openness to foreign direct investment, ADJ^i , a dummy variable for Canada and Mexico, English , a dummy variable for English Speaking countries, EC^i , a dummy variable for membership in the European Community, and PolStab^i , an index of political stability. See the Online Appendix for more details on these variables.

How well do Brainard's result stand the test of time? To answer this question, we estimate the same equation for both 1989, the year of Brainard's data, and for 2009, the most recent Benchmark Survey that is available from the BEA. Our sample is slightly different from that considered by Brainard. Specifically, we consider only manufacturing industries (excluding several primary-good industries considered by Brainard) and a broader range of countries.¹⁹ Our explanatory variables are also constructed from high-quality data that became available only after Brainard (1997) was written (see our Online Appendix for more information on these data sources).

Our baseline results are shown in columns 1 and 2 of Table 2.3. Column 1 corresponds to the year 1989 while column 2 corresponds to 2009. To conserve space we have suppressed the coefficient estimates for all country controls except for *GDP/POP*. Comparing the estimates in the two columns, there are not substantial differences across years. As was the case in Brainard (1997), we find that higher trade costs (*Freight* and *Tariffs*) are associated with a statistically significant decrease in exports in favor of affiliate sales while higher plant-level fixed costs (*PlantSC*) and lower corporate fixed costs (*CorpSC*) are also associated with more exports and less affiliate sales. These results support the predictions of the model described in Section 4.1.1. Conversely, the coefficient estimates on *GDP/POP* are both positive but are not statistically significant in either year. These results call into question the conclusion that income per capita differences (a proxy for wage differences) are associated with less affiliate sales and more exports.²⁰

A strong implication of the Markusen and Venables (2000) framework—and of our one-factor version of it—is that relative country sizes play an important role in the structure of bilateral commerce between them. To assess this implication, we added a variable *GDP*, which is the logarithm of the absolute value of the difference in the GDP of the United States and the GDP of country *i*. Running this augmented regression on 2009 data, we obtain the coefficient estimates shown in column 3. The positive and (moderately) statistically significant coefficient on *GDP* suggests that market size does matter: U.S. firms are more likely to export to smaller markets and to engage in foreign direct investment in larger countries.

To further explore the role of relative factor endowment differences, we next replace *GDP/POP* (GDP per worker differences) with absolute differences in years of schooling (*School*) and capital-labor ratios (*KL*), and estimate the coefficients on the 2009 data. The

¹⁹ The countries included in our sample are Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, China, Colombia, Costa Rica, Denmark, Ecuador, Egypt, Finland, France, Germany, Greece, Hong Kong, India, Indonesia, Ireland, Italy, Israel, Japan, Malaysia, Mexico, Netherlands, New Zealand, Norway, Peru, Philippines, Portugal, Singapore, South Africa, South Korea, Spain, Sweden, Switzerland, Thailand, Turkey, the United Kingdom, and Venezuela.

²⁰ When the country controls are suppressed, the coefficient on *GDP/POP* is statistically significant in both years. However, if we limit our sample to the countries considered in Brainard (1997) the coefficient is not statistically significant in either year. Note that although our sample is broader than considered by Brainard, it does not include all countries. It is quite possible that if the least developed countries had been included, the coefficient on *GDP/POP* might well have been positive.

Table 2.3 Proximity-Concentration Empirics

Dep. Var.: $\log \left(\frac{x_j^i}{x_j^i + s_j^i} \right)$	(1)	(2)	(3)	(4)	(5)	(6)
Freight	-0.28** [0.05]	-0.13** [0.04]	-0.12** [0.04]	-0.13** [0.04]	-0.13* [0.06]	0.01 [0.25]
Tariffs	-0.23** [0.06]	-0.28** [0.05]	-0.27** [0.05]	-0.29** [0.06]	-0.38** [0.10]	-0.04 [0.04]
GDP/POP	0.10 [0.07]	0.04 [0.08]	0.06 [0.08]			
School				0.07 [0.09]		
KL				0.08 [0.06]		
GDP			0.32 [0.17]	0.39* [0.17]		
PlantSc	0.09* [0.04]	0.13* [0.05]	0.13* [0.05]	0.14* [0.05]	0.18 [0.15]	
CorpSc	-0.18** [0.03]	-0.32** [0.04]	-0.31** [0.04]	-0.32** [0.04]	-0.35** [0.14]	
Country Fixed Effects?	No	No	No	No	Yes	Yes
Industry Fixed Effects?	No	No	No	No	No	Yes
Year	1989	2009	2009	2009	2009	2009
Observations	1,762	2,315	2,315	2,315	2,482	2,482
R-Square	0.15	0.09	0.09	0.09	0.16	0.40

Standard errors are in brackets.

*At 1%.

**Significant at 5%.

results, shown in column 4, reveal that neither variable obtains a statistically significant coefficient, which again suggests that there is no role of relative factor endowments in the substitution of affiliate production for exports that is uniform across all industries. We will show in [Section 5.4](#) that to identify the role of comparative advantage requires a model specification that allows the factor abundances of countries to interact with factor intensities of industries.

Finally, we consider the robustness of the proximity-concentration variables to country fixed effects, and country and industry fixed effects. The results are shown in columns (5) and (6). As the estimates in column (5) indicate, the proximity-concentration variables are robust to the inclusion of country fixed effects. However, when both types of fixed effects are included, we obtain the estimates for *Freight* and *Tariff* shown in column (6) that are very close to and statistically indistinguishable from zero. The limits to the robustness of

trade cost as a predictor of the export share are not new to our exercise: Brainard (1997) obtained a similar result.

4.2. Firm Heterogeneity

As noted by Fact Four in Section 2, international economic activity is highly concentrated in a handful of very large and productive firms. We now show that explicitly addressing firm heterogeneity in productivity improves the proximity-concentration model by allowing it to be consistent with microlevel facts and by expanding our understanding of the industry characteristics that explain aggregate variables. We do so by considering the framework in Helpman et al. (2004), focusing exclusively on the industry characteristics that drive the proximity-concentration tradeoff.

Consider the simple proximity-concentration model discussed in Section 4.1.1. Again countries are symmetric ($L^H = L^F$) so that we may normalize $w^H = w^F = 1$. For convenience, we limit our attention to a generic industry and suppress industry subscripts for notational simplicity. The key new assumption is that productivity φ varies across producers of differentiated goods. A firm that is contemplating entry does not initially know its productivity φ but knows that it will be drawn from a distribution $G(\varphi)$ that has support on the interval $[\varphi, \bar{\varphi}]$ where $\bar{\varphi} > \varphi$. After paying the industry-specific fixed cost f_E , the firm learns its productivity.

Upon observing its productivity an entrant may open a plant in its home country at an additional fixed cost of f_D or it may exit. If it chooses to produce, it may also choose to sell its variety in the other country, but doing so requires that the firm pay a marketing fixed cost f_X . To serve the foreign market, the firm may export from the plant in its home location and pay an industry-specific *ad valorem* shipping cost τ . Alternatively, the firm may open a foreign affiliate by incurring the marketing fixed cost f_X and an additional plant-level fixed cost f_D .

We now analyze the decisions of a firm that has entered one of the countries and has drawn a productivity of φ . The profit (net of entry cost) earned in the home country market for an active firm of productivity φ is

$$\pi_D(\varphi) = B\varphi^{\sigma-1} - f_D. \quad (12)$$

If the firm were to export its variety to the foreign market, the *additional* profit earned would be

$$\pi_X(\varphi) = B\varphi^{\sigma-1}\tau^{1-\sigma} - f_X, \quad (13)$$

and if it were instead to open an affiliate in the foreign country, its *additional* profit would instead be

$$\pi_I(\varphi) = B\varphi^{\sigma-1} - f_X - f_D. \quad (14)$$

The profit functions in (12), (13), and (14) for a given industry are plotted in Figure 2.5. This figure shows that firms sort into their modes of serving global markets. As in

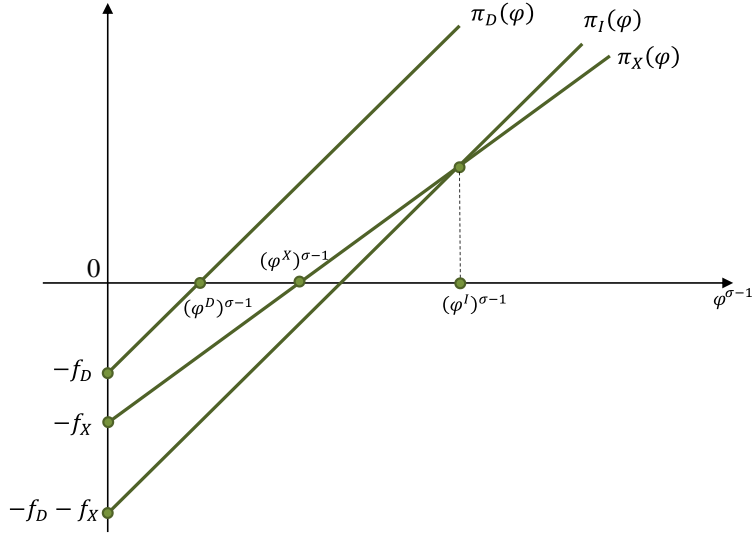


Figure 2.5 Exports, FDI and Heterogeneity

Melitz (2003), firms that receive poor draws $\varphi < \varphi^D$ exit. The least productive firms serve only their domestic market $\varphi \in (\varphi^D, \varphi^X)$ while more productive firms also sell their product abroad ($\varphi > \varphi^X$). Of the firms selling their variety abroad, the most productive do so through a foreign affiliate ($\varphi > \varphi^I$).²¹ This sorting result stems from the supermodularity of the profit function in productivity and in other marginal cost reducing measures: a highly productive firm that sells a lot of units benefits more from a reduction in marginal cost than a firm with lower productivity that sells fewer units.²²

Unlike the case in models with symmetric firms considered in Section 4.1, firms in a narrowly defined industry organize their international production through different modes and each firm *strictly prefers* its organization strategy to any alternative. The prediction that firms sort on the basis of their productivity has been confirmed in a number of studies. For instance, Girma et al. (2004) and Mayer and Ottaviano (2007) use kernel regression techniques to show that distributions of productivities across firms are consistent with this prediction, with the distribution of firm productivities of multinationals shifted to the right of the distribution of exporters, and even further to the right from the distribution of firms that do not sell their product abroad. Figure 2.6 provides an illustration of this result using 2007 data from the Spanish Encuesta sobre Estrategias Empresariales (ESEE), which also permits reporting the productivity advantage of both

²¹ Figure 2.5 has been plotted under parameter restrictions that ensure a positive mass of firms engaged in exporting and in FDI. More specifically, this requires $f_X + f_D > (\tau^{ij})^{\sigma-1} f_X > f_D$.

²² See Mrázová and Neary (2012) for a discussion of the role of supermodularity. Other preference systems can give rise to a profit function in which this feature is absent leading to different sorting implications.

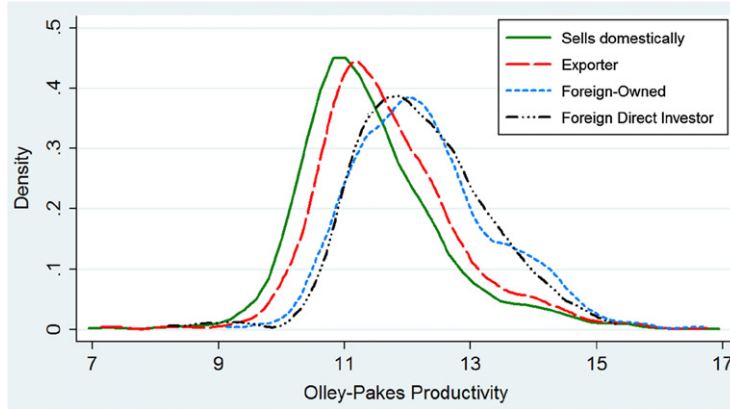


Figure 2.6 Selection into Horizontal FDI in Spain

multinational firms based in Spain as well as of foreign-owned affiliates in Spain relative to domestic producers and exporters. The differences are large and statistically significant.

We now turn to the implications of the model for the aggregate trade and multinational production data that are most widely available. To highlight the substitutability between exports and local affiliate sales (as in Brainard, 1997) we consider the share of sales by industry that are sold via exports relative to those sold through foreign affiliates. First, to fully characterize the equilibrium, we apply the free entry condition, which requires that the *expected* profit from entry is equal to zero. This condition is given by

$$B \left(V(\varphi^D) + \tau^{1-\sigma} (V(\varphi^X) - V(\varphi^I)) + V(\varphi^I) \right) - (1 - G(\varphi^D))f_D - (1 - G(\varphi^I))f_D - (1 - G(\varphi^X))f_X - f_E = 0, \quad (15)$$

where

$$V(\hat{\varphi}) = \int_{\hat{\varphi}}^{\bar{\varphi}} \varphi^{\sigma-1} dG(\varphi). \quad (16)$$

Next, integrating over firms' export sales yields aggregate exports of $S^X = \sigma B \tau^{1-\sigma} (V(\varphi^X) - V(\varphi^I))$. Repeating this procedure for affiliate sales, we obtain aggregate affiliate sales of $S^I = \sigma B V(\varphi^I)$. The ratio of aggregate exports to aggregate affiliate sales is then

$$\frac{S^X}{S^I} = \tau^{1-\sigma} \left(\frac{V(\varphi^X)}{V(\varphi^I)} - 1 \right). \quad (17)$$

It can be established through differentiation of the cutoff conditions, of the free entry condition (15), and of the function (17) that many of the standard proximity-concentration predictions are maintained in the model with heterogeneous firms. In particular S^X/S^I tends to fall as trade costs τ rise (as is also true for the fixed cost of marketing f_X that is new to this section) and tends to rise as plant-level fixed costs f_D

rise. Unlike the case of homogeneous firms as considered in [Section 4.1.1](#), the change is smooth rather than an abrupt shift to a different corner. The effect of a change in corporate fixed costs, f_E , on S^X/S^I is more subtle. An increase in f_E tends to raise the mark-up adjusted demand level B , which in turn causes all cutoffs to fall. As long as $\varphi^I < \bar{\varphi}$ the effect on S^X/S^I depends on the shape of the distribution function G (see [equation 16](#)). However, as φ^I is monotonically decreasing in B , for sufficiently small f_E the cutoff φ^I will exceed any finite upper bound $\bar{\varphi}$ and so multinational firms will disappear. In this sense, the effect of f_E on the likelihood of FDI is similar to that in the case of homogeneous firms.

To get further results, [Helpman et al. \(2004\)](#) assume that the upper limit of the support $\bar{\varphi}$ of the distribution G is infinite and that the distribution is Pareto: $G(\varphi) = 1 - \varphi^\kappa \varphi^{-\kappa}$. Note that this assumption requires the parameter restriction $\kappa > \sigma - 1$, otherwise the integrals aggregating affiliate sales and profits will not converge. The Pareto assumption has become popular in the literature because it requires few parameters, yields clean analytic solutions, and can be partially justified because it implies (along with the CES preference system) that the size distribution of firm sales is also Pareto, which is a reasonable approximation to the data for the far right tail.

The Pareto assumption means that we can solve the integrals in (17) and then substitute for the cutoffs to obtain the clean expression that relates S^X/S^I to exclusively exogenous parameters:

$$\frac{S^X}{S^I} = \tau^{1-\sigma} \left[\left(\frac{f_D}{f_X} \frac{1}{\tau^{\sigma-1} - 1} \right)^{\frac{\kappa}{\sigma-1}-1} - 1 \right]. \quad (18)$$

The parameters τ , f_X , and f_D shape the ratio as discussed in the general case above. Interestingly, with the Pareto assumption and the infinite bound on productivity, the corporate fixed cost f_E has no impact on the ratio of exports and FDI sales. Intuitively, an increase in f_E causes both φ^X and φ^I to change by the same proportion and the Pareto distribution ensures that the sales of firms that switch modes as a result of these changes in the cutoff are also in the same proportion, leaving the ratio unchanged.

Note that the distributional parameter κ , working in concert with the elasticity of substitution σ , also has an impact on the composition of commerce. A decrease in this ratio is associated with a decrease in S^X/S^I . It can be shown that if we hold fixed a particular market, the Pareto parameterization (combined with the CES demand system) imply that the standard deviation of the logarithm of sales by all active firms in this market is equal to $(\sigma - 1)/\kappa$. Hence, an increase in the variance of sales across firms within an industry should be associated with a decrease in S^X/S^I . To understand this result consider [Figure 2.7](#), where we plot the Pareto density function for two values of κ . It is clear from the figure that a decrease in κ raises the density of the function at any $\varphi > \varphi^I$ —where

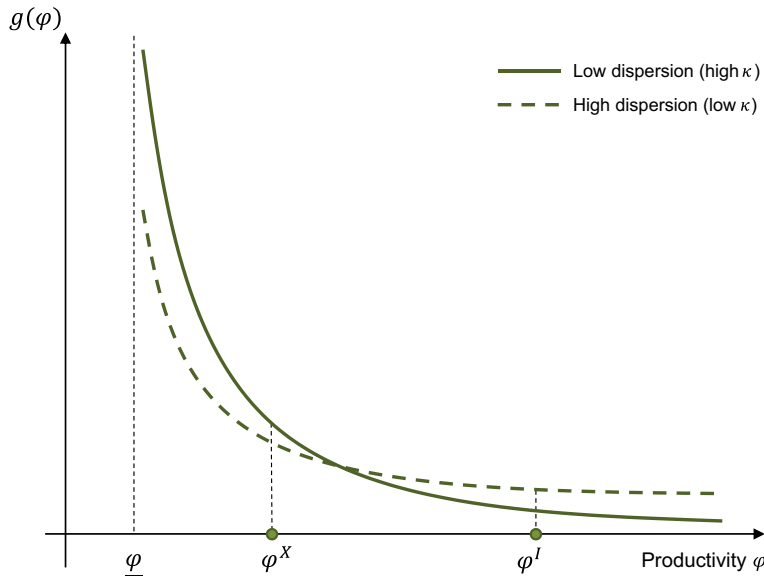


Figure 2.7 Dispersion and the Proximity-Concentration Hypothesis

firms find FDI optimal—relative to the density at any $\varphi \in (\varphi^X, \varphi^I)$ —where firms choose exporting.²³

Helpman et al. (2004) estimate a logarithm linearized version of equation (18) using U.S. data for 1994. They include as regressors country fixed effects, the standard proximity-concentration variables, and a measure of the standard deviation of the logarithm of firm sales by industry as a measure of dispersion.²⁴ They confirm Brainard's results concerning trade costs and plant-level fixed costs (measured as the average number of nonproduction workers per plant), and they find that an increase in the dispersion measure is associated with a large, statistically significant decrease in S^X/S^I . The extent of firm heterogeneity seems to matter for the aggregates. Their result must be qualified, however, because their estimates of $\frac{\kappa}{\sigma-1}$ for many industries are not consistent with the restriction $\kappa > \sigma - 1$.²⁵ This means that their results do not have a structural interpretation.²⁶

In the working paper version of the paper, Helpman et al. also show that the model can be extended to allow for modestly asymmetric geographies. When trade costs between

²³ Note that the mean of the Pareto distribution $G(\varphi) = 1 - \underline{\varphi}^\kappa \varphi^{-\kappa}$ is $\mu = \underline{\varphi} \kappa / (\kappa - 1)$, while the variance is $\varepsilon^2 = \mu^2 / (\kappa(\kappa - 2))$. A decrease in κ thus not only increases the variance of the distribution but also its mean. Nevertheless, because S^X/S^I in (18) is independent of the lower bound $\underline{\varphi}$, one can always offset the effect of a lower κ on the mean μ with a reduction in $\underline{\varphi}$, and still raise ε^2 . Thus, a mean-preserving spread raises FDI relative to exports.

²⁴ They also include measures of dispersion estimated from the size distribution of firms with similar results.

²⁵ The finding that the slope of a linear regression of log rank on log size is less than one is common. The commonly used Pareto approximation only holds for the tail of the distribution.

²⁶ There is an error in the paper that has been the source of some confusion. On page 307 it is stated that the estimators of dispersion are measuring $\kappa - \sigma + 1$ rather than $\kappa / (\sigma - 1)$.

regions are higher than within regions two tiers of multinationals arise: the most productive firms own affiliates in all countries while the less productive multinationals open a single affiliate that they use to serve the entire region. This yields an additional empirical implication: an increase in firm productivity predicts affiliate entry into a larger number of countries. Yeaple (2009) and Chen and Moore (2010) have shown this prediction to be consistent with the data (for U.S. and European multinationals, respectively).

4.3. Greenfield FDI versus Mergers and Acquisitions

We have modeled the foreign entry decision as a fixed cost of setting up a variety-specific plant in a foreign location. This has the flavor of what is called “greenfield” entry in which a firm builds its foreign production capacity from scratch. However, according to Fact Six in Section 2, in practice most firms that acquire new plants in a foreign location do so by purchasing previously existing plants through cross-border mergers and acquisitions. This fact has resulted in some discussion in the literature as to whether the models that we have described above are consistent with the actual entry modes that firms undertake.

In this section, we extend Helpman et al. (2004) in such a way as to incorporate both greenfield entry and cross-border acquisitions. In this extension, greenfield and cross-border acquisitions coexist and are in fact perfect substitutes from the perspective of the firm (although banning cross-border acquisitions would have consequences in the model). We then discuss the model’s implications in light of a small but growing empirical literature. We find that there is evidence that in fact firms do not behave as if cross-border acquisitions and greenfield investment were perfect substitutes. We close this section by describing some theory in which the two entry modes are different.

Going back to the model of Helpman et al. (2004) discussed above, suppose that prior to receiving its productivity draw, an entering firm must pay both f_E and f_D . Thus, the fixed cost of building a plant in the home country market is now sunk. The only other important deviation from the Helpman et al. (2004) model is that a firm may sell its plant to any other entrant on a perfectly competitive merger market. Firms that acquire that plant may now install their own productivity draw φ (replacing the purchased firm’s technology) and so avoid paying a fixed cost f_D . As all purchased firms will be homogeneous with respect to purchasers, they will receive the acquisition price in the market of P_A .

A firm will sell itself on the merger market if P_A exceeds the profit that it could generate independently while an acquirer would strictly prefer greenfield if $f_D < P_A$ and cross-border acquisitions otherwise. Hence, in equilibrium it must be that $P_A \leq f_D$ and that all firms with productivity $\varphi \leq \varphi^D$ would sell themselves on the merger market while all foreign firms with $\varphi \geq \varphi^I$ would either engage exclusively in cross-border acquisitions if $P_A < f_D$ or be indifferent between greenfield investment and cross-border acquisition if $P_A = f_D$. The equilibrium we have just constructed is as characterized in the previous section in spirit but we now have a mechanism that gives rise to cross-border

acquisitions. Note, however, that cross-border acquisitions do matter in the extension, for if they were not allowed, then weaker firms would have no incentive to exit.

This extension shows that it need not be the case that cross-border acquisitions are fundamentally different from what was modeled in the previous section.²⁷ It also can be used as a benchmark in interpreting the stylized facts. Let us consider the implication of the model that firms are indifferent between the two entry modes. [Nocke and Yeaple \(2008\)](#) have used data from the Bureau of Economic Analysis to show that parent firms of U.S. multinationals that enter foreign markets appear to sort into their mode of entry: more productive parent firms tend to acquire new plants via greenfield entry rather than cross-border entry. This suggests that the firms do not perceive these modes as perfect substitutes.

Now consider the implication of the extended model that the firms acquired by foreign firms are the least productive in their industries. A number of studies suggest the exact opposite. For instance, [Arnold and Javorcik \(2009\)](#), [Guadalupe et al. \(2012\)](#), and [Blonigen et al. \(2012\)](#) are examples of recent papers that suggest that it is the most productive firms within an industry that tend to be the targets of foreign acquisitions. This again suggests that cross-border acquisitions and greenfield investment are not perfect substitutes.

A number of recent papers develop equilibrium models in which there is sorting of firms into their mode of entry into foreign markets. For instance, [Nocke and Yeaple \(2007, 2008\)](#) develop equilibrium models in which firms are heterogeneous in multiple dimensions that interact in a complementary way to generate profits. The firm heterogeneity is embodied in intangible assets that can be transferred across firms only through ownership and some of these assets are imperfectly mobile across countries (such as relationships with customers and local suppliers, etc.). Equilibrium in a merger market involves improving the assignment of intangible assets to firms to exploit complementarities. Greenfield investment occurs when a firm's relatively immobile asset is so useful that it pays to move it despite high relocation costs. Similar mechanisms appear in models provided by [Guadalupe et al. \(2012\)](#), in which foreign firms possess innovative abilities and access to foreign markets that domestic firms lack and these abilities are best applied to local firms that are already highly productive.

Before leaving the topic of entry mode, it is important to mention that an alternative explanation for cross-border acquisitions is that firms acquire one another in order to reduce competition. The literature exploring this phenomenon develops oligopolist environments and explores the types of acquisitions that might arise whether between or within countries. Examples of such papers are [Horn and Persson \(2001\)](#), [Qiu and Zhou \(2006\)](#), and [Neary \(2007\)](#). An important implication of some of these models is that a

²⁷ Note that firm heterogeneity is important to generate firms that want to sell and firms that want to buy plants.

reduction in trade costs can spur waves of cross-border acquisitions. [Breinlich \(2008\)](#) provides some supportive evidence for this prediction.

5. VERTICAL EXPANSION

Some of the empirical shortcomings of horizontal FDI models are due to the fact that FDI is often motivated by factors other than the mere replication of the production processes in different countries to save on transportation costs or jump over tariffs. In reality, multinational activity might also be motivated by cost differences across countries, which induce some producers to locate different parts of the production process in different countries. Indeed, Fact One, that net FDI flows from developed to developing countries, and Fact Five that R&D expenditures within the multinational firm are concentrated in the parent firm, suggest a need to address international specialization. Further, in all industries there are exports from affiliates to their parent firms that are not easily explained in a purely horizontal setting, and in particular industries the volume of these exports is very large (see [Table 2.2](#)).

In this section, we discuss some works that formalize and empirically test these insights, starting with the simplest case of the so-called pure vertical FDI model of [Helpman \(1984\)](#).

5.1. A Factor-Proportions Model of Vertical FDI

Let us now return to the benchmark model in [Section 3](#). In order to shut down the horizontal incentive for multinational activity, let us suppose for now that there are no transport costs so exporting is the dominant strategy in the models in [Section 4](#). For simplicity, consider a two-sector ($\beta_z > 0, J = 1$), two-factor model, in which production uses capital and labor, and capital intensity is higher in the differentiated-good sector than in the traditional-good sector. For the time being, assume also that the distribution of firm productivity is degenerate within sectors and that the cost functions in each sector are identical in both countries, so there are no Ricardian differences in technology. We will relax the latter two assumptions later in this section. So far, the model is essentially a $2 \times 2 \times 2$ [Helpman and Krugman \(1985\)](#) model of international trade, with slightly stronger assumptions on preferences and technology.

The key new feature is that we will now make an explicit distinction between different production stages in the differentiated-good sector and that we will allow for a geographic separation of these stages. More specifically, the total cost function in the differentiated-good sector now follows from the aggregation of the total cost of two distinct production processes: headquarter services provision and manufacturing.

Headquarter services provision entails activities that are typically provided by the production unit from which the ownership advantage originated, i.e., the headquarters. It is useful to think of these as being closely related to our interpretation of the fixed

cost of entry in horizontal FDI models, and thus entail R&D expenditures, brand or specialized machine development, financing, etc. This is consistent with our Fact Five in [Section 2](#). These activities have important fixed cost components, but we will also allow their level to affect continuously the marginal product of manufacturing, as in [Helpman \(1984\)](#). In analogy to horizontal FDI models, headquarter services are assumed specialized and are valuable only to the firm incurring them, but within the firm, they are nonrival and can serve many manufacturing plants simultaneously, regardless of where these manufacturing plants are located. Headquarter services are produced with capital and labor according to the total cost function $C_H(w^i, r^i, h^i)$, where h^i is the number of units of services produced and $C_H(\cdot)$ is associated with an increasing returns to scale production function.

The manufacturing stage uses capital and labor to convert the available headquarter services into final goods according to an increasing returns to scale technology in which h is essential and which is represented by a cost function $C_M(w^i, r^i, \bar{h}, q^i)$, where q^i is the amount of output produced and \bar{h} is the aggregate amount of firm-level headquarter services available for production. $C_M(\cdot)$ is naturally decreasing in \bar{h} and increasing in the other arguments. Following [Helpman \(1984\)](#), we assume that the production of headquarter services is more capital intensive than manufacturing.

Because trade is costless and there are increasing returns associated with each stage, firms will not produce headquarter services or carry out manufacturing in more than one location. When both stages occur in the same country i , the total cost of production will be given by

$$C_1(w^i, r^i, q^i) = \min_{h^i} \{ C_M(w^i, r^i, h^i, q^i) + C_H(w^i, r^i, h^i) \}. \quad (19)$$

The benchmark model in [Section 3](#) is a particular case of this more general multistage model in which we forced firms to locate all production processes in the same country.²⁸ The key question that [Helpman \(1984\)](#) posed was then: under which conditions will firms want to deviate from this “Krugman-style” equilibrium and locate headquarter services and manufacturing in different countries, thus becoming multinational firms? An advantage of having related our current model to the benchmark model in [Section 3](#), which in turn is a variant of the canonical $2 \times 2 \times 2$ model studied by [Helpman and](#)

²⁸ Strictly speaking, in our benchmark model, we had specified a homothetic total cost function (see [equation \(3\)](#)) involving a fixed cost and a constant marginal cost in terms of a linearly homogeneous function of capital and labor, denoted by $c_1(w^i, r^i)$. Given our assumptions, nothing ensures that the solution to (19) will indeed give rise to a homothetic cost function of the form in (3). A simple way to make these two formulations consistent with each other (while having headquarter services be more capital intensive than manufacturing) is to assume that: (i) the fixed costs of headquarter services’ provision and manufacturing are both linear in $c_1(w^i, r^i)$; (ii) the marginal cost of headquarter services is a constant amount of capital; (iii) the marginal cost of manufacturing is a constant amount of labor; and (iv) output is a linearly homogeneous function of the level of headquarter services and manufacturing, i.e., $q^i = q(h^i, m^i)$, where $q(h^i, m^i)$ is the primal production function associated with the cost function $c_1(w^i, r^i)$. Naturally, it would be straightforward to relax these conditions to accommodate nonextreme factor intensity in marginal costs.

Krugman (1985), is that we can readily appeal to their results to conclude that multinational activity will be a (weakly) dominated strategy as long as relative factor endowment differences across countries are small. The reason is simple: for sufficiently similar relative factor endowments, factor price equalization (FPE) will attain even in the absence of fragmentation and thus firms based in one country will have no incentive to open subsidiaries (production plants) in the other country.

Outside the traditional FPE set, firms with headquarters in the relatively capital-abundant country would have an incentive to “deviate” from the equilibrium and offshore manufacturing to the relatively labor abundant country simply because wages are lower there in the absence of multinational activity. Helpman (1984) further demonstrated that, outside the traditional FPE set, there exists a set of “not-too-extreme” relative factor endowment differences for which a unique *minimum* measure of multinational firms is sufficient to bring about factor price equalization, thus eliminating the incentive for more firms to fragment production across borders. Importantly, within this set, Helpman (1984) showed that for a constant relative size of the two countries, the minimum measure of multinational firms consistent with equilibrium would be increasing in relative factor endowment differences. Although the model above does not feature intrafirm trade in physical goods, it does feature invisible exports of headquarter services from each headquarter to its subsidiary. Assuming that these services are valued at average cost, Helpman (1984) showed that, for a given relative size of countries, the share of intrafirm trade in the total volume of trade is increasing in relative factor endowment differences.

5.2. Vertical FDI and Wage Inequality

A variant of the Helpman (1984) model in which the two factors of production are interpreted as skilled and unskilled labor sheds light on the effect of vertical FDI (and offshoring, more broadly) on wage inequality in each country. Notice that in the model above, whenever fragmentation brings about factor price equalization, it does so by increasing the relative demand for and remuneration of the relatively abundant factor in each country, thereby increasing wage inequality in the relatively skilled-labor abundant country, while reducing it in the relatively unskilled-labor abundant country. This is essentially the Stolper-Samuelson theorem at work but with the movement in factor prices being shaped by changes in relative input costs rather than in relative final-good prices.

In recent years, there has been an active literature in international trade studying the robustness of this result to alternative and much richer general-equilibrium environments with offshoring. As might have been expected from the well-known limited generality of the Stolper-Samuelson theorem, the effect of fragmentation on relative factor prices and thus wage inequality has been shown to be sensitive to modeling assumptions. A large number of models have been proposed, but we will focus below on outlining two of the most popular ones developed in recent years, the ones in Feenstra and Hanson (1996)

and [Grossman and Rossi-Hansberg \(2008\)](#). While both of these models were originally cast in neoclassical environments with homogeneous goods, perfect competition, and constant returns to scale, it is straightforward to embed them in our benchmark model in [Section 3](#).

The main result in [Feenstra and Hanson \(1996\)](#) can be illustrated by considering a variant of our benchmark factor-proportions model of vertical FDI in which, instead of there being two production stages in the differentiated-good sector, there are now a continuum of stages x indexed by $s \in [0, 1]$, and the marginal cost for each activity or input is given by

$$c_x(w^i, r^i, s) = a_K(s) r^i + a_L(s) w^i,$$

where we now use the notation K and r to denote skilled workers (i.e., human capital) and their wages, while L and w are now associated with unskilled workers. This continuum of stages is arranged in increasing order of their skill intensity, so that $a_K(s)/a_L(s)$ is increasing in s , and they are combined into final output according to a Cobb-Douglas technology

$$q_1 = \exp \left(\int_0^1 \alpha(s) \ln x(s) ds \right),$$

where $x(s)$ is the value of services at stage s . As in our benchmark model, and differently than in the framework in [Feenstra and Hanson \(1996\)](#), we assume that there also exist fixed costs of production at each stage, but we assume that all these are in terms of final-good output, so that the total cost function in the differentiated good, $C_1(r^i, w^i, q_1^i)$, remains homothetic. This is isomorphic to assuming that activities or inputs are produced according to a constant-returns-to-scale technology but that there exists a fixed cost (in terms of output) associated with the assembly of intermediate inputs. We finally assume, for simplicity, that the homogeneous final-good sector uses only unskilled labor.

Despite the fact that one of the sectors features product differentiation, increasing returns to scale, and monopolistic competition, the general equilibrium of the model is very similar to that in the continuum version of the Heckscher-Ohlin model in [Dornbusch et al. \(1980\)](#). First, and as in [Helpman \(1984\)](#), if relative factor endowment differences between countries are sufficiently small, the equilibrium features factor price equalization and there is no incentive to fragment production across countries. For slightly higher relative factor endowment differences, fragmentation will occur in equilibrium and will bring about factor price equalization, but unlike in [Helpman \(1984\)](#), the actual pattern of offshoring is indeterminate given the large dimensionality of the commodity space. The most interesting results emerge when relative factor endowment differences are large enough to ensure that factor prices are not equalized even when fragmentation is costless. As in [Dornbusch et al. \(1980\)](#) or [Feenstra and Hanson \(1996\)](#), in such a case, there exists a marginal stage s^* such that the relatively skill-abundant country (call it Home) specializes in the production of stages $s > s^*$, while the unskilled-labor abundant country

(call it Foreign) specializes in the stages $s \leq s^*$ and the production of the homogeneous good z . This pattern of production is supported by an unskilled-labor wage which is strictly higher at Home ($w^H > w^F$) and a skilled-labor wage which is strictly higher in Foreign ($r^H < r^F$).

As in [Helpman \(1984\)](#), relative to a world in which fragmentation is not feasible, the factor price equalization set is larger and equilibrium factor price differences are smaller with fragmentation. Nevertheless, within a free trade equilibrium, it is easy to construct comparative static exercises such that increases in offshoring are associated with increased wage inequality *in both countries*. For instance, consider a situation in which relative factor endowment differences are so large that FPE does not attain even with frictionless fragmentation. Consider then a proportional increase in the supply of both factors in Foreign. This will naturally increase the range of tasks produced in (or offshored to) Foreign, and these marginal tasks will feature a higher skill intensity than the tasks previously produced in Foreign. As a result, the relative demand for skilled workers will increase and, in equilibrium, so will the Foreign wage premium (see the Online Appendix for details). Similarly, the relative demand for skills and the wage premium will also increase at Home because these offshored marginal tasks feature a lower skill intensity than the activities that remain at Home. It should be emphasized, however, that relative to a world in which fragmentation is not feasible at all, wage inequality in Foreign is always lower with some offshoring, just as in [Helpman \(1984\)](#).

The above example illustrates that the prediction linking increases in offshoring to reduced wage inequality in less developed countries is not a robust one (see also [Zhu and Trefler, 2005](#); [Antràs et al., 2006](#)). We next briefly outline a variant of the framework in [Grossman and Rossi-Hansberg \(2008\)](#) which demonstrates that increases in offshoring might not necessarily be associated with increased wage inequality in skill-abundant countries. To do so, let us first return to our benchmark vertical FDI model with only two inputs, headquarter services and manufacturing, and consider situations in which free trade in these inputs (or what we call above, fragmentation) brings about factor price equalization in that model. To be more specific, we consider a situation in which the unskilled-labor abundant Foreign is “large” and produces the homogeneous good as well as both inputs or stages in the differentiated-good sector, while the skill-abundant Home produces only the two inputs in the differentiated-good sector. Differently than in our benchmark model, we shall assume, however, that the technologies available to firms in both countries are now different and in particular they feature higher productivity at Home than in Foreign. More specifically, productivity levels are $\zeta < 1$ times lower in Foreign than at Home in all sectors and because we are focusing on a situation with *conditional* factor price equalization, factor prices will also be ζ times lower in Foreign than at Home.

Suppose now that each of the two inputs is produced by combining a continuum of measure one of “tasks” involving unskilled labor L and a continuum of measure one

of tasks involving skilled workers K . The two inputs continue to vary in terms of the intensity with which these continua of tasks generate value, with headquarters being more skill-intensive than manufacturing. The key innovation relative to our previous model is that we will now allow unskilled-labor tasks to be offshorable, by which we mean that firms at Home can carry them out using Foreign workers while still using the superior Home production technologies. Offshoring is however costly and tasks are ordered according to their degree of offshorability. Producing task s in Foreign inflates unit labor requirements by a multiple $\beta t(s) \geq 1$, where $t'(s) \geq 0$. For simplicity, we assume that skilled-labor tasks are prohibitively costly to offshore.²⁹

Given these assumptions, there exists a threshold task s^* such that all unskilled-labor tasks with index $s < s^*$ are offshored to Foreign, while all those with $s > s^*$ remain at Home, with the threshold being implicitly defined by

$$w^H = \beta t(s^*) w^F.$$

The marginal costs of producing headquarter services and of producing final goods are then

$$c_H(w^H, w^F, r^H) = w^H a_{LH} \Omega(s^*) + r^H a_{KH} \quad (20)$$

$$c_M(w^H, w^F, r^H) = w^H a_{LM} \Omega(s^*) + r^H a_{KM} \quad (21)$$

where

$$\Omega(s^*) \equiv 1 - s^* + \frac{\int_0^{s^*} t(s) ds}{t(s^*)} < 1.$$

Equations (20) and (21) make it clear that offshoring in this framework is isomorphic to unskilled-labor biased technological change. Because Foreign is large and produces both inputs, the marginal costs of each input are pinned down by their Foreign levels.³⁰ As a result, an increase in the offshorability of tasks, captured by a fall in β , will necessarily increase the unskilled-labor wage w^H commensurately with the fall in $\Omega(s^*)$, while leaving the skilled-labor wage r^H unaffected. In sum, the offshorability of certain types of tasks leads to a relative increase in the reward of the factor that is used intensively in precisely those tasks. The flip side of this surprising result is another counterintuitive result: increased offshorability leads to an expansion in the share of Home employment devoted to the unskilled-intensive manufacturing at the expense of employment in headquarter services.

5.3. Vertical FDI and Firm Heterogeneity

So far, we have developed vertical FDI models in which firms are homogeneous in productivity. As in the case of horizontal FDI, however, the evidence suggests that aggregate

²⁹ This assumption is relaxed in Grossman and Rossi-Hansberg (2008).

³⁰ This is true even if headquarter services are firm-specific, given that the location decision precedes production.

vertical MNE activity is accounted for, to a large extent, by a few very large and productive firms. For instance, [Ramondo et al. \(2012\)](#) report that despite the quantitative significance of the intrafirm component of U.S. trade, the median affiliate reports no shipments to its U.S. parent. This fact motivates us to introduce firm heterogeneity into the vertical model developed so far. In order to keep the analysis tractable, however, we will develop a one-factor (labor) variant of the [Helpman \(1984\)](#) model in which cross-country differences in factor prices arise due to Ricardian technological differences rather than due to differences in relative factor endowments.³¹

More specifically, assume that the cost function in the homogeneous-good sector is $c_z^i(w^i, r^i) = a_z^i w^i$ with $a_z^H < a_z^F$. Assuming further that β_z is large enough to ensure that both Home and Foreign produce the numeraire good z (i.e., $z^H, z^F > 0$), we necessarily have $w^H = 1/a_z^H > 1/a_z^F = w^F$, and producers of differentiated goods face a perfectly elastic supply of labor at those wage rates. In the differentiated-good sector, headquarter services provision first requires an initial fixed cost of entry or innovation, after which producers learn their productivity φ which is drawn from $G(\varphi)$. Firms then decide to exit or stay in the market and produce. In the latter case, headquarters need to incur an additional fixed cost to be described below. Having incurred this cost, they choose a variable amount of headquarter services h to combine with manufacturing in production. Home is assumed to be much more productive than Foreign in innovation/entry and in the production of headquarter services, so these are always produced at Home. Denote by f_E the initial fixed cost of entry and by f_D the fixed cost of headquarter services provision, and assume that units of h can be produced one-to-one with labor. Our assumptions on technology imply that in equilibrium all these costs are defined in terms of Home labor. Manufacturing entails no overhead costs and units of m can be produced one-to-one with labor in both countries. Foreign thus has comparative advantage in manufacturing. Final-good production combines h and m according to the technology

$$q^i(\varphi) = \varphi \left(\frac{h^i}{\eta} \right)^\eta \left(\frac{m^i}{1-\eta} \right)^{1-\eta},$$

where $1 < \eta < 1$ is a sectoral level of headquarter intensity, while φ measures firm-level productivity.

The lower wage in Foreign makes offshore manufacturing appealing from the point of view of the firm's headquarter at Home. We shall assume, however, that there are costs associated with the fragmentation of production. In particular, an additional fixed cost $f_I - f_D > 0$ is required from the headquarters at Home when h and m are geographically separated, and such fragmentation also entails an iceberg transportation cost $\tau > 1$, associated with shipping the manufactured input m back to the Home country (τ could also reflect communication or coordination costs). Trade in final goods remains free.

³¹ Factor-proportions models of monopolistic competition with productivity heterogeneity and trade costs are notably difficult to work with (see [Bernard et al., 2007](#)).

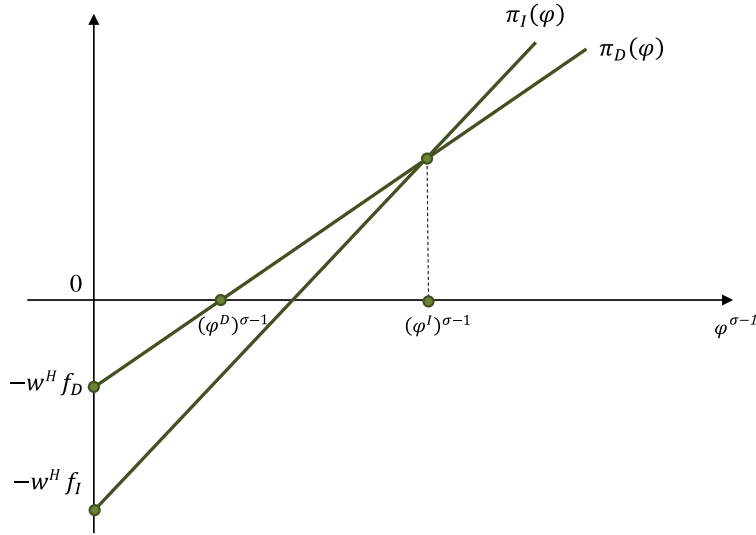


Figure 2.8 Vertical FDI and Heterogeneity

Given the assumptions above, it is straightforward to solve for the operating profits (net of entry costs) associated with domestic sourcing (or no fragmentation)

$$\pi_D(\varphi) = \varphi^{\sigma-1} (B^H + B^F) (w^H)^{1-\sigma} - w^H f_D \quad (22)$$

and those under vertical FDI or offshoring

$$\pi_I(\varphi) = \varphi^{\sigma-1} (B^H + B^F) \left((w^H)^\eta (\tau w^F)^{1-\eta} \right)^{1-\sigma} - w^H f_I, \quad (23)$$

for a firm with productivity φ , where remember B^i is defined in (5). These profit levels are plotted in Figure 2.8 under the assumption that $w^H > \tau w^F$, so some firms find it optimal to offshore.

As is clear from the figure, the model predicts selection into vertical FDI, in that only the most productive firms within an industry will find it profitable to engage in vertical FDI. This sorting pattern is consistent with the evidence on selection into importing in Bernard et al. (2009) who show that not only U.S. exporting firms but also U.S. importing firms appear to be more productive than purely domestic producers. Figure 2.9 provides further confirmation of this sorting pattern with 2007 data from the Spanish Encuesta sobre Estrategias Empresariales (ESEE). The dataset distinguishes between firms that purchase inputs only from other Spanish producers and firms that purchase inputs from abroad. As is clear from the picture, the distribution of productivity of firms that engage in foreign sourcing is a shift to the right of that of firms that only source locally.³²

³² Of course, foreign sourcing need not involve vertical FDI, as inputs can be purchased abroad from independent subcontractors. A key feature of the Spanish ESEE data is that it allows one to distinguish between foreign outsourcing

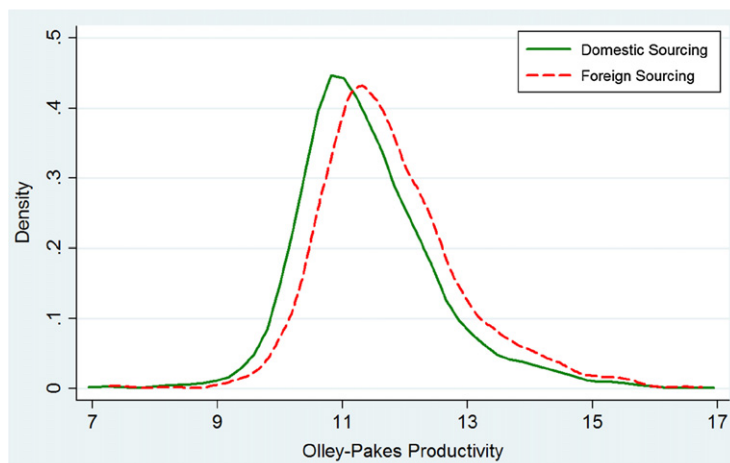


Figure 2.9 Selection into Vertical FDI in Spain

5.4. Vertical FDI: Empirical Evidence

The motives for vertical FDI are intuitive and there are thousands of examples of assembly and parts-producing affiliates in developing countries that fit the description. It is less clear how important vertical FDI is in the aggregates. Consider the facts discussed in Section 2. First, according to Fact One, most of FDI flows not to developing countries but to other developed countries. Second, only a relatively small fraction of affiliate output is exported back to the host country (see Fact Five). Third, as pointed out by [Ramondo et al. \(2012\)](#) very few foreign affiliates are engaged in international trade, selling all of their output in their host country.

A few more facts are worth mentioning before we turn to the econometric evidence. First, exports by U.S.-owned affiliates back to the United States are concentrated in a few countries. The top five affiliate export locations to the United States accounted for 60% of total U.S. affiliate exports to the U.S. while the corresponding number for affiliate sales in their host country is only 43%. This fact is perhaps unsurprising given that the point of vertical FDI is to concentrate production in low cost locations while horizontal FDI is about serial repetition of the same process across countries. Second, the list of countries that are highly engaged in hosting affiliates that export back to the United States features high variation in the level of their development. The top five countries in order of the size of U.S. affiliate exports back to the United States are Canada, Mexico, Ireland, the United Kingdom, and Singapore. This suggests that the treatment of comparative advantage in a two-country, two-factor model may be too simple to be consistent with the richness of the data. Evidently, some U.S. multinationals find Canada, a developed country, to be

and vertical FDI. As we will show later in [Figure 2.11](#), the productivity advantage of firms engaged in vertical FDI is even larger than the one observed in [Figure 2.9](#).

a relatively low cost production location for some production activities while Mexico is relatively low cost for others. The concept of comparative advantage, of course, is that being a low cost location in some activities means being a high cost location in others. Given this observation, it is perhaps not surprising that a regression of a measure of the extent of FDI on country endowments, as was done in [Section 4.1.3](#), yields a coefficient that is not statistically different from zero as the specification did not allow for the effect of endowments to vary across production activities.

[Yeaple \(2003a\)](#) allows for a more flexible treatment of comparative advantage in an effort to better capture vertical FDI in the data. He starts by noting that most of the evidence against vertical FDI comes from econometric studies that use data aggregated across industries to the country level. According to his interpretation, the [Helpman \(1984\)](#) model does *not* predict that FDI will be increasing in relative factor endowment differences. Instead, it predicts that in industries that are intensive in a particular factor, FDI should be flowing to countries that are abundant in that particular factor. In terms of our previous Brainard-style specifications, in which the relative prevalence of FDI is inversely related to the ratio of exports to exports plus the local sales of the affiliates of U.S. firms, [Yeaple \(2003a\)](#) argues that the econometric model should be specified as

$$\log \left(\frac{X_j^i}{X_j^i + S_j^i} \right) = \alpha_0 + \alpha_1 \text{Freight}_j^i + \alpha_2 \text{Tariff}_j^i + \alpha_3 \text{PlantSC}_j + \alpha_4 \text{SkillEnd}^i \\ + \alpha_5 \text{SkillInt}_j + \alpha_6 \text{SkillEnd}^i * \text{SkillInt}_j + \alpha_7 \mathbf{C}^i + \mu_{ij}, \quad (24)$$

where X_j^i is again export sales from the U.S. in industry j to country i and S_j^i are the corresponding affiliate sales. The key changes to the econometric specification relative to that in [Section 4.1.3](#) is the addition of variables SkillInt_j , which is the skill intensity of industry j , and $\text{SkillEnd}^i * \text{SkillInt}_j$, which is the interaction between skill intensity of an industry and SkillEnd^i , a measure of the skill-abundance of country i . The relevant coefficient for assessing the vertical motive of FDI is then α_6 rather than α_4 , and the predicted sign is negative: skill intensive activities should be done in skill-abundant countries and unskilled-labor intensive industries in unskilled-labor abundant countries.

We follow [Yeaple \(2003a\)](#) and estimate (24) using the most recently available data on U.S. exports and affiliate sales for the year 2009. We include all of the same control variables as in [Table 2.3](#) (coefficients suppressed below for exposition) and measure SkillEnd^i (the logarithm of average human capital per worker) and its interaction with an industry's skill intensity SkillInt_j (more details on the variables and data sources can be found in the [Online Appendix](#)). The results are shown in [Table 2.4](#).

The coefficient estimate in the first column confirms that when no interaction terms are included, skill endowments do not predict substitution between trade and affiliate production. Including skill endowments, skill intensity, and the interaction of the two (column 2) leads to coefficient estimates as anticipated: U.S. firms export more to a human capital scarce country in skill-intensive industries and substitute local production

Table 2.4 Skill Interactions

Dep. Var.: $\log \left(\frac{x_j^i}{x_j^i + s_j^i} \right)$	(1)	(2)	(3)	(4)
SkillEnd	−0.03 [0.29]	1.57* [0.63]	1.57* [0.64]	
SkillEnd * SkillInt		−10.5** [3.57]	−9.83** [3.38]	−8.77** [2.68]
SkillInt		13.7** [2.74]		
Country Fixed Effects?	No	No	No	Yes
Industry Fixed Effects?	No	No	Yes	Yes
Observations	2,315	2,315	2,315	2,482
R-Square	0.17	0.18	0.36	0.40

Standard errors are in brackets.

*At 1%.

**Significant at 5%.

for exports to human capital scarce countries in industries with low skill intensity. Note in columns 3 and 4 that the interaction term is robust to both country and industry fixed effects while the measures of *Freight* and *Tariffs* (not shown but as in Table 2.3) are not.³³

Vertical production relationships may also be more involved than in the simple model provided above. Many vertical relationships involve a parent firm shipping intermediate inputs to its foreign affiliates for further processing (see Figure 2.4). Rather than focus on factor endowment differences as the motive for this vertical specialization, Keller and Yeaple (2013) consider the input sourcing decisions of foreign assembly plants that sell their output in the host country market. They model the share of imported intermediates in the share of affiliates' total costs as the result of a tradeoff between costly technology transfer on the one hand and costly trade on the other hand. Under the maintained assumption that high R&D industries are the most burdened by technology transfer costs, their mechanism gives rise to gravity in affiliate sales (see Figure 2.3), and the effect of gravity is strongest in high R&D industries. They show that the mechanism is consistent with the data for U.S. multinationals and can explain approximately 30% of the gravity for aggregate affiliate sales that was noted as part of Fact Three.³⁴

³³ There are alternative explanations for the interactions. For instance, it could be that skill-intensive products are not consumed much in skill scarce (i.e. poorer) countries.

³⁴ See also Hanson et al. (2005) and Irrazábal et al. (2013) for additional evidence on the role of gravity. Another approach to thinking about vertical specialization appears in Alfaro and Charlton (2009). They identify vertically specialized affiliates as those that do not share an industry classification with their parent firms. According to this measure, vertical FDI begins to look quite ubiquitous but not necessarily driven by factor endowment differences across countries.

It is important to point out that the vertical model might be highly consistent with the offshoring activities of large internationally engaged firms, but that many of the firms that are engaged in such trade do not own the foreign production facility. As made clear by the OLI framework described in this chapter's Introduction ([Section 1](#)), a firm must not only see a benefit of relocating a production process in order for it to be a multinational, but it must also find integration of that facility within the firm as superior to arm's-length contracting. If the activities that are best done in developing countries are best done at arm's-length, then those production networks will not be observed within multinational firms. We will return to this idea in [Section 7](#) below.

6. MULTICOUNTRY MODELS

The two-country models discussed in [Sections 4](#) and [5](#) isolated trade costs and comparative advantage across stages of production as determinants of multinational production. In this section we consider multicountry models that incorporate both forces. These models address the large share of affiliate sales that are exported to locations other than the source country. The third country sales of multinational affiliates are quantitatively important, accounting for roughly a third of the sales of the affiliates of U.S. multinationals (see [Table 2.2](#)). These models also highlight some of the shortcomings of two-country models by showing how the characteristics of a country's neighborhood can affect the structure of its trade and multinational activity.

We consider two approaches. The first builds on the insights of the [Eaton and Kortum \(2002\)](#) model that treats locations as substitutes. In this setting an American firm might choose to serve a country like Germany either by exporting from the United States, producing in Germany, or by exporting from a plant located potentially anywhere. The firm's choice will depend on the cost of production in each country, the size of trade costs between each country and Germany, and possibly the desire to concentrate production for many markets in a single location to conserve on fixed costs of production. The second approach, which is relatively less developed, treats some sets of countries as substitutes for stages of production and other sets as complements along a vertical production chain. In this setting, a firm's decision to engage in horizontal investment in one location takes into account the proximity of that location to sources of low cost intermediates.

6.1. Locations as Substitutes

Historically, the analysis of multilateral trade treated the effects of comparative advantage and trade costs on trade patterns separately because multifactor models quickly become intractable in the absence of factor prices equalization. By modeling comparative advantage as arising from stochastic Ricardian technology shocks rather than as due to relative factor abundances, [Eaton and Kortum \(2002\)](#) made the study of comparative advantage

and trade costs in a single multicountry setting relatively tractable. Here we discuss how the insights of the Eaton and Kortum model have been applied with increasing sophistication to the analysis of trade and multinational production in a multicountry setting. As the models discussed in this section require substantial computational analysis, we sketch only their basic features and implications.

Consider a world with N countries, a continuum of industries indexed by j , and a single factor, labor. Perfect competition prevails in all markets. Each potential production location l has a labor productivity associated with production in industry j given by $\varphi_l(j)$ and (endogenous) wage rate of w_l . If the iceberg trade cost between country l and country n is τ_{ln} , then the marginal cost of serving country n from country l is $w_l \tau_{ln} / \varphi_l(j)$.³⁵ Naturally, each country purchases good j from the lowest cost location and its spending on that good depends on the marginal cost of provision. Because trade costs differ between bilateral pairs, the same good may be produced by multiple countries for different foreign markets.

The key insight of the Eaton and Kortum model is that if productivities across goods and across countries are random and drawn from a particular probability distribution, then aggregate bilateral trade flows between countries can be readily solved in general equilibrium. Moreover, the model relies on a small number of parameters that can be identified from data and equilibrium conditions. Eaton and Kortum parameterize this productivity as Fréchet

$$G_l(\varphi) = \exp(-T_l \varphi^{-\theta}),$$

where T_l governs the average productivity of country l (absolute advantage) and θ governs the dispersion of productivity across goods (comparative advantage). In the aggregate, this dispersion parameter governs how “substitutable” labor from each country is on the margin. It can then be shown that the probability that country l is the lowest cost supplier of any particular good to country n is

$$\psi_{ln} = \frac{T_l (w_l \tau_{ln})^{-\theta}}{\sum_{k=1}^N T_k (w_k \tau_{kn})^{-\theta}}, \quad (25)$$

which is also the share of country n expenditure that is spent on goods from country l . Note the sense in which θ governs the substitutability of labor from country l for labor from competing countries.

Ramondo and Rodríguez-Clare (2013) incorporate multinational production into the Eaton and Kortum model by assuming that each country receives a vector of productivity draws for each good from a multivariate Fréchet distribution where an element of the vector corresponds to a production location. A single parameter, ρ , governs the correlation of productivity draws of country i for a particular j across all possible locations.

³⁵ In the interest of simplicity, we ignore the treatment of intermediate inputs that features prominently in Eaton and Kortum (2002).

To be symmetric with the formulation of international trade costs, the authors assume that technology transfer costs from country i to country l are also iceberg and are denoted γ_{il} . When a productivity draw $\varphi_{il}(j)$ from country i in industry j is used to produce in a country l multinational production has occurred, and when that productivity draw is used to create output consumed in a country $n \neq i, l$ export platform FDI has taken place. In the paper, the authors assume that intermediate inputs from the source country are also an input into final production creating a degree of complementarity between multinational production and trade between the source and host country, while retaining the feature that production locations are substitutes for one another. Multinational production raises global output and welfare by allowing technologies to be used in their most appropriate location and by allowing technologies from productive countries to displace technologies from less technologically capable countries.³⁶

Arkolakis et al. (2013) take the insights of Ramondo and Rodríguez-Clare (2013) to a parameterized version of the Melitz model. As this model fits well into the benchmark framework of Section 2, we elaborate a little more on its structure and implications. As in the Melitz model, there is one differentiated-good industry featuring increasing returns to scale ($\beta_z = 0$, and $J = 1$) and labor is the only factor. To sell its variety in a country n the firm must first incur a fixed marketing cost f_X . As in Ramondo and Rodríguez-Clare, there is a full matrix of bilateral trade and technology transfer costs τ and γ .

To develop a new variety a firm must pay a fixed entry cost f_E . Upon paying this fixed cost a firm receives a vector of productivities from a multivariate Pareto distribution where each element corresponds to a productivity in a particular market. This distribution is

$$G_i(\varphi_1, \dots, \varphi_N) = 1 - T_i^e \left(\sum_{l=1}^N (T_l^p \varphi_l^{-\theta})^{\frac{1}{1-\rho}} \right)^{1-\rho}, \quad (26)$$

with support $\varphi_l \geq (T_i^e \sum_{l=1}^N (T_l^p)^{\frac{1}{1-\rho}})^{1/\theta}$, $\rho \in [0, 1)$, and $\theta > \sigma - 1$. As in the case of Ramondo and Rodríguez-Clare, this distribution has a parameter ρ that governs the correlation across countries in each firm's vector of draws. The parameter T_i^e governs the average quality of draws to entrants in country i , while the parameters T_l^p govern the average quality of draws for country l .

Conditional on selling its variety in country n (i.e., its variable profits justify the fixed marketing cost) it is straightforward to show that the probability that a firm that entered in country i will produce for n in country l is

$$\psi_{iln} = \frac{(T_l^p)^{\frac{1}{1-\rho}} (\gamma_{il} w_l \tau_{ln})^{-\frac{\theta}{1-\rho}}}{\sum_k (T_k^p)^{\frac{1}{1-\rho}} (\gamma_{ik} w_k \tau_{kn})^{-\frac{\theta}{1-\rho}}}. \quad (27)$$

³⁶ Global gains do not ensure that all countries gain as technology transfer has terms of trade effects.

Note how similar this expression is to [equation \(25\)](#), the one that appears in [Eaton and Kortum \(2002\)](#). The numerator of this expression contains the cost components for all firms serving country n from country l ($w_l \tau_{ln}$) adjusted for the bilateral pair-specific cost of technology transfer γ_{il} from country i to country l . The denominator summarizes the common marginal cost components of all potential locations for serving country n . In this sense, geography plays an important role in the manner in which firms choose their production locations. Notice also that the combination of parameters $\frac{\theta}{1-\rho}$ captures the substitutability of each location, and that as $\rho \rightarrow 1$ locations become perfect substitutes. Finally, [equation \(27\)](#) summarizes not only the probability that an individual firm uses country l as a production location, but also the share of all spending of country n on goods produced by country i firms (wherever they are located) that is done by affiliates in country l . It is important to note that this is not equal to the share of country n expenditure on goods from country i , which depend on the mass of firms that endogenously choose to enter in country i .

By endogenizing the development of new varieties (and their technologies) through free entry, [Arkolakis et al. \(2013\)](#) capture many of the features of the models discussed in earlier sections but now in a multicountry setting. Countries in which entrants receive good productivity draws on average (high T_i^e) will have a comparative advantage in entry (an analog to capital or skill-abundant countries in [Section 5](#)). High levels of entry drive up wages and induce firms to produce abroad with the extent of substitution between labor of different countries determined in part by ρ . Higher trade costs discourage trade in favor of multinational production as in [Section 4](#). The combination of fixed costs of entry combined with trade (τ) and technology transfer (γ) costs give rise to home market effects that can work in a number of ways.³⁷ For instance, when trade costs are large, production is attracted to large countries, thereby driving up the cost of innovation and inducing smaller countries to specialize in entry. When technology transfer costs are high, entry tends to occur in large countries as entrants are attracted by the large labor force. Given the competing forces at work in the model, the relative magnitude of the various effects depends on the model parameters including the full set of bilateral trade and technology transfer costs. The authors assess the model's empirically relevant implications by calibrating it to data on trade flows and multinational production shares across countries.

[Arkolakis et al. \(2013\)](#) achieve tractability by abstracting away from the fixed costs associated with production. In this sense, their model captures a proximity versus comparative advantage tradeoff but contains no fixed cost driven tendency toward concentration.

³⁷ In this respect, the model has implications that are similar to those of the knowledge capital model of [Markusen \(2002\)](#). In that two-country setting, comparative advantage is driven by relative factor abundances and interacts with home-market effects to determine the location of production. As is the case in the models discussed in this section, vertical (comparative-advantage driven) and horizontal (trade-cost driven) investment arise in a single framework depending on the relative size and factor abundances of countries. Finally, the knowledge capital model also delivers the possibility that countries can lose from multinational production.

Fixed costs of opening a foreign affiliate substantially complicate the analysis by making the location decisions to serve each market interdependent. For example, consider a firm that wants to sell its product in Germany and France. Minimizing marginal costs might mean selling to Germany from an Austrian affiliate and selling to France from a Spanish affiliate but the desire to concentrate production in one location that is created by large fixed costs might mean that Belgium is the best location to serve the two markets.

Tintelnot (2012) tackles these issues by further pushing the insights of Eaton and Kortum. In his model, a firm produces a continuum of products. Each time a firm opens an affiliate in a different country by incurring a fixed cost it receives a productivity draw for each of its products in that country. Now, the firm itself becomes a relevant unit of observation as the aggregation across products is at the level of the firm rather than at the level of the country. In this setting an equation of the sort of (27) now applies to the share of a firm from country i 's sales to country n that are made via an affiliate in country l given the set of countries in which the firm owns an affiliate. As the firm adds more production locations, it increases its fixed costs and the new production location reduces the share of products produced at its existing affiliates. That is, new production locations are good substitutes for some products in all existing production facilities. The firm will add new production locations until the reduction in the marginal cost of providing its portfolio of products to global markets is balanced by the additional fixed costs that need to be incurred. By pushing the smoothness from the Eaton and Kortum model into the firm, Tintelnot can structurally estimate the distribution of fixed and variable costs of multinational production using firm-level data.³⁸ He finds that these fixed costs play a large role in the geographic structure of multinational production and are necessary to explain the magnitude of export platform sales in the data.

6.2. Locations as Complements

The previous section discussed multicountry models in which comparative advantage and trade costs were analyzed in an environment in which firms perceived alternative foreign locations as substitutes: an improvement in the attractiveness of one location would tend to induce firms to substitute away from another. This section covers a small literature that demonstrates that in many instances, sets of foreign countries may be complements in multinational production.

Multinational production is motivated by some form of marginal cost reduction relative to purely national firms. By opening a foreign affiliate a firm can avoid transport costs (horizontal or replicating strategies) or reduce the cost of a stage of production by locating near key inputs (vertical or fragmentation strategies). We show how the two types of cost reductions can be complementary: a firm that replicates production of some

³⁸ To make his model tractable, Tintelnot simplifies relative to the Melitz-style models by abstracting from fixed marketing costs in each country and free entry.

activities offshore has a strong incentive to fragment production and vice versa. We refer to firms whose global organization reflects both types of multinational production as following *complex strategies*.

The framework developed here builds on the ideas in Yeaple (2003b) that were developed further in Grossman et al. (2006).³⁹ We begin with the setting used to illustrate the proximity concentration hypothesis in Section 4.1.1. There are two identical countries, H and F , that are endowed with only labor. There are two goods, one differentiated ($J = 1$) and one that is homogeneous ($\beta_z > 0$). The homogeneous good is freely traded while the differentiated industry faces transport costs $\tau > 1$ in shipping final-good varieties. Now, suppose that there is a third country, which we will refer to as S , that does not consume differentiated varieties. Choosing labor in H and F as the numeraire, we assume that the wage in S , w^S , is strictly less than one because of differences in the productivity in the outside-good sector. Entry of a differentiated-goods producer can only occur in H or F and requires f_E units of labor. Each plant requires a firm to incur fixed cost f_D .

Assume that the production of varieties of the differentiated good can be fragmented into two stages. The first is the intermediate good stage which requires one unit of labor to produce one unit of a firm-specific intermediate. The second stage is assembly that can only be done in H and F . We again choose units such that one unit of labor makes one unit of assembly possible. Building a plant to produce intermediates abroad requires the firm to incur a fixed cost f_I units of labor in its home market. Finally, while the final good is expensive to ship, the intermediate input is not subject to trade costs. The technology for producing a complete final good is such that the marginal cost is equal to $1/\varphi$ when producing both stages locally in a northern country, while it is $c(w^S)/\varphi < 1/\varphi$, when obtaining intermediates from a plant located in S .

Consider a firm from H that will serve both H and F . As in Section 4, the profits from a pure export strategy and a pure replication strategy are

$$\begin{aligned}\pi_X &= \varphi^{\sigma-1} B(1 + \tau^{1-\sigma}) - f_E - f_D, \text{ and} \\ \pi_R &= 2\varphi^{\sigma-1} B - f_E - 2f_D,\end{aligned}$$

respectively. Alternatively, the firm from H can open a plant in S to provide intermediates to its assembly plant located in H . The profits associated with this fragmenting (vertical) strategy are

$$\pi_V = \varphi^{\sigma-1} B(1 + \tau^{1-\sigma}) c(w^S)^{1-\sigma} - f_E - f_D - f_I.$$

Finally, suppose that a firm were to undertake both types of multinational production (complex FDI). In this case, the profits are

$$\pi_{VR} = 2\varphi^{\sigma-1} B c(w^S)^{1-\sigma} - f_E - 2f_D - f_I.$$

³⁹ Similar issues arise in Ekhholm et al. (2007).

It is straightforward to confirm from these equations and the assumptions over the relative costs ($1 > \tau^{1-\sigma}$ and $c(w^S)^{1-\sigma} > 1$) that the following inequalities must hold:

$$\pi_{VR} - \pi_R > \pi_V - \pi_X \Leftrightarrow \pi_{VR} - \pi_V > \pi_R - \pi_X.$$

The inequalities establish that if a firm has reduced its marginal cost in serving a foreign market through horizontal investment, it will gain more from opening an intermediate producing facility in S than if it had not, and that a firm that has opened an affiliate in S to provide intermediate inputs gains more from opening an assembly plant in the other northern country than had it not opened an intermediate plant. In short, having lowered marginal cost through one type of foreign investment, the firm optimally raises its output and so gains relatively more by lowering its marginal cost through the other type of foreign investment. Grossman et al. (2006) call this mechanism a *unit-cost complementarity*. This complementarity has interesting implications. For instance, an increase in trade costs can raise the relative payoff to horizontal investment and so induce a firm that would otherwise not do any foreign investment to simultaneously invest in F and in S as doing so only makes sense if both are done simultaneously. Similarly, although a reduction in f_I has no direct impact on the profitability of replicating strategies relative to national export strategies, it can result in local assembly in F replacing trade with H in final goods.

There have been few empirical studies motivated by models of complex strategies perhaps in part because of the conceptual complications that the models entail. Standard empirical methods treat shocks across countries as independent, but complex models warn us that this may be problematic.⁴⁰

7. MULTINATIONAL FIRM BOUNDARIES

So far we have reviewed theoretical frameworks that illustrate different types of gains associated with firms locating production processes in multiple countries. These models, however, cannot explain why these processes will be offshored within firm boundaries (thus involving FDI sales or foreign insourcing) rather than through arm's-length licensing or subcontracting. As such, and despite the terminology that we have used in the last two sections, these should not be viewed as *complete* theories of the multinational firm, but rather as theories of the *technological* drivers behind the international organization of production.⁴¹ In this section, we will review complete theories of the multinational firm that attempt to also shed light on the crucial internalization decision of multinational firms, and we will also review some empirical work testing these theories.

⁴⁰ Baldwin and Venables (2010) suggest an even greater complication. They show that the nature of interdependency across countries hinges on the temporal sequence of production stages.

⁴¹ Remember from Caves' (2007) definition in the Introduction (Section 1) that one needs to explain why multinational firms choose to *control* and *manage* their production establishments abroad.

The main unifying theme of the theoretical literature on multinational firm boundaries is the departure from the classical assumption of complete or perfect contracting. After all, and as first pointed out by Coase (1937), firm boundaries are indeterminate and irrelevant in a world in which transactions are governed by comprehensive contracts that specify (in an enforceable way) the course of action to be taken in *any* possible contingency that the contracting parties may encounter. In order to shed light on the internalization decision, this new literature on multinational firms and outsourcing has thus borrowed from the theoretical literature on incomplete contracts (cf., Williamson, 1975, 1985; Grossman and Hart, 1986), and has developed ways to incorporate these contracting frameworks into general equilibrium models. Different contributions emphasize different types of contractual frictions and they also adopt different approaches as to how the internalization of transactions affects these frictions. We will next discuss some of the key ideas and models in this literature.

7.1. Transaction-Cost Approaches

The transaction-cost theory of firm boundaries is based on the premise that firms will internalize particular transactions whenever the *transaction costs* associated with performing these transactions through the market mechanism are greater than within the firm (cf., Coase, 1937). The concept of transaction costs is somewhat vague, but it is often associated with inefficiencies that arise when transactions are not fully governed or secured by comprehensive contracts.⁴² We will discuss below two types of inefficiencies that have featured prominently in the literature: rent dissipation and hold up inefficiencies (see Garetto, 2013, for a different perspective).

7.1.1. Licensing versus FDI: Rent Dissipation

We first illustrate the relevance of rent dissipation within the horizontal model of FDI with symmetric firms and countries in Section 4.1.1. Remember that we argued above that the profits obtained by multinational firms were given by

$$\pi_I = \varphi^{\sigma-1} B^H + \varphi^{\sigma-1} B^F - f_E - 2f_D.$$

Implicit in that computation was the notion that Home firms could fully capture the net surplus associated with selling their varieties in the Foreign market, i.e., the amount $\varphi^{\sigma-1} B^F - f_D > 0$. This is a strong assumption when considering FDI transactions, and it is even harder to swallow when modeling market transactions. To be more specific, consider a licensing arrangement. In order to secure a profit flow equal to $\varphi^{\sigma-1} B^F - f_D$, the Home firm would need to be able to costlessly contract with a Foreign licensee that would commit to producing an amount of output optimally dictated by the Home firm,

⁴² Obviously, not all transaction costs are contractual in nature (take, for instance, the existence of taxes on market transactions).

collecting the sales revenue generated in Foreign, and handing the whole net surplus over to the Home firm, either via ex-ante transfers (before sale revenue is generated) or ex-post payments (after the revenue has been collected).⁴³

In practice, various types of contractual imperfections will lead to rent dissipation and the Home firm will end up sharing rents with foreign licensees. A particularly noteworthy source of frictions stems from the (partially) nonexcludable, nonrival, and noncodifiable nature of the technology that the Home firm is attempting to sell to the Foreign firm (see [Arrow, 1962](#); [Romer, 1990](#)). The partial nonexcludability of technology generates a risk of intellectual expropriation, which in turn typically limits the ability of the Home firm to appropriate surplus from the Foreign licensee *ex post*. The fact that the firm is selling intellectual property, a noncodifiable and nonrival good, will in turn limit the willingness of the Foreign firm to pay for the technology up-front (i.e., *ex ante*). Even abstracting from the risk of intellectual expropriation, moral hazard and private information constraints will typically also preclude full extraction of rents from licensees.

We next illustrate this rent dissipation phenomenon below with a simple model inspired by the work, among others, of [Ethier \(1986\)](#), [Horstmann and Markusen \(1987b\)](#), and [Ethier and Markusen \(1996\)](#). With that in mind, let us return to the Horizontal FDI model with symmetric countries, but now allow firms the option of servicing the foreign market via licensing. We will emphasize the contractual costs associated with licensing, so it is important to formally introduce the contracting assumptions and the timing of events. There are three relevant periods, $t = 0$, $t = 1$, and $t = 2$.

At $t = 0$, the firm that owns the blueprint for variety ω decides the mode via which it will service the foreign market: exporting, FDI, or licensing. The payoffs associated with exporting and FDI remain identical to those derived in [Section 4.1.1](#), reflecting the notion that these strategies can be performed with no transaction costs.

When licensing is chosen at $t = 0$, the technology owner offers a contract to a licensee abroad which stipulates an initial payment at $t = 1$ (L_1) and a second payment at $t = 2$ (L_2). We allow this initial contract to stipulate the amount of output to be produced by the licensee, but this assumption could be relaxed, thus opening the door for the type of moral hazard effects studied in [Horstmann and Markusen \(1987b\)](#). In order to produce, the licensee needs to incur an overhead cost equal to f_D , just as in our Horizontal FDI model, and we assume that this fixed cost is incurred at $t = 1$. At $t = 2$, the licensee produces the good, generates sale revenue, and uses this revenue to remunerate labor and pay the technology owner the amount L_2 . Ex-post payments that exceed the revenues generated at $t = 2$ are not allowed. In order to generate a nontrivial tradeoff between

⁴³ Strictly speaking, the tradeoff between FDI and exports could still be governed by (8), even when the technology owners do not capture the entire surplus, provided that the choice of FDI versus exporting is included in the ex-ante negotiations and that unrestricted ex-ante transfers between the licensee and the technology owner are allowed. As emphasized below, it is important thus to consider limitations on both ex-ante and ex-post transfers of surplus.

licensing and FDI, suppose that the marginal cost of production faced by the licensee is $1/(\lambda\varphi)$, where $\lambda > 1$.

With no transaction costs, it is clear that the technology owner could easily choose a transfer L_1 , or L_2 , large enough to appropriate all rents, thus making FDI dominated by this licensing strategy. Suppose, however, that due to the nonexcludable nature of the technology to produce variety ω , the technology owner cannot completely prevent the licensee from using the technology in ways that might prove detrimental to the accrual of cash flows at $t = 2$. To fix ideas, suppose that the licensee can operate the technology “on the side,” thereby diverting a share ϕ of revenues. Suppose that this diversion requires a fixed cost equal to f_D associated with setting up production of a competing variety. Then, in order to avoid cash flow diversion (or technology expropriation) on the part of the licensee, the payment L_2 needs to be low enough to satisfy

$$(\lambda\varphi)^{\sigma-1} B^F - L_2 \geq (\sigma\phi - (\sigma - 1)) (\lambda\varphi)^{\sigma-1} B^F - f_D, \quad (28)$$

where the left-hand side is the ex-post payoff of the licensee under no defection, while the right-hand side is its payoff under defection, assuming that the technology owner captures all nondiverted revenue in that case.⁴⁴

If the above was the only source of transaction costs in the model, then a simple strategy for the technology owner would be to set $L_2 = 0$, thus ensuring that (28) is met, and then setting $L_1 = (\lambda\varphi)^{\sigma-1} B^F - f_D$, thus extracting all surplus ex ante. Suppose, however, that because of the nonrival nature of technology, the technology owner cannot credibly commit at $t = 1$ to not using his/her own technology to service the foreign market at $t = 2$ via an alternative method. To be more specific, and given our assumption that all plant fixed costs are incurred at $t = 1$, assume that the only feasible alternative method at that point is exporting. This “defection” can have a significant detrimental effect on the sale revenues obtained by the licensee at $t = 2$, particularly when given the noncodifiable dimension of technology and foreseeing a future defection on his/her part, the technology owner might be inclined to transfer an inferior version of the technology to the licensee. For simplicity, we capture this in a stark way by assuming that, by exporting at $t = 2$, the technology owner would drive licensee revenues to 0.

This limited commitment constraint on the part of the technology owner effectively hinders the ability of the technology owner to extract surplus ex ante rather than ex post. More specifically, for the technology owner not to have an incentive to defect at $t = 2$, the ex-post payment L_2 has to be high enough to guarantee that

$$L_2 \geq \tau^{1-\sigma} \varphi^{\sigma-1} B^F, \quad (29)$$

where the right-hand side corresponds to the ex-post payoff of the technology owner when defecting.

⁴⁴ Equation (28) assumes that in the absence of expropriation, the choice of output maximizes total revenue net of labor costs, which is consistent with our assumptions above.

Note that both constraints (28) and (29) can only be met whenever

$$f_D > [\tau^{1-\sigma} - \sigma(1-\phi)\lambda^{\sigma-1}] \varphi^{\sigma-1} B^F, \quad (30)$$

that is when, for a given residual demand level B^F , transport costs τ are high, the scope of dissipation ϕ is low, the cost advantage λ of the licensee is large, or the defection fixed cost f_D is high. Ethier and Markusen (1996) derive similar comparative statics in a far richer framework and interpret a low f_D as reflecting an intensive use of knowledge (relative to physical) capital in production.

When condition (30) is satisfied, no defection will occur in equilibrium and the licensing option will necessarily dominate FDI at $t = 0$ since L_1 can be set equal to $(\lambda\varphi)^{\sigma-1} B^F - L_2 - f_D$ thus leaving the technology owner with all the net surplus, which exceeds the same net surplus of FDI because of $\lambda > 1$. Conversely, when condition (30) is not satisfied, licensing will be a (weakly) dominated strategy because, anticipating defection on the part of the technology owner, the licensee will anticipate zero revenue ex post, and thus it will not be willing to pay for the technology upfront, leaving the technology owner with at most its payoff under exporting.

So far we have treated the residual demand term B^F as given. In our model without licensing in Section 4.1.1, we showed that the equilibrium would be one with pervasive FDI and no exporting whenever $2f_D/f_E < \tau^{\sigma-1} - 1$. Plugging the equilibrium value $B^F = (f_D + \frac{1}{2}f_E) \varphi^{1-\sigma}$, we find that such an FDI equilibrium will dominate an equilibrium with licensing whenever

$$\frac{2f_D}{f_E} < \frac{\tau^{\sigma-1}}{\tau^{\sigma-1}(1 + \sigma(1-\phi)\lambda^{\sigma-1}) - 1} - 1.$$

Interestingly, a low ratio of plant-specific economies of scale relative to firm-specific economies of scale (f_D/f_E) favors FDI not only over exporting, as in Section 4.1.1, but also over licensing. Conversely, the positive effect of trade costs on the likelihood of FDI is now less clear-cut because high trade costs relax the temptation to defect for the technology owner. Notice that when $\tau^{\sigma-1} \rightarrow \infty$, the above condition cannot possibly hold and the equilibrium is one with licensing. This provides a potential justification for our stylized Fact Three indicating that *levels* of FDI tend to diminish with distance.

Even when $2f_D/f_E > \tau^{\sigma-1} - 1$, so FDI is dominated by exporting, licensing can still emerge in equilibrium when (30) holds and provided that λ is high enough. In this case too, licensing becomes more appealing, the higher are trade frictions τ and the cost advantage λ of licensees.⁴⁵

⁴⁵ The precise condition under which licensing dominates exporting is given by

$$\frac{\lambda^{\sigma-1} - \tau^{1-\sigma}}{(1 + \tau^{1-\sigma})} \left(\frac{f_E}{f_D} + 1 \right) < 1.$$

It is worth outlining some additional results that obtain when incorporating country asymmetries and firm heterogeneity in the model above. For the sake of brevity, we will focus only on the choice between FDI and licensing and will treat factor prices and demand levels as given. Consider first the case of wage differences across countries. Denoting by w^H and w^F , the wage at Home and in Foreign, respectively, condition (30) now becomes

$$f_D > \left[(\tau w^H)^{1-\sigma} - \sigma (1 - \phi) \lambda^{\sigma-1} (w^F)^{1-\sigma} \right] \varphi^{\sigma-1} B^F,$$

which preserves the same comparative statics as before, but notice that the right-hand side is now also decreasing in w^H and increasing in w^F . Intuitively, higher Home wages relax the incentive compatibility constraint for the technology owner by making exporting less profitable, while lower Foreign wages relax the licensee's incentive to misbehave by increasing the profitability of cooperating with the technology owner. An implication of this result is that the larger are wage differences across countries, the more likely that licensing will dominate FDI, a result that very much resonates with those obtained by [Ethier \(1986\)](#) and [Ethier and Markusen \(1996\)](#). It thus follows that in horizontal FDI models, both the location and the internalization decisions work against the prevalence of MNEs when relative factor endowment differences are high.

Next consider the choice between FDI and licensing when technology owners within sectors differ in their productivity levels, while sharing a common residual demand level as captured by B^F . Then direct inspection of (30) reveals that there exists a productivity threshold φ^L , such that all firms with productivity φ above φ^L will prefer FDI to licensing, while all firms with productivity below φ^L will prefer licensing to FDI. This result obtains because we have modeled defection as entailing a fixed cost f_D , and thus the no defection constraint will tend to be more binding for large firms than for small firms. Introducing a fixed cost of exporting f_X paid by the technology owner at $t = 2$ upon defecting would generate the same exact sorting of firms between the licensing and FDI modes and would lead to a higher threshold productivity φ^L .⁴⁶

7.1.2. Outsourcing versus FDI: Hold Up Inefficiencies

We next illustrate another source of transaction costs within the Ricardian, one-factor version of [Helpman's \(1984\)](#) vertical FDI model developed in [Section 5.3](#). There we argued a firm may find vertical FDI appealing when its productivity is high and when the Home wage is sufficiently higher than the foreign wage. Vertical FDI yields a profit

Note also that when considering licensing as a deviation from an all exporting equilibrium, the equilibrium residual demand term B^F is given in (6), but this has no qualitative effect on how parameters affect the likelihood of licensing dominating FDI.

⁴⁶ In fact, the condition becomes identical to (30) except with $f_D + f_X$ on the left-hand side of the inequality.

level (net of entry costs) equal to

$$\pi_I(\varphi) = \varphi^{\sigma-1} (B^H + B^F) \left((w^H)^\eta (\tau w^F)^{1-\eta} \right)^{1-\sigma} - w^H f_I.$$

For the Home firm to actually appropriate this entire profit flow, however, it is important that an initial contract stipulates precisely the quantity and characteristics (quality, compatibility,...) of the inputs produced in the foreign manufacturing plant, while including payments that transfer all the net surplus to the Home firm. Again, these are strong assumptions in internalized vertical relationships but they are simply untenable in arm's-length vertical transactions. The seminal work of [Williamson \(1985\)](#) has demonstrated that contractual gaps or incompleteness (and the associated renegotiation or “fine-tuning” of contracts) create inefficiencies in situations in which the parties involved in a transaction undertake relationship-specific investments or use relationship-specific assets, a realistic characteristic of offshoring relationships. Intuitively, specificity implies that, at the renegotiation stage, parties cannot costlessly switch to alternative trading partners and are partially locked into a bilateral relationship. [Williamson \(1985\)](#) illustrates how this so-called *fundamental transformation* is a natural source of ex-post inefficiencies (e.g., inefficient termination or execution of the contract) as well as ex-ante or hold-up inefficiencies (e.g., suboptimal provision of relationship-specific investments).

We next develop a simple model, along the lines of [Grossman and Helpman \(2002\)](#), that formalizes some of these insights within the vertical fragmentation model developed in [Section 5.3](#).⁴⁷ Consider then the same model as in [Section 5.3](#), but now allow for the possibility that the manufacturing stage of production is outsourced, rather than integrated. The main benefit of outsourcing is that it avoids “governance costs” and thus entails lower costs of production. To be more precise, we assume that marginal costs are lower by a factor $\lambda > 1$ under outsourcing, and that the fixed costs satisfy $f_{DO} < f_{DV}$ and $f_{IO} < f_{IV}$, where the subscripts D and I denote Domestic sourcing and International sourcing, respectively, while the subscripts O and V refer to Outsourcing and Vertical integration, respectively. In [Section 5.3](#), where all transactions were integrated, we used the simpler notation $f_{DV} = f_D$ and $f_{IV} = f_I$.

As in all contracting models, it is important to be explicit about the timing of events and the space of contracts available to agents. The model features two types of agents: headquarters H and operators of manufacturing facilities M . We shall now distinguish between four distinct periods, $t = 0$, $t = 1$, $t = 2$, and $t = 3$.

At $t = 0$, the firm observes its productivity φ , decides whether it wants to have the manufacturing part of production be done at Home or in Foreign, and whether to have it done inhouse or at arm's-length. At this stage too, all fixed costs are incurred and headquarter services h are produced. At $t = 0$, H anticipates that domestic vertical

⁴⁷ Some of the ideas to be formalized were also present in the work of [McLaren \(2000\)](#), but it is harder to map his model to the type of CES frameworks we have been restricting ourselves to in this survey.

integration and foreign vertical integration (or FDI) will be associated with the payoffs derived in Section 5.3 (see equations (22) and (23)), so below we focus on characterizing behavior in stages $t = 1$, $t = 2$, and $t = 3$ for the case of domestic and foreign outsourcing.

Under outsourcing, at $t = 1$, the headquarter H offers an initial (or ex ante) contract to a potential manufacturer for the provision of the input m . The simplest way to illustrate the effects of contractual frictions on outsourcing is to assume that the ex-ante contract is *totally incomplete* in the sense that no aspect of production can be specified in the contract. This is obviously a strong assumption, but it can easily be relaxed. The important thing is that the initial contract cannot fully discipline the behavior of agents during the production stage. The only contractible in the initial contract is a lump-sum transfer between the two agents.

At $t = 2$, M undertakes the investments necessary to produce the manufacturing input m . Because these investments are not disciplined by the initial contract, they will necessarily be set to maximize M 's private (ex-post) return from them.

Once the production of m is finalized, H and M sit down at $t = 3$ to (re-)negotiate a transaction price for the manufacturing services provided by M to H . Sale revenue is generated immediately (or shortly) after an agreement between H and M has been reached, and the receipts are divided between the agents according to the negotiated agreement. As is standard in the literature, we characterize this ex-post bargaining, using the symmetric Nash Bargaining solution and assume symmetric information between H and M at this stage. This leaves H and M with their outside options plus a share of the ex-post gains from trade (i.e., the difference between the sum of the agent's payoffs under trade and their sum under no trade). As pointed out by Williamson (1985), the fact that the division of surplus is determined ex post (after investments have been incurred) rather than ex ante becomes significant only when investments are (partly) relationship specific, so that their value inside the relationship is higher than outside of it.

In the global sourcing environments that concern us here, there are two natural sources of "lock in" for manufacturers. First, manufacturing inputs are often customized to their intended buyers and cannot easily be resold at full price to alternative buyers; and second, even in the absence of customization, search frictions typically make separations costly for both H and M . A simple way to capture these considerations in the model above is to assume that if the contractual relationship between the two parties breaks down, (i) H does not have time to turn to an alternative M for the provision of m (and is thus left with nothing), and (ii) M can resort to a secondary market but there he/she would only be able to secure a payoff equal to a share $1 - \phi < 1$ of the revenue generated when combining m with H 's headquarter services. The higher is ϕ , the higher the degree of specificity of M 's investments.

Solving for the subgame perfect equilibrium of this game under the assumption that arm's-length transactions are "totally incomplete" regardless of the location of M , we have that the profits obtained by H under domestic outsourcing and foreign (or offshore)

outsourcing are given by

$$\pi_{DO}(\varphi) = \varphi^{\sigma-1} (B^H + B^F) (w^H)^{1-\sigma} \lambda^{\sigma-1} \Gamma_O - w^H f_{DO}, \quad (31)$$

and

$$\pi_{IO}(\varphi) = \varphi^{\sigma-1} (B^H + B^F) \left((w^H)^\eta (\tau w^F)^{1-\eta} \right)^{1-\sigma} \lambda^{\sigma-1} \Gamma_O - w^H f_{IO}, \quad (32)$$

respectively, where

$$\Gamma_O = \left(\frac{\sigma - (\sigma - 1) (1 - \frac{1}{2}\phi) (1 - \eta)}{\sigma - (\sigma - 1) (1 - \eta)} \right)^{\sigma - (\sigma - 1)(1 - \eta)} \left(1 - \frac{1}{2}\phi \right)^{(1 - \eta)(\sigma - 1)} < 1. \quad (33)$$

These expressions are analogous to the insourcing ones in (22) and (23), except for the lower fixed costs, the terms with $\lambda > 1$ capturing the (governance) cost advantage of outsourcing, and the term $\Gamma_O < 1$, which reflects the transaction costs due to incomplete contracting associated with market transactions. Straightforward differentiation indicates that Γ_O is decreasing in the degree of specificity ϕ and reaches a value of 1 when ϕ goes to 0 and there is no specificity. Consequently, outsourcing will tend to dominate integration for low levels of specificity and large values of λ , while integration will instead tend to dominate outsourcing for high values of specificity ϕ and for λ close enough to 1.

Furthermore, cumbersome differentiation (see the Online Appendix) demonstrates that Γ_O in (33) is a strictly increasing function of η , and thus the transaction costs of using the market are particularly high when the input m is relatively important in production. The result is intuitive given that the source of transaction costs is the underinvestment in m , and it suggests a higher relative profitability of outsourcing in headquarter intensive sectors.

So far we have focused on studying how the different parameters of the model shape the profitability of a firm when using alternative organizational models, while treating the equilibrium demand levels B^H and B^F as given. The simplest way to close the model is by assuming that all Home firms have the same productivity level φ and that free entry brings profits down to zero. We will introduce intraindustry heterogeneity in productivity in the next section on property-rights models. With homogeneity, in any generic equilibrium all firms choose the same organizational form. Furthermore, for w^H/w^F sufficiently high, the equilibrium is necessarily one with all firms offshoring in Foreign, with these firms optimally choosing outsourcing whenever

$$\frac{f_{IV}}{f_{IO}} \lambda^{\sigma-1} \Gamma_O > 1,$$

while choosing vertical FDI whenever this inequality is reversed. From our previous discussion it thus follows that the likelihood of *internalized* fragmentation is increasing in

the degree of specificity ϕ , and decreasing in the governance costs λ and headquarter intensity η . Introducing intraindustry heterogeneity generates analogous comparative static results regarding the share of offshoring firms and activity associated with vertical FDI, except for a subtle counterbalancing force of headquarter intensity that will be discussed in the next section on property-rights models (see footnote 55).

Although we have drawn inspiration from the work of [Grossman and Helpman \(2002\)](#) in developing the model above, it is important to point out some interesting features of their framework that we have left out.⁴⁸ Most notably, key to their analysis are search frictions, which we ruled out above by assuming that when a firm chooses outsourcing instead of integration, it can simply post a contract and pick an operator M from the set of firms applying to fulfill the order. [Grossman and Helpman \(2002\)](#) instead assume that the matching between stand-alone H agents and M operators is random and depends on the relative mass of each type of agent looking for matches. Furthermore, once a match is formed, agents are not allowed to exchange transfers prior to production. Other things equal, it is clear that these features will tend to reduce the attractiveness of outsourcing vis-à-vis integration, even holding constant supplier investments. Intuitively, search frictions and the lack of transfers inhibit the ability of H producers to fully capture the rents generated in production relative to a situation in which they can make take-it-or-leave-it offers to a perfectly elastic supply of M operators. Search frictions can generate much more subtle and interesting results when allowing the matching function governing the search process to feature increasing returns to scale. In such a case, [Grossman and Helpman \(2002\)](#) show that there may exist multiple equilibria with different organizational forms (or industry systems) applying in ex-ante identical countries or industries. Furthermore, the likelihood of an equilibrium with outsourcing is enhanced by an expansion in the market size, which increases the efficiency of matching in the presence of increasing returns to scale in the matching function.

A previous paper by [McLaren \(2000\)](#) also provided an alternative framework in which the organizational decisions of firms within an industry exerted externalities on the decisions of other firms in the industry. Rather than assuming a search/congestion externality, [McLaren \(2000\)](#) focuses on the implications of market thickness on the ex-post determination of the division of surplus between H and M agents. In his framework, the thicker is the market for inputs, the larger is the ex-post payoff obtained by M producers which in turn alleviates hold-up inefficiencies. Crucially, however, the thickness of the market for inputs depends in turn on the extent to which firms rely on outsourcing rather than integration, since only outsourcers enter that market. [McLaren \(2000\)](#) demonstrates too

⁴⁸ In certain respects, the framework above is much richer than that in [Grossman and Helpman \(2002\)](#), since they focused on a closed-economy model with no producer heterogeneity and with no headquarter investments, or $\eta = 0$ in the model above. The [Grossman and Helpman \(2002\)](#) setup was later extended by the authors to open-economy environments with cross-country variation in the degree of contractibility of inputs in [Grossman and Helpman \(2003\)](#) and [Grossman and Helpman, 2005](#).

the possibility of multiple equilibria and shows that trade opening, by thickening the market for inputs, may lead to a worldwide move toward more disintegrated industrial systems, thus increasing world welfare and leading to gains from trade quite different from those emphasized in traditional trade theory.⁴⁹

7.2. The Property-Rights Approach

The transaction-cost theory of firm boundaries has enhanced our understanding of the sources and nature of inefficiencies that arise when transacting via the market mechanism, but it sheds little light on the limits or costs of vertically or laterally integrated transactions. Transaction-cost models appeal to some vague notion of “governance costs” to deliver a nontrivial tradeoff in internalization decisions, but these governance costs are treated as exogenous parameters and thus orthogonal to the sources of transaction costs in market transactions. The property-rights theory of the firm, as first expounded in [Grossman and Hart \(1986\)](#), and further developed in [Hart and Moore \(1990\)](#) and [Hart \(1995\)](#), has convincingly argued that this approach is unsatisfactory. After all, intrafirm transactions are not secured by all-encompassing contracts and there is no reason to assume that relationship specificity will be any lower in integrated relationships than in nonintegrated ones. For these reasons, opportunistic behavior and incentive provision are arguably just as important in within-firm transactions as they are in market transactions.

If one accepts the notion that within-firm transactions typically entail transaction costs and that the source of these transaction costs is not too distinct from those in market transactions, then a natural question is: what defines then the boundaries of the firm? From a legal perspective, integration is associated with the ownership (via acquisition or creation) of nonhuman assets, such as machines, buildings, inventories, patents, copyrights, etc. The central idea of the property-rights approach of Grossman, Hart, and Moore is that internalization matters because ownership of assets is a source of power when contracts are incomplete. More specifically, when parties encounter contingencies that were not foreseen in an initial contract, the owner of these assets naturally holds *residual* rights of control, and he or she can decide on the use of these assets that maximizes his/her payoff at the possible expense of that of the integrated party. [Grossman and Hart \(1986\)](#) then show that in the presence of relationship-specific investments, these considerations lead to a theory of the boundaries of the firm in which both the benefits and the costs of integration are endogenous. In particular, vertical integration entails endogenous (transactions) costs because it reduces the incentives of the integrated firm to make investments that are partially specific to the integrating firm, and that this underinvestment lowers the overall surplus of the relationship.

⁴⁹ Another aspect studied by both the [McLaren \(2000\)](#) and [Grossman and Helpman \(2002\)](#) papers but ignored in our discussion above is the choice by M producers of the degree to which they customize their intermediate products to their intended buyers (see also [Qiu and Spencer, 2002](#), and [Chen and Feenstra, 2008](#)).

The property-rights theory of the firm has featured prominently in the international trade literature in recent years, starting with the work of [Antràs \(2003\)](#). The vast majority of applications of the property-rights approach have focused on the type of vertical integration decisions inherently associated with vertical FDI models, rather than with lateral integration decisions of the type emphasized in the FDI versus licensing literature (an exception is [Chen et al., 2012](#)).

Let us thus go back to our Ricardian, one-factor, vertical FDI model first explicated in [Section 5.3](#). As in [Section 7.1.2](#), we maintain the assumption that when transacting at arm's-length, only “totally incomplete” contracts are available to agents. The key innovation in this section is that integrated transactions also entail transaction costs, and following [Grossman and Hart \(1986\)](#) and [Hart and Moore \(1990\)](#), the source of these costs is related to the fact that intrafirm transactions are also governed by totally incomplete contracts.⁵⁰ In particular, we shall assume that when H decides its mode of organization at $t = 0$, it anticipates playing an analogous game with a manufacturing operator M regardless of whether the operator is an employee of H or an independent contractor. Both the “outsourcing” and “integration” branches of the game feature an ex-ante contracting stage ($t = 1$), an investment stage ($t = 2$), and an ex-post bargaining stage ($t = 3$). The only difference between the two branches of the game is at $t = 3$, and more precisely, in the outside options available to H and M at this stage. Remember that in the outsourcing stage we assumed that in the absence of an agreement at $t = 3$, H was left with a zero payoff (since it could not create output without m and there was no time to find an alternative M for the provision of m), while M could sell the input m in a secondary market and obtain a share $1 - \phi < 1$ of the revenue generated when combining m with H 's headquarter services. In the case of integration, the above formulation of the outside options is unrealistic. It seems natural to assume instead that H will hold property rights over the input m produced by M , and thus H has the ability to fire a stubborn operator M that has refused to agree on a transfer price (leaving M with nothing), while still being able to capture part, say a fraction $\delta < 1$, of the revenue generated by combining h and m . The fact that $\delta < 1$ reflects the intuitive idea that F cannot use the input m as effectively as it can with the cooperation of its producer, i.e., M .

In the ex-post bargaining, each party will capture their outside option plus an equal share of the ex-post gains from trade. Denote by β_k the share of revenue accruing to H at $t = 3$ under organizational form $k = V, O$. Given our assumptions, we have

$$\beta_V \equiv \frac{1}{2}(1 + \delta) > \frac{1}{2}\phi \equiv \beta_O,$$

which illustrates the notion that H holds more *power* under integration than under outsourcing.

⁵⁰ As discussed below, our framework could easily accommodate variation in contractibility across organizational forms but we will refrain from doing so in the spirit of the property-rights approach.

Because we have maintained our assumptions regarding the outsourcing branches of the game, the payoffs associated with domestic and offshore outsourcing are still given by [equations \(31\) and \(32\)](#), with Γ_O given in [\(33\)](#). Following analogous steps, we can now solve for H 's profits under domestic and offshore integration in the presence of within-firm transactions costs. These are given by

$$\pi_{DV}(\varphi) = \varphi^{\sigma-1} (B^H + B^F) (w^H)^{1-\sigma} \Gamma_V - w^H f_{DV} \text{ and} \quad (34)$$

$$\pi_{IV}(\varphi) = \varphi^{\sigma-1} (B^H + B^F) \left((w^H)^\eta (\tau w^F)^{1-\eta} \right)^{1-\sigma} \Gamma_V - w^H f_{IV}, \quad (35)$$

respectively, with

$$\Gamma_V = \left(\frac{\sigma - (\sigma - 1) \frac{1}{2} (1 - \delta) (1 - \eta)}{\sigma - (\sigma - 1) (1 - \eta)} \right)^{\sigma - (\sigma - 1)(1 - \eta)} \left(\frac{1}{2} (1 - \delta) \right)^{(1 - \eta)(\sigma - 1)} < 1.$$

These expressions are very similar to the outsourcing ones, except for the cost differences in fixed and marginal costs (due to factor prices and exogenous “governance costs”) and for the different levels of transaction costs, as captured by the difference between Γ_V and Γ_O . A key object in governing the relative profitability of integration and outsourcing is the ratio Γ_V / Γ_O , which we can reduce to:

$$\frac{\Gamma_O}{\Gamma_V} = \left(\frac{\sigma - (\sigma - 1) (1 - \beta_O) (1 - \eta)}{\sigma - (\sigma - 1) (1 - \beta_V) (1 - \eta)} \right)^{\sigma - (\sigma - 1)(1 - \eta)} \left(\frac{1 - \beta_O}{1 - \beta_V} \right)^{(1 - \eta)(\sigma - 1)}. \quad (36)$$

Differentiating this expression with respect to η , we find that for any $\beta_V > \beta_O$, with $\beta_k \in (0, 1)$ for $k = V, O$, this ratio is necessarily decreasing in η (see the Online Appendix). Hence, unlike in the transaction-cost model in [Section 7.1.2](#), low levels of headquarter intensity are now associated with higher (rather than lower) relative profitabilities of outsourcing versus integration. This result resonates with one of the central results in the property-rights theory: with incomplete contracting, ownership rights of assets should be allocated to parties undertaking noncontractible investments that contribute disproportionately to the value of the relationship. The relative importance of the operator M 's investment is captured in [\(36\)](#) by the elasticity of output with respect to that agent's investment, i.e., $1 - \eta$, and thus the lower is η , the higher the need for H to give away ownership rights to M by engaging in outsourcing. As argued below, the fact that headquarter intensity shapes the integration decision in opposite ways in transaction-cost and property-rights models opens the door for empirical relative evaluations of the two models.

So far we have only discussed how η affects the relative profitability of integration and outsourcing. A problematic feature of the stylized property-rights model we have developed so far is that because M is the only agent undertaking noncontractible,

relationship-specific investments, transaction costs stemming from underinvestment will always be higher under integration than under outsourcing. In terms of our notation, we necessarily have $\Gamma_O > \Gamma_V$. Coupled with the presence of exogenous “governance costs” the model thus predicts that vertical integration of any sort will be a dominated strategy for H . Intuitively, M ’s underinvestment is minimized by conceding ownership rights to him or her, while H can still capture all the net surplus from production via ex-ante transfers in the initial contract.

In order to generate a nontrivial tradeoff between integration and outsourcing, the literature has followed one of two approaches. A first one consists in allowing some of the investments in headquarter services carried out by H to also be noncontractible. [Antràs \(2003\)](#) and [Antràs and Helpman \(2004\)](#) consider scenarios in which investments in h are completely noncontractible and are carried out at $t = 2$, simultaneously and noncooperatively with the investment in m by the manufacturing plant. [Antràs and Helpman \(2008\)](#) consider a more general framework with partial contractibility of both h and m and show that the ratio Γ_O/Γ_V is necessarily increasing in the relative importance of the *noncontractible* manufacturing investments, and decreasing in the importance of the *noncontractible* headquarter investments. It is then clear that higher headquarter intensity η continues to affect negatively the relative profitability of outsourcing. Importantly, however, for a high enough η , it now becomes possible for transaction costs to be higher under integration than under outsourcing. The reason for which integration is no longer a dominated strategy for H is that incomplete contracting is now generating a *double-sided* holdup problem and thus it is no longer always optimal to allocate as much ex-post bargaining power as possible to M . Even though H can extract rents upfront, it still needs to make sure that it will have high-powered incentives to invest at $t = 2$, when those initial transfers are bygone, and vertical integration provides a way to generate those high-powered incentives.

A second way to generate a nontrivial tradeoff between integration and outsourcing is to introduce restrictions on the ability of the firm to extract rents from M in the ex-ante contract. The literature often motivates these ex-ante constraints by appealing to financial constraints (see [Acemoglu et al., 2007](#); [Basco, 2010](#); [Carluccio and Fally, 2012](#); [Antràs and Chor, 2012](#)), but other interpretations are possible.⁵¹ It is intuitively clear that even when headquarter services are fully contractible, H will not necessarily want to maximize the ex-post bargaining power of M if by doing so it reduces the share of surplus it will end up with. Even though “financial constraints” affect the relative transactions costs of integration and outsourcing it can again be shown that the ratio Γ_O/Γ_V continues to be increasing in the relative importance of noncontractible manufacturing investments and decreasing in the relative importance of noncontractible headquarter services (see [Antràs, forthcoming](#)).

⁵¹ See also [Conconi et al. \(2012\)](#) and [Alfaro et al. \(2010\)](#).

So far we have focused on studying how the different parameters of the model shape the decision of a firm of how to optimally organize production. In practice, and as discussed in more detail below in [Section 7.3](#), these firm decisions are rarely observed by researchers, who instead often work with product-level datasets that aggregate a subset of these firm decisions into a particular observation. With that in mind, we next use the model to integrate over these firm decisions and characterize more formally the relative prevalence of different organizational forms within a sector or industry. The analysis is analogous to that in the vertical FDI model in [Section 5.3](#), although matters are now a bit more involved because H is choosing from four possible organizational modes: domestic outsourcing, domestic integration, foreign outsourcing, and foreign integration.

A first obvious observation, however, is that any form of vertical integration is necessarily a dominated strategy whenever $\lambda^{\sigma-1}\Gamma_O > \Gamma_V$, given the implied efficiency advantage of outsourcing in that case and given the lower fixed costs associated with outsourcing relative to integration. In other words, equilibria with multinational activity are only possible whenever Γ_O is sufficiently low relative to Γ_V , which from our previous discussion requires high levels of headquarter intensity η . When integration is a dominated strategy, the sorting of firms into organizational forms is analogous to that in [Figure 2.8](#) in [Section 5.3](#), but with the most productive firms engaging in foreign outsourcing (rather than vertical FDI) and the least productive firms (among the active ones) relying on domestic outsourcing.

When $\lambda^{\sigma-1}\Gamma_O < \Gamma_V$, much richer sorting patterns can emerge. In particular, the effective marginal cost is now lower under integration than under outsourcing, but outsourcing continues to be a strategy associated with lower fixed costs, and thus a subgroup of relatively unproductive firms might continue to prefer outsourcing to integration. For certain parameter configurations, one can construct an industry equilibrium in which all four organizational forms coexist in equilibrium, as depicted in [Figure 2.10](#). Firms with productivity $\varphi^{\sigma-1}$ below φ^D do not produce, those with $\varphi^{\sigma-1} \in (\varphi^D, \varphi^{DV})$ outsource domestically, those with $\varphi^{\sigma-1} \in (\varphi^{DV}, \varphi^I)$ integrate domestically, those with $\varphi^{\sigma-1} \in (\varphi^I, \varphi^{IV})$ outsource abroad, and those with $\varphi^{\sigma-1} > \varphi^{IV}$ integrate abroad, i.e., they engage in foreign direct investment. Naturally, for other configurations of parameter values, we can eliminate one or more of these regimes in equilibrium, but their ranking by productivity will not be affected.

Assuming a Pareto distribution of firm productivity with unbounded support ($G(\varphi) = 1 - \underline{\varphi}^\kappa \varphi^{-\kappa}$ for $\varphi > \underline{\varphi}$), one can explicitly solve for the relative prevalence of different organizational modes within this equilibrium and carry out various comparative statics exercises. To save space, let us focus on the implications of the model for the relative prevalence of vertical FDI versus offshore outsourcing, since this is an object that has featured prominently in the empirical literature (as discussed in [Section 7.3](#) below). There are many ways to measure the relative extent of vertical FDI and offshore outsourcing, but it is simplest to consider the share of offshoring firms (those with $\varphi \geq \varphi^I$) that

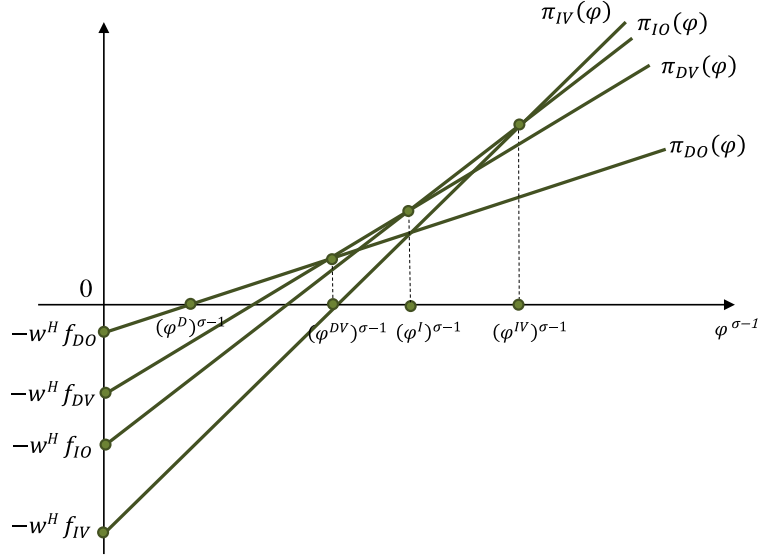


Figure 2.10 Sorting into Organizational Modes

vertically integrate. Analogous results apply when computing, for instance, the share of imports of m that are transacted within multinational firm boundaries, although these formulas become more complicated (see the Online Appendix). Manipulating (32), (34), and (35) to solve for the thresholds φ^I and φ^{IV} , and plugging $G(\varphi)$ we find that the share of offshoring firms that integrate suppliers is given by

$$\frac{N_{IV}}{N_{IV} + N_{IO}} = \frac{1 - G(\varphi^{IV})}{1 - G(\varphi^I)} = \left(\frac{f_{IO} - f_{DV}}{f_{IV} - f_{IO}} \frac{1 - \lambda^{\sigma-1} \Gamma_O / \Gamma_V}{\lambda^{\sigma-1} \Gamma_O / \Gamma_V - \left(\frac{w^H}{\tau w^F} \right)^{-(1-\eta)(\sigma-1)}} \right)^{\frac{\kappa}{\sigma-1} - 1}. \quad (37)$$

Some interesting comparative statics follow from this expression. First, and rather obviously, the prevalence of vertical FDI versus offshore outsourcing is decreasing in λ and Γ_O / Γ_V , because high values of these make arm's-length transacting more appealing. Second, the share $N_{IV} / (N_{IV} + N_{IO})$ is increasing in fragmentation barriers ($f_{IO} - f_{DV}$ and τ) and decreasing in the wage gap (w^H / w^F). This result is more subtle because these parameters do not affect directly the relative marginal efficiency of vertical FDI and foreign outsourcing (compare equations (32) and (35)). Intuitively, an increase in trade barriers or a decrease in the wage gap makes offshoring of any form relatively less productive, and will thus generate an *extensive margin* of trade response by which some firms will switch from foreign to domestic sourcing. Inspection of Figure 2.10 reveals, however, that those switchers are necessarily engaged in foreign outsourcing, thus generating an increase in the relative prevalence of FDI within the set of offshoring

firms. Third, and unlike in the transaction-cost model above, the relative prevalence of vertical FDI is increasing in headquarter intensity η . This is for two reasons: because a low headquarter intensity shifts market share from domestic integrating firms to offshore outsourcing firms (the extensive margin of trade effect), and because it also shifts market share from FDI firms to offshore outsourcers (since Γ_O/Γ_V decreases in η). A fourth unambiguous comparative static relates to a positive effect of dispersion (a low κ) on the relative prevalence of integration: the intuition here is analogous to that in [Helpman et al. \(2004\)](#), which was discussed in [Section 4.2](#).

Other contributions to the property-rights approach have suggested and formalized alternative determinants of the organizational form decisions of firms and of the relative prevalence of vertical FDI versus offshore outsourcing. [Antràs and Helpman \(2008\)](#) emphasize the possibility that improvements in the contractibility of manufacturing might increase the relative prevalence of FDI rather than of offshore outsourcing, contrary to what transaction-cost models would predict. [Acemoglu et al. \(2007\)](#) consider an environment in which the headquarters H contract with various suppliers and identify a positive effect of the degree of technological complementarity across inputs on the internalization decision of multinational firms. [Antràs and Chor \(2012\)](#) also consider a multisupplier environment and show that when the production process is sequential in nature, H might have differential incentives to integrate suppliers along the value chain, and thus the “downstreamness” of an input becomes a determinant of the ownership structure decisions related to that input. [Basco \(2010\)](#) and [Carluccio and Fally \(2012\)](#) find that multinationals are more likely to integrate suppliers located in countries with poor financial institutions, and that the effect of financial development should be especially large when trade involves complex goods. The insights of the property-rights theory have also been applied to dynamic, general-equilibrium models of international trade with the goal of understanding how ownership decisions vary along the life-cycle of a product or input (see [Antràs, 2005](#)).⁵²

7.3. Empirical Evidence

Empirically testing models of the various determinants of the internalization decision of multinational firms poses at least two important challenges. First, data on the internalization decisions of multinational firms are not readily available. The existing datasets on the operations of multinational firms generally contain only limited information on their market transactions, and even less is known about arm’s-length exchanges that do not involve multinational firms on either side. A second concern in testing theories of internalization is that the predictions from these theories are associated with subtle features of the environment—such as rent dissipation, relationship-specificity, contractibility, or the

⁵² The transaction-cost model has also been applied to dynamic endogenous-growth models of the FDI versus licensing choice. See [Glass and Saggi \(2002\)](#) for a particular example and [Saggi \(2002\)](#) for a broader review of this literature.

relative intensity of distinct noncontractible, and relationship-specific investments—that are generally unobservable in the data (see [Whinston, 2003](#)).

The empirical literature on multinational firm boundaries has circumvented the first limitation in two ways. A first approach, featuring prominently in the international business literature, consists in employing unique datasets containing a small sample of internalization decisions of multinational firms in certain industries and countries.⁵³ A second empirical approach to the study of multinational firm boundaries relies on indirect inference based on official import and export merchandise trade statistics, which in some countries identify whether transactions involve related or nonrelated parties. Although most applications of this second approach rely on product- and country-level tests, some recent contributions have made use of fairly representative firm-level datasets that contain detailed information on the sourcing strategies of firms in different countries. By their nature, the data are better suited to testing models of the FDI versus outsourcing decision underlying vertical models of multinational firms, than to testing “horizontal” models of multinational firm boundaries.

Some of the key contributions to this second approach (see [Nunn and Trefler, 2008, 2013](#); [Bernard et al., 2010](#); [Antràs and Chor, 2012](#)) have employed the “U.S. Related Party Trade” data collected by the U.S. Bureau of Customs and Border Protection, and have more specifically studied the determinants of the variation in the share of intrafirm imports in total U.S. imports across products and exporting countries. Underlying these tests is the notion that if goods in a particular product category originating from a particular country tend to be exported to the United States within firm boundaries, then one can infer that U.S. firms in that sector tend to find it optimal to internalize purchases of goods from those countries, or alternatively that firms in the exporting country choose to internalize their sales of those goods to the United States. These tests are well grounded in theory, since the new vintage models of the internalization decision featuring intraindustry heterogeneity deliver predictions about precisely the determinants of the cross-sectional variation in the share of internalized transactions within a sector, as illustrated above in [Section 7.2](#).

Even after accepting that the share of intrafirm imports provides a valid empirical proxy for the relative prevalence of integration and outsourcing in vertical models of internalization, there still remains the issue of how to proxy for the various determinants of this share, as identified by these models. This is the second main challenge facing this line of empirical work, as mentioned above. For instance, one of the central results of the property-rights theory is that the prevalence of integration should be higher in headquarter intensive sectors, that is, in sectors where noncontractible, relationship-specific investments carried out by headquarters are disproportionately more important than

⁵³ For instance, [Davidson and McFetridge \(1984\)](#) studied 1,376 internal and arm’s-length transactions involving high-technology products carried out by 32 U.S.-based multinational enterprises between 1945 and 1975. See also [Mansfield et al. \(1979\)](#), [Mansfield and Romeo \(1980\)](#), and [Kogut and Zander \(1993\)](#) for related contributions.

those undertaken by suppliers. A key question is then: how does one measure headquarter intensity in the data? Similar difficulties arise when considering the measurement of parameters governing the degree of relationship-specificity and of contractibility, which were shown above to crucially affect the incentives to integrate suppliers, and in some cases, differently so in transaction-cost and property-rights model.

Following [Antràs \(2003\)](#), a widely used proxy of headquarter intensity is a measure of capital intensity, such as the ratio of physical capital to employment. This approach is justified within Antràs's (2003) stylized model, which features two factors of production, physical capital and labor, and where it is assumed that production of headquarter services is more capital intensive than that of manufacturing, with all investments in his framework being noncontractible and fully relationship-specific. As pointed out by [Nunn and Trefler \(2013\)](#), the latter assumption is unrealistic, since standard measures of capital intensity instead embody several investments that are fairly easy to contract on or that are not particularly relationship-specific. More specifically, one would expect investments in specialized equipment to be much more relevant for the integration decision than investments in structures or in nonspecialized equipment (such as automobiles or computers), which tend to lose little value when used outside the intended production process. Measures of R&D intensity, such as the ratio of R&D expenditures over sales, have also been suggested as appropriate proxies for headquarter intensity (see for instance, [Yeaple, 2006](#), or [Nunn and Trefler, 2013](#)), given that these expenditures tend to be carried out by headquarters (see our Fact Five in [Section 2](#)) and given that they are hard to contract on. Finally, and for analogous reasons, skill intensity (the ratio of nonproduction worker to production worker employment) is sometimes posited as an alternative proxy for headquarter intensity.

The first three columns of [Table 2.5](#) report the effect of these proxies for headquarter intensity on the share of U.S. intrafirm imports. The trade dataset is disaggregated at the six-digit NAICS level and covers the period 2000–2011. Column 1 reports a positive and highly statistically significant correlation of the intrafirm trade share with R&D intensity and physical capital intensity.⁵⁴ Column 2 demonstrates that the physical capital intensity effect is explained by spending on equipment and not structures, while column 3 further disaggregates capital expenditures and shows that the effect of capital equipment is not driven by expenditures on computers and data processing equipment or on automobiles and trucks, consistent with the findings of [Nunn and Trefler \(2013\)](#). In column 4, we further exploit the available variation in the intrafirm import share across source countries and run the same specification as in column 3 but with country-year fixed effects.

Put together, the results in columns 1 through 4 of [Table 2.5](#) demonstrate the existence of a significant and robust positive relationship between the importance of

⁵⁴ See the Online Appendix for details on the variables used. We add 0.001 to R&D intensity before taking logarithms to avoid throwing away a large number of observations with zero R&D outlays.

Table 2.5 The Determinants of the U.S. Intrafirm Import Share

Dep. Var.: $\frac{\text{Intrafirm Imports}}{\text{Total Imports}}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (R&D/Sales + 0.001)	0.33** [0.05]	0.32** [0.05]	0.25** [0.06]	0.09** [0.02]	0.08** [0.02]	0.09** [0.02]	0.07** [0.02]	0.06* [0.03]
Log (Skilled/Unskilled)	0.10 [0.07]	0.16* [0.07]	0.14 [0.11]	0.07* [0.029]	0.07* [0.029]	0.01 [0.02]	0.03 [0.03]	0.03 [0.03]
Log (Capital/Labor)	0.28** [0.05]							
Log (Buildings/L)		-0.23** [0.10]	-0.11 [0.09]	-0.14** [0.03]	-0.14** [0.03]	-0.08* [0.03]	-0.10** [0.04]	-0.05 [0.04]
Log (Equipment/L)		0.46** [0.09]						
Log (Autos/L)			-0.24** [0.06]	-0.08** [0.02]	-0.07** [0.02]	-0.05** [0.01]	-0.07** [0.02]	-0.07** [0.02]
Log (Computer/L)			0.12 [0.10]	0.05 [0.03]	0.03 [0.03]	0.06* [0.03]	0.05 [0.03]	0.02 [0.04]
Log (Other Eq./L)			0.39** [0.09]	0.14** [0.03]	0.17** [0.03]	0.13** [0.03]	0.17** [0.04]	0.13** [0.04]
Seller Contractibility					-0.05* [0.02]	0.05** [0.03]	0.05** [0.03]	0.05 [0.03]
Buyer Contractibility						-0.11** [0.02]	-0.09** [0.03]	-0.12** [0.03]
Buyer Productivity Dispersion								0.03 [0.03]
Freight Costs								-0.05** [0.01]
Tariffs								-0.02** [0.000]
Headquarter controls	Seller	Seller	Seller	Seller	Seller	Buyer	Buyer	Buyer
Restricted Sample	No	No	No	No	No	No	Yes	Yes
Observations	2,888	2,888	2,888	214,694	214,694	227,829	85,691	55,161
R-Squared	0.28	0.30	0.35	0.18	0.18	0.18	0.16	0.20

Standard errors clustered at the industry level are in brackets.

*Significant at 5%

**At 1%.

Columns 1–3 include year fixed effects. Columns 4–8 include country-year fixed effects.

relatively noncontractible investments by headquarters and the prevalence of within-firm import transactions. The literature has generally interpreted this finding as an empirical validation of the property-rights model in [Section 7.2](#), but one should be cautious in interpreting these results since these patterns are not necessarily inconsistent with alternative theories of firm boundaries, such as the transaction-cost model in [Section 7.1.2](#).⁵⁵ Similarly, the significance of R&D or skill intensity for the integration decision of multinational firms could be viewed as a validation of transaction-cost theories that emphasize the importance of the nonexcludable nature of knowledge in shaping multinational firm boundaries, with intrafirm trade being simply a manifestation of complex FDI strategies.

A particularly promising way to discriminate between the property-rights theory of the multinational firm and alternative theories of firm boundaries consists in exploiting the implications of the theory for the effect of contractibility on the share of intrafirm trade. In the property-rights model, the effect of contractibility on the prevalence of integration depends crucially on the degree to which contractual incompleteness stems from noncontractibilities in the inputs controlled by the final-good producer or by his or her suppliers. Conversely, in transaction-cost models, any type of improvement in contractibility would be associated with a lower need to integrate and, in industry equilibrium, with a lower share of intrafirm imports.

Column 5 of [Table 2.5](#) indicates that adding this standard measure of contractibility (in particular, one minus the [Nunn \(2007\)](#) measure) to the previous empirical model suggests a negative and statistically significant effect of contractibility on the prevalence of intrafirm trade. Similar results are reported by [Bernard et al. \(2010\)](#) using an alternative measure of contractibility based on the idea that contracting is likely to be easier for products passing through intermediaries such as wholesalers. The observed negative correlation between the share of intrafirm trade and contractibility is intuitively in line with what one would expect from transaction-cost models of firm boundaries.

A caveat of the previous results is that all the industry controls are constructed using data related to the *selling* industry, i.e., of the good or sector being imported into the United States. In assessing the effects of headquarter intensity this may be problematic when the headquarters are based in the U.S. and import intermediate inputs under an industry classification different from their main line of business. In such a case, a more appropriate approach is to construct measures of headquarter intensity of the *buying* industry. Unfortunately, the U.S. Census Related Party data, and publicly available trade statistics more generally, do not contain information on the industry

⁵⁵ In [Section 7.1.2](#), we emphasized that a higher value of η reduces the level of transaction costs (i.e., increases Γ_O) and thus makes FDI less appealing relative to foreign outsourcing. But with intraindustry heterogeneity, the extensive margin effect illustrated in [Figure 2.10](#) would apply also in a transaction-cost model and a higher η would shrink the set of firms outsourcing abroad, thereby increasing the relative prevalence of FDI on this account. The balance of these two effects is generally ambiguous.

classification of the importer. [Antràs and Chor \(2012\)](#) compute instead measures of headquarter intensity of the *average buying* industry using interindustry flow data from the U.S. input-output data. The exact same approach can also be used to construct measures of the (Nunn) contractibility of the average buyer of U.S. imports of particular goods. Column 6 of [Table 2.5](#) reports the effects of buyer headquarter intensity as well as buyer and seller contractibility on the share of intrafirm imports. Comparing the results in columns 5 and 6, it is clear that the distinction between buyer versus seller headquarter intensity has little effect on the estimates—only the effect of skill intensity is significantly affected—while the distinction between buyer and seller contractibility turns out to matter much more. In particular, the coefficient estimates imply a negative and statistically significant effect of buyer contractibility and a positive and significant effect of seller contractibility. These results very much resonate with the subtle predictions of the property-rights model of [Antràs and Helpman \(2008\)](#), while they are inconsistent with the unambiguously negative effects of contractibility on integration predicted by transaction-cost models.

Our discussion above regarding the effects of buyer headquarter intensity and buyer versus seller contractibility relied on an interpretation of U.S. intrafirm imports as being associated with U.S. headquarters importing goods from foreign suppliers, integrated or not. An important share of these imports, however, consist of shipments from foreign headquarters to their U.S. affiliates or to U.S. unaffiliated parties, in which case the results in column 6 are harder to interpret. For these reasons, in column 7 we follow [Nunn and Trefler \(2008\)](#) in checking the robustness of our results to a restricted sample that better fits the spirit of the vertical models developed above. More specifically, our restricted sample regressions only include countries from which at least two-thirds of intrafirm U.S. imports reach parent firms in the U.S. ([Nunn and Trefler, 2008; Table 2.5](#)), while an industry category is included only if it contains an intermediate input product according to the categorization of [Wright \(2010\)](#). Although these corrections reduce our sample size in column 7 by about 60%, the results are remarkably similar to those in column 6, both qualitatively and quantitatively.

The vertical models of firm boundaries developed above generate further predictions for the relative prevalence of FDI versus foreign outsourcing. For instance, the property-rights models delivered a positive association of the share of intrafirm trade with trade barriers and productivity dispersion, and a negative association with the wage gap (w^H/w^F). Column 8 reports the effect of (buyer) productivity dispersion, freight costs, and U.S. tariffs on the share of intrafirm trade. As is clear, productivity dispersion indeed has a positive effect on the share of intrafirm trade, but the effect is statistically weak. Conversely, we find that trade costs, natural or man-made, have a negative and statistically significant effect on the share of intrafirm trade, a result that is inconsistent with the models developed in this section. Later in this section, we will revisit this issue

and will provide a potential explanation for this finding.⁵⁶ As for the negative effect of relative wages, the prediction has found some indirect support from regressions that exploit the cross-country variation in the data and have shown a positive effect of a country's aggregate capital-labor ratio on the intrafirm trade share (see Antràs, 2003; Bernard et al., 2010). These cross-country specifications have also unveiled a robust positive effect of the quality of institutions on the share of intrafirm U.S. imports, a counterintuitive result from the point of view of transaction-cost theories. Of course, the standard concerns associated with cross-country regressions (omitted variable bias, endogeneity, etc.) apply here as well, so one should be cautious in interpreting these correlations as necessarily falsifying certain models.

Due to data availability, the bulk of work using product-level data to test the theory of multinational firm boundaries has employed U.S. intrafirm import data. Feenstra and Hanson (2005) and Fernandes and Tang (2010) are two notable exceptions that instead use product-level export data from the Customs General Administration of the People's Republic of China, containing detailed information on whether the exporter is a foreign-owned plant or not.

Recent contributions to the empirical literature on multinational firm boundaries have made use of a few available datasets that include *firm-level* information on the sourcing strategies of firms. Because models of the internalization decision are essentially models of firm behavior, firm-level data are an ideal laboratory to use in testing certain aspects of the models. A notable example of this is the sorting pattern in Figure 2.10, which is key to some of the predictions of these models. There is fairly robust evidence that firms engaging in foreign vertical integration (FDI) appear to be more productive than firms undertaking foreign outsourcing with no FDI (see Tomiura, 2007, for Japan, Corcos et al., 2013, for France, and Kohler and Smolka, 2009, for Spain).⁵⁷ An interesting feature of the Spanish ESEE (Encuesta sobre Estrategias Empresariales) database is that it contains information on both the domestic and foreign sourcing strategies of firms in Spain, thus permitting a fuller evaluation of the empirical relevance of the sorting pattern in Figure 2.10, which compares domestic outsourcing, domestic integration, foreign outsourcing, and foreign integration. The probability density functions in Figure 2.11 indicate that, consistently with Figure 2.10, domestic outsourcers are (on average) the least productive firms, while foreign integrators are (on average) the most productive firms. The relative ranking of domestic integrating firms and foreign outsourcing firms is instead inconsistent with the

⁵⁶ Díez (2010) finds instead a positive correlation between the prevalence of intrafirm trade and U.S. tariffs when working with the more disaggregated six-digit HS data, and when introducing two-digit HS fixed effects into the estimation. He also finds a negative correlation between U.S. intrafirm imports and foreign tariffs and shows that it can be reconciled with a variant of the Antràs and Helpman (2004) framework.

⁵⁷ Defever and Toubal (2007) showed that in French data, the productivity advantage of FDI firms over foreign outsourcers in a given market is reversed in some sectors, but their results are based on a sample of multinational firms, and thus selection biases may complicate the interpretation of their results (see Corcos et al., 2013).

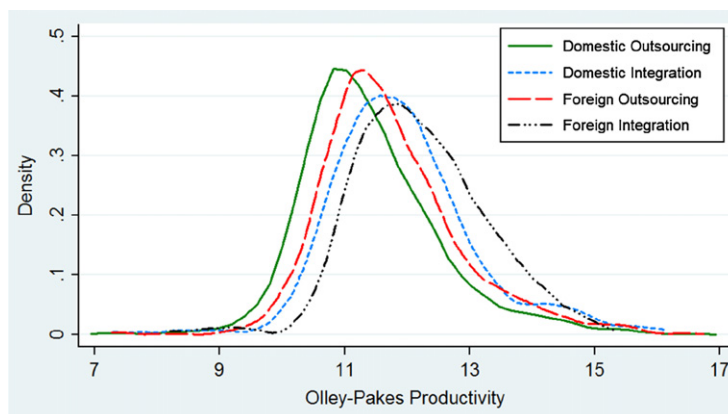


Figure 2.11 Organizational Sorting in Spain

model, and suggests that exploring the implications of a reverse ranking between these two modes is worthwhile, as it may as well help rationalize the negative effects of trade frictions on intrafirm trade shares unveiled by [Table 2.5](#).⁵⁸

Another appealing feature of accessing firm-level data is that they open the door for alternative ways to discriminate between models of the internalization decision by separately identifying the effect of certain variables on the intensive and extensive margin of offshoring (see [Corcos et al., 2013](#), for a recent attempt along these lines).

8. CONCLUSION

This chapter has reviewed the state of the international trade literature on multinational firms. We have addressed to varying degrees the answers that the literature provides to three central questions. Why do some firms become multinational? Where do these firms choose to locate production? And, why do firms own foreign affiliates rather than contract with external providers? In our exposition of the main theoretical contributions of the literature, we have adapted focal models so that they fit within a single organizing framework. It is our hope that the use of consistent tools and notation across models will allow researchers to more easily tease out what is common and, more importantly, what is different about the various approaches. With respect to the empirics, we have re-estimated the econometric models of a few influential papers using the latest available data.

In reviewing a large and diverse literature one is forced to make hard decisions concerning the scope of the coverage. In the case of this chapter, we have faced not only

⁵⁸ Intuitively, given the sorting pattern suggested in [Figure 2.11](#), a reduction in trade frictions will not only lead some marginal firms to shift to foreign outsourcing, but might also lead some particularly productive domestic integrating firms to switch to vertical FDI, and on that account increasing the share of intrafirm trade.

page constraints, but also constraints imposed by our determination to stick to a single organizational principle. For instance, our decision to focus on monopolistically competitive models has meant that we have entirely abstracted from research on certain strategic aspects of multinational production associated with the oligopolistic nature of certain industries in which multinational firms are pervasive. As another example, because the benchmark framework of [Section 3](#) was built on the assumption of CES preferences, we have omitted promising recent research on the role of nonhomothetic preferences in multinational activity. Finally, we have developed models in which capital is immobile across countries and have thus not explored the interactions between multinational activity and FDI flows, an area of utmost importance that has not been much explored in the literature.

REFERENCES

- Acemoglu, D., Antràs, P., Helpman, E., 2007. Contracts and technology adoption. *American Economic Review* 97 (3), 916–943.
- Alfaro, L., Charlton, A., 2009. Intra-industry foreign direct investment. *American Economic Review* 99 (5), 2096–2119.
- Alfaro, L., Conconi, P., Fadinger, H., Newman, A.F., 2010. Trade Policy and Firm Boundaries. NBER Working Paper 16118.
- Antràs, P., 2003. Firms, contracts, and trade structure. *Quarterly Journal of Economics* 118 (4), 1375–1418.
- Antràs, P., 2005. Incomplete contracts and the product cycle. *American Economic Review* 95 (4), 1054–1073.
- Antràs, P., forthcoming. Grossman–Hart (1986) goes global: incomplete contracts, property rights, and the international organization of production. *Journal of Law, Economics and Organization*.
- Antràs, P., Chor, D., 2012. Organizing the Global Value Chain. NBER Working Paper 18163, June.
- Antràs, P., Helpman, E., 2004. Global sourcing. *Journal of Political Economy* 112, 552–580.
- Antràs, P., Helpman, E., 2008. Contractual frictions and global sourcing. In: Helpman, E., Marin, D., Verdier, T. (Eds.), *The Organization of Firms in a Global Economy*. Harvard University Press.
- Antràs, P., Rossi-Hansberg, E., 2009. Organizations and trade. *Annual Review of Economics* 1, 43–64.
- Antràs, P., Garicano, L., Rossi-Hansberg, E., 2006. Offshoring in a knowledge economy. *Quarterly Journal of Economics* 121 (1), 31–77.
- Antràs, P., Desai, M., Foley, C.F., 2009. Multinational firms, FDI flows and imperfect capital markets. *Quarterly Journal of Economics* 124 (3), 1171–1219.
- Arkolakis, C., Ramondo, N., Rodríguez-Clare, A., Yeaple, S., 2013. Innovation and Production in the Global Economy. Mimeo, Penn State University.
- Arnold, F., Javorcik, B., 2009. Gifted kids or pushy parents? Foreign direct investment and plant productivity in Indonesia. *Journal of International Economics* 79 (1).
- Arrow, K., 1962. Economic welfare and the allocation of resources for invention. In: *The Rate and Direction of Inventive Activity: Economic and Social Factors*. National Bureau of Economic Research, pp. 609–626.
- Baldwin, R., Venables, A.J., 2010. Spiders and Snakes: Offshoring and Agglomeration in the Global Economy. CEPR Working Paper 8163.
- Barba Navaretti, G., Venables, A.J., 2004. *Multinational Firms in the World Economy*. Princeton University Press.
- Barefoot, K., Mataloni, R., 2011. Operations of U.S. Multinational Companies in the United States and Abroad: Preliminary Results from the 2009 Benchmark. *Survey of Current Business*, November, 29–48.
- Basco, S., 2010. Financial Development and the Product Cycle. Working Paper, Universidad Carlos III.
- Bernard, A.B., Redding, S., Schott, P.K., 2007. Comparative advantage and heterogeneous firms. *Review of Economic Studies* 74 (1), 31–66.

- Bernard, A.B., Jensen, J.B., Schott, P.K., 2009. Importers, exporters, and multinationals: a portrait of firms in the U.S. that trade goods. In: Dunne, T., Jensen, J.B., Roberts, M.J. (Eds.), *Producer Dynamics: New Evidence from Micro Data*. NBER.
- Bernard, A.B., Jensen, J.B., Redding, S., Schott, P.K., 2010. Intra-firm trade and product contractibility. *American Economic Review Papers and Proceedings* 100 (2), 444–448.
- Blonigen, B.A., Lionel, F., Nicholas, S., Farid, T., 2012. Cherries for Sale: The Incidence of Cross-Border M&A. Mimeo, University of Oregon.
- Brainard, S.L., 1993. A Simple Theory of Multinational Corporations with a Tradeoff between Proximity and Concentration. NBER Working Paper 4269.
- Brainard, S.L., 1997. An empirical assessment of the proximity-concentration trade-off between multinational sales and trade. *American Economic Review* 87(4), 520–544.
- Breinlich, H., 2008. Trade liberalization and industrial restructuring through mergers and acquisitions. *Journal of International Economics* 76 (2), 254–266.
- Buckley, P.J., Casson, M., 1976. *The Future of the Multinational Enterprise*. Holms and Meier, London.
- Caliendo, L., Rossi-Hansberg, E., 2012. The Impact of trade on organization and productivity. *Quarterly Journal of Economics* 127 (3), 1393–1467.
- Carluccio, J., Fally, T., 2012. Global sourcing under imperfect capital markets. *The Review of Economics and Statistics* 94 (3), 740–763.
- Caves, R.E., 1971. International corporations: the industrial economics of foreign investment. *Economica* 38 (149), 1–27.
- Caves, R.E., 2007. *Multinational Enterprise and Economic Analysis*, third ed. Cambridge University Press, Cambridge, UK.
- Chaney, T., 2008. Distorted gravity: the intensive and extensive margins of international trade. *American Economic Review* 98 (4), 1707–1721.
- Chen, M.X., Moore, M.O., 2010. Location decision of heterogeneous multinational firms. *Journal of International Economics* 80 (2), 188–199.
- Chen, Y., Feenstra, R.C., 2008. Buyer investment. Export variety and intrafirm trade. *European Economic Review* 52 (8), 1313–1337.
- Chen, Y., Horstmann, I., Markusen, J., 2012. Physical capital, knowledge capital and the choice between FDI and outsourcing. *Canadian Journal of Economics* 45, 1–15.
- Coase, R.H., 1937. The nature of the firm. *Economica* 4 (16), 386–405.
- Conconi, P., Patrick, L., Newman, A.E., 2012. Trade liberalization and organizational change. *Journal of International Economics* 86 (2), 197–208.
- Corcos, G., Irac, D.M., Mion, G., Verdier, T., 2013. The determinants of intra-firm trade. *The Review of Economics and Statistics* 95 (3), 825–838.
- Davidson, W.H., McFetridge, D.G., 1984. International technology transactions and the theory of the firm. *Journal of Industrial Economics* 32 (3), 253–264.
- Defever, F., Toubal, F., 2007. Productivity and the Sourcing Modes of International Firms: Evidence from French Firm-Level Data. CEP Discussion Paper No. 842.
- Desai, M., Foley, C.F., Hines, J., 2004. The cost of shared ownership: evidence from international joint ventures. *Journal of Financial Economics* 73 (2), 323–374.
- Díez, F.J., 2010. The Asymmetric Effects of Tariffs on Intra-Firm Trade and Offshoring Decisions, Working Paper No. 10-4, Federal Reserve Bank of Boston.
- Dixit, A.K., Stiglitz, J.E., 1977. Monopolistic competition and optimum product diversity. *American Economic Review* 67, 297–308.
- Dornbusch, R., Fischer, S., Samuelson, P.A., 1980. Heckscher-Ohlin trade theory with a continuum of goods. *Quarterly Journal of Economics* 95 (2), 203–224.
- Dunning, J.H., 1981. *International Production and the Multinational Enterprise*. Allen and Unwin, London.
- Eaton, J., Kortum, S., 2002. Technology, geography, and trade. *Econometrica* 70 (5), 1741–1779.
- Ekholm, K., Forslid, R., Markusen, J.R., 2007. Export-platform foreign direct investment. *Journal of the European Economic Association* 5 (4), 776–795.
- Ethier, W., 1986. The multinational firm. *Quarterly Journal of Economics* 101, 805–833.
- Ethier, W., Markusen, J., 1996. Multinational firms, technology diffusion, and trade. *Journal of International Economics* 41, 1–28.

- Feenstra, R.C., Hanson, G.H., 1996. Foreign investment, outsourcing and relative wages. In: Feenstra, R.C., Grossman, G.M., Irwin, D.A. (Eds.), *The Political Economy of Trade Policy: Papers in Honor of Jagdish Bhagwati*. The MIT Press, Cambridge, MA.
- Feenstra, R.C., Hanson, G.H., 2005. Ownership and control in outsourcing to China: estimating the property-rights theory of the firm. *Quarterly Journal of Economics* 120 (2), 729–761.
- Fernandes, A.P., Tang, H., 2010. Determinants of Vertical Integration in Export Processing: Theory and Evidence from China. Working Paper, Tufts University.
- Garetto, S., 2013. Input sourcing and multinational production. *American Economic Journal: Macroeconomics* 5 (2), 118–151.
- Girma, S., Gorg, H., Strobl, E., 2004. Exports, international investment, and plant performance: evidence from a non-parametric test. *Economic Letters* 83 (3), 317–324.
- Glass, A.J., Saggi, K., 2002. Licensing versus direct investment: implications for economic growth. *Journal of International Economics* 56 (1), 131–153.
- Grossman, S.J., Hart, O.D., 1986. The costs and benefits of ownership: a theory of vertical and lateral integration. *Journal of Political Economy* 94 (4), 691–719.
- Grossman, G.M., Helpman, E., 2002. Integration versus outsourcing in industry equilibrium. *Quarterly Journal Economics* 117, 85–120.
- Grossman, G.M., Helpman, E., 2003. Outsourcing versus FDI in industry equilibrium. *Journal of the European Economic Association* 1 (2–3), 317–327.
- Grossman, G., Helpman, E., 2004. Managerial incentives and the international organization of production. *Journal of International Economics* 63 (2).
- Grossman, G.M., Helpman, E., 2005. Outsourcing in a global economy. *Review of Economic Studies* 72 (1), 135–159.
- Grossman, G., Helpman, E., Szeidl, A., 2006. Optimal integration strategies for multinational firms. *Journal of International Economics* 70 (1).
- Grossman, G.M., Rossi-Hansberg, E., 2008. Trading tasks: a simple theory of offshoring. *American Economic Review* 98 (5), 1978–1997.
- Guadalupe, M., Olga, K., Catherine, T., 2012. Innovation and foreign ownership. *American Economic Review* 102 (7), 3594–3627.
- Hanson, G., Mataloni, R., Slaughter, M., 2005. Vertical production networks in multinational firms. *Review of Economics and Statistics* 87 (4), 664–678.
- Hart, O., 1995. *Firms, Contracts, and Financial Structure*. Oxford University Press, New York.
- Hart, O., Moore, J., 1990. Property rights and the nature of the firm. *Journal of Political Economy* 98, 1119–1158.
- Helpman, E., 1984. A simple theory of international trade with multinational corporations. *Journal of Political Economy* 92 (3), 451–471.
- Helpman, E., 2006. Trade, FDI and the organization of firms. *Journal of Economic Literature* 44, 589–630.
- Helpman, E., Krugman, P.R., 1985. *Market Structure and Foreign Trade*. MIT Press, Cambridge, MA (Chapter 12).
- Helpman, E., Melitz, M.J., Yeaple, S.R., 2004. Exports versus FDI with heterogeneous firms. *American Economic Review* 94 (1), 300–316.
- Horn, H., Persson, L., 2001. The equilibrium ownership of an international oligopoly. *Journal of International Economics* 53 (2), 307–333.
- Horstmann, I.J., Markusen, J.R., 1987a. Strategic investments and the development of multinationals. *International Economic Review* 28 (1), 109–121.
- Horstmann, I.J., Markusen, J.R., 1987b. Licensing versus direct investment: a model of internalization by the multinational enterprise. *Canadian Journal of Economics* 20, 464–481.
- Hymer, S.H., 1960. *The International Operations of National Firms: A Study of Direct Foreign Investment*. Ph.D. Dissertation. Massachusetts Institute of Technology, Department of Economics. Published Posthumously, 1976. MIT Press, Cambridge, MA.
- Irrazábal, A., Moxnes, A., Opromolla, L.D., 2013. The margins of multinational production and the role of intra-firm trade. *Journal of Political Economy* 121 (1), 74–126.
- Keller, W., Yeaple, S., 2013. The gravity of knowledge. *American Economic Review* 103 (4), 1414–1444.
- Kindleberger, C., 1969. *American Business Abroad: Six Lectures on Direct Investment*. New Haven, London.

- Klein, M., Peek, J., Rosengren, E., 2002. Troubled banks, impaired foreign direct investment: the role of relative access to credit. *American Economic Review* 92 (3), 664–682.
- Kogut, B., Zander, U., 1993. Knowledge of the firm and the evolutionary theory of the multinational corporation. *Journal of International Business Studies* 24 (4), 625–645.
- Kohler, W., Smolka, M., 2009. Global Sourcing Decisions and Firm Productivity: Evidence from Spain. CESifo Working Paper No. 2903, CESifo Group, Munich.
- Krugman, P.R., 1979. Increasing returns. Monopolistic competition, and international trade. *Journal of International Economics* 9 (4), 469–479.
- Krugman, P.R., 1980. Scale economies. Product differentiation, and the pattern of trade. *American Economic Review* 70 (5), 950–959.
- Mansfield, E., Romeo, A., 1980. Technology transfer to overseas subsidiaries by U.S.-based firms. *Quarterly Journal of Economics* 95 (4), 737–750.
- Mansfield, E., Romeo, A., Wagner, S., 1979. Foreign trade and U.S. research and development. *Review of Economics and Statistics* 61 (1), 49–57.
- Marin, D., Verdier, T., 2009. Power in the multinational corporation in industry equilibrium. *Economic Theory* 38, 437–464.
- Markusen, J., 1984. Multinationals, multi-plant economies, and the gains from trade. *Journal of International Economics* 16, 205–226.
- Markusen, J., 2002. *Multinational Firms and the Theory of International Trade*. MIT Press, Cambridge, MA.
- Markusen, J., Venables, A., 1998. Multinational firms and the new trade theory. *Journal of International Economics* 46 (2), 183–203.
- Markusen, J.R., Venables, A.J., 2000. The theory of endowment. Intra-industry and multinational trade. *Journal of International Economics* 52, 209–234.
- Mayer, T., Ottaviano, G., 2007. *The Happy Few: The Internationalisation of European firms*. Bruegel Blueprint Series, vol. III.
- McLaren, J., 2000. ‘Globalization’ and vertical structure. *American Economic Review* 90 (5), 1239–1254.
- Melitz, M.J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71 (6), 1695–1725.
- Mrázová, M., Neary, P.J., 2012. Selection Effects with Heterogeneous Firms. Mimeo, University of Oxford.
- Mundell, R.A., 1957. International trade and factor mobility. *American Economic Review* 47 (3), 321–335.
- Neary, P.J., 2007. Cross-border mergers as instruments of comparative advantage. *Review of Economic Studies* 74 (4), 1229–1257.
- Nocke, V., Yeaple, S., 2007. Cross-border mergers and acquisitions versus greenfield foreign direct investment: the role of firm heterogeneity. *Journal of International Economics* 72 (2), 336–365.
- Nocke, V., Yeaple, S., 2008. An assignment theory of foreign direct investment. *Review of Economic Studies* 75 (2), 529–557.
- Nunn, N., 2007. Relationship-specificity. Incomplete contracts, and the pattern of trade. *Quarterly Journal of Economics* 122 (2), 569–600.
- Nunn, N., Treffer, D., 2008. The boundaries of the multinational firm: an empirical analysis. In: Helpman, E., Marin, D., Verdier, T. (Eds.), *The Organization of Firms in a Global Economy*. Harvard University Press, pp. 55–83.
- Nunn, N., Treffer, D., 2013. Incomplete contracts and the boundaries of the multinational firm. *Journal of Economic Behavior and Organization* 94, 330–344.
- Ottaviano, T., Thisse, 2002. Agglomeration and trade revisited. *International Economic Review* 43, 409–436.
- Puga, D., Treffer, D., 2010. Wake up and smell the ginseng: the rise of incremental innovation in low-wage countries. *Journal of Development Economics* 91, 64–76.
- Qiu, L.D., Spencer, B., 2002. Keiretsu and relationship-specific investment: implications for market-opening trade policy. *Journal of International Economics* 58, 49–79.
- Qiu, L.D., Zhou, W., 2006. International mergers: incentives and welfare. *Journal of International Economics*, 68, 38–58.
- Ramondo, N., 2013. *A Quantitative Approach to Multinational Production*. Mimeo, Arizona State University.
- Ramondo, N., Rappoport, V., Ruhl, K., 2012. Horizontal vs. Revisiting Evidence from U.S. Multinationals. Mimeo, Arizona State University, Vertical FDI.

- Ramondo, N., Rodríguez-Clare, A., 2013. Trade, multinational production, and the gains from openness. *Journal of Political Economy* 121 (2), 273–322.
- Romer, P.M., 1990. Endogenous technological change. *Journal of Political Economy* 98 (5), S71–S102.
- Rugman, A.M., 1981. *Inside the Multinationals: The Economics of Internal Markets*. Columbia University Press, New York.
- Saggi, K., 2002. Trade, foreign direct investment, and international technology transfer: a survey. *World Bank Research Observer* 17 (2), 191–235.
- Spencer, B., 2005. International outsourcing and incomplete contracts. *Canadian Journal of Economics* 38, 1107–1135.
- Tintelnot, F., 2012. *Global Production with Export Platforms*. Mimeo, Pennsylvania State University.
- Tomura, E., 2007. Foreign outsourcing, exporting, and FDI: a productivity comparison at the firm level. *Journal of International Economics* 72, 113–127.
- Whinston, M.D., 2003. On the transaction cost determinants of vertical integration. *Journal of Law, Economics, and Organization* 19, 1–23.
- Williamson, O.E., 1975. *Markets, Hierarchies: Analysis. Antitrust Implications*. Free Press, New York.
- Williamson, O.E., 1985. *The Economic Institutions of Capitalism*, Free Press (Chapters 1–3).
- Wright, G.C., 2010. *Revisiting the Employment Impact of Offshoring*. Mimeo, University of Essex.
- Yeaple, S., 2003a. The role of skill endowments in the structure of U.S. outward foreign direct investment. *Review of Economics and Statistics*, 85(3): 726–734.
- Yeaple, S., 2003b. The complex integration strategies of multinational firms and cross-country dependencies in the structure of foreign direct investment. *Journal of International Economics* 60 (2), 293–314.
- Yeaple, S., 2006. Foreign direct investment, and the structure of U.S. trade. *Journal of the European Economic Association* 4, 602–611.
- Yeaple, S., 2009. Firm heterogeneity and the structure of U.S. multinational activity: an empirical analysis. *Journal of International Economics* 78, 206–215.
- Yeaple, S., 2013. The multinational firm. *Annual Review of Economics* 5, 193–217.
- Zhu, S.C., Treffer, D., 2005. Trade and inequality in developing countries: a general equilibrium analysis. *Journal of International Economics* 65 (1), 21–48.