Input Sourcing and Multinational Production

By Stefania Garetto*

I propose a general equilibrium framework where firms decide whether to outsource or integrate input manufacturing, domestically or abroad. By outsourcing, firms may benefit from suppliers' technologies, but pay mark-up prices. By sourcing intrafirm, they save on mark-ups and pay possibly lower foreign wages. Multinational corporations arise when firms integrate production abroad. The model predicts that intrafirm imports are positively correlated with the mean and variance of the firms' productivity distribution, and with the degree of input differentiation. I use the model to quantify the U.S. welfare gains from intrafirm trade, which amount to about 0.23 percent of consumption per-capita.
JEL: F12, F23, L11
Keywords: International trade, intrafirm trade, multinational firms, vertical FDI

Globalization has expanded the scope of trade, as trade in finished products is being gradually outpaced by trade in intermediates, taking place both within and across the boundaries of the firm. Many studies document the growth of multistage production, in which plants in different locations contribute to the creation of value added through processing and assembly.[1] A good example is the vertical production chain of the Barbie doll quoted by Feenstra (1998), in which U.S.-produced molds cross six Asian countries before being shipped back to the U.S. where the dolls are sold. Multinational corporations play a large role in this scenario, as a substantial share of offshore production happens within their boundaries. Bernard, Jensen and Schott (2009) report that in the year 2000 almost 50% of U.S. imports and about 30% of exports happened within firms' boundaries.

In this paper I provide a new theoretical framework to think about cost-driven, vertical multinational production and the associated flows of intrafirm trade.

* Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215. E-mail: garettos@bu.edu. I am indebted to Robert E. Lucas Jr. for motivating me and helping me in this work. This paper has benefited from comments by Fernando Alvarez, Pol Antràs, Christian Broda, Thomas Chaney, José Fillat, François Gourio, Gene Grossman, Kenneth Judd, Marc Melitz, Dilip Mookherjee, Natalia Ramondo, Esteban Rossi-Hansberg, Leena Rudanko, Nancy Stokey, Mathis Wagner, three anonymous referees, and seminar participants at the University of Chicago, Arizona State, Boston University, the Chicago Fed, the Fed Board, the New York Fed, NYU, Penn State, Princeton, UC Berkeley (Haas), the University of Iowa, the University of Toronto, the University of Virginia, Yale, the 2008 SED Meeting, and the 2008 Winter ITI NBER Meeting. I especially thank Sharif Rahman for sharing the code to compute the recursion coefficients for the Weibull measure. Gumilang A. Sahadewo provided outstanding research assistance. All remaining errors are mine.

[1] See Campa and Goldberg (1997), Hummels, Rapoport and Yi (1998), Hanson, Mataloni and Slaughter (2001), Hummels, Ishii and Yi (2001), and Yeats (2001).

Firms need to acquire a set of tradeable inputs in order to produce a non-tradeable consumption good. Input production can be outsourced to unaffiliated suppliers, generating volumes of trade in intermediates, or can be integrated by the firm itself. When a firm decides to integrate input production, it sets up a new plant, possibly in another country where factor costs are lower. This choice gives rise endogenously to the creation of multinational firms, and to vertical foreign direct investment (henceforth, FDI) in the form of integrated production abroad.[2] As a result, when inputs produced by an affiliate offshore are shipped back to the parent, we observe flows of intrafirm trade.

The novelty of this approach is the fact that the optimal sourcing strategy is achieved as a market equilibrium, while the most recent literature on this topic (notably Antràs, 2003, and Antràs and Helpman, 2004) presented it as the outcome of a contracting problem. My formulation is advantageous because it allows to explore general equilibrium effects, which are essential to study cost-driven FDI and are overlooked by the previous literature. In my model firms simply choose the sourcing options and the locations that minimize their production costs.

Firms are heterogeneous in productivity and in the type of technology they use. Intermediate goods producers have an adaptable technology with which they produce inputs that they can supply to final good producers worldwide. Final good producers are endowed with two types of technologies: a homogeneous technology to produce the consumption good and a set of heterogeneous, non adaptable technologies that they can use anywhere in the world to produce their own inputs in affiliate plants. Final good producers can either buy inputs from the suppliers or integrate input production using their "in-house" technologies. When they decide to integrate production abroad, they become the parents of a multinational corporation. Offshore integrated production takes the form of vertical FDI, and generates flows of intrafirm trade when the inputs produced offshore are shipped back to the parent.

The model delivers endogenous organizational choices without relying on the property-rights model of the boundaries of the firm. The driving forces behind the sourcing choice are technological heterogeneity and the implications of imperfect competition on prices. By outsourcing from a supplier, a final good producer can have access to a potentially better technology, but has to pay a price which is augmented by a mark-up. On the other hand, by integrating, firms have to use their own technology, but save on the mark-ups charged by the suppliers: intrafirm trade happens between a firm and itself, and is priced at marginal cost.[3]

[2]In his survey of the literature on trade and multinational production, Helpman (2006, pg. 8-9) defines vertical FDI as *"[activity done through] subsidiaries that add value to products that are not destined (...) for the host country market"*.

[3]This way of modeling the differences between intrafirm and arm's length pricing is consistent with recent transaction-level evidence. Bernard, Jensen and Schott (2006) document the existence of a large gap between the prices associated with arm's length transactions and the transfer prices associated with intrafirm transactions: by comparing related-party sales and arm's length sales of the same good by the same firm to the same destination market, they find that arm's length prices are significantly higher than intrafirm prices. Neiman (2011) also finds that arm's length prices are more responsive than intrafirm

Moreover, when intrafirm sourcing takes place abroad, the parent firm is able to transfer (at least partially) its technology to the destination country, and match it with the possibly lower labor costs of the location chosen.[4]

Integrated firms increase competition in the input market: suppliers' market power and prices are reduced by the possibility of integrated production. This effect is reinforced when integration takes the form of multinational production, as lower factor costs in FDI host countries make the integration option more attractive and reduce arm's length prices further. Because of the sensitivity of prices to the possibility of integration, FDI liberalization makes both outsourcing and FDI more attractive, increases competition and lowers prices.

The model has testable implications for the determinants of intrafirm trade flows. First, the intrafirm share of imports is positively correlated with the mean and the variance of the firms' productivity distribution. This prediction is not new in the literature: it holds (through different economic channels) also in the contract-based framework by Antràs and Helpman (2004), and has been successfully tested using data from various countries.[5] Second, the intrafirm share of imports is positively correlated with the degree of differentiation across the goods traded. This prediction is novel, and I test it exploiting variation in intrafirm trade shares at the sector level and various measures of differentiation.

The general equilibrium structure of the model makes it suitable for welfare analysis. I calibrate the model to match aggregate moments of U.S. data, and use the calibrated economy to quantify the welfare gains arising from vertical multinational production, distinguishing them from the gains from arm's length trade.[6] Firms' organizational choices depend on technological differences (through factor costs) and market structure (through price adjustments). The effect of price adjustments is more relevant in scenarios where goods are more differentiated and suppliers have a significant level of market power. The welfare gains for the U.S. economy arising from vertical multinational production are currently about 0.23% of consumption per capita, and further reductions of the costs of FDI can increase them substantially: the model predicts that a 50% drop in the calibrated barrier to FDI could generate a gain of more than 5% of consumption per capita.

The rationale behind the existence of multinational firms is similar to Helpman (1984, 1985), where multinationals emerge to exploit factor cost differences across countries. In these papers, firms choose the location of their activities to minimize production costs. The incentive to become multinational arises from the existence

prices to price changes of competing firms, suggesting that intrafirm prices are set differently from arm's length prices.

[4]Hanson, Mataloni and Slaughter (2005) document the importance of low-cost locations in vertical production networks. Grossman and Rossi-Hansberg (2008) model "tasks" offshoring as motivated by factor cost differences.

[5]See Yeaple (2006), Nunn and Trefler (2008), Corcos et al. (2011), Kohler and Smolka (2012).

[6]The model excludes horizontal FDI, *i.e.* the establishment of offshore production to serve foreign markets, and export platforms, *i.e.* the establishment of offshore production to serve third markets. Ramondo and Rodríguez-Clare (2010) present a quantitative model that captures the gains from these alternative forms of multinational production.

of an immaterial factor of production that may serve product lines without being located in their plants.[7] Similarly, I assume that firms can imperfectly transfer their productivity when they decide to integrate production abroad. In addition, the model I propose generalizes Helpman's idea to a world with heterogeneous firms, trade costs and asymmetric countries.

More recently, Antràs (2003) and Antràs and Helpman (2004) modeled the joint choice of location and organizational structure by merging existing models of trade with a contract-based theory of the firm.[8] Their approach has the advantage of analyzing separately the two choices, and matches qualitative features of the data on intrafirm trade. My model provides a complementary analysis that emphasizes the role of technological heterogeneity and market structure in explaining organizational choices. Moreover, imperfect competition allows me to analyze optimal pricing and the interactions between pricing and organizational choices, which is absent in Antràs and Helpman's work. The model shares predictions with Antràs and Helpman (2004) and delivers an additional prediction driven by the optimal pricing mechanism. Finally, the general equilibrium structure of the model makes it amenable to a calibration exercise in which I quantify the welfare effects deriving from multinational production and intrafirm trade.

Methodologically, this paper is close to the quantitative literature on the gains from trade and openness. The model I develop extends Eaton and Kortum (2002) and Alvarez and Lucas (2007) to incorporate imperfect competition and the choice of the sourcing option. Other quantitative papers on the topic, like Rodríguez-Clare (2007), Ramondo and Rodríguez-Clare (2010), and Arkolakis et al. (2011), focused on quantifying the gains from horizontal multinational production. Intrafirm trade of the vertical type is present in Irarrazabal, Moxnes and Opromolla (2010), but is assumed as the preferred organizational choice for affiliates' sourcing, and does not respond to changes in the economy. The model presented in this paper is unique in delivering a quantitative framework where intrafirm trade and vertical FDI are endogenous outcomes of firms' choices.

The rest of the paper is organized as follows. Section I lays out the closed economy model, to isolate the choice between outsourcing and integration without considering the location choice. Section II extends the model to a two-country open economy and illustrates the properties of the general equilibrium. Section III shows the testable implications of the model linking intrafirm trade shares with the moments of the productivity distribution and with the degree of differentiation across goods. Section IV contains the calibration and the computation of the welfare gains from vertical multinational production. Section V concludes.

[7]The idea of modeling multinational production through the existence of an immaterial factor is present also in Markusen (1984). In his setup, multinational corporations arise to increase efficiency by avoiding duplication of the control input, but this may come at the expense of higher market power and higher prices. Conversely, the structure of competition in my model implies that the presence of multinationals increases competition and reduces prices.

[8]Antràs (2003) is based on Helpman and Krugman (1985), while Antràs and Helpman (2004) is based on Helpman, Melitz and Yeaple (2004). Grossman and Helpman (2002) and Grossman and Helpman (2004) also model organizational choices and location choices.

I. Integration and Outsourcing in the Closed Economy

In this section I present the closed economy model. I extend Eaton and Kortum (2002) and Alvarez and Lucas (2007) to incorporate imperfect competition and the choice of the sourcing option. Given the structure of technology heterogeneity, the organization choice simply adds one dimension to the description of goods as vectors of technology draws.

The economy is organized in two sectors. There is an intermediate goods sector, where a continuum of differentiated goods is produced using labor as the only input, and a final good sector, where labor and an aggregate of intermediate goods are combined in the production of a unique, homogeneous consumption good.

Accordingly, there are two types of producers in this economy: final good producers (or buyers) and intermediate goods producers (or suppliers). There is a fixed continuum of intermediate inputs. Each input is produced by a unique supplier. Suppliers differ in their labor productivity and each of them has market power over the input he produces.[9] Let z denote a supplier's unit labor requirement. z is a random draw from a common density $\psi(z)$ defined on $\Re_+$. Each supplier can sell his good to any buyer, without having to incur any cost to adapt it to the buyer's specific production process. Suppliers cannot discriminate across buyers, and each supplier charges a price $p(z)$, which depends on his cost draw, to all buyers in the market.

The final good is produced by a continuum of identical producers of unit mass, operating in a perfectly competitive market. They all produce the same, homogeneous consumption good using labor and intermediate goods as inputs. For each input, a final good producer has two possible sourcing options: he can either produce it in-house or buy it from a supplier. When he decides to integrate production, his technology allows him to produce only for his own product line.[10]

The sourcing decision involves comparing the costs of the two options: the in-house cost of production and the outside price charged by the supplier. For each input, the final good producer has an in-house unit labor requirement x, which is a random draw from a density $\phi(x)$ defined on $\Re_+$ and indicates the number of units of labor needed to internalize production of one unit of input. All the final good producers draw the in-house costs from the same distribution $\phi(\cdot)$, but they can have different cost draws for the same input. In the closed economy the wage is normalized to one, so the unit labor requirement x is equal to the unit

[9]The assumption of market power on behalf of intermediate producers is fairly standard in models of vertical interactions the Industrial Organization literature, where vertical integration arises as a way to avoid the inefficiency induced by double marginalization – see for example Tirole (1988). Waterman and Weiss (1996) report evidence of suppliers' market power in the U.S. cable TV industry. Caves and Bradburd (1988) find that vertical integration rises with market concentration in the suppliers' industry. Bernard, Jensen and Schott (2006) argue that some form of market power of intermediate producers is necessary to account for large observed differences between arm's length prices and intrafirm prices.

[10]In principle, the final good producer could acquire an adaptable technology (at some cost) to enter the intermediate goods' market and sell inputs to other final good producers. I assume that that cost is too large to be covered by the expected profits.

cost of in-house production. Hence each final good producer observes a set of input prices $\{p(z)\}$, draws a set of in-house labor requirements $\{x\}$ and then – for each intermediate good – he chooses whether he wants to buy it or produce it. Obviously, he buys those inputs for which the selling price $p(z)$ is lower than the in-house unit cost of production x.

A. *The Final Good Producer's Problem*

Intermediate goods differ only in their labor costs. For this reason, I index each intermediate good with the pair of unit labor requirements that the two types of agents need for its production: (x, z) denotes a good for which the potential buyer has unit cost x and the supplier has unit cost z and charges a price $p(z)$. Accordingly, $q(x, z)$ denotes the quantity produced of good (x, z), regardless of who produces it. A final good producer minimizes the total cost of input sourcing:

$$(1) \qquad \begin{aligned} \min_{q(x,z)} \quad & \int_0^\infty \int_0^\infty \min\{x, p(z)\} q(x,z)\phi(x)\psi(z) dx dz \\ s.t. \quad & \left[\int_0^\infty \int_0^\infty q(x,z)^{1-1/\eta} \phi(x)\psi(z) dx dz \right]^{\eta/(\eta-1)} \geq q \end{aligned}$$

where $\eta > 1$ is the elasticity of substitution across inputs, and q denotes an aggregate of intermediate goods, which the final good producer takes as given and is determined by equilibrium conditions in the final good market.[11] The arm's length prices $p(z)$ are also taken as given.

Let $\mathcal{B}^I = \{(x, z) : x \leq p(z)\}$ be the set of inputs that the final good producer decides to internalize and $q^I(x, z)$ be the solution of (1) in $\mathcal{B}^I$. Similarly, let $\mathcal{B}^T = \{(x, z) : x \geq p(z)\}$ be the set of inputs that the final good producer decides to outsource and $q^T(x, z)$ be the solution of (1) in $\mathcal{B}^T$. Hence:

$$(2) \qquad q^I(x,z) \equiv q^I(x) = x^{-\eta} p^\eta q \qquad \forall\ (x,z) \in \mathcal{B}^I$$

$$(3) \qquad q^T(x,z) \equiv q^T(p(z)) = [p(z)]^{-\eta} p^\eta q \quad \forall\ (x,z) \in \mathcal{B}^T.$$

The term p is the price index for intermediate goods: $p = \left[p_I^{1-\eta} + p_T^{1-\eta} \right]^{1/(1-\eta)}$,

[11] The assumption of a continuum of goods implies that – by the law of large numbers – the aggregate q is the same across final good producers even if they have different cost draws for each of the goods.

where:

$$p_I = \left[\int_0^\infty \int_0^{p(z)} x^{1-\eta} \phi(x) \psi(z) dx dz \right]^{1/(1-\eta)} \tag{4}$$

$$p_T = \left[\int_0^\infty p(z)^{1-\eta} \Big[1 - \Phi(p(z)) \Big] \psi(z) dz \right]^{1/(1-\eta)} \tag{5}$$

and $\Phi(\cdot)$ is the cumulative distribution function associated with the in-house density $\phi(\cdot)$.

B. The Supplier's Problem

A supplier with cost draw z chooses the profit-maximizing price $p(z)$ by trading off the higher per-unit profits given by a higher price with the possibility of capturing a larger mass of buyers with a relatively lower price:

$$\max_{p(z)} \; [p(z) - z] \int_{p(z)}^\infty q^T(p(z)) \phi(x) dx \tag{6}$$

where z is the supplier's unit cost of production, $q^T(p(z))$ is given by (3), and $\int_{p(z)}^\infty \phi(x)dx$ is the mass of buyers that decide to buy the good at price $p(z)$. The first order condition of this problem is:

$$[p(z) - z] \left[q^T(p(z)) \phi(p(z)) - \frac{\partial q^T(p(z))}{\partial p(z)} [1 - \Phi(p(z))] \right] = q^T(p(z)) \int_{p(z)}^\infty \phi(x) dx. \tag{7}$$

Equation (7) summarizes the supplier's trade-off. For a given level of sales, the gain from increasing the mark-up over the marginal cost ($[p(z) - z]$) must be counterbalanced by the sum of the losses on both the extensive and the intensive margin. If the supplier raises the price, he is going to lose the marginal buyers (the extensive margin, captured by the term $q^T(p(z))\phi(p(z))$) and he is going to sell lower quantities to the remaining buyers (the intensive margin, captured by the term $-\frac{\partial q^T(p(z))}{\partial p(z)}[1 - \Phi(p(z))]$). Using (3), the first order condition delivers:

$$p(z) = \frac{\hat{\eta}}{\hat{\eta} - 1} z \tag{8}$$

where:

$$\hat{\eta} \equiv \eta + \frac{\phi(p(z))}{1 - \Phi(p(z))} p(z). \tag{9}$$

Equations (8)-(9) show how the buyers' possibility of integration generates a departure from the constant mark-up pricing rule implied by CES preferences associated with monopolistic competition. If integration were not possible, $\hat{\eta} = \eta$ and the optimal pricing rule would depend only on the standard intensive margin, whose magnitude depends on the elasticity of substitution. The possibility of integration introduces the extra term $\frac{\phi(p(z))}{1-\Phi(p(z))}p(z)$ (the extensive margin), which depends on the hazard rate, or the probability that – after an infinitesimal price increase – the buyer switches to integrating, conditional on buying before the price increase. Since this term is non-negative, (8) implies that prices are strictly below the standard constant mark-up ones: $p(z) < \frac{\eta}{\eta-1}z, \quad \forall z > 0$. Moreover, as the term $\frac{\phi(p(z))}{1-\Phi(p(z))}p(z)$ depends on the supplier's cost draw z, mark-ups are endogenous and variable across suppliers. For a wide range of parameterizations, the term $\frac{\phi(p(z))}{1-\Phi(p(z))}p(z)$ is increasing in the supplier's unit cost z, indicating that more productive suppliers charge lower prices but higher mark-ups than less productive suppliers.[12]

To illustrate graphically the behavior of (8), Figure 1 plots prices $p(z)$ and mark-ups $p(z)/z$ as functions of the unit cost z for Weibull-distributed in-house costs: $\Phi(x) = 1 - e^{-\lambda x^{\vartheta}}$, for $x \geq 0$, $\lambda > 0$, $\vartheta > 1$.[13] In the left panel, the dotted line is the supplier's marginal cost z, the solid line is the price $p(z)$ (solution of equation (8)), and the dashed line is the constant mark-up pricing rule of the model without possibility of integration. By comparing the pricing strategies with and without integration is evident that the integration option significantly reduces suppliers' profit margins. In the right panel of Figure 1, the solid line is the supplier's mark-up $p(z)/z$ as a function of the cost draw z. The dashed line is the constant mark-up of the standard model without possibility of integration. The model displays endogenous mark-ups, higher for more productive sellers, and lower for less productive ones.[14]

It is easy to prove that the optimal price is decreasing in η: when the degree of substitutability increases, potential buyers can more easily switch to cheaper substitutes, hence suppliers must decrease the price to keep their share of the

[12] The result of endogenous mark-ups holds for any functional specification of the cost distribution $\phi(x)$, except for the Pareto, for which the elasticity of demand is constant and hence mark-ups are constant too (but lower than in the model without integration). Garetto (2012) describes the properties of equation (8), including conditions for monotonicity of the term $\frac{\phi(p(z))}{1-\Phi(p(z))}p(z)$ and the implications of this pricing behavior for incomplete pass-through and pricing-to-market.

[13] Consistently with Eaton and Kortum (2002) and Alvarez and Lucas (2007), I will also use the Weibull cost distribution to parameterize the model in Sections III and IV. If the distribution of in-house costs is Weibull, equation (8) reduces to: $\lambda\vartheta p(z)^{\vartheta} \cdot (p(z) - z) + \eta(p(z) - z) - p(z) = 0$.

[14] Melitz and Ottaviano (2008) obtain mark-ups variability and dependence of profits on productivity by assuming linear demand systems with horizontal product differentiation. Bernard et al. (2003) obtain similar features by assuming Bertrand competition in the intermediate goods sector. In Bernard et al. (2003) aggregate mark-ups do not depend on country characteristics and geographic barriers, while they do in Melitz and Ottaviano (2008) and in this paper. The dependence of aggregate mark-ups on country characteristics creates a link between trade/FDI liberalization and competition which will be clearer in Section II.

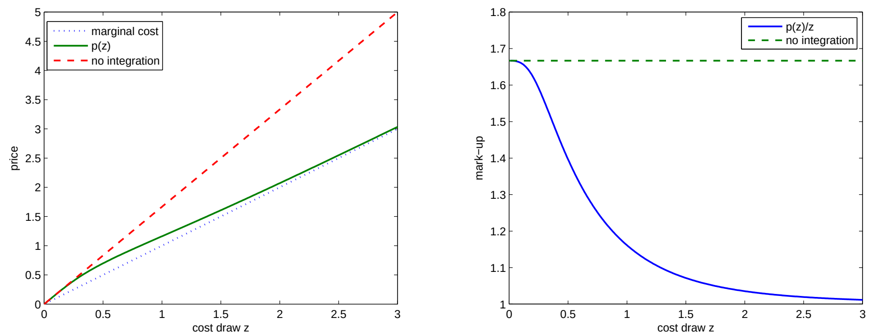

FIGURE 1. PRICES AND MARK-UPS, CLOSED ECONOMY ($\eta = 2.5$, $\lambda = 1$, $\vartheta = 3$).

market.[15] The price is decreasing in λ. When λ decreases, the mean of the buyers' cost distribution increases: potential buyers are more likely to have high costs and suppliers can charge higher prices and mark-ups. The parameter ϑ affects the concavity of the pricing function: a higher ϑ makes the pricing function more concave.

C. *Equilibrium in the Final Good Market*

Production of the final consumption good c is done through a constant returns to scale technology which requires labor and the intermediate goods aggregate q as inputs: $c = q^{\alpha} l_f^{1-\alpha}$, where $\alpha \in (0, 1)$ and l_f is the labor force employed in the final good sector. Let L denote the country's total labor force; then $l_i = L - l_f$ is the labor force working in the intermediate goods sector (for both suppliers and integrated segments of final good producers). The linearity of each intermediate good production technology implies: $q = \frac{l_i}{k}$, where k is the number of units of labor required to produce one unit of the aggregate q:

(10)
$$k = p^{\eta} \left[\int_0^{\infty} \int_0^{p(z)} x^{1-\eta} \phi(x) \psi(z) dx dz + \int_0^{\infty} z p(z)^{-\eta} \left[1 - \Phi(p(z))\right] \psi(z) dz \right].$$

[15]For $\eta \to \infty$, prices tend to marginal costs and the model reduces to a perfectly competitive framework where organizational choices are purely driven by productivity differences.

Optimality in the final good market delivers the equilibrium labor allocation and the value of q:

(11)
$$l_f = \left(\frac{(1-\alpha)p}{(1-\alpha)p+\alpha k}\right)L \;\; ; \;\; l_i = \left(\frac{\alpha k}{(1-\alpha)p+\alpha k}\right)L \;\; ; \;\; q = \left(\frac{\alpha}{(1-\alpha)p+\alpha k}\right)L.$$

The zero-profit equilibrium price of the final good is: $r = \alpha^{-\alpha}(1-\alpha)^{\alpha-1}p^{\alpha}$.

In equilibrium, the following market clearing condition holds: $L + \Pi = rc$, where L is equal to labor income (wages are normalized to one), rc denotes total expenditure in the final good, and Π are total suppliers' profits:

(12)
$$\Pi = \int_0^{\infty} [p(z) - z]p(z)^{-\eta}p^{\eta}q[1 - \Phi(p(z))]\psi(z)dz.$$

II. The Open Economy: Integration, Trade and FDI

I consider now producers' optimal choices in a world of two countries, that I denote by h (Home) and f (Foreign).[16] Each country is a replica of the economy of the previous section, in the sense that is populated by a continuum of identical final good producers and by a continuum of specialized intermediate goods producers. A final good producer in a country sources a continuum of inputs to produce a non-tradeable, homogeneous final good. As in the closed economy, each input can be either produced in an integrated facility or bought from a specialized supplier, but each of these options can be implemented domestically or abroad. Suppliers in each country can sell to buyers in each country, so a final good producer can outsource from the lowest cost supplier "worldwide". Similarly, if a final good producer decides to integrate production, he can set up a domestic integrated plant, or engage in FDI and set up an affiliate abroad. The optimal sourcing strategy is determined comparing arm's length prices and in-house costs of production worldwide, and is going to be affected also by trade costs and factor cost differences between the two countries.

When a final good producer decides to integrate production abroad, he engages in vertical FDI. Since the final good is non-tradeable and is assembled domestically, the inputs produced in foreign affiliate plants are shipped back to the parent, generating flows of intrafirm trade, precisely of imports from foreign affiliates. I restrict FDI to be only vertical in this economy, *i.e.* firms that decide to establish a plant abroad do not serve the host country. Foreign plants are only used to produce inputs for the domestic final good sector. This restriction relies on assuming that the in-house technology is not adaptable to serve other firms.

Labor is immobile, so wages may differ across countries. All producers take

[16] The model can be written for an arbitrarily large number of countries. I present here the two-country case to be consistent with the quantitative analysis in Section IV, which is bilateral for data limitations.

wages as given, and wages are determined endogenously to clear the labor market in each country. I denote with w_i the wage level in country i ($i = h, f$). A final good producer located in country i has a set of technology draws $\{x_i\}$, each drawn from a country-specific distribution $\phi_i(x_i)$. If he decides to integrate production of an input, he may choose to do so in his own country or abroad. If he decides to produce at home, his unit cost is given by his technology draw times the domestic wage, $w_i x_i$. If he decides to produce abroad, he can transfer its technology draw to the foreign country and hire foreign workers at local wages. Production abroad entails other costs, like building a new plant, dealing with foreign institutions, relocate managerial know-how, in addition to the trade costs between parent and affiliate that must be paid when repatriating the produced inputs for further manufacturing. For simplicity, I model these costs as bilateral iceberg costs, implicitly assuming that they are correlated with the size of production.[17] I denote with τ_v the unit iceberg cost for a final good producer to vertically integrate an input abroad. Hence, if a final good producer from country i decides to produce in country j an input for which he has cost draw x_i, his unit cost of production is $\tau_v w_j x_i$ ($i, j = h, f$ and $\tau_v > 1$).

We now turn to the outsourcing option. In each country there is a continuum of suppliers, each of whom produces a unique differentiated input with an adaptable technology that enables him to sell it to buyers in both countries. Each supplier in country j ($j = h, f$) has a productivity draw z_j, which affects his marginal cost and the price he charges for the good. Each z_j is drawn from the country-specific distribution $\psi_j(z_j)$, and the cost distributions $\{\phi_i(\cdot)\}_{i=h,f}$, $\{\psi_j(\cdot)\}_{j=h,f}$ are mutually independent across countries. An intermediate goods producer in country j can only hire domestic labor, hence his unit cost of production is $w_j z_j$.[18] When selling abroad, he also bears an additional cost, representing barriers to international trade, such as tariffs and transportation costs. I denote with τ_o the iceberg trade costs associated with arm's length (outsourced) transactions ($\tau_o > 1$).[19]

Given imperfect competition in the intermediate goods market, suppliers of the same input from different countries choose their optimal prices to maximize their profits, keeping into account direct competition from suppliers in the other country. I denote with $p_{ij}(z_j)$ the price charged to a potential buyer in country i by a supplier in country j who has a cost draw z_j.

In this setup, an input used by a final good producer in country i is defined by the three-dimensional vector $(x_i, \mathbf{z}) = (x_i, z_h, z_f)$, for $i = h, f$. I denote with $q_i(x_i, \mathbf{z})$ the quantity used of an intermediate good for which a final good pro-

[17] In Section IV, iceberg FDI costs are calibrated to match the intrafirm share of U.S. imports.

[18] I assume that only final good producers can have offshore production plants, while suppliers produce only domestically. This assumption can be relaxed and the unit cost of a supplier from country j producing in country i would take the form $\tau_v w_i z_j$, for $i \neq j$.

[19] I am not restricting the two parameters τ_v and τ_o to be equal, but in theory they could be the same. I will let the targeted moments from the data determine the value of the iceberg cost parameters in Section IV.

ducer in country i has cost draw x_i and suppliers in the two countries have cost draws $\mathbf{z} = \{z_h, z_f\}$.

A. Organizational Choices and Location

The analysis of the model follows almost unchanged from the previous section. A final good producer in country i observes his own set of technology draws $\{x_i\}$, a set of wages and iceberg costs $\{w_h, w_f, \tau_v\}$, a set of arm's length C.I.F. prices $\{p_{ij}(z_j)\}_{j=h,f}$, and decides whether to buy or produce each of the inputs and where to do so. The organizational choice is done by comparing the minimum cost across countries of producing an input and the minimum arm's length price of buying it. The problem is exactly as in the closed economy, but with four possible prices to shop for instead of two. Let $c_i(x_i, \mathbf{z})$ denote the minimum unit cost of good $(x_i, \mathbf{z})$:

$$c_i(x_i, \mathbf{z}) = \min\{m_i x_i, p_{ii}(z_i), p_{ij}(z_j)\} \quad \text{for} \quad i \neq j \tag{13}$$

where $m_i = \min\{w_i, \tau_v w_j\}$ for $i \neq j$.

Once a final good producer decides to integrate, the location of production is determined by the interaction between iceberg costs and wages. On the other hand, once he decides to outsource, the location of the lowest cost supplier is determined by trade costs and by the joint cost distribution of suppliers in both countries, which affect the prices charged. A final good producer with a set of cost draws $\{x_i\}$ in country i solves:

$$\begin{aligned} \min_{q_i(x_i,\mathbf{z})} \quad & \int_{\Re^3_+} c_i(x_i, \mathbf{z}) q_i(x_i, \mathbf{z}) \phi_i(x_i) \psi(\mathbf{z}) dx_i d\mathbf{z} \\ s.t. \quad & \left[\int_{\Re^3_+} q_i(x_i, \mathbf{z})^{1-1/\eta} \phi_i(x_i) \psi(\mathbf{z}) dx_i d\mathbf{z} \right]^{\eta/(\eta-1)} \geq q_i \end{aligned} \tag{14}$$

where $\psi(\mathbf{z}) = \psi_h(z_h) \cdot \psi_f(z_f)$ is the density of the vector $\mathbf{z} = (z_h, z_f)$, and the intermediate goods aggregate q_i is determined by equilibrium conditions in the final good market. Let $\mathcal{B}_i^I = \left\{(x_i, \mathbf{z}) \in \Re^3_+ : c_i(x_i, \mathbf{z}) = m_i x_i\right\}$ denote the set of goods that a final good producer in country i decides to internalize, in the location(s) with the lowest cost m_i. Let $\mathcal{B}_{ij}^T = \left\{(x_i, \mathbf{z}) \in \Re^3_+ : c_i(x_i, \mathbf{z}) = p_{ij}(z_j)\right\}$ denote the set of goods that he decides to outsource from a producer in country j. Problem (14) is solved by:

$$q_i^I(x_i, \mathbf{z}) = q_i^I(x_i) = (m_i x_i)^{-\eta} p_i^\eta q_i \quad \forall\ (x_i, \mathbf{z}) \in \mathcal{B}_i^I \tag{15}$$

$$q_i^T(x_i, \mathbf{z}) = q_i^T(p_{ij}(z_j)) = [p_{ij}(z_j)]^{-\eta} p_i^\eta q_i \quad \forall\ (x_i, \mathbf{z}) \in \mathcal{B}_{ij}^T \tag{16}$$

where p_i is the price index for intermediate goods in country i:

$$p_i = \left[(p_i^I)^{1-\eta} + \sum_{j=h,f} (p_{ij}^T)^{1-\eta} \right]^{1/(1-\eta)} \tag{17}$$

and:

$$p_i^I = \left[\int_{\mathcal{B}_i^I} (m_i x_i)^{1-\eta} \phi_i(x_i) \psi(\mathbf{z}) dx_i d\mathbf{z} \right]^{1/(1-\eta)} \tag{18}$$

$$p_{ij}^T = \left[\int_{\mathcal{B}_{ij}^T} [p_{ij}(z_j)]^{1-\eta} \phi_i(x_i) \psi(\mathbf{z}) dx_i d\mathbf{z} \right]^{1/(1-\eta)} . \tag{19}$$

It remains to determine the prices $\{p_{ij}(z_j)\}_{i,j=h,f}$. Markets are segmented. In the intermediate goods market, each supplier maximizes its expected profits from sales to potential buyers in both countries, and may charge different prices to buyers in different countries. By assuming that no resale is possible, I study the pricing problem country by country. In choosing the optimal price to charge in a market, a supplier must consider direct competition from the producer of the same good in the other country and the fact that potential buyers have the option of integrating production. I assume there is only one supplier of each input in each country.[20] Each supplier observes his own marginal cost and the parameters of the cost distributions of the potential buyers and of his competitors in other countries.[21] Based on this information, suppliers simultaneously declare a set of prices (one for each country). In each market, the supplier that declares the lowest price sells the input to all the buyers with sufficiently high costs of insourcing. The price setting mechanism has the properties of a potentially asymmetric first-price sealed-bid auction.[22] Each supplier sets the price as a function of his own marginal cost in a way that, given that all the other suppliers set their price in the same way, no individual supplier could do better by choosing the price differently.

[20]This assumption can be relaxed by interpreting the production function of a supplier as the aggregate production function of a set of "lower-level" suppliers that produce the same input in a country. By assuming that the technology for producing an input is country-specific (*i.e.*, that z_j is constant across all producers of the same input in country j) and that each lower-level producer has a decreasing returns to scale production function and pays a fixed cost to enter the market (along the lines of Rossi-Hansberg and Wright, 2007), it can be shown that the aggregation of lower-level producers generates a constant returns to scale technology that is isomorphic to the linear technology of each supplier in the model. This is achieved by appropriately redefining the technology draw z as a function of the fixed entry cost and of the parameter ruling decreasing returns for the lower-level producers.

[21]Dvir (2010) studies the final good producer's optimal procurement problem in a setting with the same informational assumptions.

[22]In their survey of the auctions literature, McAfee and McMillan (1987, pg. 701-702, 706) report that *"sealed-bid tenders are (...) used by firms procuring inputs from other firms"*. Asymmetric auctions seem a natural tool to study pricing in international markets, *"when both domestic and foreign firms submit bids and, for reasons of comparative advantage, there are systematic cost differences between domestic and foreign firms"*. The online Appendix analyzes the similarities of this problem with a first-price sealed-bid auction.

The resulting equilibrium is a Bayesian Nash equilibrium, where each supplier chooses its optimal price based on his guess (correct in equilibrium) of the pricing rules followed by suppliers of the same good in the other country.

Let $\hat{\mathcal{B}}_{ij}(p_{ij}(z_j))$ be the set of technology draws of buyers in country i and of suppliers outside country j such that the buyers in country i decide to buy good $(x_i, \mathbf{z})$ from the supplier in country j: $\hat{\mathcal{B}}_{ij}(p_{ij}(z_j)) = \{(x_i, \{z_k\}_{k\neq j}) \in \Re^2_+ : (x_i, \mathbf{z}) \in \mathcal{B}^T_{ij}\}$. A supplier in country j with productivity draw z_j maximizes his expected profits from sales in country i in the set $\hat{\mathcal{B}}_{ij}(p_{ij}(z_j))$:

$$\max_{p_{ij}(z_j)} \int_{\hat{\mathcal{B}}_{ij}(p_{ij}(z_j))} [p_{ij}(z_j) - \tau_o w_j z_j] q^T_i(x_i, \mathbf{z})\phi_i(x_i)\psi_k(z_k)dx_i dz_k \quad \text{for} \quad k \neq j. \tag{20}$$

Using (16), and due to the independence property of the cost distributions, (20) can be restated as:

$$\max_{p_{ij}(z_j)} [p_{ij}(z_j) - \tau_o w_j z_j] \left(\frac{p_{ij}(z_j)}{p_i}\right)^{-\eta} q_i A_{ij}(p_{ij}(z_j)) \tag{21}$$

where $A_{ij}(p_{ij}(z_j)) = \left[1 - \Phi_i\left(\frac{p_{ij}(z_j)}{m_i}\right)\right] \cdot [1 - G_{ik}(p_{ij}(z_j))]$ is the probability that – given the price $p_{ij}(z_j)$ – a final good producer in country i buys good $(x_i, \mathbf{z})$ from the supplier in country j. $G_{ik}(\cdot)$ denotes the cumulative distribution function of the prices charged by suppliers in country k $(k \neq j)$ to final good producers in country i. The first order condition of problem (21) is:

$$\begin{aligned} & p_{ij}(z_j)(1-\eta) + \eta\tau_o w_j z_j - ... \\ & ...p_{ij}(z_j)[p_{ij}(z_j) - \tau_o w_j z_j] \left\{ \frac{\phi_i\left(\frac{p_{ij}(z_j)}{m_i}\right)\frac{1}{m_i}}{\left[1 - \Phi_i\left(\frac{p_{ij}(z_j)}{m_i}\right)\right]} + \frac{g_{ik}(p_{ij}(z_j))}{[1 - G_{ik}(p_{ij}(z_j))]} \right\} = 0 \end{aligned} \tag{22}$$

for $k \neq j$. As each supplier competes with the supplier of the same good from the other country, $p_{ij}(\cdot)$ is determined by evaluating the effects of substitutability on demand, the average "insourcing capacity" of the potential buyers (the hazard rate term $\frac{\phi_i(p_{ij}(z_j)/m_i)/m_i}{[1 - \Phi_i(p_{ij}(z_j)/m_i)]}$), and how the optimal price compares with the expected price charged by the supplier in the other country (the hazard rate term $\frac{g_{ik}(p_{ij}(z_j))}{[1 - G_{ik}(p_{ij}(z_j))]}$). Notice that – since in the open economy sourcing possibilities have increased – prices are going to be lower than in the closed economy.

The system of equations (22) must be solved numerically for the entire distributions of prices in the two countries.[23] Here I illustrate properties of the pricing rule

[23]The algorithm to solve system (22) is described in the online Appendix and is available upon request

for two identical countries with some arbitrary trade barriers ($\tau_o = 1.3$, $\tau_v = 1.5$) and Weibull-distributed costs: $\Phi_i(x_i) = 1 - \lambda_i e^{-\lambda_i x_i^{\vartheta}}$, and $\Psi_i(z_i) = 1 - \mu_i e^{-\mu_i z_i^{\vartheta}}$, for $x_i, z_i \geq 0$, $\lambda_i, \mu_i > 0$, $\vartheta > 1$.

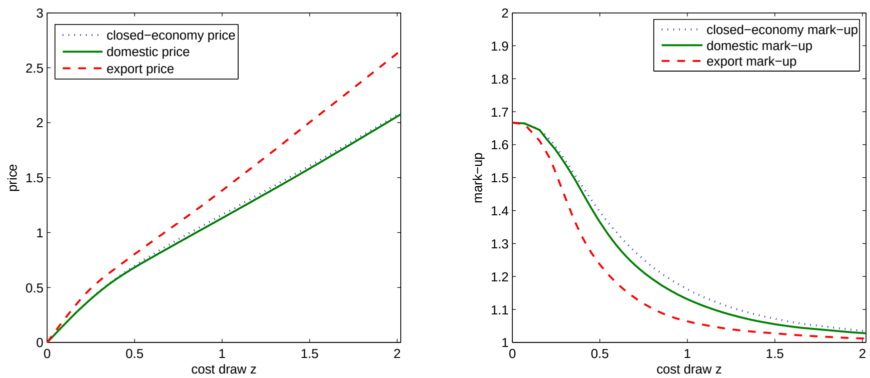

FIGURE 2. OPEN ECONOMY PRICES (2 SYMMETRIC COUNTRIES, $\tau_o = 1.3$, $\tau_v = 1.5$).

The left panel of Figure 2 shows both the closed-economy price (the dotted line) and C.I.F. open-economy prices. Trade costs create a wedge between domestic prices (the solid line) and export prices (the dashed line). Moreover, due to competition among suppliers in the two countries, domestic prices in the open economy are lower than autarky prices. The right panel of the figure shows the corresponding mark-ups: also the domestic open-economy mark-ups are lower than the autarky ones. Moreover, export mark-ups are even lower to counteract the fact that foreign buyers must also pay the transportation cost on the imported goods: firms shrink their mark-ups to be competitive in the foreign market despite the higher costs. Introducing heterogeneity in the two countries' wages and productivity distributions may create larger wedges between domestic and export prices, and may also produce export mark-ups higher than the domestic ones, if competition in the home country is tougher than in the export market.

Numerical exercises show that the price $p_{ij}(z_j)$ is increasing in the cost of integration m_i: a high minimum costs of integration (through wages or iceberg costs) makes the integration option less attractive, and a higher arm's length price still preferable for the potential buyers. Similarly, lower productivity in the other country increases the price charged, as foreign competition is weaker $(\frac{\partial p_{ij}(z_j)}{\partial \mu_k} < 0$ for $k \neq j)$. Finally, like in the closed economy, prices are decreasing in λ and η, while ϑ affects the concavity of the pricing function. These properties make explicit the dependence of prices and mark-ups on country characteristics

to the author.

and geographic barriers. Consequently, country characteristics also affect the choice of undertaking intrafirm transactions through their effects on arm's length prices.[24] The analysis of the pricing strategy confirms that when a country opens to operations with other countries, both integration and trade become cheaper: integration may be relocated in a lower-cost country, and trade becomes more attractive because the higher degree of competition has the effect of lowering prices.

Finally, market power on the side of the suppliers implies that if we compare the intrafirm and arm's length price of the same good provided by either an arm's length supplier or an intrafirm affiliate from the same country and with the same labor productivity, the price associated with the arm's length transaction is higher than the price associated with the intrafirm transaction (which is assumed to be equal to marginal cost). Bernard, Jensen and Schott (2006) document differences in arm's length and intrafirm prices that are consistent with this result: by looking at data for U.S.-based multinational corporations, they find that within firm, product, destination country, and mode of transport, prices of arm's length transactions are higher than prices of transactions between related parties.

B. *General Equilibrium*

The final good is non-tradeable, and must be produced domestically using local labor and the intermediate goods aggregate. The final good production function in country i is: $c_i = q_i^{\alpha}(l_i^f)^{1-\alpha}$, where l_i^f is the amount of labor used in the final good sector. The labor force in each country is split in the two sectors, and the share of the labor force working in the intermediate goods sector may either work for local suppliers (serving the domestic and/or the foreign market) or for affiliates of domestic or foreign integrated firms. Labor is immobile, and the following population constraint must hold in each country:

$$L_i = l_i^f + \sum_{j=h,f} (l_{ji}^I + l_{ji}^T) \quad \text{for} \quad i = h, f, \tag{23}$$

where l_{ji}^I is the labor force of country i working in integrated segments of firms from country j and l_{ji}^T is the labor force of country i working for suppliers from

[24]The properties of the pricing rule described here depend on the assumption that suppliers cannot observe their competitors' marginal costs. This is a reasonable assumption to make in the international context, where it may be too costly to monitor a foreign competitor's cost structure. Alternatively, one could remove the assumption of private information on the marginal costs and assume Bertrand competition across suppliers of the same input in different countries, as in Bernard et al. (2003). Under this alternative scenario, the lowest cost supplier would still "win" the market, and would charge a price equal to the minimum between the second-best producer's marginal cost and the unconstrained profit-maximizing price (equal to the closed economy price in (8) corrected for transportation costs and wages). In the symmetric case, this alternative formulation implies the same average prices as (22) by the Revenue Equivalence Theorem. In the asymmetric case, prices could be higher or lower than the ones implied by (22), and the limit pricing rule would weaken the link between trade/FDI liberalization and prices.

country i selling in market j. Since the intermediate goods production function is linear, the labor force segments can be expressed as linear functions of the aggregate quantities q_h, q_f:

$$L_i = \frac{(1-\alpha)p_i}{\alpha w_i} q_i + \sum_{j=h,f} (k^I_{ji} q_j + k^T_{ji} q_j) \quad \text{for} \quad i = h, f \tag{24}$$

where the proportionality factors k^I_{ji}, k^T_{ji} are functions of the wage levels w_h, w_f and of model parameters only:

$$k^I_{ji} = \frac{p_j^\eta}{w_i} \int_{\mathcal{B}^I_j} (m_j x_j)^{1-\eta} \phi_j(x_j) \psi(\mathbf{z}) dx_j d\mathbf{z} \tag{25}$$

$$k^T_{ji} = p_j^\eta \int_{\mathcal{B}^T_{ji}} \tau_o z_i p_{ji}(z_i)^{-\eta} \phi_j(x_j) \psi(\mathbf{z}) dx_j d\mathbf{z}. \tag{26}$$

Taking the wages as given, (24) is a linear system of two independent equations in two unknowns, whose solution delivers the equilibrium values of q_h, q_f as functions of w_h, w_f and model parameters. Market clearing conditions allow one to solve for the equilibrium wages. In each country, total income (labor income plus the suppliers' profits) must be equal to total expenditure in the final good:

$$r_i c_i = L_i w_i + \Pi_i \quad \text{for} \quad i = h, f \tag{27}$$

where r_i is the zero-profit price of the final good in country i: $r_i = \alpha^{-\alpha}(1-\alpha)^{(\alpha-1)} p_i^\alpha w_i^{1-\alpha}$, and Π_i denotes the total profits of suppliers from country i:

$$\Pi_i = \int_0^\infty \sum_{j=h,f} [p_{ji}(z_i) - \tau_o w_i z_i] \left(\frac{p_{ji}(z_i)}{p_j}\right)^{-\eta} q_j A_{ji}(p_{ji}(z_i)) \psi_i(z_i) dz_i. \tag{28}$$

The market clearing condition (27) is a system of two equations in the two unknowns w_h, w_f. Normalizing $w_f = 1$, by Walras' law we can solve for the equilibrium relative wage w_h by equating the excess demand to zero in country h: $ED_h = L_h w_h + \Pi_h - r_h c_h = 0$.

Figure 3 plots the excess demand correspondence in the Home country for the symmetric case, for some arbitrary values of the parameters.[25] Due to the discrete choice of where to locate integrated production, the correspondence has two kinks at $w_h/w_f = 1/\tau_v$ and $w_h/w_f = \tau_v$. The excess demand associated with each of these two points is an interval, and if the correspondence crosses the zero line at one of these points the corresponding relative wage does not necessarily clear the market. This happens because $w_h/w_f = 1/\tau_v$ and $w_h/w_f = \tau_v$ are the levels of

[25] $\eta = 2.5$, $\alpha = 0.25$, $\vartheta = 3$, $\tau_o = 1.3$ and $\tau_v = 1.5$.

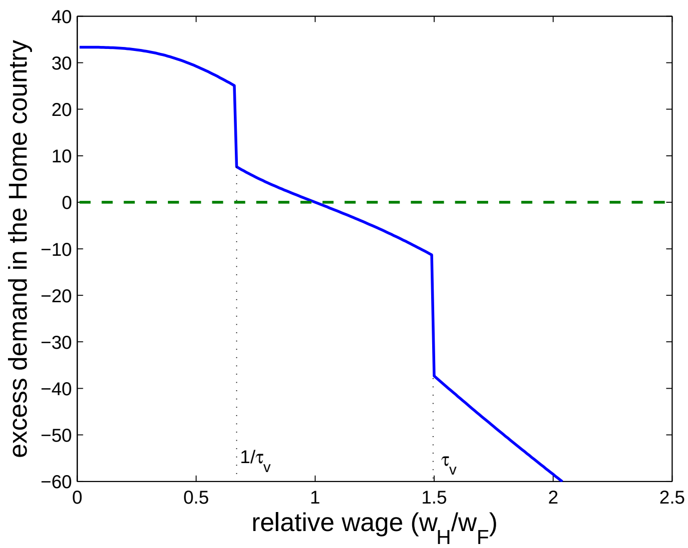

Figure 3. Excess demand correspondence in the Home country.

the relative wage such that firms change the location of their integrated activities. For $w_h \in (0, w_f/\tau_v)$, firms from both countries integrate in the Home country. For $w_h \in (w_f/\tau_v, \tau_v w_f)$, firms from both countries integrate domestically. Otherwise, for $w_h \in (\tau_v w_f, \infty)$, firms from both countries integrate in the Foreign country.

When $w_h = \tau_v w_f$, firms from h are indifferent about where to integrate production, whether domestically or abroad, while firms from f integrate domestically. The figure shows that if firms from h choose to integrate only in one country when they are indifferent, the equilibrium wage may not clear the market. Then firms from h integrate in both countries, and the allocation of labor in the integrated sectors in each country is the variable that clears the market. Similarly, when $w_h = w_f/\tau_v$, firms from f integrate in both countries, while firms from h integrate domestically. At these critical points, the excess demand correspondence is non-smooth because the cost structure of the firms suddenly changes: Figure 4 shows that the unit labor demand of integrated sectors of Home and Foreign firms has a kink at the point where firms switch from domestic to foreign integration and viceversa.

The following proposition establishes the existence of the equilibrium for the two-country case.

PROPOSITION 1: *Provided that the pricing rules $\{p_{ij}(z_j)\}$ are continuous in z_j, for $i, j = h, f$, there exists a relative wage w_h/w_f such that $ED_h = 0$.*

Proof: See Appendix.

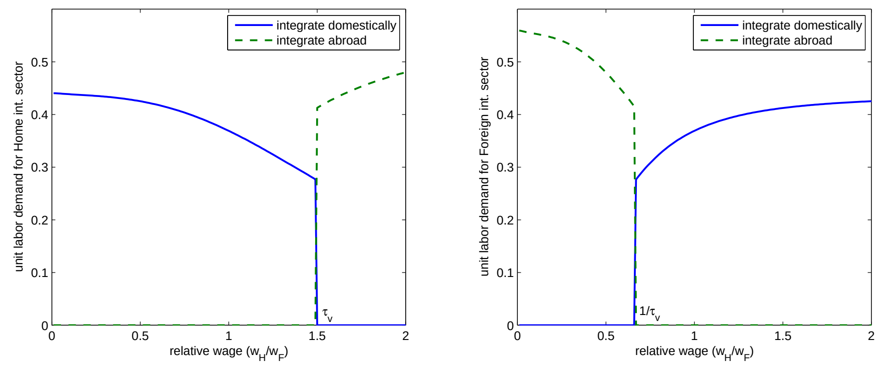

FIGURE 4. LABOR DEMAND IN THE INTEGRATED SECTORS.

III. Productivity, Differentiation, and Intrafirm Trade

The model developed above has predictions for the determinants of intrafirm trade flows. In this section I confront these predictions with existing and novel empirical evidence. I start by showing that the model is consistent with the predictions of contract theory-based models linking the mean and variance of the productivity distribution to the share of intrafirm imports in a country. Then I show a novel prediction of the model, linking the degree of input differentiation with the intrafirm share of imports, and I provide empirical evidence in support of this prediction.

The model-generated variables of interest are volumes of arm's length imports and volumes of FDI imports. Let IM_i denote arm's length imports as a fraction of GDP for country i, and $VFDI_i$ denote vertical FDI (or intrafirm imports) as a fraction of GDP for country i:

$$IM_i = \frac{\left[(p_{ij}^T)^{1-\eta}(p_i)^\eta q_i\right]}{r_i c_i} \quad \text{for} \quad i \neq j \tag{29}$$

$$VFDI_i = \begin{cases} \dfrac{(p_i^I)^{1-\eta}(p_i)^\eta q_i}{r_i c_i} & ; \text{if } w_i > \tau w_j \\ \dfrac{(1-\gamma_i)(p_i^I)^{1-\eta}(p_i)^\eta q_i}{r_i c_i} & ; \text{if } w_i = \tau w_j \\ 0 & ; \text{otherwise} \end{cases} \tag{30}$$

where γ_i is the percentage of labor force hired domestically in the integrated sector when final good producers from country i integrate production both domestically and abroad.

Most papers that study theoretically or empirically the determinants of intrafirm trade focus on the share of total imports that happens intrafirm. To compare the prediction s of my model with other results in the literature, it is useful to define this share as: $SH_i = \frac{VFDI_i}{IM_i + VFDI_i}$.

A. Productivity and Intrafirm Trade: Theory and Evidence

Antràs and Helpman (2004) nest the choice of integration versus outsourcing into a model of trade with heterogeneous firms and increasing returns to scale. By assuming that the fixed cost of vertical integration is higher than the fixed cost of outsourcing, they obtain selection by productivity: more (less) productive firms decide to integrate (outsource) input production. On aggregate, firms from industries or countries with higher mean productivity exhibit more intrafirm sourcing than outsourcing. Not only the mean but also the variance of the productivity distribution matters in Antràs and Helpman (2004)'s analysis: a higher productivity dispersion implies that there is more mass in the extremes of the distribution, and that there are relatively more very productive firms that find optimal to integrate. As a result, the intrafirm share of total imports is increasing in the mean and in the variance of the productivity distribution.

Several recent papers have tested these predictions using different datasets. Yeaple (2006) and Nunn and Trefler (2008) build sector-level measures of productivity dispersion and find support for the hypothesis that the intrafirm share of U.S. imports is larger in industries characterized by more dispersion.[26] Using French firm-level data, Corcos et al. (2011) find that more productive firms are more likely to be engaged in intrafirm transactions. Finally, Kohler and Smolka (2012) test the Antràs and Helpman (2004)'s prediction s using a sample of Spanish firms: their analysis is unique in that the data allows them to observe all four possible sourcing strategies (domestic integration, vertical FDI, domestic and foreign outsourcing). They find that more productive firms are more likely to engage in vertical integration than outsourcing.

The model I developed in this paper is consistent with the predictions of Antràs and Helpman (2004) and with this body of empirical evidence: the intrafirm share of imports is increasing in the mean and in the variance of the productivity distribution. Figures 5-7 illustrate these predictions.[27] Consistently with recent quantitative Ricardian models of trade, like Eaton and Kortum (2002) and Alvarez and Lucas (2007), I assume that unit costs are Weibull-distributed, hence productivity (the inverse of the unit cost) follows a Fréchet distribution.[28] When

[26]Nunn and Trefler (2008)'s analysis is based on the U.S. Census data on related party trade, while Yeaple (2006) uses data from the Bureau of Economic Analysis (B.E.A.) on the operations of U.S. Multinational Corporations.

[27]The comparative statics shown in this section are numerical results. Since the model does not admit a closed-form solution for the pricing rules, it is not possible to show analytically how the variables IM_i and FDI_i depend on the exogenous parameters of the model.

[28]The unit costs of integrated production in country i are distributed according to $\Phi_i(x_i) = 1 -$

the location parameters λ, μ increase and the shape parameter ϑ decrease, both the mean and the variance of the productivity distribution increase. Consistently with the predictions of Antràs and Helpman (2004) and with the empirical evidence, the intrafirm share of imports of country i is increasing in λ_i, μ_i and decreasing in ϑ.

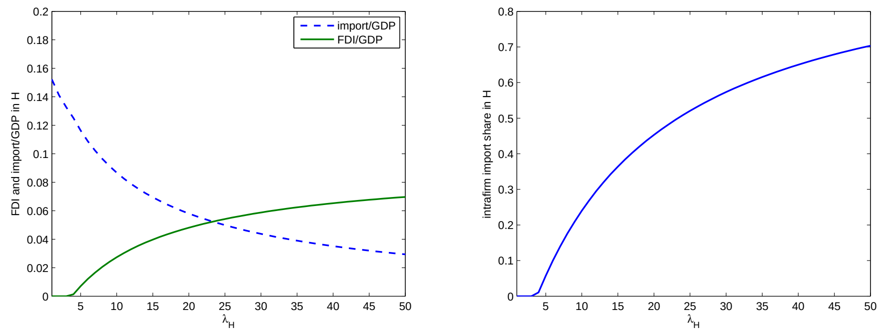

Figure 5. Arm's length imports, intrafirm imports, and intrafirm share as functions of λ.

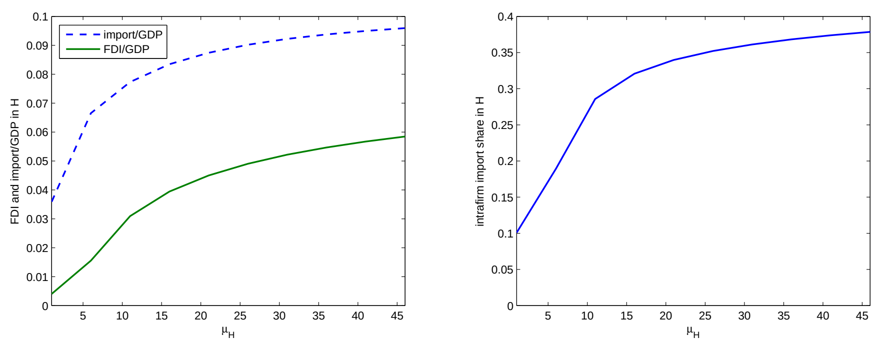

Figure 6. Arm's length imports, intrafirm imports, and intrafirm share as functions of μ.

The intuition is as follows. When λ_H increases, the productivity distribution of final good producers in country H shifts to the right and integrated production becomes relatively cheaper than outsourcing. For this reason, volumes of arm's length imports decrease and volumes of intrafirm imports increase, so that the intrafirm share increases. When μ_H increases, the productivity distribution of suppliers in country H shifts to the right. This has two effects: on the one hand,

$\lambda_i e^{-\lambda_i x_i^{\vartheta}}$, and suppliers' unit costs are distributed according to $\Psi_i(z_i) = 1 - \mu_i e^{-\mu_i z_i^{\vartheta}}$, for $x_i, z_i \geq 0$, $\lambda_i, \mu_i > 0$, $\vartheta > 1$. Figures 5-7 are drawn by computing the model at the calibrated values of the parameters (see Section IV.A).

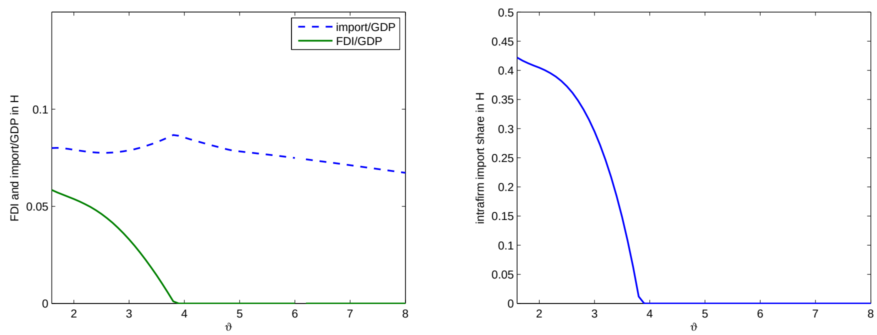

FIGURE 7. ARM'S LENGTH IMPORTS, INTRAFIRM IMPORTS, AND INTRAFIRM SHARE AS FUNCTIONS OF ϑ.

the fact that suppliers from H are more productive pushes towards an increase in domestic outsourcing. On the other hand, the increase in μ_H is accompanied by an increase in the relative wage w_H, which creates incentive for both types of foreign sourcing. In the calibrated economy this second effect dominates and both intrafirm and arm's length imports increase. Since intrafirm imports are not affected by mark-ups, they increase with μ_H at a faster rate than arm's length imports and as a result the intrafirm import share increases. When ϑ decreases, the productivity distribution becomes more disperse, comparative advantage is stronger and both types of trade flows increase. Higher prices and profit margins associated to arm's length flows have the effect that arm's length import volumes increase less than intrafirm import volumes, and the intrafirm share of import increases.[29]

These numerical results show that – albeit via different economic channels – the same predictions of Antràs and Helpman (2004) hold in my model.

B. *Differentiation and Intrafirm Trade: Theory and Evidence*

Variable mark-ups on arm's length prices have interesting implications for the effect of the elasticity of substitution on the volume of intrafirm trade. A lower elasticity of substitution is associated with more differentiation across goods, more market power on the side of the suppliers, and higher arm's length prices. These higher prices in turn provide a stronger incentive for final good producers to integrate input production. As a result, the model predicts that the intrafirm share of imports is higher the lower the elasticity of substitution.[30]

[29]These comparative statics exercises could also be performed by varying λ and ϑ so that the mean of the distribution changes while the variance stays constant, or viceversa having the mean fixed and the variance change. Obviously the results are qualitatively the same as the ones displayed in Figures 5-7.

[30]In a recent working paper, Antràs and Chor (2012) develop a model that studies integration versus outsourcing decisions along the value chain. Their paper is based on the contract theory approach to study intrafirm trade decisions, and delivers a similar prediction via a different economic channel: in

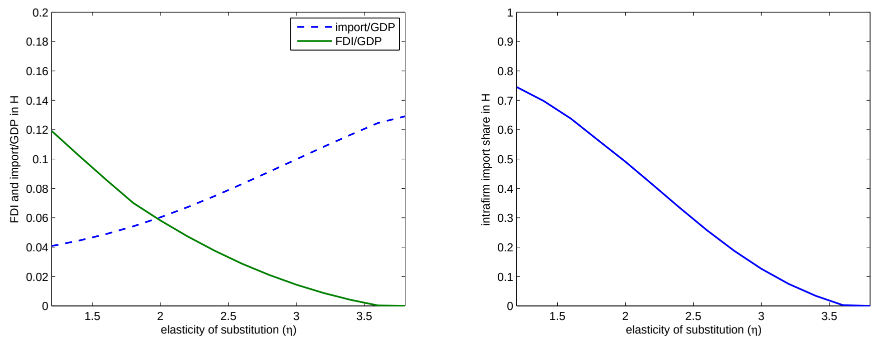

Figure 8. Arm's length imports, intrafirm imports, and intrafirm share as functions of η.

Figure 8 illustrates this result. A lower elasticity of substitution is associated to larger (smaller) volumes of intrafirm (arm's length) imports, and hence to a higher intrafirm share of imports.

In order to show the empirical validity of this prediction, I look at intrafirm import shares across different sectors of the U.S. economy. Figure 9 shows evidence from the raw data. The variable on the horizontal axis is a measure of within-sector product differentiation (described next), while the variable on the vertical axis is the volume of imports of U.S. parents from their foreign affiliates as a share of U.S. total imports. Each point in the plot is a sector-year observation. Data are plotted for 29 manufacturing industries and 16 years, from 1983 to 1998.[31] The positive correlation in the plot shows that intrafirm imports are the prevailing sourcing channel in sectors where goods are highly differentiated.

The measure of differentiation I construct is based on Broda and Weinstein's estimates of sector-level elasticities of substitution. Broda and Weinstein (2006) estimate elasticities of substitution from price and volume data on U.S. consumption of imported goods. By using data at the 10-digit Harmonized System, they estimate how much demand shifts between 10-digit varieties when relative prices vary, within each 3-digit SITC sector. In order to overcome the measurement error

industries with low elasticity of demand, profits are larger and integration is appealing in order to avoid significant rent dissipation.

[31]Data on intrafirm imports at the sector level are available on the B.E.A. website until recent years (see Bureau of Economic Analysis, 1983-2005). However, after 1998 changes in the industry classification brought the number of industries down to 22. Robustness of the results to the more recent sample period is discussed in the online Appendix. Antràs and Chor (2012) use more disaggregated data from the U.S. Census Bureau (see U.S. Census Bureau, 2002-2011). The reason why I did not use those data in my empirical analysis is the following. Intrafirm imports in my model are imports of parents from their foreign affiliates. The U.S. Census Bureau data reports total bilateral intrafirm trade flows by industry, including imports of U.S. parents from their foreign affiliates but also imports of U.S.-based affiliates from their foreign parents and imports of U.S.-based affiliates from other foreign affiliates of multinational corporations. My model has no predictions for these other kinds of intrafirm flows, so I preferred not to use these data to keep the empirical analysis as close as possible to the assumptions and results of the model.

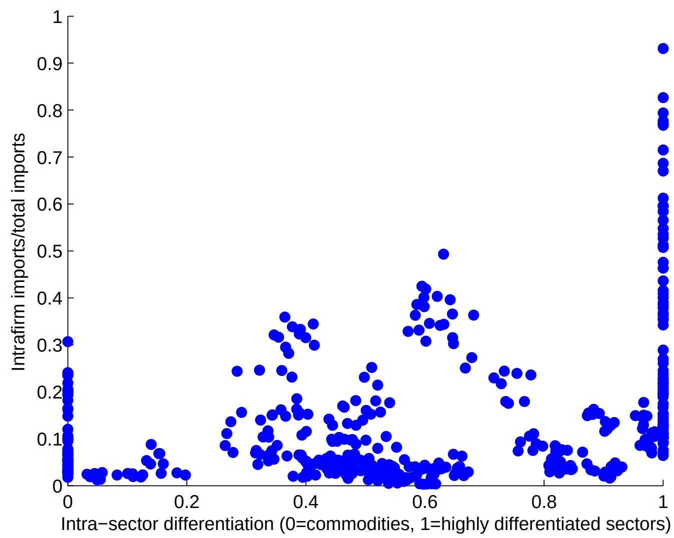

Figure 9. Intrafirm import share and product differentiation by sector.

problems associated with using their estimates as data for the empirical analysis, I build a dummy variable $D^{\eta}_{SITC_s}$ that takes value 1 (0) when the estimated sectoral elasticity is below (above) the median value of 2.54.[32]

Data on intrafirm import at the sector level are publicly available on the B.E.A. website for 34 manufacturing industries. After excluding natural resources and sectors for which data are missing, I am left with 29 industries. In order to aggregate $D^{\eta}_{SITC_s}$ at the B.E.A. classification level, each value of $D^{\eta}_{SITC_s}$ is weighted by the import share of the 3-digit SITC sector in the corresponding B.E.A. sector:

$$D^{\eta}_{BEA_s} = \sum_{SITC_s \in BEA_S} \frac{IMPORT_{SITC_s}}{IMPORT_{BEA_S}} \times D^{\eta}_{SITC_s}. \tag{31}$$

The resulting variable assigns values between zero and one to each of the 29 B.E.A. industries. A low value of $D^{\eta}_{BEA_s}$ is associated to commodities, while a high value is associated to highly differentiated sectors. To test the strength of the relationship linking intrafirm trade and product differentiation, I run regressions of the form:

$$ln(SH_{st}) = \beta_0 + \beta_1 D^{\eta}_s + \beta_2 \ln(K_{st}/L_{st}) + X'_{st}\gamma + \delta_t + \varepsilon_{st} \tag{32}$$

[32]The online Appendix shows that the results are robust to using the estimated elasticities instead of the dummy defined here.

where SH_{st} denotes imports of U.S. parents from their foreign affiliates as a fraction of total U.S. imports, by sector and year, D_s^η is the differentiation variable defined in (31) in sector s and year t, K_{st}/L_{st} is the capital/labor ratio in sector s and year t, X_{st} is a vector of additional controls, δ_t denotes year fixed effects, and ε_{st} is an orthogonal error term.

Even if the model developed in Section II does not feature capital, controlling for capital intensity of a sector is in order. Antràs (2003) has shown that the intrafirm share of U.S. imports is higher the higher the capital intensity of the exporting industry. Without this control, one could argue that the measure of differentiation I construct is in reality carrying information about capital intensity, as more differentiated sectors might be more capital-intensive. The list of regressors contained in X_{st} follows the specification in Antràs (2003): human capital intensity, value added as a share of total sales, advertising intensity, R&D intensity, and capital stock per establishment. Data sources and details on the construction of these variables can be found in the online Appendix.

Table 1— Regressions of intrafirm import shares on a measure of differentiation and other controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$D^\eta_{BEA_s}$	1.015 (0.454)**		1.466 (0.440)***	1.345 (0.496)***	1.438 (0.495)***	1.476 (0.492)***	1.400 (0.489)***	1.457 (0.532)***
$\ln(K/L)_s$		0.778 (0.293)***	0.989 (0.257)***	0.883 (0.276)***	0.786 (0.277)***	0.766 (0.303)**	0.794 (0.311)**	0.329 (0.583)
$\ln(H/L)_s$				0.277 (0.297)	0.422 (0.294)	0.414 (0.301)	0.383 (0.272)	0.265 (0.240)
$\ln(VA/sales)_s$					-0.726 (0.751)	-0.752 (0.768)	-0.729 (0.689)	-1.470 (0.712)**
$\ln(ADV/sales)_s$						0.047 (0.152)	0.062 (0.138)	0.066 (0.151)
$\ln(R\&D/sales)_s$							0.188 (0.144)	0.216 (0.149)
$\ln(K/N)_s$								0.503 (0.472)
No. of obs.	458	458	458	458	458	458	458	145
adj. R^2	0.098	0.181	0.371	0.39	0.403	0.405	0.426	0.507

Note: Standard errors (indicated in parentheses) are clustered by sector, to control for potential within-sector correlation of the error term.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

The results are displayed in Table 1, which presents the coefficients adding one regressor at a time. The first column of Table 1 is the econometric equivalent of Figure 9. As predicted by the model, the coefficient on the differentiation dummy D_s^η is positive and significant at the 5% significance level. The second column presents the results of the baseline regression in Antràs (2003), regressing intrafirm import shares on capital intensity. Columns (3)-(8) report the results of the regression with both differentiation and capital intensity, plus the different controls listed above.[33] The magnitude and significance of the coefficient on D_s^η are robust to adding the battery of controls listed above. According to the most inclusive specification in column (8), a change in D_s^η from 0 to 1 is associated to a 329 percent increase in the intrafirm share of imports, *i.e.* highly differentiated goods ($D_s^\eta = 1$) are associated to a share of intrafirm imports on average 3.29 times higher than the one associated to commodities ($D_s^\eta = 0$).

The plot in Figure 9 and the regressions in Table 1 are based on the differentiation variable defined by equation (31). An alternative way of classifying sectors according to their degree of differentiation follows the classification in Rauch (1999). The online Appendix shows that the results are robust to constructing the differentiation measure using a similar dummy based on Rauch (1999)'s classification.

The results contained in this Section provide external validation of the mechanism that is unique of this paper: sectors characterized by more differentiation across goods and more market power on the side of the suppliers tend to be associated with a larger share of intrafirm transactions.

IV. The Gains from Multinational Production

The general equilibrium structure of the model makes it amenable to welfare analysis. I calibrate the model to match aggregate volumes of trade and multinational activity for the U.S.. With the calibrated model, I quantify the gains arising from vertical FDI/intrafirm trade. Counterfactual experiments show how the gains depend on the degree of competition in the market and on the extent of barriers to foreign investment. Details about the computation of the model are relegated to the online Appendix.

A. *Calibration*

I start by describing the calibration of the parameters of the model. I identify the Home country with the United States, and the Foreign country with an aggregate of countries that I denote as the "rest of the world" (henceforth, ROW).[34] Calibration of the bilateral model requires to assign values to the parameters

[33]The value of the coefficient $\hat{\beta}_2$ is different from the one in Antràs (2003) for two reasons: first, this paper uses a more restrictive definition of intrafirm imports: consistently with the model, I include in the construction of $D^\eta_{BEA_s}$ imports of U.S. parents from their foreign affiliates only. Antràs (2003) also includes imports of U.S.-based affiliates from their foreign parents. Second, the sample period in this paper is 1983-1998 compared to the four years in Antràs (2003): 1987, 1989, 1992, and 1994. The number of observations in column (8) is smaller because data on capital per establishment are available only for 1993-1997.

[34]The ROW in the calibration is composed by 155 countries, representing (together with the U.S.) 98.66% of world GDP.

of the production functions, α and η, to the productivity parameters $\lambda_{us}/\lambda_{row}$, μ_{us}/μ_{row}, ϑ, to the relative size parameter L_{us}/L_{row}, and to the iceberg costs τ_o and τ_v.

$(1-\alpha)$ represents the labor share in final good production. As the final good in the model is non-tradeable, Alvarez and Lucas (2007) identify $(1-\alpha)$ with the fraction of employment in the non-tradeable sector, and compute α using data on agriculture, mining and manufacturing (defined as tradeables). Following calculations from different data sources, they choose $\alpha = 0.25$ as a reasonable value for industrialized countries.

Convergence of the integrals defining aggregate prices for Weibull-distributed draws requires $\vartheta > \eta - 1$. I choose $\vartheta = 3$, close to the estimates in Bernard et al. (2003).[35] The elasticity of substitution η is a measure of product differentiation and market power, and has a large effect on the computation of the welfare gains. In the baseline calibration I choose a value of $\eta = 2.5$, equal to the median value in Broda and Weinstein (2006)'s SITC 3-digit estimates. The value $\eta = 2.5$ implies mark-ups ranging from 67% to zero. Average mark-ups depend on productivity parameters and trade barriers. I also present the results for a lower value of η to show how the gains from multinational production depend on this aspect of competition in the market. The chosen value of $\eta = 1.22$ is the lower bound of B.E.A.-level estimates of the elasticity of substitution from Broda and Weinstein (2006).[36]

I calibrate jointly the remaining parameters to match a set of relevant moments in the data. I identify the ratio μ_{us}/μ_{row} with the relative average productivity of U.S. firms with respect to ROW firms. Bernard, Jensen and Schott (2009) report that multinational corporations appear to be on average more productive than non-multinational firms. Due to the scarcity of available data to quantify this productivity differential, I assume that final good producers (the potential multinational corporations) draw their productivity from the same distribution as local firms, and hence have the same relative average productivity: $\lambda_{us}/\lambda_{row} = \mu_{us}/\mu_{row}$. As in Alvarez and Lucas (2007), L_{us}/L_{row} represents labor in efficiency units in the US relative to the ROW.[37] τ_o and τ_v are average iceberg costs of trade and FDI. I choose the four parameters μ_{us}/μ_{row}, L_{us}/L_{row}, τ_o, and τ_v to match the intrafirm share of imports of U.S. multinational corporations from their foreign affiliates, U.S. total imports as a fraction of GDP, the U.S. share of world GDP, and U.S. GDP per worker relative to an average of the ROW.

[35] I use the value of ϑ from Bernard et al. (2003) because is the only model featuring imperfect competition where this parameter is estimated. Estimates of the perfectly competitive model in Eaton and Kortum (2002) deliver values of $\vartheta \in [3.6, 8.3]$. Corrections to the Eaton and Kortum (2002)'s estimation procedure performed by Simonovska and Waugh (2011) settle the value of ϑ towards the lower end of the interval.

[36] The estimates in Broda and Weinstein (2006) are at the 3-digit SITC level. I aggregate them at the B.E.A. level using import weights and constructing a logarithmic geometric average to mitigate the effect of outliers.

[37] As explained in Alvarez and Lucas (2007), the size parameter L cannot be measured directly. In the model, size and productivity parameters are directly linked to GDP and GDP per worker in each country, variables that I include in the set of moments to be matched.

All matched data are for the year 2004. The intrafirm share of imports of U.S. parents from their foreign affiliates was 13.27% in 2004, and almost constant over the last decade.[38] Notice that the share I construct is smaller than the ones reported by other papers: Antràs (2003) and Bernard, Jensen and Schott (2009) report an intrafirm share of imports of about 40 percent. This discrepancy is due to the fact that I consider only the portion of intrafirm imports that the model explains: imports of U.S. parents from their foreign affiliates.[39] The other moments are constructed using data from the Center of International Data at U.C. Davis and from the World Bank's World Development Indicators (WDI).[40] In 2004, U.S. imports were 12.9 percent of U.S. GDP. The share of U.S. GDP in the world GDP was 28 percent, and U.S. GDP per worker relative to an average of the ROW was 2.03.[41]

Let $PAR = [\mu_{us}/\mu_{row} \;\; L_{us}/L_{row} \;\; \tau_o \;\; \tau_v]$. The vector of calibrated parameters is a vector $\widehat{PAR} = \arg\min_{PAR} \sum [mom - \widehat{mom}(PAR)]^2$, where mom is the vector of moments from the data, and $\widehat{mom}(PAR)$ is the vector of moments generated by the model as function of the vector of parameters PAR. The baseline calibrated model implies $L_{us}/L_{row} = 0.125$, $\tau_o = 1.1$, $\tau_v = 1.96$, $\mu_{us}/\mu_{row} = \lambda_{us}/\lambda_{row} = 12$.[42]

Given the nonlinearities introduced in the model through the shape of the pricing functions and the discrete choice of location, it is hard to talk about identification of the parameters. The calibrated parameters must be determined jointly, as each or them affects all the matched moments. This said, sensitivity analysis reveals that the computed intrafirm share of imports is extremely sensitive to the choice of the value of the iceberg cost τ_v. To be able to match the share of intrafirm import from the data, the calibrated value of τ_v implies that producing one unit of input abroad almost doubles its unit costs. I believe that the necessity of this high cost to match the data depends on the fact that the model does not consider other types of transaction costs, fixed costs of entering the foreign market, or legal restrictions to intrafirm activities. These frictions – that the model does not consider explicitly – are reflected in the results of the calibration.

Conversely, the calibrated trade iceberg cost τ_o is low. Particularly, it is significantly lower than Eaton and Kortum (2002)'s estimates and is close to the lower bound of the estimates used by Alvarez and Lucas (2007). This depends on the fact that in my model, unlike in theirs, firms have also the option of integrat-

[38] The aggregate intrafirm share of imports of U.S. parents from their foreign affiliates is constructed using the same data used in Section III.B.

[39] I am excluding imports of U.S.-located affiliates from foreign parents (because the model does not support bilateral intrafirm transactions, more common when talking about horizontal FDI), and transactions between affiliates.

[40] See Feenstra (1972-2006), World Bank (1960-2011).

[41] I compute the average GDP per worker in the ROW as a weighted average of each country's GDP per worker, with the shares of US imports from that country as weights.

[42] Productivity differences implied by the calibrated value of μ_{us}/μ_{row} amount to say that - on average - U.S. producers are 2.29 times more productive than ROW producers.

ing production domestically, and this option is very attractive in the calibrated economy.[43] For this reason, a higher trade cost would generate substitution away from arm's length imports into domestic integrated production, and a low iceberg trade cost is necessary to match the share of import that we observe in the data.

B. Results

I compute the welfare gains that the theory implies by comparing the calibrated economy and a counterfactual world without possibility of vertical multinational production. I compute the gain in consumption per capita as:

$$\text{welfare gain} = \left(\frac{\text{consumption p.c. in calibrated model}}{\text{consumption p.c. in model without vertical FDI}} - 1 \right) \times 100$$

where the term in the denominator is obtained by computing the model with the calibrated parameters, but shutting down the possibility of vertical FDI (foreign integration).

Gains from vertical FDI for the U.S. economy arise from two sources: first, U.S. final good producers benefit from combining their higher productivity with the lower wages of the ROW, and second, consumers gain because prices decrease (more integration possibilities result in increased competition among suppliers and lower prices). On the other hand, in equilibrium, the ROW economy integrates only domestically. Hence the welfare gains for the ROW consumers come from a general equilibrium channel: the upward pressure on relative wages determined by the entry of U.S. firms.

The results are shown in the first column of Table 2. The calibrated economy implies a gain in U.S. consumption per capita of 0.23 percent with respect to a world economy with no possibility of vertical FDI. The magnitude of this number should not be underestimated. What this exercise says is that the change from a world without vertical FDI to a world where vertical FDI amounts to 13 percent of total imports generates an increase of 0.23 percent in consumption per capita. We are opening the economy only to match a volume of FDI that is relatively small in the data. Nonetheless, the induced welfare gain is sizeable.[44]

Welfare gains in the model arise from two sources: first, vertical FDI allows final good producers to use their more productive technologies but to pay lower wages, and is hence equivalent to an expansion of the production possibility frontier of the economy. Second, the possibility of integration reduces the mark-ups charged on traded intermediates and hence prices. The last line in Table 2 shows the percentage reduction in average (sales-weighted) U.S. domestic mark-ups that is

[43] In the calibrated economy $w_{us} = \tau_v w_{row}$ and 90 percent of U.S. integrated production is done domestically.

[44] Ramondo and Rodríguez-Clare (2010) compute the welfare gains from horizontal multinational production. They find welfare gains of about 3 percent for the U.S., much larger than in this exercise, and consistent with the observation that horizontal FDI flows are larger than vertical ones.

Table 2— Welfare gains from vertical FDI.

	baseline calibration ($\eta = 2.5$, $\tau_v = 1.96$)	FDI reform ($\eta = 2.5$, $\tau_v' = 1.48$)	higher market power ($\eta' = 1.22$, $\tau_v = 1.96$)
U.S. welfare gains (percent)	0.23	5.92	2.36
ROW welfare gains (percent)	0.03	0.36	0.02
U.S. intrafirm import share (percent)	13.87	82.32	62.66
percentage change in average U.S. domestic mark-up	- 0.22	-1.68	-4.1

driven by opening the economy to vertical FDI. The reduction in mark-ups is minimal in the baseline calibration, due to the small extent of vertical FDI that this exercise allows by construction.

The second column of Table 2 reports the same calculations performed in a world where the unit cost of integrated production abroad drops of 50 percent. I refer to this experiment as to an "FDI reform", like an institutional reduction of barriers to FDI, that in the model takes the form of a decrease in the parameter τ_v.[45] As expected, a drop in the cost of FDI increases the welfare gains: compared to the baseline calibration, consumption levels are significantly higher in both countries. The drop in τ_v generates a shift in the world allocation of production: all integrated activity of U.S. firms happens abroad, with an associated welfare gain in consumption per capita of 5.92 percent. Notice also the sensitivity of the computed intrafirm import share to changes in τ_v: a 50 percent drop in τ_v increases the share of intrafirm imports almost six-fold. The larger extent of foreign investment in ROW countries also increases ROW's relative wage more and induces the 0.36 percent gain in consumption per capita attributable to the entry of foreign firms. The drop in mark-ups is higher with respect to the baseline scenario: more attractive vertical FDI possibilities induce more competition across suppliers.

The third column of Table 2 reports the results of the same computations performed with a lower value of the elasticity of substitution: $\eta' = 1.22$. This version of the calibration describes a scenario where the degree of differentiation across intermediates is higher. As a result, competition is lower and suppliers have more market power. This scenario is empirically relevant as most intrafirm trade happens in sectors where the degree of differentiation is high (see Section III). Table 2 shows that the gains from opening to intrafirm trade are higher

[45] An example of such liberalization could be any legislative action *"to increase intellectual property protection and to provide the legal conditions for the participation of transnational corporations in the privatization of state industries"* (see UNCTAD, 1993). It is true that most episodes of liberalization involve measures designed to facilitate both trade and FDI. Nonetheless, there have been examples of liberalizations explicitly targeting FDI, like for example the inclusion of FDI-related issues in the Uruguay Round agreement and the Multilateral Agreement on Investment in the OECD (see UNCTAD, 1996).

than in the baseline calibration, because the possibility of integration reduces more significantly the suppliers' market power and boosts competition in the economy.[46] Notice that under this scenario there is a higher intrafirm import share than in the baseline case: more market power on the side of the suppliers gives more incentive to integrate. U.S. gains from vertical FDI are one order of magnitude larger than in the baseline scenario. The different value of η also implies that suppliers reduce their mark-ups more than in the previous cases.

The numbers reported in Table 2 quantify the current gains from vertical FDI for the U.S. economy. A related question is to ask how large these gains are compared to extreme scenarios like complete autarky and complete integration. Table 3 shows the results. Using the calibrated model, I compute consumption per capita in the U.S. in the autarky case, in which barriers to trade and FDI are prohibitively high and there is no foreign sourcing. I allow for domestic integration in the autarky economy, so that the gains from domestic integration are not reflected in the calculations. To ease the comparison, consumption under autarky is normalized to one. I then compute the welfare gains (increases in consumption per capita) under three different scenarios: the calibrated economy, an economy that is perfectly open to trade but is closed to FDI, and a frictionless economy where both trade and FDI happen at no cost.

Table 3— U.S. Welfare gains: from autarky to costless trade and FDI.

	U.S. welfare gains
autarky ($\tau_o, \tau_v \to \infty$)	1
calibrated economy ($\tau_o = 1.1, \tau_v = 1.96$)	1.063
costless trade and no FDI ($\tau_o = 1, \tau_v \to \infty$)	1.078
costless trade and costless FDI ($\tau_o = \tau_v = 1$)	1.232

The second row in the table presents the welfare gains for the calibrated economy compared to autarky. The current gains for the U.S. are sizeable, at 6.3 percent of consumption per capita. The third row reports the hypothetical gains arising from opening the economy to free trade ($\tau_o = 1$), but not allowing foreign integration ($\tau_v \to \infty$). The calibrated economy is similar to the economy with free trade and no FDI, with welfare gains of 6.3 percent and 7.8 percent, respectively. The small amount of FDI in the calibrated economy generates little gain, and the gains for the U.S. are actually lower than in the economy with free trade but no FDI at all.

[46]This result is consistent with Rauch (1999), who finds that the impact of trade barriers is lower on commodities that on differentiated goods. Accordingly, in my model the effect of the removal of barriers to FDI is larger, the larger the degree of differentiation across goods.

The gains from trade that the model generates are lower than estimates from other papers: Alvarez and Lucas (2007) estimate a gain of 10 percent, while Eaton and Kortum (2002) obtain a gain of 17 percent. The calibration exercise in this paper is bilateral, and less ambitious in scope compared to the ones in Eaton and Kortum (2002) and Alvarez and Lucas (2007). Moreover, the attractiveness of domestic integration and the changes in suppliers' profits induced by trade are features that are not present in those models. Nonetheless, the computed welfare gains for the U.S. economy are of a similar order of magnitude as in previous quantitative analyses, fact that raises one's confidence on the model's ability to generate reasonable gains from FDI.

The third row of the table presents the hypothetical gains arising in a frictionless economy ($\tau_o = \tau_v = 1$). Opening to costless FDI implies an additional increase in consumption per capita for the U.S. of about 15 percent, for a total gain with respect to autarky of 23 percent. Rodríguez-Clare (2007) estimates that the combined gains from trade and diffusion of ideas across countries can reach about 200 percent of consumption, depending on the relative importance of a country's research intensity. Given their large role in total world research, the gains for the U.S. are much lower than this upper bound, reaching about 10 percent of consumption. Compared to Rodríguez-Clare's analysis, my model concentrates the attention on a very specific channel of diffusion – vertical FDI –, nonetheless the computed total gains for the U.S. are larger. This feature depends on the different source of the gains I consider. In Rodríguez-Clare (2007), countries profit from openness because they can get access to ideas generated in other countries, so the gains are limited for a country, like the U.S., that accounts for the majority of world research. In my model, the gains arise from the match of "good ideas" (high productivity draws) with low labor costs, so even a country that accounts for the totality of world research can benefit from opening.

The frictionless economy is an ideal theoretical construct, and as such the associated welfare gains should be interpreted as an upper bound to possible welfare improvements. The results in Table 3 show that potential gains from vertical FDI are large, and that the actual economy is still far away from reaping the full potential of this aspect of globalization.

V. Conclusions

This paper proposes a new general equilibrium framework aimed at explaining the decisions of firms to fragment their production processes across national borders, both in terms of location and organizational structure, through the choice of outsourcing *versus* integrating input production. Firms' optimal sourcing strategies are the outcome of a market equilibrium, where choices are driven by technology heterogeneity and by the implications of imperfect competition on prices. Multinational corporations arise endogenously when firms decide to integrate production in foreign countries.

The possibility of integration induces downward pressure on arm's length prices, establishing a link between trade and FDI liberalization and equilibrium prices. The model has predictions for the dependence of intrafirm trade flows on the economy's fundamentals: the intrafirm share of imports is positively correlated to the mean and the variance of the firms' productivity distribution, and to the degree of differentiation across goods in the economy. These predictions find support in the data, providing external validation to the theory proposed.

I calibrate the model to match aggregate U.S. data and compute the implied gains from vertical multinational production and intrafirm trade. The welfare gains for the U.S. are currently about 0.23 percent of consumption per capita, and the model shows that further reductions of the costs of FDI would increase them substantially.

Extensions of the model should be devoted to a more flexible characterization of the FDI technology, able to reproduce multilateral patterns that we observe in the data. Nonetheless, I believe the analysis conducted here is a useful starting point to get a deeper understanding of the role of technology and market structure in shaping firms' sourcing decisions, and of the welfare consequences of this aspect of globalization.

Appendix: Existence of the Equilibrium

This Appendix contains the proof of Proposition 1. To show the existence of the equilibrium, it is sufficient to show that the excess demand correspondence ED_h is continuous and that $\exists \underline{w}_h, \bar{w}_h$ such that $ED_h(\underline{w}_h) > 0$ and $ED_h(\bar{w}_h) < 0$.

It is clear from the construction of the model that – provided that the pricing rules are continuous – the excess demand is differentiable (hence continuous) almost everywhere. The only two points where the excess demand correspondence is not differentiable are $w_h = \tau_v$ and $w_h = 1/\tau_v$. At these wage levels, firms switch the location of production, and labor demand is not differentiable. The labor demand for integrated segments of firms from country H is:

$$l_h^I \begin{cases} = \dfrac{p_h^\eta q_h}{w_h} \displaystyle\int_{\mathcal{B}_h^I} (w_h x_h)^{1-\eta} \phi_h(x_h) \psi(\mathbf{z}) dx_h d\mathbf{z} & ; \text{if } w_h < \tau_v \\ \in \left[\dfrac{p_h^\eta q_h}{w_h} \displaystyle\int_{\mathcal{B}_h^I} (w_h x_h)^{1-\eta} \phi_h(x_h) \psi(\mathbf{z}) dx_h d\mathbf{z}, \dfrac{p_h^\eta q_h}{w_s} \displaystyle\int_{\mathcal{B}_h^I} (\tau_v w_f x_h)^{1-\eta} \phi_h(x_h) \psi(\mathbf{z}) dx_h d\mathbf{z} \right] & ; \text{if } w_h = \tau_v \\ = \dfrac{p_h^\eta q_h}{w_s} \displaystyle\int_{\mathcal{B}_h^I} (\tau_v w_f x_h)^{1-\eta} \phi_h(x_h) \psi(\mathbf{z}) dx_h d\mathbf{z} & ; \text{if } w_h > \tau_v \end{cases}$$

where l_h^I takes values in a closed interval for $w_h = \tau_v$, and $\lim\limits_{w_h^- \to \tau_v} l_h^I \in l_h^I(\tau_v)$, $\lim\limits_{w_h^+ \to \tau_v} l_h^I \in l_h^I(\tau_v)$. The labor demand for integrated segments of firms from country F is constructed in the same way, with the non-differentiability at $w_h = 1/\tau_v$. This is sufficient to ensure continuity of ED_h at $w_h = \tau_v$ (and similarly at $w_h = 1/\tau_v$).

On the second point, it is sufficient to compute the limits of the excess demand correspondence for $w_h \to 0$ and $w_h \to \infty$. Using the definition of profits, the first order conditions in the final good market and the population constraint, the excess demand correspondence can be written as:

$$\begin{aligned} ED(w_h) &= L_h w_h + (p^T_{hh})^{1-\eta} p^\eta_h q_h + (p^T_{fh})^{1-\eta} p^\eta_f q_f - w_h(l^T_{hh} + l^T_{fh}) - r_h c_h \\ &= \left[\frac{(p^I_{hh})^{1-\eta} + (p^T_{hh})^{1-\eta}}{p_h^{1-\eta}} - 1\right] p_h q_h + \left[\frac{(p^I_{fh})^{1-\eta} + (p^T_{fh})^{1-\eta}}{p_f^{1-\eta}}\right] p_f q_f \end{aligned}$$

where, given the definition of the price indexes:

$$\left[\frac{(p^I_{hh})^{1-\eta} + (p^T_{hh})^{1-\eta}}{p_h^{1-\eta}} - 1\right] \in (-1, 0) \quad \text{and} \quad \left[\frac{(p^I_{fh})^{1-\eta} + (p^T_{fh})^{1-\eta}}{p_f^{1-\eta}}\right] \in (0, 1).$$

Since prices are increasing in wages:

$$\lim_{w_h \to 0} p_h = \lim_{w_h \to 0} p_f = 0 \quad \text{and} \quad \lim_{w_h \to \infty} p_h = \lim_{w_h \to \infty} p_f = \infty.$$

It remains to determine the limits of q_h, q_f. The term q_h can be rewritten as:

$$\begin{aligned} q_h &= \frac{\alpha w_h \left[(1-\alpha)p_f + \alpha w_f k_{ff}\right] L_h - \alpha^2 w_h w_f k_{fh} L_f}{\left[(1-\alpha)p_h + \alpha w_h k_{hh}\right]\left[(1-\alpha)p_f + \alpha w_f k_{ff}\right] - \alpha^2 w_h w_f k_{fh} k_{hf}} \\ &= \left\{\frac{(1-\alpha)p_h}{\alpha w_h L_h} + \frac{k_{hh}}{L_h} - \frac{\alpha w_f k_{fh} k_{hf}}{\left[(1-\alpha)p_f + \alpha w_f k_{ff}\right] L_h}\right\}^{-1} \ldots \\ &\quad \ldots - \left\{\frac{\left[(1-\alpha)p_h/w_h + \alpha k_{hh}\right]\left[(1-\alpha)p_f + \alpha w_f k_{ff}\right]}{\alpha^2 w_f k_{fh} L_f} - \frac{k_{hf}}{L_f}\right\}^{-1}. \end{aligned}$$

When $w_n \to 0$, the term $\frac{(1-\alpha)p_h}{\alpha w_h L_h} + \frac{k_{hh}}{L_h}$ is a positive constant; the term $\frac{\alpha w_f k_{fh} k_{hf}}{\left[(1-\alpha)p_f + \alpha w_f k_{ff}\right] L_h}$ is a $\frac{0}{0}$ indeterminacy; the term $\frac{\left[(1-\alpha)p_h/w_h + \alpha k_{hh}\right]}{\alpha^2 w_f k_{fh} L_f}$ is a positive constant, and the term $\left[(1-\alpha)p_f + \alpha w_f k_{ff}\right]$ tends to zero, so $\frac{\left[(1-\alpha)p_h/w_h + \alpha k_{hh}\right]\left[(1-\alpha)p_f + \alpha w_f k_{ff}\right]}{\alpha^2 w_f k_{fh} L_f}$ tends to zero; the term $\frac{k_{hf}}{L_f}$ tends to zero. As a result: $\lim_{w_h \to 0} q_h = -\infty$. Similarly, one can show that $\lim_{w_h \to \infty} q_h = \infty$, $\lim_{w_h \to 0} q_f = \infty$, $\lim_{w_h \to \infty} q_f = -\infty$. It is immediate to conclude that:

$$\lim_{w_h \to 0} ED(w_h) = \quad \text{and} \quad \lim_{w_h \to \infty} ED(w_h) = -\infty.$$

*

References

Alvarez, Fernando, and Robert E. Lucas. 2007. "General Equilibrium Analysis of the Eaton-Kortum Model of International Trade." *Journal of Monetary Economics*, 54(6): 1726–1768.

Antràs, Pol. 2003. "Firms, Contracts, and Trade Structure." *Quarterly Journal of Economics*, 118(4): 1375–1418.

Antràs, Pol, and Davin Chor. 2012. "Organizing the Global Value Chain." Mimeo, Harvard University.

Antràs, Pol, and Elhanan Helpman. 2004. "Global Sourcing." *Journal of Political Economy*, 112(3): 552–580.

Arkolakis, Costas, Natalia Ramondo, Andrès Rodríguez-Clare, and Stephen Yeaple. 2011. "Innovation and Production in the Global Economy." Mimeo, Yale University.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott. 2006. "Transfer Pricing by U.S.-Based Multinational Firms." NBER Working Paper 12493.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott. 2009. "Importers, Exporters, and Multinationals: a Portrait of Firms in the U.S. that Trade Goods." In T. Dunne, J.B. Jensen and M.J. Roberts (eds.)*, Producer Dynamics: New Evidence from Micro Data.* Chicago:University of Chicago Press.

Bernard, Andrew B., Jonathatan Eaton, J. Bradford Jensen, and Samuel Kortum. 2003. "Plants and Productivity in International Trade." *The American Economic Review*, 93(4): 1268–1290.

Broda, Christian, and David E. Weinstein. 2006. "Globalization and the Gains from Variety." *Quarterly Journal of Economics*, 121(2): 541–585.

Bureau of Economic Analysis. 1983-2005. "U.S. Direct Investment Abroad: Operations of U.S. Parent Companies and Their Foreign Affiliates. Comprehensive financial and operating data." U.S. Department of Commerce, Washington, D.C., http://www.bea.gov/international/di1usdop.htm (accessed August 2012).

Campa, José, and Linda S. Goldberg. 1997. "The Evolving External Orientation of Manufacturing Industries: Evidence from Four Countries." Federal Reserve Bank of New York Economic Policy Review, 3(2): 53–81.

Caves, Richard E., and Ralph M. Bradburd. 1988. "The Empirical Determinants of Vertical Integration." *Journal of Economic Behavior and Organization*, 9: 265–279.

Corcos, Gregory, Delphine M. Irac, Giordano Mion, and Thierry Verdier. 2011. "The Determinants of Intrafirm Trade." *The Review of Economics and Statistics*. Forthcoming.

Dvir, Eyal. 2010. "Globalization, Optimal Auctions and Exchange Rate Pass-Through." Mimeo, Boston College.

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

Feenstra, Robert C. 1972-2006. "U.S. Import and Export Data. Center for International Data at U.C. Davis." http://cid.econ.ucdavis.edu/ (accessed May 2011).

Feenstra, Robert C. 1998. "Integration of Trade and Disintegration of Production in the Global Economy." *The Journal of Economic Perspectives*, 12(4): 31–50.

Garetto, Stefania. 2012. "Firms' Heterogeneity and Incomplete Pass-Through." Mimeo, Boston University.

Grossman, Gene M., and Elhanan Helpman. 2002. "Integration Versus Outsourcing in Industry Equilibrium." *Quarterly Journal of Economics*, 117(1): 85–120.

Grossman, Gene M., and Elhanan Helpman. 2004. "Managerial Incentives and the International Organization of Production." *Journal of International Economics*, 63: 237–262.

Grossman, Gene M., and Esteban Rossi-Hansberg. 2008. "Trading Tasks: A Simple Theory of Offshoring." *The American Economic Review*, 98(5): 1978–1997.

Hanson, Gordon H., Raymond J. Mataloni, and Matthew J. Slaughter. 2001. "Expansion Strategies of U.S. Multinational Firms." In D. Rodrik and S.M. Collins (eds.), *Brookings Trade Forum 2001.*

Hanson, Gordon H., Raymond J. Mataloni, and Matthew J. Slaughter. 2005. "Vertical Production Networks in Multinational Firms." *Review of Economics and Statistics*, 87(4): 664–678.

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

Helpman, Elhanan. 1985. "Multinational Corporations and Trade Structure." *Review of Economic Studies*, 52(3): 443–457.

Helpman, Elhanan. 2006. "Trade, FDI, and the Organization of Firms." NBER Working Paper 12091.

Helpman, Elhanan, and Paul Krugman. 1985. *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy.* Cambridge, MA:MIT Press.

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

Hummels, David, Dana Rapoport, and Kei-Mu Yi. 1998. "Vertical Specialization and the Changing Nature of World Trade." Federal Reserve Bank of New York Economic Policy Review, 4(2): 79–99.

Hummels, David, Jun Ishii, and Kei-Mu Yi. 2001. "The Nature and Growth of Vertical Specialization in World Trade." *Journal of International Economics*, 54: 75–96.

Irarrazabal, Alfonso, Andreas Moxnes, and Luca D. Opromolla. 2010. "The Margins of Multinational Production and the Role of Intrafirm Trade." *Journal of Political Economy.* Forthcoming.

Kohler, Wilhelm K., and Marcel Smolka. 2012. "Global Sourcing Decisions and Firm Productivity: Evidence from Spanish Firm-Level Data." In Robert M. Stern (eds.)*, Quantitative Analysis of Newly Evolving Patterns of International Trade: Fragmentation, Offshoring of Activities, and Vertical Intra-Industry Trade.* World Scientific Studies in International Economics. 139–193.

Markusen, James R. 1984. "Multinationals, Multi-Plant Economies, and the Gains from Trade." *Journal of International Economics*, 16: 205–226.

McAfee, Preston, and John McMillan. 1987. "Auctions and Bidding." *Journal of Economic Literature*, 25(2): 699–738.

Melitz, Marc J., and Gianmarco I.P. Ottaviano. 2008. "Market Size, Trade, and Productivity." *Review of Economic Studies*, 75(1): 295–316.

Neiman, Brent. 2011. "A State-Dependent Model of Intermediate Goods Pricing." *Journal of International Economics*, 85(1): 1–13.

Nunn, Nathan, and Daniel Trefler. 2008. "The Boundaries of the Multinational Firm: An Empirical Analysis." In E. Helpman, D. Marin, and T. Verdier (eds.)*, The Organization of Firms in a Global Economy.* 55–83. Cambridge, MA:Harvard University Press.

Ramondo, Natalia, and Andrès Rodríguez-Clare. 2010. "Trade, Multinational Production, and the Gains from Openness." Mimeo, Arizona State University.

Rauch, James E. 1999. "Networks Versus Markets in International Trade." *Journal of International Economics*, 48: 7–35.

Rodríguez-Clare, Andrès. 2007. "Trade, Diffusion, and the Gains from Openness." Mimeo, University of California, Berkeley.

Rossi-Hansberg, Esteban, and Mark L. J. Wright. 2007. "Establishment Size Dynamics in the Aggregate Economy." *The American Economic Review*, 97(5): 1639–1666.

Simonovska, Ina, and Michael Waugh. 2011. "The Elasticity of Trade: Estimates and Evidence." Mimeo, University of California, Davis.

Tirole, Jean. 1988. *The Theory of Industrial Organization.* Cambridge, MA:MIT Press.

UNCTAD. 1993. *World Investment Report; Transnational Corporations and Integrated International Production.* New York:United Nations.

UNCTAD. 1996. *World Investment Report; Investment, Trade, and International Policy Arrangements.* New York:United Nations.

U.S. Census Bureau. 2002-2011. "Related Party Trade Database." Washington, D.C., http://sasweb.ssd.census.gov/relatedparty/ (accessed August 2012).

Waterman, David, and Andrew A. Weiss. 1996. "The Effects of Vertical Integration Between Cable Television Systems and Pay Cable Networks." *Journal of Econometrics*, 72: 357–395.

World Bank. 1960-2011. "World Development Indicators." http://data.worldbank.org/data-catalog/world-development-indicators (accessed August 2012).

Yeaple, Stephen R. 2006. "Offshoring, Foreign Direct Investment, and the Structure of U.S. Trade." *Journal of the European Economic Association, Papers and Proceedings of the Twentieth Annual Congress of the European Economic Association*, 4(2-3): 602–611.

Yeats, Alexander J. 2001. "Just How Big is Global Production Sharing?" In Sven W. Arndt and Henryk Kierzkowski (eds.), *Fragmentation: New Production Patterns in the World Economy.* Oxford:Oxford University Press.